Biopharmaceutical erythropoietin characterisation: critical quality attribute (CQA) mapping using LC-ESI-Qtof and automated database searching of glycan and glycopeptide analytes

Daniel I R Spencer[1]; Julia Smith[2]; Richard Gardner[1]; Bob Galvin[2]; Anja Resemann[3]; Peter Hufnagel[3]; and Daryl L Fernandes[1]

[1] *Ludger Ltd, Culham Science Centre, Oxfordshire, OX14 3EB, United Kingdom*

[2] *Bruker Ltd. Coventry, Coventry, UK*

[3] *Bruker Daltonic GmbH, Bremen, Germany*

GLYCAN WORKFLOW

UHPLC Peak ID	% Area	ESI-LC/MS Composition: Possible structure	Example Glycan surface	Hex (H)	HexNAc (N)	Fuc (F)	NeuAc (S) 2-3 Da	NeuAc (S) 7-9 Da	potential sulphate or phosphate
1	0.16	Man6-P		4	2	0	0	0	1
2	0.45	FMan5-P		4	2	1	0	0	1
3	1.23	Man7-P		5	2	0	0	0	1
4	0.15	Man8-P		5	2	0	0	0	1
		FA2G2S2(6)2		5	4	1	1	0	0
5	0.79	Man9-P		5	2	0	0	0	1
		FMan9-P		5	2	1	0	0	1
6	0.47	FMan8-P		5	2	1	0	0	1
		FA2G2S2(6)Ac1		5	4	1	1	0	0
7	1.67	Man8-P		6	2	0	0	0	1
8,9	1.16	FA2G2S1		5	4	1	0	0	0
		Man9-P		6	2	0	0	0	1
10,11	0.82	FMan9-P		6	2	1	0	0	1
		Man9-P		6	2	0	0	0	1
12	3.05	FA2G2S2		5	4	1	2	0	0
13	0.43	FA2G2S1Ac(6)2		5	4	1	1	0	1
		FA2G2S2Ac(6)2		6	5	1	1	0	0
		FA2G2S2		5	4	1	2	0	0
		FA2G2S1Ac(6)2		6	5	1	3	0	1
14,15	0.46	FA2G2S2		5	4	1	2	0	0
16	0.94	FA2G2S1		6	5	1	1	0	0
17	0.54	FA2G2S1		6	5	1	1	0	0
		FA2G4S2Ac(2)		7	6	1	1	2	0
18	1.96	FA2G2S2		6	5	1	2	0	0
		A2G2S2Ac(4)		6	5	0	0	2	0
19	1.82	FA2G2S2		6	5	1	2	0	0
		A2G2S2Ac(4)		6	5	0	0	2	0
		FA4G4S2Ac(2)		7	6	1	1	2	0
		FA4G4S4Ac(2)		7	6	1	2	2	0
20	5.32	FA2G2S1		6	5	1	3	0	0
21	1.76	FA4G4S1 or FA4G3S1Ac4Ac(1)		7	6	1	1	0	0
		FA2G2S3		6	5	1	3	0	0
		FA4G4S4Ac(2)		7	6	1	2	2	0
		FA4G4S3Ac		7	6	1	2	1	0
22	6.19	FA4G4S2 or FA4G3S2(2,6)Ac(1)		7	6	1	2	0	0
		FA4G4S4		7	6	1	3	1	0
23	11.24	FA4G4S3 or FA4G3S3(2,6)Ac(1)		7	6	1	3	0	0
		FA4G4S4Ac		7	6	1	3	1	0
24	12.96	FA4G4S4		7	6	1	4	0	0
25,26	2.78	FA4G4S5(2,6)Ac(1)		8	7	1	2	0	0
27	9.83	FA4G4S5(2,6)Ac(1)		8	7	1	3	0	0
28	1.31	FA4G4S5(2,6)Ac(1)		8	7	1	3	0	0
		FA4G4S5Ac(2,6)Ac(1)		8	7	1	1	2	0
29	13.67	FA4G4S6(2,6)Ac(1)		7	6	1	4	0	0
30	1.76	FA4G4S5(2,6)Ac(2)		7	6	1	2	0	0
31	0.50	FA4G4S5(2,6)Ac(2)		9	8	1	3	0	0
32	0.26	FA4G4S6(2,6)Ac(2)		9	8	1	4	0	0
34	1.33	FA4G4S5(2,6)Ac(3)		10	9	1	3	0	0
35	1.72	FA4G4S6(2,6)Ac(3)		10	9	1	4	0	0

Table 1: Glycans manually identified as present on our recombinant human erythropoietin expressed in Chinese hamster ovary cells. Glycans also identified automatically using our EPO database in the Bruker ProteinScape 4.0 software are highlighted in the green rows.

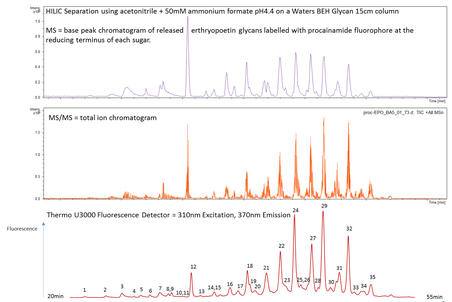

Figure 1: MS, MS/MS and FD traces of procainamide labelled EPO glycans run on a UHPLC HILIC column into Bruker impact II or amaZon Speed ETD ESI-MS. Peak numbers are ID'ed in Table 1.

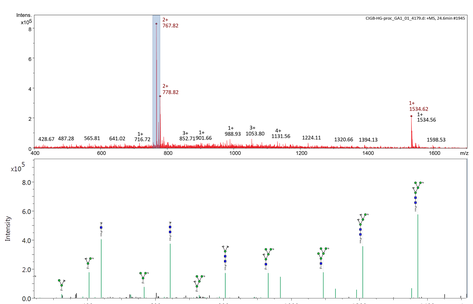

Figure 2: MS and an MS/MS of glycan mass 767.82 m/z using the Bruker amaZon Speed ETD – good fragmentation characteristics but mass accuracy not sufficient to determine whether the glycan was sulphated (767.7874 m/z) or phosphorylated (767.7922 m/z)

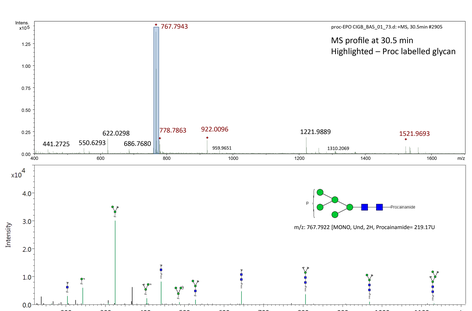

Figure 3: MS and an MS/MS of glycan mass 767.7943 m/z using the Bruker impact II QTOF – this instrument, designed for sub 1ppm mass accuracy identified the glycan to have the mass corresponding to a MAN5 glycan with a single phosphorylation. The glycan was identified using our EPO database in the Bruker ProteinScape 4.0 software.

Introduction

Glycosylation greatly influences the clinical performance and safety of recombinant erythropoietin (EPO) drugs. Consequently, biopharma companies producing EPO products must carefully optimise, accurately measure and tightly control glycosylation throughout the production lifetime of their drug.

Clinically relevant glycosylation metrics relating to EPO include the following:

1. The overall degree of drug sialylation and acetylation modifications

2. The topologies (i.e. branching) of the sialic acid bearing glycans

3. The distribution of sialylated glycans over the various glycosylation sites

Here, two workflows are presented for EPO glycan and glycopeptide mapping where LC-ESI-QTof-MS or LC-ESI-QIT-MS instruments, standard flow HILIC + fluorescence detection or wide pore nano-C18 were coupled with flexible software program for glycan or glycopeptide characterisation using a user defined database of potential glycans.

Methods

Erythropoietin or IgG sample preparation:

Glycan Analysis

Erythropoietin glycans were released from denatured protein using PNGAseF, desalted (LC-PBM-01, Ludger UK) and fluorophore labelled by reductive amination with procainamide (LC-PROC-VP24), a fluorophore that has enhanced MS signal over alternative labels[1]. Sample preparation was performed using Ludger's Longbow automated sample preparation system utilizing a Hamilton Starlet robot[2].One fluorophore label per glycan ensures relative abundances of glycans can be measured using LC-fluorescence detection. Samples were separated on a Water BEH-glycan LC column (15cm) using a Thermo U3000 SD system with a fluorescence detector (FD) coupled to a Bruker amaZon Speed ETD MS, or by LC without FD on a Bruker impact II Qtof where higher mass accuracy was required (in order to determine unambiguously phosphorylated from sulphated glycans). Glycan identification was performed both manually and automatically. Manually using retention time characteristics coupled with MS and MS/MS data, and automatically using MS/MS data coupled with the Bruker ProteinScape 4.0 GlycoQuest software against our in-house EPO glycan database.

Glycopeptide Analysis

In order to generate glycopeptides with single glycosylation sites, clinical grade recombinant EPO samples were subjected to reduction and alkylation followed by overnight trypsin digestion, C18 spin column enrichment and then a subsequent overnight Glu-C proteolysis in PBS solvent.

Impact II/ProteinScape methods:

Glycopeptides were separated by an acetonitrile gradient on a Pepmap300 Å C18, 5 µm, 75 µm x 150 mm column. Spectra were acquired using a modified standard Instant Expertise method[3] on a Bruker impact II ESI QTOF, with a fixed MS duty cycle of 3.5 s at 2 Hz and variable MSMS at 1.5 – 4 Hz depending upon precursor intensity. Peak lists were generated in DataAnalysis 4.3 and glycopeptide spectra identified in ProteinScape 4.0.

Data and Conclusions

Glycan Workflow

The combination of a sensitive glycan fluorophore preparation method that was compatible with high-throughput workflows, HILIC-FD-LC and amaZon or impact II MS resulted in the automatic assignment of the majority of the glycans previously identified by manually methods (Table 1 – green rows – ProteinScape 4.0 identified glycans). This method can be readily adopted for the monitoring of biopharmaceutical CQAs for EPO to a GMP standard.

Glycopeptide Workflow

Unlike with the smaller pore C18 columns regularly utilised in nano-lc peptide mapping methods, the wide pore C18 chromatography used here allowed for more sensitive glycopeptide analysis. The 3D spatial volume of glycopeptides is often large, particularly with the highly branched, sialylated and extended glycans regularly seen in EPO products. The wide pores (300 Å) in the column particles enabled the entire chromatographic particle volume to be accessible to the analytes and resulted in more highly resolved elution profiles when compared to standard 75-100 Å pore columns. Using polyclonal human IgG digested with trypsin we were able to identify all 4 IgG subclasses of the Fc glycopeptide constant regions (peptide sequence EEQYNSTYR – with F substitutions for Y in the different IgG subclasses). These subclass glycopeptides appear to separate nicely on the wide pore C18 column.

From the recombinant erythropoietin we identified N-glycosylation sites but it was evident that the age of the sample was coupled to endoprotease activity resulting in many unpredicted peptide sequence cleavage points. Nevertheless with the LC-MS/ProteinScape 4.0 workflow we were able to identify N-linked sites with varying degrees of glycan sialylation, poly-lactosamine extension and acetylation which were key CQAs for EPO even where the peptide cleavage points could not be predicted.

GLYCOPEPTIDE WORKFLOW

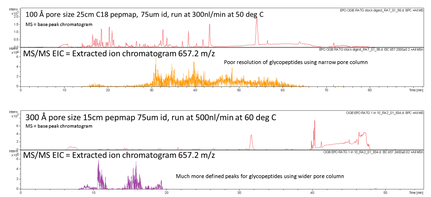

Figure 4: Comparison of EPO glycopeptide separation on a standard 25 cm nano C18 100Å pore size column (top panel) and on a 15 cm nano C18 wide pore 300A column (bottom panel) . Better glycopeptide resolution gained was with the wide pore column despite being shorter. It is frequently observed that larger compex glycans and glycopeptides appear to occupy larger spatial volumes than their mass alone implies. An EIC of 657.2m/z identifies glycans terminating in HexNac-Hex-NeuAc structures.

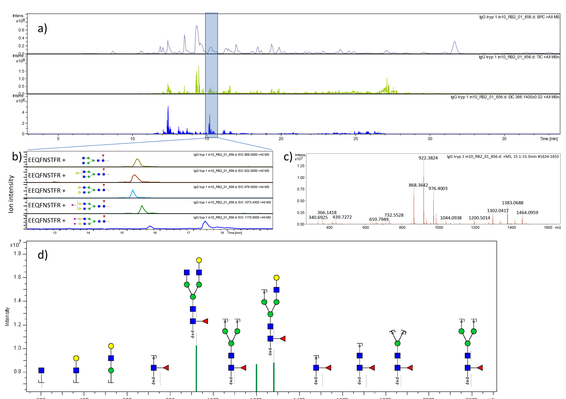

Figure 5: Human polyclonal IgG glycopeptide chromatography on the wide pore nanoC18 column and analysis on impact II MS, a) BPC of sample, TIC and EIC for glycan fragments (366.2 mu – GlcNac+Gal fragments) , b) EIC showing the IgG subclass 2 constant region glycopeptides with different glycans attached, c) MS profile of neutral glycan EEQFNSTFR glycopeptides, d) Proteinscape 4.0 automatic identification and labelling of EEQFNSTFR glycopeptide + FA2G1 glycan (G1F)

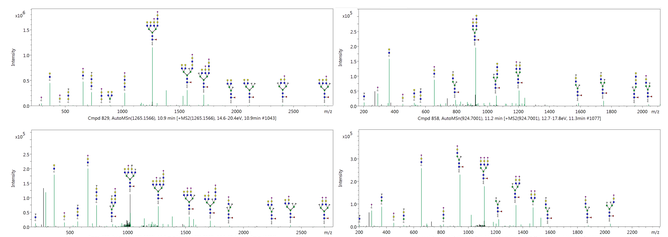

Figure 6: MS/MS profiles of four different EPO glycopeptides, same peptide backbone of NIT but with four different glycans. EPO functionality is partially tied to the type of glycans present on the protein, with large (in excess of 4000 mu), polylactosamine containing glycans considered advantageous. By adjusting glycan fragmentation parameters within the impact II settings we were able to correctly assign large glycan structure containing glycopeptides.

Future Perspectives

We feel that the glycan workflow is optimized for GMP grade analyses. Integration of the fluorescence data, providing relative quantitation, in combination with the MS data for added specificity and confidence will lead to even more powerful and automatable workflows.

For the glycopeptide workflow, fresh samples run on a longer wide pore C18 gradient will lead to more glycopeptide identifications. The very large glycans typical of EPO required non-standard fragmentation conditions and we need to optimise the fragmentation method in order to obtain a good balance of large glycan fragment identification as well as peptide backbone sequencing in order to provide the ProteinScape 4.0 software with more information for glycopeptide identification.

Acknowledgements

High throughput workflows for glycan preparation, and glycan identification was performed with the assistance of a European Union FP7 grant: HighGlycan grant no.278535. The human EPO was a kind gift from Raquel Montesino and Antonio Vallin from the Center for Genetic Engineering and Biotechnology, Cuba.

References

1. Kozak, R. P.; Tortosa, C. B.; Fernandes, D. L.; Spencer, D. I. R. *Anal. Biochem.* **2015**, 486, 38–40.
2. Ventham, N. T.; Gardner, R. a.; Kennedy, N. a.; Shubhakar, A.; Kalla, R.; Nimmo, E. R.; Fernandes, D. L.; Satsangi, J.; and Spencer, D. I. R. *PLoS One.* **2015**, 10, e0123028.
3. Bruker Poster Note PN-03 **Using multiple collision energies to improve N- and O-glycopeptide identification by LC-MS/MS**.

Contact

For more information please contact Daniel Spencer - daniel.spencer@ludger.com
You can also visit www.ludger.com and www.highglycan.eu