We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 3

Structural Biology of Glycoproteins

Joanne E. Nettleship

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48154

1. Introduction

In recent years, the field of structural biology has seen many advances in technology for the production of recombinant proteins, mainly led by the high-throughput techniques of the structural genomics community. These technologies have largely focussed on expression using *Escherichia coli* with around 88 % of the protein chains in the Protein Data Bank (PDB) in 2010 being produced using this bacterium [1]. However, many proteins require post-translational modification for correct folding and/or activity. By far the most common modification is glycosylation and it has been estimated that over half of all human proteins including many membrane proteins are glycosylated [2-3]. Therefore, the production of correctly glycosylated proteins for functional and structural studies requires the use of eukaryotic systems [4].

Glycoproteins represent a unique challenge to the structural biologist due to the size and heterogeneity of the oligosaccharide chains. Glycans can constitute from 1 % to over 80 % of the total protein mass [5] with variation in the type and number of sugars attached to a glycosylation site and also the occupancy of a site. The heterogeneity introduced by glycosylation can hinder crystallization of a glycoprotein [4] thus limiting the use of X-ray crystallography, one of the major techniques using in protein structure solution.

The following chapter reviews options for the production of glycoproteins from different types of expression system, both prokaryotic and eukaryotic, along with methods for addressing glycan heterogeneity in order to produce homogeneous glycoprotein samples which are amenable for structural studies. The use of such homogeneous glycoprotein samples in both crystallization and nuclear magnetic resonance (NMR) experiments are discussed.

2. Glycosylation in mammalian cells

In eukaryotes, glycans are added to the polypeptide chain in the endoplamic reticulum (ER) and the Golgi as the protein is secreted (see Figure 1). There are two types of

INTECH
open science | open minds

© 2012 Nettleship, licensee InTech. This is an open access chapter distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

glycosylation, one involves linkage of an oligosaccahride to an asparagine residue and the second involves linkage to a serine or threonine residue referred to as N- and O-glycosylation respectively.

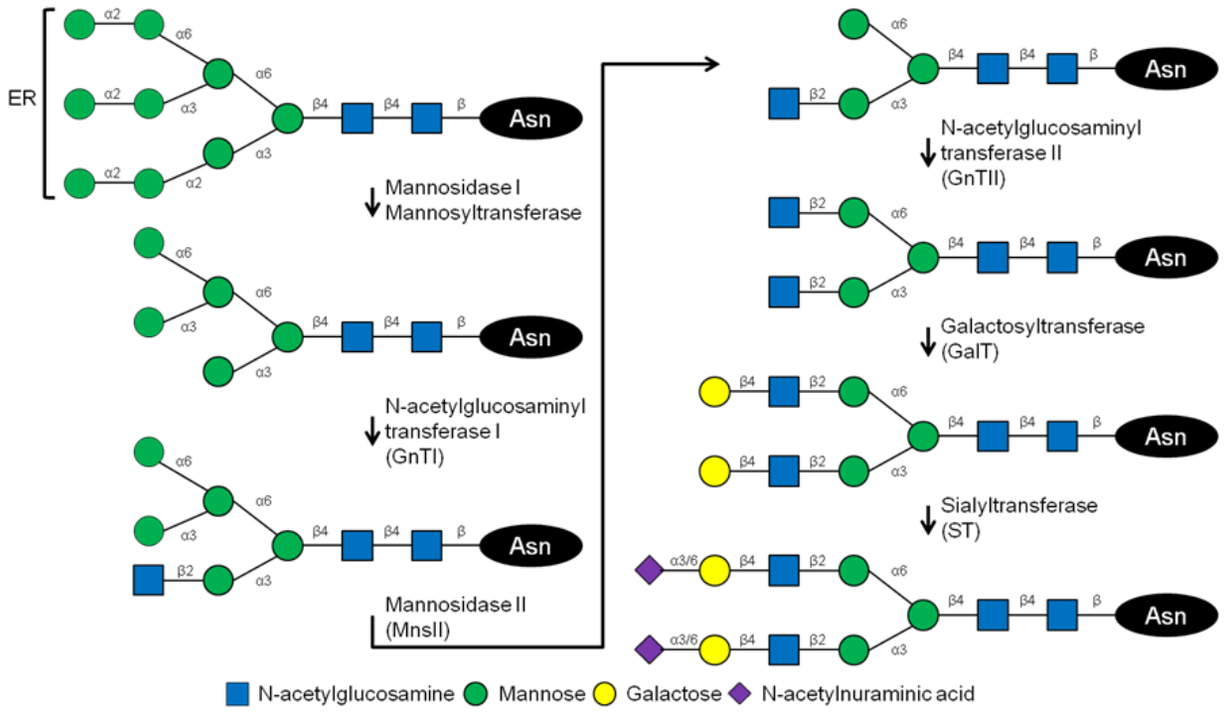


Figure 1. Simplified representation of the N-glycosylation pathway in humans with the initial core $Man_9GlcNAc_2$ sugar being formed in the endoplasmic reticulum (ER) and further modifications taking place in the Golgi.

N-glycosylation occurs co-translationally in the ER at the consensus sequon Asn-X-Ser/Thr/Cys-X where X is any amino acid except proline. The exact sequence of the sequon has a bearing on the occupancy of the glycosylation site. N-glycosylation is often essential for the expression and folding of a glycoprotein [6], with the initial glycan formed in the ER being further modified and decorated in the Golgi apparatus. This elaboration of the glycan core leads to a large number of possible structures which are classified as high mannose, complex or hybrid (Figure 2).

O-glycosylation (See Figure 3) usually occurs in regions containing large numbers of sequential serine, threonine and proline residues which are known as mucin domains. These regions show little secondary structure and are therefore often excluded from structural studies as proteins containing such disorder are unlikely to crystallize. O-glycosylation occurring outside mucin regions is difficult to predict accurately and so is usually only detected after production of a protein. As O-glycosylation occurs in the Golgi, it has little bearing on the early stages of protein folding and therefore sites found to be O-glycosylated can be engineered out of the protein by site-directed mutagenesis of the acceptor Ser/Thr residues.

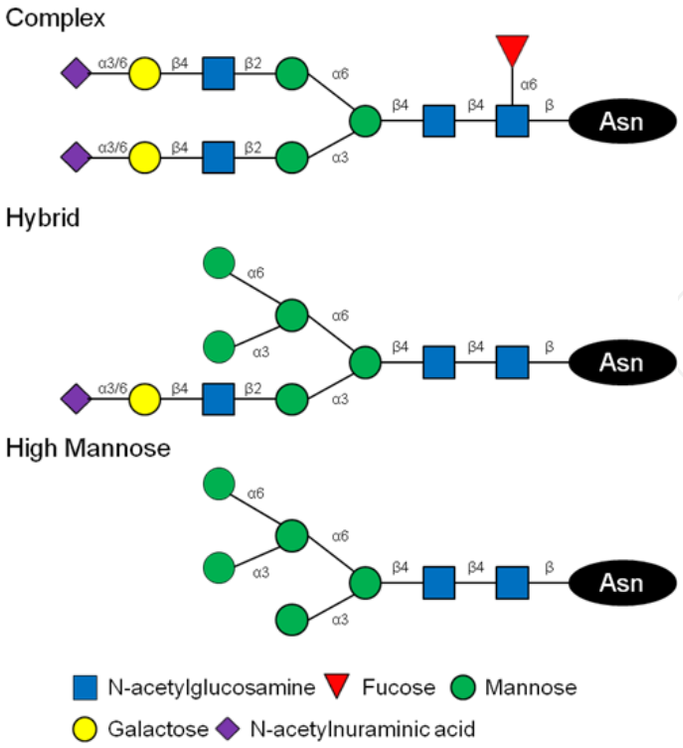


Figure 2. Diagrammatic representation of the major classes of N-linked sugars produced by mammalian cells.

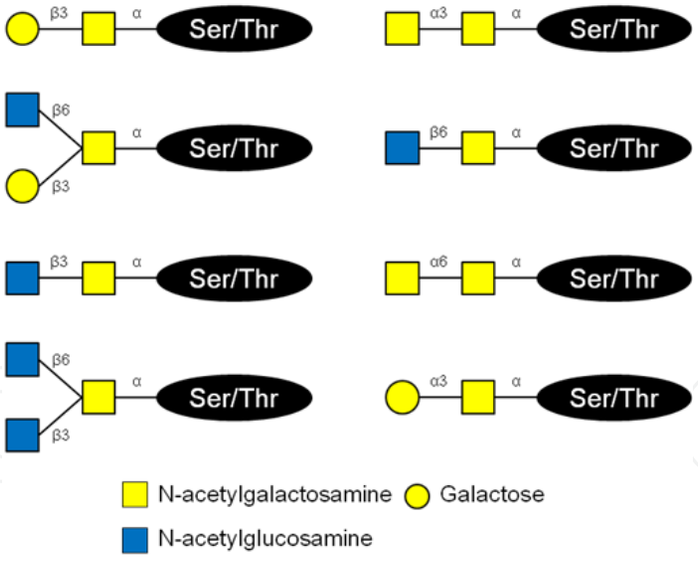


Figure 3. Representation of the core structures of the common O-glycans produced by mammalian cells.

3. Expression systems

For the correct production and folding of glycoproteins, the expression host needs to post-translationally modify the protein chain by adding sugars at the glycosylation sites. Therefore, production of glycoproteins is most often performed using eukaryotic cells, although the aglycosylated protein can be expressed as inclusion bodies in *E. coli* and refolded and also

produced using cell-free systems. Recent advances in glycosylation pathway engineering have resulted in both *E. coli* [7] and cell-free systems [8] which are capable of introducing N-glycosylation onto a protein. These methods are discussed in the relevant sections below.

In addition to finding an appropriate system for the over-expression of the target glycoprotein, the ability of this system to incorporate selenomethioine into the protein must be assessed. Selenomethioine labelling enables phasing of X-ray diffraction data by multi-wavelength anomalous diffraction (MAD) [9]. Proteins expressed using *E. coli* routinely incorporate 100 % selenomethionine, whereas incorporation is more variable for proteins produced in higher eukaryotic systems.

3.1. *Escherichia coli*

E. coli is an attractive host as it is fast, simple to use, robust and cost-effective and therefore remains the dominant host for the production of recombinant proteins. In terms of glycoprotein production, the protein is usually expressed in inclusion bodies as bacteria do not have either the endoplasmic reticulum or the Golgi apparatus needed to secrete and post-translationally modify glycoproteins. As the protein is produced in inclusion bodies, refolding is needed to produce samples for structural biology – a process which can be time consuming and inefficient. However this method has proved useful for a number of products, for instance Watson *et al.* used *E. coli* to produce crystals of the extracellular domain of human UL16-binding protein ULBP1 [10]. Here the authors investigated various refolding strategies before finding the optimum protocol of slow dilution of guanidine-solubilised protein in solution containing arginine, ethylenediaminetetraacetic acid, reduced and oxidized glutathione and phenylmethyl-sulphonyl fluoride over 48 hours. In addition, recently the crystal structures of the murine class I major histocompatibility complex H-2Kb (PDB entry 3ROL) [11] and human β-secretase I with bound inhibitors (PDB entry 3S7M and 3S7L) [12] and the NMR structure of the sterile alpha motifs of EphA2 and SHIP2 (PDB entry 2KSO) [13] have solved using production in *E. coli* followed by refolding.

Another approach is to express the glycoprotein in the periplasm of *E. coli* using a bacterial signal sequence such as OmpA. For example, human α1-microglubulin was produced in the periplasm of *E. coli* and used to obtain the crystal structure to 2.3 Å resolution which revealed a potential heme binding site (PDB entry 3QKG) [14].

Recently an engineered eukaryotic protein glycosylation pathway has been inserted into *E. coli* resulting in cells capable of producing glycoproteins with $Man_3GlcNAc_2$ sugars attached [7]. Four glycosyltransfereases from *Saccharomyces cerevisae*, the uridine diphosphate-N-acetylglucosamine transferases Alg13 and Alg14 and the mannosyltransferases Alg1 and Alg2, are used to generate the glycan. The glycan is then transferred onto an N-glycosylation site using the oligosaccharyltransferase PglB from *Campylobacter jejuni*. Valderrama-Rincon *et al.* tested production of three eukaryotic glycoproteins; the Fc domain of human IgG1, bovine RNaseA and the placental variant of human growth hormone, and detected expression of glycosylated proteins [7]. Although currently only ~1 % of the expressed proteins were found to be glycosylated, this technology represents a huge potential for the cost-effective production of glycoproteins with a defined glycosylation pattern.

3.2. Insect cells

Glycoproteins produced by insect cells, such as *Spodoptera frugiperda, Trichoplusia ni* and *Drosophila melanogaster*, express products with glycans which are oligomannose and paucimannose in nature and mainly of the form α1-6 fucosylated $Man_3GlcNAc_2$ (Figure 4A) [15]. This compact, relatively homogeneous glycoform is compatible with protein crystallization and a number of structures have been determined of glycoproteins produced using insect cells. For example, the crystal structure of FcγRIIa was solved to 1.5 Å resolution using baculovirus infected *S. frugiperda* (Sf) 21 cells (PDB entry 3RY4) and also its structure bound to human IgG1-Fc was resolved to 3.8 Å (PDB entry 3RY6) [16].

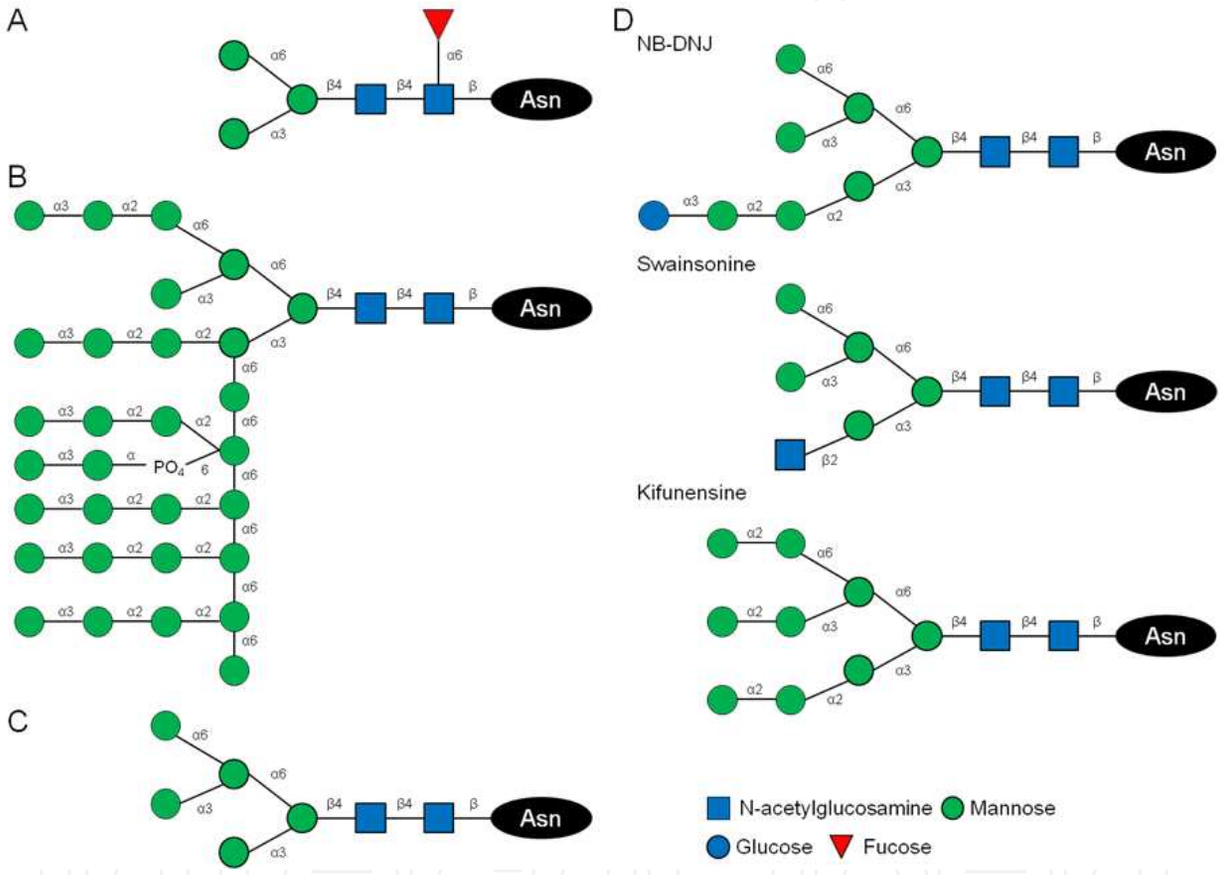


Figure 4. Representation of the N-glycans attached to proteins produced by (A) insect cells; (B) the yeast *Saccharomyces cerevisiae* (selected yeast oligosaccharide); (C) mutant CHO and HEK cell lines; and (D) mammalian cell lines in the presence of small molecule inhibitors.

The glycoforms produced by insect cells can be trimmed by treatment with endoglycosidases, for example endoglycosidase (endo) H or endo F1 and endo D will remove oligomannose and paucimannose sugars respectively leaving one GlcNAc residue on N-glcyosylation sites. In addition, endo F2 can be used in combination with endo F3 to cleave oligomannose and biantennary complex sugars and core fucosylated bi and triantennary complex gycans. Using this de-glycosylation strategy Fan *et al.* reported the successful crystallization and structure determination of a complex of human follicle stimulating hormone with its receptor (PDB entry 1XWD). Here crystals of fully

glycosylated complex diffracted to 9 Å whereas 2.9 Å resolution was obtained after de-glycosylation with endo F2 and endo F3 [17]. Also, the structure of HIV-1 envelope glycoprotein gp120 was solved to 2.2 Å using de-glycosylation with endo H and endo D (PDB entry 1G9N) [18]. Although less common, tunicamycin has been used during production to block all glycosylation in order to promote crystallization. For example, evasin-1 structure was solved to 1.63 Å in its non-glycosylated form and to 2.7 Å in its glycosylated form (PDB entries 3FPR and 3FPT respectively) [19].

Selenomethionine labelling in insect cell systems can be difficult due to the toxicity of selenomethionine and the long incubation times needed as late baculovirus promoters such as polyhedron or P10 are normally used [1]. The levels of incorporation for secreted glycoproteins is higher than for intracellular proteins as unlabelled protein is removed during the media exchange process [20]. For example, 85 % selenomethionine incorporation was achieved for envelope glycoprotein D from HSV1 [21] and 76 % for palmitoyl protein thioesterase 1 (PDB entry 1EI9 and 1EH5) [22]. In 2007, Cronin *et al.* addressed this problem of variability in levels of incorporation and developed a method which consistently gave 70-75 % selenomethionine incorporation with a yield of around 20 % of the unlabelled protein [23].

3.3. Yeast cells

Yeast have been used for the production of human glycoproteins, with the most popular expression hosts being *Saccharomyces cerevisiae* and *Pichia pastoris* [24-25]. Recombinant proteins are secreted and post-translationally glycosylated to give glycoforms that are sensitive to endo H or endo F1 treatment and are therefore suitable for crystallization (Figure 4B). For example, de-glycosylation of protein produced in *P. pastoris* was used to prepare the G-protein coupled receptor (GPCR) human β_2 adrenergic receptor for crystallization [26] and production of gastric intrinsic factor with cobalamin (vitamin B12) bound (IF-Cbl) in *P. pastoris* was followed by complex formation with the cubilin IF-Cbl-binding-region and endo H treatment prior to crystallization of the complex to give a 3.3 Å resolution crystal structure (PDB entry 3KQ4) [27].

There are examples in the literature where retaining the N-glycans on the glycosylated protein produced in yeast is important, such as the of human and mouse glutaminyl cyclases (QC), enzymes linked with Alzheimer's disease, (PDB entry 3SI0) [28]. The structure of human QC had already been solved using protein produced in *E. coli* (PDB entry 2AFO) [29]. However, the glycosylated structure was shown to be more stable and to contain a disulphide bond not present in the non-glycosylated structure. The reduced stability of the aglycosylated human QC was associated with ready loss of the catalytic zinc ion [28].

Production of selenomethionine labelled proteins is possible in yeast with incorporation levels of around 50 % routinely reported for both *P. pastoris* and *S. cerevisiae* [20]. However higher selenomethionine incorporation has been reported. For instance, incorporation of ~98 % selenomethionine was reported for production of the carboxy-terminal fragment of the RNA-dependent RNA polymerase from *Neurospora crassa*, QDE-1 using *S. cerevisiae* [30]. This labelled protein was used to generate crystals which diffracted to 3.2 Å and allowed

phasing of the native data giving a 2.3 Å resolution crystal structure (PDB entry 2J7N). More recently Malkowski *et al.* described generation of a selenomethionine-resistant strain of *S. cerevisiae* by blocking the S-adenosylmethionine synthesis pathway [31]. This strain was used to produce tryptophanyl-tRNA synthetase with >95 % selenomethionine incorporation which led to the determination of the structure of this yeast enzyme. [31]. Using the strain to produce and label heterologous proteins has so far not been reported.

3.4. Mammalian cells

Mammalian cells are a popular choice for the expression of glycoproteins, particularly for the production of human proteins. These cells have the correct machinery to fold and post-translationally modify glycoproteins. The two main cell lines used are human embryonic kidney (HEK) 293 cells and Chinese hamster ovary (CHO) cells, both of which are readily transfected with polythyleneimine (PEI), calcium phosphate or commercial lipids giving expression in 60-80 % of the cells [32-33]. There are two main variants of HEK 293 cells, (i) 293T which expresses the SV40 large-T antigen and (ii) 293E which expresses the Epstein-Barr virus (EBV) nuclear antigen 1 (EBNA1). Plasmids containing the SV40 or EBV origins of replication are amplified within these variants, giving more copies of the plasmid per cell and therefore a higher levels of protein expression [34]. An alternative method of introducing genes into mammalian cells is to use viral-mediated transduction such as the BacMam system [35], which has been shown to give milligram quantities of protein for structural studies [36].

Mammalian cells can be grown in either attached or suspension culture and automation of the culturing processes is possible in both cases [37-38]. Structural studies have been performed on proteins produced both by transient transfection, formation of a stable cell line, and using stable pools. Transient transfection is attractive due to the short timeframe between transfection and purified protein. Glycoprotein yields for non-antibodies of up to 40 mg/L have been obtained from attached HEK cells [39] and 36 mg/L from HEK cells in suspension [40], with yields of 27 mg/L from CHO cells in suspension [40].

Formation of a stable cell line producing a glycoprotein of interest often gives higher yields than utilizing transient transfection, however establishment of a stable cell line can take 2-6 months [41]. Automation can give a three-fold increase in throughput, although the timeline is the same [42]. An advantage of the system is that once the stable cell line has been established, production of the glycoprotein is fast and robust.

Use of stable transfection pools for expression are becoming increasingly popular as the protein yield is usually higher than for transient transfection but the timeline is shorter than for making a stable cell line. Post-transfection, the cells are sorted, often using a fluorescence marker, in order to enrich the pool for high producers. Using such a method, highly productive pools of cells can be obtained in 3 weeks, though yields may decline over time in culture as the transfection cells are not clonal. Using this method, expression levels of monoclonal antibodies from 100 mg/L to 1 g/L can be achieved in 2 months post transfection using CHO cells [43].

Selenomethionine labelling in mammalian cells is not as efficient as in *E. coli*, however levels of up to 90 % have been reported for stable HEK cell lines [44] and 78 % for transient glycoprotein production [1]. Generally cells are grown for a period of about 12 hours in media lacking methionine in order to deplete the intracellular methionine pools before the addition of selenomethionine. Keeping the concentration of selenomethionine used to 60 mg/L or below is important due to its toxicity which will lead to a lower yield of glycoprotein.

Unlike insect and fungi, mammalian cells produce glycoproteins with complex oligosaccharide chains (Figure 2) which are very heterogeneous and therefore not readily amenable to crystallization. Complete removal of N-linked glycans can be achieved by treatment of a purified glycoprotein with Peptide N-glycosidase (PNGase) F thus aiding crystallization [45]. In practice, two problems are encountered which limit this approach. Firstly, incomplete removal of glycans leading to partial de-glycosylation of the product; and secondly, insolubility due to protein aggregation following removal of all the sugars. Alternatively, glycosylation can be completely blocked in cells by treatment with tunicamycin. However, if glycosylation is required for proper folding and/or solubility in situ, then this approach will compromise the synthesis of the product. Two methods have been developed to get round these problems both of which depend upon manipulating N-glycosylation during glycoprotein biosynthesis by blocking the action of processing enzymes using either chemical inhibitors or null mutant cell lines.

3.4.1. Inhibitors

Three inhibitors have been used in the production of glycoproteins to manipulate the glycosylation pathway: N-butyldeoxynojirimycin (NB-DNJ), swainsonine and kifunensine. NB-DNJ inhibits α-glucosidase, thus blocking the early stages of N-glycan processing and giving products which contain high mannose or hybrid type sugars (Figure 4A). NB-DNJ has mainly being used in combination with the mutant CHO cell line, CHO Lec3.2.8.1 (Section 3.4.2), for instance in the crystallization of human costimulatory molecule B7-1 which is important in human immune response which gave crystals diffracting to 2.7 Å after treatment with endo H [46]. Swainsonine blocks α-mannosidase II resulting in high mannose or hybid type sugars shown in Figure 4D. Kifunensine strongly inhibits α-mannosidase I activity resulting in sugars of the form $Man_9GlcNAc_2$ (Figure 4D) [4]. Treatment of cells with the any of the above inhibitors results in relatively simple and chemically uniform glycoforms. Further, the high mannose and hybrid glycans resulting from the use of these drugs are cleavable using endoglycosidase (endo) H or endo F1 to leave one GlcNAc residue attached to the N-glycosylation site.

In practice, kifunensine is the most commonly used inhibitor in the production of glycoproteins for structural studies as the resulting glycans are the most homogeneous. In fact, the structures of glycoproteins can be solved following kifunensine treatment, without the use of endoglycosidase (for example, PDB entry 2WAH [47]), however more commonly the sugars are trimmed with endo H before crystallization studies. For example, Bishop *et al.* used transient transfection of HEK 293T cells in the presence of kifunensine to produce a number of glycoproteins involved in human Hedgehog signalling pathway. These were

treated with endo H before crystallization and co-crystallization to give structures of hedgehog interaction protein ectodomain (Hhip) and Desert hedgehog in isolation and Hhip in complex with Desert hedgehog and sonic hedgehog (PDB entries 2WFT, 2WFR, 2WFQ, 2WG3, 2WFX and 2WG4) [48]. Production in the presence of kifunensine followed by endo H treatment has also been used with a stable CHO cell line in the crystallization of the extracellular region of cytotoxic T-lymphocyte antigen 4 (CTLA-4) giving crystals which diffracted to 1.8 Å [49].

3.4.2. Mutant cell lines

CHO and HEK cell lines have been generated with mutations in their glycosylation pathways which disrupt the action of GlcNAc transferase I (GnTI) [50-51]. CHO Lec3.2.1 contains four mutations which are in the *Gne*, *Slc35a1*, *Slc35a2* and *Mgat1* open reading frames leading to lower activity of various glycosylation enzymes including GnTI [50]. Use of CHO Lec3.2.8.1 leads to proteins mainly containing high mannose glycans which are sensitive to removal of all but the last GlcNAc residue with endo H or endo F1 (Figure 4C). However, the mutations in CHO Lec3.2.8.1 do not completely inhibit the formation of complex glycans [50] with, in some cases, as little as 10 % of the glycoprotein produced being endoglycosidase sensitive [4]. Structural studies have been carried out using CHO Lec3.2.8.1 both in the absence and presence of NB-DNJ which has a complementary inhibition effect to the mutant cell line giving endoglycosidase sensitive sugars (For examples see [46, 52-54]).

The GnTI-deficient cell line, HEK 293 GnTI$^-$ (also known as HEK 293S) produces glycoproteins with high mannose glycans (Figure 4C) which are endo H and endo F1 sensitive. Use of HEK 293 GnTI$^-$ cells gives product containing a very uniform glycosylation pattern of the form $Man_5GlcNAc_2$ with only traces of other glycan patterns being detected [51, 55]. This cell line has proved popular, recently facilitating structure solution of NetrinG1 and NetrinG2 in complex with their respective ligands at 3.25 Å and 2.6 Å resolution respectively (PDB entries 3ZYJ and 3ZYJ) [56]; the human glutamate receptor GluR2 amino terminal domain at 1.8 Å resolution (PDB entries 2WJW and 2WJX) [57]; and the orphan domain of the membrane glycoprotein endoglin using small angle X-ray scattering (SAXS) [58].

3.5. Other

Although not currently in mainstream use for the production of glycoproteins, two other expression systems are worth mentioning, namely cell-based production in the protozoan *Leishmania tarentolae* and cell-free synthesis using coupled transcription-translation systems.

The structure of human Cu/Zn superoxide dismutase has been recently published and represents the first structure solved using *L. tarentolae* as the expression host (PDB entry 3KH3) [59]. The advantages of the *L. tarentolae* system (commercialized by Jena Biosciences as LEXSY) are that it is simple to use, gives high yields and is inexpensive compared with the higher eukaryotes. *L. tarentolae* has eukaryotic folding and post-translational machinery and has been used in a proof of principle experiment to express glycosylated human

erythropoietin [60]. These data show that use of *L. tarentolae* as a host system is applicable to the expression of glycoproteins for structural studies.

Cell-free synthesis systems have been used for structural studies for a number of years and are commonly based on lysates from *E. coli*, wheat germ and rabbit reticulocytes [61]. Glycoproteins can be produced by supplementing lysates with microsomal fractions [62], or using extracts of eukaryotes such as insect cells [63], hybridomas [64] and mammalian cells [65], although the yields are often poor. These systems give glycosylation patterns native to the host which in the case of insect cells can be modified using endoglycosidases and in the case of mammalian cells can be blocked using inhibitors (Section 3.4.1). Recently both the *E. coli* lysate cell free synthesis system and the PURE system for *in vitro* translation using purified components of *E. coli* [66] have been adapted for production of glycoproteins [8]. In this article, Guarino and DeLisa used the protein glycosylation locus from *Campylobater jejuni* to supplement both systems and produce glycoprotein with the GlcGalNAc$_5$Bac (where Bac represents bacillosamine) glycosylation pattern of *C. jejuni* (Figure 5) [8]. Cell-free systems have yet to yield a glycosylated protein for which a structure has been deposited in the PDB. However the ability to micro-engineer the components of the glycosylation pathway in an "open system" is an interesting and useful addition to the structural biologist's toolbox.

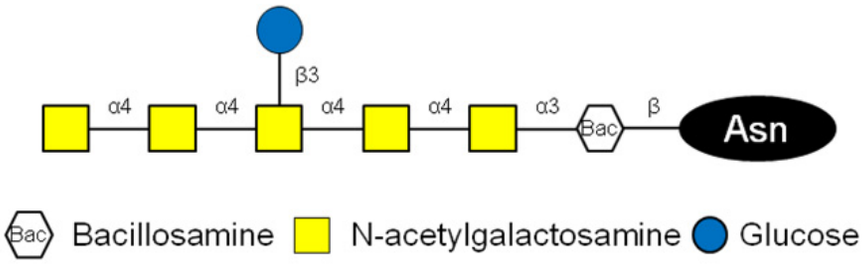


Figure 5. Representation of the glycosylation pattern added to glycoproteins using the cell-free synthesis system of Guarino and DeLisa [8].

4. Variable occupancy of glycosylation sites

In addition to the types of oligosaccharide chain attached to a glycosylation site, i.e. the glycoform, heterogeneity occurs in the occupancy of a glycosylation site. The occupancy of an N-glycosylation site depends upon the recognition sequon along with the structure of the protein in proximity to the glycosylation site. Some N-glycosylation sites have very low occupancy as a result of the local sequence composition, secondary structure and also distance to the C-terminus [67-69]. Bioinformatics programmes such as NetNGlyc and NetOGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/ and http://www.cbs.dtu.dk/services/NetOGlyc/) [70-71] can predict occupancy of N-glycosylation and O-glycosylation sites respectively; however these are sequence based predictions and so need to be verified experimentally. After experimental identification of glycosylation site occupancy, variably occupied sites can be removed. Removal of variably occupied glycosylation sites should not affect the activity of a glycoprotein as these glycans are only present in a proportion of the sample and therefore must not be important for folding and solubility.

4.1. Identification methods

Analysis of glycosylation site occupancy is usually carried out using mass spectrometry. For a review of recent developments in glycoprotein analysis and glycomics see Zaia [72]. In the most straightforward experiment, the occupancy of an N-glycosylation site can be determined using a combination of tryptic digest and PNGase F treatment, followed by analysis of the resulting peptide [73]. The peptide contains asparagine at the proposed site if no glycan is present and, due to the action of PNGase F, an aspartic acid if sugars are attached. As glycopeptides are not readily ionized within a mass spectrometer, they can be enriched in the sample using hydrophilic interaction liquid chromatography (HILIC), thus increasing the likelihood of a glycosylation site-containing peptide being detected [74-75]. Mapping of glycoproteins, including both analysis of the glycosylation site and analysis of the glycans themselves, has been miniaturized and automated, such as methods using functionalized magnetic nanoparticles [76] and integrating chromatography steps [77] which allow for rapid acquisition of results.

Determination of O-glycosylation site occupancy is more difficult than for N-glycoslation sites as there is no endoglycosidase which universally releases O-linked glycans in the same way that PNGase F does for N-linked glycans. However, Halfinger *et al.* have developed a robust protocol using partial digestion with exoglycosidase and β-elimination using a mild alkylamine base followed by Michael addition and analysis using collision induced dissociation (CID) MS/MS [78].

Variable occupancy of glycosylation sites has also been detected by mutation studies. Here all the potential sites are mutated and the resulting protein analysed for activity. For example, Garman *et al.* analysed human IgE-FcεRIα and identified three glycosylation sites, N74, N135 and T142, which were not essential for protein folding, secretion and activity [79]. These sites were later confirmed by mass spectrometric analysis to be variably occupied [74]. However, mapping glycosylation occupancy by systematic mutation studies is labour intensive and is likely to result in many mutants that do not express or fold correctly as essential glycosylation sites have been removed.

4.2. Removal of glycosylation sites

Glycosylation sites found to be non-essential are often removed using site directed mutagenesis of the asparagine codon to a glutamine codon as this amino acid is similar in size and charge, but is not an attachment site for a glycan chain.

Formation of an N174Q mutant version of human ephrinA2 was essential in obtaining crystals of the EphA4:ephrinA2 complex which resulted in the structure being solved to 2.3 Å resolution (PDB entry 2WO3) [80]. In contrast, previous attempts at crystallization using wild type ephrinA2 with EphA4 were not successful. Although the N174 was shown to be invariably glycosylated, this site is not conserved across the ephrin family, (Nettleship, Bowden, unpublished results), and therefore was presumed to be non-essential.

In the case of human alpha-N-acetylgalactosaminidase, which has five N-glycosylation sites, mutation of N201 to glutamine led to crystals giving diffraction to 1.9 Å resolution (PDB

entry 3H53) as opposed to the 8 Å resolution data collected using wild type glycoprotein crystals [81].

5. Structural studies

5.1. Crystallization and X-ray crystallography

Glycoproteins crystallize in a variety of conditions and trials are usually set up using the same screens as non-glycosylated proteins; such as those found in kits sold by Hampton Research, Molecular Dimensions and Emerald Biosystems. A number of reviews have addressed the problems connected with crystallization of glycoproteins such as the increase in surface entropy associated with large post-translational modifications and the microheterogenity of glycans [82-83]. However, it is to be noted that the presence of glycans can be an advantage for crystallization as they can form essential intermolecular contacts in crystal lattices [82]. Indeed glycoprotein crystallization has a success rate of around 50 % which is comparable to that for non-glycosylated proteins [82]. Methods around glycoprotein crystallization have developed to include automation and miniaturization using microscale crystallization techniques with as little as 65 μg of protein sample [45].

The strategy which is unique to the crystallization of glycoproteins is manipulation of the glycoform as described above. Such manipulations affect the propensity of a glycosylated protein to crystallize and also the quality of the resulting crystals. As shown in Figure 6, the type of oligosaccharide attached to the protein affects crystal morphology. Interestingly, in the case of human IgE-FcεRIα, the $Man_5GlcNAc_2$ glycoform gave better diffracting crystals than the shorter $Man_1GlcNAc_2$ (Figure 6E and 6D respectively) and therefore setting up crystallization trials with more than one glycoform can be advantageous.

X-ray crystal structures of glycoproteins often only show the initial GlcNAc residue even if more sugar residues are known to be attached to the protein because the glycan chains are flexible and so electron density for further sugar residues is not present. In rare cases, more of the oligosaccharide chain is resolved due to it either forming crystal contacts or interacting with the polypeptide chain. For example Figure 7 shows that the oligosaccharide chain attached to N297 of human IgG1-Fc could be fully resolved (PDB entry 2WAH) [47].

In solving a glycoprotein structure, it is important that the carbohydrate chains are built into the relevant electron density taking into account the restrictions on sugar conformation, including torsion angles, and linkages found in nature [87-88]. Recently, Crispin *et al.* have described the biosynthesis of glycans with emphasis on how these pathways lead to the glycosidic linkages found in glycoprotein structures [88]. If the production host gave sugars of a defined composition, for example HEK 293 GnTI$^-$ gives sugars of the form $Man_5GlcNAc_2$, modelling the glycans is relatively straightforward. For other cases, there are a number of databases where the structures of known glycans are deposited, with GlycomeDB (www.glycome-db.org) being the most comprehensive and

unified carbohydrate database containing information from all public carbohydrate databases and the PDB [89-90]. Such databases are a useful resource for checking on the possible oligosaccharides and linkages likely to be attached to a glycoprotein produced using a given host cell. For building a three-dimensional model of an oligosaccharide chain, bioinformatics resources such as Sweet-II (www.glycosciences.de/spec/sweet2) [91] and Glycam Biomolecule Builder (glycam.ccrc.uga.edu/CCRC/biombuilder/biomb_index.jsp) [92] are available. After modelling the glycan, Glycam Biomolecule Builder attaches the sugar chain to the protein structure at the glycosylation site, with glycans generated by Sweet-II being attached using the glyProt software (www.glycosciences.de/modeling/glyprot/php/main.php) [93]. Upon deposition of the glycoprotein structure into the PDB, pdb-care (PDB carbohydrate residue check) can help with problems found within glycan structures [94-95]. Overall, use of these web-based tools allows for the correct building of sugar chains into glycoprotein structures when electron density is available.

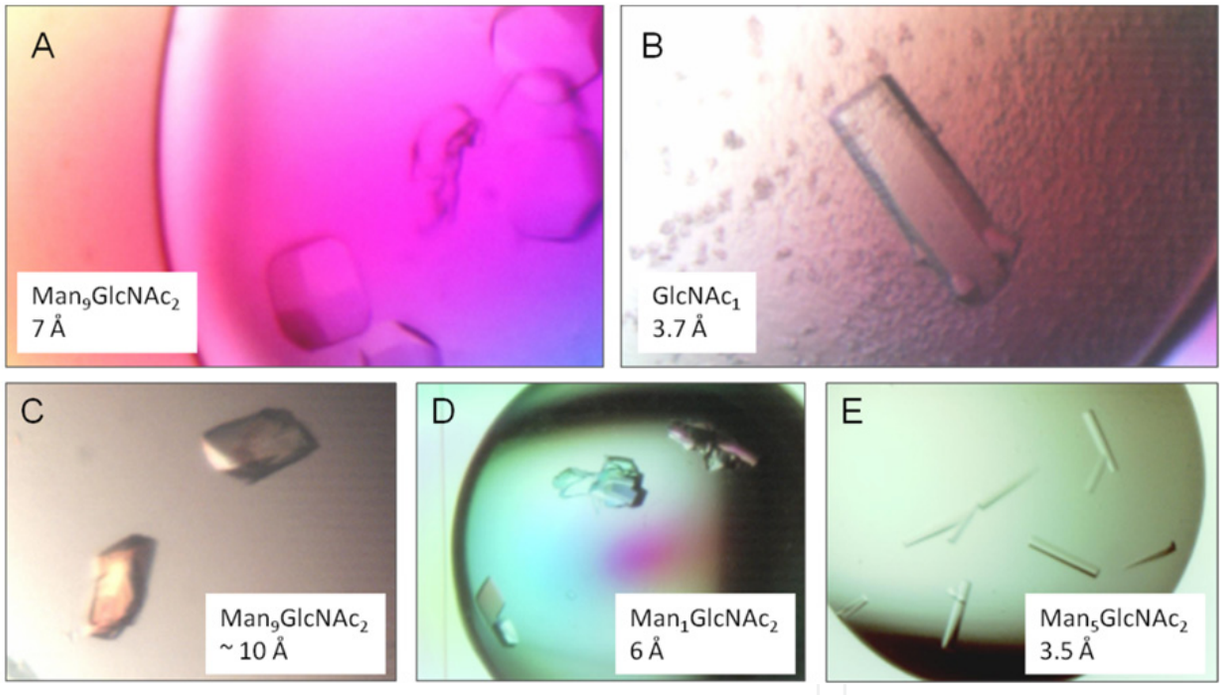


Figure 6. Crystals of human programmed cell death 1-ligand 1 formed from glycoprotein secreted by HEK 293T cells (A) in the presence of kifunensine and (B) after treatment with endo F1. Diffraction was to 7 Å and 3.7 Å respectively. Crystals of human IgE-FcεRIα (N74A, N135A, T142A) from protein produced using HEK 293T cells (C) in the presence of kifunensine; (D) after treatment with jack bean mannosidase and (E) produced by HEK 293 GnTI$^-$ cells. Diffraction was around 10 Å, to 6 Å and to 3.5 Å respectively. In each case the insert shows the major glycosylation state of the protein upon crystallization and the resolution of diffraction data collected. (Human programmed cell death 1-ligand 1 - PDB entry 3BIS [84] and human IgE-FcεRIα - PDB entry 2Y7Q [85])

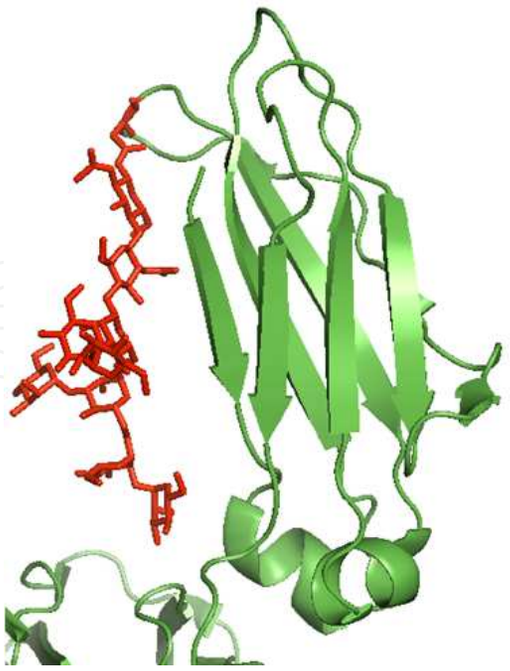


Figure 7. Cartoon representation of the human IgG1-Fc glycan attached to N297 shown as sticks in red with the protein backbone shown in green. The glycoprotein was produced in HEK 293T cells in the presence of kifunensine, giving sugars of the form $Man_9GlcNAc_2$. (The diagram was produced using PyMOL [86] and PDB entry 2WAH [47]).

5.2. NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy can be used to study protein structure, but the technique faces some obstacles in terms of its use for glycoproteins, namely the ^{1}H chemical shift overlap between carbohydrate and protein signal [96] and the difficulty of obtaining sufficient yield of labelled protein from eukaryotic systems. These challenges have led to the majority of glycoproteins structures solved by NMR being of the aglycosylated protein chains expressed in *E. coli*.

Labelling with ^{15}N or both ^{15}N and ^{13}C has been demonstrated in mammalian cells [97], insect cells [98] and yeast [99]. Such eukaryotic expression systems have been used to solve a few recombinant glycoprotein structures containing glycans such as fragments of human fibronectin (PBD entry 1E8B), human thrombomadulin (PBD entry 1DQB) and the extracellular domain of human cytotoxic T lymphocylte-associated protein (CTLA)-4 (PDB entry 1AH1) [100-102].

The problem of signal overlap between proteins and carbohydrates has been tackled by Slynko *et al.* who added unlabelled glycan chains using in vitro N-glycosylation technology [103] to AcrA from *C. jejuni* produced using *E. coli* with ^{15}N and ^{13}C labelling (PDB entry 2K32) [96]. This allowed a NOESY (Nuclear Overhauser effect spectroscopy) experiment to be performed without confusion between protein and glycan resonances which often cause difficulty in structure determination of unlabelled or ^{15}N-labelled glycoproteins [96].

In NMR structures, the protein chains and attached glycans are in solution and so are able to move around. Unlike in crystal structures, solution NMR allows the flexibility of both the oligosaccharide chain and the protein chain around a glycosylation site to be observed. In the case of AcrA (PDB entry 2K32), the flexibility was seen to be in the α-helical loop region containing the N42 glycosylation site (Figure 8A), with the heptasaccharide glycan having a well defined rod-like structure [96]. Human CTLA4 (PDB entry 1AH1) contains two N-glycosylation sites, N78 and N111 which are occupied with partially deglycosylated glycans of the form $Man_1GlcNAc_2$ (Figure 8B). The glycan attached to N78 interacts extensively with the side chains of a nearby β-sheet and is therefore ordered, whereas N111 had limited interaction with the protein chain and so only the initial GlcNAc residue is well defined [102].

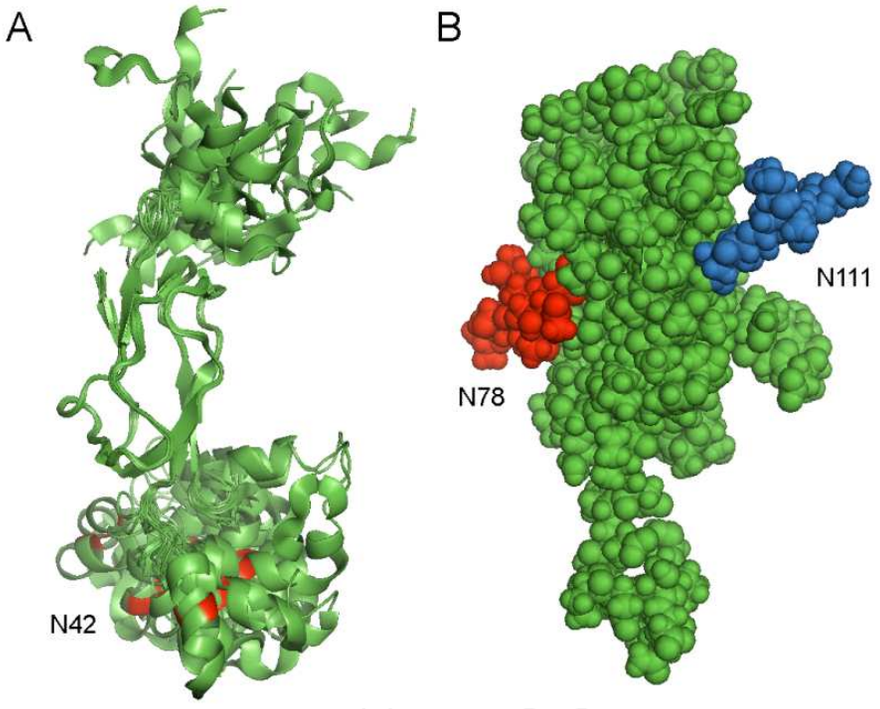


Figure 8. (A) Cartoon representation of the NMR ensemble structure of AcrA showing the N42 glycosylation site (in red) in the flexible lower domain of the protein. (B) A space filling model of CTLA4 with the glycans attached to N78 (in red) showing interaction with the nearby β-sheet and the N111 sugar chain (in blue) being remote from the polypeptide chain. (Diagrams were produced using PyMOL [86] and PDB entries 2K32 [96] and 1AH1 [102] respectively).

6. Conclusions and future prospects

This chapter has reviewed some of the recent developments in the field of glycoprotein structural biology. Increasingly, the expression system of choice for the production of recombinant glycoproteins is the mammalian cell, and in particular HEK in the presence of kifunensine which modifies glycan processing. This enables treatment of the purified glycoprotein with endoglycosidase to reduce the sugar "load" to a single GlcNAc at each N-linked attachment site. Experience shows that by preparing glycoproteins with different

glycoforms the chances of obtaining diffraction quality crystals are significantly increased. Combining this approach with the inclusion of novel additives in the crystallization experiment, for example "smart materials", such as the molecularly imprinted polymers (MIPs) recently reported by Sarkidakis *et al*. [104], adds another dimension to crystallization.

Future developments in simpler low cost expression technology as exemplified by *Leishmania tarentolae* system will also have an impact on the structural biology of glycoproteins. It is anticipated that the number of structures solved by X-ray crystallography and NMR for this important class of proteins will rapidly increase in the next few years.

Author details

Joanne E. Nettleship
OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratories, Harwell Oxford, Didcot, UK

Division of Structural Biology, Oxford University, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, UK

Acknowledgement

The author would like to thank Ray Owens and Max Crispin for critical reading of the manuscript. The OPPF-UK is funded by the UK Medical Research Council and Biotechnology and Biological Sciences Research Council.

7. References

[1] Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ. Recent advances in the production of proteins in insect and mammalian cells for structural biology. J Struct Biol. 2010 Oct;172(1):55-65.
[2] Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta. 1999 Dec 6;1473(1):4-8.
[3] Zafar S, Nasir A, Bokhari H. Computational analysis reveals abundance of potential glycoproteins in Archaea, Bacteria and Eukarya. Bioinformation. 2011;6(9):352-5.
[4] Davis SJ, Crispin M. Solutions to the Glycosylation Problem for Low- and High-Throughput Structural Glycoproteomics. In: Owens RJ, Nettleship JE, editors. Functional and Structural Proteomics of Glycoproteins2011. p. 127-58.
[5] Varki A, Sharon N. Historical Background and Overview. 2009.
[6] Parodi AJ. Protein glucosylation and its role in protein folding. Annu Rev Biochem. 2000;69:69-93.
[7] Valderrama-Rincon JD, Fisher AC, Merritt JH, Fan YY, Reading CA, Chhiba K, et al. An engineered eukaryotic protein glycosylation pathway in Escherichia coli. Nat Chem Biol. 2012;8(5):434-6.

[8] Guarino C, Delisa MP. A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. Glycobiology. 2011 Nov 8;8:8.
[9] Hendrickson WA, Horton JR, LeMaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. Embo J. 1990 May;9(5):1665-72.
[10] Watson AA, Christou CM, O'Callaghan CA. Expression, purification and crystallization of the human UL16-binding protein ULBP1. Protein Expr Purif. 2011 Sep;79(1):44-8.
[11] Madhurantakam C, Duru AD, Sandalova T, Webb JR, Achour A. Inflammation-Associated Nitrotyrosination Affects TCR Recognition through Reduced Stability and Alteration of the Molecular Surface of the MHC Complex. PLoS One. 2012;7(3):e32805.
[12] Malamas MS, Erdei J, Gunawan I, Barnes K, Hui Y, Johnson M, et al. New pyrazolyl and thienyl aminohydantoins as potent BACE1 inhibitors: exploring the S2' region. Bioorg Med Chem Lett. 2011 Sep 15;21(18):5164-70.
[13] Lee HJ, Hota PK, Chugha P, Guo H, Miao H, Zhang L, et al. NMR structure of a heterodimeric SAM:SAM complex: characterization and manipulation of EphA2 binding reveal new cellular functions of SHIP2. Structure. 2012 Jan 11;20(1):41-55.
[14] Meining W, Skerra A. The crystal structure of human alpha1-microglobulin reveals a potential haem-binding site. Biochem J. 2012 Apr 19;19:19.
[15] Harrison RL, Jarvis DL. Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. Adv Virus Res. 2006;68:159-91.
[16] Ramsland PA, Farrugia W, Bradford TM, Sardjono CT, Esparon S, Trist HM, et al. Structural basis for Fc gammaRIIa recognition of human IgG and formation of inflammatory signaling complexes. J Immunol. 2011 Sep 15;187(6):3208-17.
[17] Fan QR, Hendrickson WA. Structure of human follicle-stimulating hormone in complex with its receptor. Nature. 2005 Jan 20;433(7023):269-77.
[18] Kwong PD, Wyatt R, Desjardins E, Robinson J, Culp JS, Hellmig BD, et al. Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1). J Biol Chem. 1999 Feb 12;274(7):4115-23.
[19] Dias JM, Losberger C, Deruaz M, Power CA, Proudfoot AE, Shaw JP. Structural basis of chemokine sequestration by a tick chemokine binding protein: the crystal structure of the complex between Evasin-1 and CCL3. PLoS One. 2009;4(12):e8514.
[20] Walden H. Selenium incorporation using recombinant techniques. Acta Crystallogr D Biol Crystallogr. 2010 Apr;66(Pt 4):352-7.
[21] Carfi A, Gong H, Lou H, Willis SH, Cohen GH, Eisenberg RJ, et al. Crystallization and preliminary diffraction studies of the ectodomain of the envelope glycoprotein D from herpes simplex virus 1 alone and in complex with the ectodomain of the human receptor HveA. Acta Crystallogr D Biol Crystallogr. 2002 May;58(Pt 5):836-8.
[22] Bellizzi JJ, 3rd, Widom J, Kemp C, Lu JY, Das AK, Hofmann SL, et al. The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. Proc Natl Acad Sci U S A. 2000 Apr 25;97(9):4573-8.
[23] Cronin CN, Lim KB, Rogers J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. Protein Sci. 2007 Sep;16(9):2023-9.

[24] Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, Porro D. Recombinant protein production in yeasts. Methods Mol Biol. 2012;824:329-58.
[25] Damasceno LM, Huang CJ, Batt CA. Protein secretion in Pichia pastoris and advances in protein production. Appl Microbiol Biotechnol. 2012 Jan;93(1):31-9.
[26] Noguchi S, Satow Y. Purification of human beta2-adrenergic receptor expressed in methylotrophic yeast Pichia pastoris. J Biochem. 2006 Dec;140(6):799-804.
[27] Andersen CB, Madsen M, Storm T, Moestrup SK, Andersen GR. Structural basis for receptor recognition of vitamin-B(12)-intrinsic factor complexes. Nature. 2010 Mar 18;464(7287):445-8.
[28] Ruiz-Carrillo D, Koch B, Parthier C, Wermann M, Dambe T, Buchholz M, et al. Structures of glycosylated mammalian glutaminyl cyclases reveal conformational variability near the active center. Biochemistry. 2011 Jul 19;50(28):6280-8.
[29] Huang KF, Liu YL, Cheng WJ, Ko TP, Wang AH. Crystal structures of human glutaminyl cyclase, an enzyme responsible for protein N-terminal pyroglutamate formation. Proc Natl Acad Sci U S A. 2005 Sep 13;102(37):13117-22.
[30] Laurila MR, Salgado PS, Makeyev EV, Nettelship J, Stuart DI, Grimes JM, et al. Gene silencing pathway RNA-dependent RNA polymerase of Neurospora crassa: yeast expression and crystallization of selenomethionated QDE-1 protein. J Struct Biol. 2005 Jan;149(1):111-5.
[31] Malkowski MG, Quartley E, Friedman AE, Babulski J, Kon Y, Wolfley J, et al. Blocking S-adenosylmethionine synthesis in yeast allows selenomethionine incorporation and multiwavelength anomalous dispersion phasing. Proc Natl Acad Sci U S A. 2007 Apr 17;104(16):6678-83.
[32] Huh SH, Do HJ, Lim HY, Kim DK, Choi SJ, Song H, et al. Optimization of 25 kDa linear polyethylenimine for efficient gene delivery. Biologicals. 2007 Jun;35(3):165-71.
[33] Martin-Montanez E, Lopez-Tellez JF, Acevedo MJ, Pavia J, Khan ZU. Efficiency of gene transfection reagents in NG108-15, SH-SY5Y and CHO-K1 cell lines. Methods Find Exp Clin Pharmacol. 2010 Jun;32(5):291-7.
[34] Van Craenenbroeck K, Vanhoenacker P, Haegeman G. Episomal vectors for gene expression in mammalian cells. Eur J Biochem. 2000 Sep;267(18):5665-78.
[35] Jardin BA, Elias CB, Prakash S. Expression of a secreted protein in mammalian cells using baculovirus particles. Methods Mol Biol. 2012;801:41-63.
[36] Dukkipati A, Park HH, Waghray D, Fischer S, Garcia KC. BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. Protein Expr Purif. 2008 Dec;62(2):160-70.
[37] Zhao Y, Bishop B, Clay JE, Lu W, Jones M, Daenke S, et al. Automation of large scale transient protein expression in mammalian cells. J Struct Biol. 2011 Aug;175(2):209-15.
[38] Gonzalez R, Jennings LL, Knuth M, Orth AP, Klock HE, Ou W, et al. Screening the mammalian extracellular proteome for regulators of embryonic human stem cell pluripotency. Proc Natl Acad Sci U S A. 2010 Feb 23;107(8):3552-7.
[39] Aricescu AR, Lu W, Jones EY. A time- and cost-efficient system for high-level protein production in mammalian cells. Acta Crystallogr D Biol Crystallogr. 2006 Oct;62(Pt 10):1243-50.
[40] Bollin F, Dechavanne V, Chevalet L. Design of Experiment in CHO and HEK transient transfection condition optimization. Protein Expr Purif. 2011 Jul;78(1):61-8.

[41] Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 2004 Nov;22(11):1393-8.
[42] Lindgren K, Salmen A, Lundgren M, Bylund L, Ebler A, Faldt E, et al. Automation of cell line development. Cytotechnology. 2009 Jan;59(1):1-10.
[43] Ye J, Alvin K, Latif H, Hsu A, Parikh V, Whitmer T, et al. Rapid protein production using CHO stable transfection pools. Biotechnol Prog. 2010 Sep-Oct;26(5):1431-7.
[44] Barton WA, Tzvetkova-Robev D, Erdjument-Bromage H, Tempst P, Nikolov DB. Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. Protein Sci. 2006 Aug;15(8):2008-13.
[45] Lee JE, Fusco ML, Saphire EO. An efficient platform for screening expression and crystallization of glycoproteins produced in human cells. Nat Protoc. 2009;4(4):592-604.
[46] Davis SJ, Ikemizu S, Collins AV, Fennelly JA, Harlos K, Jones EY, et al. Crystallization and functional analysis of a soluble deglycosylated form of the human costimulatory molecule B7-1. Acta Crystallogr D Biol Crystallogr. 2001 Apr;57(Pt 4):605-8.
[47] Crispin M, Bowden TA, Coles CH, Harlos K, Aricescu AR, Harvey DJ, et al. Carbohydrate and domain architecture of an immature antibody glycoform exhibiting enhanced effector functions. J Mol Biol. 2009 Apr 17;387(5):1061-6.
[48] Bishop B, Aricescu AR, Harlos K, O'Callaghan CA, Jones EY, Siebold C. Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. Nat Struct Mol Biol. 2009 Jul;16(7):698-703.
[49] Yu C, Crispin M, Sonnen AF, Harvey DJ, Chang VT, Evans EJ, et al. Use of the alpha-mannosidase I inhibitor kifunensine allows the crystallization of apo CTLA-4 homodimer produced in long-term cultures of Chinese hamster ovary cells. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2011 Jul 1;67(Pt 7):785-9.
[50] North SJ, Huang HH, Sundaram S, Jang-Lee J, Etienne AT, Trollope A, et al. Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. J Biol Chem. 2010 Feb 19;285(8):5759-75.
[51] Chang VT, Crispin M, Aricescu AR, Harvey DJ, Nettleship JE, Fennelly JA, et al. Glycoprotein structural genomics: solving the glycosylation problem. Structure. 2007 Mar;15(3):267-73.
[52] Wilke S, Groebe L, Maffenbeier V, Jager V, Gossen M, Josewski J, et al. Streamlining homogeneous glycoprotein production for biophysical and structural applications by targeted cell line development. PLoS One. 2011;6(12):e27829.
[53] Fischer PB, Karlsson GB, Butters TD, Dwek RA, Platt FM. N-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with changes in antibody recognition of the V1/V2 region of gp120. J Virol. 1996 Oct;70(10):7143-52.
[54] Butters TD, Sparks LM, Harlos K, Ikemizu S, Stuart DI, Jones EY, et al. Effects of N-butyldeoxynojirimycin and the Lec3.2.8.1 mutant phenotype on N-glycan processing in Chinese hamster ovary cells: application to glycoprotein crystallization. Protein Sci. 1999 Aug;8(8):1696-701.
[55] Crispin M, Harvey DJ, Chang VT, Yu C, Aricescu AR, Jones EY, et al. Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. Glycobiology. 2006 Aug;16(8):748-56.

[56] Seiradake E, Coles CH, Perestenko PV, Harlos K, McIlhinney RA, Aricescu AR, et al. Structural basis for cell surface patterning through NetrinG-NGL interactions. Embo J. 2011 Nov 2;30(21):4479-88.
[57] Clayton A, Siebold C, Gilbert RJ, Sutton GC, Harlos K, McIlhinney RA, et al. Crystal structure of the GluR2 amino-terminal domain provides insights into the architecture and assembly of ionotropic glutamate receptors. J Mol Biol. 2009 Oct 9;392(5):1125-32.
[58] Alt A, Miguel-Romero L, Donderis J, Aristorena M, Blanco FJ, Round A, et al. Structural and functional insights into endoglin ligand recognition and binding. PLoS One. 2012;7(2):e29948.
[59] Gazdag EM, Cirstea IC, Breitling R, Lukes J, Blankenfeldt W, Alexandrov K. Purification and crystallization of human Cu/Zn superoxide dismutase recombinantly produced in the protozoan Leishmania tarentolae. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2010 Aug 1;66(Pt 8):871-7.
[60] Breitling R, Klingner S, Callewaert N, Pietrucha R, Geyer A, Ehrlich G, et al. Non-pathogenic trypanosomatid protozoa as a platform for protein research and production. Protein Expr Purif. 2002 Jul;25(2):209-18.
[61] Katzen F, Chang G, Kudlicki W. The past, present and future of cell-free protein synthesis. Trends Biotechnol. 2005 Mar;23(3):150-6.
[62] Walter P, Blobel G. Preparation of microsomal membranes for cotranslational protein translocation. Methods Enzymol. 1983;96:84-93.
[63] Tarui H, Murata M, Tani I, Imanishi S, Nishikawa S, Hara T. Establishment and characterization of cell-free translation/glycosylation in insect cell (Spodoptera frugiperda 21) extract prepared with high pressure treatment. Appl Microbiol Biotechnol. 2001 May;55(4):446-53.
[64] Mikami S, Kobayashi T, Yokoyama S, Imataka H. A hybridoma-based in vitro translation system that efficiently synthesizes glycoproteins. J Biotechnol. 2006 Dec 15;127(1):65-78.
[65] Shibutani M, Kim E, Lazarovici P, Oshima M, Guroff G. Preparation of a cell-free translation system from PC12 cell. Neurochem Res. 1996 Jul;21(7):801-7.
[66] Ohashi H, Kanamori T, Shimizu Y, Ueda T. A highly controllable reconstituted cell-free system--a breakthrough in protein synthesis research. Curr Pharm Biotechnol. 2010 Apr;11(3):267-71.
[67] Rao RS, Bernd W. Do N-glycoproteins have preference for specific sequons? Bioinformation. 2010;5(5):208-12.
[68] Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology. 2004 Feb;14(2):103-14.
[69] Bano-Polo M, Baldin F, Tamborero S, Marti-Renom MA, Mingarro I. N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. Protein Sci. 2011 Jan;20(1):179-86.
[70] Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002:310-22.
[71] Julenius K, Molgaard A, Gupta R, Brunak S. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology. 2005 Feb;15(2):153-64.

[72] Zaia J. Mass spectrometry and glycomics. Omics. 2010 Aug;14(4):401-18.
[73] Nettleship JE, Aplin R, Aricescu AR, Evans EJ, Davis SJ, Crispin M, et al. Analysis of variable N-glycosylation site occupancy in glycoproteins by liquid chromatography electrospray ionization mass spectrometry. Anal Biochem. 2007 Feb 1;361(1):149-51.
[74] Nettleship JE. Hydrophilic Interaction Liquid Chromatography in the Characterization of Glycoproteins. In: Wang PG, He W, editors. Hydrophilic Interaction Liquid Chromatography (HILIC) and Advanced Applications: CRC Press; 2011. p. 523-50.
[75] Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P. A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J Proteome Res. 2004 May-Jun;3(3):556-66.
[76] Kuo CW, Wu IL, Hsiao HH, Khoo KH. Rapid glycopeptide enrichment and N-glycosylation site mapping strategies based on amine-functionalized magnetic nanoparticles. Anal Bioanal Chem. 2012 Jan 29;29:29.
[77] Qu Y, Xia S, Yuan H, Wu Q, Li M, Zou L, et al. Integrated sample pretreatment system for N-linked glycosylation site profiling with combination of hydrophilic interaction chromatography and PNGase F immobilized enzymatic reactor via a strong cation exchange precolumn. Anal Chem. 2011 Oct 1;83(19):7457-63.
[78] Halfinger B, Sarg B, Lindner HH. Evaluation of non-reductive beta-elimination/Michael addition for glycosylation site determination in mucin-like O-glycopeptides. Electrophoresis. 2011 Dec;32(24):3546-53.
[79] Garman SC, Wurzburg BA, Tarchevskaya SS, Kinet JP, Jardetzky TS. Structure of the Fc fragment of human IgE bound to its high-affinity receptor Fc epsilonRI alpha. Nature. 2000 Jul 20;406(6793):259-66.
[80] Bowden TA, Aricescu AR, Nettleship JE, Siebold C, Rahman-Huq N, Owens RJ, et al. Structural plasticity of eph receptor A4 facilitates cross-class ephrin signaling. Structure. 2009 Oct 14;17(10):1386-97.
[81] Clark NE, Garman SC. The 1.9 a structure of human alpha-N-acetylgalactosaminidase: The molecular basis of Schindler and Kanzaki diseases. J Mol Biol. 2009 Oct 23;393(2):435-47.
[82] Mesters JR, Hilgenfeld R. Protein Glycosylation, Sweet to Crystal Growth? Cryst Growth Des. 2007;7(11):2251-3.
[83] Stura EA, Nemerow GR, Wilson IA. Strategies in the crystallization of glycoproteins and protein complexes. J Cryst Growth. 1992;122(1–4):273-85.
[84] Lin DY, Tanaka Y, Iwasaki M, Gittis AG, Su HP, Mikami B, et al. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. Proc Natl Acad Sci U S A. 2008 Feb 26;105(8):3011-6.
[85] Holdom MD, Davies AM, Nettleship JE, Bagby SC, Dhaliwal B, Girardi E, et al. Conformational changes in IgE contribute to its uniquely slow dissociation rate from receptor FcvarepsilonRI. Nat Struct Mol Biol. 2011 May;18(5):571-6.
[86] Schrödinger LLC. The PyMOL molecular graphics system. 1.3 ed.
[87] Crispin M, Stuart DI, Jones EY. Building meaningful models of glycoproteins. Nat Struct Mol Biol. 2007 May;14(5):354; discussion -5.
[88] Crispin M, Scanlan CN, Bowden TA. Structure and Biosynthesis of Glycoprotein Carbohydrates. In: Moo-Young M, editor. Comprehensive Biotechnology. 2 ed: Elsevier; 2011. p. 73-90.

[89] Ranzinger R, Maaβ K, Lütteke T. Bioinformatics Databases and Applications Available for Glycobiology and Glycomics. In: Owens RJ, Nettleship JE, editors. Functional and Structural Proteomics of Glycoproteins2011. p. 59-90.
[90] Ranzinger R, Herget S, von der Lieth CW, Frank M. GlycomeDB--a unified database for carbohydrate structures. Nucleic Acids Res. 2011 Jan;39(Database issue):D373-6.
[91] Bohne A, Lang E, von der Lieth CW. SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides. Bioinformatics. 1999 Sep;15(9):767-8.
[92] Woods Group. GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA 2005-2012.
[93] Bohne-Lang A, von der Lieth CW. GlyProt: in silico glycosylation of proteins. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W214-9.
[94] Lutteke T, von der Lieth CW. pdb-care (PDB carbohydrate residue check): a program to support annotation of complex carbohydrate structures in PDB files. BMC Bioinformatics. 2004 Jun 4;5(69):69.
[95] Read RJ, Adams PD, Arendall WB, 3rd, Brunger AT, Emsley P, Joosten RP, et al. A new generation of crystallographic validation tools for the protein data bank. Structure. 2011 Oct 12;19(10):1395-412.
[96] Slynko V, Schubert M, Numao S, Kowarik M, Aebi M, Allain FH. NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. J Am Chem Soc. 2009 Jan 28;131(3):1274-81.
[97] Egorova-Zachernyuk TA, Bosman GJ, Degrip WJ. Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. Appl Microbiol Biotechnol. 2011 Jan;89(2):397-406.
[98] Saxena K, Dutta A, Klein-Seetharaman J, Schwalbe H. Isotope labeling in insect cells. Methods Mol Biol. 2012;831:37-54.
[99] Pickford AR, O'Leary JM. Isotopic labeling of recombinant proteins from the methylotrophic yeast Pichia pastoris. Methods Mol Biol. 2004;278:17-33.
[100] Pickford AR, Smith SP, Staunton D, Boyd J, Campbell ID. The hairpin structure of the (6)F1(1)F2(2)F2 fragment from human fibronectin enhances gelatin binding. Embo J. 2001 Apr 2;20(7):1519-29.
[101] Wood MJ, Sampoli Benitez BA, Komives EA. Solution structure of the smallest cofactor-active fragment of thrombomodulin. Nat Struct Biol. 2000 Mar;7(3):200-4.
[102] Metzler WJ, Bajorath J, Fenderson W, Shaw SY, Constantine KL, Naemura J, et al. Solution structure of human CTLA-4 and delineation of a CD80/CD86 binding site conserved in CD28. Nat Struct Biol. 1997 Jul;4(7):527-31.
[103] Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, et al. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science. 2006 Nov 17;314(5802):1148-50.
[104] Saridakis E, Khurshid S, Govada L, Phan Q, Hawkins D, Crichlow GV, et al. Protein crystallization facilitated by molecularly imprinted polymers. Proc Natl Acad Sci U S A. 2011 Jul 5;108(27):11081-6.