Chemical and biological approaches to glycoprotein synthesis

Roslyn M Bill[1] and Sabine L Flitsch[2]

Protein glycosylation is a common posttranslational modification that produces glycoproteins that are highly complex in terms of both their structure and in their function. Systematic structure–function studies of such glycoproteins require synthetic methods that can produce homogeneous glycoproteins with defined oligosaccharide sidechains.

Addresses: [1]Department of Chemistry, The University of Michigan, Ann Arbor, MI 48109-1055, USA and [2]Department of Chemistry, The University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK.

Chemistry & Biology March 1996, **3**:145–149

 ISSN 1074-5521

The ubiquity of protein glycosylation is well established: glycoproteins are widely distributed in eukaryotic cells and may also be found in prokaryotes. The best characterized examples of protein–oligosaccharide bonds in glycoproteins are (i) *N*-glycosidic linkages between *N*-acetylglucosamine at the reducing end of an oligosaccharide chain and certain asparagines in the sequence Asn-X-Ser/Thr, and (ii) *O*-glycosidic linkages between a reducing *N*-acetylgalactosamine and certain serine or threonine residues. Three examples of the wide range of oligosaccharide chains typically associated with *N*-glycosidic linkages are illustrated in Figure 1. Combining biosynthetic pathways of glycosylation with chemical synthesis can provide homogeneous preparations of glycoproteins for structure–function studies and, perhaps, for the clinic.

Biosynthesis

The biosynthesis of glycoproteins is well documented and is similar in all eukaryotes. For example, *N*-glycosylation is known to be concurrent with synthesis of the nascent polypeptide chain into the endoplasmic reticulum, where a common oligosaccharide precursor is transferred from a glycolipid to the nascent protein. Modification of this oligosaccharide precursor also occurs in the endoplasmic reticulum before the protein is transported to the Golgi apparatus for further modification by the host cell's complement of glycosidases and glycosyltransferases. Each organism has its own 'glycosylation machinery' and thus expression of the same gene in different organisms produces different glycoproteins. Moreover, the efficiency of one or more biosynthetic steps may be low, resulting in a mixture of oligosaccharide sidechains of different length and composition at each glycosylation site. Therefore a glycoprotein is generally a heterogeneous mixture of 'glycoforms'; protein molecules which differ only in the structure of the bound oligosaccharide.

The functional role of oligosaccharides in glycoproteins

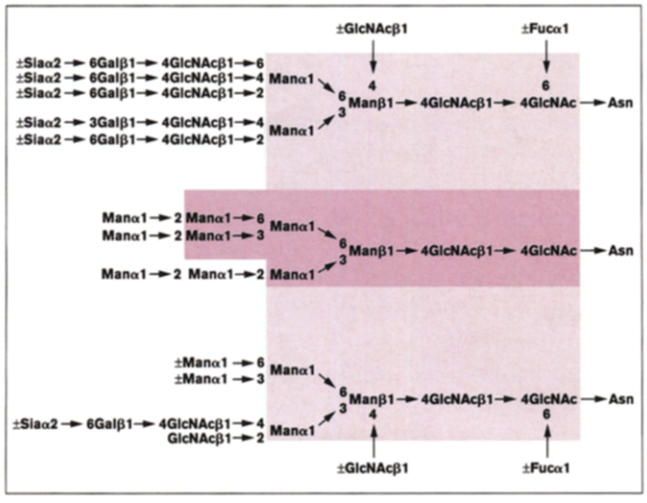
The reason for the heterogeneity of protein glycosylation is still not well understood. But how much do the oligosaccharide chains influence the properties of glycoproteins? It is now established that the different glycoforms of a protein can have a range of different physical, biochemical and biological properties [1]. Thus, the heterogeneous glycosylations seem to constitute a subtle and sophisticated mechanism of biological control. Regulatory authorities in the United States have recognized the significance of such glycoforms in the case of tissue plasminogen activator, a recombinant fibrinolytic agent. The glycoforms have been shown to have differing fibrin-dependent plasminogenolytic activities [2], and the United States Patent Office has therefore ruled that particular glycoforms must be specified in patents, even though all forms have the same amino-acid sequence.

The most obvious way to determine the importance of an oligosaccharide chain is to remove it. Fully deglycosylated proteins can be prepared either by site-directed mutagenesis of the gene to remove the glycosylation sites or by chemical and enzymatic deglycosylation of the protein. Partial deglycosylation of the protein can be achieved using glycosidases. Comparison of the structure and activity of this range of different proteins suggests that the biological roles of oligosaccharides vary from those that are seemingly trivial to those that are crucial for the development, growth, function and survival of the organism [3]. These roles include: (i) increasing the stability of the glycoprotein to heat or proteolysis, (ii) modulating cell–cell interactions, (iii) altering protein–protein interactions, (iv) modulating biological activity and (v) targeting of the protein to particular sites within the cell or organism.

Removing the oligosaccharide chains completely is a rather crude way to assess their function, however. There have been relatively few studies so far that have compared different homogeneous glycoforms of a glycoprotein. One recent example is that of bovine pancreatic ribonuclease, which has been isolated as five electrophoretically pure glycoforms; these molecules have a four-fold variation in enzymatic activity towards double-stranded RNA [4].

A dramatic recent demonstration of the importance of complex or hybrid *N*-glycosylation of glycoproteins *in vivo* came from studies of a mouse *N*-acetylglucosaminyltransferase gene (*Mgat-I*) [5]. This transferase initiates the formation of complex and hybrid *N*-linked carbohydrates (Fig. 1), and mouse embryos lacking the *Mgat-I* gene die at mid-gestation. Thus, the presence of these *N*-linked

Figure 1



Three major subgroups of mammalian *N*-linked oligosaccharides. From the top the structures are complex-type, high mannose-type and hybrid-type. The structure within the large shaded box is common to all *N*-linked oligosaccharides. The structure within the small shaded box is the common core of the majority of high mannose oligosaccharides. Outside these lines the structures are variable. Adapted from [18].

carbohydrates is essential, and lack of them is lethal at a very early stage of development.

So far then, the only thing we can say with certainty about the role of oligosaccharides is that they mediate specific recognition events or modulate biological processes [3]. The presence of oligosaccharides expands the possibilities for specific recognition far beyond those allowed by the limited number of proteins encoded in the genome of an organism [3].The functional diversity thus generated is crucial for the evolution and development of an organism.

How can we understand glycoproteins better?

To understand the relationship between oligosaccharide structure and glycoprotein function, homogeneous glycoproteins containing oligosaccharide sidechains of defined structure must be obtained either by isolation from glycoprotein mixtures or by *de novo* synthesis. Chromatographic methods, in particular capillary electrophoresis [4], can now give access to pure glycoforms, but such methods are limited to glycoforms that are naturally available and abundant. Synthetic approaches, however, can produce a much wider range of both natural and unnatural glycoproteins. The resultant synthetic glycoproteins can have better physical stability than the natural form, or, in the case of therapeutic glycoproteins, may show reduced toxicity and desirable pharmacokinetic and pharmacodynamic properties.

The complex structure of the glycoproteins themselves and the complexity of the biosynthetic pathways make either chemical or biological synthesis of glycoproteins a formidable task. Nonetheless, some significant advances using both biological and chemical methodologies have been made in recent years, including glycosylation engineering, enzymatic synthesis and chemical synthesis, which are discussed below.

Glycosylation engineering

The concept of 'glycosylation engineering', that is, the manipulation of protein oligosaccharide sidechains by expression of genes in different hosts, has been discussed by Stanley [6]. Since each organism or cell has its own complement of glycosylation enzymes, glycosylation patterns will be biased in a predictable way. For example, when a gene is expressed in yeast, the resulting glycoprotein has highly mannosylated oligosaccharides, whereas when the same gene is expressed in most mammalian cell lines, the product is a glycoprotein with diverse complex oligosaccharides capped with sialic acids and galactosides (see Fig. 1). This approach can be extended by using mutant cell lines in which the biosynthetic pathways have been interrupted; the products of such cells are biosynthetic intermediates of the glycosylation pathway, containing truncated oligosaccharide sidechains. Interruption of the biosynthetic pathway can also be achieved by inhibitors such as tunicamycin, which prevents all asparagine-linked glycosylation.

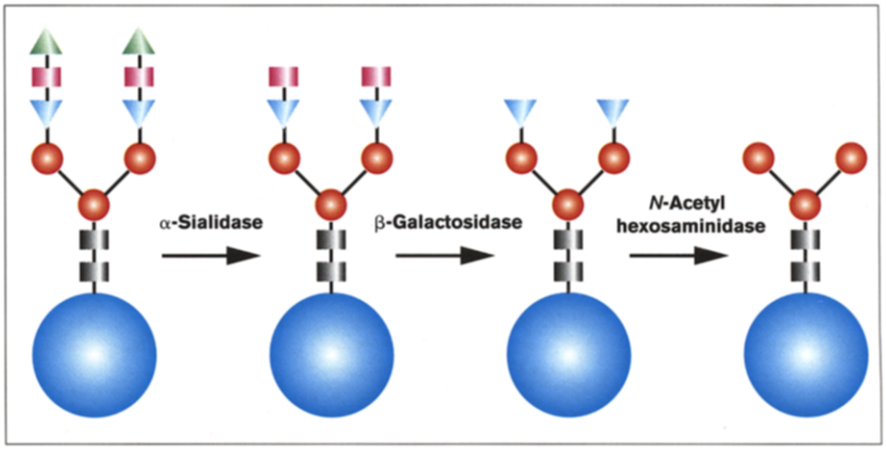
Mutant cell lines can also be produced by adding key glycosyltransferases to the complement of a cell's glycosidation enzymes. This has already been successful with several glycosyltransferases, such as the addition of α2–6 sialyl transferase to a Chinese hamster ovary cell line [7]. But all such *in vivo* approaches have the disadvantage that their products are heterogeneous. Thus, if we wish to study homogeneous (or unnatural) glycoproteins, we must turn to *in vitro* methods.

Enzymatic synthesis of glycoproteins

Glycosidases and glycosyltransferases are not only integral to *in vivo* glycosylation, but are also useful for *in vitro* processing. A range of these enzymes has been made available through isolation from natural sources or by heterologous overexpression in foreign hosts; these enzymes have been shown to accept saccharides, glycolipids and glycoproteins as substrates. By using an excess of enzyme, high yielding reactions are possible, even with poor substrates. This approach has been used successfully to modify glucocerebrosidase so that it can be used as an enzyme replacement therapy for Gaucher's disease, (an inherited deficiency). Glucocerebrosidase expressed in a mammalian system contains oligosaccharides terminating in sialic acids. But to target glucocerebrosidase to the mannose receptors of reticulo-endothelial cells, where the enzyme function is required, the protein must have terminal mannose residues. To expose the mannosides, the glycoprotein is

Figure 2

Glycoprotein remodelling of glucocerebrosidase for the treatment of Gaucher's disease [8,9]. For efficient targeting to the reticulo-endothelial system, the oligosaccharides found on glucocerebrosidase made in tissue culture must be trimmed with glycosidases to expose mannoside residues.



trimmed with glycosidases (sialidase, galactosidase and hexosaminidase), as shown in Figure 2 [8,9].

It is also possible to add to the complexity of an oligosaccharide using glycosyltransferases, which attach monosaccharides in a regio- and stereo-selective manner to oligosaccharide sidechains. For example, β1,4-galactosyltransferase from bovine colostrum primarily recognizes terminal *N*-acetyl glucosaminide (GlcNAc) as an acceptor substrate and catalyzes the selective formation of the terminal Galβ1→4GlcNAc disaccharide sequence. Such enzymatic *in vitro* synthesis is not limited to natural saccharides but can also be used for introducing labelled saccharides into glycoproteins or to generate analogues. The recognition process is often reasonably flexible, for instance, β1,4-galactosyltransferase can transfer galactose, glucose or *N*-acetyl galactosamine to oligosaccharide chains ending in GlcNAc [10].

Another important enzyme in the *N*-linked glycosylation pathway is oligosaccharyltransferase, which catalyses the glycosylation of asparagine residues on the nascent protein chain as it leaves the ribosome [11] (Fig. 3). This enzyme has been purified and can be obtained in large enough quantities for *in vitro* glycosylation studies of peptides containing asparagine sites for *N*-glycosylation. Glycosylation of

Figure 3

The transfer of a tetradecasaccharide from a lipid precursor to the nascent polypeptide chain is a key step in the formation of *N*-linked oligosaccharides [12].

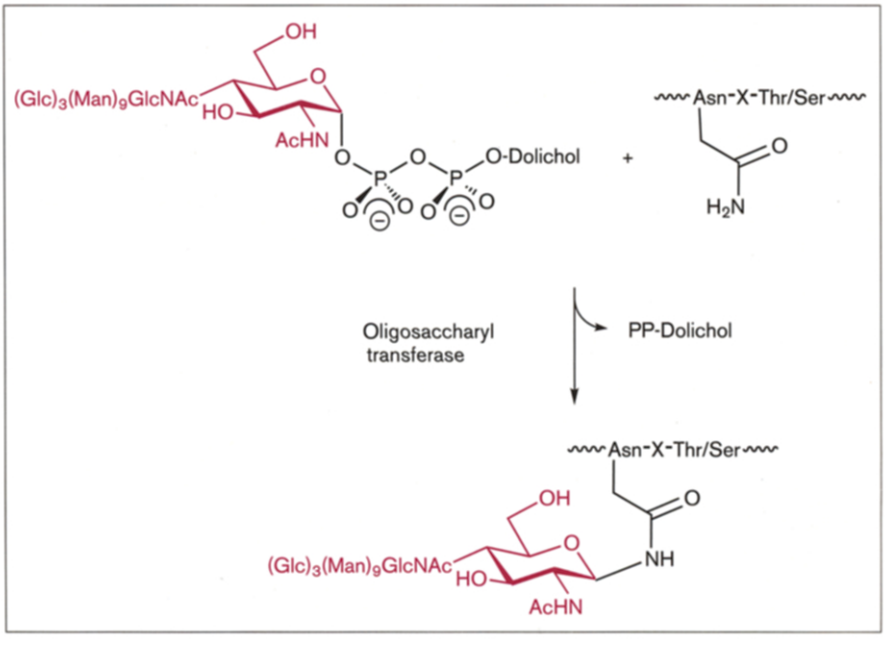
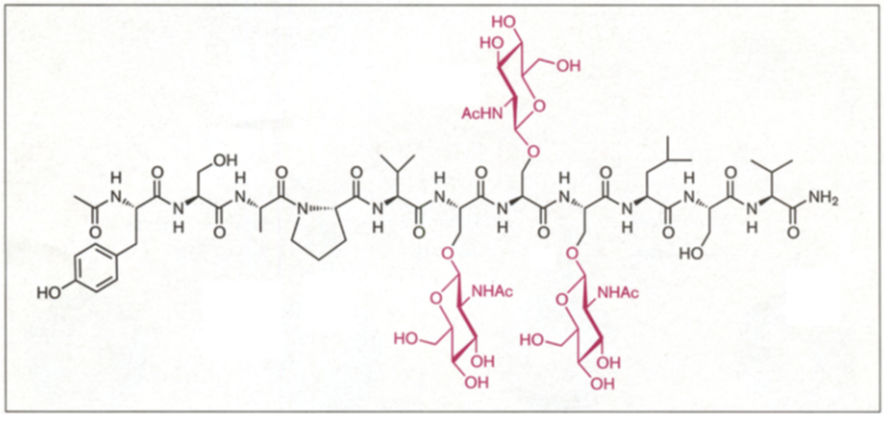


Figure 4



Chemically synthesized undecapeptide containing *O*-linked *N*-acetyl glucosamine residues [13].

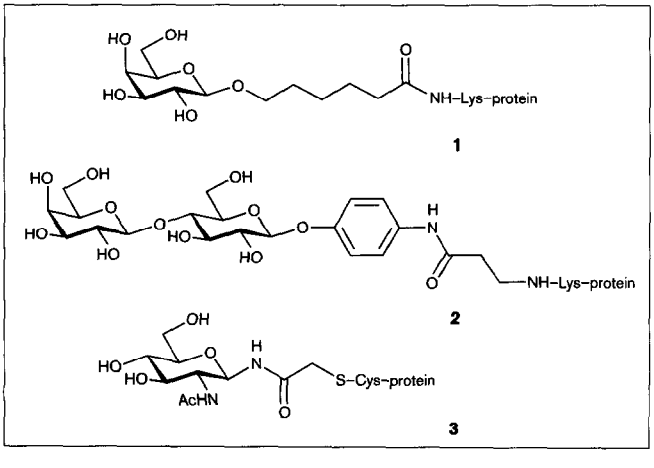
whole proteins has proved to be more difficult [12], however, except when the proteins are in a denatured state.

Chemical synthesis of glycoproteins

The chemical synthesis of glycopeptides has become possible in recent years as milder deprotection methods have been developed for glycopeptide synthesis. Thus, it is now possible to incorporate glycosylated amino acids during solid-phase synthesis. This is particularly useful for small oligosaccharide sidechains, such as those found in *O*-linked glycans. A peptide that has recently been prepared by solid-phase peptide synthesis is shown in Figure 4 [13].

Alternatively, suitably protected peptides can be synthesized first, then chemically glycosylated. This is particularly useful for *N*-linked oligosaccharides, since the glycosyl amide linkage can be synthesized by coupling an aspartate sidechain of the peptide with easily available glycosylamines [14]. Neither of these methods has been applied to chemical synthesis of large glycoproteins to date. But as the size limit for solid phase protein synthesis has now been extended to about 170 amino acids, the chemical synthesis of glycoproteins such as erythropoietin (166 amino acids), an important glycoprotein hormone that is used in the treatment of anemia, should now be feasible.

Figure 5



Neoglycoprotein linkages **1** [19], **2** [20] and **3** [16,17].

Chemical glycosylation methods are not restricted to the synthesis of natural glycoproteins but can also be used for generating 'neoglycoproteins', which contain unnatural linkages between protein and saccharides, and are often chosen because they are more easy to synthesize [15]. Such glycosylation can dramatically improve the physical and biological properties of proteins. Neoglycoproteins have been synthesized by using the inherent reactivity of lysine and cysteine sidechains in proteins to achieve selective glycosylation. The resulting linkages are often structurally very different from the natural linkage (such as linkages **1** and **2** in Figure 5), and thus neoglycoproteins might have limited applications in structure-function studies but are useful as novel protein based therapeutics. They can also provide carbohydrate antigens and immunogens from which immunodiagnostic and therapeutic reagents can be derived. Some more native-like linkages (**3**) have recently been generated by the reaction of glycosyl-*N*-iodoacetamides with cysteine residues [16,17].

What next?

The study of protein glycosylation integrates several disciplines, including chemistry, biochemistry and molecular and cell biology. In the field of chemistry, recent advances in the organic synthesis of peptides and carbohydrates suggest that the total synthesis of glycoproteins could soon be possible, yielding homogeneous products of defined structure. Complementing such efforts, glycosyltransferases and glycosidases have been studied intensively and can now be used either in cell culture or as isolated catalysts for glycoprotein synthesis. These chemical and enzymatic approaches should synergize; organic synthesis

can provide unnatural glycoproteins and can also supply suitable natural and unnatural substrates for the glycosyltransferases and glycosidases. Where enzymes with the appropriate specificity exist and are available in sufficient quantities, enzymatic synthesis is faster and much more selective than chemical synthesis. We predict that, a practical route to the synthesis of glycoproteins will probably involve a mixture of chemical and biological techniques.

References

1. Rademacher, T.W., Parekh, R.B. & Dwek, R.A. (1988). Glycobiology. *Annu. Rev. Biochem.* **57**, 785–838.
2. Parekh, R.B., *et al.*, & Tiemeier, D.C. (1989). *N*-glycosylation and *in vitro* enzymatic activity of human recombinant tissue plasminogen activator expressed in Chinese hamster ovary cells and a murine cell line. *Biochemistry* **28**, 7670–7679.
3. Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130.
4. Rudd, P.M., *et al.*, & Dwek, R.A. (1994). Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **33**, 17–22.
5. Ioffe, E. & Stanley, P. (1994). Mice lacking *N*-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid *N*-linked carbohydrates. *Proc. Natl. Acad. Sci. USA* **91**, 728–732.
6. Stanley, P. (1992). Glycosylation engineering. *Glycobiology* **2**, 99–107.
7. Lee, E.U., Roth, J. & Paulson, J.C. (1989). Alteration of terminal glycosylation sequences on *N*-linked oligosaccharides of Chinese hamster ovary cells by expression of β galactoside α2,6-sialyltransferase. *J. Biol. Chem.* **264**, 13848–13855.
8. Barton, N.W., *et al.*, & Yu, K.T. (1991). Replacement therapy for inherited enzyme deficiency – macrophage targetted glucocerebrosidase for Gaucher's disease. *N. Engl. J. Med.* **324**, 1464–1470.
9. Friedman, B., Hubbard, S.C. & Rasmussen, J.R. (1993). Development of a recombinant form of Ceredase® (glucocerebrosidase) for the treatment of Gaucher's disease. *Glycoconjugate J.* **10**, 257–257.
10. Palcic, M.M. & Hindsgaul, O. (1991). Flexibility in the donor substrate specificity of β1,4-galactosyltransferase: application in the synthesis of complex carbohydrates. *Glycobiology* **1**, 205–209.
11. Imperiali, B. & Henrickson, T.L. (1995). Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* **3**, 1565–1578.
12. Liu, Y.-L., Hoops, G.C., Coward, J.K. (1994). A comparison of proteins and peptides as substrates for microsomal and solubilised oligosaccharyltransferase. *Bioorg. Med. Chem.* **2**, 1133–1141.
13. Meinjohanns, E., Vargas-Berenguel, A., Meldal, M., Paulsen, H. & Bock, K. (1995). Comparison of *N*-Dts and *N*-Aloc in the solid-phase syntheses of *O*-GlcNAc glycopeptide fragments of RNA-polymerase II and mammalian neurofilaments. *J. Chem. Soc., Perkin Trans.* **I** 2165–2175.
14. Anisfeld, S.T. & Lansbury, P.T. (1990). A convergent approach to the chemical synthesis of asparagine-linked glycopeptides. *J. Org. Chem.* **55**, 5560–5562.
15. Stowell, C.P. & Lee, Y.C. (1980). Neoglycoproteins. The preparation and application of synthetic glycoproteins. *Adv. Carbohydr. Chem. Biochem.* **37**, 225–281.
16. Davis, N.J. & Flitsch, S.L. (1991). A novel method for the specific glycosylation of proteins. *Tetrahedron Lett.* **32**, 6793–6796.
17. Wong, S.Y.C., Guile, G.R., Dwek, R.A. & Arsequell, G. (1994). Synthetic glycosylation of proteins using *N*-(beta-saccharide) iodoacetamides – applications in site-specific glycosylation and solid-phase enzymatic oligosaccharide synthesis. *Biochem. J.* **300**, 843–850.
18. Kobata, A. (1993). Glycobiology: an expanding research area in carbohydrate chemistry. *Accounts Chem. Res.* **26**, 319–324.
19. Sabesan, S. & Linna, T.J. (1994). Chemical glycosylation of recombinant interleukin-2. *Methods Enzymol.* **242**, 46–56.
20. Roy, R., Tropper, F.D., Morrison, T. & Boratynski, J. (1991). Synthesis and transformation of glycohydrolase substrates into protein conjugates based on Michael additions. *J. Chem. Soc., Chem. Commun.* 536–538.