CHARACTERIZATION OF GLYCOPROTEINS AND N-GLYCANS OF RECOMBINANT HUMAN ERYTHROPOIETIN BY LIQUID CHROMATOGRAPHY AND MASS SPECTROMETRY

Waters

Mario Cindrić[1], Nick Tomczyk[2], Gerold Schwarz[2], Iain Campuzano[3], Johannes P.C. Vissers[2]
[1]Pliva-Research and Development, Prilaz baruna Filipovića 29, Zagreb, Croatia
[2]Waters Corporation - Almere, The Netherlands, [3]Waters Corporation – Manchester, UK

OBJECTIVES

- The objective of this study was to confirm the amino acid sequence, and the glycan structures, expressed on a recombinant human Erythropoieten (rhEPO)
- In addition, we wished to identify the degree of heterogeneity present within the sample and obtain the relative abundance of the isoforms.

INTRODUCTION

Glycoproteins are biomolecular macromolecules in which carbohydrates are attached covalently to asparagine, serine or threonine residues of protein amino acid backbones. Glycosylation and deglycosylation of mammalian proteins are pivotal for many biological processes; hence, detailed structure elucidation is of major interest. In this study, direct analysis by mass spectrometry mass spectrometry in combination with a dual enzymatic digestion of the macromolecule, allows characterization of the peptide fragments and released glycans [1].
Combining reverse phase chromatography, allowing separation of the peptides and glycans, provides for a more comprehensive characterization of rhEPO [2,3].
In this work we have used both MALDI TOF, with parallel PSD (PSD MX) and nanoscale LC-MS/MS, in both positive and negative ion modes, for the analysis of rhEPO.
High mass accuracy was achieved with NanoLockSpray™ and internal calibration, where applicable, providing reliable structure identification of the rhEPO.

EXPERIMENTAL

MALDI micro MX™ – An axial MALDI-TOF mass spectrometer utilizing stainless steel target plates with the matrices Alpha-Cyano-4-hydroxy cinnamic acid (CHCA) and 2,4,6-trihydrdoxyacetophenone (THAP) used for peptide mass fingerprinting (PMF) and N-glycan analysis respectively.

Q-Tof Premier™ – Equipped with NanoLockSpray source and directly connected to a nanoACQUITY UPLC system for the Peptide DDA and PID analysis. Peptide data was corrected with a solution of [Glu[1]]-Fibrinopeptide B (GFP).

LCT Premier™ - Equipped with LockSpray™ source and directly connected to an ACQUITY UPLC™ system for the Peptide mapping experiment, Data was corrected with [Glu[1]]-Fibrinopeptide B (GFP) solution.

RESULTS AND DISCUSSION

MALDI PMF

The peptide mass fingerprint with mass range 500-3500 m/z (Figure 1) of the trypsin digested rhEPO was obtained and searched against the SwissProt database using ProteinLynx™ Global SERVER. The details of the observed peptides results are shown in Table 1. This analysis shows that we observe all possible tryptic fragments that lie within the acquired mass range.

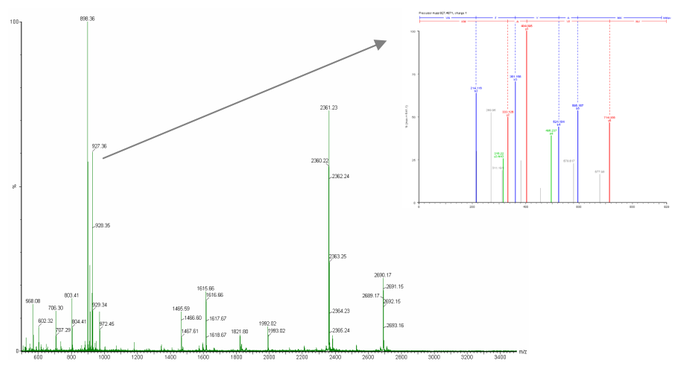

Figure 1. MALDI micro MX PMF of Trypsin digested rhEPO, shown inset is the PSD spectra of peptide 927.3 m/z with assigned fragments

residue #	sequence	fragment #	M	[M+H]	[M+2H]	[M-3H]	[M+H] (MALDI-TOF)	ppm (MALDI-TOF)	[M+H] (MALDI-TOF-PSD)	ppm (MALDI-TOF-PSD)	sequence (MALDI-TOF-PSD)
140-140	K	T15	146.11	147.1128							
53-53	R	T7	174.11	175.1189							
151-152	GK	T18	203.13	204.1342							
153-154	LK	T19	259.19	260.1968							
141-143	LPR	T16	434.26	435.2714							
1-4	APPR	T1	439.25	440.2616							
163-166	TQDR	T21	447.21	448.2150							
11-14	VLER	T3	515.31	516.3140			516.30	-34.09			
111-116	ALGAGK	T12	586.34	587.3511			587.36	12.09			
98-103	AVSGLR	T10	601.35	602.3620			602.37	19.92			
5-10	LKDSR	T2	705.35	706.3552			706.36	5.24	706.38	35.11	
15-20	YLLEAK	T4	735.42	736.4239			736.43	6.38			
104-110	SLTTLLR	T11	802.49	803.4965			803.49	-4.73	803.49	-10.58	SLTTLLR
144-150	VYSNFLR	T17	897.47	898.4781			898.46	-16.69	898.48	0.68	VYSNFLR
155-162	LYTGEACR	T20	911.42	912.4244			912.41	-15.78			
132-139	TTFACPTR	T14	923.47	924.4705			924.46	-17.74			
49-52	VNFYAWK	T6	926.47	927.4723			927.46	-18.11	927.49	15.42	VNFYAWK
117-131	EAISPPDAASAAPLR	T13	1484.76	1485.7645			1485.75	-6.82	1485.79	17.40	EAISPPDAASAAPLR
77-97	GQALLVNESQFWEPLQLHVDK	T9	2358.23	2360.2244			2360.24	8.91	2360.30	32.03	GQALLVNESQFWEPLQLHVDK
54-76	MEVGQGAVEYWGGLALLSEAVLR	T8	2525.33	2526.3384			2526.35	3.96	2526.33	-3.32	
21-46	EAENITTGCAEHCSLNENITVPDTK	T5	2688.20	2689.2079			2689.17	-15.02	2689.05	-58.72	EAENITTGCAEHCSLNENITVPDTK
							RMS (ppm)	15.36		26.22	

Table 1. Summary of MALDI Micro MX PMF and PSD results

LCMS PMF

An LCMS run was conducted using 50 fmol rhEPO on the Q-Tof Premier and nanoACQUITY UPLC using the novel trapping scheme . A 75 µm C18 BEH nanoACQUITY column was used and spectra acquired in the range 400-1500 m/z. The LCMS chromatogram is shown in Figure 2. Inset is an example mass measured peptide peak showing a mass error of 0.34 ppm. This experiment was repeated with the LCT premier and ACQUITY UPLC on a 1mm id C18 BEH ACQUITY column at 150 uL/min and operating at a pressure of >10,000 psi. with a total load of 1 pmol rhEPO. This analysis provided very comparable results (data not shown) however additional species were observed due to direct loading onto analytical column where two hydrophilic peptides (T2 and T20) were lost using the trapping scheme.

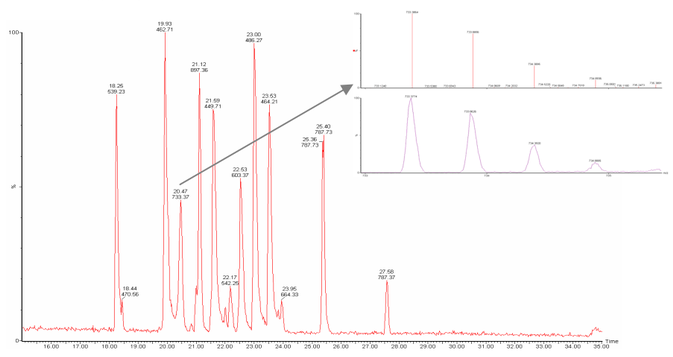

Figure 2. LCMS peptide mapping, Q-Tof Premier. 50 fmol rhEPO on column with a 20 min 5%-60% acetonitrile gradient.

residue #	sequence	fragment #	M	[M+H]	[M+2H]	[M-3H]	[M+H] (Q-TOF)	ppm (Q-TOF)	[M+3H] (Q-TOF)	ppm (Q-TOF)	[M+3H] (Q-TOF)	ppm (Q-TOF)
140-140	K	T15	146.11	147.1128	74.0603	49.7095						
53-53	R	T7	174.11	175.1189	88.0634	59.0446						
151-152	GK	T18	203.13	204.1342	102.5710	68.7166						
153-154	LK	T19	259.19	260.1968	130.6023	87.4041						
141-143	LPR	T16	434.26	435.2714	218.1396	145.7523						
1-4	APPR	T1	439.25	440.2616	220.6347	147.4257						
163-166	TQDR	T21	447.21	448.2150	224.6114	150.0759						
11-14	VLER	T3	515.31	516.3140	258.6609	172.7765	516.3115	-4.8424				
111-116	ALGAGK	T12	586.34	587.3511	294.1795	196.4556	587.3508	-0.5108				
98-103	AVSGLR	T10	601.35	602.3620	301.6849	201.4592	602.3723	17.0996				
5-10	LKDSR	T2	705.35	706.3552	353.6815	236.1238						
15-20	YLLEAK	T4	735.42	736.4239	368.7159	246.1465	736.4235	-0.5432				
104-110	SLTTLLR	T11	802.49	803.4965	402.2532	268.5047	803.5051	3.2358	402.2519	-3.1075		
144-150	VYSNFLR	T17	897.47	898.4781	449.7430	300.1646	898.4805	2.6712	449.7403	-5.9923		
155-162	LYTGEACR	T20	911.42	912.4244	456.7181	304.8133						
132-139	TTFACPTR	T14	923.47	924.4705	462.7432	308.8314	924.4897	12.1149	462.7456	5.2945		
46-52	VNFYAWK	T6	926.47	927.4723	464.2401	309.8293	927.4720	-0.3235	464.2379	-4.6312		
117-131	EAISPPDAASAAPLR	T13	1484.76	1485.7645	743.3862	495.2600			**743.3864**	**0.3489**		
77-97	GQALLVNESQFWEPLQLHVDK	T9	2358.23	2360.2244	1180.6161	787.4133			1180.6136	-2.1175	787.4114	-2.4553
54-76	MEVGQGAVEYWGGLALLSEAVLR	T8	2525.33	2526.3384	1263.6731	842.7847					842.7829	-1.9589
21-46	EAENITTGCAEHCSLNENITVPDTK	T5	2688.20	2689.2079	1345.1079	897.0745					897.0594	-17.9472
							RMS (ppm)	7.7530	RMS (ppm)	4.9588	RMS (ppm)	10.5133

Table 2. Summary of LCMS peptide mapping on Q-Tof Premier

Peptide MSMS fragmentation

MALDI PSD

rhEPO fragmentation data was acquired using the novel multiplexed PSD methodology, the results of which is shown in Table 1. An example PSD spectra is inset in Figure 1. Six peptides provided high quality fragmentation data with a range of mass accuracy's.

LC-MSMS Data Directed Analysis™

A DDA run was conducted on the Q-Tof Premier and nanoACQUITY UPLC system with a total load of 50 fmol digested rhEPO. Seven of the possible 9 peptides (based on m/z range) were identified with validated MSMS spectra (containing ≥ 3 consecutive b or y" ions) which correspond to 93% sequence coverage of the amino acid sequence, discounting peptides with a mass below m/z 400.
Figure 3 shows assigned MSMS spectra for the 733 m/z precursor (monoisotopic mass 1464.75 m/z) and the obtained protein coverage.

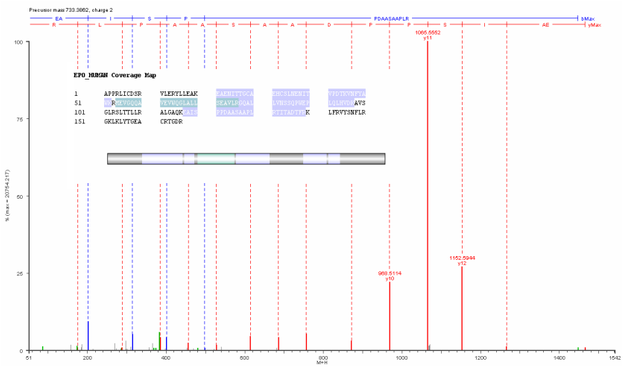

Figure 3. Obtained MSMS spectrum for peptide precursor 733 m/z. Shown inset is the coverage map for rhEPO identification

Submitted Mass	Submitted Charge	Experimental Mass	mW	Delta (Da)	Delta (ppm)	Probability	Start	End	Sequence	Modifications	Substitutions
897.055	3	2688.141	2688.201	0.06	22.251	43.079	48	72	(K)EAENITTGCAEHCSLNENITVPDTK(V)		
464.233	2	926.451	926.465	0.014	15.153	97.689	73	79	(K)VNFYAWK(R)		
842.782	3	2525.322	2526.331	0.01	3.77	63.635	81	103	(R)MEVGQGAVEYWGGLALLSEAVLR(G)		
848.114	3	2541.32	2541.326	0.006	2.498	72.625	81	103	(R)MEVGQGAVEYWGGLALLSEAVLR(G)	Oxidation M (1)	
486.294	2	970.572	970.581	0.009	9.496	108.251	96	103	(L)ALLSEAVLR(G)		
787.409	3	2359.205	2359.217	0.012	5.071	18.54	104	124	(R)GQALLVDSSQFWEPLQLHVDK(A)		D for N (7)
402.237	2	802.457	802.491	0.034	41.909	96.037	131	137	(R)SLTTLLR(A)		
733.386	2	1464.757	1464.757	0	0.083	203.493	144	158	(K)EAISPPDAASAAPLR(T)		
462.74	2	923.464	923.471	0.007	8.063	95.011	159	166	(R)TTADTFR(K)		
449.727	2	897.439	897.471	0.032	35.501	115.651	171	177	(R)VYSNFLR(G)		

Table 3. Details of peptide MSMS database search leading to rhEPO identification

LC-MSMS Parent Ion Discovery

To confirm attachment position and structure of the O-linked glycans a PID experiment was performed utilizing the enhance duty cycle option of the Q-TOF Premier focusing the MS survey on the glycan product ion 366.1 m/z. During the PID experiment 2 MSMS spectra were obtained on precursor ions 1061.61 m/z and 1207.04 m/z. These species correspond to the T13 fragment with the O-linked glycan isoforms. Table 4 outlines the observed fragments in the MSMS spectra based on the theoretical structure shown in Figure 4.

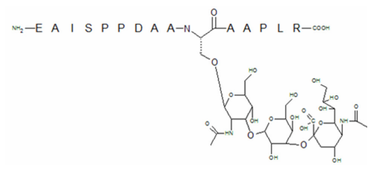

Figure 4. Structure of one isoform of the O-linked glycopeptide observed at 1061.61 m/z

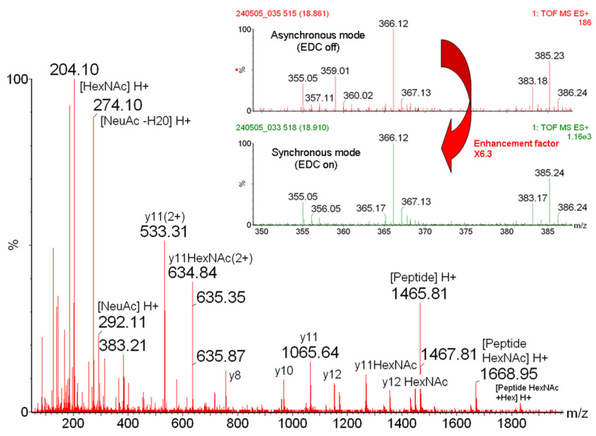

Figure 5. MSMS of glyco-peptide of 1061.61 m/z during EDC PID experiment – inset shows the 6 fold increase in sensitivity for the 366 m/z product ion due to Enhanced Duty Cycle.

Amino acid sequence	y_n	b_m	Fragment	Glycopetide or peptide ion Theoretical (M+xH)	Observed fragment of T13-Ser126-GalNAcGalNeuAc (1061.9 m/z)	Error (ppm)
R	1	15	y_1^+	175.1195	/	
L	2	14	b_{14}^+	1447.7600	1447.7539	-4.2134
P	3	13	y_3^+	385.2563	385.2544	-4.9318
A	4	12	$y_4^+()$	456.2934	456.288	-11.8345
A	5	11	$b_{11}^+()$	1010.4794	1010.4692	-10.0942
S	6	10	$y_6^+()$	614.3625	614.3661	5.8597
A	7	9	$y_7^+()$	685.3997	685.3982	-2.1885
			y_7-GalNAc$^+()$	888.4790	888.4799	1.0130
A	8	8	$y8^+()$	756.4368	756.4364	-0.5288
			y_8-GalNAc$^+()$	959.5162	959.5229	6.9827
D	9	7	$y^+()$	871.4637	871.4644	0.8032
P	10	6	$y_{10}^+()$	968.5165	968.5098	-6.9178
			y_{10}-GalNAc$^+()$	1171.5959	1171.5879	-6.8283
P	11	5	$y_{11}^{2+}()$	533.2885	533.2813	-13.5011
			y_{11}-GalNAc$^{2+}()$	634.8282	634.8256	-4.0956
			$y_{11}^+()$	1065.5692	1065.5676	-1.5015
			y_{11}-GalNAc$^+()$	1268.6486	1268.6461	-1.9706
S	12	4	$y_{12}^+()$	1152.6012	1152.6034	1.9087
			y_{12}-GalNAc$^+()$	1355.6807	1355.6844	-12.0235
I	13	3	$y_{13}^+()$	1265.6853	1265.6901	3.7924
A	14	2	$y_{14}^+()$	1336.7224	1336.7058	-12.4184
			$b_2^+()$	288.2000	/	/
E	15	1	T13$^{2+}()$	733.7700	/	/
			T13-GalNAc$^{2+}()$	834.9261	/	/
			T13-GalNAcNeuAc$^{2+}()$	980.4738	/	/
			T13$^+()$	1465.7700	1465.7458	-16.5101
			T13-GalNAc$^+()$	1668.8444	/	/
				RMS (ppm)		7.75

Table 4. Details of glycol-peptide 1061.61 m/z MSMS during EDC PID experiment

N-linked Glycan analysis

MALDI TOF

The purifed N-linked glycans were analysed by MALDI TOF MS in negative mode and data was internally corrected. Figure 6 shows the MS spectra of the N-glycans and Table 5 shows the observed species and mass error.

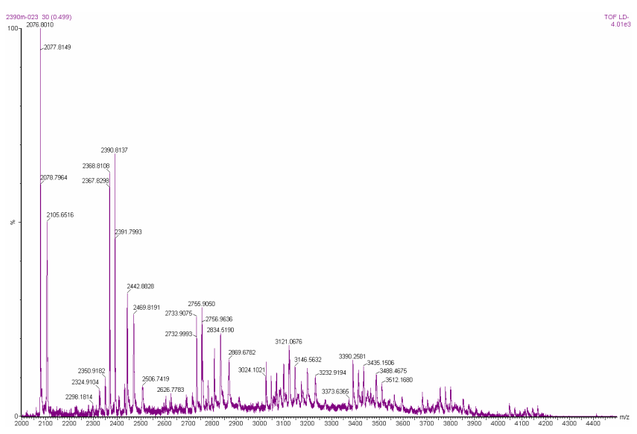

Figure 6. MALDI micro MX spectrum of cleaved N-glycans

N-glycans	Theor monoisotopic mass (m/z)	[M-H]-	[M-2H]2-	[M-3H]3-	[M-4H]4-	NanoESI Identified m/z	Error ppm	MALDI Identified m/z	Error ppm
Biantennary									
N2/A	2368.8409	2367.8331	1183.4126	/	/	1183.4138	1.0140	2367.8335	0.1689
Triantennary									
N2/B	2733.9731	2732.9653	1365.9787	910.3165	/	1365.9954	12.2257	2732.9653	0.0000
N2/C									
N3/A	3025.0685	3024.0607	1511.5264	1007.3483	/	1511.5098	-10.9823	3024.0549	-1.9180
N3/B									
Tetraantennary									
N2/D	3099.1052	3098.0975	1548.5448	1032.0273	/	1548.5542	6.0702	3098.0945	-0.9683
N3/C	3390.2007	3389.1929	1694.0925	1129.0591	/	1129.0593	0.1771	3389.2192	7.7600
						1694.0977	3.0695		
N2/E	3464.2375	3463.2297	1731.1103	1153.738	/	1731.1185	4.7368	3463.2847	15.8811
N4/A	3681.2961	3680.2883	1839.6402	1226.0909	919.3162	1839.6442	2.1743	3680.3765	23.9655
						1226.0922	1.0603		
						919.3227	7.0705		
N3/D	3755.3329	3754.3251	1876.6584	1250.7698	937.8254	1876.6706	6.5009	3754.4153	24.0256
						1250.7645	-4.2374		
N4/B	4046.4283	4045.4205	2022.2063	1347.8016	1010.5992	1347.8048	2.3742	4045.6406	54.4072
						1010.6002	0.9895		
N3/E	4120.4651	4119.4573	2059.2247	1372.4805	1029.1084	1372.4536	-19.5995	4119.7139	62.2898
N4/C	4411.5605	4410.5527	2204.7724	1469.5123	1101.8823	1469.5092	-2.1095	/	/
						1101.8875	4.7192		
N4/D	4776.6927	4775.6849	2387.3385	1591.2231	1193.1653	1591.2002	-14.3914	/	/
						1193.1535	-9.8897		
					RMS (ppm)		7.88		28.82

Table 5. Summary of NanoESI and MALDI analysis of cleaved N-glycans showing observed species and mass errors

NanoESI MS

The same batch of purifed N-linked glycans were analysed by nanoESI using the Q-Tof Premier in negative ion mode. The obtained spectrum is shown in Figure 7 and the mass assignments shown in Table 5. All theoretical glycan species were observed with good mass accuracies (several species observed in more than one charge state).

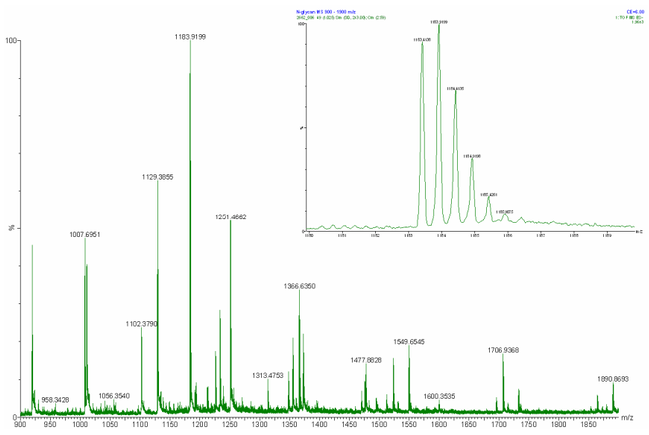

Figure 7. Q-Tof Premier nanoESI spectrum of cleaved N-glycans

Intact rhEPO analysis

Undigested rhEPO was analysed by LC-MS on the Q-TOF Premier using a 300µm C4 column. Figure 8 shows the obtained spectra for a dilution series ranging from 500 fmol to 10 pmol on column. These complex spectra were deconvoluted by the MaxEnt1 algorithm. From the deconvoluted spectra the heterogeneity of the rhEPO sample and differences in global gylcan and phosphate composition can be ascertained

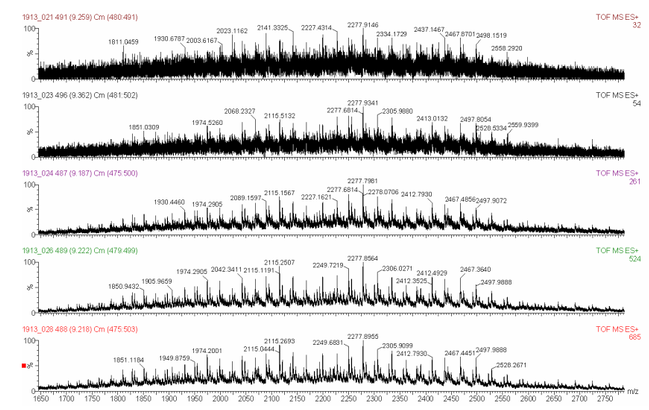

Figure 8. Spectra obtained during LCMS dilution series of intact EPO, 500 fmol, 1 pmol, 2.5 pmol, 5 pmol and 10 pmol on column

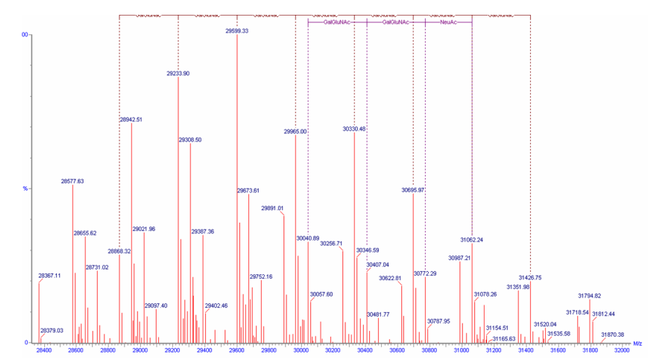

Figure 9. MaxEnt1 deconvoluted spectra corresponding to a total protein load of 10 pmol on column

CONCLUSIONS

- Results of peptide mapping and MSMS conclusively confirm the theoretical amino acid sequence with high mass accuracy and coverage.
- The PID experiment shows the location and composition of the O-glycan forms are consistent with those theoretically explained.
- MS analysis of N-glycans confirms the presence of all the theoretical N-glycan isoforms.
- Intact analysis shows a complex sample mix and provides relative quatification of the different forms of rhEPO.

FUTURE WORK

We will focus on the full characterization of the N-glycans using nanoscale MSMS and also further analysis of intact rhEPO by means of nanoscale LCMS

[1] S. A. Carr, *et al*, Protein Sci. 2 (1993) 183-196.
[2] M. Nimtz, *et al.* Eur. J. Biochem. 213 (1993) 39-56.
[3] D. Sagi, *et al.* J. Am. Soc. Mass. Spectrom. 13 (2002) 1138-1148.

TO DOWNLOAD A COPY OF THIS POSTER VISIT WWW.WATERS.COM/POSTERS

Lit Code Number 720001343EN
©2005 Waters Corporation