TITLE OF THE INVENTION

METHOD FOR PRODUCING THERAPEUTIC PROTEINS IN *PICHIA PASTORIS* LACKING DIPEPTIDYL AMINOPEPTIDASE ACTIVITY

FIELD OF THE INVENTION

The present invention is directed to methods and compositions for the production of glycoproteins in yeast cell lines having no dipeptidyl aminopeptidase (DAP) activity, which are useful as human or animal therapeutic agents.

BACKGROUND OF THE INVENTION

Yeast are important production platforms for the generation of recombinant proteins. In that yeasts are eukaryotes, they share common evolutionary processes with those of higher eukaryotes, including many of the post-translational modifications that occur in the secretory pathway. Recent advances in glycoengineering have resulted in cell lines of the yeast strain *Pichia pastoris* with genetically modified glycosylation pathways that allow them to carry out a sequence of enzymatic reactions, which mimic the process of glycoproteins in humans. See, for example, US Pat. Nos. 7,029,872 and 7,326,681 that describe methods for producing a recombinant glycoprotein in a lower eukaryote host cell that are substantially identical to their human counterparts. Human-like sialylated bi-antennary complex N-linked glycans like those produced in *Pichia pastoris* from the aforesaid methods have demonstrated utility for the production of therapeutic glycoproteins.

Similar to higher eukaryotes, yeast also express numerous proteases, many of which are either localized to the secretory pathway or pass through it on route to their final destination. As a consequence, non-desirable proteolysis of some recombinant proteins may occur with the specific cleavage being dependent on the class of protease involved. Dipeptidyl aminopeptidases (DAPs) are a class of proteolytic enzymes which remove a two amino acid peptide from the N-terminus of a protein. In *Saccharomyces cerevisiae* genes for the enzymes *STE13* and *DAP2* have been identified as having DAP activity, see, Julius et al., Cell, 32: 839-852, 1983; Rendueles et al., J. Bacteriology, 169: 4041-4048, 1987. Applicants herein have developed methods for the elimination of DAP activity in *Pichia pastoris*, which will allow for the production of full length therapeutic proteins.

SUMMARY OF THE INVENTION

In one embodiment the invention herein is a method for producing therapeutic proteins in yeast cell lines lacking dipeptidyl aminopeptidase (DAP) activity. This embodiment comprises transforming a genetically modified *Pichia pastoris* cell line in which the DAP activity has been eliminated with a polynucleotide vector encoding the therapeutic protein and culturing the transformed host cell to produce the therapeutic protein. DAP activity can be eliminated by modifying a *Pichia pastoris* cell line such that *STE13*, *DAP2*, and *DPPIII* have been deleted or disrupted. In another embodiment DAP activity is eliminated by modifying a *Pichia pastoris* cell line such that *STE13* and *DAP2* have been deleted or disrupted.

In one embodiment, the invention is a genetically modified yeast cell line lacking DAP activity that can be used for the production of therapeutic proteins comprising a *Pichia pastoris* cell line that has been recombinantly modified through the deletion of *STE13* and *DAP2*.

In still other embodiments, the invention is a method for the production of therapeutic proteins from *Pichia pastoris*, such as TNFRII-Fc, a recombinant fusion protein comprising the ectodomain of tumor necrosis factor receptor 2 (TNFRII) fused to an IgG1 Fc domain (TNFRII-Fc), or a recombinant granulocyte colony-stimulating factor (GCSF) polypeptide, comprising the secreted plasma form of GCSF.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graphic representation of the vectors used in the methods herein. The Ura5 marked Pp*STE13* and Pp*DAP2* knockout vectors are shown as pGLY4520 and pGLY4521 in Figures 1A and 1B, respectively. The Pp*STE13* and Pp*DAP2* nourseothricin marked vectors are shown as pGLY5018 and pGLY5019 in Figures 1C and 1D, respectively. The SfiI restriction sites used to excise the knockout fragments used in the transformations are underlined. Pp*STE13* and Pp*DAP2* flanking regions are highlighted in black.

Figure 2 is a graphic depiction of the N-terminal cleavage of TNFRII-Fc associated with DAP activity in yeast. Figure 2A shows the seven N-terminal amino acids (SEQ ID NO: 1) of intact secreted TNFRII-Fc (SEQ ID NO: 3) and Figure 2B shows the five N-terminal amino acids (SEQ ID NO: 2) of the truncated product. The arrow illustrates the cleavage site recognized by both of the dipeptidyl aminopeptidases, Dap2p and Ste13p.

Figure 3 is the amino acid sequence for the fusion protein, TNFRII-Fc, comprising tumor necrosis factor receptor 2 (TNFRII) and an IgG1 Fc region (Fc) (SEQ ID NO: 3).

Figure 4 is a graphic depiction of the N-terminal cleavage of GCSF associated with DAP activity in yeast. Figure 4A shows the seven N-terminal amino acids (SEQ ID NO: 4) of intact secreted GCSF (SEQ ID NO: 6) and Figure 4B shows the five N-terminal amino acids (SEQ ID NO: 5) of the truncated product. The arrow illustrates the cleavage site recognized by both of the dipeptidyl aminopeptidases, Dap2p and Ste13p.

Figure 5 is the amino acid sequence for a recombinant granulocyte colony-stimulating factor (GCSF) protein (SEQ ID NO: 6).

Figure 6 is a representation of a western blot of GCSF produced in a *Pichia pastoris* strain where neither the *STE13* nor the *DAP2* genes have been deleted or disrupted, i.e. lower molecular weight GCSF due to DAP cleavage (lanes 27-29), as compared to intact higher molecular GCSF produced in a *Pichia pastoris* strain when Ste13p and Dap2p activity has been eliminated (lanes 32-34).

Figures 7A-7C represents a flow diagram for the glycoengineered strain, YGLY7406, used for generation of the knockout of *STE13* and *DAP2* (Example 3). Figures 7C represents a flow diagram for the single knockout glycoengineered strains, *ste13* (YGLY8084) and *dap2* (YGLY8090), as well as the subsequent double knock-out strain (YGLY8096).

Figures 8A and 8B are the cDNA (SEQ ID NO: 37) and amino acid (SEQ ID NO: 38) sequences for *Pichia pastoris DAP2*. ORF is shown in boldface type, +/- approximately 1 kb flanking sequence.

Figures 9A and 9B are the cDNA (SEQ ID NO: 39) and amino acid (SEQ ID NO: 40) sequences for *Pichia pastoris STE13*. ORF is shown in boldface type, +/- approximately 1 kb flanking sequence.

Figures 10A and 10B are the 5' (SEQ ID NO: 41) and 3' (SEQ ID NO: 42) DNA flanking regions for Pp*STE13* amplified to generate pGLY4511 and pGLY4512, respectively. The flanking regions (underlined) are themselves flanked by nucleotides constituting EcoRI restriction sites for the Pp*STE13* 5' flanking region or nucleotides constituting HindIII restriction sites for the Pp*STE13* 3' flanking region.

Figures 11A and 11B are the 5' (SEQ ID NO: 43) and 3' (SEQ ID NO: 44) DNA flanking regions for Pp*DAP2* amplified to generate pGLY4513 and pGLY4514, respectively. The flanking regions (underlined) are themselves flanked by nucleotides constituting EcoRI restriction sites for the Pp*DAP2* 5' flanking region or nucleotides constituting HindIII restriction sites for the Pp*DAP2* 3' flanking region.

Figure 12 is the cDNA sequence for a nourseothricin marker cassette (SEQ ID NO: 45) amplified from pAG25, with the ORF shown in boldface type.

Figure 13 is the cDNA (SEQ ID NO: 54) sequence for the *Pichia pastoris DPPIII.* ORF is shown in boldface type, +/- approximately 1 kb flanking sequence.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

As used herein the term "dipeptidyl aminopeptidase activity" or "DAP activity" refers to the enzymatic cleavage of a polypeptide produced by the genes designated *STE13*, *DAP2*, or *DPPIII.*

As used herein the phrase "elimination of dipeptidyl aminopeptidase activity" or "elimination of DAP activity" refers to the absence of the enzymatic activity produced by the genes designated *STE13, DAP2*, and *DPPIII.*

The term "therapeutic protein" as used herein refers to a full length, i.e. non-truncated form, biologically active polypeptide than can be utilized as a therapeutic to treat a disease or condition in an animal or human. Examples of this term as used herein are the fusion protein TNFRII-Fc, which comprises the tumor necrosis factor receptor 2 (TNFRII) and the Fc region of IgG1 (Fc), and a recombinant granulocyte colony-stimulating factor (GCSF) protein.

The term "N-terminal recognition site" as used herein refers to a polypeptide having an N-terminal sequence of the motif X-Pro or X-Ala, where X is any amino acid and the second position relative to the N-terminal is either a proline or an alanine.

The term "glycoengineered *Pichia pastoris* strain" as used herein refers to a *Pichia pastoris* strain that has been engineered to express human glycoproteins. Representative strains include YJN201 (Choi et al., PNAS, 100 (9): 5022-5027, 2003); YSH44 (Hamilton et al., Science, 301 (5637): 1244-1246, 2003); RDP36-1 (Davidson et al., Glycobiology, 14 (4): 1-9, 2004); PBP6-5 (Bobrowicz et al., Glycobiology, 14 (9): 757-766, 2004); YSH597 (Hamilton et al., Science, 313 (5792): 1441-1443, 2006).

The term "wild type strain" as used herein refers to a glycoengineered *Pichia pastoris* strain in which the genes, *STE13*, *DAP2*, or *DPPIII* have not been altered, disrupted or deleted from the genome.

The term "deletion strain" as used herein refers to a glycoengineered *Pichia pastoris* strain in which the *Pichia pastoris* homologue of one, two or all DAP genes, i.e. *STE13*, and *DAP2*, and *DPPIII*, have been modified at the genomic level to eliminate functional DAP

activity. This includes, but is not limited to, complete or partial deletion of the gene (comprising the promoter, open reading frame and terminator); introduction of one or more mutations that alter transcription or translation of the gene or encoded mRNA, respectively; and introduction of one or more mutations that inactivate the protein activities. An example of such a deletion strain is YGLY8084.

The term "deleted or disrupted" and "deletion or disruption" as used herein refers to the inhibition of the activity of an enzyme, said enzyme produced from a yeast cell genome, in which the inhibition of the enzyme activity is to the extent that a substrate protein has an intact N-terminus. Examples of which are yeast host cells in which enzyme activity can be abrogated or disrupted including, but not limited to, 1) deletion or disruption of the upstream or downstream regulatory sequences controlling expression of the gene; 2) mutation of the gene encoding the enzyme activity to render the gene non-functional, where "mutation" includes deletion, substitution, insertion, or addition into the gene to render it incapable of enzymatic activity; 3) abrogation or disruption of the enzymatic activity by means of a chemical, peptide, or protein inhibitor; 4) abrogation or disruption of the enzymatic activity by means of nucleic acid-based expression inhibitors, such as antisense DNA and siRNA; 5) abrogation or disruption of the enzymatic activity by means of transcription inhibitors or inhibitors of the expression or activity of regulatory factors that control or regulate expression of the gene encoding the enzyme activity; and 6) by any means in which the product obtained, even if expressed, is not identical to the secreted protein and the function is attenuated.

Abbreviations

The following abbreviations are used throughout this description:

URA5	Orotate phosphoribosyltransferase (OPRTase) isozyme
Sc*SUC2*	*S. cerevisiae* Invertase
OCH1	Alpha-1,6-mannosyltransferase
KlGlcNAcTr	*K. lactis* UDP-GlcNAc transporter
BMT2:	Beta-mannose-transfer (beta-mannose elimination)
MNN4B:	MNN4A-like gene (charge elimination)
MmGlcNAcTr	Mouse homologue of UDP-GlcNAc transporter
PNO1:	Phosphomannosylation of *N*-glycans (charge elimination)
MNN4A:	Mannosyltransferase (charge elimination)
ADE1	N-succinyl-5-aminoimidazole-4-carboxamide ribotide synthetase

MNS1	Mouse mannosidase IA catalytic domain fused to Sc*SEC12* leader
GnTI	Human GlcNAc transferase I catalytic domain fused to Pp*SEC12* leader
HIS1	ATP phosphoribosyltransferase
GalTI	Truncated human galactosyltransferase 1 catalytic domain fused to Sc*KRE2* leader
GalE	*S. cerevisiae* UDP-glucose 4-epimerase
UDP-GalTr	UDP-Galactose transporter
ARG1	Arginosuccinate synthetase
MNSII	*Drosophila* mannosidase II catalytic domain fused to Sc*MNN2* leader
GnTII	Rat GlcNAc transferase II catalytic domain fused to Sc*MNN2* leader
PRO1	Gamma-glutamyl kinase
TrMNS1	Secreted *T. reesei* mannosidase I catalytic domain fused to ScαMAT
AOX1	Alcohol oxidase I
TNFRII-Fc	Human Tumour Necrosis Factor Receptor II fused to Fc domain of IgG1
Zeo	Zeocin resistance marker
STE13	Dipeptidyl aminopeptidase
DAP2	Dipeptidyl aminopeptidase
DPPIII	Dipeptidyl aminopeptidase
Nat	Nourseothricin resistance marker

Production of therapeutic proteins in yeast

A significant fraction of proteins isolated from humans or other animals are glycosylated. Among proteins used therapeutically about 70% are glycosylated. If a therapeutic protein is produced in a microorganism host such as yeast and is glycosylated utilizing an endogenous pathway, typically its therapeutic efficiency is greatly reduced. Notwithstanding, such glycoproteins can be immunogenic in humans, and demonstrate reduced half-lives *in vivo* after administration, Takeuchi, Trends in Glycoscience and Glycotechnology, 9: S29-S35, 1997.

Specific receptors in humans and animals can recognize terminal mannose residues and promote the rapid clearance of the protein from the bloodstream. Additional adverse effects may include changes in protein folding, solubility, susceptibility to proteases, trafficking, transport, compartmentalization, secretion, recognition by other proteins or factors, antigenicity, or allergenicity. Accordingly, it has been necessary to produce therapeutic glycoproteins in animal host systems, so that the pattern of glycosylation is identical or at least

similar to that occurring in humans or in the intended recipient species. In most cases a mammalian host system, such as mammalian cell culture, is used. Systems used have included Chinese hamster ovary cells (CHO), transgenic animals, plants, or insect cells. Recombinant human proteins expressed in such systems may still include non-human glycoforms, Raju et al., Annals Biochem., 283(2): 123-132, 2000. All of these systems have significant drawbacks, including but not limited to, that only certain therapeutic proteins are suitable for expression in animal or plant systems, that the cell culture systems are typically very slow, that protein yields are low relative to microbial fermentation processes and may require complex and expensive nutrients or additives, or that such systems may be susceptible to contamination.

The lack of a suitable expression is thus an obstacle to the production of any therapeutic glycoprotein. Production of glycoproteins via the fermentation of microorganisms would offer advantages over the existing systems including, but not limited to, rapid production of high concentrations of protein, the ability to use sterile, well-controlled production conditions or chemically defined growth media, the ability to express a wide variety of proteins and easy recovery of the therapeutic protein. However, as noted above, bacterial systems do not produce glycosylated therapeutic proteins like eukaryotes. To that end, yeast cell lines and in particular, *Pichia pastoris*, with genetically modified glycosylation pathways that produce human glycoproteins have been developed. See, for example, US Pat. Nos. 7,029,872 and 7,326,681 and US 2006-0286637 that describe methods for producing recombinant glycoproteins in a lower eukaryote host cell that are substantially identical to their human counterparts. Human-like sialylated bi-antennary complex N-linked glycans like those produced in *Pichia pastoris* from the aforesaid methods have demonstrated utility for the production of therapeutic glycoproteins.

While the aforesaid *Pichia pastoris* cell lines can produce proteins having a human-like glycosylation pattern making them appropriate for use as therapeutics, expression of therapeutic proteins in *Pichia pastoris* does not necessarily result in the production of the full length polypeptide. Non-desirable enzymatic activity, such as proteolysis that truncates the therapeutic protein, needs to be eliminated in order to use the yeast expression system efficiently.

Methods for producing therapeutic proteins in *Pichia* lacking dipeptidyl aminopeptidase activity

Analysis of the peptide sequence of a recombinant fusion protein TNFRII-Fc, the sequence of which is shown in Figure 3 (SEQ ID NO: 3), secreted from the yeast cell line, *Pichia pastoris*, indicated that all of the peptide produced had been truncated at the N-terminus by two

amino acids. Applicants herein identified that the deletion of two genes, *STE13 and DAP2*, in a *Pichia pastoris* cell line eliminated all DAP activity and resulted in the production full length TNFRII-Fc. Applicants have also identified that *Pichia pastoris* has a third dipeptidyl aminopeptidase, *DPPIII*. Thus, in one embodiment the invention herein is a method for producing therapeutic proteins in yeast cell lines lacking DAP activity. This embodiment comprises transforming a genetically modified *Pichia pastoris* cell line, in which the DAP activity has been eliminated, with a polynucleotide vector encoding a therapeutic protein and culturing the transformed host cell to produce the therapeutic protein. DAP activity is eliminated by modifying the *Pichia pastoris* cell line such that *STE13* and *DAP2* and/or *DPPIII* have been deleted or disrupted. In a specific embodiment of the invention herein, DAP activity was eliminated by modifying the *Pichia pastoris* cell line such that *STE13* and *DAP2* have been deleted or disrupted.

In another embodiment, the invention is a yeast cell line used for the production of therapeutic proteins which comprises a *Pichia pastoris* cell line that has been recombinantly modified such that the *STE13* and *DAP2* genes were deleted or disrupted and all resultant DAP activity was eliminated.

In still other embodiments, the invention is a method for the production from *Pichia pastoris* of a therapeutic protein having the amino acid proline or alanine in the second position relative to the N-terminus of the polypeptide, such as the TNFRII-Fc fusion protein or the GCSF protein described herein.

Dipeptidyl aminopeptidase activity

Genes that are associated with dipeptidyl aminopeptidase (DAP) activity, a subclass of proteases known to truncate two amino acids from the amino terminus of a polypeptide, have been identified in yeast. Non-mating alpha-cell mutants of *S. cerevisiae* have been characterized and mutations in the dipeptidyl aminopeptidase gene *STE13*, have been attributed to incomplete processing of the alpha mating factor pheromone, Julius et al., Cell, 32(3), 839-52, 1983. Furthermore, a second dipeptidyl peptidase, Dap2p, was identified by screening *S. cerevisiae* mutant strains deficient in Ste13p activity, Suarez Rendueles and Wolf, Journal of Bacteriology, 169 (9), 4041-48, 1987. It has also been reported that knockout of the *Pichia* homolog of *S. cerevisiae STE13* prevents *in vivo* proteolytic cleavage of proteins having the amino acids HG (His-Gly) at the N-terminus and allowed production of a full length insulinotropic peptide, Melarkode et al., WO 2007/148345; Prabha et al., Protein Expression and

Purification, 64, 155-161, 2009. Conversely, disruption of the *DAP2 Pichia* homolog of *Saccharomyces cerevisiae* did not prevent N-terminal proteolytic cleavage, Melarkode et al., WO 2007/148345; Prabha et al., Protein Expression and Purification, 64, 155-161, 2009.

Ste13p and Dap2p individually have been shown to cleave N-terminal peptides having the motif X-P/A, where X is any amino acid and the second position is either proline (P) or alanine (A), Misumi and Ikehara, Handbook of Proteolytic Enzymes, 2nd edition, pp. 1910-1911, Elsevier, London, 2004. The TNFRII-Fc and GCSF proteins used herein to illustrate the present invention conform to the conserved X-Pro Ste13p and Dap2p motif, suggesting that for production of therapeutic proteins in *Pichia pastoris*, the DAP activity of Ste13p or Dap2p must be eliminated. In contrast, the insulinotropic peptide of Melarkode et al, WO 2007/148345, possesses a novel non-conserved N-terminus with a His-Gly motif that is only recognized by Ste13p.

As demonstrated herein by Applicants, elimination of *Pichia pastoris* Ste13p activity resulted in only a partial reduction in DAP activity for the TNFRII-Fc protein, with somewhat greater than half of the TNFRII-Fc protein produced having the full length sequence. Applicants surprisingly found that for complete elimination of DAP activity, i.e. for 100% occurrence of full length TNFRII-Fc protein, the activity of both Ste13p and Dap2p must be eliminated from the *Pichia pastoris* cell line. Similarly, only when both DAP genes are eliminated from *Pichia pastoris* did the Applicants observe intact production of a recombinant GCSF protein. Thus, one skilled in the art would recognize and appreciate that the present invention differs from the prior art by requiring the elimination of both Ste13p and Dap2p activity in *Pichia pastoris* for production of full length proteins having the N-terminal X-P/A motif.

Based on homologies amongst known DAP genes, one skilled in the art may design PCR primers, examples of which are shown in Table 1, or use genes or gene fragments as probes to identify homologues in DNA libraries of the target organism. Applicants' analysis of the *Pichia pastoris* genome to identify homologues having DAP activity has resulted in the identification of a homologue for *STE13* and *DAP2*, designated Pp*STE13* and Pp*DAP2*, (Figures 8A and 8B, SEQ ID NOS: 37 and 38, and Figures 9A and 9B, SEQ ID NOS: 39 and 40), respectively. Melarkode et al., WO 2007/148345, describes the DNA sequence for a *Pichia pastoris STE13* homologue, while Prabha et al., Protein Expression and Purification, 64: 155-161, 2009, describes the protein sequences for *Pichia pastoris* Ste13p and Dap2p homologues. The Ste13p homologue generated herein by Applicants appears to be in agreement with that

reported previously in Melarkode et al. While major portions of the Dap2p protein sequence generated herein by Applicants agrees with that reported by Prabha et al., the Dap2p homologue identified herein differs at the C-terminus. Applicants' homologue contains the following C-terminus sequence:

GLENYDIHVFPDSDHSIRYHNGNVIVYDKLFHWIRRAFKAGK (SEQ ID NO: 46).

Whereas the Prabha et al. homologue has the following C-terminus sequence:

GLENYDIHVFPDTIPLD (SEQ ID NO: 47).

The N-terminal sequences up to and including the underlined portion of the C-terminal sequences above are conserved in both homologues.

Table 1 lists sequences for representative primers used for the generation of knockout vectors in *Pichia pastoris*. Regions underlined in the primer sequence represent restriction enzyme recognition sites that have been introduced to facilitate gene knockout fragment generation. The specific restriction enzyme recognition sites introduced are named in the adjacent primer description column. Table 2 lists primers used to confirm knockout of the *Pichia STE13* and *DAP2* from the genome following transformation with the knock-out vectors generated using the primers in Table 1. Successful deletion is confirmed when the 5' and 3' primer sets give PCR products of the desired size and the knockout primer set gives no product.

Table 1

Primer	Sequence (5' to 3')	Description
SH85	GGCTCGAGGATCTGTTTAGCTTGCCTCGTCC (SEQ ID NO: 7)	NAT^R cass *Xho*I for
SH86	GGCTCGAGGGAGCTCGTTTTCGACACTGGATGG (SEQ ID NO: 8)	NAT^R cass *Xho*I rev
SH379	CATGCCCCTGAGCTGCGCACGTCAAG (SEQ ID NO: 9)	pTEF (NAT^R) outwards
SH380	CAGAAAGTAATATCATGCGTCAATCG (SEQ ID NO: 10)	TEF tt (NAT^R) outwards
SH491	GGCGATTACCGTTGATGTTGAAGTGGCGAG (SEQ ID NO: 11)	LacZ 5'-3' screen out
SH558	CATCCAGAGGCACTTCACCGCTTGCCAGCG (SEQ ID NO: 12)	LacZ 3'-5' screen out

SH774	GGAATTCGGCCTTGGGGGCCTCCAGGACTTGCTG (SEQ ID NO: 13)	Pp*STE13* 5' *Eco*RI for
SH775	GGAATTCCTCGAGCTGTTTGAATCTGGAACGTAC TCG (SEQ ID NO: 14)	Pp*STE13* 5' *Eco*RI rev
SH776	GAAGCTTCTCGAGCTACTGGGAACCACGAGACAT CAC (SEQ ID NO: 15)	Pp*STE13* 3' *Hind*III for
SH777	GCAAGCTTGGCCCATTAGGCCCACCTACAATCAT TACC (SEQ ID NO: 16)	Pp*STE13* 3' *Hind*III rev
SH778	CAAGGCACATTAAAAGTCCGCCAAAGG (SEQ ID NO: 17)	Pp*STE13* pre 5'
SH779	GTGGCCCTTGTATTGATAGAAGTATTCAG (SEQ ID NO: 18)	Pp*STE13* post 3'
SH780	CACGTCTATCGTTGAACCAAAACAGAC (SEQ ID NO: 19)	Pp*STE13* KO for
SH781	GTAACCAATGGTATCTCCAACGACAG (SEQ ID NO: 20)	Pp*STE13* KO rev
SH782	GGAATTCGGCCACCTGGGCCTGTTGCTGCTGGTA CTG (SEQ ID NO: 21)	Pp*DAP2* 5' *Eco*RI for
SH783	CGAATTCCTCGAGCGTTGTAAGTGATTGTAGACT CG (SEQ ID NO: 22)	Pp*DAP2* 5' *Eco*RI rev
SH784	GAAGCTTCTCGAGGGCAGCAAAGCCTTACGTTG (SEQ ID NO: 23)	Pp*DAP2* 3' *Hind*III for
SH785	GCAAGCTTGGCCTAGGTGGCCGACCCATTTTTAG AGG (SEQ ID NO: 24)	Pp*DAP2* 3' *Hind*III rev
SH786	CACTTTCATCCTGAGGATCTTGGTCCTG (SEQ ID NO: 25)	Pp*DAP2* pre 5'
SH787	CATATACCAAAGCAATTGATATCTGGTC (SEQ ID NO: 26)	Pp*DAP2* post 3'
SH788	CGGATAAGAGACATAATTGGCGCCATTC (SEQ ID NO: 27)	Pp*DAP2* KO for
SH789	CTTTCTATTGAGGATTTCTTGGTTGCTG (SEQ ID NO: 28)	Pp*DAP2* KO rev
SH801	CGCCATCCAGTGTCGAAAACGCTGTTTGAATCTG GAACGTACTC (SEQ ID NO: 29)	*STE13* 5' (*NAT*^{R}) rev
SH802	GAGTACGTTCCAGATTCAAACAG*CGTTTTCGACAC TGGATGGCG* (SEQ ID NO: 30)	*NAT*^{R} (*STE13* 5') for

SH803	GTGATGTCTCGTGGTTCCCAGTAG*TGTTTAGCTTG CCTCGTCCCCG* (SEQ ID NO: 31)	*NAT*^{R} (*STE13* 3') rev
SH804	CGGGGACGAGGCAAGCTAAACACTACTGGGAAC CACGAGACATCAC (SEQ ID NO: 32)	*STE13* 3' (*NAT*^{R}) for
SH805	CGCCATCCAGTGTCGAAAACGCGTTGTAAGTGAT TGTAGACTCGTTG (SEQ ID NO: 33)	*DAP2* 5' (*NAT*^{R}) rev
SH806	CAACGAGTCTACAATCACTTACAACG*CGTTTTCGA CACTGGATGGCG* (SEQ ID NO: 34)	*NAT*^{R} (*DAP2* 5') for
SH807	CAACGTAAGGCTTTGCTGCC*TGTTTAGCTTGCCTC GTCCCCG* (SEQ ID NO: 35)	*NAT*^{R} (*DAP2* 3') rev
SH808	CGGGGACGAGGCAAGCTAAACAGGCAGCAAAGC CTTACGTTG (SEQ ID NO: 36)	*DAP2* 3' (*NAT*^{R}) for

Table 2

Knock-out	Vector	Region	Primer pair	Product size (kb)
ste13::*URA5*	pGLY4520	5' cross-over	SH778 (SEQ ID NO:17) + SH558 (SEQ ID NO: 12)	1.0
		3' cross-over	SH779 (SEQ ID NO: 18) + SH491 (SEQ ID NO: 11)	1.1
		Knock-out*	SH780 (SEQ ID NO: 19) + SH781 (SEQ ID NO: 20)	no product
dap2::*URA5*	pGLY4521	5' cross-over	SH786 (SEQ ID NO: 25) + SH558 (SEQ ID NO: 12)	1.2
		3' cross-over	SH787 (SEQ ID NO: 26) + SH491 (SEQ ID NO: 11)	1.4
		Knock-out*	SH788 (SEQ ID NO: 27) + SH789 (SEQ ID NO: 28)	no product
ste13:: *NAT*^{R}	pGLY5018	5' cross-over	SH778 (SEQ ID NO: 17) + SH380	1.0

			(SEQ ID NO: 10)	
		3' cross-over	SH779 (SEQ ID NO: 18) + SH379 (SEQ ID NO: 9)	1.1
		Knock-out*	SH780 (SEQ ID NO: 19) + SH781 (SEQ ID NO: 20)	no product
dap2:: NAT^R	pGLY5019	5' cross-over	SH786 (SEQ ID NO: 25) + SH380 (SEQ ID NO: 10)	1.2
		3' cross-over	SH787 (SEQ ID NO: 26) + SH379 (SEQ ID NO: 9)	1.4
		Knock-out*	SH788 (SEQ ID NO: 27) + SH789 (SEQ ID NO: 28)	no product

*1 Kb product obtained with presence of wild-type loci

To assess the role of each of these genes with respect to the N-terminal truncation observed for TNFRII-Fc, knockout vectors containing the *URA5* gene were designed with the 5' and 3' regions of either the *Pichia pastoris DAP2* or *STE13* genes to target the vector for disruption of the *DAP2* or *STE13* locus of the genome (Figures 1A and 1B). These knockout vectors were transformed into the same *ura* minus parent strain (YGLY7406) expressing the recombinant fusion protein, TNFRII-Fc. Following PCR screening of the transformants, a Pp*ste13* and a Pp*dap2* knockout strain were identified, YGLY8084 and YGLY8090, respectively. To generate a double knockout strain, i.e. knockout of both Pp*STE13* and Pp*DAP2* in the same strain, the Pp*ste13* deletion strain, YGLY8084, was transformed with the Pp*DAP2* nourseothricin dominant marker knockout vector, pGLY5019 (Figure 1D). Following PCR screening of the transformants, several Pp*dap2* knockouts were obtained in the Pp*ste13*::*URA5* background parent strain. A representative Pp*ste13*/Pp*dap2* double knockout strain was designated YGLY8096.

Peptide analysis of the TNFRII-Fc secreted from each of these strains shows that while Ste13p plays a major role in the proteolysis of this polypeptide, deletion of both DAP genes is necessary to eliminate DAP activity. For the wild-type strain, 100% of the protein produced was truncated by two amino acids at the N-terminus. The single deletion of *STE13*

resulted in 59% of the peptide population having an intact N-terminus, while deletion of *DAP2* resulted in the production of intact TNFRII-Fc, but to a much lesser degree, with only 15% of the secreted polypeptides having the full length sequence. Analysis of TNFRII-Fc secreted from the double knock-out strain, YGLY8096, indicated that 100% of the peptide population had an intact N-terminus. Table 3 is a summary of the data generated from this analysis showing the molar percentages of truncated versus full length peptides isolated. Data is representative of three replicate samples for each strain. Due to *ura* minus strains being poor secretors of recombinant protein, the prototrophic parent strain of YGLY7406, YGLY6646, was used to produce the material from a strain with both Pp*STE13* and Pp*DAP2* genes un-interrupted. The genotype of this strain is described in Table 3 as wild-type, referring to both intact Pp*STE13* and Pp*DAP2* loci.

Table 3

Strain	Relevant Genotype	N-Terminal Sequence of Truncated/Full Length Peptides Isolated	Truncated/Full Length Peptide Occurence (Molar %)
YGLY6646	Wild-type	AQVAF (SEQ ID NO: 2) and LPAQV (SEQ ID NO: 1)	100/0
YGLY8084	*ste13*::*URA5*	AQVAF (SEQ ID NO: 2 and LPAQV (SEQ ID NO: 1)	41/59
YGLY8090	*dap2*::*URA5*	AQVAF (SEQ ID NO: 2) and LPAQV (SEQ ID NO: 1)	85/15
YGLY8096	*ste13*::*URA5*, *dap2*::*NAT*^{R}	AQVAF (SEQ ID NO: 2) and LPAQV (SEQ ID NO: 1)	0/100

Based on this analysis Applicants determined that PpSte13p and PpDap2p are the sole proteases involved in the N-terminal proteolysis of recombinant TNFRII-Fc expressed in *Pichia pastoris* and that deletion of both are required to eliminate DAP activity. Similarly, when a wild-type granulocyte colony-stimulating factor (GCSF) polypeptide (protein sequence in Figure 5; SEQ ID NO: 6) was expressed in *Pichia pastoris* the N-terminus was truncated. N-terminal sequencing of this material indicated that it lacked two amino acids at the N-terminus (Figure 4A and 4B, SEQ ID NOS: 4 and 5, respectively). Production of this truncated form of GCSF was also completely eliminated by producing the GCSF in a double *Δ*Pp*ste13*/Pp*dap2* knockout background (Figure 6). In both instances, cleavage of these recombinant proteins

(TNFRII-Fc and CSF) occurs C-terminal to a proline residue in the second position (relative to the N-terminal) of the polypeptide. As shown herein, deletion of both DAP associated genes in *Pichia pastoris* prevented truncation of a therapeutic protein having a proline in the second position from the N-terminus of the polypeptide and that deletion of both Pp*STE13* and Pp*DAP2* was required for elimination of DAP activity.

Both therapeutic protein examples described above have the N-terminal motif Xaa-Pro, where X can be any amino acid. As described above, Ste13p and Dap2p individually have been shown to cleave N-terminal peptides also having the motif Xaa-Ala, where Xaa is any amino acid and alanine (Ala) is the second amino acid, Misumi and Ikehara, Handbook of Proteolytic Enzymes, 2nd edition, pp. 1910-1911, Elsevier, London, 2004. To demonstrate that knockouts of both Pp*STE13* and Pp*DAP2* are required to prevent truncation of proteins with the Xaa-Ala motif, one skilled in the art can mutate the second amino acid in the secreted form of TNFR-Fc from a proline to an alanine. Based on the results herein it appears that deletion of either Pp*STE13* or Pp*DAP2* would reduce the percentage of N-terminal cleavage observed. Complete elimination of DAP activity and the resulting 100% intact protein would only be observed from a double *Δ*Pp*ste13*/Pp*dap2* knockout host strain, i.e. mutation to knockout both *STE13* and *DAP2*.

Although the combined elimination of PpSte13p and PpDap2p has demonstrated the production of therapeutic proteins with intact N-termini, Applicants have found that *Pichia pastoris* has a third dipeptidyl aminopeptidase, herein referred to as Pp*DPPIII*. This enzyme was identified by annotation of the *Pichia pastoris* genome and has 43% sequence identity to the human enzyme, dipeptidyl peptidase III (DppIIIp). The *Pichia pastoris* DppIIIp homologue has yet to be characterized, but is distinct from PpSte13p and PpDap2p at the sequence level, having only 10% and 12% sequence identity, respectively. While this DAP gene has not been fully characterized, Applicants believe that PpDppIIIp, when expressed under the proper conditions, will result in truncated proteins, notwithstanding that no activity was observed under the conditions evaluated herein. Those of skill in the art would recognize and appreciate that, under other growth conditions or through expression with other reporter proteins, PpDppIIIp activity may be evidenced. As such, under other appropriate conditions or reporters, to completely eliminate DAP activity in a *Pichia pastoris*, Pp*DppIII* may need to be deleted or disrupted, either alone or in conjunction with PpSte13p and/or PpDap2p activity. Those skilled in the art would be able to delete or disrupt PpDPPIII according to the methods set forth in the examples herein, specifically those of Example 5.

Pichia strains for elimination of DAP activity

While any commercially available *Pichia pastoris* strain can be utilized for the present invention, such as NRRL-Y11430 (American Type Culture Collection (ATCC), Manassas, VA, Catalog No. 76273), in a preferred embodiment the strain employed for the invention herein would be a glycoengineered *Pichia pastoris* strain, for example, the GS5.0 strain described in Example 3 or the glycoengineered strains described *infra*, which incorporates modifications to generate a human-like glycosylation profile upon expression. For example, in the case of the GS5.0 strain of Example 3, such modifications include those directed to the deletion of *Δoch1, Δpno1, Δmnn4B, Δbmt2*, and *Δura5*, or the insertion of *K. lactis* and *M. musculus* UDP-GlcNAc transporters, *M. musculus* α-1,2-MnsI, *H. sapiens* β-1,2,-GlcNAc transferase 1 activity, *R. norvegicus* β-1,2-GlcNAc transferase II activity, *D. melanogaster* MnsII activity, *S. pombe* Gal epimerase, *D. melanogaster* UDP-Gal transporter and *H. sapiens* β-1,4-galactosyltransferase activity. A GS5.0 strain is capable of producing glycoproteins that have galactose-terminated N-glycans, e.g., $GalGlcNAc_2Man_3GlcNA_{c2}$, $Gal_2GlcNAc_2Man_3GlcNAc_2$, or mixtures thereof. Other representative glycoengineered strains include YJN201 (Choi et al., PNAS, 100 (9): 5022-5027, 2003); YSH44 (Hamilton et al., Science, 301 (5637): 1244-1246, 2003); RDP36-1 (Davidson et al., Glycobiology, 14 (4): 1-9, 2004); PBP6-5 (Bobrowicz et al., Glycobiology, 14 (9): 757-766, 2004); YSH597 (Hamilton et al., Science, 313 (5792): 1441-1443, 2006).

In addition to *Pichia pastoris*, host cells useful in the present invention include yeast that express homologues to the *Pichia pastoris DAP2* and *STE13* genes. Such yeast may be selected from among *Pichia finlandica, Pichia trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia minuta (Ogataea minuta, Pichia lindneri), Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia guercuum, Pichia pijperi, Pichia stiptis, Pichia methanolica, Pichia sp., Saccharomyces cerevisiae, Saccharomyces sp., Hansenula polymorpha, Kluyveromyces sp., Kluyveromyces lactis, Candida albicans*. Various yeasts, such as *K. lactis, Pichia pastoris, Pichia methanolica,* and *Hansenula polymorpha* are particularly suitable for cell culture because they are able to grow to high cell densities and secrete large quantities of recombinant protein.

Yeast can be genetically modified so that they express glycoproteins in which the glycosylation pattern is human-like or humanized. In this manner, glycoprotein compositions

can be produced in which a specific desired glycoform is predominant in the composition. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or genetically engineering the host cells and/or supplying exogenous enzymes to mimic all or part of the mammalian glycosylation pathway as described in US 2004/0018590. If desired, additional genetic engineering of the glycosylation can be performed, such that the glycoprotein can be produced with or without core fucosylation. Use of lower eukaryotic host cells is further advantageous in that these cells are able to produce highly homogenous compositions of glycoprotein, such that the predominant glycoform of the glycoprotein may be present as greater than thirty mole percent of the glycoprotein in the composition. In particular aspects, the predominant glycoform may be present in greater than forty mole percent, fifty mole percent, sixty mole percent, seventy mole percent and, most preferably, greater than eighty mole percent of the glycoprotein present in the composition.

Yeast, can be genetically modified so that they express glycoproteins in which the glycosylation pattern is human-like or humanized. Human-like or humanized N-glycans include both hybrid and complex *N*-glycans. Such can be achieved by eliminating selected endogenous glycosylation enzymes and/or supplying exogenous enzymes as described by Gerngross *et al.*, US 20040018590. For example, a host cell can be selected or engineered to be depleted in 1,6-mannosyl transferase activities, which would otherwise add mannose residues onto the *N*-glycan on a glycoprotein.

In one embodiment, the host cell further includes an α1,2-mannosidase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target the α1,2-mannosidase activity to the ER or Golgi apparatus of the host cell. Passage of a recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $Man_5GlcNAc_2$ glycoform, for example, a recombinant glycoprotein composition comprising predominantly a $Man_5GlcNAc_2$ glycoform. For example, U.S. Patent No, 7,029,872 and U.S. Published Patent Application Nos. 2004/0018590 and 2005/0170452 disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $Man_5GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a GlcNAc transferase I (GnT I) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase I activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through

the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAcMan_5GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a hybrid $GlcNAcMan_5GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872 and U.S. Published Patent Application Nos. 2004/0018590 and 2005/0170452 disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $GlcNAcMan_5GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexosaminidase to produce a recombinant glycoprotein comprising a $Man_5GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a mannosidase II catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target mannosidase II activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAcMan_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a complex $GlcNAcMan_3GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872 and U.S. Published Patent Application No. 2004/0230042 discloses lower eukaryote host cells that express mannosidase II enzymes and are capable of producing glycoproteins having predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexosaminidase to produce a recombinant glycoprotein comprising a $Man_3GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes GlcNAc transferase II (GnT II) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target GlcNAc transferase II activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a complex $GlcNAc_2Man_3GlcNAc_2$ glycoform. U.S. Patent No, 7,029,872 and U.S. Published Patent Application Nos. 2004/0018590 and 2005/0170452 disclose lower eukaryote host cells capable of producing a glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a hexosaminidase to produce a recombinant glycoprotein comprising a $Man_3GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a complex $GalGlcNAc_2Man_3GlcNAc_2$ (G1) or complex $Gal_2GlcNAc_2Man_3GlcNAc_2$ (G2) glycoform, or mixture thereof for example a recombinant glycoprotein composition comprising predominantly a $GalGlcNAc_2Man_3GlcNAc_2$ glycoform or $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof. U.S. Patent No, 7,029,872 and U.S. Published Patent Application No. 2006/0040353 discloses lower eukaryote host cells capable of producing a glycoprotein comprising a $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform. The glycoprotein produced in the above cells can be treated *in vitro* with a galactosidase to produce a recombinant glycoprotein comprising a $GlcNAc_2Man_3GlcNAc_2$ glycoform, for example a recombinant glycoprotein composition comprising predominantly a $GlcNAc_2Man_3GlcNAc_2$ glycoform.

In a further embodiment, the immediately preceding host cell further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialytransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly a complex $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or complex $NANAGal_2GlcNAc_2Man_3GlcNAc_2$ glycoform or mixture thereof. It is useful that the host cell further include a means for providing CMP-sialic acid for transfer to the *N*-glycan. U.S. Published Patent Application No. 2005/0260729 discloses a method for genetically engineering lower eukaryotes to have a CMP-sialic acid synthesis pathway and U.S. Published Patent Application No. 2006/0286637 discloses a method for genetically engineering lower eukaryotes to produce sialylated glycoproteins. The glycoprotein produced in the above cells can be treated *in vitro* with a neuraminidase to produce a recombinant glycoprotein comprising predominantly a $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoform.

Any one of the preceding host cells can further include one or more GlcNAc transferase selected from the group consisting of GnT III, GnT IV, GnT V, GnT VI, and GnT IX to produce glycoproteins having bisected (GnT III) and/or multiantennary (GnT IV, V, VI, and

IX) *N*-glycan structures such as disclosed in U.S. Published Patent Application Nos. 2004/074458 and 2007/0037248.

In further embodiments, the host cell that produces glycoproteins that have predominantly hybrid GlcNAcMan5GlcNAc2 *N*-glycans further includes a galactosyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target galactosyltransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising predominantly the hybrid GalGlcNAcMan5GlcNAc2 glycoform.

In a further embodiment, the immediately preceding host cell that produced glycoproteins that have predominantly the GalGlcNAcMan5GlcNAc2 *N*-glycans further includes a sialyltransferase catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain and selected to target sialytransferase activity to the ER or Golgi apparatus of the host cell. Passage of the recombinant glycoprotein through the ER or Golgi apparatus of the host cell produces a recombinant glycoprotein comprising a hybrid NANAGalGlcNAcMan5GlcNAc2 glycoform.

Various of the preceding host cells further include one or more sugar transporters such as UDP-GlcNAc transporters (for example, *Kluyveromyces lactis* and *Mus musculus* UDP-GlcNAc transporters), UDP-galactose transporters (for example, *Drosophila melanogaster* UDP-galactose transporter), and CMP-sialic acid transporter (for example, human sialic acid transporter). Because lower eukaryote host cells such as yeast and filamentous fungi lack the above transporters, it is preferable that lower eukaryote host cells such as yeast and filamentous fungi be genetically engineered to include the above transporters.

Host cells further include lower eukaryote cells (e.g., yeast such as *Pichia pastoris*) that are genetically engineered to eliminate glycoproteins having α-mannosidase-resistant *N*-glycans by deleting or disrupting one or more of the β-mannosyltransferase genes (e.g., *BMT1*, *BMT2, BMT3*, and *BMT4*)(*See*, U.S. Published Patent Application No. 2006/0211085) and glycoproteins having phosphomannose residues by deleting or disrupting one or both of the phosphomannosyl transferase genes *PNO1* and *MNN4B* (*See* for example, U.S. Patent Nos. 7,198,921 and 7,259,007), which in further aspects can also include deleting or disrupting the *MNN4A* gene. Disruption includes disrupting the open reading frame encoding the particular enzymes or disrupting expression of the open reading frame or abrogating translation

of RNAs encoding one or more of the β-mannosyltransferases and/or phosphomanno-syltransferases using interfering RNA, antisense RNA, or the like. The host cells can further include any one of the aforementioned host cells modified to produce particular *N*-glycan structures.

Host cells further include lower eukaryote cells (e.g., yeast such as *Pichia pastoris*) that are genetically modified to control *O*-glycosylation of the glycoprotein by deleting or disrupting one or more of the protein *O*-mannosyltransferase (Dol-P-Man:Protein (Ser/Thr) Mannosyl Transferase genes) (*PMT*s) (*See* U.S. Patent No. 5,714,377) or grown in the presence of Pmtp inhibitors and/or an alpha-mannosidase as disclosed in Published International Application No. WO 2007061631, or both. Disruption includes disrupting the open reading frame encoding the Pmtp or disrupting expression of the open reading frame or abrogating translation of RNAs encoding one or more of the Pmtps using interfering RNA, antisense RNA, or the like. The host cells can further include any one of the aforementioned host cells modified to produce particular *N*-glycan structures.

Pmtp inhibitors include but are not limited to a benzylidene thiazolidinediones. Examples of benzylidene thiazolidinediones that can be used are 5-[[3,4-bis(phenylmethoxy) phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic acid; 5-[[3-(1-phenylethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic acid; and 5-[[3-(1-phenyl-2-hydroxy)ethoxy)-4-(2-phenylethoxy)]phenyl]methylene]-4-oxo-2-thioxo-3-thiazolidineacetic acid.

In particular embodiments, the function or expression of at least one endogenous *PMT* gene is reduced, disrupted, or deleted. For example, in particular embodiments the function or expression of at least one endogenous *PMT* gene selected from the group consisting of the *PMT1*, *PMT2, PMT3,* and *PMT4* genes is reduced, disrupted, or deleted; or the host cells are cultivated in the presence of one or more PMT inhibitors. In further embodiments, the host cells include one or more *PMT* gene deletions or disruptions and the host cells are cultivated in the presence of one or more Pmtp inhibitors. In particular aspects of these embodiments, the host cells also express a secreted alpha-1,2-mannosidase.

PMT deletions or disruptions and/or Pmtp inhibitors control *O*-glycosylation by reducing *O*-glycosylation occupancy, that is, by reducing the total number of *O*-glycosylation sites on the glycoprotein that are glycosylated. The further addition of an alpha-1,2-mannosidase that is secreted by the cell controls *O*-glycosylation by reducing the mannose chain length of the *O*-glycans that are on the glycoprotein. Thus, combining *PMT* deletions or disruptions and/or

Pmtp inhibitors with expression of a secreted alpha-1,2-mannosidase controls *O*-glycosylation by reducing occupancy and chain length. In particular circumstances, the particular combination of *PMT* deletions or disruptions, Pmtp inhibitors, and alpha-1,2-mannosidase is determined empirically as particular heterologous glycoproteins (antibodies, for example) may be expressed and transported through the Golgi apparatus with different degrees of efficiency and thus may require a particular combination of *PMT* deletions or disruptions, Pmtp inhibitors, and alpha-1,2-mannosidase. In another aspect, genes encoding one or more endogenous mannosyltransferase enzymes are deleted. This deletion(s) can be in combination with providing the secreted alpha-1,2-mannosidase and/or *PMT* inhibitors or can be in lieu of providing the secreted alpha-1,2-mannosidase and/or *PMT* inhibitors.

Therefore, the host cell can be any host cell that has been genetically modified to produce glycoproteins wherein the predominant *N*-glycan is selected from the group consisting of complex *N*-glycans, hybrid *N*-glycans, and high mannose *N*-glycans wherein complex *N*-glycans are selected from the group consisting of $Man_3GlcNAc_2$, $GlcNAc_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}Man_3GlcNAc_2$; hybrid *N*-glycans are selected from the group consisting of $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, and $NANAGalGlcNAcMan_5GlcNAc_2$; and high mannose *N*-glycans are selected from the group consisting of $Man_6GlcNAc_2$, $Man_7GlcNAc_2$, $Man_8GlcNAc_2$, and $Man_9GlcNAc_2$. Examples of *N*-glycan structures include but are not limited to $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_2Man_3GlcNAc_2$, $GlcNAc_3Man_3GlcNAc_2$, $GlcNAc_4Man_3GlcNAc_2$, $GalGlcNAc_2Man_3GlcNAc_2$, $Gal_2GlcNAc_2Man_3GlcNAc_2$, $Gal_2GlcNAc_3Man_3GlcNAc_2$, $Gal_2GlcNAc_4Man_3GlcNAc_2$, $Gal_3GlcNAc_3Man_3GlcNAc_2$, $Gal_3GlcNAc_4Man_3GlcNAc_2$, $Gal_4GlcNAc_4Man_3GlcNAc_2$, $NANAGal_2GlcNAc_2Man_3GlcNAc_2$, $NANA_2Gal_2GlcNAc_2Man_3GlcNAc_2$, $NANA_3Gal_3GlcNAc_3Man_3GlcNAc_2$, and $NANA_4Gal_4GlcNAc_4Man_3GlcNAc_2$.

In the examples that follow, the glycoengineered *Pichia pastoris* strain has been modified to eliminate Ste13p and Dap2p dipeptidyl aminopeptidase activity. In general, Ste13p and/or Dap2p activity can be eliminated from any *Pichia pastoris* strain using the methods described herein provided that the markers used for selection can be utilized with the strain. Alternatively another auxotropic or dominant marker, through which selection is available, may

be substituted. For example, *STE13* can be deleted from the commercially available strain, NRRL-Y11430, using the vector pGLY5018 (Example 2C) and nourseothricin as a selection marker. In that the markers for neither of the *DAP2* knockout vectors described below (Example 2B) are compatible with NRRL-Y11430, a hygromycin marked vector can be generated from pAG32 (Goldstein, et al., Yeast, 15(6): 507-511,1999; Erratum: Yeast, 15(12): 1297, 1999), using primers corresponding to the italicized fragments of primers SH806 and SH807 (Table 1) (SEQ ID NO: 48 and SEQ ID NO: 49, respectively) to amplify the 1654bp Hyg marker from pAG32. This latter fragment can then be fused to the *DAP2* 5' and *DAP2* 3' flanking regions as described in Example 2C for the generation of pGLY5019. Following digestion with *Sfi*I (New England BioLabs, Ipswich, MA), the vector can be transformed into either NRRL-Y11430 or the *ste13* knockout derivative thereof and selected on hygromycin containing plates to produce either the single *dap2Δ* or double *Δste13/dap2* knockout strains, respectively. An example is also provided as to how one skilled in the art would be able to use a glycoengineered *Pichic pastoris* strain that has been modified to eliminate DppIIIp dipeptidyl aminopeptidase activity using the methods described herein.

EXAMPLE 1

Strains, culture conditions, and reagents.

Escherichia coli strains TOP10 (Invitrogen, Carlsbad, CA) or XL10-Gold (Stratagene, Santa Clara, CA) were used for recombinant DNA work. Restriction and modification enzymes were obtained from New England BioLabs, Ipswich, MA, and used as directed by the manufacturer. Oligonucleotides were obtained from Integrated DNA Technologies, Coralville, IA. Salts and buffering agents were from Sigma, St. Louis, MO. Minimal medium used herein comprised 1.4% yeast nitrogen base, 2% dextrose, 1.5% agar and 4 X 10^{-5}% biotin and amino acids supplemented as appropriate. YMD rich media is 1% yeast extract, 2% martone, 2% dextrose, and 1.5% agar for plates. Nourseothricin is obtained from US Biologicals, Swampscott, MA (Catalogue number N5375-74) and is added to the YMD rich media to a final concentration of 100μg/ml.

EXAMPLE 2

Generation of knock-out vectors

A. Generation of *ste13*::*URA5* knock-out vector

DNA fragments corresponding to 5' and 3' flanking regions of the *STE13* open reading frame (SEQ ID NO: 41 and 42) were amplified using PfuUltra™ DNA polymerase (Stratagene, Santa Clara, CA) and genomic DNA from the *Pichia pastoris* strain NRRL-Y11430 as template. The primer pairs SH774 (SEQ ID NO: 13) and SH775 (SEQ ID NO: 14) and SH776 (SEQ ID NO: 15) and SH777 (SEQ ID NO: 16), shown in Table 1, were used to amplify the 771bp and 949bp fragments for *STE13* 5' and 3', respectively. Following incubation with ExTaq™ (TaKaRa, Bio. Inc., Japan) for ten minutes at 72°C, the amplified fragments were cloned into pCR2.1 (Invitrogen, Carlsbad, CA) and transformed into TOP10 competent cells. DNA sequencing confirmed the *STE13* 5' and *STE13* 3' flanking regions were correct and the resultant vectors were designated pGLY4511 and pGLY4512, respectively.

A 763bp *STE13* 5' flanking region fragment (represented by the underlined region of Figure 10A, SEQ ID NO: 50) was digested from pGLY4511 using EcoRI and subcloned into a *P. pastoris URA5*-blaster vector pGLY13b, similar to pJN396 (Nett and Gerngross, Yeast, 20, 1279-1290, 2003), which was previously digested with the same restriction enzyme and treated with calf intestinal alkaline phosphatase (CIAP). The *URA5*-blaster vector possesses a pUC19 vector backbone containing a functional gene fragment of the *Pichia URA5* gene flanked by LacZ direct repeats to facilitate recovery of the *URA5* marker through counter selection on 5-fluoroorotic acid. Following transformation of the newly ligated vector into XL10 Gold competent cells and confirmation by restriction analysis, the resultant vector was designated pGLY4518. The vector pGLY4512 was digested with HindIII to release a 940bp (represented by the underlined region of Figure 10B, SEQ ID NO: 51) fragment encoding the *STE13* 3' flanking region and subcloned into pGLY4518, which was previously digested with the same enzyme and CIAP treated. The ligation product was transformed into XL10 Gold competent cells and designated pGLY4520 following restriction analysis. This final *ste13*::*Ura5* knockout vector is shown graphically in Figure 1A.

B. Generation of *dap2*:: *URA5* knock-out vector

The *DAP2* 5' and 3' flanking regions (SEQ ID NO: 43 and 44) were amplified from *Pichia pastoris* genomic DNA as described above using the primer sets SH782 (SEQ ID

NO: 21) and SH783 (SEQ ID NO: 22) and SH784 (SEQ ID NO: 23) and SH785 (SEQ ID NO: 24), shown in Table 1, to generate 1003bp and 1142bp fragments, respectively. Following cloning into pCR2.1 and sequencing, the vectors were designated pGLY4513 and pGLY4514, encoding the *DAP2* 5' and *DAP2* 3' regions, respectively. Following a similar approach to that described above in Example 2A, the 995bp *DAP2* 5' region (represented by the underlined region of Figure 11A, SEQ ID NO: 52) was subcloned into the EcoRI site in the Ura5-blaster vector pGLY13b, resulting in the intermediate construct pGLY4519. Subsequently the 1133bp *DAP2* 3' region (represented by the underlined region of Figure 11B, SEQ ID NO: 53) was subcloned into the HindIII site of pGLY4519, resulting in the *dap2::URA5* knockout vector pGLY4521, shown graphically in Figure 1B.

C. Generation of dominant marker *STE13* and *DAP2* knockout vectors

PCR fusion was used to generate both *STE13* and *DAP2* knockout vectors. The *STE13* 5' and *STE13* 3' fragments were amplified from pGLY4520 with the primer pairs SH774 (SEQ ID NO: 13) and SH801 (SEQ ID NO: 29) and SH804 (SEQ ID NO: 32) and SH777 (SEQ ID NO: 16), shown in Table 1, using PfuUltra™ DNA polymerase. The nourseothricin (NAT^R) marker cassette (SEQ ID NO: 45), which contains the open reading frame for the nourseothricin acetyltransferase (represented by the bold text of Figure 12, SEQ ID NO: 54) from *Streptomyces noursei* under the expressional control of the *Ashbya gossypii* Transcription Elongation Factor (TEF) promoter and terminator (represented by the regular text of Figure 12), which are 5' and 3' to the highlighted open reading frame, respectively), was amplified from pAG25 (Goldstein and McCusker, Yeast, 15, 1541-1553, 1999) using the primers SH802 (SEQ ID NO: 30) and SH803 (SEQ ID NO: 31). The PCR reactions were run on a DNA agarose gel and the 779bp, 958bp, and 1249bp fragments, corresponding to *STE13* 5', *STE13* 3' and the Nat marker, respectively, were isolated. Subsequently 20 ng of each were combined and fused together using PfuUltra™ DNA polymerase and the primer pair SH774 (SEQ ID NO: 13) and SH777 (SEQ ID NO: 16). Following incubation for ten minutes at 72°C with ExTaq ™ DNA polymerase (TaKaRa, Bio. Inc., Japan), the amplified (2896bp) fragment was cloned into pCR2.1 and transformed into TOP10 competent cells. DNA sequencing confirmed that the *ste13*:: NAT^R fusion was correct and the resultant vector was designated pGLY5018. This vector is shown graphically in Figure 1C.

In a similar manner, the primer sets SH782 (SEQ ID NO: 21) and SH805 (SEQ ID NO: 33), SH808 (SEQ ID NO: 36) and SH785 (SEQ ID NO: 24), and SH806 (SEQ ID NO: 34) and SH807 (SEQ ID NO: 35), shown in Table 1, were used to amplify 1011bp, 1151bp, and 1248bp fragments corresponding to *DAP2* 5', *DAP2* 3', and the NAT^R marker from pGLY4521 and pAG25. Following isolation 20 ng of each fragment was used with the primer pair SH782 (SEQ ID NO: 21) and SH785 (SEQ ID NO: 24) to generate the 3321bp fragment that was cloned into pCR2.1, sequenced and designated pGLY5019. This vector is shown graphically in Figure 1D.

EXAMPLE 3

Generation of *STE13* and *DAP2* knockout strains

Pichia pastoris auxotrophic glycoengineered cell line YGLY7406 [Δ*och1*, Δ*pno1*, Δ*mnn4B*, Δ*bmt2*, Δ*ura5*, *K. lactis* and *M. musculus* UDP-GlcNAc transporters, *M. musculus* α-1,2-MnsI, *H. sapiens* β-1,2-GlcNAc transferase I, *R. norvegicus* β-1,2-GlcNAc transferase II, *D. melanogaster* MnsII, *S. pombe* Gal epimerase, *D. melanogaster* UDP-Gal transporter and *H. sapiens* β-1,4-galactosyltransferase] expressing GS5.0 glycans (*See* for example, Bobrowicz et al., Glycobiology, 14(9): 757-766, 2004; Hamilton et al., Science, 313 (5792): 1441l-1443, 2006); U.S. Published Application No. 20060040353, was used as the starting strain for all manipulations. See Figure 7 (A-C) for a flow diagram of how this and the subsequent strains were generated. GS5.0 strains can produce glycoproteins having bi-antennary afucosylated *N*-linked glycans terminating in β-1,4 galactose residues on the non-reducing ends of one or both termini (Bobrowicz et al., Glycobiology, 14(9): 757-766, 2004; Hamilton et al., Science, 313 (5792): 1441l-1443, 2006). The strain YGLY7406 expresses full length human tumor-necrosis factor receptor II fused to the Fc domain of IgG1 (TNFRII-Fc), which will be used as a reporter protein for dipeptidyl peptidase activity.

For transformation 20μg of the vectors pGLY4520 and pGLY4521 were digested with the restriction enzyme SfiI (New England Biolabs, Ipswich, MA) to release a 4091bp *STE13*::Ura5 knock-out fragment or a 4516bp *DAP2*:: *URA5* knock-out fragment which were transformed into YGLY7406 by electroporation to delete *STE13* or *DAP2*, respectively, and selected on *ura* minus minimal plates. Successful knockouts of each gene were confirmed using the 5', 3' and knock-out primer sets shown in Table 2. The *ste13Δ* and *dap2Δ* knockout strains were named YGLY8084 and YGLY8090, respectively. Subsequently, the double *Δste13/dap2* knockout strain was generated by transforming the *ste13Δ* knockout strain, YGLY8084, with

pGLY5019, which was previously digested with SfiI to release the 3290 bp *DAP2*::Nat knock-out fragment. Transformants were plated on 100 μg/ml Nat YMD plates and successful double knockouts were confirmed using the 5', 3' and knockout primer sets shown in Table 2. A representative double knockout strain was designated YGLY8096. Although this strain was generated by knocking-out the *STE13* first followed by knocking-out the *DAP2*, we have experimentally demonstrated that a strain with the same desired phenotype can be obtained if the *DAP2* gene is knocked-out first followed knock-out of the *STE13* gene. To obtain such a strain, pGLY5018 is digested with SfiI and the 2865bp *ste13*:: NAT^R knock-out fragment generated can be transformed into a strain such as YGLY8090, where the *DAP2* gene has already been knocked-out.

EXAMPLE 4

Generation, isolation and analysis of reporter protein

Protein expression was carried out for 72 hours at 26°C in 200ml buffered glycerol-complex medium (BMGY) consisting of 1% yeast extract, 2% peptone, 100mM potassium phosphate buffer, pH 6.0, 1.34% yeast nitrogen base, 0.00004% biotin, and 1% glycerol as a growth medium. Induction was performed for 48 hours in 20 ml of buffered methanol-complex medium (BMMY), consisting of 1.5% methanol instead of glycerol in BMGY.

Following expression, the cells were removed by centrifugation at 2000 rpm for fifteen minutes. The TNFRII-Fc fusion protein was captured by affinity chromatography from the supernatant using Streamliner Protein A resin from GE Healthcare (Chalfont St. Giles, UK, Cat. #17-1281-01). The cell free supernatant medium was loaded on to Streamliner Protein A column (XK 16/20 1.6cm x 10.0 cm) and pre-equilibrated with 3 column volumes of 20 mM Tris-HCl pH 7.0 at a flow rate of 5.0 ml/min. The column was washed with 3 column volumes of the same buffer and the TNFRII-Fc fusion protein was eluted with 7 column volumes of 40 mM sodium citrate pH 3.5. The eluted fusion protein was neutralized immediately with 1M Tris-HCl pH 8.0.

CHT® Hydroxyapatite type I 40 μm resin (Bio-Rad Laboratories, Hercules, CA, Cat #157-0040) was used as a second purification step. The Hydroxyapatite column was equilibrated with 3 column volumes of 5 mM sodium phosphate pH 6.5 and the Streamliner Protein A purified TNFRII-Fc fusion protein was buffer exchanged into the equilibration buffer and applied to the column. After loading, the column was washed with 3 column volumes of the

equilibration buffer and elution was performed by developing a gradient over 20 column volumes ranging from 0 to 1000 mM sodium chloride. TNFRII-Fc fusion protein elutes around 550-650 mM sodium chloride. The pooled TNFRII-Fc fusion protein was sterile filtered using 0.2 μm Polyethersulfone (PES) membrane filter and stored at 4°C.

SDS-PAGE (4-20% Tris-HCl gels, Bio-Rad Laboratories, Hercules, CA, Cat. #161-1123) was run for the two column purified TNFRII-Fc fusion protein fractions, transferred onto a polyvinylidene difluoride (PVDF) membrane (ProBlott™ Membranes, Applied Biosystems, Foster City, CA, Cat # 400994) at 55 volts for thirty minutes and stained with ProBlott stain (Applied Biosystems, Foster City, CA). Bands corresponding to TNFRII-Fc fusion protein were cut from the PVDF membrane and sent to Tufts Core Facility, Boston, MA, for N-terminal sequencing.

Expression of a full length recombinant granulocyte colony-stimulating factor (GCSF) protein was carried out in a similar manner. However, instead of the GS5.0 host cell above that was used to express the TNFRII-Fc, the host cell used for expressing the GCSF was a *Pichia pastoris* GS2.0 host cell in which the *STE13* and *DAP2* genes had been disrupted following the process described in Example 3. GS2.0 host cells are host cells that have been genetically engineered to produce glycoproteins comprising $Man_5GlcNAc_2$ *N*-glycans. These strains are not capable of producing glycoproteins that have galactose-terminated *N*-glycans. Examples of such strains have been disclosed in Nett and Gerngross, Yeast 20:1279 (2003); Choi *et al.*, Proc. Natl. Acad. Sci. USA 100:5022 (2003); and Hamilton *et al.*, Science 301:1244 (2003). Figure 6 shows that these host cells are capable of producing an intact GCSF when the *STE13* and *DAP2* genes were disrupted as described above for the GS5.0 strain expressing TNFRII-Fc (compare lanes 27-29 showing GCSF produced in *STE13/DAP2* strain to lanes 32-34 showing GCSF produced in an *ste13/dap2* knock-out strain).

EXAMPLE 5

Deletion of PpDppIIIp

Similarly, the methods described in the examples above to eliminate PpSte13p and PpDap2p can be used to eliminate PpDppIIIp. One such method would be to design a knock-out vector in a manner similar to that provided for Pp*STE13* and Pp*DAP2* in Example 2. Primers would be designed to amplify the 5' and 3' flanking regions of the Pp*DPPIII* gene (Figure 13, SEQ ID NO: 54), for an essential region of the genomic sequence, for example, a region required for functional PpDppIIIp activity is omitted. PpDppIIIp activity can be

eliminated by combining these flanking regions with a selectable marker. In combination with compatible selectable markers, PpDppIIIp activity can be eliminated individually or in combination with the elimination of PpSte13p and/or PpDap2 activity. A strain in which PpSte13p, PpDap2p and PpDppIIIp activity has been eliminated would produce a strain devoid of any potential dipeptidyl aminopeptidase activity.

The present invention is not to be limited by the specific embodiments disclosed in the examples that are intended as illustrations of a few aspects of the invention and any embodiments that are functionally equivalent are within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art and are intended to fall within the scope of the appended claims.

WHAT IS CLAIMED:

1. A method for producing a recombinant protein in a yeast host cell in which said host yeast cell lacks dipeptidyl aminopeptidase (DAP) activity comprising:

a. transforming a genetically modified yeast cell in which DAP activity has been eliminated with a polynucleotide vector encoding the protein;

b. culturing the transformed host cell in a medium under conditions which induce expression of the protein; and

c. isolating the protein from the transformed host cell or culture medium.

2. The method of claim 1 where the yeast cell is *Pichia pastoris*.

3. The method of claim 1 or 2 wherein DAP activity has been eliminated by deletion or disruption of the *STE13* and *DAP2* genes from the yeast cell genome.

4. The yeast of claim 1, wherein the host cell has been genetically engineered to produce glycoproteins comprising human-like *N*-glycans.

5. The yeast of claim 4, wherein the host cell wherein the human-like N-glycans are selected from the group consisting of hybrid and complex *N*-glycans.

6. The yeast of claim 1, wherein the host cell is genetically engineered to produce glycoproteins that have predominantly an *N*-glycan selected from $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, $NANAGalGlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$.

7. The method of claim 1, wherein the yeast further expresses a human glycoprotein.

8. A genetically modified yeast cell lacking dipeptidyl aminopeptidase (DAP) activity comprising a yeast cell line in which the genomic DNA encoding *STE13* and *DAP2* has been deleted or disrupted from the yeast cell genome.

9. The yeast cell of claim 8 where the yeast cell is *Pichia pastoris*.

10. The method of claim 6, wherein the host cell has been genetically engineered to produce glycoproteins comprising human-like *N*-glycans.

11. The method of claim 10, wherein the host cell wherein the human-like N-glycans are selected from the group consisting of hybrid and complex *N*-glcyans.

12. The method of claim 6, wherein the host cell is genetically engineered to produce glycoproteins that have predominantly an *N*-glycan selected from $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, $NANAGalGlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$.

13. The method of claim 6, wherein the protein is a human glycoprotein.

14. A method for producing a therapeutic protein having an N-terminal recognition site Xaa-Pro or Xaa-Ala, where Xaa represents any one amino acid, in a yeast cell comprising:

a. transforming a genetically modified yeast cell in which DAP activity has been eliminated with a polynucleotide vector encoding the therapeutic protein with the N-terminal recognition site selected from the group consisting of Xaa-Pro and Xaa-Ala;

b. culturing the transformed host cell under conditions which induce expression of the therapeutic protein; and

c. isolating the therapeutic protein from the transformed host cell.

15. The method of claim 8 where the yeast cell is *Pichia pastoris*.

16. The method of claim 14 or 15 wherein DAP activity has been eliminated by deletion or disruption of *STE13* and *DAP2* from the yeast cell genome.

17. The method of claim 14, wherein the host cell has been genetically engineered to produce glycoproteins comprising human-like *N*-glycans.

18. The method of claim 17, wherein the human-like N-glycans are selected from the group consisting of hybrid and complex *N*-glycans.

19. The method of claim 14, wherein the host cell is genetically engineered to produce glycoproteins that have predominantly an *N*-glycan selected from $Man_5GlcNAc_2$, $GlcNAcMan_5GlcNAc_2$, $GalGlcNAcMan_5GlcNAc_2$, $NANAGalGlcNAcMan_5GlcNAc_2$, $GlcNAcMan_3GlcNAc_2$, $GlcNAc_{(1-4)}Man_3GlcNAc_2$, $Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$, and $NANA_{(1-4)}Gal_{(1-4)}GlcNAc_{(1-4)}Man_3GlcNAc_2$.

20. The method of claim 14, wherein the protein is a human glycoprotein.