

HYDROLYSIS OF MANNOSE-1-PHOSPHO-6-MANNOSE LINKAGE TO PHOSPHO-6-MANNOSE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Application Serial No. 61/246,847, filed on September 29, 2009. The disclosure of the prior application is incorporated by reference in its entirety.

TECHNICAL FIELD

The invention relates to methods of hydrolyzing mannose-1-phospho-6-mannose linkages on glycoproteins, and more particularly, to using a mannosidase to hydrolyze mannose-1-phospho-6-mannose linkages to uncap the phospho-6-mannose residues on the glycoprotein.

BACKGROUND

High performance expression systems are required to produce most biopharmaceuticals (e.g., recombinant proteins) currently under development. The biological activity of many of these biopharmaceuticals is dependent on their post-translational modification (e.g., phosphorylation or glycosylation). A yeast-based expression system combines the ease of genetic manipulation and fermentation of a microbial organism with the capability to secrete and to modify proteins. However, recombinant glycoproteins produced in yeast cells exhibit mainly heterogeneous high-mannose and hyper-mannose glycan structures, which can be detrimental to protein function, downstream processing, and subsequent therapeutic use, particularly where glycosylation plays a biologically significant role.

U.S. Application Serial No. 12/062,469 is incorporated by reference in its entirety.

SUMMARY

The present invention is based, at least in part, on the discovery of a mannosidase that is capable of hydrolyzing mannose-1-phospho-6-mannose linkages on glycoproteins. As such, the mannosidase can be used to obtain glycoproteins containing uncapped

terminal mannose-6-phosphate residues. *In vitro* and *in vivo* methods of obtaining such glycoproteins are described herein. Genetically engineered cells can be used in the methods to produce target molecules having uncapped terminal mannose-6-phosphate residues.

In one aspect, this document features a method for uncapping a mannose-6-phosphate residue on an oligosaccharide. The method includes providing the oligosaccharide having a mannose-1-phospho-6-mannose linkage; and contacting the oligosaccharide with a mannosidase capable of hydrolyzing the mannose-1-phospho-6-mannose linkage to phospho-6-mannose. The contacting step can be performed using a purified mannosidase, a recombinant mannosidase, a cell lysate containing the recombinant mannosidase, or a fungal cell containing the recombinant mannosidase. The mannosidase can include a targeting sequence. The oligosaccharide can be attached to a protein (e.g., a human protein expressed in a fungal organism).

In another aspect, this document features a method of producing a target protein having terminal phospho-6-mannose residues. The method includes providing a fungal cell genetically engineered to include a nucleic acid encoding a mannosidase, the mannosidase capable of hydrolyzing a mannose-1-phospho-6-mannose linkage to phospho-6-mannose; and introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein comprising the terminal phospho-6-mannose residues.

This document also features a method of producing a target protein having terminal phospho-6-mannose residues in a fungal organism. The method includes providing a fungal cell genetically engineered to include a nucleic acid encoding a mannosidase capable of hydrolyzing a mannose-1-phospho-6-mannose linkage to phospho-6-mannose, wherein the fungal cell further includes a nucleic acid encoding a target protein; and isolating the target protein having the terminal phospho-6-mannose residues. The fungal cell further can include a nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation and/or can be genetically engineered to be deficient in OCH1 activity.

This document also features an isolated fungal cell genetically engineered to produce glycoproteins comprising terminal phospho-6-mannose residues. The fungal cell

includes a nucleic acid encoding a mannosidase, wherein expression of the mannosidase in the fungal cell produces glycoproteins comprising the terminal phospho-6-mannose residues. The fungal cell further can include a nucleic acid encoding a target glycoprotein protein.

In another aspect, this document features a substantially pure culture of *Yarrowia lipolytica*, *Pichia pastoris*, *Hansenula polymorpha*, *Arxula adeninivorans*, *Pichia methanolica*, *Oogataea minuta*, or *Aspergillus niger* cells, a substantial number of which are genetically engineered to produce glycoproteins comprising a terminal phospho-6-mannose residue, the cells comprising a nucleic acid encoding a mannosidase capable of hydrolyzing a mannose-1-phospho-6-mannose linkage to phospho-6-mannose.

In any of the embodiments described herein, the fungal organism can be *Yarrowia lipolytica* or *Arxula adeninivorans*. The fungal organism can be a methylotrophic yeast such as *Pichia pastoris*, *Pichia methanolica*, *Oogataea minuta*, or *Hansenula polymorpha*. The fungal organism can be a filamentous fungus (e.g., a filamentous fungus selected from the group consisting of *Aspergillus caesiellus*, *Aspergillus candidus*, *Aspergillus carneus*, *Aspergillus clavatus*, *Aspergillus deflectus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus glaucus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus ochraceus*, *Aspergillus oryzae*, *Aspergillus parasiticus*, *Aspergillus penicilloides*, *Aspergillus restrictus*, *Aspergillus sojae*, *Aspergillus sydowi*, *Aspergillus tamari*, *Aspergillus terreus*, *Aspergillus ustus*, and *Aspergillus versicolor*).

In any of embodiments described herein, the protein can be a pathogen protein, a lysosomal protein, a growth factor, a cytokine, a chemokine, an antibody or antigen-binding fragment thereof, or a fusion protein. The lysosomal protein can be a lysosomal enzyme (e.g., a lysosomal enzyme associated with a lysosomal storage disorder (LSD) such as acid alpha glucosidase or alpha galactosidase). The LSD can be Fabry's disease, mucopolysaccharidosis I, Farber disease, Gaucher disease, GM1-gangliosidosis, Tay-Sachs disease, Sandhoff disease, GM2 activator disease, Krabbe disease, metachromatic leukodystrophy, Niemann-Pick disease, Scheie disease, Hunter disease, Sanfilippo disease, Morquio disease, Maroteaux-Lamy disease, hyaluronidase deficiency, aspartylglucosaminuria, fucosidosis, mannosidosis, Schindler disease, sialidosis type 1, Pompe disease, Pycnodysostosis, ceroid lipofuscinosis, cholesterol ester storage disease,

Wolman disease, Multiple sulfatase deficiency, galactosialidosis, mucopolipidosis, cystinosis, sialic acid storage disorder, chylomicron retention disease with Marinesco-Sjögren syndrome, Hermansky-Pudlak syndrome, Chediak-Higashi syndrome, Danon disease, or Geleophysic dysplasia. For example, the LSD can be Pompe disease or Fabry's disease.

In any of the embodiments described herein, for the mannosidase, the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.

In any of the embodiments described herein, the mannosidase can include an amino acid sequence having at least 90% identity (e.g., at least 95% or 98% identity) to the amino acid sequence set forth in residues 1 to 774 of SEQ ID NO:50 or to the amino acid sequence set forth in SEQ ID NO:50.

In any of the embodiments described herein, the mannosidase can include an amino acid sequence having (i) a GVGXXGXGG motif, where X is Gly, Ala, Ser, Thr, or Cys; (ii) a VRXE motif, where X is any amino acid other than Pro; (iii) an X₁YQGX₂ motif, where X₁ is Leu, Ile, Val, Ala, Phe, Tyr or Met, and X₂ is Thr, Ser, or Asn; or (iv) a GDXGN motif, where X can be any amino acid other than Pro.

In any of the embodiments described herein, the mannosidase can be a *C. cellulans*, *Streptomyces coelicolor*, or *Streptomyces lividans* mannosidase.

In any of the embodiments described herein, the fungal cell further can include a nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation (e.g., a MNN4 polypeptide such as *Yarrowia lipolytica*, *S. cerevisiae*, *Ogataea minuta*, *Pichia pastoris*, or *C. albicans* polypeptide) and/or can be genetically engineered to be deficient in OCH1 activity. For example, the polypeptide capable of promoting mannosyl phosphorylation can be a *P. pastoris* PNO1 polypeptide.

In any of the embodiments described herein, the mannosidase can include a secretion signal and/or a targeting signal to target the mannosidase to an intracellular compartment. The target protein and the mannosidase can be co-secreted.

This document also features an isolated glycoprotein that includes terminal phospho-6-mannose residues, wherein the protein is produced by the methods described herein.

In yet another aspect, this document features a composition that includes a glycoprotein, wherein at least 47% of the N-glycans on the glycoprotein have terminal phospho-6-mannose residues. For example, at least 50%, 75%, 80%, 85%, or 90% of the N-glycans on the glycoprotein can have terminal phospho-6-mannose residues.

This document also features an isolated nucleic acid that includes a nucleotide sequence set forth in SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14, or a nucleotide sequence that is at least 90% identical to SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:20. This document also features a vector that includes a promoter operably linked to such a nucleic acid, wherein the nucleic acid encodes a mannosidase. The nucleic acid further can include a secretion signal or targeting sequence to target the mannosidase to an intracellular compartment.

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 belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the exemplary methods and materials are described below. All publications, patent applications, patents, Genbank® Accession Nos, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present application, including definitions, will control. The materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings⁷

FIG. 1 is a schematic of the pYLTmAX and pYLTmAXMnn4 constructs.

FIG. 2 is a series of electroferograms depicting sugar analysis of MTLY60Δoch1(1 wild type copy of Mnn4), MTLY60Δoch1+Hp4dMnn4 (1WT +1 extra

copy of Mnn4) and MTLY60 Δ och1+Hp4dMnn4+TEFMnn4. P represents the monophosphorylated peak, PP represents the diphosphorylated peak, and Man8 represents the Man₈GlcNAc₂ peak.

FIG. 3 is a schematic of mammalian and yeast glycan phosphorylation pathways. The mammalian glycan phosphorylation pathway involves addition of a phospho-GlcNAc catalyzed by GlcNAc-phosphotransferase to Man₈GlcNAc₂ glycans followed by decapping of the GlcNAc to expose the phosphate by an uncovering enzyme. In contrast, yeast glycan phosphorylation involves addition of a phospho-mannose to Man₈GlcNAc₂ glycans, but no endogenous enzyme is present to uncap the mannose to expose the phosphate.

FIG. 4 is a series of electroferograms depicting N-glycans derived from strain MTLY60 Δ och1+Hp4dMnn4+TEFMnn4 treated for different time frames (7 hrs, 8 hrs, or overnight (ON)) with supernatants from *C. cellulans* medium.

FIG. 5 is a series of electroferograms depicting N-glycans derived from an MNN4 overexpressing strain treated with *C. cellulans* supernatant (SN) with and without phosphatase (CIP) incubation.

FIG. 6 is a graph of the absorbance units (mAU) of elution fractions at the indicated MW. Each elution fraction contained ~500 μ l.

FIG. 7 is a representation of a SDS-polyacrylamide gel after electrophoresis of elution fractions from silica-based gel filtration (250 μ l of each fraction was DOC/TCA precipitated). The boxed bands were cut out for peptide mass fingerprinting and de novo sequencing using tandem mass spectrometry (MS/MS).

FIG. 8A is the nucleotide sequence (SEQ ID NO:6) encoding CcMan1 (i.e., mannosidase candidate 1 from *C. Cellulans*) (on contig 1003), which was identified in the MS/MS de novo sequencing. FIG. 8B is the amino acid sequence (SEQ ID NO:7) of CcMan 1, including the signal sequence (in bold). The predicted molecular weight of the CcMan 1 polypeptide without the signal sequence is 92.6 kDa.

FIG. 9A is the nucleotide sequence (SEQ ID NO:8) encoding CcMan2 (on contig 774) and FIG. 9B is the amino acid sequence of CcMan2 (SEQ ID NO:9) with signal sequence (in bold). The predicted molecular weight of the CcMan2 polypeptide without the signal sequence is 121.6 kDa.

FIG. 10A is the nucleotide sequence (SEQ ID NO:10) encoding CcMan3 (on contig 774) and FIG. 10 B is the amino acid sequence of CcMan3 (SEQ ID NO:11) with signal sequence (in bold). The predicted molecular weight of the CcMan3 polypeptide without the signal sequence is 116 kDa.

FIG. 11A is the nucleotide sequence (SEQ ID NO:12) encoding CcMan4 (on contig 1237) and FIG. 11B is the amino acid sequence of CcMan4 (SEQ ID NO:13) with signal sequence (in bold). The predicted molecular weight of the CcMan4 polypeptide without the signal sequence is 184 kDa.

FIG. 12A is the nucleotide sequence (SEQ ID NO:14) encoding CcMan5 (on contig 896). FIG. 12B is the amino acid sequence of CcMan5 with signal sequence (in bold) (SEQ ID NO:15) and FIG. 12C is the amino acid sequence of CcMan5 without signal sequence (SEQ ID NO:50). The predicted molecular weight of the CcMan5 polypeptide without the signal sequence is 173 kDa.

FIG. 13 contains examples of expression plasmids for the expression of CcMan1-5 in the periplasm of *E. coli* (pET25-Man), as secreted proteins in *Yarrowia lipolytica* (pYLP_{Sec}CcMan1-5), as proteins targeted to the secretory pathway of *Yarrowia lipolytica*, tagged to the N-terminus (pYLP_{Nt}CcMan1-5) or tagged to the C-terminus (pYLP_{Ct}CcMan1-5).

FIG. 14 is the nucleotide sequence of CcMan1 that has been codon optimized for expression in *E. coli* (SEQ ID NO:16).

FIG. 15 is the nucleotide sequence of CcMan2 that has been codon optimized for expression *E. coli* (SEQ ID NO:17).

FIG. 16 is the nucleotide sequence of CcMan3 that has been codon optimized for expression in *E. coli* (SEQ ID NO:18).

FIG. 17 is the nucleotide sequence of CcMan4 that has been codon optimized for expression in *E. coli* (SEQ ID NO:19).

FIG. 18 is the nucleotide sequence of CcMan5 that has been codon optimized for expression in *E. coli* (SEQ ID NO:20).

FIG. 19 is a schematic of the pLSAH36 and pLSH36 vectors and the cloning strategy for introducing the *C. cellulans* genes into the vectors.

FIG. 20 is a series of electroferograms depicting analysis of the periplasmic fraction of CcMan4 and CcMan5 expressing *E. coli* cells. Analysis was performed using DNA sequencer-assisted, fluorophore-assisted carbohydrate electrophoresis (DSA-FACE). The first and second panels represent the dextran ladder and the sugars from RNaseB, respectively. The third panel is the untreated Mnn4 sugars with “P” corresponding to the mono mannophosphorylated Man₈GlcNAc₂ peak, “PP” corresponding to the double mannophosphorylated Man₈GlcNAc₂ peak, and “Man8” corresponding to the Man₈GlcNAc₂ peak. Panels 4 to 9 are the results obtained with Mnn4 glycans incubated with the indicated periplasm, with or without a subsequent calf intestinal phosphatase (CIP) digest.

FIG. 21 is a schematic alignment of CcMan4 (1759 AA) and CcMan5 (1650 AA) with Bt3990 (744 AA) and Bt2199 (739 AA) mannosidases described in Zhu *et al.*, *Nat. Chem. Biol.*, 6(2):125-32. Epub 2009 Dec 27 (2010).

FIG. 22 is a series of electroferograms depicting the analysis of the CcMan4 and CcMan5 enzymes obtained from expressing *E. coli* cells. Analysis was performed using DSA-FACE using MNN4 overexpressing strain derived glycans (referred to as MNN4 glycans or MNN4 sugars) as a substrate. The first panel represents the dextran ladder and the second panel represents the untreated Mnn4 sugars. In the third through sixth panels, the sugars were incubated with the CcMan4domain periplasmic fraction not induced, induced overnight at 18°C with IPTG, the CcMan5domain periplasmic fraction not induced, and induced overnight at 18°C with IPTG, respectively. The last panel represents the sugars from RNaseB.

FIG. 23 is a ribbon representation of CcMan5₁₋₇₇₄. CcMan5₁₋₇₇₄ consists of a N-terminal β -sandwich domain (residues 8-271; light gray), an α -helical linker (residues 272-290; black) and a ($\alpha\alpha$)₆ barrel domain (residues 291-771; dark gray). The catalytic Ca²⁺ is shown as a sphere.

FIG. 24 is a ribbon representation of the CcMan5₁₋₇₇₄ protein backbone with side chains lining the substrate binding site shown in stick representation. Carbon, oxygen and nitrogen atoms are colored light gray, gray, and dark gray, respectively. The Ca²⁺ ion and waters W1, W2, W3 and W4 in the catalytic center are shown as spheres.

FIG. 25 is a ribbon representation of the CcMan5₁₋₇₇₄ protein backbone with side chains lining the substrate binding site and the modeled position of mannose-1-phospho-6-mannose (labeled Man-P-Man) shown in stick representation. Carbon, oxygen and nitrogen atoms are coloured light gray, gray, and dark gray, respectively. The Ca²⁺ ion and water molecules W1, W2, W3 and W4 in the catalytic center are shown as spheres (for comparison, the positions of W2 and W3 which will be displaced by the substrate O2 and O3 hydroxyl groups are still shown). Yellow, red and black dashed lines indicate coordination bonds with Ca²⁺, H-bonds with the proposed nucleophilic water (W4) and H-bonds with the -1 site mannose and phosphate, respectively. The -1 site mannose is modeled in its ground state chair conformation. During catalysis, its O2 hydroxyl will occupy a position nearer that seen for W2, in the equatorial coordination plane of Ca²⁺ ion, thereby leading to a distortion to a half-chair conformation in the mannose -1 ring and facilitating in line attack of the nucleophilic water (W4) on the C1 carbon (arrow).

FIG. 26A is the *Y. lipolytica* codon optimized nucleotide sequence encoding α -GalactosidaseA with lip2 pre sequence in bold and the Myc His tag underlined (SEQ ID NO:22). FIG. 26B is the amino acid sequence of α -GalactosidaseA with lip2 pre sequence in bold and the Myc His tag underlined (SEQ ID NO:23).

FIG. 27A is the codon optimized nucleotide sequence of human alpha glucosidase (GAA) with lip2 pre sequence in bold (SEQ ID NO:24). FIG. 27B is the amino acid sequence of human GAA with lip2 pre sequence in bold (SEQ ID NO:25), where the * represents the stop codon.

FIG. 28 is a schematic of a *Y. lipolytica* expression vector used for cloning of huGAA.

FIG. 29 is a series of electroferograms depicting analysis of treatment of huGAA with CcMan5 derived from the periplasmic fraction of *E. coli* cells. Analysis was performed using DSA-FACE.

FIG. 30 is a depiction of the minimal catalytic center of CcMan5. The numbering of equivalent residues in SEQ ID NO:50 is given in parenthesis. 1: Q (Q536); 2: N/D-E/Q (N588-Q589); 3: D/E (D355); 4: R (R405); 5: D/E-X-D/E (D660-X-D662); 6: G-G (G71-G72); and 7: T/S/G (T626).

FIG. 31 is an alignment of the amino acid sequence of CcMan5 (SEQ ID NO:50, the amino acid sequence set forth in SEQ ID NO:15 without the signal peptide) and 10 of its homologs using MUSCLE (Multiple Sequence Comparison by Log-Expectation). NP_630514 *Streptomyces*, SEQ ID NO:26; ZP_02866543 *Clostridium*, SEQ ID NO:27; NP_812442 *Bacteroides*, SEQ ID NO:28; YP_003584502 *Zunongwangia*, SEQ ID NO:29; YP_003120664 *Chitinophaga*, SEQ ID NO:30; AAK22560 *Caulobacter*, SEQ ID NO:31; ACL94075 *Caulobacter*, SEQ ID NO:32; ACT03290 *Paenibacillus*, SEQ ID NO:33; ACU59240 *Chitinophaga*, SEQ ID NO:34; ACU05553 *Pedobacter*, SEQ ID NO:35.

FIG. 32 is an alignment of the amino acid sequence of CcMan5 (SEQ ID NO:50) and 19 of its homologs using MUSCLE. *Streptomyces* NP_630514, SEQ ID NO:26; *Streptomyces* ZP_02866543, SEQ ID NO:36; ZP_06527366 *Streptomyces*, SEQ ID NO:37; YP_003013376 *Paenibacillus*, SEQ ID NO:38; NP_812442 *Bacteroides*, SEQ ID NO:28; ZP_04848482 *Bacteroides*, SEQ ID NO:39; ZP_03677957 *Bacteroides*, SEQ ID NO:40; YP_003584502 *Zunongwangia*, SEQ ID NO:29; ZP_01061975 *Leeuwenhoekiella*, SEQ ID NO:41; ZP_07083984 *Sphingobacterium*, SEQ ID NO:42; YP_003120664 *Chitinophaga*, SEQ ID NO:30; ZP_01885202 *Pedobacter*, SEQ ID NO:43; ZP_02866543 *Clostridium*, SEQ ID NO:27; XP_367221 *Magnaporthe*, SEQ ID NO:44; ZP_07042437 *Bacteroides*, SEQ ID NO:45; ZP_05759807 *Bacteroides*, SEQ ID NO:46; ZP_05287524 *Bacteroides*, SEQ ID NO:47; ZP_06076108 *Bacteroides*, SEQ ID NO:48; YP_001302992 *Parabacteroides*, SEQ ID NO:49.

FIG. 33 contains the structural coordinates of the residues surrounding the active site of CcMan5₁₋₇₇₄.

FIG. 34 contains the protein C alpha atoms and the catalytic Ca²⁺ atoms of the two CcMan5₁₋₇₇₄ molecules in the asymmetric unit in PDB entry 2xsg and describes the overall fold of the protein.

DETAILED DESCRIPTION

In general, this document provides methods and materials for hydrolyzing mannose-1-phospho-6-mannose linkages on glycoproteins to produce target molecules (e.g., target proteins) having uncapped phospho-6-mannose (M6P) residues. The

methods and materials described herein are particularly useful for producing agents for treating patients with lysosomal storage disorders (LSDs), a diverse group of hereditary metabolic disorders characterized by the accumulation of storage products in the lysosomes due to impaired activity of catabolic enzymes involved in their degradation. The build-up of storage products leads to cell dysfunction and progressive clinical manifestations. Deficiencies in catabolic enzymes can be corrected by enzyme replacement therapy (ERT), provided that the administered enzyme can be targeted to the lysosomes of the diseased cells. Lysosomal enzymes typically are glycoproteins that are synthesized in the endoplasmic reticulum (ER), transported via the secretory pathway to the Golgi, and then recruited to the lysosomes. One way in which lysosomal enzymes are delivered to the lysosome is via a cation-dependent (CD) mannose 6-phosphate receptor (MPR). M6P terminal glycans are recognized in the trans-Golgi network (TGN) by two MPRs that mediate the sorting of lysosomal enzymes from the secretory pathway and deliver the enzyme to the lysosome. Using the methods and materials described herein, a microbial based production process can be used to obtain therapeutic proteins with uncapped M6P glycans, which can be delivered to lysosomes by exploiting the same M6P dependent pathway. Thus, the methods and materials described herein are useful for preparing glycoproteins for the treatment of metabolic disorders such as LSDs.

Mannosidases

This document provides isolated nucleic acids encoding mannosidase polypeptides capable of hydrolyzing terminal mannose-1-phospho-6-mannose linkages on oligosaccharides, as well as isolated mannosidases capable of hydrolyzing terminal mannose-1-phospho-6-mannose linkages on oligosaccharides. The terms “nucleic acid” and “polynucleotide” are used interchangeably herein, and refer to both RNA and DNA, including cDNA, genomic DNA, synthetic DNA, and DNA (or RNA) containing nucleic acid analogs. Polynucleotides can have any three-dimensional structure. A nucleic acid can be double-stranded or single-stranded (i.e., a sense strand or an antisense strand). Non-limiting examples of polynucleotides include genes, gene fragments, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, siRNA, micro-RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids,

vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers, as well as nucleic acid analogs.

"Polypeptide" and "protein" are used interchangeably herein and mean any peptide-linked chain of amino acids, regardless of length or post-translational modification. Typically, a polypeptide described herein (e.g., a mannosidase or a target protein having uncapped M6P residues) is isolated when it constitutes at least 60%, by weight, of the total protein in a preparation, e.g., 60% of the total protein in a sample. In some embodiments, a polypeptide described herein consists of at least 75%, at least 90%, or at least 99%, by weight, of the total protein in a preparation.

An "isolated nucleic acid" refers to a nucleic acid that is separated from other nucleic acid molecules that are present in a naturally-occurring genome, including nucleic acids that normally flank one or both sides of the nucleic acid in a naturally-occurring genome (e.g., a yeast genome). The term "isolated" as used herein with respect to nucleic acids also includes any non-naturally-occurring nucleic acid sequence, since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome.

An isolated nucleic acid can be, for example, a DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences as well as DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., any paramyxovirus, retrovirus, lentivirus, adenovirus, or herpes virus), or into the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include an engineered nucleic acid such as a DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not considered an isolated nucleic acid.

The term "exogenous" as used herein with reference to nucleic acid and a particular host cell refers to any nucleic acid that does not occur in (and cannot be

obtained from) that particular cell as found in nature. Thus, a non-naturally-occurring nucleic acid is considered to be exogenous to a host cell once introduced into the host cell. It is important to note that non-naturally-occurring nucleic acids can contain nucleic acid subsequences or fragments of nucleic acid sequences that are found in nature provided that the nucleic acid as a whole does not exist in nature. For example, a nucleic acid molecule containing a genomic DNA sequence within an expression vector is non-naturally-occurring nucleic acid, and thus is exogenous to a host cell once introduced into the host cell, since that nucleic acid molecule as a whole (genomic DNA plus vector DNA) does not exist in nature. Thus, any vector, autonomously replicating plasmid, or virus (e.g., retrovirus, adenovirus, or herpes virus) that as a whole does not exist in nature is considered to be non-naturally-occurring nucleic acid. It follows that genomic DNA fragments produced by PCR or restriction endonuclease treatment as well as cDNAs are considered to be non-naturally-occurring nucleic acid since they exist as separate molecules not found in nature. It also follows that any nucleic acid containing a promoter sequence and polypeptide-encoding sequence (e.g., cDNA or genomic DNA) in an arrangement not found in nature is non-naturally-occurring nucleic acid. A nucleic acid that is naturally-occurring can be exogenous to a particular cell. For example, an entire chromosome isolated from a cell of yeast x is an exogenous nucleic acid with respect to a cell of yeast y once that chromosome is introduced into a cell of yeast.

A nucleic acid encoding a mannosidase can have at least 70% sequence identity (e.g., at least 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity) to a nucleotide sequence set forth in SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14. In some embodiments, nucleic acids described herein can encode mannosidase polypeptides that have at least 70% (e.g., at least 75, 80, 85, 90, 95, 99, or 100 percent) identity to an amino acid sequence set forth in SEQ ID NOs: 7, 9, 11, 13, 15, 50. For example, a nucleic acid can encode a mannosidase having at least 90% (e.g., at least 95 or 98%) identity to the amino acid sequence set forth in SEQ ID NO:15 or SEQ ID NO:50, or a portion thereof. For example, a nucleic acid can encode a mannosidase having at least 90% identity to amino acid residues 1 to 774 of SEQ ID NO:50. The percent identity between a particular amino acid sequence and the amino acid sequence set forth in SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13,

SEQ ID NO:15, or SEQ ID NO:50 is determined as follows. First, the amino acid sequences are aligned using the BLAST 2 Sequences (Bl2seq) program from the stand-alone version of BLASTZ containing BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained from Fish & Richardson's web site (e.g., www.fr.com/blast/) or the U.S. government's National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov). Instructions explaining how to use the Bl2seq program can be found in the readme file accompanying BLASTZ. Bl2seq performs a comparison between two amino acid sequences using the BLASTP algorithm. To compare two amino acid sequences, the options of Bl2seq are set as follows: -i is set to a file containing the first amino acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C:\output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. If the two compared sequences share homology, then the designated output file will present those regions of homology as aligned sequences. If the two compared sequences do not share homology, then the designated output file will not present aligned sequences. Similar procedures can be following for nucleic acid sequences except that blastn is used.

Once aligned, the number of matches is determined by counting the number of positions where an identical amino acid residue is presented in both sequences. The percent identity is determined by dividing the number of matches by the length of the full-length mannosidase polypeptide amino acid sequence followed by multiplying the resulting value by 100. For example, an amino acid sequence that has 700 matches when aligned with the sequence set forth in SEQ ID NO:7 is 77.8 percent identical to the sequence set forth in SEQ ID NO:7 (i.e., $700 \div 900 * 100 = 77.8$).

It is noted that the percent identity value is rounded to the nearest tenth. For example, 78.11, 78.12, 78.13, and 78.14 is rounded down to 78.1, while 78.15, 78.16, 78.17, 78.18, and 78.19 is rounded up to 78.2. It also is noted that the length value will always be an integer.

It will be appreciated that a number of nucleic acids can encode a polypeptide

having a particular amino acid sequence. The degeneracy of the genetic code is well known to the art; i.e., for many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. For example, codons in the coding sequence for a given mannosidase polypeptide can be modified such that optimal expression in a particular species (e.g., bacteria or fungus) is obtained, using appropriate codon bias tables for that species. For example, the nucleic acids set forth in SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, or SEQ ID NO:14 can be codon optimized for *E. coli* expression as set forth in FIGs. 14-18 (see SEQ ID NOs:16-20).

Hybridization also can be used to assess homology between two nucleic acid sequences. A nucleic acid sequence described herein, or a fragment or variant thereof, can be used as a hybridization probe according to standard hybridization techniques. The hybridization of a probe of interest (e.g., a probe containing a portion of a CcMan5 nucleotide sequence) to DNA or RNA from a test source is an indication of the presence of DNA or RNA (e.g., a CcMan5 nucleotide sequence) corresponding to the probe in the test source. Hybridization conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6, 1991. Moderate hybridization conditions are defined as equivalent to hybridization in 2X sodium chloride/sodium citrate (SSC) at 30°C, followed by a wash in 1 X SSC, 0.1% SDS at 50°C. Highly stringent conditions are defined as equivalent to hybridization in 6X sodium chloride/sodium citrate (SSC) at 45°C, followed by a wash in 0.2 X SSC, 0.1% SDS at 65°C.

Mannosidase polypeptides capable of hydrolyzing terminal mannose-1-phospho-6-mannose linkages on oligosaccharides also can be identified based on the three dimensional structure described herein for a portion of a mannosidase from *C. cellulans* (residues 1 to 774 of SEQ ID NO:50, also referred to as CcMan5₁₋₇₇₄). The three dimensional structure can be determined by, for example, X-ray diffraction of a crystal of CcMan5₁₋₇₇₄. Structural coordinates of CcMan5₁₋₇₇₄ (e.g., the coordinates of CcMan5₁₋₇₇₄ deposited with the Protein Data Bank (world wide web at PDB.org under PDB ID No. 2xs), the coordinates set forth in FIG. 33 for the catalytic center of CcMan5, or the coordinates set forth in FIG. 34 for the protein C alpha atoms and the catalytic Ca²⁺ atoms of the two CcMan5₁₋₇₇₄ molecules in the asymmetric unit in PDB entry 2xsg) are

useful for a number of applications, including, but not limited to, the characterization of a three dimensional structure of a mannosidase capable of hydrolyzing terminal mannose-1-phospho-6-mannose linkages on oligosaccharides, as well as the visualization, identification and characterization of regions of the mannosidase that are involved in acceptance of mannose-6-phosphate- α ,1-mannose (hereafter referred to as Man-P-Man) as a substrate, and conferring its ability to hydrolyse Man-P-Man to produce a terminal phospho-6-mannose. "Structural coordinates" are the Cartesian coordinates corresponding to an atom's spatial relationship to other atoms in a molecule or molecular complex. Structural coordinates can be obtained using x-ray crystallography techniques or NMR techniques, or can be derived using molecular replacement analysis or homology modeling. Various software programs allow for the graphical representation of a set of structural coordinates to obtain a three dimensional representation of a molecule or molecular complex. The structural coordinates of the structures described herein can be modified from the original set provided in FIG. 33 or FIG. 34 by mathematical manipulation, such as by inversion or integer additions or subtractions. As such, it is recognized that the structural coordinates of the present invention are relative, and are in no way specifically limited by the actual x, y, z coordinates of FIG. 33 or FIG. 34.

As set forth in Example 8, the structure of CcMan5₁₋₇₇₄ consists of two domains, an N-terminal β -sandwich domain (residues 8 to 271 of SEQ ID NO:50) and a C-terminal ($\alpha\alpha$)6 barrel domain (residues 291 to 771 of SEQ ID NO:50), connected via an α -helical linker (residues 272 to 290 of SEQ ID NO:50). The interface between both domains gives shape to a shallow cavity that harbors a conserved catalytic Ca²⁺ ion and gives shape to the -1 substrate binding site (nomenclature as described by Davies *et al.*, *Biochem. J.* 321:557-9 (1997)) and the catalytic center. Residues 22, 25, 71, 72, 195, 196, 354, 405, 535, 536, 588, 589, 626, 658, 660, and 662 of SEQ ID NO: 50 form the substrate binding site.

The three dimensional structure of CcMan5₁₋₇₇₄ can be characterized in part, or all, using the structural coordinates of PDB ID No. 2xs, or an extract of which that is presented in FIG. 33, comprising the residues surrounding the active site of CcMan5₁₋₇₇₄, or an extract of which that is presented in FIG. 34, comprising the protein C alpha atoms and the catalytic Ca²⁺ atoms of the two CcMan5₁₋₇₇₄ molecules in the asymmetric unit in

PDB entry 2xsg, and describing the overall fold of the protein. For example, the three-dimensional structure of CcMan5₁₋₇₇₄ can be characterized by the structural coordinates of amino acid residues 7 to 771 according to PDB ID No. 2xs, $\pm$ a root mean square deviation from the conserved backbone atoms of said amino acids of not more than 2Å. In some embodiments, the three dimensional structure of CcMan5₁₋₇₇₄ comprises the complete structural coordinates of the amino acids according to PDB ID No. 2xs, $\pm$ a root mean square deviation from the conserved backbone atoms of said amino acids of not more than 2Å (e.g., not more than 1.5 Å, 1.0Å or 0.5Å). As used herein, “root mean square deviation” is the square root of the arithmetic mean of the squares of the deviations from the mean, and is a way of expressing deviation or variation from the structural coordinates described herein. The present disclosure includes all embodiments comprising conservative substitutions of the noted amino acid residues resulting in same structural coordinates within the stated root mean square deviation.

The structural coordinates provided herein can be used to characterize a three dimensional structure of a mannosidase polypeptide. From such a structure, substrate binding sites, for example, can be computationally visualized, identified and characterized based on the surface structure of the molecule, surface charge, steric arrangement, the presence of reactive amino acids, regions of hydrophobicity or hydrophilicity, etc.

In order to use the structural coordinates generated for a structure described herein as set forth in FIGs. 33, FIG. 34, or PDB ID No. 2xs, the relevant coordinates can be displayed as, or converted to, a three dimensional shape or graphical representation. Software programs are commercially available that are capable of generating three dimensional graphical representations of molecules or portions thereof from a set of structural coordinates. Examples of commercially available software programs include, without limitation, the following: GRID (Oxford University, Oxford, UK) ; MCSS (Molecular Simulations, San Diego, CA); AUTODOCK (Scripps Research Institute, La Jolla, CA); DOCK (University of California, San Francisco, CA); Flo99 (Thistlesoft, Morris Township, NJ); Ludi (Molecular Simulations, San Diego, CA); QUANTA (Molecular Simulations, San Diego, CA); Insight (Molecular Simulations, San Diego,

CA); SYBYL (TRIPOS, Inc., St. Louis, MO); and LEAPFROG (TRIPOS, Inc., St. Louis, MO).

The structural coordinates described herein can be used with standard homology modeling techniques in order to determine the unknown three-dimensional structure of a molecule or molecular complex. Homology modeling involves constructing a model of an unknown structure using structural coordinates of one or more related protein molecules, molecular complexes or parts thereof. Homology modeling can be conducted by fitting common or homologous portions of the protein whose three dimensional structure is to be solved to the three dimensional structure of homologous structural elements in the known molecule, specifically using the relevant (i.e., homologous) structural coordinates provided by FIGs. 33 and 34 herein. Homology may be determined using amino acid sequence identity, homologous secondary structure elements, and/or homologous tertiary folds. Homology modeling can include rebuilding part or all of a three dimensional structure with replacement of amino acids (or other components) by those of the related structure to be solved. Accordingly, a three dimensional structure for the unknown molecule may be generated using the three dimensional structure of CcMan5₁₋₇₇₄ described herein, and refined using a number of techniques well known in the art.

Based on the three dimensional structure described herein, substitutions can be made in some of the atoms or side groups of CcMan5₁₋₇₇₄ or other mannosidases in order to improve or modify its selectivity. For example, CcMan5 contains a non-acidic residue at positions 536 and 588, which may allow the mannosidase to tolerate the phosphate linkage to the anomeric oxygen in Man-P-Man substrates. As such, corresponding residues in other mannosidases can be changed to non-acidic residues to increase the ability of the mannosidase to accept Man-P-Man substrates.

Other mannosidase polypeptide candidates suitable for use herein can be identified by analysis of nucleotide and polypeptide sequence alignments. For example, performing a query on a database of nucleotide or polypeptide sequences can identify homologs and/or orthologs of mannosidase polypeptides. Sequence analysis can involve BLAST, Reciprocal BLAST, or PSI-BLAST analysis of nonredundant databases using known mannosidase amino acid sequences. Those polypeptides in the database that have

greater than 40% sequence identity can be identified as candidates for further evaluation for suitability as a mannosidase polypeptide. Amino acid sequence similarity allows for conservative amino acid substitutions, such as substitution of one hydrophobic residue for another or substitution of one polar residue for another. If desired, manual inspection of such candidates can be carried out in order to narrow the number of candidates to be further evaluated. Manual inspection can be performed by selecting those candidates that appear to have domains suspected of being present in mannosidases capable of hydrolyzing terminal mannose-1-phospho-6-mannose linkages, e.g., one or more (e.g., 1, 2, 3, 4 or more) conserved domains or functional regions (e.g., substrate binding cavity). Such domains can include a glycine-rich motif GVGXXGXGG, where X is Gly, Ser, Thr, Val, Ala, Cys or Gln (or other amino acid with a small side chain). This motif is found at residues 69-77 of SEQ ID NO:50. This region makes a loop that provides essential hydrogen bonds to the -1 mannose and phosphate-binding subsite in the active site of the enzyme.

Another example of a conserved motif includes a VRXE motif, where Arg (R) makes a hydrogen bond to the -1 ring and possibly the +1 ring, Glu (E) is in a salt bridge to this R residue, probably shaping this motif; and X is Trp or any of the 20 amino acids except Pro. This motif is found at residues 404-407 of SEQ ID NO:50.

A suitable motif also can be an X₁YQGX₂ motif, where X₁ can be Leu, Ile, Val, Ala, Phe, Tyr or Met, and X₂ can be Thr, Ser or Asn. This motif is found at residues 534-538 of SEQ ID NO:50. The Gln (Q) in this motif is important as an E is present in mannosidases that do not have the ability to hydrolyze terminal mannose-1-phospho-6-mannose linkages on oligosaccharides. The Tyr (Y) in this motif also is thought to be important for the +1 site formation.

In addition, a region defined by residues 22, 25, 71, 72, 195, 196, 354, 405, 535, 536, 588, 589, 626, 658, 660, and 662 of SEQ ID NO:50 forms the substrate binding cavity of CcMan5. As a minimal requirement, G71, G72, D355, R405, Q536, N588, Q589, T626, D660, D662 form the catalytic center, where N588, Q589 and D660 are involved in coordinating the catalytic Ca²⁺ ion, D662 and D660 are involved in activating the nucleophilic water, Q536 stabilizes the anomeric oxygen during the transition state and G71, G72, D355, R405 and T626 are involved in substrate binding at

the -1 site. See FIG. 30 for a representation of this minimal catalytic center. As such, a mannosidase can be selected as a candidate mannosidase capable of hydrolyzing terminal mannose-1-phospho-6-mannose linkages when the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center (e.g., as set forth in FIG. 30) fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.

A conserved motif also can be a GDXGN motif in the N-terminal domain of the protein, where X can be any amino acid except P. This motif is found at residues 21-25 of SEQ ID NO:50 and forms part of the substrate binding pocket of the enzyme as shown in FIG. 24. In particular, the side chains of the D and N line the substrate binding cavity and may shape an alternative subpocket to bind the +1 mannose.

As set forth in Example 14, performing a query on a database of polypeptide sequences identified homologs of CcMan5 in the following organisms: *Streptomyces coelicolor* (GenBank Accession No. NP_630514), *Streptomyces lividans* (GenBank Accession No. ZP_05522540); *Streptomyces lividans* (GenBank Accession No. ZP_06527366); *Clostridium spiroforme* (GenBank Accession No. ZP_02866543), *Bacteroides thetaiotaomicron* (GenBank Accession No. NP_812442), *Zunongwangia profunda* (GenBank Accession No. YP_003584502); *Chitinophaga pinensis* (GenBank Accession No. YP_003120664); *Paenibacillus sp* (GenBank Accession No. YP_003013376); *Bacteroides sp.* (GenBank Accession No. ZP_04848482); *Bacteroides cellulosilyticus* (GenBank Accession No. ZP_03677957); *Leeuwenhoekiella blandensis* (GenBank Accession No. ZP_01061975); *Sphingobacterium spiritivorum* (GenBank Accession No. ZP_07083984); and *Pedobacter sp.* (GenBank Accession No. ZP_01885202). The mannosidases from *Streptomyces coelicolor* and *Streptomyces lividans* are similar (66% sequence identity to the CcMan5 GH92 domain, with 501 identities over 765 aligned residues by BLASTP), not only in the above motifs but also in many the loops of the three dimensional structure.

Isolated nucleic acid molecules encoding mannosidase polypeptides can be produced by standard techniques. For example, polymerase chain reaction (PCR) techniques can be used to obtain an isolated nucleic acid containing a nucleotide sequence described herein. PCR can be used to amplify specific sequences from DNA as

well as RNA, including sequences from total genomic DNA or total cellular RNA. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. Various PCR strategies also are available by which site-specific nucleotide sequence modifications can be introduced into a template nucleic acid. Isolated nucleic acids also can be chemically synthesized, either as a single nucleic acid molecule (e.g., using automated DNA synthesis in the 3' to 5' direction using phosphoramidite technology) or as a series of oligonucleotides. For example, one or more pairs of long oligonucleotides (e.g., >100 nucleotides) can be synthesized that contain the desired sequence, with each pair containing a short segment of complementarity (e.g., about 15 nucleotides) such that a duplex is formed when the oligonucleotide pair is annealed. DNA polymerase is used to extend the oligonucleotides, resulting in a single, double-stranded nucleic acid molecule per oligonucleotide pair, which then can be ligated into a vector. Isolated nucleic acids of the invention also can be obtained by mutagenesis of, e.g., a naturally occurring DNA.

This document also provides (i) biologically active variants and (ii) biologically active fragments or biologically active variants thereof, of the mannosidases described herein. Biologically active variants of mannosidases can contain additions, deletions, or substitutions relative to the sequences set forth in SEQ ID NOs: 7, 9, 11, 13, 15, or 50. Proteins with substitutions will generally have not more than 50 (e.g., not more than one, two, three, four, five, six, seven, eight, nine, ten, 12, 15, 20, 25, 30, 35, 40, or 50) conservative amino acid substitutions. A conservative substitution is the substitution of one amino acid for another with similar characteristics. Conservative substitutions include substitutions within the following groups: valine, alanine and glycine; leucine, valine, and isoleucine; aspartic acid and glutamic acid; asparagine and glutamine; serine, cysteine, and threonine; lysine and arginine; and phenylalanine and tyrosine. The non-polar hydrophobic amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Any substitution of

one member of the above-mentioned polar, basic or acidic groups by another member of the same group can be deemed a conservative substitution. By contrast, a non-conservative substitution is a substitution of one amino acid for another with dissimilar characteristics. The sequence alignments set forth in FIGs. 31 and 32 provide numerous examples of amino acid substitutions that can be made.

Deletion variants can lack one, two, three, four, five, six, seven, eight, nine, ten, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acid segments (of two or more amino acids) or non-contiguous single amino acids.

Additions (addition variants) include fusion proteins containing: (a) a mannosidase set forth in SEQ ID NOs: 7, 9, 11, 13, or 15, or a fragment thereof; and (b) internal or terminal (C or N) irrelevant or heterologous amino acid sequences. In the context of such fusion proteins, the term "heterologous amino acid sequences" refers to an amino acid sequence other than (a). A heterologous sequence can be, for example a sequence used for purification of the recombinant protein (e.g., FLAG, polyhistidine (e.g., hexahistidine), hemagglutinin (HA), glutathione-S-transferase (GST), or maltose-binding protein (MBP)). Heterologous sequences also can be proteins useful as diagnostic or detectable markers, for example, luciferase, green fluorescent protein (GFP), or chloramphenicol acetyl transferase (CAT). In some embodiments, the fusion protein contains a signal sequence from another protein. In certain host cells (e.g., yeast host cells), expression and/or secretion of the target protein can be increased through use of a heterologous signal sequence. In some embodiments, the fusion protein can contain a carrier (e.g., KLH) useful, e.g., in eliciting an immune response for antibody generation) or endoplasmic reticulum or Golgi apparatus retention signals. Heterologous sequences can be of varying length and in some cases can be a longer sequences than the full-length target proteins to which the heterologous sequences are attached.

Biologically active fragments or biologically active variants of the mannosidases have at least 40% (e.g., at least: 50%; 60%; 70%; 75%; 80%; 85%; 90%; 95%; 97%; 98%; 99%; 99.5%, or 100% or even greater) of the mannosidase activity (e.g., uncapping of M6P residues) of the wild-type, full-length, mature protein. For example, a biologically active fragment of a mannosidase can contain residues 1 to 774 of SEQ ID NO:50.

The mannosidases described herein can be used to produce molecules (e.g., target proteins) having uncapped terminal phospho-6-mannose (M6P) residues. The methods can be performed *in vitro* or *in vivo*.

In Vitro Methods of Uncapping M6P Residues

A mannosidase described herein can be recombinantly produced and used *in vitro* to uncapped terminal M6P residues on oligosaccharides. To recombinantly produce a mannosidase, a vector is used that contains a promoter operably linked to nucleic acid encoding a mannosidase polypeptide. As used herein, a "promoter" refers to a DNA sequence that enables a gene to be transcribed. The promoter is recognized by RNA polymerase, which then initiates transcription. Thus, a promoter contains a DNA sequence that is either bound directly by, or is involved in the recruitment, of RNA polymerase. A promoter sequence can also include "enhancer regions," which are one or more regions of DNA that can be bound with proteins (namely, the trans-acting factors, much like a set of transcription factors) to enhance transcription levels of genes (hence the name) in a gene-cluster. The enhancer, while typically at the 5' end of a coding region, can also be separate from a promoter sequence and can be, e.g., within an intronic region of a gene or 3' to the coding region of the gene.

As used herein, "operably linked" means incorporated into a genetic construct (e.g., vector) so that expression control sequences effectively control expression of a coding sequence of interest.

Expression vectors can be introduced into host cells (e.g., by transformation or transfection) for expression of the encoded polypeptide, which then can be purified. Expression systems that can be used for small or large scale production of mannosidase polypeptides include, without limitation, microorganisms such as bacteria (e.g., *E. coli*) transformed with recombinant bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors containing the nucleic acid molecules, and fungal (e.g., *S. cerevisiae*, *Yarrowia lipolytica*, *Arxula adeninivorans*, *Pichia pastoris*, *Hansenula polymorpha*, or *Aspergillus*) transformed with recombinant fungal expression vectors containing the nucleic acid molecules. Useful expression systems also include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the

nucleic acid molecules, and plant cell systems infected with recombinant virus expression vectors (e.g., tobacco mosaic virus) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing the nucleic acid molecules. Mannosidase polypeptides also can be produced using mammalian expression systems, which include cells (e.g., immortalized cell lines such as COS cells, Chinese hamster ovary cells, HeLa cells, human embryonic kidney 293 cells, and 3T3 L1 cells) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., the metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter and the cytomegalovirus promoter), along with the nucleic acids described herein.

Typically, recombinant mannosidase polypeptides are tagged with a heterologous amino acid sequence such FLAG, polyhistidine (e.g., hexahistidine), hemagglutinin (HA), glutathione-S-transferase (GST), or maltose-binding protein (MBP) to aid in purifying the protein. Other methods for purifying proteins include chromatographic techniques such as ion exchange, hydrophobic and reverse phase, size exclusion, affinity, hydrophobic charge-induction chromatography, and the like (see, e.g., Scopes, *Protein Purification: Principles and Practice*, third edition, Springer-Verlag, New York (1993); Burton and Harding, *J. Chromatogr. A* 814:71-81 (1998)).

To produce molecules having uncapped terminal M6P residues *in vitro*, a target molecule containing a mannose-1-phospho-6 mannose linkage is contacted under suitable conditions with a purified mannosidase or a cell lysate containing a recombinantly produced mannosidase. The cell lysate can be from any genetically engineered cell, including a fungal cell, a plant cell, or animal cell. Non-limiting examples of animal cells include nematode, insect, plant, bird, reptile, and mammals such as a mouse, rat, rabbit, hamster, gerbil, dog, cat, goat, pig, cow, horse, whale, monkey, or human. Upon contacting the target molecule (e.g., an oligosaccharide or glycoprotein) with the purified mannosidase or cell lysate, the mannosidase hydrolyzes the mannose-1-phospho-6 mannose linkage and produces a target molecule having one or more uncapped terminal M6P residues. The methods described in Example 2 can be used to determine if the terminal M6P residues have been uncapped. Following processing by the mannosidase, the target molecule having uncapped terminal M6P residues can be isolated.

Suitable methods for obtaining cell lysates that preserve the activity or integrity of the mannosidase activity in the lysate can include the use of appropriate buffers and/or inhibitors, including nuclease, protease and phosphatase inhibitors that preserve or minimize changes in N-glycosylation activities in the cell lysate. Such inhibitors include, for example, chelators such as ethylenediamine tetraacetic acid (EDTA), ethylene glycol bis(P-aminoethyl ether) N,N,N1,N1-tetraacetic acid (EGTA), protease inhibitors such as phenylmethylsulfonyl fluoride (PMSF), aprotinin, leupeptin, antipain and the like, and phosphatase inhibitors such as phosphate, sodium fluoride, vanadate and the like. Appropriate buffers and conditions for obtaining lysates containing enzymatic activities are described in, e.g., Ausubel et al. *Current Protocols in Molecular Biology* (Supplement 47), John Wiley & Sons, New York (1999); Harlow and Lane, *Antibodies: A Laboratory Manual* Cold Spring Harbor Laboratory Press (1988); Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Press (1999); Tietz *Textbook of Clinical Chemistry*, 3rd ed. Burtis and Ashwood, eds. W.B. Saunders, Philadelphia, (1999).

A cell lysate can be further processed to eliminate or minimize the presence of interfering substances, as appropriate. If desired, a cell lysate can be fractionated by a variety of methods well known to those skilled in the art, including subcellular fractionation, and chromatographic techniques such as ion exchange, hydrophobic and reverse phase, size exclusion, affinity, hydrophobic charge-induction chromatography, and the like.

In some embodiments, a cell lysate can be prepared in which whole cellular organelles remain intact and/or functional. For example, a lysate can contain one or more of intact rough endoplasmic reticulum, intact smooth endoplasmic reticulum, or intact Golgi apparatus. Suitable methods for preparing lysates containing intact cellular organelles and testing for the functionality of the organelles are described in, e.g., Moreau *et al.* (1991) *J. Biol. Chem.* 266(7):4329-4333; Moreau *et al.* (1991) *J. Biol. Chem.* 266(7):4322-4328; Rexach *et al.* (1991) *J. Cell Biol.* 114(2):219-229; and Paulik *et al.* (1999) *Arch. Biochem. Biophys.* 367(2):265-273.

Target molecules, as used herein, refer to any molecule containing terminal mannose-1-phospho-6 mannose linkages or any molecule, when expressed in a cell of

fungal origin, that contains mannose-1-phospho-6 mannose linkages. Suitable target proteins include pathogen proteins such as tetanus toxoid or diphtheria toxoid; viral surface proteins such as cytomegalovirus (CMV) glycoproteins B, H and gCIII, human immunodeficiency virus 1 (HIV-1) envelope glycoproteins, Rous sarcoma virus (RSV) envelope glycoproteins, herpes simplex virus (HSV) envelope glycoproteins, Epstein Barr virus (EBV) envelope glycoproteins, varicella-zoster virus (VZV) envelope glycoproteins, human papilloma virus (HPV) envelope glycoproteins, Influenza virus glycoproteins, and Hepatitis family surface antigen; lysosomal proteins (e.g., acid alpha glucosidase, alpha galactosidase, glucocerebrosidase, cerebrosidase, or galactocerebrosidase); insulin; glucagons; growth factors; cytokines; chemokines; and antibodies or fragments thereof. Growth factors include, e.g., vascular endothelial growth factor (VEGF), Insulin-like growth factor (IGF), bone morphogenic protein (BMP), Granulocyte-colony stimulating factor (G-CSF), Granulocyte-macrophage colony stimulating factor (GM-CSF), Nerve growth factor (NGF); a Neurotrophin, Platelet-derived growth factor (PDGF), Erythropoietin (EPO), Thrombopoietin (TPO), Myostatin (GDF-8), Growth Differentiation factor-9 (GDF9), basic fibroblast growth factor (bFGF or FGF2), Epidermal growth factor (EGF), Hepatocyte growth factor (HGF). Cytokines include, for example, interleukins such as IL-1 to IL-33 (e.g., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, or IL-15)). Chemokines include, e.g., I-309, TCA-3, MCP-1, MIP-1 α , MIP-1 β , RANTES, C10, MRP-2, MARC, MCP-3, MCP-2, MRP-2, CCF18, MIP-1 γ , Eotaxin, MCP-5, MCP-4, NCC-1, Ck β 10, HCC-1, Leukotactin-1, LEC, NCC-4, TARC, PARC, or Eotaxin-2. Also included are tumor glycoproteins (e.g., tumor-associated antigens), for example, carcinoembryonic antigen (CEA), human mucins, HER-2/neu, and prostate-specific antigen (PSA) [Henderson and Finn, *Advances in Immunology*, 62, pp. 217-56 (1996)].

In some embodiments, the target protein can be one associated with a lysosomal storage disorder, which target proteins include, e.g., acid alpha glucosidase, alpha galactosidase, alpha-L-iduronidase, beta-D-galactosidase, beta-glucosidase, beta-hexosaminidase, beta-D-mannosidase, alpha-L-fucosidase, arylsulfatase B, arylsulfatase A, alpha-N-acetylgalactosaminidase, aspartylglucosaminidase, iduronate-2-sulfatase, alpha-glucosaminide-N-acetyltransferase, beta-D-glucoronidase, hyaluronidase, alpha-L-

mannosidase, alpha-neuraminidase, phosphotransferase, acid lipase, acid ceramidase, sphingomyelinase, thioesterase, cathepsin K, and lipoprotein lipase.

In some embodiments, the target proteins are fusion proteins in which the target protein is fused to another polypeptide sequence, or to a polymer, a carrier, an adjuvant, an immunotoxin, or a detectable (e.g., fluorescent, luminescent, or radioactive) moiety. For example, a target protein can be joined to a polymer such as polyethyleneglycol to increase the molecular weight of small proteins and/or increase circulation residence time.

In Vivo Methods of Uncapping M6P Residues

Genetically engineered cells described herein can be used to produce target molecules containing uncapped M6P residues. For example, a cell based method can include the steps of introducing into a fungal cell genetically engineered to include a nucleic acid encoding a mannosidase, a nucleic acid encoding a target molecule, wherein the cell produces the target molecule containing uncapped terminal M6P residues. In some embodiments, the nucleic acids encoding the mannosidase and target molecule contain a secretion sequence such that the mannosidase and target molecule are co-secreted.

Genetically engineered cells described herein contain a nucleic acid encoding a mannosidase and are useful for producing one or more target molecules having uncapped terminal M6P residues. Cells suitable for *in vivo* production of uncapped M6P residues can be of fungal origin, including *Yarrowia lipolytica*, *Arxula adeninivorans*, methylotrophic yeast (such as a methylotrophic yeast of the genus *Candida*, *Hansenula*, *Oogataea*, *Pichia* or *Torulopsis*) or filamentous fungi of the genus *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, or *Chrysosporium*. Exemplary fungal species include, without limitation, *Pichia anomala*, *Pichia bovis*, *Pichia canadensis*, *Pichia carsonii*, *Pichia farinose*, *Pichia fermentans*, *Pichia fluxuum*, *Pichia membranaefaciens*, *Pichia membranaefaciens*, *Candida valida*, *Candida albicans*, *Candida ascalaphidarum*, *Candida amphixiae*, *Candida Antarctica*, *Candida atlantica*, *Candida atmosphaerica*, *Candida blattae*, *Candida carpophila*, *Candida cerambycidarum*, *Candida chauliodes*, *Candida corydalis*, *Candida dosseyi*, *Candida dubliniensis*, *Candida ergatensis*, *Candida*

fructus, *Candida glabrata*, *Candida fermentati*, *Candida guilliermondii*, *Candida haemulonii*, *Candida insectamens*, *Candida insectorum*, *Candida intermedia*, *Candida jeffresii*, *Candida kefyr*, *Candida krusei*, *Candida lusitanae*, *Candida lyxosophila*, *Candida maltosa*, *Candida membranifaciens*, *Candida milleri*, *Candida oleophila*, *Candida oregonensis*, *Candida parapsilosis*, *Candida quercitrusa*, *Candida shehatei*, *Candida temnochilae*, *Candida tenuis*, *Candida tropicalis*, *Candida tsuchiyae*, *Candida sinolaborantium*, *Candida sojae*, *Candida viswanathii*, *Candida utilis*, *Oogataea minuta*, *Pichia membranaefaciens*, *Pichia silvestris*, *Pichia membranaefaciens*, *Pichia chodati*, *Pichia membranaefaciens*, *Pichia menbranaefaciens*, *Pichia minuscula*, *Pichia pastoris*, *Pichia pseudopolymorpha*, *Pichia quercuum*, *Pichia robertsii*, *Pichia saitoi*, *Pichia silvestrisi*, *Pichia strasburgensis*, *Pichia terricola*, *Pichia vanriji*, *Pseudozyma Antarctica*, *Rhodospiridium toruloides*, *Rhodotorula glutinis*, *Saccharomyces bayanus*, *Saccharomyces bayanus*, *Saccharomyces momdshuricus*, *Saccharomyces uvarum*, *Saccharomyces bayanus*, *Saccharomyces cerevisiae*, *Saccharomyces bisporus*, *Saccharomyces chevalieri*, *Saccharomyces delbrueckii*, *Saccharomyces exiguus*, *Saccharomyces fermentati*, *Saccharomyces fragilis*, *Saccharomyces marxianus*, *Saccharomyces mellis*, *Saccharomyces rosei*, *Saccharomyces rouxii*, *Saccharomyces uvarum*, *Saccharomyces willianus*, *Saccharomycodes ludwigii*, *Saccharomycopsis capsularis*, *Saccharomycopsis fibuligera*, *Saccharomycopsis fibuligera*, *Endomyces hordei*, *Endomycopsis fobuligera*, *Saturnispora saitoi*, *Schizosaccharomyces octosporus*, *Schizosaccharomyces pombe*, *Schwanniomyces occidentalis*, *Torulaspora delbrueckii*, *Torulaspora delbrueckii*, *Saccharomyces dairensis*, *Torulaspora delbrueckii*, *Torulaspora fermentati*, *Saccharomyces fermentati*, *Torulaspora delbrueckii*, *Torulaspora rosei*, *Saccharomyces rosei*, *Torulaspora delbrueckii*, *Saccharomyces rosei*, *Torulaspora delbrueckii*, *Saccharomyces delbrueckii*, *Torulaspora delbrueckii*, *Saccharomyces delbrueckii*, *Zygosaccharomyces mongolicus*, *Dorulaspora globosa*, *Debaryomyces globosus*, *Torulopsis globosa*, *Trichosporon cutaneum*, *Trigonopsis variabilis*, *Williopsis californica*, *Williopsis saturnus*, *Zygosaccharomyces bisporus*, *Zygosaccharomyces bisporus*, *Debaryomyces disporua*, *Saccharomyces bisporas*, *Zygosaccharomyces bisporus*, *Saccharomyces bisporus*, *Zygosaccharomyces mellis*, *Zygosaccharomyces priorianus*, *Zygosaccharomyces rouxiim*, *Zygosaccharomyces rouxii*, *Zygosaccharomyces*

barkeri, *Saccharomyces rouxii*, *Zygosaccharomyces rouxii*, *Zygosaccharomyces major*, *Saccharomyces rousii*, *Pichia anomala*, *Pichia bovis*, *Pichia Canadensis*, *Pichia carsonii*, *Pichia farinose*, *Pichia fermentans*, *Pichia fluxuum*, *Pichia membranaefaciens*, *Pichia pseudopolymorpha*, *Pichia quercuum*, *Pichia robertsii*, *Pseudozyma Antarctica*, *Rhodospiridium toruloides*, *Rhodospiridium toruloides*, *Rhodotorula glutinis*, *Saccharomyces bayanus*, *Saccharomyces bayanus*, *Saccharomyces bisporus*, *Saccharomyces cerevisiae*, *Saccharomyces chevalieri*, *Saccharomyces delbrueckii*, *Saccharomyces fermentati*, *Saccharomyces fragilis*, *Saccharomycodes ludwigii*, *Schizosaccharomyces pombe*, *Schwanniomyces occidentalis*, *Torulaspora delbrueckii*, *Torulaspora globosa*, *Trigonopsis variabilis*, *Williopsis californica*, *Williopsis saturnus*, *Zygosaccharomyces bisporus*, *Zygosaccharomyces mellis*, or *Zygosaccharomyces rouxii*. Exemplary filamentous fungi include various species of *Aspergillus* including, but not limited to, *Aspergillus caesiellus*, *Aspergillus candidus*, *Aspergillus carneus*, *Aspergillus clavatus*, *Aspergillus deflectus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus glaucus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus ochraceus*, *Aspergillus oryzae*, *Aspergillus parasiticus*, *Aspergillus penicilloides*, *Aspergillus restrictus*, *Aspergillus sojae*, *Aspergillus sydowi*, *Aspergillus tamari*, *Aspergillus terreus*, *Aspergillus ustus*, or *Aspergillus versicolor*. Such cells, prior to the genetic engineering as specified herein, can be obtained from a variety of commercial sources and research resource facilities, such as, for example, the American Type Culture Collection (Rockville, MD). Target molecules include proteins such as any of the target proteins described herein (see above).

Genetic engineering of a cell can include, in addition to an exogenous nucleic acid encoding a mannosidase, one or more genetic modifications such as: (i) deletion of an endogenous gene encoding an Outer CHain elongation (OCH1) protein; (ii) introduction of a recombinant nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation (e.g, a MNN4 polypeptide from *Yarrowia lipolytica*, *S. cerevisiae*, *Ogataea minuta*, *Pichia pastoris*, or *C. albicans*, or PNO1 polypeptide from *P. pastoris*) to increasing phosphorylation of mannose residues; (iii) introduction or expression of an RNA molecule that interferes with the functional expression of an OCH1 protein; (iv) introduction of a recombinant nucleic acid encoding a wild-type (e.g., endogenous or

exogenous) protein having a N-glycosylation activity (i.e., expressing a protein having an N-glycosylation activity); (v) introduction of a recombinant nucleic acid encoding a target molecule described above; or (v) altering the promoter or enhancer elements of one or more endogenous genes encoding proteins having N-glycosylation activity to thus alter the expression of their encoded proteins. RNA molecules include, e.g., small-interfering RNA (siRNA), short hairpin RNA (shRNA), anti-sense RNA, or micro RNA (miRNA). Genetic engineering also includes altering an endogenous gene encoding a protein having an N-glycosylation activity to produce a protein having additions (e.g., a heterologous sequence), deletions, or substitutions (e.g., mutations such as point mutations; conservative or non-conservative mutations). Mutations can be introduced specifically (e.g., by site-directed mutagenesis or homologous recombination) or can be introduced randomly (for example, cells can be chemically mutagenized as described in, e.g., Newman and Ferro-Novick (1987) *J. Cell Biol.* 105(4):1587).

Genetic modifications described herein can result in one or more of (i) an increase in one or more activities in the genetically modified cell, (ii) a decrease in one or more activities in the genetically modified cell, or (iii) a change in the localization or intracellular distribution of one or more activities in the genetically modified cell. It is understood that an increase in the amount of a particular activity (e.g., promoting mannosyl phosphorylation) can be due to overexpressing one or more proteins capable of promoting mannosyl phosphorylation, an increase in copy number of an endogenous gene (e.g., gene duplication), or an alteration in the promoter or enhancer of an endogenous gene that stimulates an increase in expression of the protein encoded by the gene. A decrease in one or more particular activities can be due to overexpression of a mutant form (e.g., a dominant negative form), introduction or expression of one or more interfering RNA molecules that reduce the expression of one or more proteins having a particular activity, or deletion of one or more endogenous genes that encode a protein having the particular activity.

To disrupt a gene by homologous recombination, a “gene replacement” vector can be constructed in such a way to include a selectable marker gene. The selectable marker gene can be operably linked, at both 5' and 3' end, to portions of the gene of sufficient length to mediate homologous recombination. The selectable marker can be one of any

number of genes which either complement host cell auxotrophy or provide antibiotic resistance, including URA3, LEU2 and HIS3 genes. Other suitable selectable markers include the CAT gene, which confers chloramphenicol resistance to yeast cells, or the lacZ gene, which results in blue colonies due to the expression of β -galactosidase. Linearized DNA fragments of the gene replacement vector are then introduced into the cells using methods well known in the art (see below). Integration of the linear fragments into the genome and the disruption of the gene can be determined based on the selection marker and can be verified by, for example, Southern blot analysis. A selectable marker can be removed from the genome of the host cell by, e.g., Cre-loxP systems (see below).

Alternatively, a gene replacement vector can be constructed in such a way as to include a portion of the gene to be disrupted, which portion is devoid of any endogenous gene promoter sequence and encodes none or an inactive fragment of the coding sequence of the gene. An "inactive fragment" is a fragment of the gene that encodes a protein having, e.g., less than about 10% (e.g., less than about 9%, less than about 8%, less than about 7%, less than about 6%, less than about 5%, less than about 4%, less than about 3%, less than about 2%, less than about 1%, or 0%) of the activity of the protein produced from the full-length coding sequence of the gene. Such a portion of the gene is inserted in a vector in such a way that no known promoter sequence is operably linked to the gene sequence, but that a stop codon and a transcription termination sequence are operably linked to the portion of the gene sequence. This vector can be subsequently linearized in the portion of the gene sequence and transformed into a cell. By way of single homologous recombination, this linearized vector is then integrated in the endogenous counterpart of the gene.

Expression vectors can be autonomous or integrative. A recombinant nucleic acid (e.g., one encoding a mannosidase) can be introduced into the cell in the form of an expression vector such as a plasmid, phage, transposon, cosmid or virus particle. The recombinant nucleic acid can be maintained extrachromosomally or it can be integrated into the yeast cell chromosomal DNA. Expression vectors can contain selection marker genes encoding proteins required for cell viability under selected conditions (e.g., URA3, which encodes an enzyme necessary for uracil biosynthesis or TRP1, which encodes an enzyme required for tryptophan biosynthesis) to permit detection and/or selection of

those cells transformed with the desired nucleic acids (see, e.g., U.S. Pat. No. 4,704,362). Expression vectors can also include an autonomous replication sequence (ARS). For example, U.S. Pat. No. 4,837,148 describes autonomous replication sequences which provide a suitable means for maintaining plasmids in *Pichia pastoris*.

Integrative vectors are disclosed, e.g., in U.S. Pat. No. 4,882,279. Integrative vectors generally include a serially arranged sequence of at least a first insertable DNA fragment, a selectable marker gene, and a second insertable DNA fragment. The first and second insertable DNA fragments are each about 200 (e.g., about 250, about 300, about 350, about 400, about 450, about 500, or about 1000 or more) nucleotides in length and have nucleotide sequences which are homologous to portions of the genomic DNA of the species to be transformed. A nucleotide sequence containing a gene of interest (e.g., a gene encoding a protein having N-glycosylation activity) for expression is inserted in this vector between the first and second insertable DNA fragments whether before or after the marker gene. Integrative vectors can be linearized prior to yeast transformation to facilitate the integration of the nucleotide sequence of interest into the host cell genome.

An expression vector can feature a recombinant nucleic acid under the control of a yeast (e.g., *Yarrowia lipolytica*, *Arxula adeninivorans*, *P. pastoris*, or other suitable fungal species) promoter, which enables them to be expressed in fungal cells. Suitable yeast promoters include, e.g., ADC1, TPI1, ADH2, hp4d, POX, and Gal10 (see, e.g., Guarente *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79(23):7410) promoters. Additional suitable promoters are described in, e.g., Zhu and Zhang (1999) *Bioinformatics* 15(7-8):608-611 and U.S. Patent No. 6,265,185.

A promoter can be constitutive or inducible (conditional). A constitutive promoter is understood to be a promoter whose expression is constant under the standard culturing conditions. Inducible promoters are promoters that are responsive to one or more induction cues. For example, an inducible promoter can be chemically regulated (e.g., a promoter whose transcriptional activity is regulated by the presence or absence of a chemical inducing agent such as an alcohol, tetracycline, a steroid, a metal, or other small molecule) or physically regulated (e.g., a promoter whose transcriptional activity is regulated by the presence or absence of a physical inducer such as light or high or low

temperatures). An inducible promoter can also be indirectly regulated by one or more transcription factors that are themselves directly regulated by chemical or physical cues.

It is understood that other genetically engineered modifications can also be conditional. For example, a gene can be conditionally deleted using, e.g., a site-specific DNA recombinase such as the Cre-loxP system (see, e.g., Gossen *et al.* (2002) *Ann. Rev. Genetics* 36:153-173 and U.S. Application Publication No. 20060014264).

A recombinant nucleic acid can be introduced into a cell described herein using a variety of methods such as the spheroplast technique or the whole-cell lithium chloride yeast transformation method. Other methods useful for transformation of plasmids or linear nucleic acid vectors into cells are described in, for example, U.S. Patent No. 4,929,555; Hinnen *et al.* (1978) *Proc. Nat. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163; U.S. Patent No. 4,879,231; and Sreekrishna *et al.* (1987) *Gene* 59:115, the disclosures of each of which are incorporated herein by reference in their entirety. Electroporation and PEG1000 whole cell transformation procedures may also be used, as described by Cregg and Russel, *Methods in Molecular Biology: Pichia Protocols*, Chapter 3, Humana Press, Totowa, N.J., pp. 27-39 (1998).

Transformed fungal cells can be selected for by using appropriate techniques including, but not limited to, culturing auxotrophic cells after transformation in the absence of the biochemical product required (due to the cell's auxotrophy), selection for and detection of a new phenotype, or culturing in the presence of an antibiotic which is toxic to the yeast in the absence of a resistance gene contained in the transformants. Transformants can also be selected and/or verified by integration of the expression cassette into the genome, which can be assessed by, e.g., Southern blot or PCR analysis.

Prior to introducing the vectors into a target cell of interest, the vectors can be grown (e.g., amplified) in bacterial cells such as *Escherichia coli* (*E. coli*) as described above. The vector DNA can be isolated from bacterial cells by any of the methods known in the art which result in the purification of vector DNA from the bacterial milieu. The purified vector DNA can be extracted extensively with phenol, chloroform, and ether, to ensure that no *E. coli* proteins are present in the plasmid DNA preparation, since these proteins can be toxic to mammalian cells.

In some embodiments, the genetically engineered fungal cell lacks the OCH1 gene or gene products (e.g., mRNA or protein) thereof, and is deficient in OCH1 activity. In some embodiments, the genetically engineered cell expresses a polypeptide capable of promoting mannosyl phosphorylation (e.g., a MNN4 polypeptide from *Yarrowia lipolytica*, *S. cerevisiae*, *Ogataea minuta*, *Pichia pastoris*, or *C. albicans*, or a PNO1 polypeptide from *P. pastoris*). For example, the fungal cell can express a MNN4 polypeptide from *Y. lipolytica* (Genbank® Accession Nos: XM_503217, Genolevures Ref: YALI0D24101g). In some embodiments, the genetically engineered cell is deficient in OCH1 activity and expresses a polypeptide capable of promoting mannosyl phosphorylation.

Following uncapping of the M6P residues, the target molecule can be isolated. In some embodiments, the target molecule is maintained within the yeast cell and released upon cell lysis. In some embodiments, the target molecule is secreted into the culture medium via a mechanism provided by a coding sequence (either native to the exogenous nucleic acid or engineered into the expression vector), which directs secretion of the molecule from the cell. The presence of the uncapped target molecule in the cell lysate or culture medium can be verified by a variety of standard protocols for detecting the presence of the molecule. For example, where the altered target molecule is a protein, such protocols can include, but are not limited to, immunoblotting or radioimmunoprecipitation with an antibody specific for the altered target protein (or the target protein itself), binding of a ligand specific for the altered target protein (or the target protein itself), or testing for a specific enzyme activity of the altered target protein (or the target protein itself).

In the target molecules produced using the methods described herein, at least 47% (e.g., at least 50, 55, 60, 65, 70, 75, 80, 85, or 90%) of the N-glycans on the glycoprotein have terminal phospho-6-mannose residues. The percentage of N-glycans having terminal phospho-6-mannose residues can be estimated from the peak areas in the DSA-FACE electropherograms. See Example 13.

In some embodiments, following isolation, the uncapped target molecule can be attached to a heterologous moiety, e.g., using enzymatic or chemical means. A “heterologous moiety” refers to any constituent that is joined (e.g., covalently or non-

covalently) to the altered target molecule, which constituent is different from a constituent originally present on the altered target molecule. Heterologous moieties include, e.g., polymers, carriers, adjuvants, immunotoxins, or detectable (e.g., fluorescent, luminescent, or radioactive) moieties. In some embodiments, an additional N-glycan can be added to the altered target molecule.

Methods for detecting glycosylation of a target molecule include DNA sequencer-assisted (DSA), fluorophore-assisted carbohydrate electrophoresis (FACE) or surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS). For example, an analysis can utilize DSA-FACE in which, for example, glycoproteins are denatured followed by immobilization on, e.g., a membrane. The glycoproteins can then be reduced with a suitable reducing agent such as dithiothreitol (DTT) or β -mercaptoethanol. The sulfhydryl groups of the proteins can be carboxylated using an acid such as iodoacetic acid. Next, the N-glycans can be released from the protein using an enzyme such as N-glycosidase F. N-glycans, optionally, can be reconstituted and derivatized by reductive amination. The derivatized N-glycans can then be concentrated. Instrumentation suitable for N-glycan analysis includes, e.g., the ABI PRISM® 377 DNA sequencer (Applied Biosystems). Data analysis can be performed using, e.g., GENESCAN® 3.1 software (Applied Biosystems). Optionally, isolated mannoproteins can be further treated with one or more enzymes to confirm their N-glycan status. Additional methods of N-glycan analysis include, e.g., mass spectrometry (e.g., MALDI-TOF-MS), high-pressure liquid chromatography (HPLC) on normal phase, reversed phase and ion exchange chromatography (e.g., with pulsed amperometric detection when glycans are not labeled and with UV absorbance or fluorescence if glycans are appropriately labeled). See also Callewaert *et al.* (2001) *Glycobiology* 11(4):275-281 and Freire *et al.* (2006) *Bioconjug. Chem.* 17(2):559-564.

Cultures of Engineered Cells

This document also provides a substantially pure culture of any of the genetically engineered cells described herein. As used herein, a “substantially pure culture” of a genetically engineered cell is a culture of that cell in which less than about 40% (i.e., less than about : 35%; 30%; 25%; 20%; 15%; 10%; 5%; 2%; 1%; 0.5%; 0.25%; 0.1%; 0.01%;

0.001%; 0.0001%; or even less) of the total number of viable cells in the culture are viable cells other than the genetically engineered cell, e.g., bacterial, fungal (including yeast), mycoplasmal, or protozoan cells. The term "about" in this context means that the relevant percentage can be 15% percent of the specified percentage above or below the specified percentage. Thus, for example, about 20% can be 17% to 23%. Such a culture of genetically engineered cells includes the cells and a growth, storage, or transport medium. Media can be liquid, semi-solid (e.g., gelatinous media), or frozen. The culture includes the cells growing in the liquid or in/on the semi-solid medium or being stored or transported in a storage or transport medium, including a frozen storage or transport medium. The cultures are in a culture vessel or storage vessel or substrate (e.g., a culture dish, flask, or tube or a storage vial or tube).

The genetically engineered cells described herein can be stored, for example, as frozen cell suspensions, e.g., in buffer containing a cryoprotectant such as glycerol or sucrose, as lyophilized cells. Alternatively, they can be stored, for example, as dried cell preparations obtained, e.g., by fluidized bed drying or spray drying, or any other suitable drying method.

Metabolic Disorders

Molecules having uncapped terminal M6P residues can be used to treat a variety of metabolic disorders. A metabolic disorder is one that affects the production of energy within individual human (or animal) cells. Most metabolic disorders are genetic, though some can be "acquired" as a result of diet, toxins, infections, etc. Genetic metabolic disorders are also known as inborn errors of metabolism. In general, the genetic metabolic disorders are caused by genetic defects that result in missing or improperly constructed enzymes necessary for some step in the metabolic process of the cell. The largest classes of metabolic disorders are disorders of carbohydrate metabolism, disorders of amino acid metabolism, disorders of organic acid metabolism (organic acidurias), disorders of fatty acid oxidation and mitochondrial metabolism, disorders of porphyrin metabolism, disorders of purine or pyrimidine metabolism, disorders of steroid metabolism disorders of mitochondrial function, disorders of peroxisomal function, and lysosomal storage disorders (LSDs).

Examples of metabolic disorders that can be treated through the administration of one or more molecules having uncapped terminal M6P residues (or pharmaceutical compositions of the same) can include hereditary hemochromatosis, oculocutaneous albinism, protein C deficiency, type I hereditary angioedema, congenital sucrase-isomaltase deficiency, Crigler-Najjar type II, Laron syndrome, hereditary Myeloperoxidase, primary hypothyroidism, congenital long QT syndrome, tyroxine binding globulin deficiency, familial hypercholesterolemia, familial chylomicronemia, abeta-lipoproteinemia, low plasma lipoprotein A levels, hereditary emphysema with liver injury, congenital hypothyroidism, osteogenesis imperfecta, hereditary hypofibrinogenemia, alpha-1 antichymotrypsin deficiency, nephrogenic diabetes insipidus, neurohypophyseal diabetes insipidus, adenosine deaminase deficiency, Pelizaeus Merzbacher disease, von Willebrand disease type IIA, combined factors V and VIII deficiency, spondylo-epiphyseal dysplasia tarda, choroideremia, I cell disease, Batten disease, ataxia telangiectasias, ADPKD-autosomal dominant polycystic kidney disease, microvillus inclusion disease, tuberous sclerosis, oculocerebro-renal syndrome of Lowe, amyotrophic lateral sclerosis, myelodysplastic syndrome, Bare lymphocyte syndrome, Tangier disease, familial intrahepatic cholestasis, X-linked adrenoleukodystrophy, Scott syndrome, Hermansky-Pudlak syndrome types 1 and 2, Zellweger syndrome, rhizomelic chondrodysplasia puncta, autosomal recessive primary hyperoxaluria, Mohr Tranebjaerg syndrome, spinal and bulbar muscular atrophy, primary ciliary dyskinesia (Kartagener's syndrome), gigantism and acromegaly, galactorrhea, Addison's disease, adrenal virilism, Cushing's syndrome, ketoacidosis, primary or secondary aldosteronism, Miller Dieker syndrome, lissencephaly, motor neuron disease, Usher's syndrome, Wiskott-Aldrich syndrome, Optiz syndrome, Huntington's disease, hereditary pancreatitis, anti-phospholipid syndrome, overlap connective tissue disease, Sjögren's syndrome, stiff-man syndrome, Brugada syndrome, congenital nephritic syndrome of the Finnish type, Dubin-Johnson syndrome, X-linked hypophosphosphatemia, Pendred syndrome, persistent hyperinsulinemic hypoglycemia of infancy, hereditary spherocytosis, aceruloplasminemia, infantile neuronal ceroid lipofuscinosis, pseudoachondroplasia and multiple epiphyseal, Stargardt-like macular dystrophy, X-linked Charcot-Marie-Tooth disease, autosomal dominant retinitis

pigmentosa, Wolcott-Rallison syndrome, Cushing's disease, limb-girdle muscular dystrophy, mucopolysaccharidosis type IV, hereditary familial amyloidosis of Finnish, Anderson disease, sarcoma, chronic myelomonocytic leukemia, cardiomyopathy, faciogenital dysplasia, Torsion disease, Huntington and spinocerebellar ataxias, hereditary hyperhomocysteinemia, polyneuropathy, lower motor neuron disease, pigmented retinitis, seronegative polyarthritis, interstitial pulmonary fibrosis, Raynaud's phenomenon, Wegner's granulomatosis, preproteinuria, CDG-Ia, CDG-Ib, CDG-Ic, CDG-Id, CDG-Ie, CDG-If, CDG-IIa, CDG-IIb, CDG-IIc, CDG-IId, Ehlers-Danlos syndrome, multiple exostoses, Griscelli syndrome (type 1 or type 2), or X-linked non-specific mental retardation. In addition, metabolic disorders can also include lysosomal storage disorders such as, but not limited to, Fabry disease, mucopolysaccharidosis I, Farber disease, Gaucher disease, GM₁-gangliosidosis, Tay-Sachs disease, Sandhoff disease, GM₂ activator disease, Krabbe disease, metachromatic leukodystrophy, Niemann-Pick disease (types A, B, and C), Scheie disease, Hunter disease, Sanfilippo disease, Morquio disease, Maroteaux-Lamy disease, hyaluronidase deficiency, aspartylglucosaminuria, fucosidosis, mannosidosis, Schindler disease, sialidosis type 1, Pompe disease, Pycnodysostosis, ceroid lipofuscinosis, cholesterol ester storage disease, Wolman disease, Multiple sulfatase deficiency, galactosialidosis, mucopolipidosis (types II, III, and IV), cystinosis, sialic acid storage disorder, chylomicron retention disease with Marinesco-Sjögren syndrome, Hermansky-Pudlak syndrome, Chediak-Higashi syndrome, Danon disease, or Geleophysic dysplasia.

Symptoms of a metabolic disorder are numerous and diverse and can include one or more of, e.g., anemia, fatigue, bruising easily, low blood platelets, liver enlargement, spleen enlargement, skeletal weakening, lung impairment, infections (e.g., chest infections or pneumonias), kidney impairment, progressive brain damage, seizures, extra thick meconium, coughing, wheezing, excess saliva or mucous production, shortness of breath, abdominal pain, occluded bowel or gut, fertility problems, polyps in the nose, clubbing of the finger/toe nails and skin, pain in the hands or feet, angiokeratoma, decreased perspiration, corneal and lenticular opacities, cataracts, mitral valve prolapse and/or regurgitation, cardiomegaly, temperature intolerance, difficulty walking, difficulty swallowing, progressive vision loss, progressive hearing loss, hypotonia, macroglossia,

areflexia, lower back pain, sleep apnea, orthopnea, somnolence, lordosis, or scoliosis. It is understood that due to the diverse nature of the defective or absent proteins and the resulting disease phenotypes (e.g., symptomatic presentation of a metabolic disorder), a given disorder will generally present only symptoms characteristic to that particular disorder. For example, a patient with Fabry disease can present a particular subset of the above-mentioned symptoms such as, but not limited to, temperature intolerance, corneal whirling, pain, skin rashes, nausea, or diarrhea. A patient with Gaucher syndrome can present with splenomegaly, cirrhosis, convulsions, hypertonia, apnea, osteoporosis, or skin discoloration.

In addition to the administration of one or more uncapped molecules described herein, a metabolic disorder can also be treated by proper nutrition and vitamins (e.g., cofactor therapy), physical therapy, and pain medications.

Depending on the specific nature of a given metabolic disorder, a patient can present these symptoms at any age. In many cases, symptoms can present in childhood or in early adulthood. For example, symptoms of Fabry disease can present at an early age, e.g., at 10 or 11 years of age.

As used herein, a subject “at risk of developing a metabolic disorder” is a subject that has a predisposition to develop a disorder, i.e., a genetic predisposition to develop metabolic disorder as a result of a mutation in an enzyme such as acid alpha glucosidase, alpha galactosidase, alpha-L-iduronidase, beta-D-galactosidase, beta-glucosidase, beta-hexosaminidase, beta-D-mannosidase, alpha-L-fucosidase, arylsulfatase B, arylsulfatase A, alpha-N-acetylgalactosaminidase, aspartylglucosaminidase, iduronate-2-sulfatase, alpha-glucosaminide-N-acetyltransferase, beta-D-glucuronidase, hyaluronidase, alpha-L-mannosidase, alpha-neuraminidase, phosphotransferase, acid lipase, acid ceramidase, sphingomyelinase, thioesterase, cathepsin K, or lipoprotein lipase. Clearly, subjects “at risk of developing a metabolic disorder” are not all the subjects within a species of interest.

A subject “suspected of having a disorder” is one having one or more symptoms of a metabolic disorder such as any of those described herein.

Pharmaceutical Compositions and Methods of Treatment

A target molecule having uncapped M6P residues can be incorporated into a pharmaceutical composition containing a therapeutically effective amount of the molecule and one or more adjuvants, excipients, carriers, and/or diluents. Acceptable diluents, carriers and excipients typically do not adversely affect a recipient's homeostasis (e.g., electrolyte balance). Acceptable carriers include biocompatible, inert or bioabsorbable salts, buffering agents, oligo- or polysaccharides, polymers, viscosity-improving agents, preservatives and the like. One exemplary carrier is physiologic saline (0.15 M NaCl, pH 7.0 to 7.4). Another exemplary carrier is 50 mM sodium phosphate, 100 mM sodium chloride. Further details on techniques for formulation and administration of pharmaceutical compositions can be found in, e.g., Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.). Supplementary active compounds can also be incorporated into the compositions.

Administration of a pharmaceutical composition containing molecules with uncapped M6P residues can be systemic or local. Pharmaceutical compositions can be formulated such that they are suitable for parenteral and/or non-parenteral administration. Specific administration modalities include subcutaneous, intravenous, intramuscular, intraperitoneal, transdermal, intrathecal, oral, rectal, buccal, topical, nasal, ophthalmic, intra-articular, intra-arterial, sub-arachnoid, bronchial, lymphatic, vaginal, and intra-uterine administration.

Administration can be by periodic injections of a bolus of the pharmaceutical composition or can be uninterrupted or continuous by intravenous or intraperitoneal administration from a reservoir which is external (e.g., an IV bag) or internal (e.g., a bioerodable implant, a bioartificial organ, or a colony of implanted altered N-glycosylation molecule production cells). See, e.g., U.S. Pat. Nos. 4,407,957, 5,798,113, and 5,800,828. Administration of a pharmaceutical composition can be achieved using suitable delivery means such as: a pump (see, e.g., *Annals of Pharmacotherapy*, 27:912 (1993); *Cancer*, 41:1270 (1993); *Cancer Research*, 44:1698 (1984); microencapsulation (see, e.g., U.S. Pat. Nos. 4,352,883; 4,353,888; and 5,084,350); continuous release polymer implants (see, e.g., Sabel, U.S. Pat. No. 4,883,666); macroencapsulation (see, e.g., U.S. Pat. Nos. 5,284,761, 5,158,881, 4,976,859 and 4,968,733 and published PCT patent applications WO92/19195, WO 95/05452); injection, either subcutaneously,

intravenously, intra-arterially, intramuscularly, or to other suitable site; or oral administration, in capsule, liquid, tablet, pill, or prolonged release formulation.

Examples of parenteral delivery systems include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, pump delivery, encapsulated cell delivery, liposomal delivery, needle-delivered injection, needle-less injection, nebulizer, aerosolizer, electroporation, and transdermal patch.

Formulations suitable for parenteral administration conveniently contain a sterile aqueous preparation of the altered N-glycosylation molecule, which preferably is isotonic with the blood of the recipient (*e.g.*, physiological saline solution). Formulations can be presented in unit-dose or multi-dose form.

Formulations suitable for oral administration can be presented as discrete units such as capsules, cachets, tablets, or lozenges, each containing a predetermined amount of the altered N-glycosylation molecule; or a suspension in an aqueous liquor or a non-aqueous liquid, such as a syrup, an elixir, an emulsion, or a draught.

A molecule having uncapped M6P residues suitable for topical administration can be administered to a mammal (*e.g.*, a human patient) as, *e.g.*, a cream, a spray, a foam, a gel, an ointment, a salve, or a dry rub. A dry rub can be rehydrated at the site of administration. Such molecules can also be infused directly into (*e.g.*, soaked into and dried) a bandage, gauze, or patch, which can then be applied topically. Such molecules can also be maintained in a semi-liquid, gelled, or fully-liquid state in a bandage, gauze, or patch for topical administration (see, *e.g.*, U.S. Patent No. 4,307,717).

Therapeutically effective amounts of a pharmaceutical composition can be administered to a subject in need thereof in a dosage regimen ascertainable by one of skill in the art. For example, a composition can be administered to the subject, *e.g.*, systemically at a dosage from 0.01 µg/kg to 10,000 µg/kg body weight of the subject, per dose. In another example, the dosage is from 1 µg/kg to 100 µg/kg body weight of the subject, per dose. In another example, the dosage is from 1 µg/kg to 30 µg/kg body weight of the subject, per dose, *e.g.*, from 3 µg/kg to 10 µg/kg body weight of the subject, per dose.

In order to optimize therapeutic efficacy, a molecule having uncapped M6P residues can be first administered at different dosing regimens. The unit dose and

regimen depend on factors that include, *e.g.*, the species of mammal, its immune status, the body weight of the mammal. Typically, levels of a such a molecule in a tissue can be monitored using appropriate screening assays as part of a clinical testing procedure, *e.g.*, to determine the efficacy of a given treatment regimen.

The frequency of dosing for a molecule having uncapped M6P residues is within the skills and clinical judgement of medical practitioners (*e.g.*, doctors or nurses). Typically, the administration regime is established by clinical trials which may establish optimal administration parameters. However, the practitioner may vary such administration regimes according to the subject's age, health, weight, sex and medical status. The frequency of dosing can be varied depending on whether the treatment is prophylactic or therapeutic.

Toxicity and therapeutic efficacy of such molecules or pharmaceutical compositions thereof can be determined by known pharmaceutical procedures in, for example, cell cultures or experimental animals. These procedures can be used, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Pharmaceutical compositions that exhibit high therapeutic indices are preferred. While pharmaceutical compositions that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to normal cells (*e.g.*, non-target cells) and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in appropriate subjects (*e.g.*, human patients). The dosage of such pharmaceutical compositions lies generally within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For a pharmaceutical composition used as described herein (*e.g.*, for treating a metabolic disorder in a subject), the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the

concentration of the pharmaceutical composition which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

As defined herein, a “therapeutically effective amount” of a molecule having uncapped M6P residues is an amount of the molecule that is capable of producing a medically desirable result (e.g., amelioration of one or more symptoms of a metabolic disorder) in a treated subject. A therapeutically effective amount (i.e., an effective dosage) can include milligram or microgram amounts of the compound per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram).

The subject can be any mammal, e.g., a human (e.g., a human patient) or a non-human primate (e.g., chimpanzee, baboon, or monkey), a mouse, a rat, a rabbit, a guinea pig, a gerbil, a hamster, a horse, a type of livestock (e.g., cow, pig, sheep, or goat), a dog, a cat, or a whale.

A molecule or pharmaceutical composition thereof described herein can be administered to a subject as a combination therapy with another treatment, e.g., a treatment for a metabolic disorder (e.g., a lysosomal storage disorder). For example, the combination therapy can include administering to the subject (e.g., a human patient) one or more additional agents that provide a therapeutic benefit to the subject who has, or is at risk of developing, (or suspected of having) a metabolic disorder (e.g., a lysosomal storage disorder). Thus, the compound or pharmaceutical composition and the one or more additional agents can be administered at the same time. Alternatively, the molecule can be administered first and the one or more additional agents administered second, or vice versa.

It will be appreciated that in instances where a previous therapy is particularly toxic (e.g., a treatment for a metabolic disorder with significant side-effect profiles), administration of a molecule described herein can be used to offset and/or lessen the amount of the previously therapy to a level sufficient to give the same or improved therapeutic benefit, but without the toxicity.

Any of the pharmaceutical compositions described herein can be included in a container, pack, or dispenser together with instructions for administration.

The following are examples of the practice of the invention. They are not to be construed as limiting the scope of the invention in any way.

EXAMPLES

EXAMPLE 1

Creation of a *Yarrowia lipolytica* strain with a higher degree of phosphorylated N-glycans

To upregulate the phosphorylation of glycans in *Y. lipolytica*, strain MTLY60 was transformed with 2 extra copies of the MNN4 gene, each in a separate expression vector. The MNN4 gene is involved in increasing glycan phosphorylation in yeast. FIG. 1 contains a schematic of the pYLTmAX plasmid into which the MNN4 gene was cloned to produce pYLTmAXMnn4, which contains the MNN4 open reading frame under control of the TEF promoter. A strain was made that contains two extra copies of the MNN4 gene, one under control of the hp4d promoter and one under control of the TEF1 promoter. N-glycans were prepared from strain MTLY60 Δ och1 (1 wild type copy of MNN4), strain MTLY60 Δ och1+Hp4dMNN4 (1WT +1 extra copy of MNN4) and strain MTLY60 Δ och1+Hp4dMNN4+TEFMNN4 (1WT +2 extra copies of Mnn4) and assayed by DNA sequencer-assisted (DSA), fluorophore-assisted carbohydrate electrophoresis (FACE). See, Callewaert *et al.*, *Glycobiology* 11(4):275-281 (2001). Based on the results in FIG. 2, it can be deduced that the mono phosphorylated peak is upregulated in the strain with 1 extra copy and that a peak of double phosphorylation appears. In the strain with 2 extra copies, the double phosphorylated peak was much higher and the peak of neutral Man₈GlcNAc₂ sugars was much lower.

EXAMPLE 2

Identification of a mannosidase activity that can uncap the capping mannose residue present on phosphorylated glycans of fungal origin.

The phosphorylation of sugars by yeast and filamentous fungi results in a mannose-phospho-mannose di-ester linkage (FIG. 3). To obtain a structure where the

phosphate is in a mono-ester linkage, a mannosidase is required that is able to hydrolyze the mannose-phosphate linkage, leaving the phosphate attached to the 6 position of a mannose of the high mannose glycan structure. Chiba *et al.*, *Glycobiology*, 12(12):821-8 (2002) indicate that a mannosidase from a *Cellulomonas* species is capable of decapping the mannose. However, Chiba *et al.* only partially purified the mannosidase protein and could not identify the gene encoding the protein.

A *Cellulosimicrobium cellulans* (also known as *Oerskovia xanthineolytica* and *Arthrobacter luteus*) isolate was obtained from the LMG bacteria collection and tested for production of mannosidase activity. The bacteria were grown at 30°C and in mannan containing medium to secrete the mannosidase in the medium. Bacterial supernatants (SN) were obtained from the cultures and tested for the desired mannosidase activity by incubating the SN with isolated N-glycans derived from the MNN4 overexpressing strain described in Example 1. After incubation, the glycans were assayed by DSA-FACE (FIG. 4).

After treatment with the SN, glycans gain an additional charge and migrate faster in the electric field and shift to the left hand side of the electroferogram. If these fast-running structures are indeed phosphomonoester-substituted high mannose glycans, they would be larger in size than the neutral products running at the same position. Treatment of such glycans with a phosphatase would result in neutral oligosaccharides that run much slower. As shown in FIG. 5, treatment with calf intestine phosphatase (CIP) resulted in the peaks displaying lower electrophoretic mobility, proving that the phosphates are terminal and that the mannose was decapped.

EXAMPLE 3

Partial purification and further identification of a mannosidase

To purify the mannosidase, *C. cellulans* was grown in 1L of medium B (Bagiyan *et al.*, *Eur. J. Biochem.* 249(1):286-92 (1997)) or medium A (Chiba *et al.*, 2002, *supra*). See Table 1. Thereafter, the medium was precipitated with 40% and 80% ammonium sulphate and the samples were analysed by SDS-PAGE. The ammonium sulphate fractions were dialyzed against 20mM Na-phosphate buffer pH 6.5 with 1 mM CaCl₂, and then tested for activity on oligosaccharides derived from a MNN4 overexpressing strain (Example 1).

TABLE 1
Medium components

Medium A (1 liter)	Medium B (1 liter)
2 g mannan	2 g mannan
0.5 g (NH ₄) ₂ SO ₄	2 g (NH ₄) ₂ SO ₄
0.4 g MgSO ₄ .7H ₂ O	0.02 g MgSO ₄ .7H ₂ O
20 mg FeSO ₄ .7H ₂ O	1 mg FeSO ₄ .7H ₂ O
60 mg CaCl ₂ .2H ₂ O	1 g yeast extract
1 g yeast extract	4.2 g KOH
7.54 g K ₂ HPO ₄	14 g KH ₂ PO ₄
2.32 g KH ₂ PO ₄	

Both cultivation conditions resulted in the production of the uncapping activity. Only the 40% ammonium sulphate fraction derived from medium B showed activity, whereas all fractions of the medium A supernatant displayed activity.

The 40% ammonium sulphate sample derived from the medium A cultivation was further purified over a silica-based gel filtration column (FIG. 6). This resulted in a peak with a shoulder around 670 kDa.

All elution fractions were incubated with oligosaccharides derived from a MNN4 overexpressing *Yarrowia lipolytica* strain (described in Example 1) (with or without a following CIP-digest) to test for the phosphate uncapping activity. The decapping and mannosidase activity was observed in all of the samples. Samples were also analyzed on SDS-PAGE (FIG. 7), which showed not just one protein band, but several protein bands. Several bands were cut out from the gel and portions of their sequence analyzed by de novo peptide sequencing using Mass Spectrometry.

The de novo sequencing results revealed several peptide sequences, which were compared against the sequences in the non redundant database using BLAST. Peptides with homology to the following proteins were identified: a phosphodiesterase, a hypothetical protein, a putative alpha-1,2 mannosidase (identified peptides shown in Table 2) (homology to a mannosidase from *Magnetospirillum*), and an aminopeptidase Y. The phosphodiesterase was a possible candidate, but here only 2 of the 6 peptides gave a hit. The mannosidase also was a candidate with 3/5 and 5/5 hits for 2 different mannosidases.

TABLE 2

Peptide Sequences

Peptide Sequence	SEQ ID NO
SAYQSFTTR	1
VWGFSHR	2
VEGGWLPR	3
TQGNNFALLPER	4
DVHAELTAMAR	5

EXAMPLE 4

Identification of mannosidases with the desired sequence
based on whole genome sequencing

To identify the mannosidase gene coding for the desired activity, the genome of *C. cellulans* was sequenced using a Titanium 454 sequencing (Eurofins MWG Operon). Due to the high GC content, the sequencing was only partial (1.96 Mbases) and of poor quality (with only a low average contig size). The high GC content of the genome that causes loop formation during the emulsion PCR (emPCR), resulting in deletions and very short sequences.

This problem was overcome using new sequencing chemistry for the emPCR that was made available in beta testing by Roche. This gave a much improved sequence (4.7 Mbases), allowing the identification of 5 mannosidase genes belonging to glycosyl hydrolase family 92, one of which (CcMan1, SEQ ID NO:6) corresponds to the sequence from which the peptides described in Example 3 were obtained. No mannosidases from family 38 or 47 were found. The start codon of each of CcMan1-CcMan4 was predicted by MetaGeneAnnotator (see the world wide web at metagene.cb.k.u-tokyo.ac.jp/metagene/) and compared to Blast results with known genes. The start codon of CcMan5 could not be predicted since it is missing from the sequence. The signal sequence of each gene was predicted with signal P (see the world wide web at cbs.dtu.dk/services/SignalP/) by two methods (neural networks and hidden markov models).

FIGs. 8-12 contain the nucleotide and encoded amino acid sequences of the 5 mannosidase genes from *C. cellulans*.

EXAMPLE 5

Heterologous expression of mannosidase for *in vitro* or *in vivo* mannose decapping.

In order to allow decapping of the yeast type phosphorylation by the mannosidase, it has to be expressed either heterologously in a different host or in the same fungal host from which the protein for therapeutic use is expressed. In the latter case it can be co-secreted or targeted to an intracellular compartment (e.g., Golgi apparatus or endoplasmic reticulum). This can be accomplished by cloning the gene (be it codon optimised for the target host or not) operably linked after a promoter in an expression vector. The mannosidase can be tagged with an epitope tag to allow easy detection and purification or expressed as such. It can be secreted in the periplasm of a bacterial cell or expressed intracellularly. In case of expression in the fungal host, the sequence can contain a secretion signal or a targeting signal to target the protein to an intracellular compartment or both. Table 3 contains a list of secretion and targeting signals for expression in fungal organisms. Examples of such expression vectors are presented in FIG. 13.

TABLE 3

Secretion and targeting signals for expression in fungal organisms

Secretion signals	Golgi targeting signal	
	N-terminal	C-terminal
<i>LIP2</i> prepro	<i>MNN2</i>	<i>KEX2</i>
<i>LIP2</i> pre	<i>MNN4</i>	
S.c. α mating factor	<i>MNN6</i>	
<i>XPR2</i> prepro	<i>MNN1</i>	
<i>XPR2</i> pre	<i>MNN9</i>	
	<i>OCH1</i>	
	<i>SEC12</i>	
	<i>KRE2</i>	

The CcMan1-Man5 genes were codon optimized for expression in *E. coli*. See FIGs. 14-18 for the codon optimized sequences. Table 4 contains the length of each

codon optimized nucleotide sequence and the predicted molecular weight of each polypeptide without the signal sequence.

TABLE 4
Codon Optimized Genes

	Length (bp)	SEQ ID NO	Size (kDa) of encoded product
CcMan1	2613	16	92.6
CcMan2	3483	17	121.6
CcMan3	3363	18	116
CcMan4	5283	19	184
CcMan5	4956	20	173

EXAMPLE 6

Cloning and Activity of *C. cellulans* Glycosyl Hydrolase (GH) Family 92 Enzymes

The CcMan1-CcMan5 codon optimized nucleic acids were cloned into *E. coli* vectors pLSH36, which contains a Spy signal sequence, and/or pLSAH36, which contains a DsbA signal sequence for periplasmic expression. Both pLSH36 and pLSAH36 result in the encoded polypeptide having a polyhistidine tag and a murine caspase-3 site, which can be used for the removal of the His6-tag during purification. FIG. 19 contains a schematic of the pLSH36 and pLSAH36 vectors as well as the cloning strategy for introducing the *C. cellulans* GH92 genes into the vectors. After cloning, the different mannosidases were transformed into *E. coli* BL21 + pICa2 expression strain. The transformed strains were grown to an optical density (OD) of 0.5 to 1 and induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Different cell fractions (medium, periplasm, soluble and insoluble fraction) were isolated and analyzed by SDS PAGE and Western blotting with an anti-His6 antibody. For CcMan1, CcMan2, and CcMan3, expression was detected in all fractions. For CcMan4 and CcMan5, expression was the highest in the soluble fraction, but some expression also was detected in the other fractions.

To determine the activity of the CcMan1-CcMan5 proteins, activity tests were performed using methylumbelliferyl alpha mannoside (MUM) as set forth in Chiba *et al.*, 2002, *supra*. For CcMan1 and CcMan2, the medium and periplasm samples were able to hydrolyze MUM weakly, whereas CcMan3 and CcMan5 were not able to hydrolyze

MUM. The soluble fraction of CcMan4 gave the highest fluorescent signal, indicating that CcMan4 is the only mannosidase with α 1,2-mannosidase activity.

All medium and periplasmic samples of the 5 different *C. cellulans* mannosidases also were tested on sugars derived from the MNN4 overexpressing strain of Example 1 (referred to herein as MNN4 sugars) to see if they were able to degrade the sugars and uncap the mannose of the mannose-6-phosphate. The sugars were incubated overnight and analysed by DNA sequencer-assisted, fluorophore-assisted carbohydrate electrophoresis (DSA-FACE). The sugar profiles of the medium samples could not be analyzed due to fluorophoric molecules in the medium presentation resulting in irrelevant peaks in the electroferogram. The sugar profiles of the periplasm of CcMan1, CcMan2 and CcMan3 showed neither degradation nor decapping, CcMan4 showed degradation, and CcMan5 showed decapping activity (FIG. 20). A CIP-digest on the decapped sugars confirmed the decapping activity of CcMan5 as the dephosphorylated peaks moved to neutral Man8.

The active mannosidases CcMan4 and CcMan5 were aligned with Bt3990 (744 AA) and Bt2199 (739 AA), family 92 mannosidases with known structure (see Zhu *et al.*, *Nat. Chem. Biol.*, 6(2):125-32. Epub 2009 Dec 27 (2010)). See FIG. 21. Since only the first part of CcMan4 and CcMan5 aligned with Bt3990 and Bt2199, and because they are large proteins, it was decided to clone the first domain of each protein separately and test the activity. CcMan4domain (1-3357 bp, i.e., nucleotides 1-3357 of SEQ ID NO:20) and CcMan5domain (1-2322 bp, i.e., nucleotides 1-2322 of SEQ ID NO:20) were cloned into the pLSAH36 *E. coli* expression vector. See, FIG. 19 for a schematic of the pLSAH36 cloning vector. The expression vectors were transformed into the *E. coli* BL21 + pICa2 expression strain, which was grown to an OD of 0.5 to 1, and induced with 1 mM IPTG. Different cell fractions (medium, periplasm, soluble and insoluble fraction) were isolated and analyzed by SDS PAGE and Western blotting with an anti-His6 antibody. Expression was detected in all 4 cell fractions.

The activity of the domains was tested on Mnn4 sugars. Hereto, the periplasmic fraction of each of the CcMan4domain and CcMan5domain was incubated in the presence of Mnn4 sugars (FIG. 22) and analyzed by DSA-FACE. This experiment showed that the CcMan4domain lost its mannosidase activity since no degradation could

be detected (FIG. 22, panel 4). In contrast, the CcMan5 domain kept its uncapping activity (FIG. 22, panel 6).

EXAMPLE 7

Production and Purification of CcMan5 and its family 92 homologous domain

The recombinant CCman5 (nucleotides 1-4995 of SEQ ID NO:20 and CcMan5 domain (nucleotides 1-2322 of SEQ ID NO:20) were expressed in *E. coli* strain BL21codon + pICA2 that were transformed with the expression vectors pLSAHCcMan5 and pLSAHCcMan5domain. Expression was induced by IPTG under control of a λ pL-promotor (see WO 98/48025 and WO 04/074488). See Example 6 and FIG. 19 for a description of pLSAH. The transformed bacteria were grown in Luria Bertani (LB) medium supplemented with ampicillin (100 μ g/ml) and kanamycin (50 μ g/ml) overnight at 28 °C before 1/100 inoculation in a 20 liter fermenter provided with LB medium supplemented with ampicillin (100 μ g/ml) and 1 % glycerol. The initial stirring and airflow was 200 rpm and 1.5 l/min., respectively, and was automatically adapted to keep the pO₂ at 30 %. The temperature was kept at 28°C. The cells were grown to an optical density of $A_{600\text{ nm}} = 1.0$, transferred at 20°C, and expression was induced by addition of 1 mM IPTG overnight. Cells were then harvested and frozen at -20 °C. After thawing, the cells were gently resuspended at a concentration of 3 ml/g in 50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 1 mM PMSF and 10 μ g/ml DNaseI. The periplasmic fraction was prepared by stirring the cell suspension for 1 h at 4 °C and was isolated by centrifugation at 18,000 x g for 30 min. All steps were conducted at 4 °C. The clear supernatant was applied to a 20 ml Ni-Sepharose 6 FF column (GE Healthcare), equilibrated with 20 mM NaH₂PO₄ pH 7.4, 300 mM NaCl, 20 mM imidazole, 0.1 % CHAPS. The column was eluted with 20 mM NaH₂PO₄ pH 7.4, 20 mM NaCl, 400 mM imidazole, 0.1 % CHAPS after an extra wash step with 50 mM of imidazole in the same buffer. The elution fraction was diluted 1/10 with 20 mM Tris pH 8.0, 0.1 % CHAPS and loaded on an 14 ml Source 15Q column (GE Healthcare) to remove contaminants. After equilibration, the protein of interest was eluted by a linear gradient over 10 column volumes of NaCl from 0 to 1 M in 20 mM Tris, 0.1 % CHAPS. The CcMan5 and CcMan5 domain containing fractions were further injected on a HiLoad 26/60 Superdex 200 prep grade with PBS as running

solution. The obtained fractions were analyzed by SDS-PAGE and western blotting with an anti-His6 antibody. Finally, the concentration was determined using the BCA assay (Pierce). The purified yield for the full-length CcMan5 protein was 5.7 mg, and for the CcMan5 family 92 domain, it was 110 mg from these 20 L fermentations, showing that the family 92 domain alone can be produced and purified in higher yield. The activity of the purified CcMan5 domain was tested on the Mnn4 isolated sugars as set forth in Example 6. A decapped sugar profile was obtained.

EXAMPLE 8

Structure of CcMan5domain

CcMan5₁₋₇₇₄ (residues 1 to 774 of SEQ ID NO:50, encoded by nucleotides 1-2322 of SEQ ID NO:20; corresponding to the mature protein after removal its natural leader sequence) was expressed in *E. coli* BL21 (DE3) periplasm as a fusion product starting with an N-terminal 6xHis tag followed by a 9 amino acid linker (VGPGSDEV D, SEQ ID NO:21) after the DsbA leader sequence. Cells were cultured in M9 medium containing 100 µg/ml of kanamycin and 100 µg/ml ampicillin at 28 °C. At an OD₆₀₀ of 0.4, CcMan5₁₋₇₇₄ expression was induced by addition of 1mM IPTG and the culture was further grown overnight at 18°C. Cells from the overnight culture were harvested by centrifugation, washed and incubated for 20 min at 4°C with buffer containing 20 mM Tris/HCl pH 8.0, 20% sucrose, 5 mM EDTA, and 0.1 mg/ml lysozyme to make spheroplasts. Periplasmic proteins were isolated from spheroplasts by centrifugation at 20,000xg for 20 min.

CcMan5₁₋₇₇₄ was purified from the periplasmic extract by metal ion affinity chromatography (HisTrap HP, GE Healthcare, loading under a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and eluted using an imidazole gradient up to 400 mM), ion exchange chromatography (HiTrap Q FF, GE Healthcare, buffer: 20 mM Tris-HCl pH 8.0, 40 mM NaCl and a NaCl gradient up to 1 M) and hydrophobic interaction chromatography (HiTrap Phenyl HP, GE Healthcare, loading buffer: 20 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 M (NH₄)₂SO₄ and eluted using a (NH₄)₂SO₄ gradient up to 0 mM).

Purified CcMan5₁₋₇₇₄ was concentrated to 130 mg/ml in 10 mM Tris– HCl pH 8.0, 10mM NaCl and plate-like crystals (0.2x0.07x0.01 mm³) were grown by vapor diffusion using a crystallization solution containing 0.2 M Na fluoride, 0.1 M Bis-Tris propane pH

7.5 and 20% PEG 3350. Crystals were briefly transferred into a cryoprotecting solution containing the crystallization solution supplemented with 10%(v/v) glycerol and flash-cooled in liquid nitrogen. Single crystal diffraction data were collected at 100 K at the PXIII beamline at the Swiss Light Source (SLS, Villigen, Swiss) and beamline BM30A at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The structure of CcMan5₁₋₇₇₄ was solved using a KAuCl₄-soaked crystal for the calculation of experimental phases from a SAD experiment at 11.958 keV, corresponding to the Au L-III absorption edge. FIG. 33 contains the structural coordinates of the catalytic center. The CcMan5 model built from the experimental phases was refined by maximum likelihood methods against 2 Å resolution data collected on a native crystal to a final R- and freeR-factor of 19.3 and 23.9 %, respectively. The final model contains 2 CcMan5₁₋₇₇₄ molecules per asymmetric unit (residues 8 to 771), comprising 11.513 protein atoms, 860 solvent atoms, 2 Ca²⁺ ions and 1 bis-tris-propane and glycerol molecule each.

Based on sequence similarity, CcMan5 falls within family 92 of glycosyl hydrolases (GH92), which are defined as exo-acting alpha-mannosidases. The X-ray structures for two GH92 family members with α 1, 2-mannosidase activity are available: Bt3990 and Bt2199 (PDB access codes 2WVX and 2WVY, respectively). The overall fold seen from the CcMan5₁₋₇₇₄ structure solved here, and deposited as PDB entry 2xsg, corresponds well with that seen in both Bt3990 and Bt2199, with r.m.s.d (root mean standard deviation) values of 1.99 Å and 2.12 Å over 624 and 621 matched C α atoms, respectively. CcMan5₁₋₇₇₄ consists of two domains, an N-terminal β -sandwich domain (residues 8 to 271) and a C-terminal ($\alpha\alpha$)₆ barrel domain (residues 291 to 771), connected via an α -helical linker (residues 272 to 290). The interface between both domains gives shape to a shallow cavity that harbors a conserved catalytic Ca²⁺ ion and gives shape to the -1 substrate binding site (nomenclature: Davies *et al.*, *Biochem. J.* 321:557-9 (1997)) and the catalytic center (FIGs. 23 and 24).

GH92 family glycosyl hydrolases are Ca²⁺-dependent alpha-mannosidases that catalyse glycosidic bond hydrolysis through a single displacement mechanism, leading to inversion of the anomeric configuration in the released mannose (Zhu *et al.*, 2010, *supra*). In CcMan5₁₋₇₇₄, the catalytic Ca²⁺ is octahedrally coordinated via the carbonyl oxygen of Asn 588, a carboxyl oxygen of Glu589 and Asp662 each, and three water

molecules (W1, W2, W3 – see FIG. 23) that lie in the equatorial coordination plane. An additional water molecule (W4) is present near the catalytic center, bound to the carboxyl groups of the conserved pair Asp 660 and Asp 662. The substrate binding cavity surrounding the catalytic Ca^{2+} is lined by the residues Asn 588, Gln 589, Thr 626, Thr 658, Asp 22, Asn 25, Gly 71, Gly 72, Phe 195, Tyr 196, Arg 405, Trp 354, Tyr 535, and Gln536 (FIG. 23).

CcMan5 sets itself apart from other alpha-mannosidases in the GH92 family because of its unique ability to accept mannose-alpha-1-phospho-6-mannose (Man-P-Man) as a substrate and a lack of alpha-1,2-, alpha-1,3-, alpha-1,4- or alpha-1,6-mannosidase activity. In order to obtain insight in the discriminating residues in the CcMan5 active site that give rise to this unique substrate specificity, Man-P-Man was modeled into the CcMan5₁₋₇₇₄ active site of molecule B of the asymmetric unit (FIG. 25). Positioning of the -1 mannose was based on the gross binding conformation observed in Bt3990 and guided by the positions of two water molecules (W2 and W3) and a glycerol molecule present in the apo active site. In this way, the O2, O3, O4 and O6 hydroxyl groups of the -1 mannose take equivalent positions to those observed for the water molecules W2, W3 and the O1 and O3 hydroxyl groups of the glycerol molecule, respectively. Thus, the mannose -1 O2 and O3 hydroxyl group position in the equatorial plane of the actohedral Ca^{2+} coordination sphere. O3 makes an additional hydrogen bond to the Asp 355 carboxyl group. The latter is further provides a H-bond to the O4 hydroxyl, which also comes within H-bonding distance of the Arg 405 guanidinium group. The O6 hydroxyl and O5 oxygen can be involved in H-bonding with the Gly 71 amide. For modeling, the -1 mannose was retained in its ground state chair conformation. As observed for Bt3990, positioning of the O2 hydroxyl group to come into idealized coordination with Ca^{2+} will lead to a distortion of the sugar ring to a half chair conformation (see FIG. 25). This is in line with the general acceptance that a distortion of the sugar ring during catalysis is required for the nucleophilic substitution at the acetal center in α -mannosides in order to break the 1,2-diaxial interaction of the incoming nucleophile with the O2 hydroxyl (Vocadlo *et al.*, *Curr. Opin. Chem. Biol.* 12:539-55 (2008)). The obtained model for substrate binding in the -1 site further shows that water molecule W4 lies in a good position to act as nucleophile for in line attack on the acetal

carbon. W4 is in H-bond interaction with the carboxyl groups of Asp 660 and Asp 662, which are conserved throughout GH92 enzymes and are proposed to form the base catalyst(s) for activation of the nucleophile. Therefore, the modeled substrate binding at the -1 site and the position of catalytic residues and nucleophile are consistent with the mechanistic requirements for nucleophilic substitution with inversion of the anomeric center in the released mannose. As discussed above, CcMan5 distinguishes itself by the ability to bind and hydrolyse Man-P-Man. The obtained model for Man-P-Man binding to the CcMan5 active site now provides a rationale for these observations. In known GH92 family members, the anomeric oxygen making the glycosidic bond, is in electrostatic interaction with the carboxyl group of a conserved glutamic acid residue (Glu 533 in Bt3990). The glutamic acid residue has been shown to serve as catalytic acid, stabilizing the transition intermediate by binding the anomeric oxygen and protonating the leaving group (Zhu *et. al.*, 2010, *supra*). In CcMan5, the equivalent residue to Bt3990 Glu 533 is mutated to glutamine, which is not able to serve as a proton donor and therefore explains the loss-of-function in CcMan5 for hydrolysis of mannosides. In Man-P-Man substrates, however, the phosphate bound to the anomeric oxygen constitutes a much stronger leaving group that would not require an acid catalyst to protonate the anomeric oxygen, explaining why enzymes like CcMan5 can retain catalytic activity for Man-P-Man substrates. Concomitant with substitution of the catalytic acid, the equivalent of Glu 585 in Bt3990 is replaced by Thr in CcMan5 (Thr 626). In Bt3990, Glu 585 interacts with Glu 533 and has been suggested to regulate the latter's pK_a and/or play a role in binding the leaving group in 2-linked mannosides (Zhu *et. al.*, 2010, *supra*).

It appears that mutation to non-acidic residues in the Gln 536 and Thr 626 pair alleviates part of the negative electrostatic potential in the binding site, thereby tolerating the phosphate linkage to the anomeric oxygen in Man-P-Man substrates. In CcMan5, the modeled phosphate binding site (P in FIG. 25) is shaped by Thr 626 and the amide of Gly 72, both of which appear able to donate a H-bond to the non-glycosidic oxygens in the phosphate.

Finally, based on the modeled binding of Man-p-Man in the CcMan5₁₋₇₇₄ active site, the reducing end mannose comes in the vicinity of two tyrosine residues, Tyr 535 and Tyr 196, suggesting the latter to residues form part of the +1 mannose binding site.

Both residues lay at the edge of a shallow cleft that could be involved in further interactions with glycans at the reducing end of the glycan tree.

EXAMPLE 9

Expression of α -GalactosidaseA in *Y. lipolytica*

A nucleic acid encoding human α -GalactosidaseA, without pre and pro sequence, was synthesized with codon optimization for *Y. lipolytica* and addition of a Myc-His tag. The obtained sequence was cloned in frame after the pre sequence of the lip2 gene. The nucleotide sequence of the codon optimized nucleotide sequence (SEQ ID NO:22) is set forth in FIG. 26A; amino acid sequence (SEQ ID NO:23) is presented in FIG. 26B.

Y. lipolytica MTLY60 with 2 extra copies of MNN4 and one copy of α -GalactosidaseA was induced in a larger culture to purify over a Ni-NTA column. Thus, they were grown in YTG and induced in oleic acid medium in 2 x 225 ml (2L shake flask) during 48 hours. The culture was centrifuged, followed by filtration of the medium over a 0.22 μ m filter. The filtered medium was desalted on a sephadex G25 XK50/100 column (GE Healthcare) to 20 mM NaH₂PO₄ pH 7.4, 0.5 M NaCl, 20 mM imidazole to remove non-protein disturbing contaminants before purification on Ni-sepharose 6 FF. The desalted protein fraction was loaded on a 4.3 ml Ni-sepharose 6 FF column (GE Healthcare), equilibrated with 20 mM NaH₂PO₄ pH 7.4, 0.5 M NaCl, 20 mM imidazole, washed with 50 mM imidazole in the same buffer and eluted with 20 mM NaH₂PO₄ pH 7.4, 20 mM NaCl, 400 mM imidazole. Samples 3-10 and 36-49 after the Ni-sepharose column were analysed on SDS-PAGE and Western blotting using an anti-His6 antibody. A protein band of around 50 kDa and of 65 kDa was present on coomassie in samples 40 and 41 was revealed by Coomassie blue staining of the SDS-PAGE gel. In the western blot, only a band of 50 kDa was detected and is most likely the α -GalactosidaseA. The estimated yield of the purified α -GalactosidaseA was 100-125 μ g/L culture medium.

The purified sample was used to determine the type of sugars on the recombinant α -GalactosidaseA. The sugars were removed in solution and afterwards labelled with APTS. After cleaning the sample by gel filtration, the sugars were analyzed on DSA-FACE. The expected sugars, the mono mannophosphorylated Man₈GlcNAc₂ peak (P)

and the double mannophosphorylated Man₈GlcNAc₂ peak (PP) were present as major peaks.

EXAMPLE 10

Expression of human alpha glucosidase in *Y. lipolytica*

Y. lipolytica strain OXYY1589 was constructed that contained three copies of the human alpha glucosidase (also known as acid alpha glucosidase (GAA) or acid maltase EC3.2.1.3) and two copies of the *Y. lipolytica* MNN4 gene. The genotype of strain OXYY1589 is as follows:

MatA, *leu2-958*, *ura3-302*, *xpr2-322*,
gut2-744, *ade2-844*

POX2-Lip2pre-huGAA:URA3Ex::zeta
POX2-Lip2pre-huGAA:LEU2Ex::zeta
POX2-Lip2pre-hGM-CSF:GUTEx::zeta
YIMNN4-POX2-hp4d-YLMNN4 :ADE2::PT targeted

All transformations were carried out according to well established protocols with modifications for the different selective markers. In all cases (unless otherwise specified), a huGAA integration fragment has been obtained by NotI restriction digestion in order to remove the kanamycin resistance gene from the expression plasmids. The resulting fragments were all separated by agarose gel electrophoresis followed by Qiagen column purification of the correct huGAA fragment. Strain OXYY1589 was constructed by first cloning human GAA (huGAA) into a *Y. lipolytica* expression vector and constructing a *Y. lipolytica* MNN4 tandem expression vector. Three stable integrative transformations then were performed in order to obtain the final huGAA production strain OXYY1589.

Y. lipolytica codon optimized huGAA expression vector: The nucleotide sequence encoding the 110 kDa human GAA (huGAA) precursor was chemically synthesized and codon optimized for *Y. lipolytica* expression. In the synthetic construct, the pre- and the pro- huGAA signal peptides were eliminated such that the protein starts at amino acid 57. The synthetic ORF of huGAA (FIG. 27A) is fused in frame at the 5' end to the 3' end of the *Y. lipolytica* LIP2 signal sequence (pre), followed by the coding sequence of two Xxx-Ala cleavage sites and flanked by BamHI and AvrII restriction sites

for cloning in expression vector. The construct is under the control of the inducible POX2 promoter. The complete amino acid sequence of the fusion construct is shown on FIG. 27B.

A general scheme of an expression vector is presented in FIG. 28. The bacterial moiety is derived from the plasmid pHSS6, and comprises a bacterial origin of replication (ori) and the kanamycin-resistant gene conferring resistance to kanamycin (KanR). The integration cassette comprises a) the selection marker for transformation to *Yarrowia lipolytica* (URA3; LEU2; GUT2), b) the expression cassette composed of a promoter, c) a multiple cloning site (MCS) to insert huGAA in frame with signal sequence and d) the terminator of the LIP2 gene. The integration cassette is flanked by zeta sequences for stable non-homologous integration into the *Y. lipolytica* genome. Two NotI restriction sites enable the isolation of the expression cassette before transformation. Plasmids pRAN034, pRAN036 and OXYP183 have been used to generate huGAA expression vectors pRAN058, pRAN059 and pRAN060, respectively, containing URA3, LEU2 and GUT2 transformation markers, respectively.

Tandem YIMNN4 expression vector: The YIMNN4 gene was cloned under control of the inducible pPOX2 promoter and the (semi)constitutive hp4d promoter. These two expression cassettes of YIMNN4 were subcloned in one vector as a tandem construct carrying flanking regions (PT) of the ADE2 gene for targeted integration into the ADE2 locus of the genome and the ADE2 gene as a selection marker.

Intermediate Strain OXY1569: The first transformation was a co-transformation of the expression cassette purified from pRAN058 and pRAN059 vectors using URA3 and LEU2 marker to produce intermediate recombinant strain OXY1569. OXY1569 carries two expression constructs of huGAA under control of the pPOX2 promoter randomly integrated in the genome of strain G014.

OXY1569 was selected as follows. PCR screening of genomic DNA was performed in order to confirm the integration of the foreign huGAA DNA into the genome of *Y. lipolytica*. Primers were designed to amplify a fragment of 2552bp from huGAA nucleotide sequence. Southern blot analysis of the genomic DNA also was performed in order to confirm the integration of at least 2 copies of huGAA DNA. In

particular, genomic DNA from OXY1569 clones were digested with Hind III and probed with huGAA DIG labeled specific probe.

In order to select a clone secreting high levels of huGAA, several randomly selected clones that were identified as positive in the PCR screening and Southern blot were grown in shake flasks under POX2 inducing conditions according to a standard procedure. In all cases, the culture supernatant was collected 72h post-induction and screened in a standard Western blot and enzyme activity assay analysis. N-Glycan analysis of OXY1569 indicated the predominant structure in OXY1569 is $\text{Man}_8\text{GlcNAc}_2$.

Intermediate Strain OXY1584: Recombinant strain OXY1569 was transformed in order to integrate two copies of the *Y. lipolytica* MNN4 gene into its genome to produce OXY1584. The transformation was performed with a SacII/XmaI derived expression cassette excised from plasmid OXP1479B. The expression cassette was designed for targeted integration into the ADE2 locus of *Y. lipolytica* genome. The recombinant strain was selected after Southern blotting and glycan analysis to evaluate the strain behavior with respect to the increased phosphorylation. Genomic DNA of several arbitrary chosen transformants was SpeI digested and probed with MNN4 specific DIG labeled probe. Correct targeted integration of MNN4 expression cassette into the ADE2 locus of *Y. lipolytica* genome should give 4207bp and 5683bp bands. Southern blot positive clones were grown in a standard shake flask procedure. N-glycan analysis of secreted proteins was performed in order to select the intermediate clone OXY1584. Compared to the parent strain OXY1569, the predominant structures after MNN4 over-expression are $\text{Man}_8\text{GlcNAc}_2(\text{PMan})_1$ and $\text{Man}_8\text{GlcNAc}_2(\text{PMan})_2$.

Production strain OXY1589: To generate the final prototrophic production strain OXY1589, a third copy of huGAA was integrated into the genome of recombinant OXY1584 strain. The transformation was performed with Not I excised expression cassette from pRAN069. Transformants were first screened by PCR on gDNA for presence of the additional copy of huGAA. To evaluate huGAA production arbitrary selected PCR positive clones were further analyzed for expression after a standard shake flask cultivation. The clone expressing the highest level of huGAA (OXY1589) was chosen after Western blot analysis and enzymatic activity assay. It also was reconfirmed

that the conversion levels of M8 to MP2-M8 and MP-M8 N-glycans was not influenced by the presence of the additional huGAA expression cassette.

EXAMPLE 11

Fed Batch Cultivation of Strain OXYY1589

To produce huGAA from strain OXYY1589 (Example 10), a fed batch process was established using a 10 L stirred tank, with a working volume of 6-8 liters. The process was divided in two phases:

- 1) Batch growth on glucose for biomass formation
- 2) Product formation by induction with help of a limited oleic acid feed.

Typically the batch phase was about 20 hours (h) and the production phase approximately 72 hours. At the end of the process, the culture broth was centrifuged and the supernatant was collected. The supernatant was used as starting material for the purification of the GAA (see Example 12).

The following parameters were controlled during the fermentation. Aeration was maintained at a constant value of 1.5 vvm air (volume per volume per minute). Dissolved oxygen (DO) was initially kept at 30%. The stirring was increased from 600 to 1200rpm depending on the DO levels. Once it reached the maximum of 1200 rpm, this speed was kept constant and the DO-setpoint was set to 10%. To maintain 10% DO, oxygen was spiked into the reactor with a maximal percentage of 50%. Foam evolution was controlled by a foam probe. In case of foam detection, antifoam was added to the bioreactor. The pH was controlled by adding 14% (v/v) ammonia (base) or 10% phosphoric acid to maintain a constant value of pH 6.8. The temperature was kept constant at 28°C throughout the whole process.

Biomass was monitored by measurement of optical density at 600 nm (OD600). The samples were diluted 2 – 1000 times in distilled water to obtain values in the linear range of the spectrophotometer. Product formation was detected by Western blot analysis and specific enzymatic activity tests.

EXAMPLE 12

Purification of recombinant huGAA (rhGAA)

The supernatant after cultivation (see Example 11) was clarified via depth filtration. The resulting material was then concentrated 20 times via TFF and diafiltered against 20 mM sodium phosphate pH 6 and 100 mM NaCl on a 10kDa MWCO membrane (Millipore).

Purification of rhGAA was start by adding ammonium sulphate up to a concentration of 1 M. After centrifugation, the supernatant was loaded on a Toyopearl-Phenyl 650M (Tosoh Biosciences) packed XK16/40 column. A linear gradient from 1 to 0 M ammonium sulphate was applied for elution. Those fractions that contain rhGAA were then pooled and subjected to a buffer exchange into 10 mM BIS-TRIS pH 6. Further purification was achieved via anion exchange chromatography on a source 30Q packed Tricorn 10/50 or XK25/20 column (GE Healthcare) using a linear salt gradient from 0 to 1 M NaCl. The resulting GAA-containing fractions were then concentrated before loading onto a final Hiload 16/60 superdex 200 gel filtration column (GE Healthcare) that was pre-equilibrated with 50 mM sodium phosphate pH 6 and 200 mM NaCl. Fractions were selected on the basis of specific activity and purity on Coomassie-stained SDS-PAGE gels and then combined and concentrated to a final concentration of 5-10 mg/ml. Protein concentration was done on 15 ml Amicon Ultra centrifugal devices (Millipore) with a MWCO of 10 kDa.

The reactions for the qualitative screening for rhGAA were started by adding the reaction buffer consisting of 0.35 mM 4-MUG, 0.1% BSA and 100 mM sodium acetate pH 4 in a 10:1 or 20:1 volume proportion to 10 or 5 µl of elution fraction. All reactions were done in 96-well flat-bottom microtiter plates. After an incubation period of 30 minutes to 1 hour at 37°C, an equal volume of 100 mM glycine pH11 was added to stop the reaction and the release of the fluorogenic reaction product 4-methylumbelliferone was observed under UV-light. Specific activities (units/mg protein) were determined using a colorimetric assay with the synthetic substrate p-nitrophenyl- α -D-glucopyranoside (PNPG) that measures the enzymatic release of the yellow coloured p-nitrophenolate reaction product. The reactions were started by mixing 10 µl of enzyme solution and 90 µl of substrate reaction buffer (2 mM PNPG in 150mM citrate-phosphate buffer pH4, 1% BSA) in reaction wells of a microtiterplate and were subsequently incubated at 37°C. After 1 to 2 hours an equal volume of stop buffer, 10% sodium

carbonate pH 12, was added to quench the reaction and bring the released p-nitrophenol (PNP) in its ionized state. Background-corrected absorbances and p-nitrophenolate standards were measured at a wavelength of 405 nm and specific activities were calculated. Protein concentrations were determined with the bicinchoninic acid (BCA) method. One unit was defined as the amount of enzyme that catalyzes the conversion of 1 nmol of PNPG to 1 nmol PNP and D-glucose per min at 37°C at a final substrate concentration of 2 mM in a citrate-phosphate buffer, pH 4.0.

EXAMPLE 13

Phosphate uncapping activity of heterologously expressed CcMan5 on glycoproteins expressed in a *Y. lipolytica* strain with a higher degree of phosphorylated N-glycans

The huGAA was expressed in *Y. lipolytica* strain OXY1589 to yield a glycoprotein with a high degree of phosphorylated N-glycan structures (see Example 10). The huGAA was purified as described in Example 12.

CcMan5 (1 and 5 μ l respectively at a concentration of 70 μ g/ml) was added to a solution of 4 μ g huGAA in 100 mM HEPES buffer pH 7.0 with 2 mM CaCl_2 . The 20 μ l reaction mixture was incubated overnight at room temperature. The N-glycans were released with PNGaseF, labelled with APTS and subsequently analysed on DSA-FACE, essentially as described in Laroy W. *et al.*, *Nature Protocols*, 1: 397-405 (2006). The N-glycan profiles before and after CcMan5 treatment are shown in Figure 29. The N-glycan mixture released from purified huGAA is mainly composed of ManP-Man8GlcNAc₂ and (ManP)₂-Man8GlcNAc₂ (FIG. 29, panel B). A peak running slightly faster than ManP-Man8GlcNAc₂ can be assigned to ManP-Man7GlcNAc₂. Only very minor amounts of Man₈GlcNAc₂ and Man₇GlcNAc₂ are present. After incubation of huGAA with CcMan5 the conversion of ManP-Man8GlcNAc₂ and (ManP)₂-Man8GlcNAc₂ to P-Man8GlcNAc₂ and P₂-Man8GlcNAc₂ respectively is observed (FIG. 29, panel C and D). The peak in the electropherogram running between P-Man8GlcNAc₂ and P₂-Man8GlcNAc₂ corresponds to the partially uncapped bi-phosphorylated (ManP)₂-Man8GlcNAc₂ with a phosphodiester- and a phosphomonoester-linkage present ((MP)-M8-P in FIG. 29, panel

C and D). This product is further hydrolyzed to the fully uncapped P2-Man8GlcNAc₂ when using a higher concentration of CcMan5 or a longer incubation time.

The percentage of phosphorylated N-glycans versus neutral N-glycans was estimated from measuring the peak areas in the DSA-FACE electropherograms (FIG. 29). The figures related to the area under the curve are presented for the different N-glycans present on huGAA before (Panel B) and after CcMan5 treatment (Panel D). In huGAA (Panel B), (ManP)₂-Man8GlcNAc₂ (11597), ManP-Man6GlcNAc₂ (1261), ManP-Man7GlcNAc₂ (5901), ManP-Man8GlcNAc₂ (15576), Man6GlcNAc₂ (680), Man7GlcNAc₂ (1716), Man8GlcNAc₂ (1572) were present. Approximately 90 % of the N-glycans on recombinant huGAA were composed of mannose-phosphate containing structures.

After an overnight treatment of recombinant huGAA with CcMan5 (Panel D), P₂-Man8GlcNAc₂ (16182), (ManP)P-Man8GlcNAc₂ (1997), P-Man7GlcNAc₂ (8254), P-Man8GlcNAc₂ (17893), ManP-Man6GlcNAc₂ (500), ManP-Man7GlcNAc₂ (2495), ManP-Man8GlcNAc₂ (1326), Man6GlcNAc₂ (1097), Man7GlcNAc₂ (2143), Man8GlcNAc₂ (1599) were present. The N-glycans released from huGAA were composed of 83 % uncapped phosphorylated structures, 8 % is still mannose-phosphate capped and 9 % neutral N-glycans are present. The percentage of uncapped phosphorylated structures can be increased when using a higher concentration of CcMan5 or a longer incubation time.

EXAMPLE 14

Identification of Homologs Likely to Have Uncapping Activity

To identify other GH92 family members with similar predicted catalytic site topology and functionality, curated GH92 family members, as mined from the world wide web at cazy.org/GH92_all.html, were analyzed as were the top 500 hits obtained by Blastp search with the CcMan5 domain sequence on the Non Redundant Protein Sequences database at NCBI. Subsequently, these 392 sequences were used as the input for the multiple sequence alignment package MUSCLE (MUltiple Sequence Comparison by Log-Expectation), which also ranks the sequences in order of 'phylogenetic' distance (from closest related to furthest related).

Based on the curated GH92 family members from the Cazy database, MUSCLE alignment of all GH92 protein sequences (392) and the CcMan5 domain sequence identified the following as the closest homologs of CcMan5:

Streptomyces coelicolor CAA18915 (GenBank Accession No. NP_630514)

Clostridium spiroforme (GenBank Accession No. ZP_02866543)

Bacteroides thetaiotaomicron AAO78636 (GenBank Accession No. NP_812442)

Zunongwangia profunda ADF52306 (GenBank Accession No. YP_003584502)

Chitinophaga pinensis ACU58463 (GenBank Accession No. YP_003120664)

Their sequences and those of the next 5 closest homologs are aligned in FIG. 31.

Based on MUSCLE alignment of the 500 best scoring blastp protein hits versus the CcMan5 domain, the following were considered the closest homologs of CcMan5

Streptomyces coelicolor (GenBank Accession No. NP_630514)

Streptomyces lividans (GenBank Accession No. ZP_05522540)

Streptomyces lividans (GenBank Accession No. ZP_06527366)

Paenibacillus sp (GenBank Accession No. YP_003013376)

Bacteroides thetaiotaomicron (GenBank Accession No. NP_812442)

Bacteroides sp. (GenBank Accession No. ZP_04848482)

Bacteroides cellulosilyticus (GenBank Accession No. ZP_03677957)

Zunongwangia profunda (GenBank Accession No. YP_003584502)

Leeuwenhoekiella blandensis (GenBank Accession No. ZP_01061975)

Sphingobacterium spiritivorum (GenBank Accession No. ZP_07083984)

Chitinophaga pinensis (GenBank Accession No. YP_003120664)

Pedobacter sp. (GenBank Accession No. ZP_01885202)

Clostridium spiroforme (GenBank Accession No. ZP_02866543)

Alignment of these and the 5 next-best homologs can be found in FIG. 32. All 5 best hits from the annotated GH92 database are also found in these 13 best hits from the Blast search on the entire sequence database.

The top 5 hits in FIG. 31 and the top 13 hits in FIG. 32 uniquely share the following three motifs, which were shown in the crystal structure of Example 8 to be different from the alpha-1,2-mannosidase GH92 family members of which the structure was reported in Zhu *et. al.*, 2010, *supra*.

1) a glycine-rich motif GVGxxGxGG, with each X being G, S, T, V, A, C or Q (small side chains), numbering of crystal structure residues of CcMan5 domain: 69-77. This region makes a loop that provides essential hydrogen bonds to the -1 and phosphate-binding subsite in the active site of the enzyme.

2) a VRxE motif. The R makes a hydrogen bond to the -1 ring and possibly the +1 ring. E is in a salt bridge to this R residue, probably shaping this motif. x is W in the closest-related subfamily (top 3 homologs to CcMan5), or could be any of the 20 amino acids except P. This motif is found at residues 404-407 of SEQ ID NO:50.

3) a LYQGT motif, containing the Q which is an E in the mannosidases (proton donor), and which contains Y535, which is important for the +1 site formation. In some of the sequences, the L is A or Y and could reasonably be expected to also be I, V, A, F or M, and in some of them the T is N and can be expected to also tolerate S. Two *Caulobacter* sequences have an E instead of Q and would thus be predicted not to work on phosphorylated glycans.

4) a GDXGN motif. The D and N make part of the substrate binding cavity and may shape an alternative subpocket to bind the +1 mannose. X can be any amino acid other than P. This motif is found at residues 21-25 of SEQ ID NO:50.

Based on the above bioinformatics workflow and motif search based on the structure, it is thus possible to filter the GH92 sequences present in the non-redundant proteins sequence database (currently containing over 1220 sequences) for those rare family members that are good candidates for having the same substrate specificity to CcMan5, i.e., to be capable of uncapping Man-6-Pi-Man structures. In particular, the 3 sequences from *Streptomyces coelicolor* and *Streptomyces lividans* are similar to CcMan5, not only in the above motifs but also in many of the loops of the structure.

A search with Hidden Markov Models based on the sequence elements unique to CcMan5 and its closest homologs, reveals no further GH92 sequences which contain all of these elements, strongly indicating that no such GH92 members have obtained these elements through convergent evolution (these are the ones that would not be top-ranked in multiple-sequence alignments).

EXAMPLE 15

The presence of phosphate uncapping activity in GH92 glycosidases from *Bacteroides thetaiotaomicron*

An enzymatic analysis of 23 family GH92 α -mannosidases from *Bacteroides thetaiotaomicron* has been reported by Zhu, Y. *et al*, 2010, *supra*. Enzymes with α 1,2-, α 1,4-, α 1,3- or α 1,6-mannosidase activity are present in this group of enzymes, although some variants display very low activity. The three-dimensional structure of two α -1,2-mannosidases (Bt3990 and Bt2199) allowed to identify key amino-acid residues which seem to be a signature motif for α -1,2-mannosidase activity, i.e. His584-Glu585 and Trp99 in Bt3990. The activity on phosphorylated N-glycans (MNN4 sugars described in Example 1) of three GH92 enzymes from *B. thetaiotaomicron*, Bt3530 (Genbank nr AAO78636.1), Bt3965 (Genbank nr AAO79070.1) and Bt3994 (Genbank nr AAO79099.1) was tested. These enzymes display low α 1,4-mannosidase activity and lack the His-Glu and Pro-Trp motif.

Bt3530, Bt3965 and Bt3994 were expressed in *E.coli* and purified as described in Zhu *et al*, 2010, *supra*. Samples (1 μ l enzyme at a concentration of 0.1 mg/ml) were incubated with 7 μ l APTS-labeled MNN4 sugars dissolved in 10 mM HEPES buffer pH 7.0 with 2 mM CaCl_2 in an overnight assay at room temperature. A control assay with CcMan5 was included. To confirm the presence of a terminal phosphate the reaction mixture was incubated with CIP. An N-glycan preparation containing Man8GlcNAc₂ (M8) and the monophosphorylated ManP-Man8GlcNAc₂ (MP-M8) was used as substrate. No uncapping activity for Bt3530, Bt3965 and Bt3994 was detected under the above assay conditions. No shift in electrophoretic mobility of the peaks was observed compared to the CcMan5 control reaction (appearance of fast running P-M8 peak), followed by CIP treatment (disappearance of P-M8).

In an additional experiment, 1 μ l of enzyme, i.e. Bt3530 (0.1 mg/ml), Bt3965 (4.75 mg/ml) and Bt3994 (1.37 mg/ml) respectively, was incubated with MNN4 N-glycans at pH 7.0 (10 mM HEPES buffer pH 7.0 with 2 mM CaCl_2) and at pH 5.0 (10 mM Ammonium Acetate pH 5.0 with 2 mM CaCl_2) during 60 hours at room temperature. Very minor α 1,2-mannosidase activity was observed with Bt3530 at pH 7.0, as a small Man5GlcNAc₂ (M5) peak appears in the electropherogram. At pH 5.0, on

the other hand, no α 1,2-mannosidase activity is present, but a fast running peak at the left hand side of the electropherogram appears. This peak has the same electrophoretic mobility as P-Man8GlcNAc₂ (P-M8) and the terminal phosphate is hydrolyzed after incubation with CIP. CcMan5 (used at same concentration as Bt3530) is fully uncapping ManP-Man8GlcNAc₂ within 20 hours incubation at room temperature and at pH 7.0; therefore the observed activity of Bt3530 is rather low. After purification, the Bt3530 sample slowly precipitates when stored at 4 °C in 20 mM TRIS buffer, pH 8.0 with 300 mM NaCl. Therefore it is possible that instability of the Bt3530 protein influences the activity under the assay conditions used. Bt3965, which was used at a 40 times higher concentration, gave a similar result as Bt3530 at pH 7.0 (Panel G and H) and pH 5.0 (Panel I and J). No activity at all was observed with Bt3994 under the same reaction conditions (Panel K till N).

From these experiments can be concluded that phosphate uncapping activity is only a minor side activity of two of the three *B. thetaiotaomicron* GH 92 enzymes tested on MNN4 sugars.

OTHER EMBODIMENTS

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. A method for uncapping a mannose-6-phosphate residue on an oligosaccharide, said method comprising
 - a) providing said oligosaccharide having a mannose-1-phospho-6-mannose linkage; and
 - b) contacting said oligosaccharide with a mannosidase capable of hydrolyzing said mannose-1-phospho-6-mannose linkage to phospho-6-mannose.
2. The method of claim 1, wherein said mannosidase comprises an amino acid sequence having at least 90% identity to the amino acid sequence set forth in residues 1 to 774 of SEQ ID NO:50 or to the amino acid sequence set forth in SEQ ID NO:50.
3. The method of claim 1, wherein said mannosidase comprises an amino acid sequence having at least 95% identity to the amino acid sequence set forth in residues 1 to 774 of SEQ ID NO:50 or to the amino acid sequence set forth in SEQ ID NO:50.
4. The method of claim 1, wherein said mannosidase comprises an amino acid sequence having at least 98% identity to the amino acid sequence set forth in residues 1 to 774 of SEQ ID NO:50 or to the amino acid sequence set forth in SEQ ID NO:50.
5. The method of claim 1, wherein for said mannosidase, the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.
6. The method of claim 1, wherein said mannosidase comprises an amino acid sequence having (i) a GVGXXGXGG motif, where X is Gly, Ala, Ser, Thr, or Cys; (ii) a VRXE motif, where X is any amino acid other than Pro; (iii) an X₁YQGX₂ motif, where X₁ is Leu, Ile, Val, Ala, Phe, Tyr or Met, and X₂ is Thr, Ser, or Asn; or (iv) GDXGN, where X can be any amino acid other than Pro.

7. The method of any one of claim 1-6, wherein said contacting step is performed using a purified mannosidase, a recombinant mannosidase, a cell lysate containing said recombinant mannosidase, or a fungal cell containing said recombinant mannosidase.
8. The method of any one of claims 1-7, wherein said oligosaccharide is attached to a protein.
9. The method of claim 7, wherein said protein is a human protein expressed in a fungal organism.
10. The method of claim 9, wherein said fungal organism is *Yarrowia lipolytica* or *Arxula adenivorans*.
11. The method of claim 9, wherein said fungal organism is a methylotrophic yeast.
12. The method of claim 11, wherein said methylotrophic yeast is *Pichia pastoris*, *Pichia methanolica*, *Oogataea minuta*, or *Hansenula polymorpha*.
13. The method of claim 9, wherein said fungal organism is a filamentous fungus.
14. The method of claim 13, wherein said filamentous fungus is selected from the group consisting of *Aspergillus caesiellus*, *Aspergillus candidus*, *Aspergillus carneus*, *Aspergillus clavatus*, *Aspergillus deflectus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus glaucus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus ochraceus*, *Aspergillus oryzae*, *Aspergillus parasiticus*, *Aspergillus penicilloides*, *Aspergillus restrictus*, *Aspergillus sojae*, *Aspergillus sydowi*, *Aspergillus tamari*, *Aspergillus terreus*, *Aspergillus ustus*, and *Aspergillus versicolor*.
15. The method of claim 8, wherein said protein is a pathogen protein, a lysosomal protein, a growth factor, a cytokine, a chemokine, an antibody or antigen-binding fragment thereof, or a fusion protein.

16. The method of claim 15, wherein said lysosomal protein is a lysosomal enzyme.
17. The method of claim 16, wherein said lysosomal enzyme is associated with a lysosomal storage disorder (LSD).
18. The method of claim 17, wherein said LSD is Fabry's disease, mucopolysaccharidosis I, Farber disease, Gaucher disease, GM1-gangliosidosis, Tay-Sachs disease, Sandhoff disease, GM2 activator disease, Krabbe disease, metachromatic leukodystrophy, Niemann-Pick disease, Scheie disease, Hunter disease, Sanfilippo disease, Morquio disease, Maroteaux-Lamy disease, hyaluronidase deficiency, aspartylglucosaminuria, fucosidosis, mannosidosis, Schindler disease, sialidosis type 1, Pompe disease, Pycnodysostosis, ceroid lipofuscinosis, cholesterol ester storage disease, Wolman disease, Multiple sulfatase deficiency, galactosialidosis, mucopolipidosis, cystinosis, sialic acid storage disorder, chylomicron retention disease with Marinesco-Sjögren syndrome, Hermansky-Pudlak syndrome, Chediak-Higashi syndrome, Danon disease, or Glucocorticoid deficiency.
19. The method of claim 18, wherein said LSD is Pompe disease or Fabry's disease.
20. The method of any one of claims 1-6, wherein said mannosidase comprises a targeting sequence.
21. A method of producing a target protein having terminal phospho-6-mannose residues, said method comprising:
- providing a fungal cell genetically engineered to comprise a nucleic acid encoding a mannosidase, said mannosidase capable of hydrolyzing a mannose-1-phospho-6-mannose linkage to phospho-6-mannose; and
 - introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said terminal phospho-6-mannose residues.

22. The method of claim 21, wherein for said mannosidase, the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.

23. The method of claim 21, wherein said mannosidase comprises a) an amino acid sequence having at least 90% sequence identity to amino acids 1 to 774 of SEQ ID NO:50 or to the amino acid sequence of SEQ ID NO:50; or (b) an amino acid sequence having (i) a GVGXXGXGG motif, where X is Gly, Ala, Ser, Thr, or Cys; (ii) a VRXE motif, where X is any amino acid other than Pro; (iii) an X_1YQGX_2 motif, where X_1 is Leu, Ile, Val, Ala, Phe, Tyr or Met, and X_2 is Thr, Ser, or Asn; or (iv) GDXGN, where X can be any amino acid other than Pro.

24. The method of any one of claims 21-23, wherein said fungal cell further comprises a nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation.

25. The method of any one of claims 21-24, wherein said fungal cell is genetically engineered to be deficient in OCH1 activity.

26. The method of any one of claims 21-25, wherein said mannosidase is a *C. cellulans*, *Streptomyces coelicolor*, or *Streptomyces lividans* mannosidase.

27. A method of producing a target protein having terminal phospho-6-mannose residues in a fungal organism, said method comprising

a) providing a fungal cell genetically engineered to comprise a nucleic acid encoding a mannosidase capable of hydrolyzing a mannose-1-phospho-6-mannose linkage to phospho-6-mannose, and said fungal cell further comprising a nucleic acid encoding a target protein; and

b) isolating said target protein having said terminal phospho-6-mannose residues.

28. The method of claim 27, wherein for said mannosidase, the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.

29. The method of claim 27, wherein said mannosidase comprises (i) an amino acid sequence having at least 90% sequence identity to amino acids 1 to 774 of SEQ ID NO:50 or to the amino acid sequence set forth in SEQ ID NO:50; or (ii) an amino acid sequence having (i) a GVGXXGXGG motif, where X is Gly, Ala, Ser, Thr, or Cys; (ii) a VRXE motif, where X is any amino acid other than Pro; (iii) an X_1YQGX_2 motif, where X_1 is Leu, Ile, Val, Ala, Phe, Tyr or Met, and X_2 is Thr, Ser, or Asn; or (iv) GDXGN, where X can be any amino acid other than Pro.

30. The method of any one of claims 27-29, wherein said target protein and said mannosidase are co-secreted.

31. An isolated fungal cell genetically engineered to produce glycoproteins comprising terminal phospho-6-mannose residues, said fungal cell comprising a nucleic acid encoding a mannosidase, wherein expression of said mannosidase in said fungal cell produces glycoproteins comprising said terminal phospho-6-mannose residues.

32. The fungal cell of claim 31, wherein for said mannosidase, the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.

33. The fungal cell of claim 31, wherein said mannosidase comprises a) an amino acid sequence having at least 90% sequence identity to amino acids 1 to 774 of SEQ ID NO:50 or to the amino acid sequence of SEQ ID NO:50; or b) an amino acid sequence having (i) a GVGXXGXGG motif, where X is Gly, Ala, Ser, Thr, or Cys; (ii) a VRXE motif, where X is any amino acid other than Pro; (iii) an X_1YQGX_2 motif, where X_1 is

Leu, Ile, Val, Ala, Phe, Tyr or Met, and X₂ is Thr, Ser, or Asn; or (iv) GDXGN, where X can be any amino acid other than Pro.

34. The fungal cell of any one of claims 31-33, said fungal cell further comprising a nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation.

35. The fungal cell of any one of claims 31-33, wherein said fungal cell is genetically engineered to be deficient in OCH1 activity.

36. The fungal cell of any one of claims 31-33, said fungal cell further comprising a nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation, and wherein said fungal cell is genetically engineered to be deficient in OCH1 activity.

37. The fungal cell of any one of claims 31 to 36, said fungal cell further comprising a nucleic acid encoding a target protein, wherein said target protein is a glycoprotein.

38. The fungal cell of any one of claims 31 to 36, wherein said target protein is a human protein.

39. The fungal cell of claim 38, wherein said target protein is a pathogen protein, a lysosomal protein, a growth factor, a cytokine, a chemokine, an antibody or antigen-binding fragment thereof, or a fusion protein.

40. The fungal cell of claim 39, wherein said lysosomal protein is a lysosomal enzyme.

41. The fungal cell of claim 40, wherein said lysosomal enzyme is acid alpha glucosidase or alpha galactosidase.

42. The fungal cell of claim 38, wherein said target protein is a protein associated with a LSD.

43. The fungal cell of claim 42, wherein said LSD is Fabry's disease, mucopolysaccharidosis I, Farber disease, Gaucher disease, GM1-gangliosidosis, Tay-Sachs disease, Sandhoff disease, GM2 activator disease, Krabbe disease, metachromatic leukodystrophy, Niemann-Pick disease, Scheie disease, Hunter disease, Sanfilippo disease, Morquio disease, Maroteaux-Lamy disease, hyaluronidase deficiency, aspartylglucosaminuria, fucosidosis, mannosidosis, Schindler disease, sialidosis type 1, Pompe disease, Pycnodysostosis, ceroid lipofuscinosis, cholesterol ester storage disease, Wolman disease, Multiple sulfatase deficiency, galactosialidosis, mucopolipidosis, cystinosis, sialic acid storage disorder, chylomicron retention disease with Marinesco-Sjögren syndrome, Hermansky-Pudlak syndrome, Chediak-Higashi syndrome, Danon disease, or Geleophysic dysplasia.

44. The fungal cell of any one of claims 31 to 43, wherein said fungal cell is a *Yarrowia lipolytica* or *Arxula adenivorans* cell.

45. The fungal cell of claim 32, wherein said polypeptide capable of promoting mannosyl phosphorylation is a MNN4 polypeptide.

46. The fungal cell of claim 45, wherein said MNN4 polypeptide is a *Yarrowia lipolytica*, *S. cerevisiae*, *Ogataea minuta*, *Pichia pastoris*, or *C. albicans* polypeptide.

47. The fungal cell of claim 46, wherein said polypeptide capable of promoting mannosyl phosphorylation is a *P. pastoris* PNO1 polypeptide.

48. The fungal cell of any one of claims 31 to 47, wherein said mannosidase is a *C. cellulans*, *Streptomyces coelicolor*, or *Streptomyces lividans* mannosidase.

49. The fungal cell of any one of claims 31 to 47, wherein said mannosidase comprises a secretion signal.

50. The fungal cell of any one of claims 31 to 47, wherein said mannosidase comprises a targeting signal to target said mannosidase to an intracellular compartment.

51. The fungal cell of any one of claims 31 to 47, wherein said mannosidase comprises a secretion signal and a targeting signal to target said mannosidase to an intracellular compartment.

52. A substantially pure culture of *Yarrowia lipolytica*, *Pichia pastoris*, *Hansenula polymorpha*, *Arxula adeninivorans*, *Pichia methanolica*, *Oogataea minuta*, or *Aspergillus niger* cells, a substantial number of which are genetically engineered to produce glycoproteins comprising a terminal phospho-6-mannose residue, said cells comprising a nucleic acid encoding a mannosidase capable of hydrolyzing a mannose-1-phospho-6-mannose linkage to phospho-6-mannose.

53. The culture of claim 52, wherein for said mannosidase, the three dimensional protein coordinates of the atoms in the amino acid side chains located in the minimal catalytic center fall within 1.5 Å deviation of the coordinates of the equivalent atoms in FIG. 33.

54. The culture of claim 52, wherein said mannosidase comprises a) an amino acid sequence having at least 90% sequence identity to amino acids 1 to 774 of SEQ ID NO:50 or to the amino acid sequence of SEQ ID NO:50; or b) an amino acid sequence having (i) a GVGXXGXGG motif, where X is Gly, Ala, Ser, Thr, or Cys; (ii) a VRXE motif, where X is any amino acid other than Pro; (iii) an X₁YQGX₂ motif, where X₁ is Leu, Ile, Val, Ala, Phe, Tyr or Met, and X₂ is Thr, Ser, or Asn; or (iv) GDXGN, where X can be any amino acid other than Pro.

55. The culture of any one of claims 52-54, wherein said cells further comprise a nucleic acid encoding a polypeptide capable of promoting mannosyl phosphorylation.

56. The culture of claim 55, wherein said cells are genetically engineered to be deficient in OCH1 activity.

57. An isolated glycoprotein comprising terminal phospho-6-mannose residues, wherein the protein is produced by the method of any one of claims 1-30.

58. A composition comprising a glycoprotein, wherein at least 47% of the N-glycans on said glycoprotein have terminal phospho-6-mannose residues.

59. The composition of claim 58, wherein at least 50% of the N-glycans on said glycoprotein have terminal phospho-6-mannose residues.

60. The composition of claim 58, wherein at least 75% of the N-glycans on said glycoprotein have terminal phospho-6-mannose residues.

61. The composition of claim 58, wherein at least 90% of the N-glycans on said glycoprotein have terminal phospho-6-mannose residues.