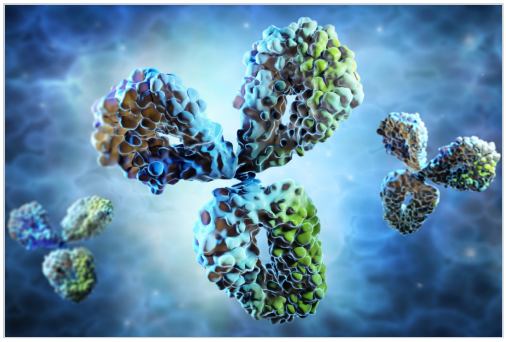
Waters™

Application Note

High Sequence Coverage Peptide Mapping of a Monoclonal Antibody with UPLC/MS^{E}

Hongwei Xie, Martin Gilar, John C. Gebler

Waters Corporation



Abstract

UPLC-MS^E is capable of characterizing digests of large, complex proteins such as monoclonal antibodies, with very high sequence coverage for both the heavy and light chains. Additionally, covalent modifications such as N-deamidation and N-linked glycosylation, can be identified. Peptide identifications are also supported by the high-quality MS^E data.

Introduction

Monoclonal antibodies are a class of recombinant proteins. They are subject to a variety of enzymatic or chemical modifications during expression, purification, and long-term storage. Liquid chromatography (LC)-based peptide mapping is the standard approach to establish protein identity, to identify post-translational modifications (PTMs), and to analyze potential impurities.

Peptide mapping with UltraPerformance LC (UPLC) Technology coupled with data-independent acquisition tandem mass spectrometry (UPLC-MS^E) has been used to obtain high sequence coverage, and to identify PTMs and variants of enolase.[1-2] The combination of MSE for unbiased identification[3] along with the resolution and sensitivity of UPLC[4-5] overcomes difficulties in protein characterization encountered with traditional LC-MS peptide mapping methods and data-dependent acquisition (DDA) LC-MS/MS sequencing tools, including co-elutions, analytical artifacts, and confusion of isobaric or nearly-isobaric peptides. MS^E has the advantage over DDA in that it does not require peak detection or prior knowledge of the precursor ions.

In this application note, we demonstrate the use of UPLC-MS^E to characterize a monoclonal antibody digest. The method provides high sequence coverage for both heavy and light chains of the antibody, and identifies modifications in a single analysis. MS^E spectra were used to identify peptide sequences. The unexpected peptides resulting from non-specific digestion were also identified.

Experimental

Sample and materials

Commercial monoclonal antibody was in solution (21 µg/µL). Iodoacetamide (IAM), dithiothreitol (DTT) and ammonium bicarbonate (NH_4HCO_3) were purchased from Sigma Chemical Co. (St. Louis, Mo., U.S.). Sequence-grade trypsin was from Promega Corp. (Madison, Wisc., U.S.), Formic acid (FA) was from EM Science (Gibbstown, N.J., U.S.) and Optima-grade acetonitrile (ACN) was from Fisher Scientific (Pittsburg, Penn., U.S.). RapiGest SF was from Waters Corp. (Milford, Mass., U.S.). The water used in all procedures was from a Millipore Milli-Q purification system (Bedford, Mass., U.S.).

Preparation of protein digests

The monoclonal antibody solution was diluted to 5 µg/µL with 100 mM NH_4HCO_3 containing 0.1% RapiGest SF. Proteins in 50 µL freshly-prepared solution were digested by adding 5 µg of sequencing-grade trypsin and incubating at 37 °C overnight. Before the digestion, 5 µL 100 mM DTT was added to reduce disulfide bonds at 60 °C for 30 min. Free cysteine residues were alkylated with 5 µL of 200 mM IAM at room temperature for 30 min in the dark. Formic acid (0.5% v/v) was used to quench enzymatic reaction and degrade RapiGest SF. After adding ACN (to 30% v/v), the digest was diluted to a 1.5 pmol/µL final concentration with 5:95 ACN/water containing 0.1% FA and analyzed by UPLC-MS^E.

UPLC-MS^E experiments

All analyses were performed using a Waters SYNAPT MS system controlled by MassLynx 4.1 Software. An ACQUITY UPLC System equipped with 2.1 x 150 mm BEH300Å 1.7 µm Peptide Separation Technology C_{18} Column was used for the separation. Peptides were eluted with a 90 min gradient (0 to 50 %B). Mobile phase A was 0.1% FA in water, B was 0.1% FA in ACN. The flow rate was 0.2 mL/min, and the column temperature was 40 °C. An auxiliary pump delivered a lockmass solution (100 fmol/µL (GLu1)-fibrinopeptide B (GFP) in 50:50 ACN/water containing 0.1% FA) for mass accuracy reference.

The instrument was operated in the positive ion V-mode. An alternating low collision energy (5 V) and elevated collision energy (ramping from 20 to 40 V) acquisition was used to acquire peptide precursor (MS) and fragmentation (MS^E) data. Scan time was 0.5 sec (1 sec total duty cycle). The capillary voltage was 3.0 kV, source temperature 100 °C, cone voltage 37 V, cone gas flow 10 L/h. Sampling of the lock spray channel was performed every 1 min.

Data processing

The acquired data were processed with Identity^E Software of ProteinLynx Global SERVER 2.3. The

processed data were first searched against a database consisting of monoclonal antibody light and heavy chain sequences, human, rat and mouse proteomes, with trypsin specificity and one potential miscleavage. Then, the data were searched again against the light and heavy chain sequences, without enzyme specificity. Cysteine (C) carbamidomethylation, asparagine (N) deamidation, and methionine (M) oxidation were allowed as optional modifications in these searches.

Results and Discussion

Figure 1 shows the UPLC-MS^E chromatogram of monoclonal antibody digest (30 pmol injected on-column). The chromatogram features more than 100 detected peaks.

In order to investigate potential protein contaminants in the antibody, the data were searched against a database consisting of human, rat, and mouse proteomes, plus the light and heavy chain sequences. Only the light and heavy chains of the antibody were identified, demonstrating that the sample was not contaminated with other proteins. It has been shown previously that UPLC-MS^E can identify protein contaminants at levels above 0.1 to 1%.[1]

The tryptic peptides identified are listed in Table 1 and assigned in Figure 1. They comprise 97% sequence of heavy chain and 97.2% sequence of light chain.

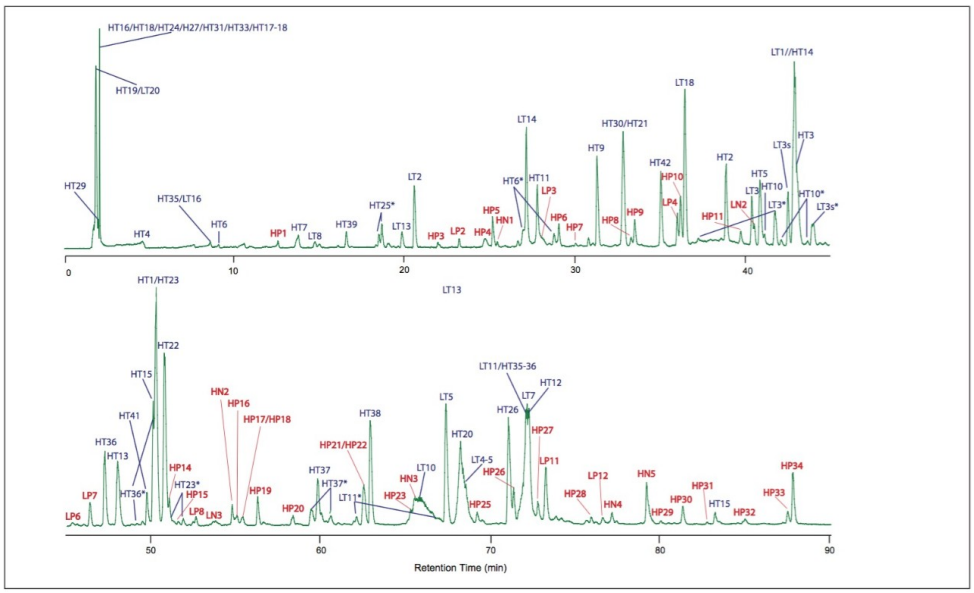


Figure 1. Chromatograms of monoclonal antibody tryptic digest. Tryptic peptides originating from heavy (HTxx) and light chain (LTxx) are annotated in blue. Partially tryptic (HPxx or LPxx) and non-tryptic peptides (HNxx or LNxx) are labeled in red. The asterisk indicates peptide with modification. For details, see Tables 1 and 2. Note that some peptides eluted as more than one peak. This is due to separation of glycosilation and deamidation peptide isoforms.

Protein	Peptide[1]	Sequence[2]	Modification	RT (Min)
Heavy Chain	HT19	SC#DK		2.06
Heavy Chain	HT29	VSNK		2.14
Heavy-Chain	HT24	TKPR		2.18
Heavy-Chain	HT16	VDK		2.2
Heavy-Chain	HT18	VEPK		2.2
Heavy Chain	HT27	EYK		2.2
Heavy-Chain	HT31	TISK		2.2
Heavy-Chain	HT33	GQPR		2.2
Heavy-Chain	HT4	QAPGK		4.5
Heavy-Chain	HT35	EEMTK		8.63
Heavy Chain	HT7	YADSVK		13.88
Heavy-Chain	HT39	LTVDK		16.71
Heavy-Chain	***HT25****	*EEQYNSTYR*	*Glycosylation N(300)*	*18.7*
Heavy-Chain	***HT25****	*EEQYNSTYR*	*Glycosylation N(300)*	*18.8*
Heavy-Chain	***HT6****	*IYPTNGYTR*	*Deamidation N(55)*	*27.19*
Heavy Chain	HT11	AEDTAVYYC#SR		27.93
Heavy-Chain	***HT6****	*IYPTNGYTR*	*Deamidation N(55)*	*28.92*
Heavy-Chain	HT9	FTISADTSK		31.44
Heavy-Chain	HT30	ALPAPIEK		32.95
Heavy-Chain	HT21	DTLMISR		32.99
Heavy Chain	HT42	SLSLSPG		35.19
Heavy-Chain	HT2	LSC#AASGFNIK		39.01
Heavy-Chain	HT5	GLEWVAR		41.02
Heavy-Chain	***HT2****	*LSC#AASGFNIK*	*Deamidation N(28)*	*41.26*
Heavy-Chain	HT10	NTAYLQMNSLR		41.3
Heavy-Chain	***HT10****	*NTAYLQMNSLR*	*Deamidation N(84)*	*42.62*
Heavy-Chain	***HT2****	*LSC#AASGFNIK*	*Deamidation N(28)*	*43*
Heavy-Chain	HT14	STSGGTAALGC#LVK		43.07
Heavy-Chain	HT3	DTYIHWVR		43.24
Heavy-Chain	***HT10****	*NTAYLQMNSLR*	*Deamidation N(84)*	*43.84*
Heavy Chain	HT36	NQVSLTC#LVK		47.31
Heavy-Chain	HT13	GPSVFPLAPSSK		48.08
Heavy-Chain	***HT36****	*NQVSLTC#LVK*	*Deamidation N(364)*	*49.2*
Heavy-Chain	HT41	WQQGNVFSC#SVMHEALHNHYTQK		49.78
Heavy-Chain	***HT36****	*NQVSLTC#LVK*	*Deamidation N(364)*	*50.22*
Heavy-Chain	HT1	EVQLVESGGGLVQPGGSLR		50.31
Heavy-Chain	HT23	FNWYVDGVEVHNAK		50.36
Heavy-Chain	HT22	TPEVTC#VVVDVSHEDPEVK		50.84
Heavy-Chain	***HT23****	*FNWYVDGVEVHNAK*	*Deamidation N(289)*	*51.3*
Heavy-Chain	***HT23****	*FNWYVDGVEVHNAK*	*Deamidation N(289)*	*51.9*
Heavy-Chain	***HT37****	*GFYPSDIAVEWESNGQPENNYK*	*Deamidation N(387)*	*59.46*
Heavy-Chain	HT37	GFYPSDIAVEWESNGQPENNYK		59.87
Heavy-Chain	***HT37****	*GFYPSDIAVEWESNGQPENNYK*	*Deamidation N(387)*	*60.6*
Heavy-Chain	HT38	TTPPVLDSDGSFFLYSK		62.97
Heavy Chain	HT20	THTC#PPC#PAPELLGGPSVFLFPPKPK		68.26
Heavy-Chain	HT26	VVSVLTVLHQDWLNGK		71.24
Heavy-Chain	***HT26****	*VVSVLTVLHQDWLNGK*	*Deamidation N(318)*	*71.9*
Heavy-Chain	HT35-36	EEMTKNQVSLTC#LVK		72.1
Heavy-Chain	***HT26****	*VVSVLTVLHQDWLNGK*	*Deamidation N(318)*	*72.3*
Heavy Chain	HT12	WGGDGFYAMDYWGQGTLVTVSSASTK		72.34
Heavy-Chain	HT15	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLY SLSSVVTVPSSSLGTQTYIC#NVNHKPSNTK		82.26
Light-Chain	LT20	GEC#		2
Light Chain	LT16	ADYEK		8.66
Light-Chain	LT6	FSGSR		9.12
Light-Chain	LT19	SFNR		10.62
Light-Chain	LT8	VEIK		14.87
Light Chain	LT13	VQWK		19.99
Light Chain	LT2	VTITC#R		20.71
Light-Chain	LT14	VDNALQSGNSQESVTEQDSK		27.26
Light-Chain	LT18	VYAC#EVTHQGLSSPVTK		36.6
Light-Chain	***LT3****	*ASQDVNTAVAWYQQKPGK*	*Deamidation N(30)*	*37.75*
Light Chain	LT3	ASQDVNTAVAWYQQKPGK		40.53
Light-Chain	***LT3****	*ASQDVNTAVAWYQQKPGK*	*Deamidation N(30)*	*41.86*
Light-Chain	LT3s	ASQDVNTAVAWYQQK		42.64
Light-Chain	LT1	DIQMTQSPSSLSASVGDR		42.97
Light-Chain	***LT3s****	*ASQDVNTAVAWYQQK*	*Deamidation N(30)*	*44.03*
Light Chain	LT15	DSTYSLSSTLTLSK		50.16
Light-Chain	***LT11****	*SGTASVVC#LLNNFYPR*	*Deamidation N(137)*	*62.16*
Light-Chain	LT10	TVAAPSVFIFPPSDEQLK		65.9
Light-Chain	***LT11****	*SGTASVVC#LLNNFYPR*	*Deamidation N(137)*	*67.1*
Light-Chain	LT5	LLIYSASFLYSGVPSR		67.42
Light Chain	LT4-5	APKLLIYSASFLYSGVPSR		68.59
Light-Chain	LT11	SGTASVVC#LLNNFYPR		72.1
Light-Chain	LT7	SGTDFTLTISSLQPEDFATYYC#QQHYTTPPTFGQGTK		72.16

[1] HT - Tryptic peptide from Heavy chain, LT - Tryptic peptide from light chain;

[2] C# - Carbamidomethyl cysteine; * Peptide with modification as shown in the table.

Table 1. Tryptic peptides identified in the mAb digest. Monoclonal antibody was digested after a prolonged storage in solution.

Peptide HT25 was identified as N-linked glycosylated forms. In addition, nine deamidated N sites (seven in heavy chain and two in light chain, two isoforms each as shown in Table 1) were also identified.

After assignment of the tryptic peptides, unknown peaks remain in the chromatogram. These peaks are partially tryptic and non-tryptic peptides, as identified in a subsequent database search against light and heavy chain sequences without enzyme specificity. The identified partially/non-tryptic peptides are presented in Table 2 and shown in Figure 1.

Protein	Peptide[1]	Sequence[2]	Start	End	Cleavage[3]	RT (min)
Heavy Chain	**HP1**	IC#NVNHKPSNTK	202	213	Partial	12.70
Heavy Chain	**HP2**	HKPSNTK	207	213	Partial	13.88
Heavy Chain	**HP3**	STSGGTAALGC#	137	147	Partial	22.06
Heavy Chain	**HP4**	EPQVY	348	352	Partial	24.78
Heavy Chain	**HP5**	MHEALHNHYTQK	431	442	Partial	25.32
Heavy Chain	**HP6**	SVMHEALHNHYTQK	429	442	Partial	29.19
Heavy Chain	**HP7**	SGFNIK	25	30	Partial	30.19
Heavy Chain	**HP8**	AEDTAVYY	88	95	Partial	33.43
Heavy Chain	**HP9**	LSC#AASGF	20	27	Partial	33.65
Heavy Chain	**HP10**	NTAYLQMN	77	84	Partial	36.34
Heavy Chain	**HP11**	SC#SVMHEALHNHYTQK	427	442	Partial	39.73
Heavy Chain	**HP12**	WQQGNVFSC#	420	428	Partial	41.91
Heavy Chain	**HP13**	WQQGNVF	420	426	Partial	44.16
Heavy Chain	**HP14**	DYWGQGTLVTVSSASTK	108	124	Partial	51.12
Heavy Chain	**HP15**	SLSSVVTVPSSSLGTQTYIC#NVNHKPSNTK	184	213	Partial	51.63
Heavy Chain	**HP16**	GPSVFPLAPS	125	134	Partial	55.07
Heavy Chain	**HP17**	WGGDGFYAM	99	107	Partial	55.41
Heavy Chain	**HP18**	FNWYVDGVEVHN	278	289	Partial	55.48
Heavy Chain	**HP19**	FNWYVDGVEVH	278	288	Partial	56.32
Heavy Chain	**HP20**	YAMDYWGQGTLVTVSSASTK	105	124	Partial	58.38
Heavy Chain	**HP21**	DYFPEPVTVSWN	151	162	Partial	62.52
Heavy Chain	**HP22**	THTC#PPC#PAPELLGGPSVF	226	244	Partial	62.61
Heavy Chain	**HP23**	DYFPEPVTVSWNSGALTSGVH	151	171	Partial	65.33
Heavy Chain	**HP24**	PELLGGPSVFLFPPKPK	235	251	Partial	68.30
Heavy Chain	**HP25**	DYFPEPVTVSWNSGAL	151	166	Partial	69.25
Heavy Chain	**HP26**	TTPPVLDSDGSFFLY	396	410	Partial	71.41
Heavy Chain	**HP27**	SVLTVLHQDWLNGK	307	320	Partial	72.29
Heavy Chain	**HP28**	VVSVLTVLHQDWLN	305	318	Partial	75.98
Heavy Chain	**HP29**	VVSVLTVLHQDWL	305	317	Partial	80.06
Heavy Chain	**HP30**	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLY	151	183	Partial	81.36
Heavy Chain	**HP31**	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLS	151	186	Partial	82.89
Heavy Chain	**HP32**	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSL	151	185	Partial	85.12
Heavy Chain	**HP33**	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYIC#NVN	151	206	Partial	87.55
Heavy Chain	**HP34**	DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTY	151	201	Partial	87.86
Heavy Chain	**HN1**	TVPSSSLGTQTY	190	201	Non	25.15
Heavy Chain	**HN2**	SLSSVVTVPSSSLGTQTY	184	201	Non	54.81
Heavy Chain	**HN3**	SGALTSGVHTFPAVLQSSGLY	163	183	Non	65.55
Heavy Chain	**HN4**	TSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTY	167	201	Non	77.20
Heavy Chain	**HN5**	SGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTY	163	201	Non	79.24
Light Chain	**LP1**	SQESVTEQDSK	159	169	Partial	13.90
Light Chain	**LP2**	SASVGDR	12	28	Partial	23.34
Light Chain	**LP3**	PPTFGQGTK	95	103	Partial	28.11
Light Chain	**LP4**	VYAC#EVTHQGL	191	201	Partial	36.15
Light Chain	**LP5**	ASQDVNTAVAWYQQK	25	39	Partial	42.65
Light Chain	**LP6**	LLNNFYPR	135	142	Partial	45.40
Light Chain	**LP7**	TVAAPSVF	109	116	Partial	46.44
Light Chain	**LP8**	SGTDFTLTIS	67	76	Partial	52.68
Light Chain	**LP9**	LLIYSASFLYSGVP	170	181	Partial	59.55
Light Chain	**LP10**	SGTDFTLTISSLQPEDFATYYC#QQH	67	91	Partial	71.71
Light Chain	**LP11**	SGTDFTLTISSLQPEDFATYYC#QQHY	67	92	Partial	73.29
Light Chain	**LP12**	SGTDFTLTISSLQPEDFATYY	67	87	Partial	76.65
Light Chain	**LN1**	TAVAWYQQK	31	39	Non	35.42
Light Chain	**LN2**	DIQMTQSPSSLSA	1	13	Non	40.68
Light Chain	**LN3**	SLQPEDFATYYC#QQHY	77	92	Non	53.84

[1] HP - Partially tryptic peptide from heavy chain, HN - Non-tryptic peptide from heavy chain, LP - Partially tryptic peptide from light chain, LN - Non-tryptic peptide from light chain.

[2] C# - Carbamidomethyl cysteine.

[3] Partial - Partially tryptic cleavage; Non - Non-tryptic cleavage.

Table 2. Sequences and retention times of identified partially tryptic and non-tryptic peptides.

The second search suggests that the longest tryptic peptide HT15 in the heavy chain (63 amino acids; 6712.3 Da) was fragmented into a series of partially tryptic peptides (HP1, HP2, HP15, HP21, HP23, HP25, HP30, HP31, HP32, HP33, HP34) and non-tryptic peptides (HN1, HN2, HN3, HN4, HN5). All the MS^E spectra of peptides listed in Table 2 were manually inspected, confirming the validity of identifications.

Figure 2 shows the MS^E spectra for partially tryptic peptides HP30 and HP15, products of chymotryptic cleavage of tryptic HT15 peptide. When combined, they cover the entire sequence of the HT15.

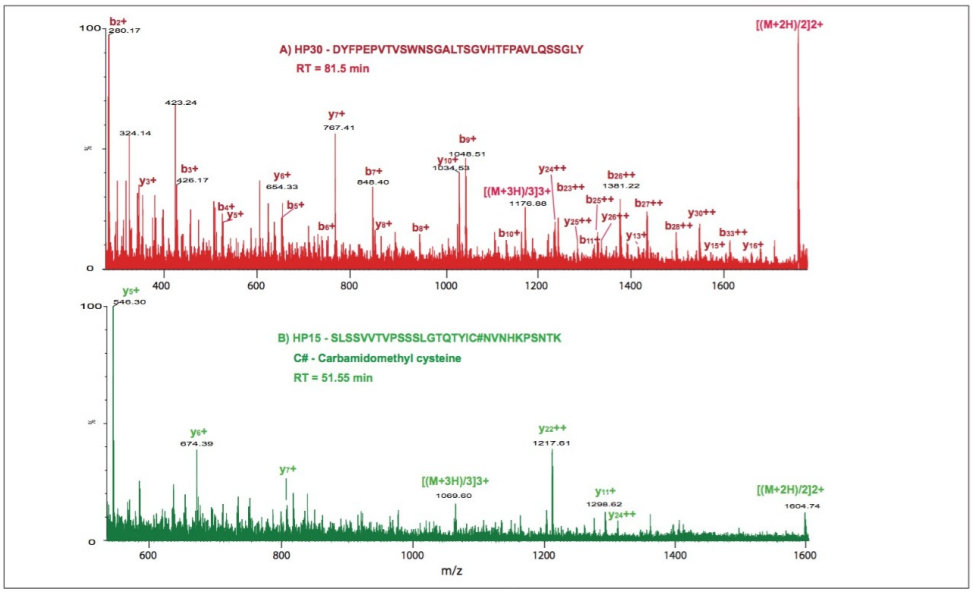


Figure 2. MS^E spectra of two partially-tryptic peptides, which cover the whole sequence of the longest tryptic peptide HT15 in heavy chain when combined together. A) HP30, first part of HT15; B) HP15, second part of HT15.

Two broad peaks were observed in the chromatogram eluting between 65 and 70 minutes (Figure 1). They are proline-rich peptides LT10 (65.5 min) and HT20 (68.3 min). Their peak shapes could be improved by using higher column temperature. MS^E allows for identification of peptides without complete resolution because it acquires fragment ion data of multiple peptide precursors in parallel.

Conclusion

The results of the analysis of a monoclonal antibody, with very high sequence coverage for both the heavy and light chains, illustrate that UPLC-MS^E is capable of characterizing digests of large, complex proteins such as monoclonal antibodies. The unexpected peptides in the sample resulting from non-specific digestion were also identified with high confidence.

MS^E enables sampling low-abundant components and acquires indiscriminately MS^E spectrum. The specific conclusions from this study show that:

1. High sequence coverage can be obtained for both heavy and light chains of a monoclonal antibody
2. Covalent modifications, for example, N-deamidation and N-linked glycosylation, can be identified
3. Peptide identifications are supported by high-quality MS^E spectra

MS^E collects precursor and fragment ion data of peptides within a single LC run, which enables the method to sequence the LC-eluted peptides. This ensures identification of unknown protein contaminants and PTMs in a single analysis. Therefore, the method may be used for detailed characterization of such proteins without additional MS/MS analyses and sacrificing the analytical efficiency. The method should increase the analytical efficiency in the characterization of therapeutic proteins.

References

1. Xie HW, Gilar M, Gebler JC. Characterization of Protein Impurities by Peptide Mapping with UPLC/MSE. Waters Application Note. 2009; 720002809en.

2. Chakraborty AB, Berger SJ, Gebler JC. Use of an Integrated MS-Multiplexed MS/MS Data Acquisition for High-coverage Peptide Mapping Studies. Rapid Comm. *Mass Spec*. 2007; 21: 730.

3. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li G-Z, McKenna T, Nold M, Richardson K, Young P, Geromanos S. Quantitative Proteomic Analysis by Mass Retention Time Pairs. *Anal. Chem*. 2005; 77: 2187.

4. Mazzeo JR, Wheat TE, Gillece-Castro BL, Lu Z. Next Generation Peptide Mapping with UPLC. *BioPharm International*, January 1, 2006.

5. Wheat TE, Lu Z, Gillece-Castro BL, Mazzeo JR. Quantitative Aspects of UPLC Peptide Mapping. Waters Application Note. 2007; 720001839en.

Featured Products

ACQUITY UPLC System

MassLynx MS Software

ProteinLynx Global SERVER (PLGS)

720002821, January 2009