

GLYCOSYLATION OF PROTEINS – STRUCTURE, FUNCTION AND ANALYSIS

AUTHOR: RICHARD EASTON, PH.D., TEAM LEADER, CARBOHYDRATE ANALYSIS, SGS M-SCAN LTD

Glycosylation is one of the most widely observed, and structurally diverse, forms of post translational modification (PTM) of proteins. Animal, plant, fungal and bacterial cell systems all have the ability to glycosylate proteins and, whilst similarities do exist between these groups, there are also significant differences in terms of monosaccharides used and carbohydrate structures produced. Glycosylation is therefore an immensely diverse form of PTM, with different cell lines exhibiting significantly different patterns of protein glycosylation due largely, but not exclusively, to the expression of different repertoires of glycosidases and glycosyltransferases within the ER and Golgi apparatus of each cell. This variety of profiles gives cells the flexibility to produce glycoproteins with specific properties related to the function of that cell but from a limited number of monosaccharides. Glycans have been shown to have a range of specific biological roles: by targeting of a glycoprotein to specific receptors both intra- and intercellularly, by controlling of residence time in the bloodstream, and by assisting in correct molecular folding¹.

The control of these functions is all defined through the use of only a handful of building blocks; the monosaccharide residues. But, as with proteins, it is how they are structurally put together that provides the key to their ability to have such exquisite functional control. Unlike amino acids which are linked in a linear form via amide bonds, monosaccharides are cyclic structures that can be linked at different positions to one another around the ring and also create branched oligosaccharides. It is this ability to link in different ways that can produce different functional structures and can induce such diversity.

There are two basic types of protein glycosylation: N-glycosylation and O-glycosylation. These have significant differences in terms of their biosynthesis and structure, as well as location within the protein chain. N-glycans are produced from a 14-mer precursor structure that is added to Asparagine residues in the consensus sequence Asn-X-Ser/Thr, where X can be any amino acid except

Proline. Not all consensus sequences are glycosylated due to the 3-dimensional structure of the protein in the region of the consensus sequence which can prevent the oligosaccharyl transferase from binding. The precursor is trimmed down to an invariant core structure (three mannose and two N-Acetylglucosamine residues) which is then built up using the repertoire of glycosyltransferases present in the cell in concert with the overall 3-D structure of the molecule. O-glycans are present on Serine or Threonine residues and do not have a linear consensus sequence from which their location in a protein can be identified. Mammalian O-glycans are small structures (typically composed of 3 to 6 monosaccharides) which are attached to the protein via the monosaccharide N-Acetylgalactosamine. Examples of N- and O-glycans are shown in Figures 1 and 2. Key characteristics are shown in Table 1.

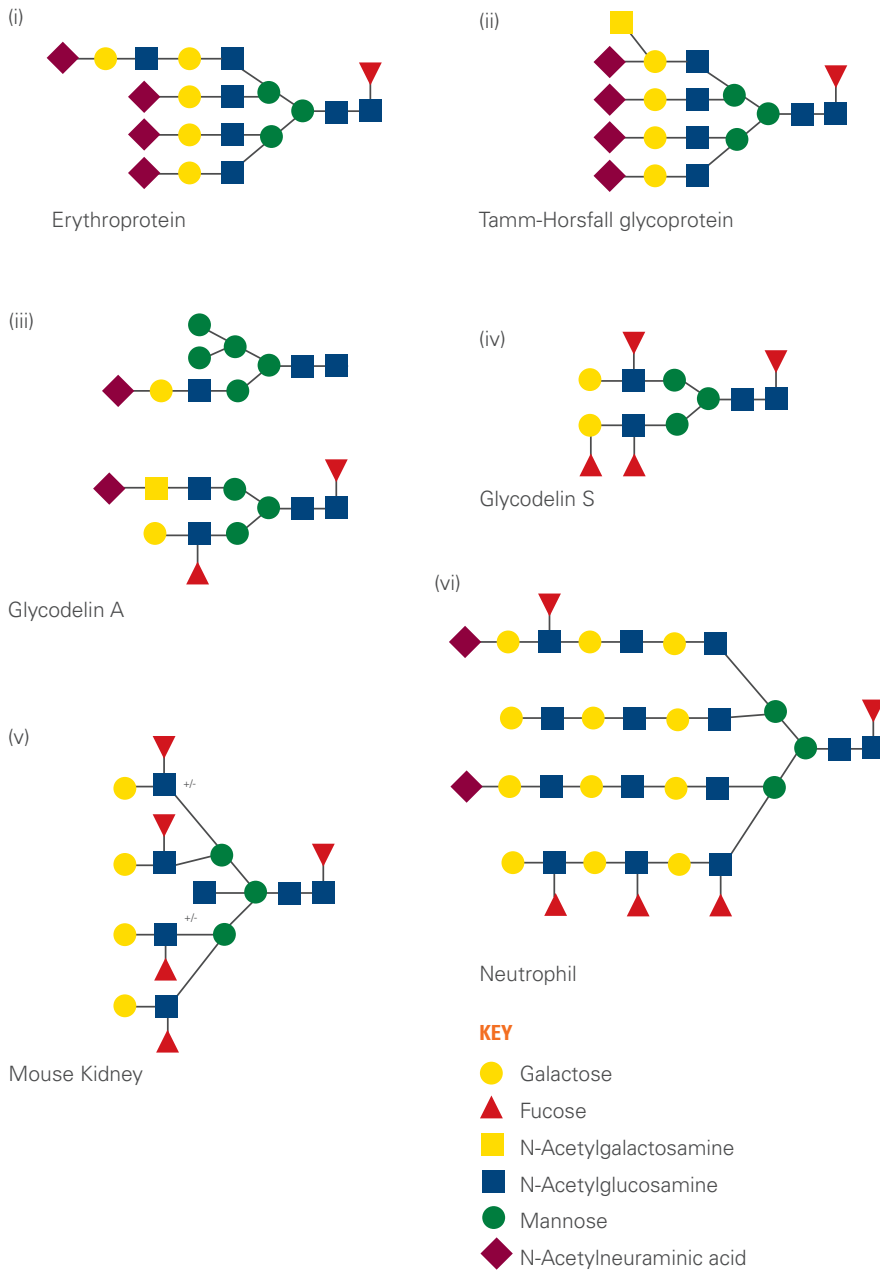
This diversity of structure and the significance of the glycosylation means that close attention must be paid to the

nature of glycans on any glycoprotein biopharmaceutical. Therefore, a detailed investigation should be performed to determine the structures of any glycans present, to confirm that they are suitable for the activity of the drug and to verify the glycans present are unlikely to cause an immune response (because they are foreign to the recipient). This is true for the development of novel products and also for biosimilars. The analysis of glycans is an important issue for regulatory agencies and is discussed in the ICH guideline Q6B².

Because of this diversity, the structural characterisation of glycans is a complex process that relies on several different techniques, each geared towards analysing different facets of the structure: overall profile, composition of specific units and linkage analysis between the residues.

Mass spectrometry has proven to be a highly valuable tool in this characterisation process. The diversity of techniques

FIGURE 1: EXAMPLES OF N-GLYCANS AND THE GLYCOPROTEINS OR CELL TYPES IN WHICH THEY ARE FOUND

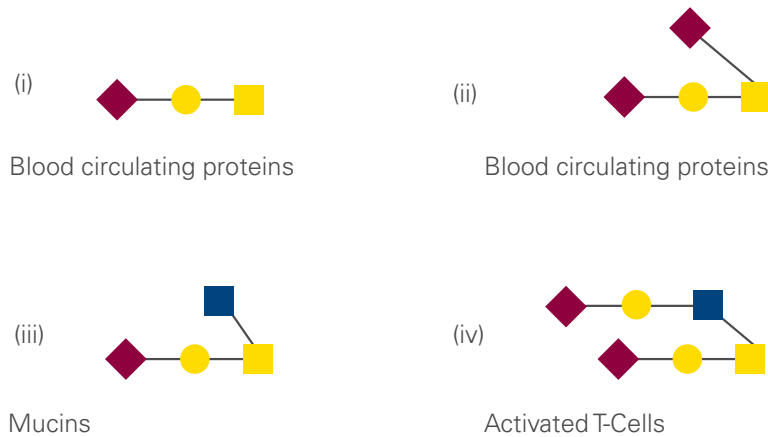


that are now available, coupled with chemical and biochemical treatment of glycans, means that all aspects of glycan structure can be examined. The mass spectrometric techniques employed use so called “soft” and “hard” ionisation processes to obtain either intact or fragmentation information respectively, all of which can be used to build a picture of the structure of the glycan profile on any glycoprotein. Furthermore, it is necessary to define which glycans are present at each glycosylation site in a glycoprotein and to this end on-line liquid chromatography-mass spectrometry can be used, in conjunction with the methods detailed below, to isolate and characterise glycopeptides.

Initial studies carried out during product development do not necessarily require full glycan analysis but should address basic questions so that the optimal cell line can be selected and bioreactor conditions for production can be assessed. Once such initial basic questions have been addressed, then more complex studies can be performed to determine fuller glycan structural characteristics.

The simplest test that can be performed is a determination of the presence or absence of glycosylation on a glycoprotein. This test can be carried out using Gas Chromatography-Mass Spectrometry (GC-MS) of released and derivatised monosaccharides or may also be performed using liquid chromatography. The data obtained can be used to quantitate the levels of each monosaccharide present and also determine what type(s) of glycosylation are likely to be found. In a simple sense, the data can be used to show that a glycoprotein is glycosylated to the level expected or if certain key monosaccharides (such as N-Acetylneuraminic acid) are at the level expected for optimal activity. For example, low levels of N-Acetylneuraminic acid could result in accelerated clearance of the glycoprotein from the bloodstream via the asialoglycoprotein receptor in the liver and thus impact efficacy³.

In conjunction with monosaccharide analysis initial studies should also focus on determining the overall profile of any N-and/or O-glycans present on the sample. This can be carried out using mass spectrometry or using chromatographic

FIGURE 2: EXAMPLES OF O-GLYCANS AND THE GLYCOPROTEINS OR CELL TYPES IN WHICH THEY ARE FOUND**KEY**

- Galactose
- N-Acetylgalactosamine
- N-Acetylglucosamine
- ◆ N-Acetylneuraminic acid

methods such as high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) or chromatography of 2-aminobenzamide (2-AB) labelled glycans coupled with fluorescence detection and/or mass spectrometric detection. A mass spectrometric approach using Matrix Assisted Laser Desorption Ionisation (MALDI) carried out on released and chemically derivatised (permethylated) glycans provides compositional information on the structures present. These chromatographic approaches give profiles that can be used as comparators for data between batches. The use of 2-AB glycan tagging, in conjunction with fluorescence detection and electrospray mass spectrometry, provides very useful information on the structures of the glycans present giving overall mass and fragment ion information as well as a chromatographic profile.

Following the initial work to obtain glycan profiles, it may soon become apparent that some structural compositions observed can be assembled biosyn-

TABLE 1: COMPARISON OF THE PROPERTIES OF N-GLYCOSYLATION AND O-GLYCOSYLATION

N-GLYCOSYLATION	O-GLYCOSYLATION
Attached to Asparagine residues	Attached to Serine and Threonine residues (mostly)
Linear protein consensus sequence for attachment	No linear protein consensus sequence for attachment
Conserved core saccharide structure	No conserved core saccharide structure (but family of structures)
N-Acetylglucosamine linker	N-Acetylgalactosamine linker (mostly)
Biosynthesised by remodeling of a specific attached oligosaccharide	Biosynthesised by sequential addition of single monosaccharides
Can be relatively large (12-25 monosaccharide residues)	Usually relatively small (3-6 monosaccharide residues)
Extracellular, vesicular or cell surface	Extracellular or cell surface (mostly)

thetically in more than one way, each of which is wholly consistent with the known biosynthetic pathway of glycan synthesis. For example the composition $\text{NeuAc}_2\text{Hex}_6\text{HexNAc}_5\text{Fuc}$ suggests either a biantennary structure with one lactosamine repeat or a triantennary structure with no lactosamine repeats (Figure 3). Furthermore, the Fucose residue can be present on the core or on an antenna. In these instances it is useful to obtain fragment ion information from permethylated glycans using electrospray mass spectrometry. This technique produces fragment ions from the non-reducing end of the molecule and these fragment ions can be used to determine whether particular antennal epitopes are present. In the case of the example above, the

location of the Fucose could prove critical as certain Fucose containing structures have been shown to act as ligands for receptors and thus target the glycoprotein to a specific location.

The final aspect of glycan structure that needs to be considered is the determination of the linkages between the monosaccharides. The nature of these linkages can play a critical role in the activity or targeting of a glycoprotein, and can also have possible immunological issues if not found in the host system. An example of this is N-Acetylneuraminic acid (sialic acid), which can be linked to either carbon 3 or carbon 6 of Galactose. These different linkages can produce structures capable of binding to different receptors.

Mass spectrometry can again be used for this type of analysis in conjunction with a multistep chemical hydrolysis and derivatisation process. The end result of the chemistry is that glycans are converted to their constituent monosaccharides containing chemical labels on unbound hydroxyl groups and different chemical labels on hydroxyls that were originally linked to other monosaccharides. These derivatives display characteristic fragmentation spectra when analysed by GC-MS and the linkages can be determined based on the retention times and fragmentation patterns observed (Figure 4).

The use of chemistry and mass spectrometry together provides a powerful approach to the structural elucidation of glycans. However there are instances when other techniques need to be incorporated into a structural investigation. One example of this is assessing a glycoprotein for the $\text{Gal}\alpha 1\text{-}3\text{Gal}$ epitope. An example of an N-glycan carrying this epitope is shown in Figure 5. This epitope is found on glycoproteins in most mammals including New World primates but is not found in humans. Humans have been shown to carry antibodies to this antennal structure and therefore therapeutics carrying this epitope run the risk of producing an immune response. As such, glycoproteins produced in cell lines capable of biosynthesising this

FIGURE 3: A DISIALYTED BIANTENNARY STRUCTURE CONTAINING A LACTOSAMINE REPEAT (i) HAS THE SAME COMPOSITION AND THEREFORE MASS AS A DISIALYLATED TRIANTENNARY STRUCTURE (ii)

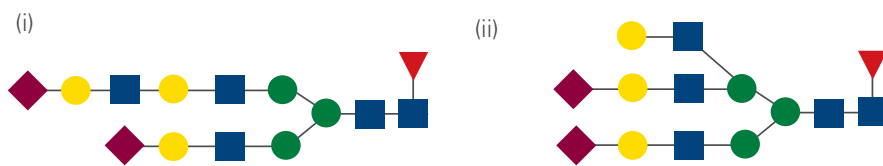


FIGURE 4: GC-MS ANALYSIS OF DERIVATISED MONOSACCHARIDES PRODUCED FROM AN N-GLYCAN POPULATION

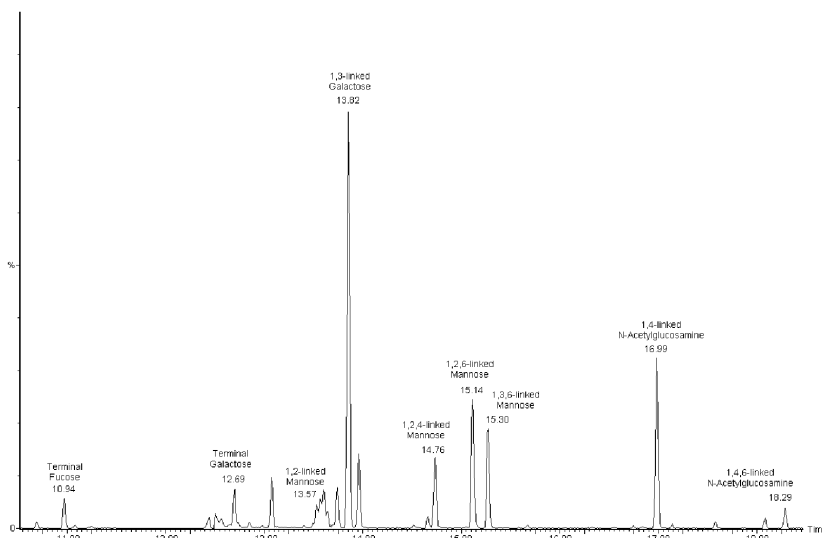
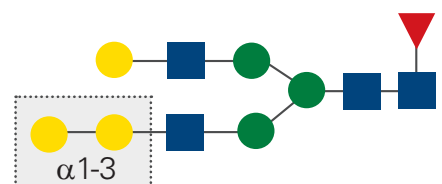


FIGURE 5 EXAMPLE OF AN N-GLYCAN CARRYING THE $\text{Gal}\alpha 1\text{-}3\text{Gal}$ EPITOPE (BOXED REGION)



structure must be analysed for its presence. This can be performed using mass spectrometric analysis of the released glycans prior to and following digestion with α -Galactosidase, an enzyme that is capable of removing the terminal α -linked Galactose residue. Analysis of the glycan population after digestion will show a mass shift if Galactose residues have been removed.

The data generated from all of the above studies, plus knowledge of the known biosynthetic pathways for glycan biosyn-

thesis, allow detailed oligosaccharide structures to be postulated. Whilst this level of detail is needed to determine the glycans present on a glycoprotein, it may not be necessary to perform this level of analysis on all batches produced, providing there are no changes in the manufacturing process. A liquid chromatographic study (using for example HPAEC-PAD or HPLC of 2-AB tagged glycans) carried out on a fully characterised glycan population can be used as a simple reference profile against which other batches of product can be compared.

In summary, mass spectrometry based techniques provide an extremely powerful tool for glycan characterisation and can help to assess and modify the developmental pathway of a therapeutic glycoprotein. The techniques available can be used to provide as much detail as is needed at each stage of product development and can then be extended, running hand in hand with the development of the product itself.

REFERENCES

1. *Essentials of Glycobiology* 2nd Edition. Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E. Eds. Cold Spring Harbor Laboratory Press, (2009)
2. ICH Harmonised Tripartite Guideline, Topic Q6B, *Specifications: Test Procedures and Acceptance Criteria for Biotechnological/ Biological products. Step 5, Consensus Guideline*, (CPMP/ICH/365/96), (September 1999)
3. Ashwell, G. and Harford, J. Carbohydrate-specific receptors of the liver. *Annu. Rev. Biochem.* **51**:532-554 (1982)

To receive future articles on current trends and regulatory updates, subscribe to SGS' Life Science News at www.sgs.com/lss_subscribe

CONTACT INFORMATION

EUROPE

BELGIUM

+32 10 42 11 11
be.pharmaqc@sgs.com

FRANCE (PARIS)

+33 1 41 06 95 93
fr.pharmaqc@sgs.com

FRANCE (POITIERS)

+33 (0) 5 49 57 04 04
clinicalresearch@sgs.com

GERMANY (BERLIN)

+49 30 3460 7500
de.pharmaqc@sgs.com

GERMANY (FREIBURG)

+49 761 6116 7760
de.pharmaqc@sgs.com

GERMANY (TAUNUSSTEIN)

+49 6128 744 245
de.pharmaqc@sgs.com

SWITZERLAND (GENEVA)

+41 22 794 8374
pharmaqc@sgs.com

UK (WOKINGHAM)

+44 (0) 1189 896940
pharmaqc@sgs.com

ASIA

INDIA

+91 44 2254 2601
in.pharmaqc@sgs.com

SINGAPORE

+65 677 53 034
sg.pharmaqc@sgs.com

CHINA

+86 21 6115 2197
cn.pharmaqc@sgs.com

TAIWAN

+886 2 2299 3279 ext 2500
tw.pharmaqc@sgs.com

NORTH AMERICA

CANADA

+ 1 905 364 3757
ca.pharmaqc@sgs.com

USA (FAIRFIELD, NJ)

+ 1 888 747 8782
us.pharmaqc@sgs.com

USA (LINCOLNSHIRE, IL)

+1 847 821 8900
us.pharmaqc@sgs.com

USA (WEST CHESTER, PA)

+ 1 610 696 8210
us.pharmaqc@sgs.com

WWW.SGS.COM/PHARMAQC

WHEN YOU NEED TO BE SURE

SGS