White Paper

Orthogonal Approaches for the Analysis of Protein Sequence and Post Translational Modifications of a Monoclonal Antibody

Authors

Mr Nicholas Michael, Protein Senior Scientist	RSSL
Dr David Neville, Glycosylation Technical Specialist	RSSL
Dr Pat Easton, Biomolecular Analysis Manager	RSSL
Dr Li Phing Liew, Glycoprofiling Scientist	Ludger Ltd
Dr Louise Royle, Head of Glycoprofiling	Ludger Ltd

Ludger

rssl
science with service

Orthogonal Approaches for the Analysis of Protein Sequence and Post Translational Modifications of a Monoclonal Antibody

Introduction

Monoclonal antibodies are an important class of biopharmaceuticals. They are expressed from living cells and therefore, are subject to complex biochemical pathways. Not all of these pathways are fully understood and many are known to be sensitive to subtle environmental changes during production. These changes may affect the final biopharmaceutical sequence, structure and post-translational modifications. This is in addition to any changes that may occur during subsequent purification. This means that the final product from one batch may be subtly different from another batch. Furthermore, each batch is a heterogeneous mix of similar molecules. Analysis of the degree of batch-to-batch variation, and batch heterogeneity, is therefore, very important to establish in order to be confident that the drug is safe and effective for medicinal use.

Post-translational modifications (PTMs) are responsible for a significant proportion of the micro-heterogeneity observed in proteins, and common modifications encountered with monoclonal antibodies are:

- Glycosylation
- Oxidation
- Deamidation
- Cyclisation of the heavy chain N-terminal glutamine to pyro-glutamic acid
- Clipping of the heavy chain C-terminal lysine

The extent of modifications such as deamidations and oxidations should be evaluated since they may affect final antibody function (1, 2). It is also essential to characterise the glycan content, structure and composition of therapeutic proteins since aberrant glycosylation may result in a reduced half-life (3), incorrect effector function or more importantly, greater immunogenicity (4).

'ICHQ6B, Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products' (5) sets out guidelines that are used when setting specifications for biomolecules. This document states 'A physicochemical characterisation program will generally include a determination of the composition, physical properties, and primary structure of the desired product. In some cases, information regarding higher-order structure of the desired product (the fidelity of which is generally inferred by its biological activity) may be obtained by appropriate physicochemical methodologies.' The complexity of proteins means that a number of techniques must be used to characterise them fully.

This paper describes four orthogonal analytical techniques used to analyse monoclonal antibody PTMs with special emphasis on glycosylation. These analyses provide examples of the types of data that are generated and deemed necessary by regulators to describe originator molecule specifications and demonstrate the fidelity of the final product from different batches.

The antibody was analysed by: (i) intact mass analysis and (ii) peptide mapping using a liquid-chromatography mass spectrometer (LC-MS).

Released N-glycans were analysed by (i) hydrophilic interaction liquid chromatography (HILIC-UPLC) and (ii) exoglycosidase sequencing and HILIC-UPLC.

A generic monoclonal antibody is depicted in Figure 1 which consists of two light and two heavy chains complexed via disulphide bridges.

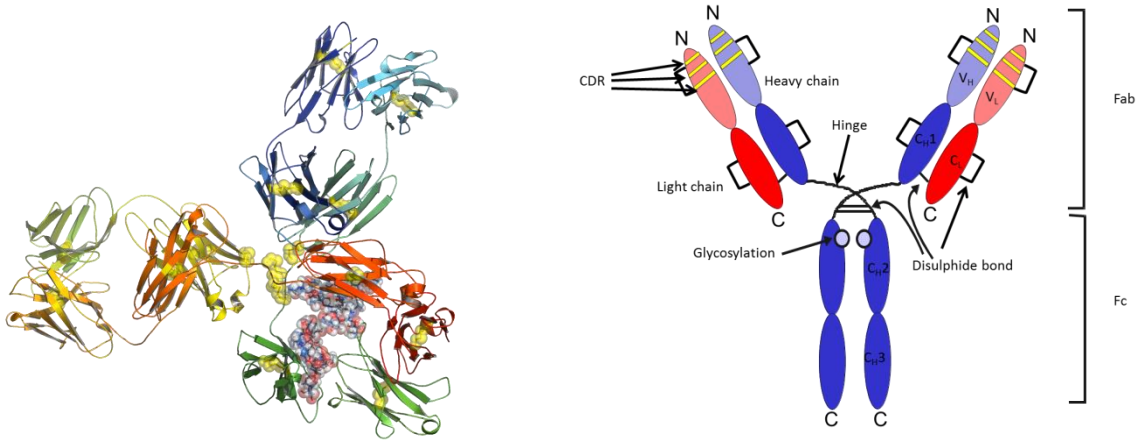


Figure 1: Generic IgG structure

N-glycans shown as balls	V_H and V_L	Variable heavy and light chain
Protein shown as ribbons	C_H1-3	Constant heavy regions 1-3
Disulphide shown in yellow	C_L	Constant light
	Fab	Fragment antigen-binding
	Fc	Fragment crystallisable
	CDR	Complementarity determining region

Antibody Details
Ref: 186006552-IgG
The antibody used as a model protein was a murine IgG monoclonal antibody purified from Mouse Myeloma NS-1 cell culture media, with a known amino acid sequence and the following known modifications:

Glutamine to pyro-glutamic acid on heavy chain N-termini
Glycosylation of asparagine residue 292
Clipping of heavy chain C-terminal lysine
17 disulphide bridges (2 intra light, 4 intra heavy, 5 interchain)

LC-MS
The intact mass and peptide mapping experiments were carried out using a Thermo QExactive LC-MS; a high resolution accurate mass (HRAM) instrument that utilises Fourier Transform (FT)-orbitrap technology.

Intact Mass Analysis

Antibody Treatment
This antibody was reduced with dithiothreitol (DTT) or deglycosylated with Peptide-*N*-Glycosidase F (PNGase F), as well as both treatments in conjunction, and the intact and the treated samples were analysed by LC-MS. Acquired spectra were deconvoluted into neutral masses using commercial software. The reduced, glycosylated heavy chain analysis is described below as an example. The deconvoluted spectrum (Figure 2) showed that the heavy chain contained three significant isoforms. These represent the potential different glycoforms of the antibody.

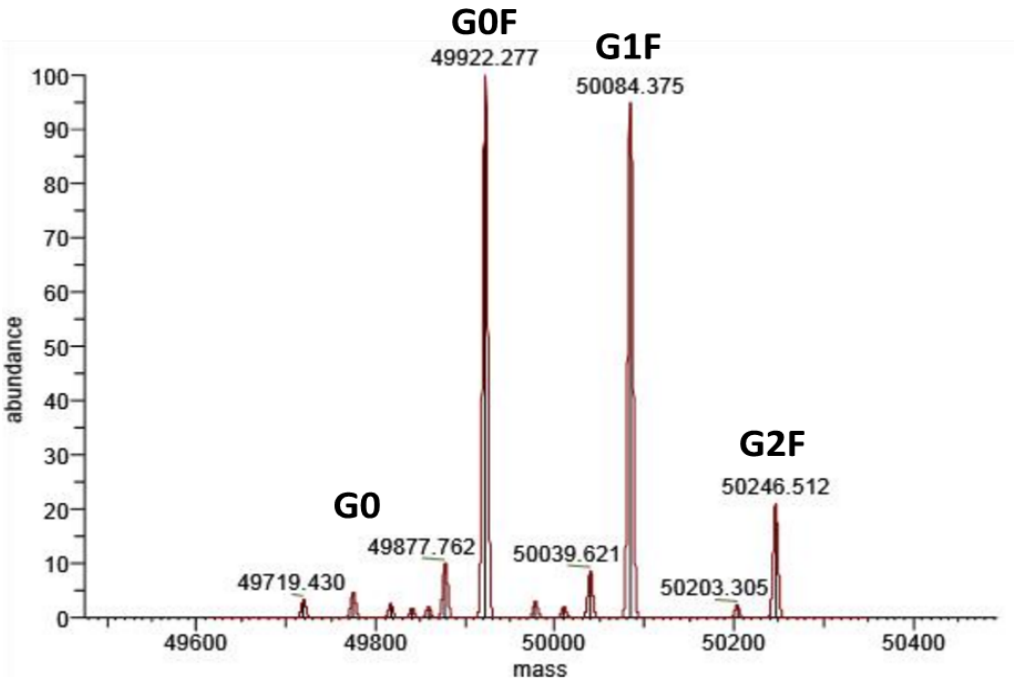


Figure 2: Deconvoluted mass spectrum of the reduced, glycosylated heavy chain

The value generated by the deconvolution software for the first major member was 49,922.3 Da., as shown in Figure 2. This is in close agreement of the theoretical value of 49,921.6 Da obtained for the protein plus one agalactosyl biantennary, core-fucosylated (G0F) glycan. The most abundant glycoform is G0F, followed closely by G1F. The abundance of the G2F glycoform is significantly lower and the G0 glycoform is only just detectable. G0, G1 and G2 correspond to the number of galactose residues present (zero, one or two, respectively) on the glycans. See table 5 for detail of structures. When the mass spectrum of the PNGase F-treated sample was analysed the peaks due to differential glycosylation disappeared, leaving only the de-glycosylated mass. This demonstrated that the previously observed masses were due to the presence of N-linked glycans on the heavy chain.

A summary of the experimentally-derived and theoretical masses obtained from the intact mass analysis is shown in Table 1. All masses are within 1.5 Da. of the expected values, giving confidence that there are no major unexpected truncations, extensions or modifications of the protein.

Molecule	**Condition**	**Predicted Average Mass (Da.)**	**Experimental Average Mass (Da.)**
Light chain	Reduced	24,193.6	24,192.3
Heavy chain	Reduced (G0F)	49,921.6	49,922.3
	Reduced & deglycosylated	48,477.3	48,478.4
Whole molecule	Intact (G0F + G0F)	148,220.4	148,221.9
	Deglycosylated	145,331.7	145,332.4

Table 1: Summary of experimentally-derived data against theoretical masses

Peptide Mapping by Mass Spectrometry

Peptide mapping is a well-established technique for studying the primary structure of proteins. It is useful for confirming the amino acid sequence and degree of modification, and it is commonly used to demonstrate lot-to-lot comparability.

The antibody was denatured, reduced and alkylated to prevent reformation of disulphide bonds or other cysteine modifications. The modified protein was quantified by measuring UV absorbance at 280 nm and using an in-house calculated extinction coefficient. The sample was then processed separately using trypsin and chymotrypsin in order to maximise the probability of achieving 100% sequence coverage.

The generated peptides were separated over a C18 reverse-phase HPLC column prior to entering the QExactive mass spectrometer. An MS1 scan was carried out, giving the mass of the peptide molecular ions. Subsequently, an MS2 experiment (fragmentation of the MS1-obtained molecular ion) was performed. MS2 fragment ion data provides confirmatory peptide sequence information and site of modification of a peptide. Using commerically available software the raw mass spectral data was matched against the theoretical light chain and heavy chain enzyme-digested peptide sequences, depending on the enzyme used. The controls were treated in the same way. In addition, the results were scrutinised and sense checked, as certain artefactual or false-positives assignments can arise and must be excluded from the final reported findings

Standard search parameters were employed using a mass accuracy of 5ppm. These included:

- Glycosylation of asparagine
- Oxidation of methionine
- Deamidation of asparagine and glutamine
- Loss of lysine from the C-terminus
- Loss of ammonia from aspargine and glutamine
- N-terminal glutamine cyclisation to pyro-glutamic acid
- Trypsin and chymotrypsin autolysis products

An example of raw data and subsequent peptide assignment is shown in Table 2 for a native peptide and its oxidised counterpart. The peptide is N388-K404 of the heavy chain (NTQPIMDTDGSYFVYSK) with M393 being oxidised. This peptide had the following molecular ion characteristics:

Charge State	Native Peptide Theoretical Monoisotopic m/z	Native Peptide Experimental Monoisotopic m/z	Mass Error (ppm)	Retention Time (mins)	Oxidised Peptide Theoretical Monoisotopic m/z	Oxidised Peptide Experimental Monoisotopic m/z	Mass Error (ppm)	Retention Time (mins)
Neutral	1964.8825	-	-	-	1980.8775	-	-	-
+1	1965.8898	1965.89236	+1.30	27.0	1981.8847	1981.88499	+0.15	24.1
+2	983.4486	983.44809	-0.52	27.0	991.4460	991.44663	+0.64	24.1
+3	655.9681	655.96747	-0.96	27.0	661.2998	661.29972	-0.06	24.1

Table 2: Example of a typical MS-1 mass assignment for an oxidised vs native peptide

The MS1 scans showed that the peptides detected existed in three charge states (1^+, 2^+ and 3^+) for both the native and oxidised species. The retention time of the native peptide was 27 minutes and the oxidised form eluted three minutes earlier, which is typical for an oxidised peptide. There is only one potential oxidisable residue in this sequence. Both assignments were confirmed by MS2 data .

A coverage map, based on MS1 data was generated for each digest. A second set of coverage maps was also generated based on MS2 confirmatory sequence data that had >95% confidence in the obtained data. An example of a coverage map for the light chain is shown in Figure 3.

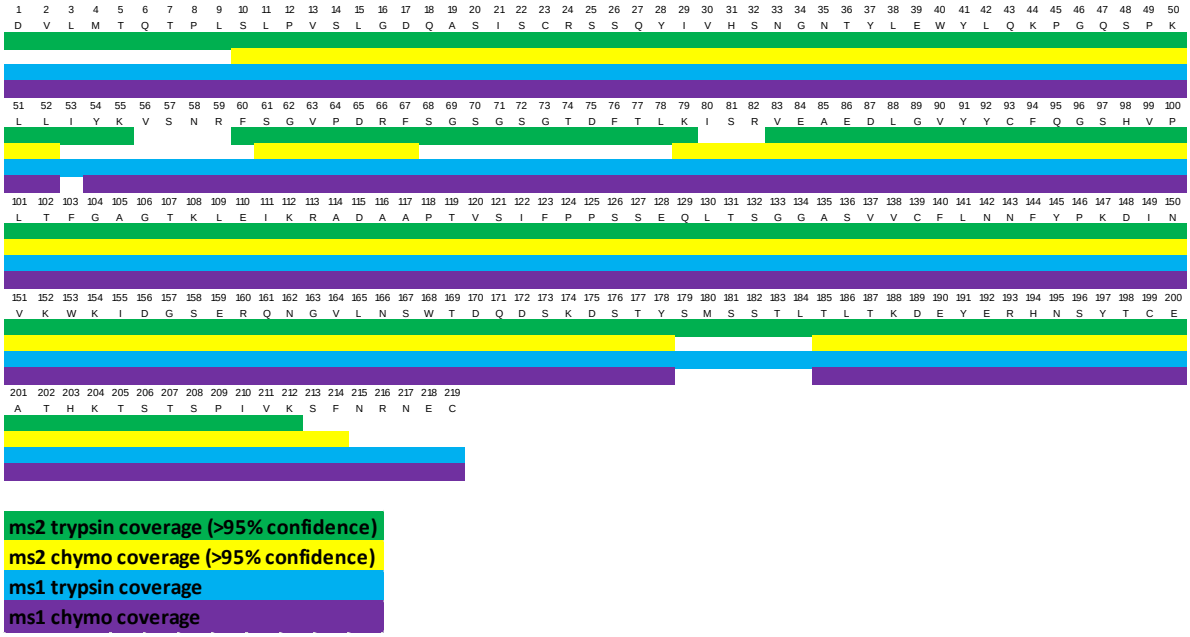


Figure 3: Coverage map of the light chain

A summary of the data obtained from the coverage maps is shown in Table 3. Full sequence coverage, based on detected and assigned peptide masses, was obtained for both the heavy and light chains by MS1 when data from both enzyme digests was combined. This demonstrates the necessity of using both approaches to maximise the data obtained to ensure successful experimental outcomes.

Protein		Trypsin	Chymotrypsin	Combined
Light Chain	MS1 coverage	100%	96.8%	100%
	MS2 coverage	93.6%	82.2%	95.9%
Heavy Chain	MS1 coverage	98.6%	98.4%	100%
	MS2 coverage	92.5%	88.9%	98.2%

Table 3: Summary of the Total Coverage Obtained with Trypsin and Chymotrypsin

The relative abundance of detected protein modifications, compared with their non-modified counterparts, can be calculated by using the MS1 peak area data. The data generated is shown in Table 4.

Protein	Modification	Abundance
Light chain	M4+Oxidation	1.0%
Light chain	N33+Deamidation*	15.2%
Light chain	N33+NH3 loss*	3.2%
Light chain	Q161+Deamidation	2.9%
Light chain	N162+Deamidation	11.6%
Heavy chain	Q1+NH3 loss (Gln to Pyro Glu)	100.0%
Heavy chain	M49+Oxidation	0.9%
Heavy chain	M82+Oxidation	0.9%
Heavy chain	N83+Deamidation	1.4%
Heavy chain	N292+G0	1.9%
Heavy chain	N292+G0F	51. 5%
Heavy chain	N292+G1F	40.2%
Heavy chain	N292+G2F	6.5%
Heavy chain	M304+Oxidation	5.5%
Heavy chain	N310+Deamidation	4.5%
Heavy chain	N310+NH3 loss	4.6%
Heavy chain	N319+Deamidation	0.6%
Heavy chain	M393+Oxidation	3.8%
Heavy chain	N429+Deamidation*	2.5%
Heavy chain	K442+Lys loss	100.0%

* = some doubt over precise residue modified

Table 4: Summary of the relative abundances of protein modifications

Except in the case of the N-terminus of the heavy chain where cyclisation of glutamine to pyro-glutamic acid is an expected modification, the loss of ammonia (NH_3 loss) is not relevant. N-terminal cyclisation was detected at 100%, as was clipping of the lysine from the C-terminus of the heavy chain. A reporting cut-off of 0.5% abundance was applied to the data. Levels of oxidations and deamidations that are greater than 5% are likely to be of concern if this was a pharmaceutical product, although the actual levels would be defined in the specification of the product.

Glycosylation Analysis

Two levels of analyses were performed on the glycans that had been released by PNGase F and subsequently fluorescently labelled with 2-aminobenzamide (2AB):

Level 1: The labelled glycans were analysed by HILIC-UPLC to obtain relative proportions of peaks for batch-to-batch comparisons. Separation is based approximately on glycan size. The retention times are expressed relative to a glucose homopolymer (GHP) as glucose units (GU) to allow comparison to standards or samples run at different times and published GU values (7).

Level 2: Detailed sequence information on the glycan structures and their relative proportions was obtained by digestion of the N-glycans with a range of exoglycosidases followed by HILIC-UPLC analysis. The exoglycosidases used were Sialidase, Beta-galactosidase, Alpha-galactosidase, Alpha-Fucosidase and Beta-*N*-acetylglucosaminidase. Combinations of enzymes are used to see those peaks whose relative retention times are shifted when compared with the undigested glycan chromatogram. The structures can then be deduced by cross-correlation of this data (8). As the Galα1-3Gal sequence is highly immunogenic, it is important that any Galα1-3Gal sequences are detected with a high degree of confidence. The presence of αGal residues directly linked to non-reducing terminal βGal residues is confirmed by specific enzyme digests using Alpha-galactosidase.

Glycoprofiling Results

- The list of N-glycan structures are shown in Figures 4 and 5 and Table 5 below. The relative amounts of the different glycan structures are given in table 6.
- No sialylated structures were detected.
- The majority (95%) of structures were core-fucosylated.
- The majority of the structures were biantennary (98%) with 1% triantennary and 0.3% Man5.
- A small percentage (3%) of structures contained α-galactose.

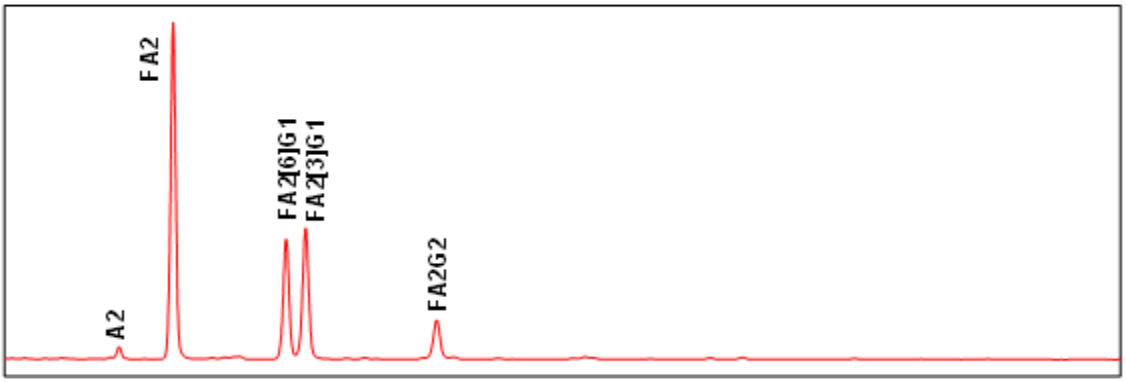


Figure 4: HILIC-UPLC profile for released, 2AB-labelled N-glycans from the model IgG

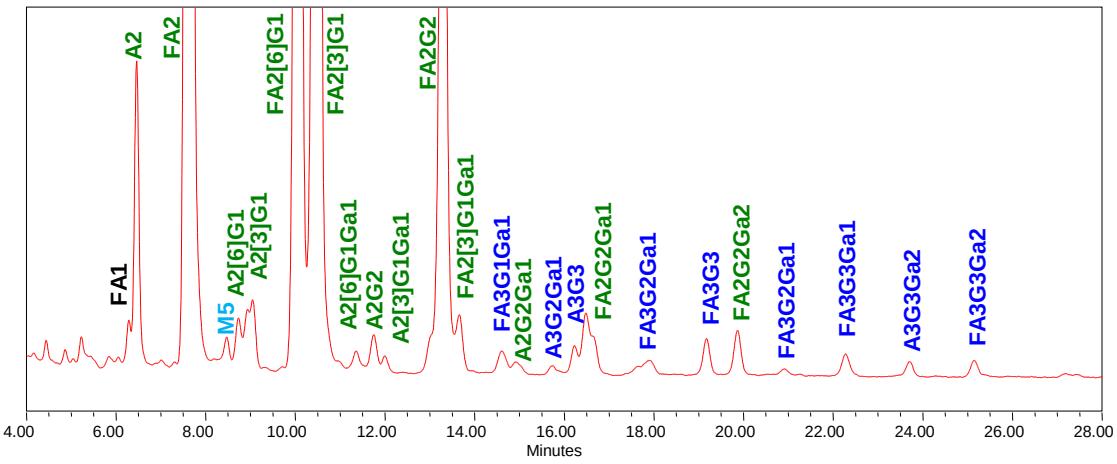
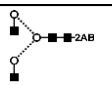
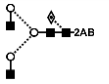
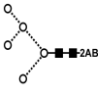
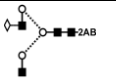
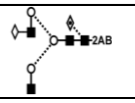
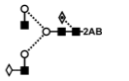
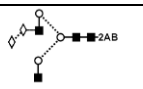
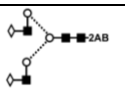
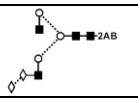


Figure 5: Zoom-in of HILIC-UPLC profile for released, 2AB-labelled N-glycans from the model IgG. The αGal-containing species are denoted by peak identifiers terminating in Ga1 and Ga2. (*Biantennary glycans labelled in green, triantennary glycans in blue*)

Peak id	Structure		Average	
			GU	% Area
1		FA1	5.42	0.24
2		A2	5.50	1.59
3		FA2	5.97	45.41
4		Man5	6.28	0.28
5		A2(6)G1	6.37	0.34
6		A2(3)G1	6.48	0.75
7		FA2(6)G1	6.80	18.61
8		FA2(3)G1	6.92	21.51
9		A2[6]G1Ga1	7.18	0.19
10		A2G2	7.30	0.28
11		A2[3]G1Ga1	7.37	0.11
12		FA2G2	7.73	7.33
13		FA2[3]G1Ga1	7.83	0.46

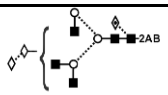
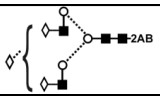
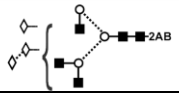
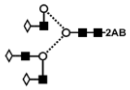
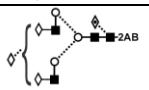
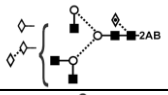
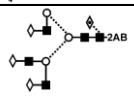
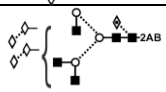
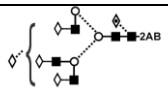
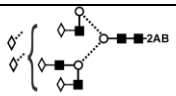
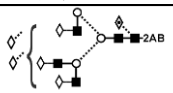
Peak id	Structure		Average	
			GU	% Area
14		FA3G1Ga1	8.10	0.24
15		A2G2Ga1	8.18	0.13
16		A3G2Ga1	8.40	0.09
17		A3G3	8.53	0.22
18		FA2G2Ga1	8.60	0.74
19		FA3G2Ga1	8.98	0.24
20		FA3G3	9.32	0.29
21		FA2G2Ga2	9.51	0.37
22		FA3G2Ga2	9.80	0.10
23		FA3G3Ga1	10.17	0.20
24		A3G3Ga2	10.56	0.13
25		FA3G3Ga2	10.96	0.15

Table 5: N-Glycan structures identified in the total undigested glycan pool

Combined Summary of MS and HPLC-HILIC Glycan Results

The data from the MS and glycan sequencing analyses are all in broad agreement for assignment of the major glycans (Table 6). It is clear that care must be taken in the interpretation and use of MS1 mass spectrometry results alone, and they should not be used as the sole basis for quantitative analysis, because many glycans are isobaric (equal mass) as illustrated in Table 6. Several other glycan species were tentatively identified in the peptide mapping experiment at abundances below 1% (not shown). However, the presence of these glycans was not confirmed by the sequencing experiment, and it is most likely that these tentative structures were produced by in-source fragmentation of the parent masses or incorrect software assignments. This illustrates the need to use orthogonal techniques to ensure correct glycan structure assignments.

Glycan Structure	MS assignment	Abundance - Glycosylation analysis	Abundance - Peptide mapping	Intact mass
FA2	G0F	45.4%	51.5%	Detected
FA2[3]G1	G1F	21.5%	40.2%	Detected
FA2[6]G1	G1F	18.6%		
FA2G2	G2F	7.3%	6.5%	Detected
FA2[6]G1Ga1	G2F	0.5%		
A2	G0	1.6%	1.9%	Detected

Table 6: Summary of the detection of the most abundant glycoforms by the 3 different methods used

Conclusions

Intact mass analysis can give a much greater level of precision and accuracy than electrophoretic techniques. Experimental error can, however, be up to 2 Da. when looking at such large molecules. Therefore, intact mass analysis is not suitable for detecting very small differences such as a single deamidation (+0.98 Da. mass shift) or a single disulphide bridge reduction/formation (+/- 2.02 Da. mass shift). It is very good for detecting larger mass differences, such as the C-terminal lysine clipping of the heavy chain. Although intact mass analysis is the quickest experiment to perform no precise sequence information is derived. Peptide mapping is required to obtain full sequence coverage, and information on protein PTMs.

Peptide mapping by high resolution LC-MS with MS2 fragmentation analysis is essential in order to identify the exact protein/peptide amino acid sequence, and to provide unambiguous assignment of any modified amino acids in a given peptide sequence. In order to obtain a semi-quantitative value for the extent of modification MS1 is needed. LC-MS protein/ peptide mapping analysis using RP-HPLC is a useful screening tool for the confirmation of the presence of glycan structures and other PTMs, although detailed sequence data from exoglycosidase sequencing should be used to obtain full information on the types of glycans present.

The relatively high levels of oxidation and deamidation found were likely due to the sample used being a commercially available monoclonal antibody that was supplied as a molecular weight standard. It was, therefore, probable that there would be more modifications then would be found on a pharmaceutical grade antibody. The choice of this antibody was based on ease of availability and a published record of its' primary sequence.

Glycosylation assignments should be confirmed and further defined by orthogonal experiments such as HILIC-UPLC with exoglycosidase sequencing of released glycans. It is possible to obtain glycan

sequence information using MS2 experiments, but experimental conditions are difficult to define and data interpretation is more complex than that for peptide sequence confirmation. Several other glycan species were tentatively identified in the peptide mapping experiment at low abundances but not confirmed by the HPLC-based sequencing experiment. This illustrates the need to use orthogonal techniques to ensure and confirm correct glycan structure assignments.

Higher levels of glucose, galactose and mannose were seen by monosaccharide analysis (data not shown) than would be expected for the defined oligosaccharides. These were possibly derived from components in the formulation buffer. It is critical that samples are clean (without additives such as formulation proteins or sugars) if monosaccharide (neutral monosaccharides or sialic acids) analysis is to provide accurate quantitation of protein-derived monosaccharides.

Mass spectrometry is a well-established technique for the analysis of (i) intact protein molecular weight, (ii) peptide digests for amino acid sequence and, (iii) measurement of major modifications such as oxidation and deamidation. Mass spectrometry can also be used as a screen to confirm the presence of the most abundant glycan structures. It is possible to use mass spectrometry to obtain more detailed structural information on the glycans, but the experiments are difficult and a high level of expertise is required. The orthogonal approach using exoglycosidase sequencing and HILIC-UPLC on released glycans gives more accurate information on the structures and relative abundance of the individual glycan species present. Neither technique is sufficient to fully characterise the whole antibody: it is the combination of both MS and HILIC-UPLC approaches that provides the most accurate information.

References

1. Gaza-Bulseco G, Faldu S, Hurkmans K, Chumsae C, Liu H. Effect of methionine oxidation of a recombinant monoclonal antibody on the binding affinity to protein A and protein G. J Chromatogr B Analyt Technol Biomed Life Sci. 2008 Jul 1;870(1):55-62..
2. Timm V, Gruber P, Wasiliu M, Lindhofer H, Chelius D. Identification and characterization of oxidation and deamidation sites in monoclonal rat/mouse hybrid antibodies. J Chromatogr B Analyt Technol Biomed Life Sci. 2010 Mar 15;878(9-10):777-84.
3. Ha S, Ou Y, Vlasak J, Li Y, Wang S, Vo K, Du Y, Mach A, Fang Y, Zhang N. Isolation and characterization of IgG1 with asymmetrical Fc glycosylation. Glycobiology. 2011 Aug;21(8):1087-96.
4. Büttel IC1, Chamberlain P, Chowers Y, Ehmann F, Greinacher A, Jefferis R, Kramer D, Kropshofer H, Lloyd P, Lubiniecki A, Krause R, Mire-Sluis A, Platts-Mills T, Ragheb JA, Reipert BM, Schellekens H, Seitz R, Stas P, Subramanyam M, Thorpe R, Trouvin JH, Weise M, Windisch J, Schneider CK . Taking immunogenicity assessment of therapeutic proteins to the next level. Biologicals. 2011 Mar;39(2):100-9.
5. ICH Harmonised Tripartite Guideline, Specifications: Test procedures and acceptance criteria for biotechnological/biological products Q6B, 1999.
6. Zhang Z, Pan H, Chen X. Mass spectrometry for structural characterization of therapeutic antibodies. Mass Spectrom Rev. 2009 Jan-Feb;28(1):147-76.
7. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. GlycoBase and autoGU: tools for HPLC-based glycan analysis.Bioinformatics. Bioinformatics. 2008 May 1;24(9):1214-6.
8. L. Royle, C.M. Radcliffe, R.A. Dwek, P.M. Rudd (2006) Detailed Structural Analysis of N-Glycans Released From Glycoproteins in SDS-PAGE Gel Bands Using HPLC Combined With Exoglycosidase Array Digestion. Methods in Molecular Biology 347: 125-144.

Reading Scientific Services Ltd Tel: +44 (0) 118 918 4076 Email: enquiries@rssl.com

Ludger Ltd Tel: +44 (0) 870 085 7011 Email: info@ludger.com