# CHARACTERIZATION OF IMPURITIES AND MODIFICATIONS IN PROTEIN PRODUCTS USING LIQUID CHROMATOGRAPHY AND DATA INDEPENDENT ACQUISITION TANDEM MASS SPECTROMETRY

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

*Hongwei Xie, Martin Gilar, and John C. Gebler*
*Waters Corporation, 34 Maple Street, Milford, MA 01757*

## INTRODUCTION

**Heterogeneity, and sequence modifications such as oxidation and deamidation are common in recombinant protein products. They have the potential to affect the safety, activity and stability of therapeutic protein drugs. Sensitive methods for effective monitoring such modifications and controlling protein quality (purity) are required.**

**We have applied Ultra Performance Liquid Chromatography–Data Independent Acquisition Tandem Mass Spectrometry (UPLC-MS$^E$) approach for analysis of peptide maps. Yeast enolase and alcohol dehydrogenase (ADH) digest were chosen to demonstrate the proof of concept for identification and quantification of low-level impurities and modifications in protein products using such an approach.**

## FEATURES OF UPLC-MS$^E$

**UPLC**

- **Improved peptide resolution**
- **Improved detection sensitivity**
- **Improved speed and efficiency**

**MS$^E$**

- **Acquiring MS/MS data in parallel**
- **Data independent acquisition (DIA)**
- **Unbiased sampling of low-abundant peptides**

### Identified Partially/Non-tryptic Peptides in Enolase

| Protein | 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** |
| 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 | **E2P1** | AILGVSMAAAR | 109 | 11 | Partial | **21.89** |
| Eno2p | **E2P2*** | VLNGGSHAGGALALQEFMIAPTGAK | 153 | 25 | Partial | **28.42** |
| Eno2p | **E2N1*** | VLNGGSHAGGALALQEF | 153 | 17 | Non | **25.25** |
| Sod1p | **S1P1** | AGVSGVVK | 11 | 8 | Partial | **13.25** |
| Sod1p | **S1P2** | IHAGQDDLGK | 119 | 10 | Partial | **15.39** |
| Sod1p | **S1P3** | VQAVAVLK | 1 | 8 | Partial | **17.35** |

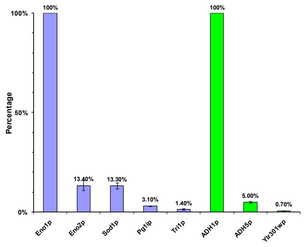
* N marked in red was deaminated.

### Identified Target Proteins and Impurities

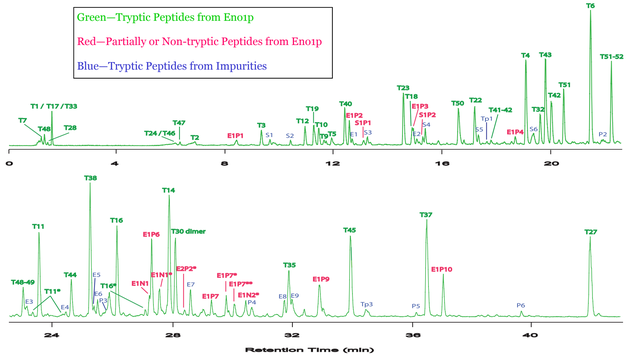
| Sample Name | Protein Name | Identified Tryptic Peptides | Intensity_ 3most x $10^3$ |
|---|---|---|---|
| **Enolase Digest** | **Enolase 1; Eno1p** | **42*** | **464.0 35** |
| | Enolase 2; Eno2p | 9 | 61.7 ± 5 |
| | Cu-Zn Superoxide Dismutase; Sod1p | 7 | 61.5 ± 2.3 |
| | Glucose-6-phosphate Isomerase; Pgi1p | 6 | 14.4 ± 1.1 |
| | Triosephosphate Isomerase; Tpi1p | 3 | 6.3 ± 0.7 |
| **ADH Digest** | **Alcohol Dehydrogenase 1; ADH1p** | **40*** | **390.9 ± 9.5** |
| | Alcohol Dehydrogenase 5; ADH5p | 3 | 19.5 ± 1.2 |
| | Ylr301wp | 5 | 2.9 ± 0.4 |

* Including modified tryptic peptides and tryptic peptides with 1 miscleavage.

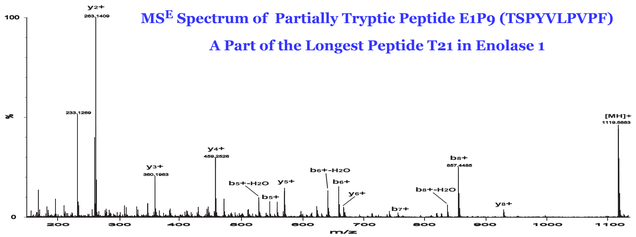
### Relative Concentration of Impurities (normalized, target protein = 100%)




### Complete LC Peak Assignment of Enolase Digest




### MS$^E$ Spectrum of Partially Tryptic Peptide E1P9 (TSPYVLPVPF) A Part of the Longest Peptide T21 in Enolase 1

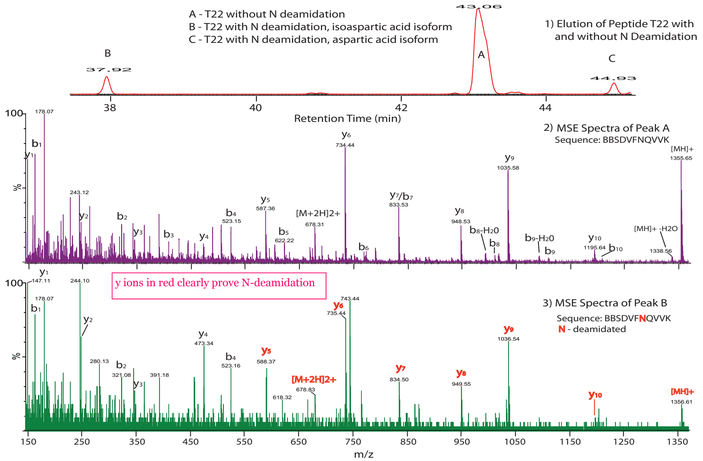



### Modification Type, Site and Stoichiometry of Modified Peptides Identified from Eno1p and ADH1p

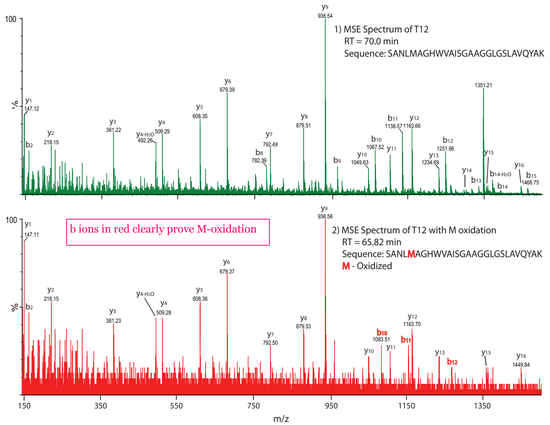
| Protein | Peptide | Start | End | Modification Type | Sequence[1] & Modification Site | Cleavage[2] | MH+ | RT (min) | Stoichiometry ± SD[3] |
|---|---|---|---|---|---|---|---|---|---|
| **Eno1p** | T11 | 67 | 78 | Deamidation N69 | NVNDVIAPAFVK | Fully | 1286.6869 | 23.33 | 6.97 ± 0.32 |
| | T11 | 67 | 78 | Deamidation N69 | NVNDVIAPAFVK | Fully | 1286.6869 | 24.4 | 1.94 ± 0.15 |
| | T16 | 105 | 119 | Deamidation N108 | LGANAILGVSLAASR | Fully | 1412.7986 | 25.95 | 3.92 ± 0.16 |
| | T16 | 105 | 119 | Deamidation N108 | LGANAILGVSLAASR | Fully | 1412.7986 | 26.95 | 1.37 ± 0.07 |
| **ADH1p** | T1 | 1 | 7 | N-terminal Acetylation | SIPETQK | Fully | 844.4411 | 25.07 | 99.15 ± 0.07 |
| | T7 | 60 | 80 | Oxidation M75 | LPLVGGHEGAGVVVGMGENVK | Fully | 2035.0641 | 46.59 | 1.8 ± 0.28 |
| | T12 | 164 | 191 | Oxidation M168 | SANLMAGHWVAISGAAGGLGSLAVQYAK | Fully | 2716.3875 | 65.82 | 1.65 ± 0.35 |
| | T21 | 261 | 275 | Deamidation N262 + Oxidation M270 | ANGTTVLVGMPAGAK | Fully | 1403.7198 | 34.03 | 1.0 ± 0.0 |
| | T21 | 261 | 275 | Deamidation N262 + Oxidation M270 | ANGTTVLVGMPAGAK | Fully | 1403.7198 | 43.99 | 8.85 ± 1.34 |
| | T21 | 261 | 275 | Deamidation N262 | ANGTTVLVGMPAGAK | Fully | 1387.725 | 43.97 | 84.1 ± 1.70 |
| | T21 | 261 | 275 | Deamidation N262 | ANGTTVLVGMPAGAK | Fully | 1387.725 | 44.27 | 2.95 ± 0.35 |
| | T5 | 30 | 38 | Deamidation N31 | ANELLINVK | Fully | 1014.583 | 43.54 | 1.35 ± 0.07 |
| | T5 | 30 | 38 | Deamidation N31 | ANELLINVK | Fully | 1014.583 | 47.18 | 1.6 ± 0.0 |
| | T22 | 276 | 286 | Deamidation N282 | BBSDVFNQVVK | Fully | 1356.5923 | 37.93 | 4.8 ± 0.0 |
| | T22 | 276 | 286 | Deamidation N282 | BBSDVFNQVVK | Fully | 1356.5923 | 44.95 | 3.3 ± 0.0 |
| | P1 | 92 | 128 | Deamidation N94 | WLNGSBMABEYBELGNESNBPHADLSGYTHDGSFQQY | Partially | 4358.689 | 59.67 | 100[4] |
| | P2 | 92 | 121 | Deamidation N94 | WLNGSBMABEYBELGNESNBPHADLSGYTH | Partially | 3533.3596 | 55.82 | 100 |
| | P3 | 92 | 119 | Deamidation N94 | WLNGSBMABEYBELGNESNBPHADLSGY | Partially | 3295.2529 | 58.8 | 100 |
| | P4 | 92 | 113 | Deamidation N94 | WLNGSBMABEYBELGNESNBPH | Partially | 2688.988 | 52.51 | 100 |
| | P5 | 92 | 98 | Deamidation N94 | WLNGSBM | Partially | 868.3328 | 42.63 | 100 |

[1] B - Carbamidomethyl C; Amino acid marked in red is with modification; Each deamidated N results in two isoforms, isoaspartic acid and aspartic acid.
[2] Fully - Fully Tryptic; Partially - Partially Tryptic.
[3] In percentage (%) , calculated by
**Intensity of the Modified Peptide / (Intensity of the Modified Peptide + Intensities of Related Unmodified and Other Modified Peptides)**
SD - Standard Deviation.
[4] Identified partially tryptic peptides with N94 deamidation from the longest peptide T10 (28 amino acids, M.W. 7601.4);
No corresponding peptides without N94 deamination was identified.

### Elution Pattern and MS$^E$ Spectra of Peptide T22 in ADH1p Before and After N-deamidation




### MS$^E$ Spectra of Peptide T12 in ADH1p Before and After M-oxidation




## CONCLUSIONS

1. **Complete assignment of all peaks in UPLC-MS$^E$ peptide maps.**
2. **Identification of expected as well as unexpected peptides in protein tryptic digest.**
3. **Identity establishment for target proteins with high sequence coverage (97% for Enolase 1, 98% for ADH1).**
4. **Identification and quantification of low-level protein contaminants; LOD of contaminants is ~0.5%.**
5. **Separation, identification and quantification of protein modifications (e.g. M-oxidation and N-deamidation) with stoichiometry as low as 1%.**
6. **De novo sequencing of LC major and minor peaks is possible (using MS$^E$ spectra).**
7. **High throughput and increased speed of analysis.**

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

720002792EN

©2008 Waters Corporation