

## 4.32 Glycoengineering: Recombinant Glycoproteins

---

**M. J. Betenbaugh, N. Tomiya, and S. Narang**, Johns Hopkins University, Baltimore, MD, USA  
© 2007 Elsevier Ltd. All rights reserved.

<b>4.32.1 Introduction</b>	607
<b>4.32.2 N-Glycan Expression in Mammalian Hosts</b>	608
4.32.2.1 Generation of LLOs	608
4.32.2.1.1 Biosynthesis of dolichol phosphate	608
4.32.2.1.2 Assembly of oligosaccharide moiety onto dolichol phosphate	608
4.32.2.2 Transfer of the Oligosaccharide Moiety to Nascent Polypeptide	609
<b>4.32.3 N-Glycan Site Occupancy</b>	609
4.32.3.1 Importance of N-Glycan Site Occupancy	609
4.32.3.2 Engineering N-Glycan Site Occupancy	611
4.32.3.2.1 Control of dolichol phosphate synthesis	611
4.32.3.2.2 Control of LLO synthesis	612
4.32.3.2.3 Control of oligosaccharyltransfer activity	613
4.32.3.2.4 Acceptor tripeptide sequon	614
<b>4.32.4 Microheterogeneity: Control of N-Glycan Processing in the Secretory Compartments</b>	615
4.32.4.1 Glycosidases and Lectins in the ER	615
4.32.4.2 Glycosidases and Glycosyltransferases in the Golgi	617
4.32.4.2.1 Trimming of the terminal mannose	617
4.32.4.2.2 Synthesis of GlcNAcMan <sub>5</sub> GlcNAc <sub>2</sub> intermediate	617
4.32.4.2.3 Addition of galactose and sialic acid	618
<b>4.32.5 Glycoprotein Expression in Nonmammalian Hosts</b>	619
4.32.5.1 Insect	619
4.32.5.1.1 N-Glycosylation in insect cells	619
4.32.5.1.2 Engineering of N-glycosylation pathway	627
4.32.5.2 Plant	630
4.32.5.2.1 N-glycan processing in plant cells	630
4.32.5.2.2 Engineering of N-glycosylation pathway	631
4.32.5.3 Yeast	632
4.32.5.3.1 N-Glycan processing in yeast	632
4.32.5.3.2 Engineering of N-glycosylation pathway	632

---

### 4.32.1 Introduction

A number of eukaryotic cell lines, including mammalian, insect, plant, and fungal, are used to produce numerous recombinant proteins in laboratories and in bioreactors. The capacity of mammalian cells to perform complex post-translational modifications resulting in improved biological activity and increased circulatory lifetimes *in vivo* have made them the system of choice for the commercial production of many therapeutic proteins.<sup>1</sup> A major fraction of these therapeutic proteins, both membrane bound and secreted, are glycosylated. Indeed, glycosylation, a post-translational modification involving oligosaccharide attachment and modification to the protein, can play a very important role in the *in vivo* efficacy of therapeutic proteins.

The two types of glycosylation observed on proteins expressed in eukaryotic systems are N-glycosylation and O-glycosylation, which are based on the linkage between oligosaccharide and the protein. During O-glycosylation, an *N*-acetylgalactosamine residue is linked to the hydroxyl of specific serine or threonine residues through an O-glycosidic bond.<sup>2</sup> This sugar group can then be modified further by addition of sugar groups as the polypeptide

traverses the endoplasmic reticulum (ER) and Golgi compartments. In N-glycosylation, a preformed oligosaccharide moiety is transferred *en bloc* from a long-chain isoprenoid lipid (dolichol) onto the specific N-glycosylation site via an N-glycosidic linkage to the asparagine (Asn) residue of a nascent polypeptide. The N-glycosidic site has Asn-X-Ser/Thr as the consensus sequence,<sup>3</sup> in which X is any amino acid except proline. N-Glycosylation heterogeneity can be observed at the so-called macro- and microlevel. Macroheterogeneity, that is, variability in the number of potential sites glycosylated among different molecules of the same protein,<sup>4</sup> occurs as a result of the variation in site availability, enzyme kinetics, and substrate concentrations as the polypeptide is generated in the ER. Alternatively, variations associated with the glycan-processing reactions that take place following the initial addition step, primarily in the Golgi apparatus, result in microheterogeneity, that is, the variability in structures of the oligosaccharide attached to each glycosylation site. The macro- and microheterogeneity of N-glycosylation may depend on cell culture conditions, the location of the potential glycosylation sites, and the particular expression system used.<sup>4</sup> Due to the influence of N-glycosylation on the physiological and biochemical properties of the glycoprotein including stability, solubility, biological activity, and *in vivo* clearance rate,<sup>1,5</sup> this heterogeneity in protein glycosylation presents a major challenge in the development of therapeutic glycoproteins with consistent properties.<sup>1</sup> In the following sections, we will focus on the factors that significantly affect the macro- and microheterogeneity of N-glycans and will describe methods underway to alter the glycoforms produced for particular host organisms.

## 4.32.2 N-Glycan Expression in Mammalian Hosts

N-Glycosylation begins with the synthesis of donor lipid-linked oligosaccharides (LLOs), such as Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-dolichol (GlcNAc=*N*-acetylglucosamine), followed by its *en bloc* transfer onto an acceptor polypeptide in the presence of the multi-subunit enzyme, oligosaccharide transferase (OST).

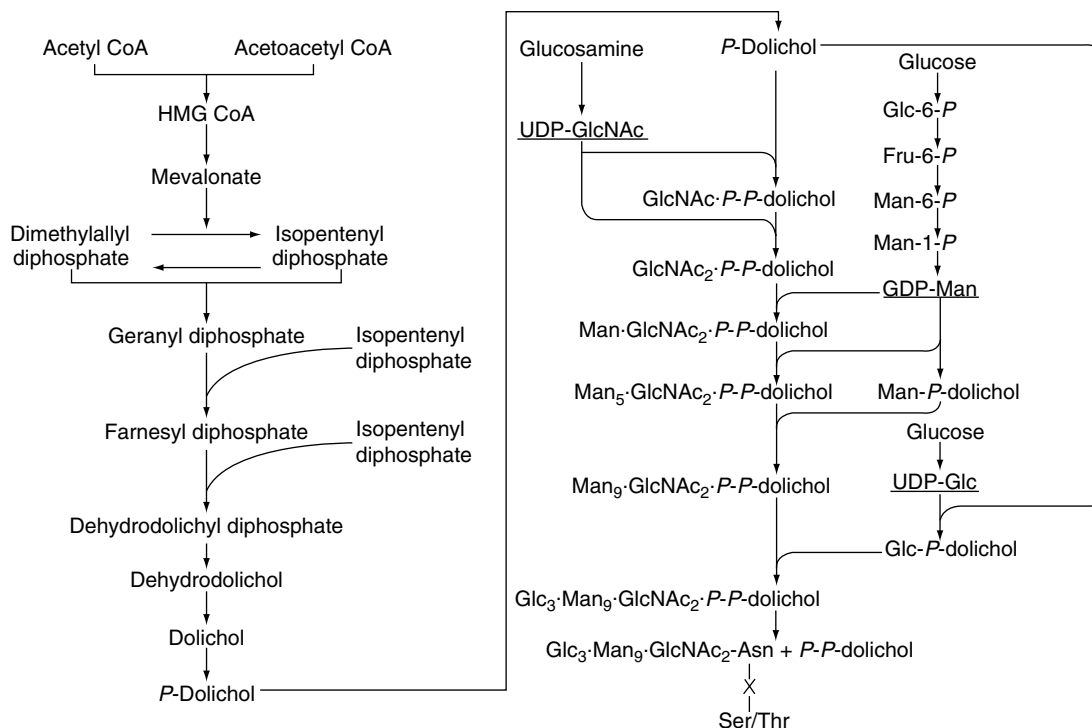
### 4.32.2.1 Generation of LLOs

#### 4.32.2.1.1 Biosynthesis of dolichol phosphate

The generation of LLOs begins *in vivo* with the synthesis of a lipid carrier, dolichol (Dol), followed by the progressive addition of monosaccharides onto a growing chain to form the donor substrate, Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-Dol (DLO).<sup>6–9</sup> The lipid glycosyl carrier, dolichol, is the longest aliphatic molecule known to be synthesized in mammalian cells. It is a polyisoprenoid lipid (16–20 isoprene units)<sup>9–11</sup> formed as an end product of the prenol biosynthesis pathway in eukaryotes<sup>3,6</sup> and prokaryotes.<sup>12</sup> The polyprenol precursor of dolichol is embedded into the cytoplasmic surface of the ER membrane. Synthesis of the dolichol phosphate (Dol-*P*) occurs in a multistep biosynthetic pathway from acetyl CoA as shown in **Figure 1**.<sup>8,9</sup> Dolichol synthesis begins with the elongation of farnesyl pyrophosphate (FPP), using isopentenyl pyrophosphate (IPP) into dehydrodolichol diphosphate by *cis*-prenyl transferase (CPT).<sup>13</sup> This polyprenyl-diphosphate is hydrophobic and is embedded in the ER membrane.<sup>14</sup> Mammalian dolichol, unlike its prokaryotic counterpart, has a saturated  $\alpha$ -isoprene subunit. Finally, the phosphatase<sup>14–16</sup> and reductase<sup>17</sup> activities dephosphorylate and reduce the  $\alpha$ -isoprene units<sup>13,14,18</sup> to form dolichol with saturated  $\alpha$ -isoprene units. The presence of the saturated  $\alpha$ -isoprene unit is important in the recognition of the lipid as a glycosyl donor.<sup>19</sup> In addition, dolichol kinase rephosphorylates free dolichol produced in the pathway.<sup>20,21</sup>

#### 4.32.2.1.2 Assembly of oligosaccharide moiety onto dolichol phosphate

Once Dol-*P* is formed, the core structure glycan begins to assemble on the cytosolic side of the ER. A membrane-associated glycosyl transferase, UDP-*N*-acetylglucosamine-dolichyl-phosphate *N*-acetylglucosaminophosphotransferase, in presence of a soluble cytosolic nucleotide-activated sugar donor, uridine-5'-diphospho-*N*-acetylglucosamine (UDP-GlcNAc), transfers *N*-acetylglucosamine-phosphate (GlcNAc-*P*) to form *N*-acetylglucosaminyl-diphosphodolichol (GlcNAc-*P-P*-Dol)<sup>22</sup> (**Figure 1**). Subsequently, *N*-acetylglucosaminyl diphosphodolichol *N*-acetylglucosaminyltransferase followed by GDP-mannose: chitobiosyldiphosphodolichol  $\beta$ -D-mannosyltransferase and several  $\alpha$ -mannosyltransferases sequentially add the second GlcNAc and five mannose residues from UDP-GlcNAc and guanosine-5'-diphosphomannose (GDP-Man) to generate the intermediate, Man<sub>5</sub>GlcNAc<sub>2</sub>-*P-P*-Dol.<sup>23</sup> This intermediate is then 'flipped' across the ER membrane into the luminal side.<sup>23–25</sup> It has been suggested that this 'flipping' is due in part to bilayer instability<sup>23,26–28</sup> or catalyzed by a protein flippase<sup>29</sup> on the membrane. The instability might be caused by high local concentrations of polyisoprenols with a phosphate head group such as Dol-*P* which can decrease fluidity and disrupt the membrane bilayer. Recently, it has been reported that in yeast, a membrane protein, Rft1, is



**Figure 1** Metabolic pathway for synthesis of DLO donor substrate,  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-dolichol}$ .

required for the flipping of a heptasaccharide from the cytoplasmic to the luminal leaflet.<sup>23,25,30,31</sup> On the luminal side of the ER, mannosyltransferases and glucosyltransferases sequentially add four more mannose and three glucose residues using Dol-*P*-Man and Dol-*P*-Glc as donor substrates, to form the complete LLO,  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$ .<sup>6</sup> Finally, the transfer of the oligosaccharide to the growing polypeptide (see Section 4.32.2.2) generates Dol-*PP*, which is converted to Dol-*P* to begin another N-glycosylation cycle.<sup>8,9</sup>

#### 4.32.2.2 Transfer of the Oligosaccharide Moiety to Nascent Polypeptide

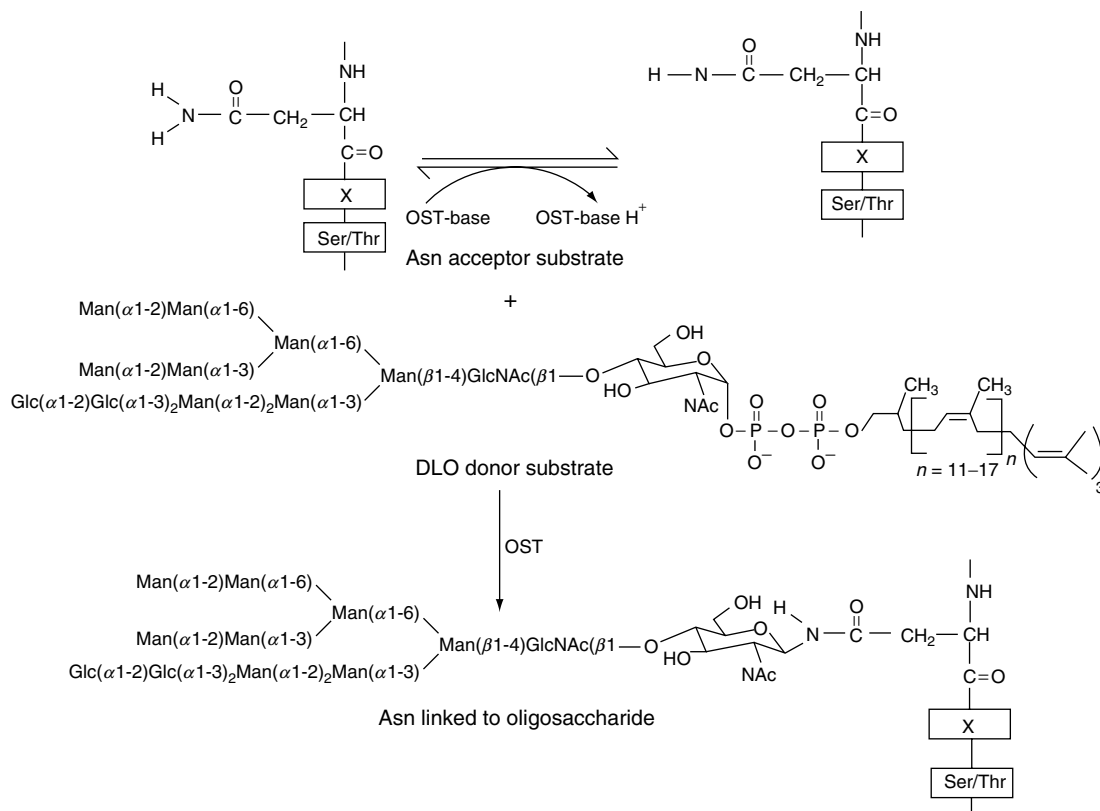
The potential N-glycosylation site of a protein is the asparagine residue in the tripeptide sequon, Asn-X-Ser/Thr, where X can be any amino acid except for proline.<sup>32</sup> Once the LLO is formed, oligosaccharyltransferase (OST) complex<sup>6,31</sup> binds to it and cleaves the phosphoglycosidic bond,  $\text{GlcNAc-P}$ , to mediate the *en bloc* transfer of  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  oligosaccharide onto the nascent polypeptide along with the release of dolichol pyrophosphate (Dol-*PP*)<sup>3,6</sup> (Figure 2). Although the composition of OST complex is still subject to debate, the roles, interactions, and structural characteristics of individual proteins of the complex have been elucidated over the past decade and many are described in a later section.<sup>35–37</sup>

### 4.32.3 N-Glycan Site Occupancy

#### 4.32.3.1 Importance of N-Glycan Site Occupancy

The N-glycan attachment influences many properties of a glycoprotein, including its folding, stability, solubility, *in vivo* antigenicity, clearance rate, and biological activity. In addition, oligosaccharide attachments facilitate folding, mediate receptor–ligand and cell–cell interactions, and play a role in directing intracellular trafficking, providing resistance to proteolysis, and altering *in vivo* protein clearance. Several studies have investigated the influence of site occupancy, including one study examining its effect on the stability, intracellular trafficking, and cell surface expression of KV1.4, a member of the potassium channel family.<sup>38</sup>

A commercially important biopharmaceutical drug used to dissolve blood clots, human tissue-type plasminogen activator (t-PA), has three potential N-glycosylation sites (Asn-117, -184, and -448), each of which is important for its physiological properties and biological activity. When expressed in mammalian cells, two glycoforms of t-PA are

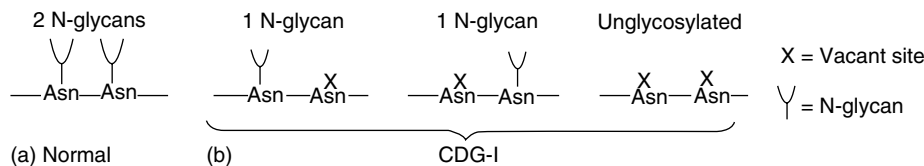


**Figure 2** OST catalyzes transfer of oligosaccharide, Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>, onto an Asn substrate.

observed: type I (all sites glycosylated) and type II (Asn-184 unglycosylated).<sup>39,40</sup> While the presence of complex oligosaccharides at both Asn-184 and Asn-448 is responsible for fibrin-stimulated enzymatic activity and binding affinity for lysine,<sup>2,40,41</sup> the N-glycan occupancy at Asn-117 is known to affect its clearance from circulation. In the case of hemagglutinin (HA) of influenza virus, which has six N-glycosylation sites, glycosylation at Asn-81 influenced the ability of the protein to acquire its native disulfide bonds in order to fold properly and exit the ER.<sup>42,43</sup> In some cases, like that of erythropoietin (EPO), higher site occupancy, coupled to further modifications such as sialylation, serves to improve the secretion, half-life, and solubility of glycoproteins. Although, nonglycosylated EPO was active *in vitro* compared to its glycosylated counterpart *in vitro*,<sup>44-46</sup> removal of three N-glycans on EPO lowered its production level by 90% and the *in vivo* biological activity by more than 90%.<sup>46</sup> Recent studies have shown that a second-generation EPO with five N-glycan sites was more effective *in vivo* than the native form of EPO. This was due to an increased circulatory half-life of the engineered variant and an increased efficacy and *in vivo* activity with an increase in the number of N-glycan attachments.<sup>47</sup> Furthermore, elimination of the glycosylation sites from transferrin (Tf) reduced its secretion level by nearly an order of magnitude<sup>48</sup> and unglycosylated Tf underwent rapid aggregation and precipitation.<sup>49</sup>

The site occupancy has also been shown to affect physiological properties of glycoproteins. For example, site occupancy of interferon-gamma (IFN- $\gamma$ ) at one of the two N-glycosylation sites, Asn-25, affects the protease resistance of the protein.<sup>50-54</sup> Studies involving site-directed mutagenesis have showed a drastic reduction in the secretion of IFN- $\gamma$  in insect cells and also in the biological activity in its dimeric form<sup>55</sup> with the elimination of the Asn-25 glycosylation site. The attachment of an N-glycan increases the overall stability of ribonuclease A, and lowers its susceptibility to proteolysis.<sup>56</sup> A mutation in tyrosinase that eliminates one N-glycan attachment results in oculocutaneous albinism of the skin, eyes, and hair.<sup>57</sup> In addition, various studies have revealed the importance of site occupancy in folding, lysosomal targeting (human tripeptidyl-peptidase I),<sup>58</sup> resistance to proteases,<sup>59-62</sup> and cell surface expression and surface stability (human dopamine transporter).<sup>63</sup>

More recently, a collection of diseases called congenital disorders of glycosylation (CDGs)<sup>64-66</sup> have been linked to site occupancy limitations. These diseases have been found so far to be primarily defects in ability to generate the complete dolichol-linked oligosaccharide (DLO) substrate, Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-P-P-Dol (CDG-I) or in the



**Figure 3** (a) Normal hTf and (b) underglycosylated hTf from CDG-I patients.

subsequent processing of protein-bound glycans (CDG-II).<sup>64</sup> A number of biochemical steps have been implicated in CDG-I disorders including 11 different steps in the DLO biosynthesis pathway (CDG-Ia through CDG-Ik) as well as other unspecified defects in the pathway (CDG-X). Clinical manifestations can vary including childhood mortality, organ failure, neurological dysfunction, and developmental delays.<sup>64–66</sup> The most widely used clinical marker for CDG-I is the accumulation of abnormal forms of Tf in serum and cerebrospinal fluid. While healthy humans generate human transferrin (hTf) with two occupied N-glycosylation sites, CDGs patients have increased levels of hTf with one occupied N-glycosylation site or nonglycosylated hTf<sup>49,64,67</sup> (**Figure 3**). Interestingly, alcoholics have also been observed to include similar defects in their transferrin glycosylation.<sup>8,68</sup>

#### 4.32.3.2 Engineering N-Glycan Site Occupancy

An understanding of the factors that control site occupancy or macroheterogeneity can be useful in controlling N-glycan site occupancy. In recent years, scientists have worked toward (1) regulating the molecular flux in the dolichol pathway, (2) understanding and controlling the function, structural organization, and other characteristics of the proteins in the OST complex, and (3) predicting the best possible residues in and around the acceptor tripeptide sequon, Asn-X-Ser/Thr, as well as the location of this tripeptide sequon within the protein structure. However, there are still many elements of the site occupancy process that are not yet understood.

##### 4.32.3.2.1 Control of dolichol phosphate synthesis

The level of Dol-*P* has been shown to be an important factor affecting N-glycosylation.<sup>69–72</sup> There have been several studies in this regard, which have been presented in numerous reviews.<sup>7,14</sup> A summary of the pathway, along with studies describing its limitations, is presented in the following sections. (For details on biosynthesis of N-glycans, see Chapter 3.01).

###### 4.32.3.2.1.1 Mevalonate synthesis

The initial steps of the dolichol synthesis pathway involve formation of acetoacetyl-coenzyme A (CoA) from acetyl-CoA, followed by a condensation reaction which results in formation of  $\beta$ -hydroxy- $\beta$ -methylglutaryl CoA (HMG-CoA). Subsequently, HMG-CoA reductase in the presence of reduced nicotinamide adenine dinucleotide phosphate (NADPH) reduces HMG-CoA to form mevalonate as shown in the pathway in **Figure 1**. This reduction step plays a critical role in the early regulation of dolichol synthesis pathway.<sup>18</sup> Studies have shown that inhibition of HMG-CoA reductase decreased the level of dolichol phosphate, N-glycosylation, and inhibited cell growth in SK-MEL-2, a melanoma cell line, and Ewing's sarcoma cells.<sup>73,74</sup> Similar results were observed in aortic smooth muscle cells,<sup>75</sup> rat liver slices,<sup>76</sup> developing sea urchin embryos,<sup>77</sup> and human embryonic kidney 293 (HEK-293) cells (J. Jones, S. Krag, and M. Betenbaugh, unpublished data). Using exogenously supplemented mevalonate with L-cells, dolichol phosphate synthesis was observed to be not regulated by the level of intracellular mevalonate but by CPT, another enzyme in the pathway.<sup>74,78</sup>

###### 4.32.3.2.1.2 *cis*-Prenyl transferase

The elongation reaction in *de novo* Dol-*P* synthesis is catalyzed by CPT. Several studies in the past have suggested the reaction catalyzed by CPT to be an important regulatory step in the biosynthesis of dolichol (**Figure 1**).<sup>79,80</sup> More recently, it has been shown that the induction of CPT activity by bacterial lipopolysaccharide (LPS) increased the rate of Dol-*P* and lipid intermediate biosynthesis in proliferating murine B lymphocytes,<sup>81</sup> embryonic rat brain,<sup>82</sup> and cyclic AMP-treated JEG-3 choriocarcinoma cells.<sup>83</sup> The recent cloning of cDNAs encoding the gene from *Arabidopsis thaliana*,<sup>84,85</sup> *Saccharomyces cerevisiae*,<sup>86,87</sup> and human cells<sup>88,89</sup> has facilitated a better understanding of the importance of this enzyme for N-glycan site occupancy.<sup>90</sup> Indeed, the overexpression of human CPT cDNA was able to

complement defects in growth, dolichol synthesis, and site occupancy of carboxypeptidase Y (CPY) expressed in yeast *rer2* mutant cells.<sup>89</sup> Another group showed that expression of the human CPT gene reverted the temperature sensitivity of mutant *rer2-2* yeast cells deficient in dehydrodolichyl diphosphate (Dedol-*P-P*) synthase activity.<sup>88</sup> Also, the overexpression of the yeast CPT increased dolichol synthesis and protein glycosylation of secreted proteins from the fungal host, *Trichoderma reesei*.<sup>91</sup> We have shown that overexpression of heterologous CPT increases the levels of dolichol in mammalian cell lines as well.<sup>92</sup>

#### 4.32.3.2.1.3 Dolichol kinase

In the final step of the *de novo* dolichol synthesis pathway, free dolichol is phosphorylated by dolichol kinase to form Dol-*P* (Figure 1). Dolichol kinase activity was initially found in sea urchin embryos,<sup>93</sup> and its activity has also been detected in mammalian cells.<sup>21,94,95</sup> Dolichol kinase may play a vital role in controlling the level of Dol-*P*,<sup>18</sup> and also in altering Dol-*P* levels through recycling of dolichol following the complete dephosphorylation of Dol-*P-P*.<sup>6,14</sup>

The level of intracellular dolichol phosphate clearly affects rate of the N-glycosylation reaction. Several studies have showed an increase in the rate and level of N-glycosylation in mammalian cell culture when exogenous dolichol phosphate is added.<sup>7,98–101</sup> However, a recent study showed that exogenous Dol-*P* was incorporated into LLO, but there was no marked effect on cellular glycosylation levels.<sup>102</sup> This may indicate that the effect of exogenous Dol-*P* on intracellular levels of lipid intermediates and N-glycosylation may vary with cell and tissue type and the efficiency with which they are able to incorporate the exogenous Dol-*P*. The dolichol kinase gene is required in yeast for cell viability, lipid synthesis, and protein N-glycosylation. Furthermore, a human dolichol kinase has also been found to complement growth defects, increase Dol-*P* synthesis, and restore CPY N-glycosylation in a yeast mutant with defect endogenous dolichol kinase.<sup>103</sup>

### 4.32.3.2.2 Control of LLO synthesis

#### 4.32.3.2.2.1 GlcNAc-*P-P*-Dol, Man-*P*-Dol, and Glc-*P*-Dol

Studies suggest that the level of Dol-*P* influences the rate of formation of GlcNAc-*P-P*-Dol, Man-*P*-Dol, and Glc-*P*-Dol.<sup>71</sup> The three corresponding enzymes involved in the formation of these intermediates, *N*-acetylglucosaminyl-1-*P* transferase (GPT), mannosylphosphoryldolichol (MPD) synthase, and glucosylphosphoryldolichol (GPD) synthase, utilize the same pool of Dol-*P*,<sup>18,104</sup> and all have similar  $K_m$  values.<sup>18</sup> In addition, biosynthesis of LLO intermediates is also regulated by feedback mechanisms. Man-*P*-Dol regulates GPT which catalyzes the formation of GlcNAc-*P-P*-Dol and the level of GlcNAc-*P-P*-Dol in turn controls the biosynthesis of Man-*P*-Dol<sup>105,106</sup> as well as inhibits its own biosynthesis by product inhibition of GPT. Similarly, GlcNAc-GlcNAc-*P-P*-Dol also inhibits its own biosynthesis by inhibiting the enzyme responsible for the transfer of GlcNAc from UDP-GlcNAc to GlcNAc-*P-P*-Dol.<sup>107</sup> However, it has been shown that Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-Dol does not stimulate feedback inhibition of LLO and LLO intermediates by GPT, MPD synthase, GPD synthase, and GDP-mannose-dependent transferases.<sup>104</sup> This study contradicts previous work which proposed that the LLO biosynthesis inhibition was caused by the lack of availability of acceptor polypeptides and ‘free’ Dol-*P* as substrates<sup>108</sup> and that the accumulation of Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-Dol acted as a feedback inhibitor of GPT.<sup>101,109</sup>

#### 4.32.3.2.2.2 Conversion of Man<sub>5</sub>GlcNAc<sub>2</sub>-*P-P*-Dol to Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-Dol

During the LLO biosynthesis pathway, Man<sub>5</sub>GlcNAc<sub>2</sub>-*P-P*-Dol is flipped from the cytoplasmic to the luminal surface of the ER, and is sequentially catalyzed by luminal glycosyltransferases to form the final oligosaccharide moiety, Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-Dol<sup>3,110</sup> (Figure 1). The seven glycosyltransferases that add the four mannose and three glucose residues, and three enzymes involved in the ‘flipping’ of Man<sub>5</sub>GlcNAc<sub>2</sub>-*P-P*-Dol and the two donor substrates, Man-*P*-Dol and Glc-*P*-Dol control the rate of this conversion.<sup>14</sup> Man-*P*-Dol and Glc-*P*-Dol, which donate the monosaccharides on the luminal surface of the ER, are synthesized on the cytoplasmic side and thus their transport may be yet another rate-determining factor.<sup>6</sup> The flipping of molecules to the luminal leaflet may also be controlled by a ‘bulk flow’ effect, activated by their constant consumption in every round of N-glycosylation and providing translocation directionality.<sup>14,23</sup> After transferring the oligosaccharide moiety to the protein, the resulting Dol-*P* and Dol-*P-P* return to the cytosolic face of the membrane. This decrease in the concentration of the lipids on the luminal surface stimulates the influx of more Man-*P*-Dol and Glc-*P*-Dol.<sup>14,23</sup> In addition, the availability of sugar nucleotides also affects the levels of lipid-linked intermediates and the LLO donor substrate and can induce an accumulation of the truncated LLO donor substrate.<sup>111</sup> Many of the genetic disorders causing CDGs contain defects in the pathways



leading to the synthesis of Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-*P-P*-Dol from GlcNAc-*P-P*-Dol. Unfortunately, there is currently little effective treatment yet for any of the diseases except CDG-Ib, which is treated with mannose supplementation.<sup>65</sup> However, as the genes responsible for these enzymatic conversions are identified, the prospects for genetic therapy and other metabolic interventions will become increasingly possible. Furthermore, as more research is done toward elucidating the kinetics and enzymology of this DLO pathway, the potential exists that provides a better understanding of the critical steps in the N-glycosylation of proteins.

#### 4.32.3.2.3 Control of oligosaccharyltransfer activity

The transfer of the preformed oligosaccharide moiety (Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>) from LLO onto nascent polypeptides as it is translocated to the lumen of the ER is catalyzed by the action of the OST complex.<sup>30,31</sup> The mechanism of this enzymatic step starts with the scanning of the polypeptide for possible N-glycosylation sites bearing the tripeptide sequon Asn-X-Ser/Thr and directing it to the OST active site in the proper conformation. This is followed by positioning the active site of OST complex near the translocon complex, recognizing and moving Dol-*P-P*-oligosaccharide into the OST active site<sup>112</sup> and finally catalyzing the transfer. Thus the availability and location of this multi-unit OST complex relative to the nascent polypeptide is yet another possible limitation to N-glycan site occupancy. Studies in *S. cerevisiae* have revealed eight membrane-bound subunits of OST (Ost1p, Ost2p, Ost3p/Ost6p, Ost4p, Ost5p, Wbp1p, Swp1p, and Stt3p)<sup>113–118</sup> that exist in three subcomplexes: Wbp1p–Swp1p–Ost2p, Ost1p–Ost5p, and Stt3p–Ost3p–Ost4p.<sup>31,113</sup> It was suggested that interactions between the OST subunits occur through either their transmembrane domains and/or a region proximal to it.<sup>118</sup> It has been found that Ost1, Ost2, Stt3, Wbp1, and Swp1 are essential for viability in yeast, and deletion of Ost3, Ost5, and Ost6 results in viable cells with significant underglycosylation.<sup>113,116,118</sup> The mammalian OST subunits OST48, ribophorin II, DAD 1, and ribophorin I are homologous to Wbp1p, Swp1p, Ost2p, and Ost1p, respectively.<sup>112,113,117,119,120</sup> Other mammalian homologs like STT3-A/STT3-B, homologs of yeast Stt3p are expressed in two different isoforms depending on the cell and tissue type<sup>121</sup> (Table 1).

Wbp1p may be involved in recognition and binding to the LLO substrate<sup>122</sup> and may be essential for association of the LLO with other subunits in the OST complex through its ER transmembrane domain.<sup>117</sup> Of the other two subunits of the subcomplex containing Wbp1p, the Ost2p is almost entirely embedded in the ER membrane<sup>123</sup> and Swp1p has three transmembrane domains.<sup>113,124</sup> Studies with Ost5p, a 9 kDa protein with two transmembrane domains,<sup>113</sup> have suggested that its absence is not lethal to cells but lowers glycosylation efficiency.<sup>125</sup> The smallest of all the yeast OST subunits, Ost4p, contains only 36 amino acids, most of which are within the ER membrane. Kim *et al.* have reported its interaction with the Stt3p and Ost3p via a portion of the transmembrane domain facing the cytosol.<sup>115</sup> The other two subunits in the subcomplex containing Ost4p are Ost3p and Stt3p. Proteins expressed in mutant yeast strains lacking Ost3p, a 34kDa subunit with many transmembrane domains, displayed underglycosylation.<sup>126</sup> Ost6p, which is highly homologous to Ost3p, is proposed to have the same function.<sup>113</sup> Recent studies have indicated that two different OST isoforms are present in yeast and differ only in the inclusion of Ost6p or Ost3p.<sup>127,128</sup> These subunits are not essential for OST activity but alter affinity for different substrates. Furthermore, the Ost4p subunit facilitates the assembly of Ost6p and Ost3p into the OST complex. Understanding the functional and structural interactions between each subunit will be invaluable to understanding and regulating N-glycosylation and site occupancy of glycoproteins.

Of the subunits, Stt3p appears to play a central conserved role in N-glycosylation as it is conserved across kingdoms.<sup>6,113,129–133</sup> Stt3p, an essential protein, is the largest member of OST with 11 transmembrane domains and contains a binding site responsible for peptide recognition and/or catalysis of the glycosylation process.<sup>132,134</sup> The mammalian homolog to Stt3p may be responsible for recognizing the consensus sequence on the nascent

**Table 1** Comparison of OST subunits in yeast and mammals

<i>Yeast (S. cerevisiae) subunits</i>	<i>Mammalian (H. sapiens) subunits</i>
Ost1p (62kDa)	Ribophorin I (66kDa)
Swp1p (30kDa)	Ribophorin II (63/64kDa)
Wbp1p (45kDa)	Ost48 (48kDa)
Ost2p (15kDa)	DAD1 (10kDa)
Stt3p (82kDa)	STT3-A and STT3-B
Ost3p/Ost6p	N33 and IAP

polypeptide.<sup>112</sup> Mammalian cells contain two STT3 isoforms, STT3A and STT3B, and the levels of STT3A vary significantly in different tissues.<sup>121</sup> Interestingly, STT3A appears to be highest in tissues such as liver and pancreas with higher secretion rates as compared to levels in brain, lung, and kidney. Also, STT3B exhibits higher catalytic activity, but STT3A is more selective toward the proper DLO substrate. Furthermore, the total amount of OST is also relatively higher in B-cell lines as compared to fibroblasts to indicate that transferase levels vary according to the processing demands of a specific cell type.

#### 4.32.3.2.4 Acceptor tripeptide sequon

The availability of the consensus sequence, Asn-X-Ser/Thr, where X is any amino acid except for proline, on the nascent polypeptide chain is necessary for N-glycosylation.<sup>32</sup> At times, atypical tripeptide sequons, Asn-X-Cys and Asn-Gly-Gly-Thr, have also been observed to be N-glycosylated.<sup>32,135,136</sup> It has also been shown that the glycosylation efficiency may be influenced by cysteine residues near the tripeptide sequon.<sup>137,138</sup> The variability in occupancy of the tripeptide sequon with the oligosaccharide, called macroheterogeneity, presents a major challenge for the biopharmaceutical industry. It may be desirable to eliminate macroheterogenic properties in a product mixture and produce a pharmaceutical product with consistent site occupancy and quality. The efficiency of oligosaccharide addition onto the asparagine residue is dependent on the accessibility of the sequon to the active site of OST complex. The efficiency of oligosaccharide transfer onto the consensus sequence is affected by a number of factors described below.

##### 4.32.3.2.4.1 The presence of serine or threonine on the hydroxyl amino acid

Studies have revealed that the tripeptide sequon, Asn-X-Thr, is 2–3 times more efficiently glycosylated than Asn-X-Ser.<sup>32,139</sup> An interesting study involving the site-directed mutagenesis of serine to threonine showed a dramatic increase in the efficiency and cell surface expression of rabies virus glycoprotein (RGP), a protein well known to be variably glycosylated at one of the glycosylation sites, Asn-37.<sup>137</sup> Even in the presence of residues in the X position believed to be inhibitory to glycosylation, the protein was efficiently glycosylated when threonine, but not serine, was present in the hydroxy position.<sup>140</sup>

##### 4.32.3.2.4.2 The amino acid in the X position of the consensus sequence

Using site-directed mutagenesis to study the effect of the amino acid at the X position, variants of RGP protein containing each of the 20 amino acids in the X position were created next to Asn-37.<sup>138</sup> Recently, a bioinformatics study confirmed that the presence of proline at the X position in the tripeptide sequon does indeed inhibit N-glycosylation.<sup>141</sup> It has been reported that in addition to proline, the presence of tryptophan (Trp), aspartic acid (Asp), glutamic acid (Glu), and leucine (Leu) at the X-position has led to inefficient glycosylation.<sup>142</sup>

##### 4.32.3.2.4.3 The location of the sequence on polypeptide chain and its proximity to the C-terminus

Statistical studies indicated that nonglycosylated sequons tend to occur more frequently toward the C-terminus of a protein.<sup>32</sup> The N-glycosylation of the tripeptide sequon of prion protein (PrP) lacking the GPI anchor expressed in neuronal cells increased as the distance from the C-terminus increased.<sup>143</sup> These findings suggest that the utilization of the tripeptide sequon in PrP is a co-translational event and that the sequon must reach the active site of the OST complex prior to chain termination.<sup>143,144</sup> Alternatively, PrP sequons introduced into membrane dipeptidase (MDP), another GPI-anchored protein, were N-glycosylated in the same cell type even when they were fewer than 60 residues from the C-terminus. Another bioinformatics study<sup>141</sup> examined the occupancy of N-glycosylation sites of well-characterized glycoproteins from the protein database and found that a significant proportion of the glycoproteins studied contained an occupied site within 10 residues of the C-terminus. Thus, the OST complex utilizes the same sequon in similar positions near the C-terminus in two proteins differently, and N-glycosylation might occur co-translationally (as previously thought) as well as post-translationally.

##### 4.32.3.2.4.4 Local amino acids flanking the consensus sequence

The transfer of the oligosaccharide by OST complex to the nascent polypeptide requires a hydrogen bond between the amide group of Asn and the hydroxy group of Ser or Thr in the tripeptide sequon. This makes the side chain more nucleophilic and is facilitated by a loop in the polypeptide backbone.<sup>145</sup> Because of proline's inhibitory role in the formation of a loop, it would be unfavorable in the X position of the tripeptide sequon.<sup>146</sup> Thus, the residues near the



sequon, which affect the conformation of the sequon, and hydrogen bonding are important for site occupancy efficiency.<sup>147–149</sup> Studies suggest that residues at the Y position of Asn-X-Ser/Thr-Y also influence the glycosylation efficiency of a protein.<sup>32,146,149</sup> The presence of proline at the Y position abolishes N-glycosylation.<sup>32,141</sup> Whereas cysteine residues in place of the Ser/Thr residues in the tripeptide sequon will allow N-glycosylation, its presence in the Y position immediately following the sequon is inhibitory to glycosylation due to the possible formation of disulfide bonds in the polypeptide chain that hinder desirable conformations.<sup>137,146</sup> Although studies have also been done up to 15 and 20 amino acid positions flanking the tripeptide sequons on either side,<sup>32,141</sup> no concrete conclusion regarding their influence on site occupancy has been made.

#### 4.32.3.2.4.5 Conformation of the polypeptide chain near the consensus sequence

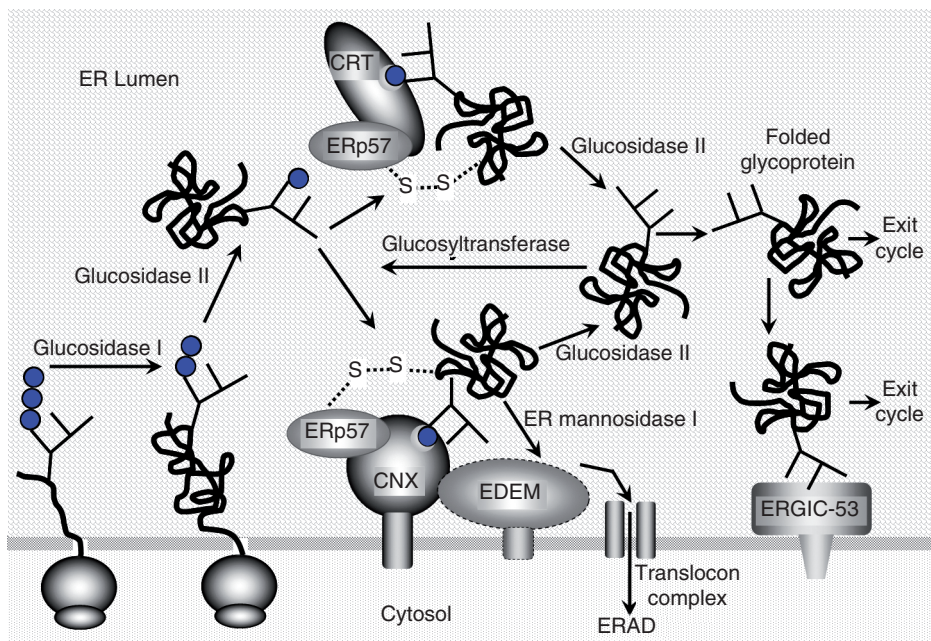
The accessibility, and consequently N-glycosylation, of the tripeptide sequon is influenced by conformation of the local region and especially the presence of disulfide bonds.<sup>153,154</sup> Co-translational disulfide bond formation in t-PA inhibits N-glycosylation at a variably occupied site (Asn-184) on the protein,<sup>39</sup> which may be due to steric hindrance inhibiting the accessibility of the sequon to the OST complex. Statistical analysis of the region flanking the N-glycosylation sites revealed that there was an increase in probability of finding an aromatic amino acid in close proximity to the tripeptide sequon. Indeed, glycosylation sites are biased toward turns and bends, where secondary structure changes, and occupied sites are mostly found more frequently on exposed convex surfaces of the polypeptide structure. This structural characteristic might play a role in the control of site occupancy and evolutionary selection of possible glycosylation sites.<sup>155</sup> Several other studies have contributed to our insights on the conformation around the tripeptide sequon,<sup>32,59,156</sup> but none has claimed any direct influence on site occupancy.

### 4.32.4 Microheterogeneity: Control of N-Glycan Processing in the Secretory Compartments

Once the oligosaccharide has been transferred to the polypeptide, its processing begins in the ER. (For details on N-glycan processing, see Chapter 3.02.) The first few steps in the pathway are conserved among most eukaryotes including mammals, insects, and yeast. First, the terminal  $\alpha$ -1-2-linked glucose is removed by the membrane-bound ER resident protein, glucosidase I,<sup>157,158</sup> which is conserved in most species except trypanosomes.<sup>159</sup> This is followed by the removal of the two remaining glucose residues by a soluble resident enzyme, glucosidase II.<sup>34,153,158</sup> Interestingly, a recent study suggested that glucosidase II removes these terminal glucose residues more efficiently from peptides or human pancreatic ribonuclease with more than one N-glycans than those with zero or one glycan.<sup>160</sup> Furthermore, the processing of the terminal glucose residues and a few other residues involves critical signals for controlling the interaction of proteins with calnexin and calreticulin to mediate the correct folding and processing of the protein as below (for review, see Refs: 161 and 162).

#### 4.32.4.1 Glycosidases and Lectins in the ER

The ER is a key checkpoint for quality control and assisting proper folding of proteins. This is accomplished with the help of ER-resident homologous lectins, calnexin (membrane bound) and calreticulin (soluble). Following the formation of the monoglucosylated oligosaccharide, Glc<sub>1</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>, the glycoprotein enters the ‘calnexin/calreticulin’ cycle during which the ER lectins specifically interact and bind to the glycoprotein.<sup>163–166</sup> The cysteines of calnexin- and calreticulin-bound glycoproteins are able to form a complex with Erp57, a thiol oxidoreductase from the protein disulfide isomerase (PDI) family, via transient disulfide bonds (Figure 4). This facilitates the glycoprotein to fold into its native conformation.<sup>167–170</sup> Subsequently, two soluble luminal ER enzymes, glucosidase II and UDP-glucose:glycoprotein glucosyltransferase (GT), further mediate the interactions of the glycoprotein with calnexin and calreticulin through opposing actions. Glucosidase II removes the remaining glucose residue on the monoglucosylated glycan structure, thereby dissociating the substrate glycoprotein from the lectins<sup>153,171–173</sup> (Figure 4). If the glycoprotein is improperly folded, GT, which serves as a folding sensor, reglucosylates the glycoprotein to renew its interaction with calnexin and calreticulin.<sup>173–175</sup> This on–off cycle continues until the glycoprotein has acquired its native conformation, at which point the glycoprotein exits the calnexin/calreticulin cycle and continues along the secretory pathway. Thus, the interaction between calnexin or calreticulin and its partner proteins with the substrate glycoprotein assists glycoproteins in increasing their stability and secretion rates.<sup>172</sup> Studies have shown that enhancing calnexin association by adding castanospermine, a glucosidase I inhibitor, alters the secretion rates and promotes

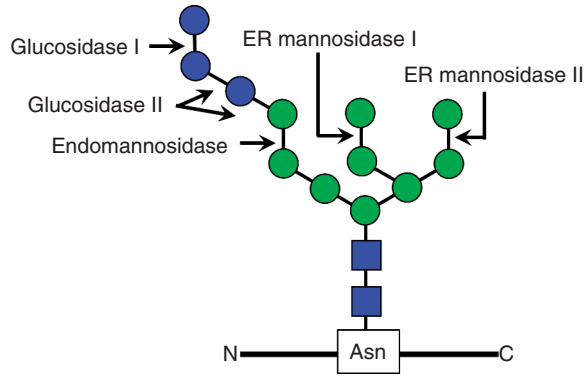


**Figure 4** The calnexin/calreticulin cycle for monitoring folding of glycosylated proteins in the ER.

degradation of glycoproteins like rat hepatic lipase and nicotinic choline receptors.<sup>176,177</sup> Furthermore, the interaction with calnexin may be rate limiting for the folding and secretion of target glycoproteins. Consequently, overexpression of heterologous calnexin has been shown to increase the processing and yields of complex membrane and secreted proteins in mammalian and insect cell lines.<sup>178,179</sup> However, the binding of calnexin and calreticulin is not required for generating an enzymatically active final structure for certain proteins such as t-PA.<sup>180</sup> As mentioned above, GT reglucosylates improperly folded glycoprotein to renew its interaction with calnexin and calreticulin. However, after repeated cycles, if the protein is persistently improperly folded, it is retrotranslocated through the ER membrane into the cytosol.<sup>172,181,182</sup> The protein is then deglycosylated by a cytosolic glycoamidase (peptide-*N*-glycosidase), ubiquitinated, and degraded in the proteasome by ER-associated degradation (ERAD) process<sup>182–185</sup> (Figure 4).

Following the removal of all three glucose residues from  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ , the ER and Golgi  $\alpha$ -1,2-mannosidases act together to remove four ( $\alpha$ 1-2)-linked mannose residues from  $\text{Man}_9\text{GlcNAc}_2$  to form  $\text{Man}_5\text{GlcNAc}_2$ . The  $\alpha$ -1,2-mannosidases, essential for processing to complex and hybrid N-glycans in mammalian cells,<sup>3,161,186</sup> generate a series of oligomannose-type N-glycans in the ER and the Golgi. One of the  $\alpha$ -1,2-mannosidases, the ER-resident  $\alpha$ -1,2-mannosidase I specifically removes a single mannose residue from  $\text{Man}_9\text{GlcNAc}_2$  to form  $\text{Man}_8\text{GlcNAc}_2$  isomer B that lacks a single terminal ( $\alpha$ 1-2)-mannose from the middle branch of the N-glycan<sup>187,188</sup> (Figure 5). The ER contains another  $\alpha$ -mannosidase (ER  $\alpha$ -mannosidase II) which generates the  $\text{Man}_8\text{GlcNAc}_2$  isomer C.<sup>189,190</sup> Htm1/Mln1 or EDEM, an ER mannosidase I homolog lacking mannosidase activity, has been implicated as a lectin that can recognize the  $\text{Man}_8\text{GlcNAc}_2$  B-isomer<sup>191–193</sup> and target misfolded glycoproteins to ERAD<sup>194</sup> (Figure 5). The elimination of Htm1/Mln 1 in yeast resulted in retardation of glycoprotein degradation,<sup>193</sup> while the overexpression of its mammalian counterpart, EDEM, resulted in the acceleration of ERAD of misfolded glycoproteins but not of nonglycosylated proteins.<sup>195–197</sup> Furthermore, the combined expression of both EDEM and ER Man I accelerated the degradation of unfolded proteins in mammalian HEK 293 cell lines.<sup>198</sup>

It has been postulated that ERGIC-53 (ER-Golgi intermediate compartment protein), a homo-oligomeric type I transmembrane protein with mannose-specific lectin activity, functions as a transport receptor for the export of certain glycoproteins from ER to the Golgi.<sup>153,158,158,158</sup> It accumulates in the ER/Golgi intermediate compartment,<sup>158</sup> and shuttles between the ER/Golgi intermediate compartment and the *cis*-Golgi.<sup>201,202</sup> Another recently identified homolog of ERGIC-53, the vesicular integral membrane protein 36 (VIP 36),<sup>203</sup> is concentrated early in the secretory pathway.<sup>204,205</sup> This is a membrane protein with lectin specificity for oligomannose structure and is believed to mediate the intracellular transport of glycoproteins through the secretory pathway.<sup>206,207</sup> However, its precise role in glycoprotein sorting and trafficking is not clear.



**Figure 5** The core N-linked glycan structure. The core oligosaccharide is assembled onto dolichol-*P* and transferred en bloc to an Asn residue in the tripeptide sequon Asn-X-Ser/Thr on the nascent polypeptide. Glycosidases involved in trimming of the oligosaccharides are indicated.

#### 4.32.4.2 Glycosidases and Glycosyltransferases in the Golgi

#### 4.32.4.2.1 Trimming of the terminal mannose

Once the oligosaccharide has been trimmed to  $\text{Man}_8\text{GlcNAc}_2$ , processing diverges between higher eukaryotes and lower eukaryotes as will be described subsequently for yeast. In mammalian cells, three different Golgi resident  $\alpha$ -1,2-mannosidases termed Golgi  $\alpha$ -1,2-mannosidase IA, IB, and IC play a key role to form  $\text{Man}_5\text{GlcNAc}$ .<sup>208,209</sup> However, the terminal ( $\alpha$ 1-2)-mannose attached to the middle branch of the N-glycan precursor is relatively resistant to these Golgi enzymes. In this way, the specificity of ER  $\alpha$ -1,2-mannosidase I, which cleaves this particular mannose residue, is complementary to that of the Golgi mannosidases.<sup>187,208,209</sup>

#### 4.32.4.2.2 Synthesis of GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> intermediate

Following the synthesis of  $\text{Man}_5\text{GlcNAc}_2$  intermediate, a single GlcNAc residue is transferred by the action of *N*-acetylglucosaminyltransferase I (GnT I) to the  $\text{Man}(\alpha 1-3)$ -branch of  $\text{Man}_5\text{GlcNAc}_2$  from UDP-GlcNAc.<sup>210</sup> The reaction catalyzed by GnT I is the initial step in the eventual biosynthesis of complex-type N-glycan. At this point, two enzymes participate in modification of the  $\text{GlcNAcMan}_5\text{GlcNAc}_2$ . In the case of mammalian cell-expressed N-glycans, the Asn-linked GlcNAc residue is often modified with an  $(\alpha 1-6)$ -linked fucose. This fucosylation first takes place on the  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  by the action of the core  $\alpha$ -1,6-fucosyltransferase,<sup>211,212</sup> which requires the presence of  $(\beta 1-2)$ -linked GlcNAc on the  $\text{Man}(\alpha 1-3)$ -branch of the core  $\text{Man}_3\text{GlcNAc}_2$  glycan. Other biantennary oligosaccharides can serve as substrates for this mammalian core  $\alpha$ -1,6-fucosyltransferase.<sup>211,213,214</sup> A second modification that depends on prior GnT I activity is the action of  $\alpha$ -mannosidase II, which cleaves off the remaining  $(\alpha 1-3)$ - and a terminal  $(\alpha 1-6)$ -linked Man residues from the  $\text{GlcNAcMan}_5[\pm\text{Fuc}(\alpha 1-6)]\text{GlcNAc}_2$  to yield  $\text{GlcNAcMan}_3[\pm\text{Fuc}(\alpha 1-6)]\text{GlcNAc}_2$ .<sup>215</sup> Mutant Chinese hamster ovary (CHO) and human embryonic kidney HEK293 cell lines have been used in which GnT I is downregulated in order to express heterologous proteins terminating in  $\text{Man}_5\text{GlcNAc}_2$ .<sup>216,217</sup> The generation of N-glycans with such consistent oligomannose structures can be advantageous for structural and functional studies of secreted and membrane proteins. In addition, less common processing pathways have been proposed in which particular  $\alpha$ -mannosidases can remove the Man residues initially to yield  $\text{Man}_3\text{GlcNAc}_2$  followed by the addition of a GlcNAc residue using GnT I.<sup>210,218–221</sup>

In mammalian cells, GlcNAcMan<sub>3</sub>GlcNAc synthesized through either the classical pathway or the alternative pathway serves as an acceptor for *N*-acetylglucosaminyltransferase II (GnT II), which adds a second GlcNAc to the Man(α1-6)-branch to yield biantennary complex-type N-glycans terminating with two (α1-2)-linked GlcNAc residues. GnT II gene has been cloned from human<sup>222</sup> and rat.<sup>223</sup> In addition to GnT I and II, three other *N*-acetylglucosaminyltransferases, GnT IV, V, and VI, further add additional GlcNAc to the (α1-3)- or (α1-6)-linked Man to fabricate multiantennary N-glycans (**Figure 6**). Alternatively, *N*-acetylglucosaminyltransferase III (GnT III) adds a bisecting GlcNAc to the β-linked Man.

The addition of a bisecting GlcNAc serves to inhibit the action of other enzymes including  $\alpha$ -1,6-fucosyltransferase.<sup>221</sup> Umana *et al.*<sup>224</sup> overexpressed GnT III, which is downregulated in CHO cells, and found that the