Comprehensive Analysis of Recombinant Human Erythropoietin Glycoforms by Capillary Electrophoresis and Nanoflow Liquid Chromatography Coupled with Intact/Middle-Down Mass Spectrometry

Sensitive and high-resolution glycoform profiling of intact rhEPO using CESI-MS with the Neutral OptiMS cartridge

M. Santos,[1] R. Viner,[2] D. Horn,[2] J. Saba,[2] M. Bern,[2] D. Bush,[4] H. Dewald,[1] A.A.M. Heemskerk,[1] B.L. Karger[3] and A.R. Ivanov[3]

1 SCIEX, Brea, CA
2 Thermo Fisher Scientific, San Jose, CA
3 Northeastern University, Boston, MA
4 Genedata, Lexington, MA

Introduction

Erythropoietin (EPO) is a naturally occurring red blood cell stimulating hormone produced in the kidney, and it was one of the first therapeutic recombinant proteins. EPO is a heavily glycosylated protein with a molecular weight of approximately 30 kDa, and 40% of its weight is due to glycosylation.1 It contains 3 N-glycosylation sites on Asn24, Asn38 and Asn83, and 1 O-glycosylation site on Ser126, which makes glycoform profiling challenging due to the high heterogeneity (Figure 1). Recombinant human EPO (rhEPO) has been studied extensively using different approaches, including capillary electrophoresis and nanoflow liquid chromatography coupled to advanced mass spectrometry (MS) detection.[2-3] However, these analytical approaches frequently involve some type of enzymatic digestion, which can induce artifacts. Therefore, the characterization of these biomolecules under intact conditions is important. Capillary electrophoresis is a well-established analytical tool for the analysis of intact native proteins. The recent development of CESI, which integrates capillary electrophoresis (CE) and electrospray ionization (ESI) into a single process within the same device, enabled the characterization of the glycosylation of intact EPO at pH2.[1] In this work, we show that by using CESI-MS and a neutral coated capillary coupled to high-resolution, high mass accuracy mass spectrometry, we were able to identify and quantify multiple glycoforms of rhEPO under native conditions. Additionally, we used a middle-down approach to assess site-specific N- and O-glycosylation.

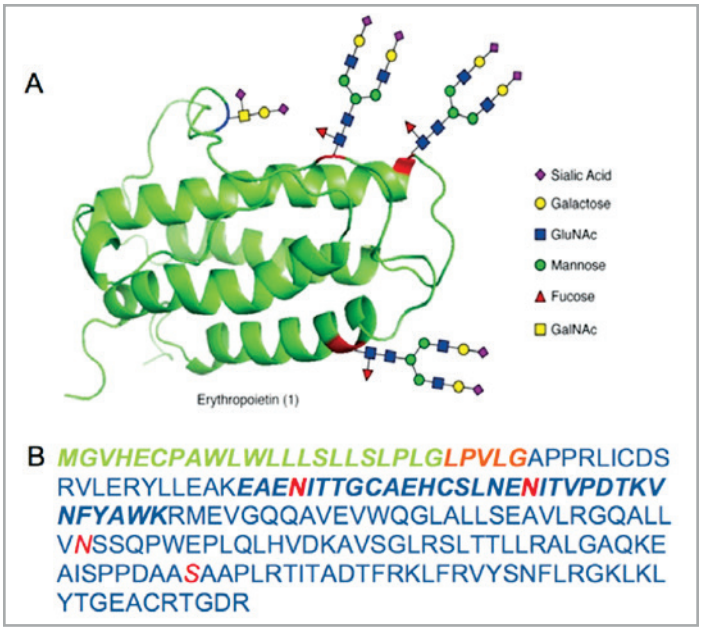

Figure 1. Erythropoietin amino acid sequence and glycosylation sites.

Materials and methods

Sample preparation

Reduced and alkylated rhEPO expressed in CHO or HEK cells (Erythropoietin-Alpha, ProSpec, NJ) was digested with LysC (enzyme:protein ratio of 1:200, for 2 hr at 37 °C in 20 mM ammonium acetate, pH 6.0); trypsin (enzyme:protein ratio of 1:100, for 4 hr at 37 °C in 20 mM ammonium bicarbonate, pH 8.0) or proteinase K (enzyme:protein ratio of 1:50, for 1 hr at 37 °C in 20 mM ammonium acetate, pH 6.0). All enzymes were from Roche, IN.

Liquid chromatography

The rhEPO tryptic or LysC digests were separated using the Thermo Scientific EASY-nLC 1000 HPLC system with a Magic C18 spray tip 20 cm x 75 µm I.D. column (Michrom). Gradient elution was performed from 4–30% over 60 min and from 30–85% over 10 min with ACN in 0.1% formic acid at a flow rate of 300 nL/min.

CESI-MS conditions:

Intact rhEPO was separated using a CESI 8000 Plus High Performance Separation-ESI Module (SCIEX) equipped with a Neutral OptiMS cartridge consisting of a porous sprayer operating in an ultra-low flow regime, and detected using an Exactive Plus EMR (Thermo Fisher Scientific) at 35 K or 140 K FWHM resolution at m/z 200. The middle-down experiments were carried out using the OptiMS Silica Surface cartridge on a Thermo Scientific Orbitrap Elite mass spectrometer and Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer using FT/IT HCD, CID or ETD MS2 fragmentations in DDDT or HCDpdETD/CID methods. FT MS1 was acquired at resolution settings of 60–120 K at m/z 200 and FTMS2 at resolution of 30–60 K at m/z 200.

Important:

- A separation current above 5 µA might cause permanent damage to the separation capillary.
- Generally, please do not apply >2000V to generate electrospray as it may result in capillary damage.

Data analysis

The Thermo Scientific ProSightPC 3.0, Protein Deconvolution 3.0, Pinpoint 1.4, and Proteome Discoverer 2.0 software with the Byonic search node (Protein Metrics) were used for glycopeptide data analysis and glycoform quantification. SimGlycan 4.5 software (PREMIER Biosoft) was used for proteinase K digest glycopeptide and glycan composition identification.

Results

Reproducibility of middle-down analysis

CESI-MS is a sensitive and reproducible analytical technique employed for the separation of glycopeptides due to differences in charge states and Stokes radii.[4] We were able to identify and quantify multiple EPO glycopeptides using 200 ng of sample with excellent S/N. Migration times and peak areas demonstrated good reproducibility with less than 10% RSD across runs (Figure 2). Glycopeptides were well separated by CESI and resolved within 20 min of a 50 min long run (Figure 2).

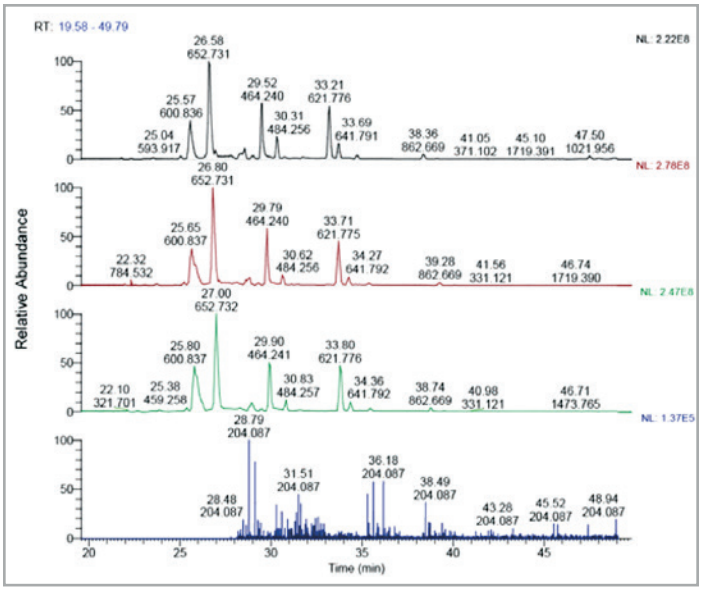

Figure 2. Base peak electropherograms for three constitutive runs of rhEPO LysC digest (200 ng) and MS2 XIC for HexNac oxonium peak.

Characterization of O-glycoforms

Limited LysC digest yielded peptides of different lengths containing Ser126 (where the O-glycan site is located) in a range of 3–9 kDa (Figure 3). All O-linked glycopeptides can only be identified in the FT ETD experiments because the O-linked glycans are very labile and do not survive collisional activation (Figure 3). As expected, CESI glycoform separation was mostly based on differences in the number of sialic acid residues (Figure 4). The 2 predominant O-glycosylated peptides (N-acetylhexosamine-hexose with 1 or 2 sialic acids) migrated as baseline resolved peaks in CESI but not in nLC (Figure 4A vs. Figure 4B) using similar analysis times. The relative abundances of major rhEPO O-glycoforms are shown in Table 1. We detected several unmodified Ser126 peptides (total relative abundance 9%, Table 1), which means that the site was only partially glycosylated in this sample. Additionally, we observed partial O-glycosylation on Ser9 and Ser120 residues.

Looking at the N-glycosylation of EPO, we found the most abundant glycopeptide containing Asn24 and Asn38 was A1–K45 with sum glycan composition for both sites of $HexNAc_{12}Hex_{14}dHex_2NeuAc_8$ (Figure 5 and Table 2). We were not able to unambiguously assign the glycan composition for

each individual site. For this reason, Table 2 presents the glycan composition as a sum of different glycan composition for both sites on this particularly large peptide. The most abundant glycopeptide containing the Asn83 site was R53–K97 (Figure 6). It is worth noting that the difference of a single sialic acid causes a small migration time shift (~0.4 min) due to a change in mobility of the peptide as a result of a change in the overall negative charge. To a lesser extent, differences in the number of HexHexNAc residues also affect the mobility due to a possible change in the Stokes radii, but not on the overall charge of the peptide as this species is neutral (Figure 7). As expected,[5] the main glycan compositions for Asn83 site were tetra-acidic oligosaccharides (Figure 8).

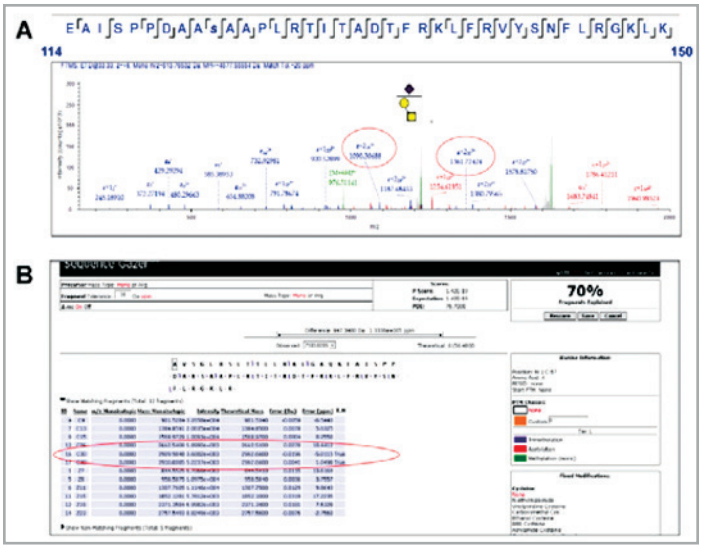

Figure 3. Identification of rhEPO LysC O-linked glycopeptides by Byonic node in Proteome Discoverer 2.0 software (114–150, A) or by ProSightPC 3.0 software using biomarker and delta m search (95–150, B).

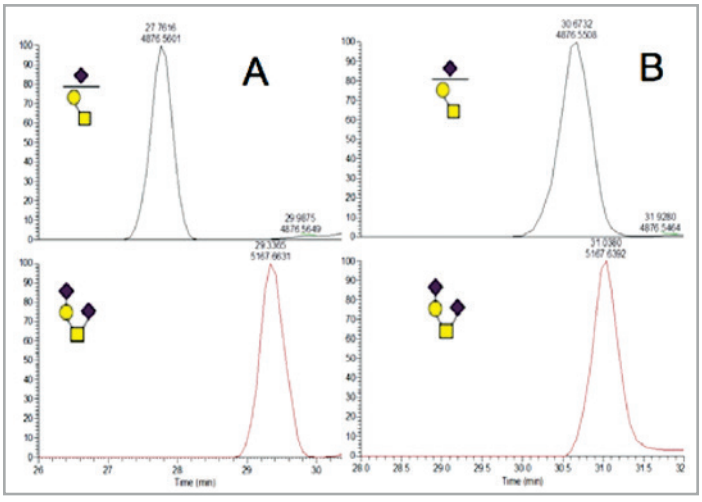

Figure 4. CESI-MS (A) and nLC-MS (B) separation of O-linked E_{114}–K_{150} glycopeptide.

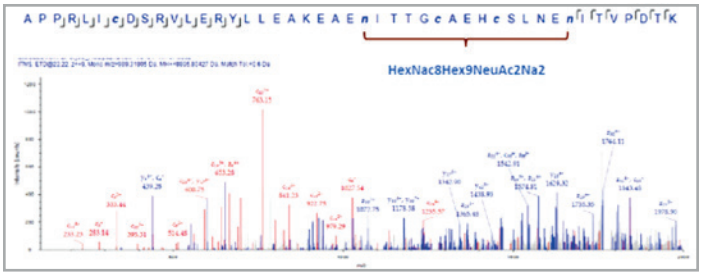

Figure 5. Identification of rhEPO LysC double N-linked glycopeptide (A_1–K_{45}) by Byonic node in Proteome Discoverer 2.0 software.

Glycoform	Relative abundance (%)
	42.5
	37
	6.5
	4
—	9

Table 1. Peptide quantification of major rhEPO glycoforms for Ser126 site. Each glycoform was calculated as a sum of all detected peptides.

Glycoform	Relative abundance (%)
HexNAc12Hex14dHex2NeuAc8	35
HexNAc13Hex15dHex2NeuAc8	18
HexNAc14Hex16dHex2NeuAc8	17
HexNAc12Hex14dHex2NeuAc7	17
HexNAc6Hex7dHex1NeuAc4	12

Table 2. Peptide quantification of major rhEPO glycoforms for Asn23 and Asn28 sites. Each glycoform was calculated as a sum of all detected peptides, including acetylated and sodiated species.

We also performed a comparative glycopeptide profiling of this site for rhEPO expressed in both CHO and HEK cells (Figure 8). It has been reported that oligosaccharide structural features of recombinant proteins are specific to cell line, culture condition and species.[6] Assuming equal detection response for all glycoforms, CESI-MS analysis demonstrated significant differences in relative abundance of glycoforms expressed in CHO vs. HEK cells. The main glycoform for this site was $HexNAc_6Hex_7dHex_1NeuAc_4$ in HEK cells, and for CHO cells several larger tetra-sialylated species offset by HexNacHex units were dominant.

Characterization of rhEPO under native conditions.

For the analysis of rhEPO under native conditions, we used a neutral coated capillary, which prevented the protein from sticking to the capillary surface, enabling a better separation of the isoforms. The CESI 8000 Plus System was coupled to an Exactive Plus EMR system at different resolution settings. Figure 9A shows a separation of EPO using 40 mM ammonium acetate (pH 7.5) as the background electrolyte and a corresponding ion density map on Figure 9B. The analysis of the deconvoluted spectra (Figure 10B) of each peak (Figure10A) revealed the presence of 167 proteoforms at a resolution of 140 K (at m/z 200) and 428 glycoforms at a resolution of 35 K (at m/z 200) with only 2.25 ng of sample injected in both cases, with the main form being $HexNAc_{19}Hex_{22}dHex_3NeuAc_{13}$ migrating in peak 4 at a resolution of 35 K. These results also show that the different glycoforms of rhEPO separate mostly based on differences in the number of sialic acid residues (Figure 10) correlating well with results from the middle-down approach (Table 3). Due to the high heterogeneity, the separation is not baseline resolved. However, in combination with HRAM-MS, the CESI-MS of native rhEPO still provided better dynamic range and more accurate quantification than direct infusion.

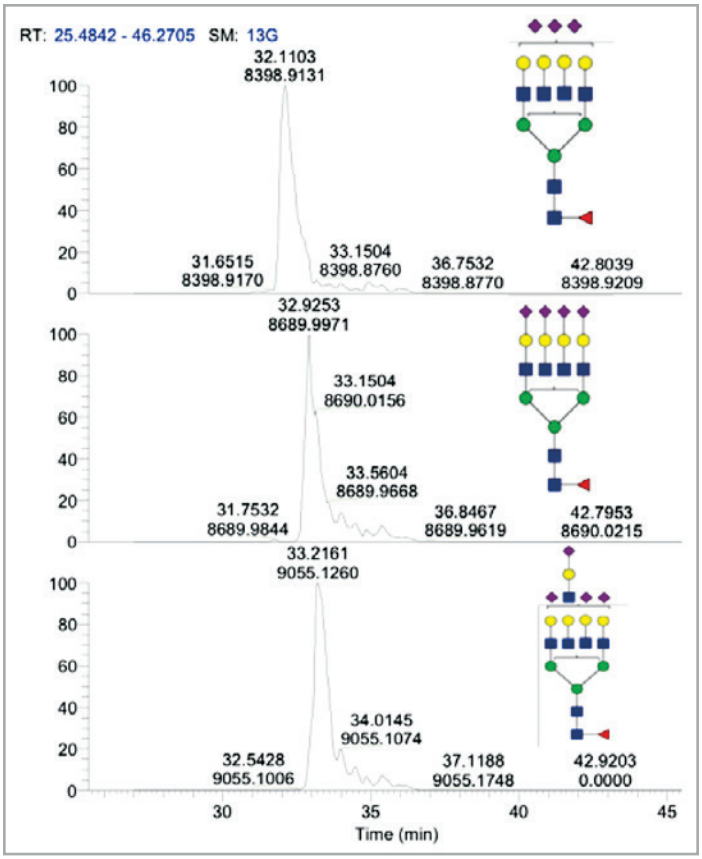

Figure 7. CESI-MS separation of Asn83 glycoforms (R53–K97).

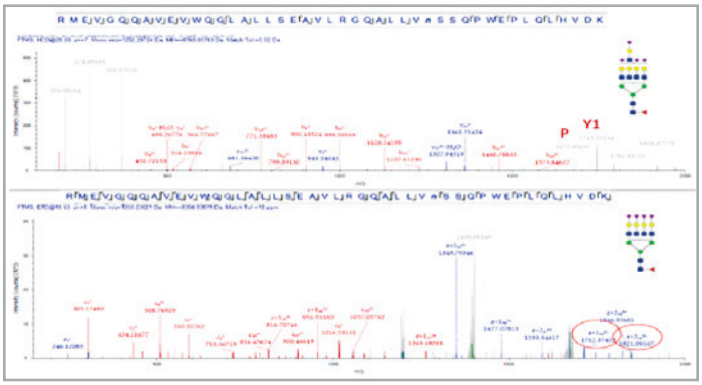

Figure 6. Identification of rhEPO LysC N-linked glycopeptide (R_{53}–K_{97}) by Byonic node in Proteome Discoverer 2.0 software using HCD (A) or ETD (B) fragmentation.

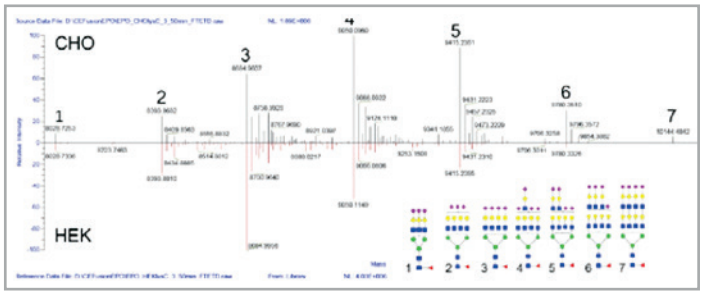

Figure 8. Comparison of CHO and HEK rhEPO Asn83 N-glycoforms using the Xtract deconvolution algorithm in Protein Deconvolution 3.0 software.

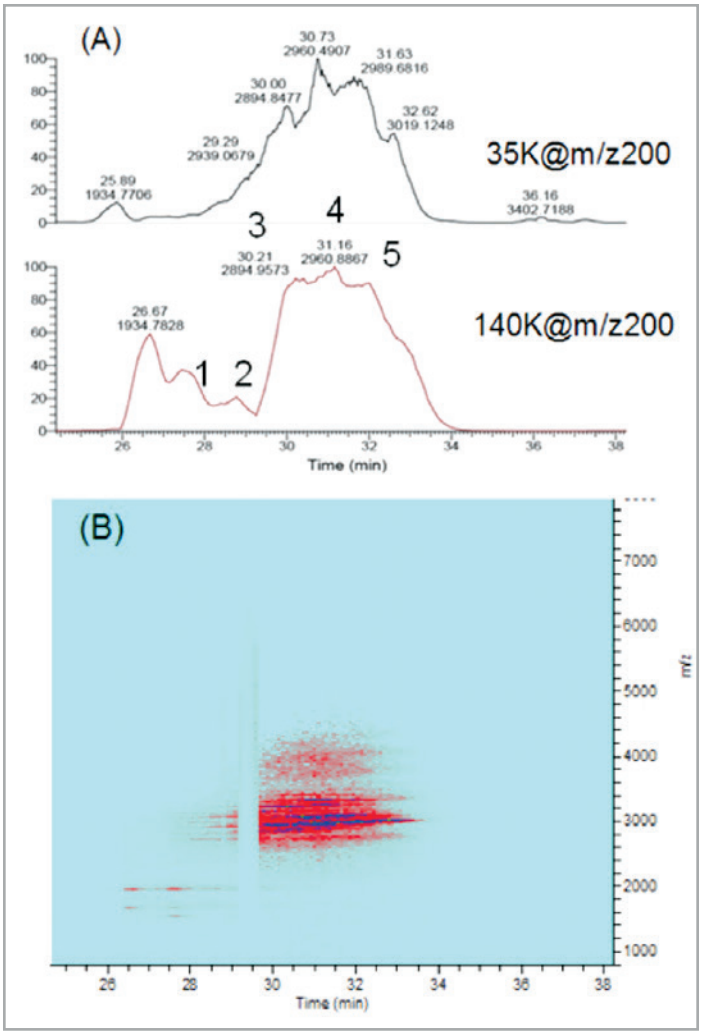

Figure 9. Base ion electropherogram of EPO under native conditions (A) and corresponding ion density map (B).

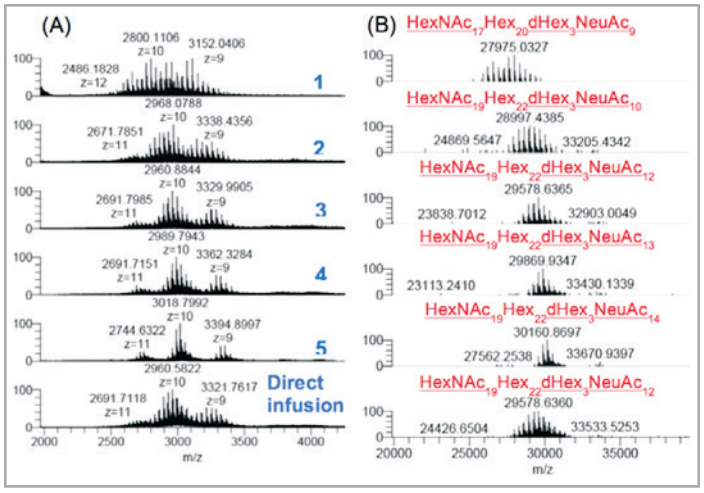

Figure 10. EPO profiling under native conditions, pH 7.5 (A) and corresponding deconvoluted spectra (B) at 140 K @ m/z 200 resolution.

EPO	Relative abund-ance,%	Ser 126	Relative abund-ance,%	Asn 23/38	Relative abund-ance,%	Asn 83	Relative abund-ance,%
HexNAc19Hex22dHex-3NeuAc13	44.79	HexNAcHexNeuAc	42.5	HexNAc12Hex14dHex-2NeuAc8	35	HexNAc6Hex7dHex-1NeuAc4	46.7
HexNAc19Hex22dHex-3NeuAc14	18	HexNAcHexNeuAc2	37	HexNAc13Hex15dHex-2NeuAc8	18	HexNAc7Hex8dHex-1NeuAc4	28
HexNAc19Hex22dHex-3NeuAc12	17	HexNAc2NeuAc	6.5	HexNAc12Hex14dHex-2NeuAc7	17	HexNAc6Hex7dHex-1NeuAc3	14
HexNAc19Hex22dHex-3NeuAc10	17	HexNAc2Hex	4	HexNAc13Hex15dHex-2NeuAc8	17	HexNAc5Hex7dHex-1NeuAc3	11.6
HexNAc6Hex7dHex-1NeuAc4	**12**	**—**	**9**	**HexNAc6Hex7dHex-1NeuAc4**	**12**	**HexNAc6Hex7dHex-1NeuAc3**	**4.7**

Table 3. Relative quantification of major hEPO glycoforms using intact and middle-down approaches. Each glycoform was calculated as a sum of all detected, including acetylated and sodiated species.

Conclusions

- The CESI-MS technique is reproducible (RSD <10%) and sensitive, obtaining the same sequence coverage and number of glycopeptides with 5 times lower amount of sample than for nLC-MS experiments.
- CESI separation of glycoforms is clearly based on differences in the number of sialic acid residues (that is, difference in charge) and the peptide to glycan mass ratios.
- The primary O-linked glycoforms of CHO rhEPO are HexNacHex+1(2)NeuAc with the relative abundance of unglycosylated Ser126 of approximately 9%.
- Comparative glycoprofiling of Asn83 site for rhEPO expressed in CHO vs. HEK using CESI-HRAM middle-down demonstrates clear differences in glycoform abundances and validates the utility of this approach for in-depth characterization of glycoproteins.
- Comparative glycoprofiling of rhEPO using CESI-HRAM middle-down and intact strategies demonstrates remarkable correlation in glycoform abundances and validates the utility of this combined approach for in-depth characterization of glycoproteins with multiple glycosites.

References

1. Haselberg, R.; de Jong, G.J.; Somsen, G.W. *Anal. Chem.* 2013, 85 (4), 2289–2296.
2. Balaguer, E.; Demelbauer, U.; Pelzing, M.; Sanz-Nebot, V.; Barbosa. J.; Neusüss, C.; *Electrophoresis* 2006, 13, 2638– 2650.
3. Moini, M. *Anal. Chem.* 2007, 79, 4241–4246.
4. Kolarich, D.; Jensen, P.H.; Altmann, F.; Packer, N.H. *Nat. Protoc.* 2012, 7 (7), 1285–1298.
5. Higgins, E. *Glycoconjugate J.* 2010, 27, 211–225.
6. Rush, R.S.; Derby, P.L.; Smith, D.M.; Merry, C.; Rogers, G.; Rohde, M.F.; Katta, V. *Anal. Chem.* 1995, 67, 1442−1452.

SCIEX

Headquarters
500 Old Connecticut Path | Framingham, MA 01701, USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at
sciex.com/offices