

Glycosylated Bioproducts – Breaking Down the Benefits

Protein glycosylation patterns are extremely important for many reasons, and with biotechnology now a major source of new therapeutic protein drugs, their characterisation can be a valuable resource; Dr Fiona M Greer and Dr Richard L Easton, M-Scan (Mass Spectrometry Consultants and Analysts) Ltd explain

Fiona is a Founder and Director of M-Scan, a group of contract service laboratories specialising in bio/pharmaceutical analysis. With over 25 years' experience in the characterisation of glycoproteins, and many publications in this area, she is involved in the analysis of a diverse range of biotechnology products and is particularly interested in regulatory requirements. She regularly gives talks at international meetings and has presented various technical training courses.

Richard obtained his PhD from Imperial College in 1997 on the structural characterisation of glycoproteins using mass spectrometry. He subsequently spent several years there as a Postdoctoral Research Scientist, working on glycoprotein characterisation with an emphasis on glycan elucidation. He joined M-Scan Limited in 2003 as a Senior Biochemist, and is involved in all aspects of carbohydrate characterisation.

Many of the biotechnology-derived molecules in production exist in their native form as glycoproteins. When expressed in recombinant form in mammalian systems, the protein backbone may be identical to the native. However, the glycosylation pattern will depend on the cell type used and the physiological status of that cell. Although there are examples where glycosylation of the recombinant molecule does not appear to affect its activity, it is now widely agreed that protein glycosylation patterns and their degree of microheterogeneity are extremely important for many reasons, ranging from biological activity and clinical efficacy to patentability. These factors, in part, have stimulated much interest in the field of carbohydrate and glycoconjugate analysis. This has led to a symbiosis between glycobiology research and the development of new analytical technologies – particularly those based on mass spectrometry – capable of addressing the structural problems posed.

The biotechnology industry has now established itself as a major source of new therapeutic protein drugs, which can be described as complex biological entities. The majority of these, such as erythropoietin (EPO), interleukin-2 (IL-2), interferon (IFN), granulocyte-colony stimulating factor (G-CSF), granulocyte macrophage-colony stimulating factor (GM-CSF) and tissue plasminogen activator (TPA), occur naturally in a glycosylated form. In some cases, for example, IL-2, γ -INF, G-CSF and GM-CSF, glycosylation of the recombinantly produced protein does not appear to be required for *in vivo* activity. However, in others – most notably TPA and EPO – glycosylation is certainly required for biological activity. In addition, glycosylation has been shown to play a major role in bioactivity in recombinant antibodies (rMAbs).

The role that the carbohydrate moiety performs may vary; for example, the glycosylation may target the molecule to a specific location, thus influencing the clearance rate from the blood, or simply perform a structural function such as holding an active

site in the 'correct' three-dimensional shape. It follows that loss or alteration of the saccharide(s) present may have a significant effect on the activity.

Glycosylation is arguably the most important of the numerous post-translational modifications, but what is undeniable is that it presents a unique challenge for in-depth analysis. The population of sugar units attached to individual glycosylation sites on any protein will depend on the host cell type used. All glycoproteins exist as a mixture of different 'glycoforms'.

REGULATORY PERSPECTIVES

The development of glycoproteins as human therapeutics has also raised a number of issues with regulatory agencies. These have included concerns about potential immunogenicity of the carbohydrates when the proteins are produced in non-human mammalian host cells, as well as issues related to the inherent heterogeneity of the glycans.

Characterisation of the carbohydrate structure of a biopharmaceutical is an essential regulatory requirement. The ICH guideline Q6B (1) states:

“For glycoproteins, the carbohydrate content (neutral sugars, amino sugars and sialic acids) is determined. In addition, the structure of the carbohydrate chains, the oligosaccharide pattern (antennary profile) and the glycosylation site(s) of the polypeptide chain is analysed, to the extent possible.”

Therefore, structural analysis of the carbohydrate portion of a recombinant glycoprotein is essential to assess the types of glycosylation present on the protein, allow a comparison of the glycosylation on the recombinant product with the natural glycoprotein (if available) and to assess the antennae of the oligosaccharides present for incomplete or antigenic motifs.

ANALYSIS OF GLYCOSYLATION

The protein or carbohydrate chemist analysing samples which may be glycosylated must therefore answer several questions. Firstly, is the sample glycosylated? If so, what is the type of glycosylation, N- or O-linked, or both? Where are the sites of glycosylation? What is the carbohydrate structure and composition at each site? Are the methods that will be employed suitable to probe these structures?

In addition, the emergence of new expression systems – insect and plant cells for example – will necessitate analysis of non-mammalian carbohydrates. So are the methods that will be employed capable of novel structure determination?

Several of the methods which are now being applied to such carbohydrate structural problems have been discussed in previous issues of *EBR* (2,3). For a decade, the most important procedure for biopolymer research has been mass spectrometry. As mass spectrometric techniques have advanced, the instrumentation has become more accessible. The same types of instruments can be used in the detailed study of both the protein and the carbohydrate moieties of the molecule, so that the overall structure of glycoproteins can be elucidated using mass spectrometry. Apart from the ability to study non-protein modifications, such as sulphation and phosphorylation, the other major strength of an MS approach is the analysis of mixtures. This has obvious applications in the analysis of heterogeneous glycoforms and glycan populations.

Various ‘soft’ ionisation MS techniques have been utilised for glycoprotein analysis, including electrospray mass spectrometry (ES-MS), on-line ES-MS (where the MS is coupled to an HPLC system), matrix assisted laser desorption ionisation mass spectrometry (MALDI-MS) and, for derivatised carbohydrates, MALDI-MS, fast atom bombardment mass spectrometry (FAB-MS) and gas chromatography mass spectrometry (GC-MS).

Powerful strategies can be used to analyse both free (underderivatised) and derivatised glycans to determine sites of glycosylation of both N- and O-linked structures, the identity of terminal non-reducing ends (potentially the most antigenic structures) and identify the type of oligosaccharide present (4).

Due to the high degree of glycan heterogeneity frequently encountered on glycoproteins, a variety of mass spectrometric techniques must be employed in order to fully characterise various aspects of the components present in these complex mixtures.

An initial study of a glycoprotein will often involve a determination of the molecular weight of the intact glycoprotein. This is most readily achieved using either MALDI-MS or electrospray mass spectrometry. The latter technique is favoured due to the nature of the ionisation process and mass analyser, which can

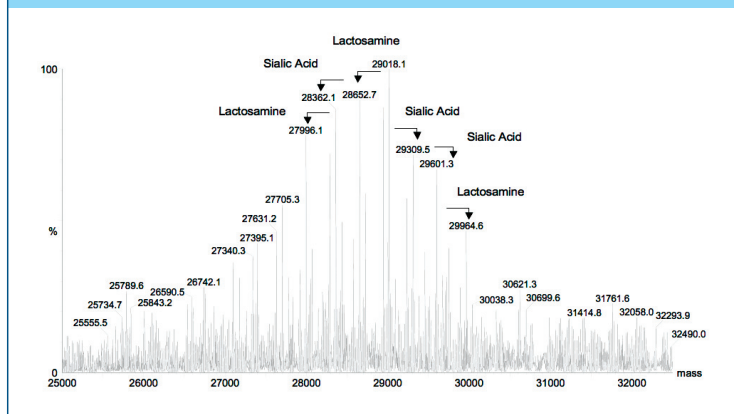
provide superior resolution of the molecular species present. The data generated can be compared to the expected molecular weight based on the known amino acid sequence and any mass discrepancies investigated as either direct variations on the peptide chain (such as amino acid substitutions or deletions), or post-translational modifications of the protein. The heterogeneity often encountered in glycosylation manifests itself in the mass spectrum as a series of signals separated by discrete mass intervals that can be attributed to individual glycoforms of the molecule. Furthermore, the presence of more than one glycosylation site, and the resulting increase in heterogeneity, can produce a complex spectrum that is very difficult to interpret and from which there can be no guarantee of identifying all the glycan types (an example of which is shown in Figure 1, following transformation of an electrospray-time of flight mass spectrum to produce an intact mass map).

Detection of potential glycosylation from intact mass analysis is only the first stage of a glycan characterisation study. Subsequent mass spectrometric techniques are then normally employed to define the structures of the glycans present.

Following intact mass measurement, N- and O-linked glycan population analysis is usually performed. This is most readily carried out following enzymic release of the N-glycans (usually using the broad specificity enzyme Peptide N-glycosidase F) followed by a chemical release of the O-glycans from the de-N-glycosylated glycoprotein. Chemical release is favoured for removal of O-glycans due to the very limited specificity of the available de-O-glycosylating enzymes. It should be noted that methods used for analysis of glycans must be tailored to the system that is being examined. For example, the enzyme Peptide N-glycosidase F will very readily cleave N-glycans from glycoproteins produced in mammalian cell lines, but will only potentially release a subset of the N-glycans produced in plant and insect glycoproteins. This is due to the presence, on plant and insect N-glycans, of a Fucose residue alpha1-3 linked to the core N-Acetylglucosamine residue. The presence of this linkage renders the structure resistant to cleavage by PNGase F. Once released, the N- and O-glycans can then be derivatised

Figure 1: Transformed mass spectrum of a glycoprotein containing three N-glycan chains and one O-glycan chain

The degree of complexity produced by the oligosaccharide chains can be readily seen. Mass differences observed can be assigned to glycan heterogeneity (as shown).



and analysed by MALDI-MS. The data obtained, plus a knowledge of the biosynthetic pathway, allows oligosaccharide compositions to be postulated for the signals observed (an example of this is shown in Figure 2). Minor components that would possibly be missed in the analysis of the intact sample are more readily observed in these data.

These glycan-mapping processes allow much information to be obtained relating to the structures of the oligosaccharides but, since only compositional data is obtained, questions can still remain with regard to the precise nature of the oligosaccharides

Figure 2: Matrix assisted laser desorption ionisation (MALDI) mass spectrum of derivatised N-glycans, released from the same glycoprotein as shown in Figure 1

With a knowledge of monosaccharide residue masses and mammalian N-linked oligosaccharide biosynthetic pathways the mass values can be translated into oligosaccharide compositions. The complexity has decreased from that seen in Figure 1 since the additive effect of having multiple N-linked glycosylation sites and the contribution from O-linked glycosylation have been removed.

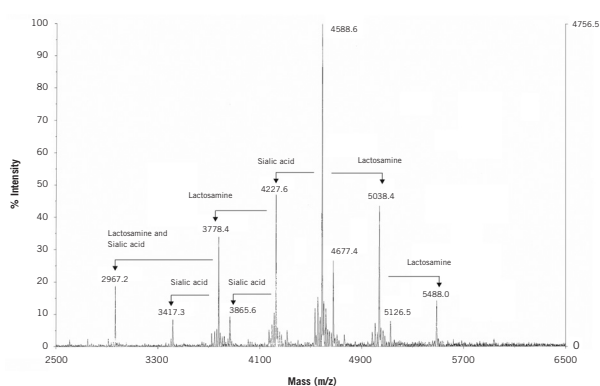
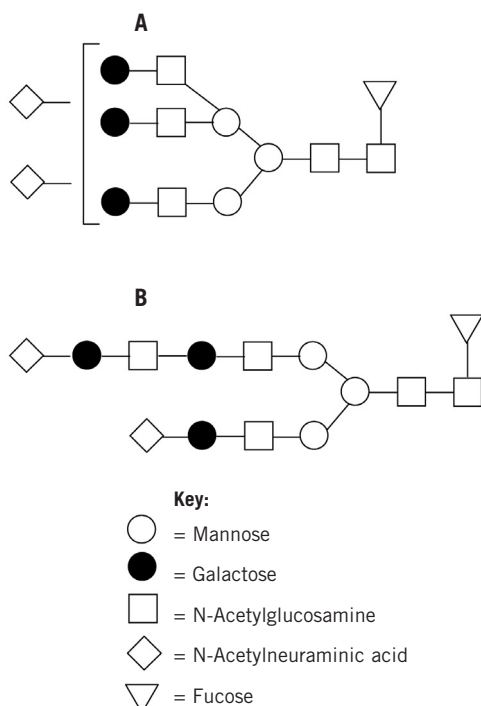


Figure 3: Structures A and B are the same mass as they have the same number and type of monosaccharides. The triantennary nature of structure A will give different information from structure B (biantennary with lactosamine repeat) in both linkage analysis by GC-MS and as antennary fragment ions by electrospray.



present. For example, signals can be observed that are consistent with both multi-antennary structures, or structures with less antennae but carrying lactosamine repeat units (see Figure 3). In order to answer these more detailed questions, the following mass spectrometric techniques can be employed.

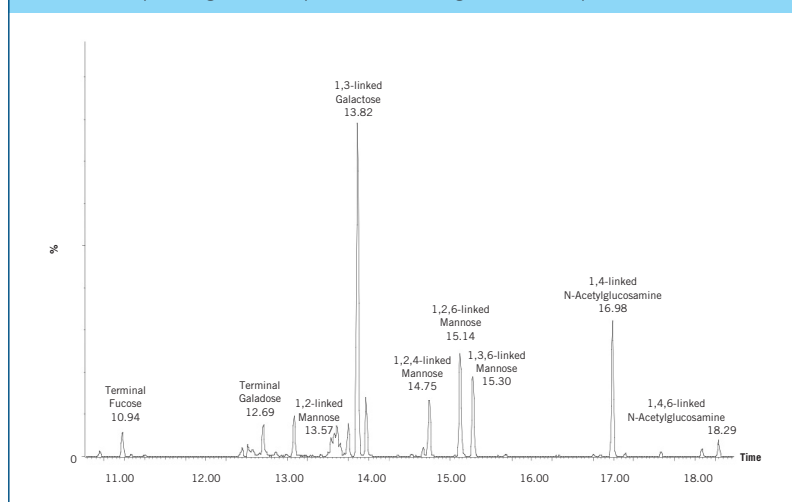
A key piece of information required to assist the structural characterisation is the determination of the structures present on the antennae of the glycans. For this, mass spectrometry can again be used, but in this instance an alternate ionisation technique is employed, such as FAB or, more commonly, electrospray. Both these techniques, with the correct use of voltages during ionisation, can be used to fragment the glycans under investigation via specific and defined pathways to produce ions derived from the antennal structures. Thus the presence of lactosamine repeat units can be identified as can, for example, Lewis^x, sialyl Lewis^x and Gal α Gal epitopes. This evidence is necessary to distinguish between possible alternate locations for specific monosaccharide residues. An example of this is fucosylation which may be found on either the core N-Acetylglucosamine, or on the antenna of an N-linked glycan. An assessment of the fragment ions produced allows more conclusions to be drawn about the structural nature of the glycans.

The techniques described above allow the initial identification of the presence of glycosylation, a study of the population present and an examination of the antennal structures present. One area of investigation that still needs to be addressed is that of linkage of the specific monosaccharides to one another. Unlike the linear chains of proteins, where a repeating amide bond links the amino function of one amino acid to the carboxyl function of the adjacent amino acid, monosaccharides can be linked together through different positions around the ring structure. This anabolic process is under the control of specific glycosyltransferases that are designed to specifically link monosaccharides together. This variety of linkage gives rise to a higher level of glycan heterogeneity for a specific population of oligosaccharides than would otherwise be possible and, due to the differing structural properties of the resultant oligosaccharides, different functions can be produced simply by altering the linkage of the monosaccharide units. This exquisite use of simple monosaccharide 'building blocks' allows precise regulation of the resultant glycoprotein and provides an extra level of complexity for manufacturers of recombinant glycoproteins. An example of this is the linkage of N-Acetylneuraminic acid (NeuAc) to Galactose (Gal), where two different linkages, and therefore structures, can be produced by mammalian cells – the second carbon of terminal NeuAc residues can either be linked to the third (NeuAc α 2-3Gal) or sixth carbon (NeuAc α 2-6Gal) of Galactose.

Identification of the various linkages of the monosaccharide units can be performed using methylation analysis, (also known as linkage analysis). This involves the production, through the use of several chemical processes, of monosaccharides that are specifically chemically labelled at each hydroxyl functional group. Initially free hydroxyl groups (those not involved in linkages between monosaccharides) are derivatised and,

Figure 4: Gas chromatography-mass spectrometry (GC-MS) analysis of derivatised monosaccharides produced during linkage analysis

The chromatogram shown has been produced by summing abundant ions produced during the electron impact fragmentation processes occurring in the mass spectrometer source.



following hydrolysis of the oligosaccharides, the hydrolysed glycosidic bonds are derivatised using a different modifier. The resultant chemical derivatives can then be separated and analysed by GC-MS. The column phase used allows each monosaccharide component to be separated on the basis of the type of monosaccharide (either Mannose or Galactose) and also the nature of the linkage. As the components elute from the GC column, they enter the source of the mass spectrometer and are fragmented by Electron Impact ionisation. The resultant mass spectrum is a 'fingerprint' for the specific linkage present in the original oligosaccharide. The combination of GC-retention time and mass spectrum therefore allows the specific monosaccharide and linkage to be defined (see Figure 4).

GC-MS is also used as a technique for identifying and quantifying monosaccharides present in a sample. This is used to determine whether levels of glycosylation are as expected (when combined with a known protein concentration), and can be also used as an early indicator of the presence or absence of N- or O-glycans.

Mass spectrometric techniques allow the characterisation of many aspects of carbohydrate structure, including population profile, structural characterisation and linkage analysis. A complementary technique to these studies is the use of chromatography to separate glycans and provide profiles based on the physical response of a carbohydrate structure to its chemical environment. This is often achieved through the use of high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). This technique separates carbohydrates via specific interactions with the stationary phase at high pH. The glycans chromatograph as anionic species and interact with the column based on glycan size, composition and linkage. The data obtained provide a profile of the glycosylation present on a product, which can be used for batch comparison studies. Identification of each component can be performed either by analysis of well-characterised standards (these standards are not always available and co-elution of components cannot be ruled out), or by

collecting and characterising the peaks of interest using mass spectrometry.

HPAEC-PAD can also be used for monosaccharide composition analysis in an analogous manner to GC-MS. An extension of this approach allows the identification and quantitation of specific sialic acid species, thus N-Acetylneuraminic acid can be distinguished from N-Glycolylneuraminic acid – evidence that can help confirm postulated oligosaccharide structures.

Frequently, the degree of characterisation sought is not limited to a study of the overall glycosylation of a molecule, but requires the investigation of glycosylation at each individual site. For these types of experiments, the same methodologies can be employed but on chromatographically purified and collected glycopeptides. The combination of online

chromatography and mass spectrometry can be used to identify the glycopeptides on which further studies can be performed.

CONCLUSION

In essence, the structural analysis of highly complicated molecules such as glycoproteins requires a battery of techniques, chemical and instrumental. Mass spectrometric methods have become an indispensable tool in protein analysis and are now being widely used in the study of complex carbohydrates. Instrumental advances have increased the sensitivity of the MS techniques, which have facilitated their use in the study of natural biological glycoconjugates as well as recombinant products, particularly in the determination of novel structures. It is now up to glycobiologists to exploit the capabilities of mass spectrometry, which in turn will stimulate the never-ending drive for refinement and improvement of current instrumentation. In the analysis of recombinant glycoproteins, these techniques have been applied at research, quality control and regulatory approval stages. ♦

The authors can be contacted at services.ltd@m-scan.com

References

1. ICH Harmonised Tripartite Guideline, Topic Q6B, Specifications: Test Procedures and Acceptance Criteria for Biotechnological/ Biological products. Step 4, Consensus Guideline, (CPMP/ICH/365/96), March 1999
2. Greer F, Reason A and Rogers M, Post-translational modifications of biopharmaceuticals – a challenge for analytical characterisation, *European BioPharmaceutical Review*, pp106-111, Summer 2002
3. Greer FM, Biopharmaceutical characterisation – considering key questions. *European BioPharmaceutical Review*, pp64-67, 2004
4. Dell A and Morris HR, Glycoprotein Structure determination by mass spectrometry, *Science*, 291, pp2,351-2,356, 2001