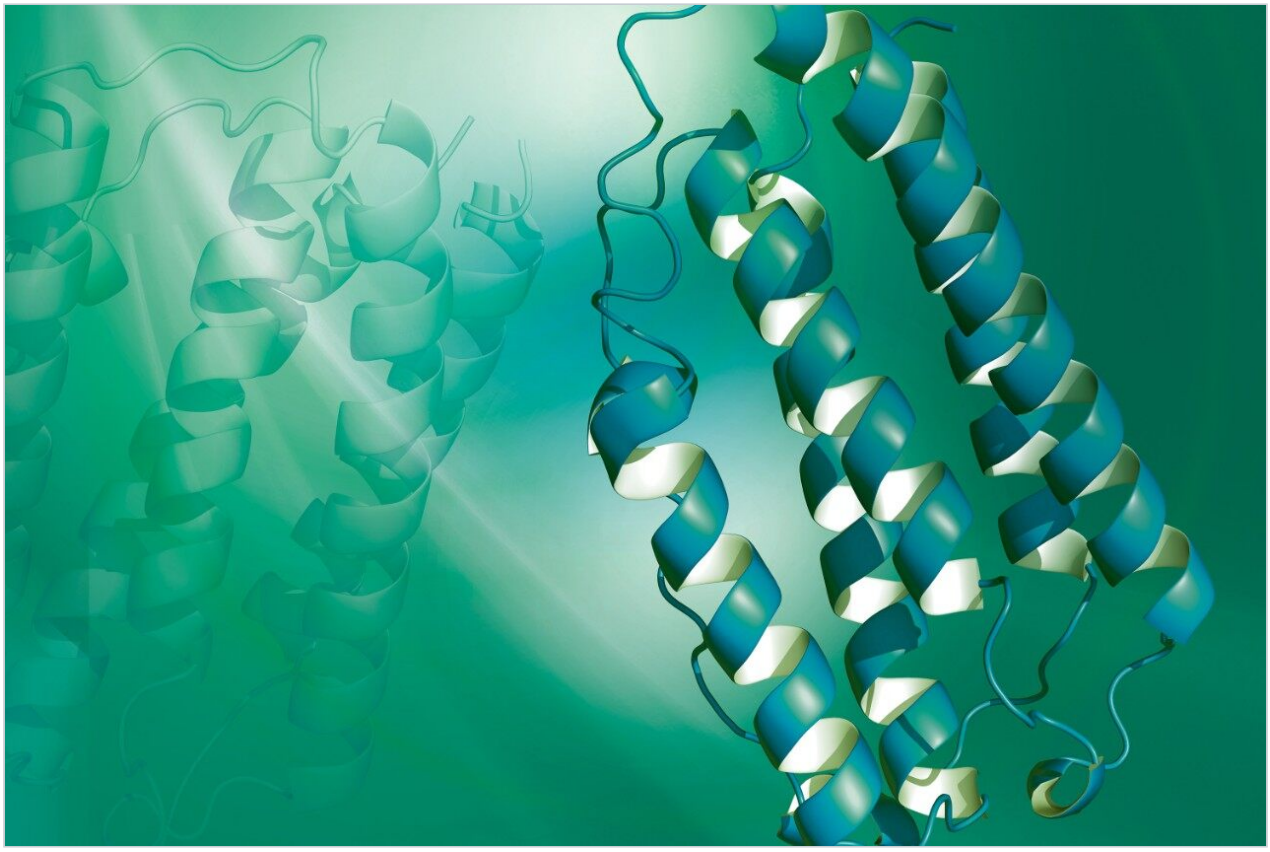
Waters™

Applikationsbericht

Characterization of Protein Impurities by Peptide Mapping with UPLC-MS^{E}

Hongwei Xie, Martin Gilar, John C. Gebler

Waters Corporation



Abstract

In this application note, we demonstrate an application of MS^E, using the Waters SYNAPT MS System coupled with UltraPerformance LC (UPLC), for characterization of a yeast enolase tryptic digest.

Introduction

Impurities such as modifications and sequence variants generated from transcriptional/translational errors are common in recombinant protein products, and may affect their safety and activity. Effective control of these variants requires sensitive and reproducible methods for protein production monitoring.

Liquid Chromatography (LC)-based peptide mapping is a key method for protein structure characterization and purity analysis. However, ultra violet (UV) or mass spectrometry (MS) detectors in traditional LC/UV or LC-MS peptide mapping methods are unable to characterize unexpected contaminants, although they are sensitive for detection of low-level impurities in recombinant proteins. Additional time-consuming tandem mass spectrometry (MS/MS) measurements are required for the elucidation of unknown sequences. Furthermore, the presence of peptides resulting from unexpected proteolytic cleavages often makes the LC separation and the assignment of LC peaks more difficult.

Recently, LC combined with data-independent acquisition mass spectrometry (MS^E) has been employed to analyze peptide maps with a high sequence coverage (>90%).[1] Excellent analytical reproducibility was obtained from replicate analyses of the protein digest. In MS^E [2-3], the parallel and unbiased data acquisition mode not only overcomes the repeatability limitations of data-dependent acquisition (DDA) LC-MS/MS experiments, but also ensures the sampling of low-abundance peptides from low-level impurities. The obtained MS and MS^E spectra of such peptides allow for identification of unknown impurities in the sample.

In this application note, we demonstrate an application of MS^E, using the Waters SYNAPT MS System coupled with UltraPerformance LC (UPLC), for characterization of a yeast enolase tryptic digest. Multiple protein contaminants as well as unexpected peptides resulting from non-specific digestion were identified. The results demonstrate that UPLC-MS^E methodology is capable of identifying and quantifying low-level impurities in protein products. LC peaks from unexpected partially tryptic and non-tryptic cleavages are assigned and distinguished from peptides originating from impurity proteins.This methodology may also be used to accelerate the development of protein purification strategies.

Experimental

Sample and Materials

Yeast enolase and ammonium bicarbonate (NH_4HCO_3) were purchased from Sigma Chemical Co. (St. Louis, MO, U.S.); sequencegrade trypsin from Promega Corp. (Madison, WI, U.S.), formic acid (FA) from EM sciences (Gibbstown, NJ, U.S.) and optima-grade acetonitrile (ACN) from Fisher Scientific (Pittsburg, PA, U.S.). The water used in all procedures was supplied by a Millipore Mili-Q purification system (Bedford, MA, U.S.).

Preparation of Protein Digests

Yeast enolase was dissolved in 100 mM NH_4HCO_3 to prepare a 5 µg/µL protein solution. Fifty microliters of this solution were utilized for tryptic digestion, accomplished by adding 5 µg of sequencing-grade trypsin and incubating at 37 °C overnight. The digest was diluted to 1.5 pmol/µL with 5:95 ACN/water containing 0.1% FA prior to UPLC/MS^E analysis.

UPLC-MS^E Experiments

All analyses were performed using a SYNAPT MS System and controlled by MassLynx 4.1 Software. An ACQUITY UPLC System equipped with a 2.1 x 100 mm BEH 300Å 1.7 µm Peptide Separation Technology Column was used for analysis. The separation was performed at 40 °C. Peptides were eluted with a 60 min gradient (0 to 50% B). Mobile phase A was 0.1% FA in water, B was 0.1% FA in ACN. Flow rate was 0.2 mL/min. An auxiliary pump delivered a lockmass solution (100 femtomole (GLu1)- fibrinopeptide B (GFP) in 50:50 ACN/water containing 0.1% FA for mass accuracy reference.

Data-independent MS acquisition in the positive ion V-mode was accomplished by alternating the collision cell energy between low (5 V, transfer cell energy 5 V) and elevated-energy setting (energy ramped from 20 to 40 V, transfer cell energy 10 V). The scan time was 0.5 sec in both acquisition modes (1 sec total duty cycle). In this configuration, both peptide precursor ion spectrum (MS) and fragmentation (MS^E) data can be obtained in a single LC analysis.

Capillary voltage of 3.0 kV, source temperature of 100 °C, cone voltage of 37 V, and cone gas flow of 10 L/h were maintained during the analyses. Sampling of the lock spray channel was performed every 1 min. The system was tuned for a minimum resolution of 10,000 and calibrated to mass error of less than 3 ppm using a 100 femtomole GFP infusion before the experiments.

Data Processing

The acquired data were processed using Identity^{E} Software. The low-energy (MS) and elevated-energy (MS^{E}) data were backgroundsubtracted, de-isotoped and charge-state-reduced to corresponding monoisotopes, lockmass-corrected, and aligned (fragment ions with corresponding precursor ions) by the retention time profile of each ion. The processed data were first searched against a yeast database with trypsin specificity and one potential missed cleavage. Then, the processed data were searched again against the protein sequences returned from the first search with no enzyme specified for partially/non-tryptic peptides. Asparagine (N) deamidation and methionine (M) oxidation were allowed as variable modifications in these searches.

Results and Discussion

In the UPLC-MS^{E} experiment, two sets of MS data are collected: lowenergy (MS) and elevated-energy (MS^{E}) chromatograms. Low-energy LC-MS data comprise accurate MS data for peptide precursors, while elevated-energy LC-MS data contain fragment ions to their corresponding peptide precursors. Identity^{E} Software is used to combine the data into MS/MS spectra, and search for peptide sequences.

Figure 1 shows the LC-MS chromatogram of enolase tryptic digest for a 30 pmole sample injected on-column. The chromatogram features more than 100 resolved and detected peaks. The peak assignment (as shown in Figure 1) was made after identification of the peptides via a database search using MS^{E} data (see Data Processing section for details).

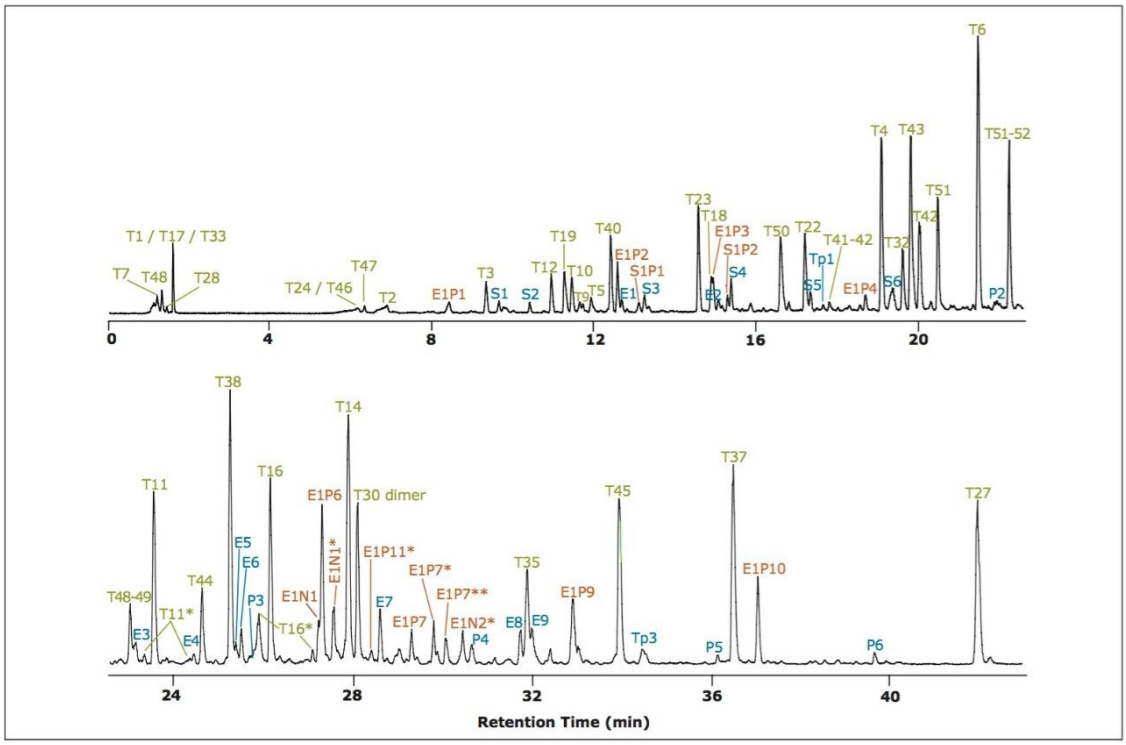


Figure 1. Chromatograms of enolase tryptic digest. Enolase tryptic peptides (Txx) are annotated in green. Tryptic peptides from enolase 2, Cu-Zn superoxide dismutase, glucose-6-phosphate isomerase, and triosephosphate isomerase are annotated in blue. Partial tryptic and non-tryptic peptides are labeled in orange. The asterisk indicates peptide with N-deamidation. For details, see Tables 2, 3, and 4.

Five proteins listed in Table 1 were identified via a search against the yeast proteome database (using trypsin specificity). As expected, enolase 1 was the top hit. 42 tryptic peptides (84% sequence coverage) were identified from this target protein, including three miscleaved peptides and two N-deamidated peptides (two isoforms each). These tryptic peptides (see Table 2) were assigned to the major peaks in the chromatogram, as labeled in Figure 1 using the Txx convention (green labels).

Protein name	Identified peptides	Intensity_ 3most x 10^3
Enolase 1; Eno1p	42	464.0 ± 35
Enolase 2; Eno2p	9	61.7 ± 5*
Cu-Zn superoxide dismutase; Sod1p	7	61.5 ± 2.3
Glucose-6-phosphate isomerase; Pgi1p	6	14.4 ± 1.1
Triosephosphate isomerase; Tpi1p	3	6.3 ± 0.7

Table 1. Identified proteins from enolase sample.

** Only peptides unique to enolase 2 were counted.*

Protein	Peptide	Sequence	RT (min)
Eno1p	**T7**	DGDK	1.24
Eno1p	**T48**	SER	1.33
Eno1p	**T28**	AAGHDGK	1.45
Eno1p	**T1**	AVSK	1.6
Eno1p	**T17**	AAAAEK	1.61
Eno1p	**T33**	NPNSDK	1.61
Eno1p	**T24**	SLTK	6.13
Eno1p	**T46**	TGQIK	6.19
Eno1p	**T47**	TGAPAR	6.34
Eno1p	**T2**	VYAR	6.87
Eno1p	**T3**	SVYDSR	9.34
Eno1p	**T12**	ANIDVK	10.95
Eno1p	**T19**	HLADLSK	11.28
Eno1p	**T10**	GVLHAVK	11.46
Eno1p	**T9**	WMGK	11.65
Eno1p	**T5**	GVFR	11.94
Eno1p	**T40**	IATAIEK	12.42
Eno1p	**T23**	IGSEVYHNLK	14.58
Eno1p	**T18**	NVPLYK	14.92
Eno1p	**T50**	LNQLLR	16.62
Eno1p	**T22**	TFAEALR	17.22
Eno1p	**T41-42**	KAADALLLK	17.83
Eno1p	**T4**	GNPTVEVELTTEK	19.1
Eno1p	**T32**	YDLDFK	19.63
Eno1p	**T43**	VNQIGTLSESIK	19.83
Eno1p	**T42**	AADALLLK	20.05
Eno1p	**T51**	IEEELGDNAVFAGENFHHGDK	20.5
Eno1p	**T6**	SIVPSGASTGVHEALEMR	21.48
Eno1p	**T51-52**	IEEELGDNAVFAGENFHHGDKL	22.26
Eno1p	**T48-49**	SERLAK	23.01
Eno1p	**T11***	NV**N**DVIAPAFVK	23.33
Eno1p	**T11**	NVNDVIAPAFVK	23.55
Eno1p	**T11***	NV**N**DVIAPAFVK	24.4
Eno1p	**T44**	AAQDSFAAGWGVMVSHR	24.62
Eno1p	**T38**	TAGIQIVADDLTVTNPK	25.27
Eno1p	**T16***	LGA**N**AILGVSLAASR	25.95
Eno1p	**T16**	LGANAILGVSLAASR	26.15
Eno1p	**T16***	LGA**N**AILGVSLAASR	26.95
Eno1p	**T14**	AVDDFLISLDGTANK	27.88
Eno1p	**T30-dimer**	IGLDCASSEFFK	28.09

data and Identity[E] Software provide for more rigorous quantification using a method described by Silva and colleagues.[3] The relative protein concentration was determined based on the ratio of the sum of intensities of the three most abundant peptides identified from each protein (with exception of enolase 2, a homolog of enolase 1). For the quantification of enolase 2, the three most abundant peptides unique to enolase 2 were used and compared with three peptides from enolase 1 with similar sequences. Figure 2 shows relative concentrations of the proteins normalized to enolase 1; two of the protein contaminants are present at levels above 10%.

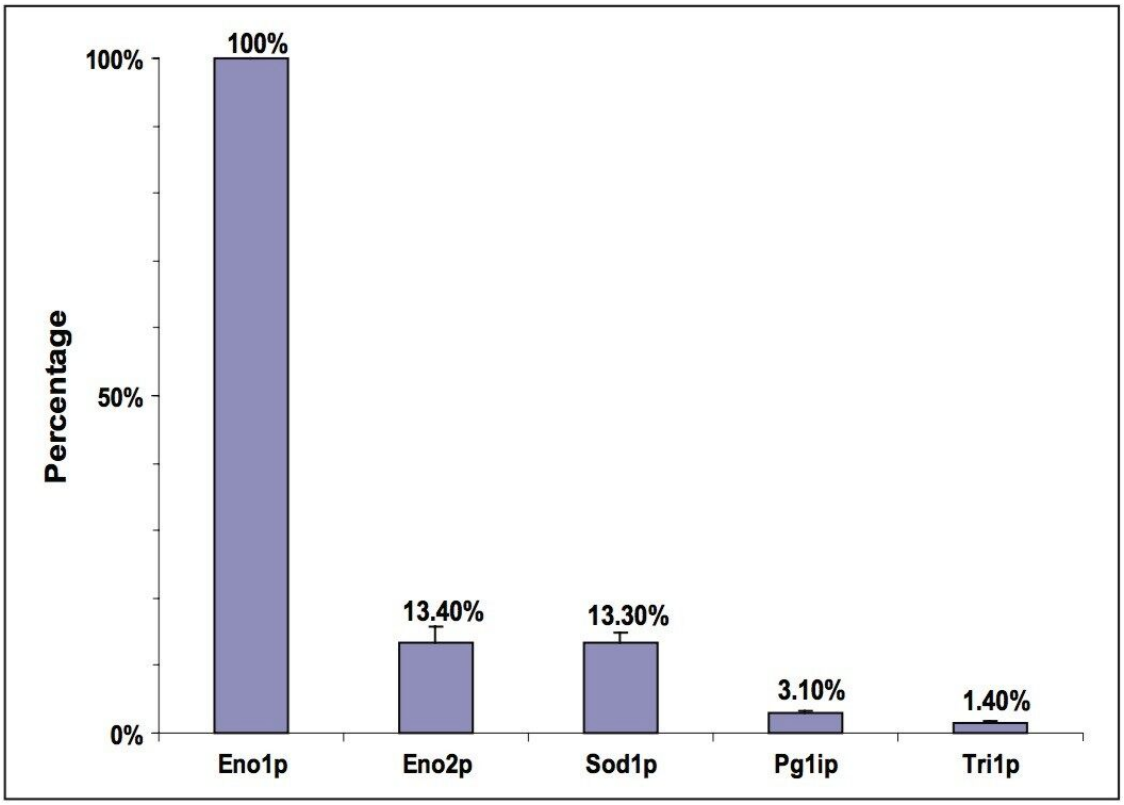


Figure 2. Normalized quantification of protein contaminants found in enolase sample.

Even after the assignment of peptides originated from protein contaminants, some peaks in Figure 1 remained unknown. An additional search for peptides with no enzyme specificity was performed using a truncated database consisting of only proteins identified previously. This allows for identification of unusual protein cleavages and unexpected sequences. Indeed, several major LC peaks were identified as partially or non-tryptic peptides. Although the longest enolase 1 tryptic peptide (T21, 3736.97 Da) was not found in the LC chromatogram, series of partially-tryptic peptides (E1P9, E1P7, E1P7*, E1P7**, E1P1, E1P2) and non-tryptic peptides (E1N1, E1N1*, E1N2*) related to T21 sequence were present in the sample. Other partially/non-tryptic

peptides, as listed in Table 4, were also identified with high confidence. All MS^E spectra of partially- and non-tryptic peptides listed in Table 4 were validated by additional manual inspection.

Protein	Peptide	Peptide Sequence	Start	Length	Cleavage	RT (min)
Eno1p	**E1P1**	PTGAK	173	5	Partial	**8.46**
Eno1p	**E1P2**	APTGAK	172	6	Partial	**12.59**
Eno1p	**E1P3**	YGASAGNVGDEGGVAPN	201	17	Partial	**14.96**
Eno1p	**E1P4**	GWGVMVSHR	366	9	Partial	**18.71**
Eno1p	**E1P5**	LGVSLAASR	111	9	Partial	**23.01**
Eno1p	**E1P6**	WLTGPQLADLYH	272	12	Partial	**27.30**
Eno1p	**E1P7**	LNVLNGGSHAGGALALQEFMIAPTGAK	151	27	Partial	**29.31**
Eno1p	**E1P7***	LNVLNGGSHAGGALALQEFMIAPTGAK	151	27	Partial	**29.80**
Eno1p	**E1P7****	LNVLNGGSHAGGALALQEFMIAPTGAK	151	27	Partial	**30.09**
Eno1p	**E1P8**	WLTGPQLADLYHSL	272	14	Partial	**31.99**
Eno1p	**E1P9**	TSPYVLPVPF	141	10	Partial	**32.91**
Eno1p	**E1P10**	IQTAEEALDLIVDAIK	218	16	Partial	**37.05**
Eno2p	**E1P11***	VLNGGSHAGGALALQEFMIAPTGAK	153	25	Partial	**28.42**
Eno1p	**E1N1**	LNVLNGGSHAGGALALQEF	151	19	Non	**27.23**
Eno1p	**E1N1***	LNVLNGGSHAGGALALQEF	151	19	Non	**27.56**
Eno1p	**E1N2***	LNVLNGGSHAGGALALQEFM	151	20	Non	**30.46**
Eno2p	**E1N3***	VLNGGSHAGGALALQEF	153	17	Non	**25.25**
Eno2p	**E2P1**	AILGVSMAAAR	109	11	Partial	**21.89**
Sod1p	**S1P1**	AGVSGVVK	11	8	Partial	**13.25**
Sod1p	**S1P2**	IHAGQDDLGK	119	10	Partial	**15.39**
Sod1p	**S1P3**	VQAVAVLK	1	8	Partial	**17.35**

* N marked in red was deaminated.

Table 4. Sequences and retention times of identified partially-tryptic and nontryptic peptides.

When including all identified peptides, the enolase 1 sequence coverage increased from 84% to 96%. The

partially tryptic and non-tryptic peptides are labeled in red in Figure 1. Figure 3 shows an example of MS^E spectrum for TSPYVLPVPF, a partially tryptic peptide originating from T21 of enolase 1.

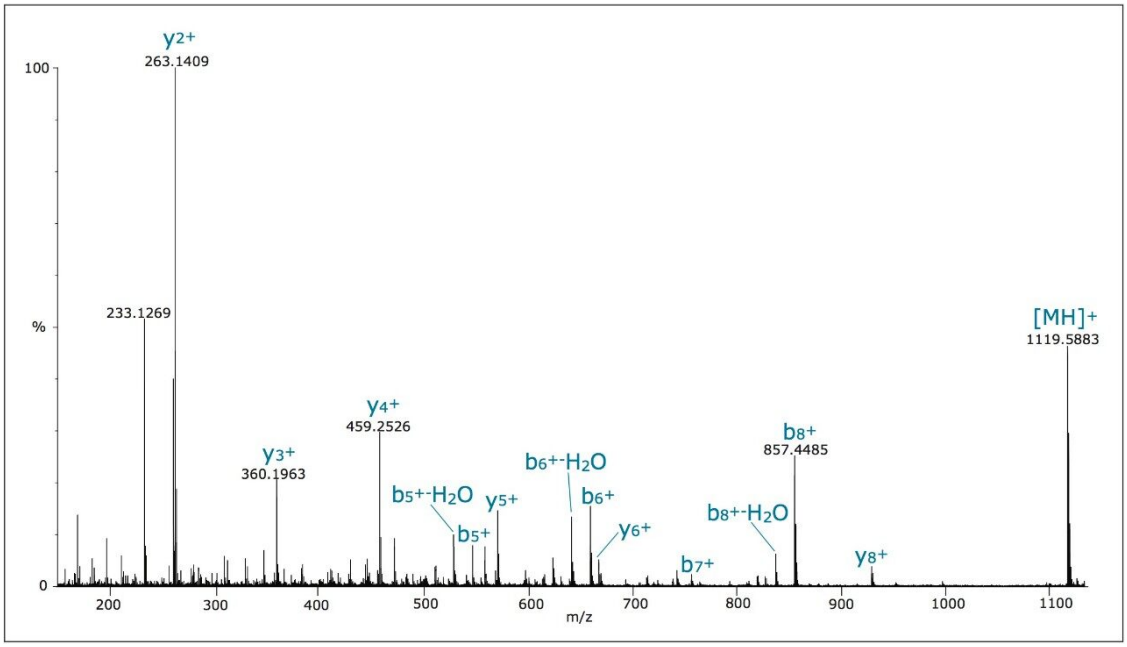


Figure 3. MS^E spectrum of the partially-tryptic peptide E1P9 (TSPYVLPVPF), a part of the longest peptide T21 of enolase 1.

Conclusion

The results presented here demonstrate the advantages of a UPLC-MS^E method for characterization of protein digests. Dataindependent MS acquisition in conjunction with high-resolution UPLC and specialized informatics tools allows for:

- Identification of expected as well as unexpected peptides in protein tryptic digest
- Providing high sequence coverage of the target protein n Identification of unknown impurities, originating from unknown protein contaminants
- Quantification of low-level peptide/protein impurities. Impurities at ~1% level have been successfully identified and quantified
- Identification and quantification of protein modifications; for example, N-deamidation. Relative

quantification of posttranslational modifications will be discussed in subsequent work4

- Assignment of all peaks in the peptide maps without the need for multiple analyses of the same sample as is common in DDA-based LC-MS/MS experiments
- High-quality sequencing of all LC peaks present in sufficient abundance (e.g. T30 dimer linked via S-S bond was identified in enolase 1)

UPLC-MS^E meets the requirements for a robust and flexible method needed to monitor the safety of biopharmaceutical proteins. The UPLC-MS^E method has the potential to expedite recombinant protein drug development.

In conclusion, UPLC-MS^E is a powerful tool that provides solutions for protein characterization challenges that are difficult to address with traditional peptide mapping and DDA-based LC-MS/MS methods.

References

1. Chakraborty AB, Berger SJ, Gebler JC. *Rapid Comm. Mass Spec*. 2007; 21: 730.
2. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young O, Geromanos SJ. *Anal. Chem.* 2005; 77: 2187.
3. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. *Mol. Cell. Proteomics.* 2006; 5: 144.
4. Xie H, Gilar M, Gebler JC. Analysis of Deamidation and Oxidation in Monoclonal Antibody using Peptide Mapping with UPLC-MS^E. Waters. 2009; 720002897EN < https://www.waters.com/nextgen/us/en/library/application-notes/2009/Deamidation-oxidation-monoclonal-antibody-using-peptide-mapping-uplc-mse.html> .

Featured Products

ACQUITY UPLC System <https://www.waters.com/514207>

MassLynx MS Software <https://www.waters.com/513662>

720002809, April 2009