Anal. Chem. **2009,** *81,* 5699–5708

Characterization of Protein Impurities and Site-Specific Modifications Using Peptide Mapping with Liquid Chromatography and Data Independent Acquisition Mass Spectrometry

Hongwei Xie,* Martin Gilar, and John C. Gebler

Biopharmaceutical Sciences, Waters Corporation, 34 Maple Street, Massachusetts 01757

Liquid chromatography (LC)-based peptide mapping is extensively used for establishing protein identity, assessing purity, and detecting post-translational modifications (PTMs) of recombinant proteins in the biopharmaceutical industry. However, current LC-UV/MS peptide mapping methods require multiple analyses and MS/MS experiments to identify protein contaminants and site-specific PTMs. This manuscript evaluated an alternative approach for protein characterization via peptide mapping employing a data independent LC-MS acquisition strategy with an alternate low and elevated collision energy scanning. The acquired peptide precursor and fragment information was utilized for effective identification of peptide sequences and site-specific modifications within a single LC run. The peptide MS signal intensities were reliably measured and used to estimate relative concentrations of PTMs and/or proteins contaminating the target protein. The method was evaluated using tryptic digests of yeast enolase and alcohol dehydrogenase. LC-eluted peptides were successfully sequenced and covered 97% target protein sequences. Protein impurities and site-specific modifications (e.g., M-oxidation and N-deamidation) were identified and quantified.

Sequence variants and posttranslational modifications (PTMs) such as glycosylation, deamidation, and oxidation are common in recombinant protein pharmaceuticals.[1−4] The modified motifs potentially affect the activity, stability, and safety of protein drugs. Effective control of these variants requires sensitive and reproducible methods for protein production monitoring.[4−7]

Liquid chromatography (LC)-based peptide mapping is a key method for protein structure characterization and purity analysis.[5,7−10] However, the ultraviolet (UV) or mass spectrometry (MS) detectors in traditional LC-UV/MS peptide mapping methods have limitations for identification of peptide sequences. Although these methods are suitable for detection of low-level protein contaminants or site-specific modifications, they cannot unambiguously confirm the peptide identity. Additional tandem mass spectrometry (either data dependent acquisition (DDA)-MS/MS or targeted MS/MS) measurements are required for elucidation of peptide sequence.

Peptide maps are often complicated by the presence of peptides resulting from unexpected proteolytic cleavages (e.g., partially tryptic and nontryptic peptides), which makes the protein characterization more challenging. If the LC-eluted peptides are not effectively identified, the sequence coverage may suffer and some protein PTMs may be missed.

Intact protein MS analysis has been used as an alternative to peptide mapping for protein characterization. Despite the recent progress of high-resolution MS-based mass measurements and "top-down" structural analysis of intact proteins,[11−17] LC-MS analysis of intact proteins has limitations. The determination of PTMs with small mass shift (e.g., +0.98 Da for deamidation) at intact protein level is a challenging task even for state-of-the-art MS instruments.[18] This is particularly the case for large proteins such as monoclonal antibodies. Therefore, the "bottom-up" based peptide mapping methods remain to be used for protein characterization.

* To whom correspondence should be addressed. Tel: 508-4823670. Fax: 508-4823085. E-mail: hongwei_xie@waters.com.

(1) Axford, J. *Trends Immunol.* **2001**, *22*, 237–239.
(2) Robinson, N. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5283–5288.
(3) Barelli, S.; Canellini, G.; Thadikkaran, L.; Crettaz, D.; Quadroni, M.; Rossier, J. S.; Tissot, J.-D.; Lion, N. *Proteomics* **2008**, *2*, 142–157.
(4) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. *Nat. Methods* **2007**, *4*, 798–806.
(5) Srebalus Barnes, C. A.; Lim, A. *Mass Spectrom. Rev.* **2007**, *26*, 370–388.
(6) Chelius, D.; Rehder, D. S.; Bondarenko, P. V. *Anal. Chem.* **2005**, *77*, 6004–6011.
(7) Chen, G.; Pramanik, B. N. *Expert Rev. Proteomics* **2008**, *5*, 435–444.
(8) Dougherty, J.; Mhatre, R.; Moore, S. *BioPharm. Intern.* **2003**, 54–58.
(9) Bongers, J.; Cummings, J. J.; Ebert, M. B.; Federici, M. M.; Gledhill, L.; Gulati, D.; Hilliard, G. M.; Jones, B. H.; Lee, K. R.; Mozdzanowski, J.; Naimoli, M.; Burman, S. *J. Pharm. Biomed. Anal.* **2000**, *21*, 1099–1128.
(10) Kannan, K.; Mulkerrin, M. G.; Zhang, M.; Gray, R.; Steinharter, T.; Sewerin, K.; Baffi, R.; Harris, R.; Karunatilake, C. *J. Pharm. Biomed. Anal.* **1997**, *16*, 631–640.
(11) Bogdanov, B.; Smith, R. D. *Mass Spectrom. Rev.* **2005**, *24*, 168–200.
(12) Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spectrom. Rev.* **2005**, *24*, 201–222.
(13) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233–6237.
(14) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 197A–203A.
(15) Parks, B. A.; Jiang, L.; Thomas, P. M.; Wenger, C. D.; Roth, M. J.; Boyne, M. T., 2nd; Burke, P. V.; Kwast, K. E.; Kelleher, N. L. *Anal. Chem.* **2007**, *79*, 7984–7991.
(16) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528–9533.
(17) Zubarev, R. A. *Curr. Opin. Biotechnol.* **2004**, *15*, 12–16.
(18) Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212–217.

10.1021/ac900468j CCC: $40.75
Published on Web 06/11/2009

Recently, data independent acquisition LC-MS called LC-MS^E has been developed for proteomics studies.[19−21] The application of this approach for simultaneous qualitative and quantitative information of a proteome was demonstrated on an *Escherichia coli* sample.[19] Chakraborty and colleagues have adapted LC-MS^E for analyzing peptide maps of a protein digest.[22] High sequence coverage (>90%) and good analytical reproducibility were achieved in replicate analyses of protein digest.

Data independent acquisition LC-MS utilizes an alternate MS scanning with low and elevated collision energy settings. Both peptide precursor and fragmentation data are collected within a single LC-MS experiment.[20,21] The acquired data are suitable for the identification of primary sequences of LC-eluted peptides.[23] The unbiased nature of the acquisition[24] overcomes the limitations of DDA LC-MS/MS experiments in terms of repeatability[25] and ensures the sampling of low-abundance peptides, especially when analyzing complex samples. Another advantage of data independent acquisition LC-MS over DDA LC-MS/MS is suitable for reliable quantification using MS data (MS signal intensities) unbiasedly collected from all detected chromatographic peaks.[21,24] In addition, the method is potentially able to detect low-level protein contaminants and substoichiometric PTMs present in the target protein digest.

In this work we evaluate an alternative peptide mapping approach with data-independent, alternate scanning MS^E approach for characterization of peptide maps in a single LC-MS experiment. The goal was to identify peptides including unexpected modifications such as partially/nontryptic peptides and PTMs (for example, asparagine (N)-deamidation and methionine (M)-oxidation, two of the most common covalent modifications in therapeutic proteins). The method was utilized for both characterization and quantification purposes.

MATERIALS AND METHODS

Materials. Yeast enolase and alcohol dehydrogenase (ADH), iodoacetamide (IAM), dithiothereitol (DTT), and ammonium bicarbonate (NH_4HCO_3) were purchased from Sigma Chemical Co. (St. Louis, MO), sequence-grade trypsin from Promega Corp. (Madison, WI), Formic acid (FA) from EM Sciences (Gibbstown, NJ), Optima-grade acetonitrile (ACN) from Fisher Scientific (Pittsburgh, PA), and RapiGest SF from Waters (Milford, MA) were used. The water used in all procedures was supplied by a Millipore Milli-Q purification system (Bedford, MA).

Preparation of Peptide Mixtures. Yeast enolase or ADH was prepared to 5 μg/μL with 100 mM NH_4HCO_3 containing 0.1% RapiGest SF. Aliquots of 50 μL were used for digestion. Before digestion, 5 μL 100 mM DTT was first added to reduce disulfide bonds by incubating at 60 °C for 30 min. The resulting free cysteine residues were alkylated by adding 5 μL of 200 mM IAM and incubating at room temperature for 30 min in the dark. Digestion was accomplished by adding 5 μg of sequencing-grade trypsin and incubating at 37 °C overnight. Formic acid (0.5% v/v) was used to degrade RapiGest SF and quench enzymatic reactions. The procedures for reducing disulfide bonds and alkylated free cysteine residues were omitted for the enolase sample. After adding ACN (to 30% v/v), the digests were diluted to 1.5 pmol/μL with 5% ACN/95% water containing 0.1% FA prior to LC-MS^E analysis.

LC-MS Experiments. All analyses were performed using a SYNAPT MS system and controlled by MassLynx 4.1 software. An ACQUITY UPLC system was used for the online reversed-phase separation of the tryptic digests. The separation was performed on a 2.1 × 100 mm BEH300 Å 1.7 μm Peptide Separation Technology column at 40 °C. Mobile phase A was aqueous 0.1% FA and B was 0.1% FA in ACN. The enolase digest was eluted with a 60-min gradient (0−50% B) and a 90-min gradient (0−50% B) employed for the ADH digest. Flow rate was 0.2 mL/min. An auxiliary pump was used to spray a solution of 100 fmol/μL Glu1-fibrinopeptide B (GFP) in 50/50 ACN/water containing 0.1% FA for mass accuracy reference (lockmass channel), with a flow rate of 25 μL/min.

MS acquisition was operated in the positive ion V-mode. An alternating low-energy (collision cell energy 5 V) and elevated-energy (collision cell energy ramped from 20 to 40 V) acquisition was used to obtain the precursor ions (MS) and their fragmentation data (MS^E), respectively. Scan time was 0.5 s (1 s total duty cycle). A 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 using a 100 fmol/μL GFP infusion.

Sequence Database Searching and Data Analysis. The acquired data were processed by Identity^E software of ProteinLynx Global Server (PLGS).[23,26] MS data were background-subtracted, deisotoped, and charge-state-reduced to the corresponding monoisotoped, 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 containing all 6139 protein entries with trypsin specificity and one potential miscleavage for impurity check. Then, the data were searched again against the sequence of target proteins (enolase 1 or ADH1) with no enzyme specified. N-deamidation (+0.98 Da), M-oxidation (+15.99 Da), N-terminal acetylation (+42.01 Da), and C-carbamidomethylation (57.02 Da) were allowed as variable modifications in these searches.

The search parameters were set as follows: (i) precursor monoisotopic ion intensity greater than 200 counts and mass tolerance less than 10 ppm; (ii) fragment monoisotopic ion intensity greater than 75 counts and mass tolerance less than 20

(19) Silva, J. C.; Denny, R.; Dorschel, C.; Gorenstein, M. V.; Li, G. Z.; Richardson, K.; Wall, D.; Geromanos, S. J. *Mol. Cell. Proteomics* **2006**, *5*, 589–607.
(20) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G. Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. *Anal. Chem.* **2005**, *77*, 2187–2200.
(21) Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J. *Mol. Cell. Proteomics* **2006**, *5*, 144–156.
(22) Chakraborty, A. B.; Berger, S. J.; Gebler, J. C. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 730–744.
(23) Li, G. Z.; Vissers, J. P.; Silva, J. C.; Golick, D.; Gorenstein, M. V.; Geromanos, S. J. *Proteomics* **2009**, *9*, 1696–1719.
(24) Geromanos, S. J.; Vissers, J. P.; Silva, J. C.; Dorschel, C. A.; Li, G. Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I. *Proteomics* **2009**, *9*, 1683–1695.
(25) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. *Mol. Cell. Proteomics* **2004**, *3*, 311–326.
(26) Skilling, J.; Denny, R.; Richardson, K.; Young, P.; McKenna, T.; Campuzano, I.; Ritchie, M. *Comp. Funct. Genomics* **2004**, *5*, 61–68.

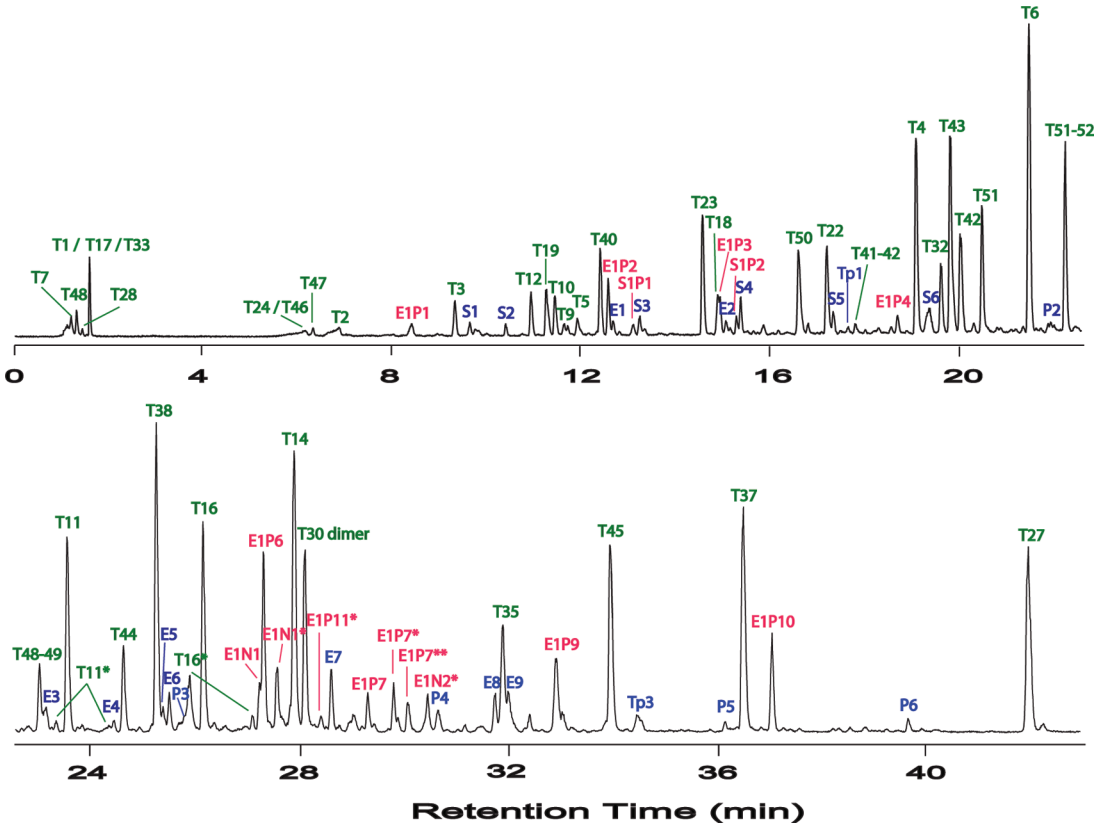

Figure 1. LC-MS chromatogram (TIC) of enolase tryptic digest. Tryptic peptides (Txx) of enolase 1 (Eno1p) are annotated in green. Tryptic peptides from enolase 2 (Eno2p), Cu−Zn superoxide dismutase (Sod1p), glucose-6-phosphate isomerase (Pgi1p), and triosephosphate isomerase (Tpi1p) are annotated in blue and with a corresponding label Ex, Sx, Px, and Tpx, respectively. Partially tryptic and nontryptic peptides are labeled in red, with E1Px representing partially tryptic peptide from Eno1p; E1Nx for nontryptic peptide from Eno1p; E2Px for partially tryptic peptide from Eno2p; and S1Px for partially tryptic peptide from Sod1p. The asterisk indicates peptide with N-deamidation. For details, see Tables 2−4.

ppm. A match was accepted when at least three fragment ions were identified for the peptide.

RESULTS AND DISCUSSION

Peptide Identification in Data Independent LC-MS Chromatogram. The features and operational aspects of data independent, alternate scanning LC-MS have been previously described.[20−22] Briefly, two sets of data are collected in parallel during LC-MS^E acquisition: low collision energy and elevated collision energy chromatograms. While low collision energy LC-MS (peptide precursors) data are used for quantitation, the combined information in low collision energy and elevated collision energy data (peptide fragments) serves for peptide sequence identification.[23,26]

In this study, we analyzed tryptic digests of two standard proteins, enolase and ADH. The freshly prepared digests were analyzed by data independent LC-MS. Each sample was analyzed in triplicate which yielded essentially identical chromatograms. The employed LC and MS method provided highly reproducible data (data not shown) similar to those reported earlier.[22]

Figure 1 shows a LC-MS total ion chromatogram (TIC) of 30 pmol of enolase tryptic digest injected on column. The chromatogram features over 100 resolved and detected peaks. The identification and sequencing of peptides was accomplished by software which utilizes data from both low collision energy LC-MS and elevated collision energy LC-MS.[20−22] The software reconstructed MS/MS spectra of all detected precursor ions in Figure 1 were searched against a protein sequence database to establish protein identities (including target protein and potential contaminants). Since the LC-MS^E method resulted in a comprehensive data set, the resulting data analysis also yielded high peptide map coverage of enolase.

The data were initially searched against the yeast protein database, because the enolase sample originates from baker's yeast. The strict tryptic specificity was used to enhance the search speed and minimize potential for false positive identifications.[27] Forty-one peptides were identified to belong to enolase 1 (Eno1p) protein, along with three miscleaved peptides and two N-deamidated peptides (two isoforms each). The number of identified peptides is given in Table 1. The peptide sequences, masses, MS intensities, and retention times are listed in Table 2. The average root mean square (rms) mass error of these identified peptides was 4.9 ppm.

The search also identified tryptic peptides belonging to four additional yeast proteins (see Table 1; sequences are shown in Table 3). Each protein was identified by numerous unique peptides. Nine, seven, six, and three unique tryptic peptides originated from enolase 2 (Eno2p), Cu−Zn superoxide dismutase (Sod1p), glucose-6-phosphate isomerase (Pgi1p), and triosephosphate isomerase (TPi1p), respectively. As a result of the sequence

(27) Xie, H.; Griffin, T. J. *J. Proteome Res.* **2006**, *5*, 1003–1009.

Table 1. Identified Proteins from Enolase and ADH Samples

sample name	protein name	PLGS score[a]	peptides	Sum_3_Intensity[b]
enolase digest	enolase 1; Eno1p	3246.1	42	464.0 ± 35
	enolase 2; Eno2p	2384.5	9[c]	61.7 ± 5
	Cu−Zn superoxide dismutase; Sod1p	485.4	7	61.5 ± 2.3
	glucose-6-phosphate ispmerase; Pgi1p	376.4	6	14.4 ± 1.1
	triosephosphate isomerase; Tpi1p	140.7	3	6.3 ± 0.7
ADH digest	ADH1; ADH1p	3237	40	390.9 ± 9.5
	ADH5; ADH5p	132.4	3	19.5 ± 1.2
	Ylr301wp	195	5	2.9 ± 0.4

[a] Identity^E database search score. [b] Intensity sum of the three most intense tryptic peptides, $\times 10^3$, an average from three replicate analyses. [c] Identified unique tryptic peptides.

homology, some tryptic peptides did not distinguish between Eno1p and Eno2p proteins; these peptides were all assigned to Eno1p.

The identified tryptic peptides were annotated in the LC-MS chromatogram shown in Figure 1. Eno1p peptides are in green, while the unique peptides from other proteins are labeled in blue. The Eno1p peptides are the most dominant peaks in the chromatogram, while the remaining minor peptide peaks belong to the other four proteins. It appears that the Eno1p sample we used in this experiment is contaminated with a small amount of protein impurities.

After labeling the identified tryptic peptides, some significant peaks in Figure 1 remained unknown. This is a typical scenario in peptide mapping. An unexpected proteolytic cleavage or modification may produce artificial peptides that are difficult to identify from LC-MS data even with using high resolution MS instruments. In a conventional LC-MS experiment, additional targeted MS/MS analyses are required to elucidate the sequence information of those unknown peaks. The benefit of data independent LC-MS data acquisition is that the existing MS data contain precursor and fragment information for all peaks, and additional data mining can supply the desirable information.

The acquired data were searched again, allowing for nonspecific tryptic cleavages, against a truncated database consisting of only the proteins identified earlier. The rationale for using the truncated database is to avoid potential false positive hits from a large-scale sequence database search.[28] We reasonably assume that force hits (if any exists) of partially and nontryptic peptides for the limited number of identified proteins from the truncated database searching should be much less than the potential false positives of partially and nontryptic peptides from thousands of proteins in large-scale proteome database searching. Therefore, this approach avoids both large potential of false partially/nontryptic hits and the resulting unnecessary large amount of manual confirmation work.

The search returned 21 confident peptide identifications, including 17 partially tryptic and 4 nontryptic peptides (Table 4). As expected, a majority (17) of these peptides originated from the target protein Eno1p.

The average rms mass error of the partially tryptic/nontryptic peptides was 4.4 ppm, which is comparable with that accuracy of earlier identified tryptic peptides (4.9 ppm), demonstrating the quality of the identifications. A manual inspection of the MS^E spectra confirmed the peptide identifications. Figure 2 illustrates an example of the MS^E spectrum of **TSPYVLPVPF**, a partially tryptic peptide originating from Eno1p tryptic peptide T21 (**TSPYVLPVPF**LNVLNGGSHAGGALALQEFMIAPTGAK, 37 amino acids, MW = 3736.97 Da). The T21 peptide was detected only as a minor peak in the LC chromatogram. It appears that tryptic peptide T21 was converted into a series of partially tryptic peptides E1P9, E1P7, E1P7*, E1P7**, E1P11*, E1P1, and E1P2 and nontryptic peptides E1N1, E1N1*, E1N2*, and E1N3* (see Table 4).

The partially and nontryptic peptides identified in the database search were annotated in Figure 1 (red label). Consistent with previous findings, the dominant partially/nontryptic peptides in LC-MS chromatogram originate from protein Eno1p. Four partially tryptic peptides from impurity proteins Eno2p and Sod1p were the other minor peaks.

The data independent LC-MS analysis and database searches allowed for identification of all significant peaks in the chromatogram, with the exception of a dominant peak eluting at 28.09 min. After manually examining the MS (Figure 3a) and MS^E (Figure 3b) spectra we concluded that this peak is a disulfide-linked Eno1p T30 dimer (T30-S-S-T30). The MS spectrum contained three dominant ions with monoisotopic mass 1315.6 Da (+2), 877.38 Da (+3), and 658.29 Da (+4), corresponding to the expected masses of double, triple, and quadruple charged precursor ions of T30-S-S-T30. The MS^E spectrum contains a series of y-ions consistent with T30 peptide sequence (as well as double and triple charged T30-S-S-T30 ions; see Figure 3b). Peptide T30 (**IGLDCASSEFFK**) contains the only cysteine residue present in the Eno1p sequence, and because the protein was not reduced and alkylated, it is not surprising that the T30 peptide formed a dimer. This is further supported by the fact that the monomeric T30 peptide was not detected in the chromatogram.

Despite the comprehensive analysis of data from the chromatogram in Figure 1, there are some minor peaks without assignment. The masses of these peptides belong to tryptic peptides of the four low-abundance impurity proteins. Because of low MS intensity, the MS^E spectra were not of sufficient quality to allow for good identification via a database search.

In addition to the enolase sample we performed data independent LC-MS analysis of yeast ADH trytic digest. Similarly, as the analysis demonstrated for enolase, tryptic, partially/nontryptic, and modified peptides were identified, including 40 tryptic peptides from ADH1 (ADH1p), 3 from ADH5 (ADH5p), and 5 from Ylr301wp (Table 1). The peptide sequences and other details are available in Supplemental Table 1 (Supporting Information). We observed N-terminal acetylation, three M-oxidations sites, and six N-deamination sites in ADH1, which will be discussed in detail later on.

Sequence Coverage. High sequence coverage (96%) was obtained for Eno1p protein using the data independent LC-MS method. The nonobserved peptides are short with less than three amino acids that were not well retained on the reversed-phase column. These short peptides (four with one amino acid residue,

(28) Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. *J. Proteome Res.* **2004**, *3*, 1082–1085.

Table 2. Fully Tryptic Peptides (Txx) Identified from Enolase 1 (Eno1p)

protein	peptide	sequence	theoretical mass (Da)	mass error (ppm)	intensity (counts)	RT (min)
Eno1p	T7	(R)DGDK(S)	433.1808	−2.3	6033	1.24
Eno1p	T48	(R)SER(L)	390.1863	−6.2	4388	1.33
Eno1p	T28	(K)AAGHDGK(V)	654.3085	−6.6	3392	1.45
Eno1p	T1	(-M)AVSK(V)	403.2431	−8.2	3464	1.6
Eno1p	T17	(R)AAAAEK(N)	559.2966	−2.7	6898	1.61
Eno1p	T33	(K)NPNSDK(S)	673.3031	−4.9	5087	1.61
Eno1p	T24	(K)SLTK(K)	447.2693	1.3	2410	6.13
Eno1p	T46	(R)TGQIK(T)	545.3173	−7.7	1878	6.19
Eno1p	T47	(K)TGAPAR(S)	571.3078	−2.1	3223	6.34
Eno1p	T2	(K)VYAR((S)	507.2805	−1.2	2331	6.87
Eno1p	T3	(R)SVYDSR(G)	725.3344	0	4896	9.34
Eno1p	T12	(K)ANIDVK(D)	658.365	0.2	8450	10.95
Eno1p	T19	(K)HLADLSK(S)	782.4286	−4.3	33522	11.28
Eno1p	T10	(K)GVLHAVK(N)	722.4439	−6.2	8835	11.46
Eno1p	T9	(K)WMGK(G)	520.2468	−3.8	5771	11.65
Eno1p	T5	(K)GVFR(S)	477.2699	3.4	8474	11.94
Eno1p	T40	(R)IATAIEK(K)	744.4382	−4.4	12839	12.42
Eno1p	T23	(R)IGSEVYHNLK(S)	1158.6033	−5.6	41086	14.58
Eno1p	T18	(K)NVPLYK(H)	732.4171	−7.6	7818	14.92
Eno1p	T50	(K)LNQLLR(I)	755.4653	−6.8	16957	16.62
Eno1p	T22	(K)TFAEALR(I)	806.4286	−3.2	14930	17.22
Eno1p	T41−42	(K)KAADALLLK(V)	941.5909	6.6	1178	17.83
Eno1p	T4	(R)GNPTVEVELTTEK(G)	1415.7146	4	116475	19.1
Eno1p	T32	(K)YDLDFK(N)	799.3752	−0.6	23065	19.63
Eno1p	T43	(K)VNQIGTLSESIK(A)	1287.7035	0.5	120477	19.83
Eno1p	T42	(K)AADALLLK(V)	813.496	−1.1	22574	20.05
Eno1p	T51	(R)IEEELGDNAVFAGENFHHGDK(L)	2327.0452	−6.1	29767	20.5
Eno1p	T6	(R)SIVPSGASTGVHEALEMR(D)	1839.915	−7.8	93821	21.48
Eno1p	T51−52	(R)IEEELGDNAVFAGENFHHGDKL(−)	2440.1292	−7.6	54950	22.26
Eno1p	T48−49	(R)SERLAK(L)	702.4024	1.7	1127	23.01
Eno1p	T11[a]	(K)NVN DVIAPAFVK(A)	1286.6869	2.6	6997	23.33
Eno1p	T11	(K)NVNDVIAPAFVK(A)	1285.703	−6.1	100270	23.55
Eno1p	T11[a]	(K)NVN DVIAPAFVK(A)	1286.6869	−0.7	1943	24.4
Eno1p	T44	(K)AAQDSFAAGWGVMVSHR(S)	1788.8365	−13	20866	24.62
Eno1p	T38	(K)TAGIQIVADDLTVTNPK(R)	1754.9415	−5.7	134860	25.27
Eno1p	T16[a]	(K)LGAN AILGVSLAASR(A)	1412.7986	0.3	4071	25.95
Eno1p	T16	(K)LGANAILGVSLAASR(A)	1411.8147	−4.3	103918	26.15
Eno1p	T16[a]	(K)LGAN AILGVSLAASR(A)	1412.7986	−2.4	1422	26.95
Eno1p	T14	(K)AVDDFLISLDGTANK(S)	1577.7938	−0.6	153207	27.88
Eno1p	T35	(K)WLTGPQLADLYHSLMK(R)	1871.9606	−2.1	37800	31.9
Eno1p	T45	(R)SGETEDTFIADLVVGLR(T)	1820.9155	−6.6	121847	33.96
Eno1p	T37	(R)YPIVSIEDPFAEDDWEAWSHFFK(T)	2827.2803	−8.8	126346	36.5
Eno1p	T27	(R)YGASAGNVGDEGGVAPNIQTAEEALDLIVDAIK(A)	3256.6091	−4.5	158967	42

[a] N is deamidated asparagine, with two isoforms: isoaspartic acid (eluting earlier) and aspartic acid (eluting later).

four with two amino acid residues, and two with three amino acid residues) were not detected in our experiment. They comprise the remaining 4% of uncovered sequence.

The other four proteins were low-level impurities; their sequence coverage calculated from identified unique peptides was 29.3%, 55.2%, 17.3%, and 23.0% for Eno2p, Sod1p, Pgi1p, and Tpi1p, respectively.

Similarly, ADH1 was identified to be the major protein (Table 1) in the ADH sample. Two different ADH1 sequences were reported in literature by Jörnvall[29,30] and by Bennetzen and colleagues[31] due to variants from different yeast strains. The protein sequences determined from the two different groups differ in 5 out of 347 amino acid residues (Y20, T58, E147, V151, and I313 reported by Jörnvall[29] and H20, V58, Q147, I151, and V313 reported by Bennetzen et al.[31]). After searching the database containing both ADH1 sequences, the results indicated that our ADH1 sample matched Jörnvall's sequence, with a 98% sequence coverage. As shown in Table 1, ADH5 and Ylr301wp were two low-level impure proteins found in our ADH sample. The identified unique peptides covered 9.4% and 29.1% sequences of ADH5 and Ylr301wp, respectively.

Data independent acquisition LC-MS method was well suited for distinguishing between the different candidate ADH1 sequences. The method is useful for identifying the target protein sequence from the list of homologues, as well as indicating the presence of additional impurity proteins in a protein product. High sequence coverage was useful in determining sequence variants and site-specific modifications.

Relative Concentration of Impurity Proteins. In addition to identification, the data independent acquisition LC-MS method provided information about the relative concentration of protein impurities. The peak intensities in the LC-MS chromatogram (Figure 1) correlate to the protein relative abundance in the sample. As protein concentration decreases, the total number of observed tryptic peptides and their corresponding signal decrease in a predictable fashion.[21]

(29) Jörnvall, H. *Eur. J. Biochem.* **1977**, *72*, 443–452.
(30) Wills, C.; Jornvall, H. *Eur. J. Biochem.* **1979**, *99*, 323–331.
(31) Bennetzen, J.; Hall, B. *J. Biol. Chem.* **1982**, *257*, 3018–3025.

Table 3. Unique Tryptic Peptides Identified from Proteins Contaminating Enolase 1 (see Table 1)

protein name	peptide	sequence	theoretical mass (Da)	mass error (ppm)	intensity (counts)	RT (min)
enolase; Eno2p	E1	(R)IEEELGDK(A)	932.4571	1.5	1743	12.69
enolase; Eno2p	E2	(K)TFAEAMR(I)	825.3923	−6.0	1885	15.08
enolase; Eno2p	E3	(K)NVPLYQHLADLSK(S)	1497.806	4.1	3112	23.13
enolase; Eno2p	E4	(K)AAQDSFAANWGVMVSHR(S)	1846.8654	−7.2	1615	24.43
enolase; Eno2p	E5	(K)LGANAILGVSMAAAR(A)	1414.7834	−1.9	12168	25.50
enolase; Eno2p	E6	(K)TAGIQIVADDLTVTNPAR(I)	1854.9919	−0.1	23451	25.88
enolase; Eno2p	E7	(K)AVDDFLLSLDGTANK(S)	1578.801	−1.7	26615	28.61
enolase; Eno2p	E8	(K)GVMNAVNNVNNVIAAAFVK(A)	1945.0323	6.5	6637	31.74
enolase; Eno2p	E9	(K)WLTGVELADMYHSLMK(R)	1893.9237	−5.3	4261	32.01
Cu, Zn superoxide dismutase; Sod1p	S1	(K)THGAPTDEVR(H)	1082.5225	1.0	2658	9.65
Cu, Zn superoxide dismutase; Sod1p	S2	(R)HVGDMGNVK(T)	956.4619	6.3	6222	10.41
Cu, Zn superoxide dismutase; Sod1p	S3	(K)GDAGVSGVVK(F)	888.4785	−3.0	2174	13.24
Cu, Zn superoxide dismutase; Sod1p	S4	(R)SVVIHAGQDDLGK(G)	1338.7012	−6.4	6395	15.38
Cu, Zn superoxide dismutase; Sod1p	S5	(−M)VQAVAVLK(G)	827.5349	−2.2	8917	17.35
Cu, Zn superoxide dismutase; Sod1p	S6	(K)LIGPTSVVGR(S)	998.5993	0.1	34407	19.38
Cu, Zn superoxide dismutase; Sod1p	S7	(K)FEQASESEPTTVSYEIAGNSPNAER(G)	2713.2224	−2.3	15963	19.90
glucose-6-phosphate isomerase; Pgi1p	P1	(K)TFTTAETITNANTAK(N)	1583.7911	9.9	2512	17.22
glucose-6-phosphate isomerase; Pgi1p	P2	(K)LATELPAWSK(L)	1115.6095	−1.0	2546	21.95
glucose-6-phosphate isomerase; Pgi1p	P3	(K)HYAGVLDVHFVSNIDGTHIAETLK(V)	2636.3467	−6.2	3081	25.90
glucose-6-phosphate isomerase; Pgi1p	P4	(K)VVDPETTLFLIASK(T)	1532.8571	3.5	6407	30.63
glucose-6-phosphate isomerase; Pgi1p	P5	(K)MLASNFFAQAEALMVGK(D)	1827.9132	2.3	2784	36.15
glucose-6-phosphate isomerase; Pgi1p	P6	(K)NLVNDEIIAALIELAK(E)	1738.9949	2.2	5446	39.67
triosephosphate isomerase; Tpi1p	Tp1	(K)ASGAFTGENSVDQIK(D)	1523.7336	5.1	2320	17.66
triosephosphate isomerase; Tpi1p	Tp2	(R)TFFVGGNFK(L)	1016.52	3.9	1570	23.07
triosephosphate isomerase; Tpi1p	Tp3	(K)DWTNVVVAYEPVWAIGTGLAATPEDAQDIHASIR(K)	3665.8235	6.1	1828	34.56

Table 4. Sequences and Retention Times of Identified Partially Tryptic and Non-Tryptic Peptides from Enolase Digest

protein	peptide	peptide sequence	theoretical mass (Da)	mass error (ppm)	intensity (counts)	cleavage[a]	RT (min)
Eno1p	E1P1	(A)PTGAK(T)	472.2645	−1.1	2501	partial	8.46
Eno1p	E1P2	(I)APTGAK(T)	543.3016	−1.6	25422	partial	12.59
Eno1p	E1P3	(R)YGASAGNVGDEGGVAPN(I)	1533.6695	−4.7	5093	partial	14.96
Eno1p	E1P4	(A)GWGVMVSHR(S)	1027.5021	−1.7	2853	partial	18.71
Eno1p	E1P5	(I)LGVSLAASR(A)	872.5079	−0.6	8413	partial	23.01
Eno1p	E1P6	(K)WLTGPQLADLYH(S)	1412.7088	1.7	93532	partial	27.30
Eno1p	E1P7	(F)LNVLNGGSHAGGALALQEFMIAPTGAK(T)	2636.3744	−4.4	7657	partial	29.31
Eno1p	E1P7[b]	(F)LNVL$\underline{N}$GGSHAGGALALQEFMIAPTGAK(T)	2637.3544	−8.3	15721	partial	29.80
Eno1p	E1P7[b]	(F)L$\underline{N}$VL$\underline{N}$GGSHAGGALALQEFMIAPTGAK(T)	2637.3544	−6.6	5711	partial	30.09
Eno1p	E1P8	(K)WLTGPQLADLYHSL(M)	1612.8249	2.8	13252	partial	31.99
Eno1p	E1P9	(K)TSPYVLPVPF(L)	1118.6011	−2.6	24893	partial	32.91
Eno1p	E1P10	(N)IQTAEEALDLIVDAIK(A)	1740.9509	5.6	52526	partial	37.05
Eno1p	E1P11[b]	(N)VL$\underline{N}$GGSHAGGALALQEFMIAPTGAK(T)	2410.2274	−7.6	2657	partial	28.42
Eno1p	E1N1	(F)LNVLNGGSHAGGALALQEF(M)	1866.9587	1.2	27040	non	27.23
Eno1p	E1N1[b]	(F)LNVL$\underline{N}$GGSHAGGALALQEF(M)	1867.9387	−1.4	32050	non	27.56
Eno1p	E1N2[b]	(F)LNVL$\underline{N}$GGSHAGGALALQEFM(I)	1998.9792	−1.4	22051	non	30.46
Eno1p	E1N3[b]	(N)VLNGGSHAGGALALQEF(M)	1639.8318	−5.0	2661	non	25.25
Eno2p	E2P1	(N)AILGVSMAAAR(A)	1058.5906	−2.1	1657	partial	21.89
Sod1p	S1P1	(D)AGVSGVVK(F)	715.4228	−2.0	2071	partial	13.25
Sod1p	S1P2	(V)IHAGQDDLGK(G)	1052.525	1.3	2069	partial	15.39
Sod1p	S1P3	(M)VQAVAVLK(G)	827.5349	−2.2	8917	partial	17.35

[a] Partial, partially tryptic; non, non-tryptic. [b] $\underline{N}$, deaminated asparagine.

Data independent LC-MS analysis provides a reliable quantification through precursor intensity measurements.[19,21] The data we collected does not suffer from bias of DDA MS/MS analysis gas phase preselection of candidate peptide precursors. An inventory of all the tryptic peptide components (precursor and fragments) above the limit of detection of the instrument is produced. The software employed provided an accurate-mass measured peptide ion list with integrated peptide MS signal intensities (deisotoped and charge state-reduced).[21] The MS signal intensity of identified peptides can be directly used to estimate the protein concentration as described by Silva and colleagues.[20,21] This is based on empirical observation[21] that the sum of the MS signal intensities of the three most dominant tryptic peptides closely correlates with relative concentration of protein present in the sample, regardless of the protein sequence. We adapted the method here to estimate the relative concentration of proteins present in our samples.

Figure 4 shows the relative concentration of the impurity proteins normalized to target protein in the sample (Eno1p or ADH1p), calculated by comparing the sum of intensities of the three most intense tryptic peptides (as shown in Table 1). The impurity proteins were present at levels ranging from 1% to 13%, assuming the concentration of target proteins is 100%. Contaminants below 1% (e.g., 0.7% for Ylr301wp) were successfully detected and quantified. By spiking a digest of known quantity of standard six-proteins with a serial of dilutions into the digest of

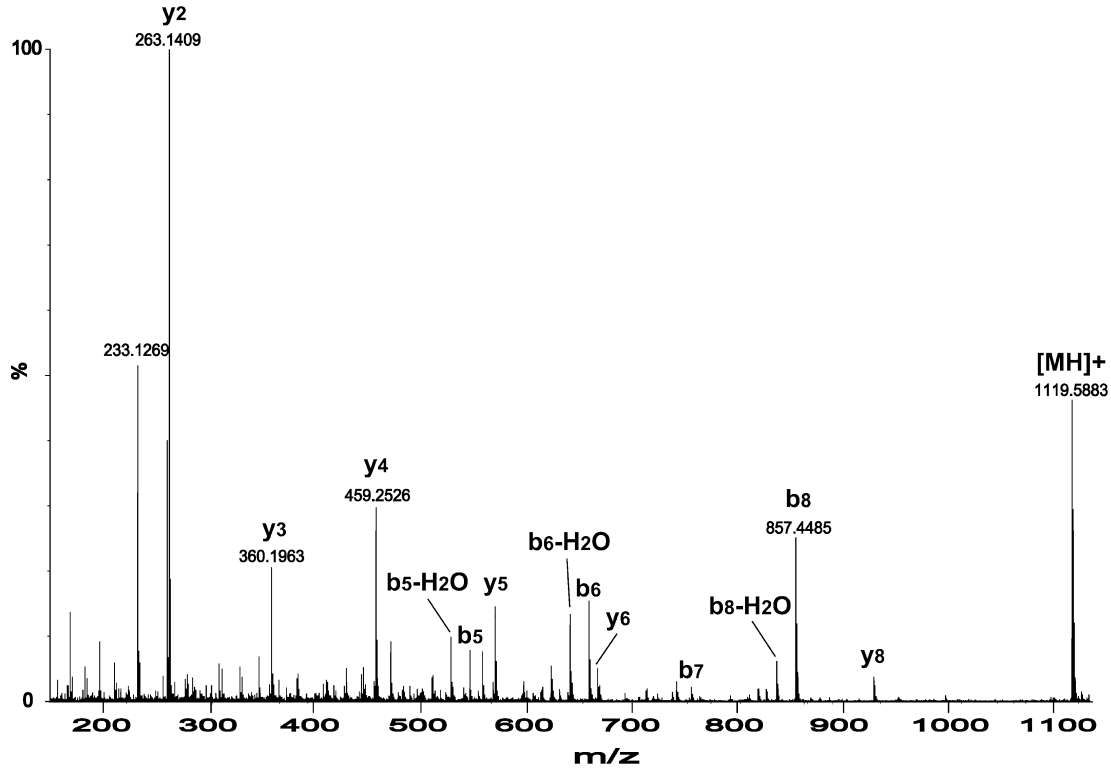

Figure 2. MS^E spectrum of partially tryptic peptide E1P9 (**TSPYVLPVPF**), a part of the unidentified tryptic peptide T21 (**TSPYVLPVPF**LNVLNGGSHAGGALALQEFMIAPTGAK, 37 amino acids, MW = 3736.97 Da) of enolase 1 (Eno1p).

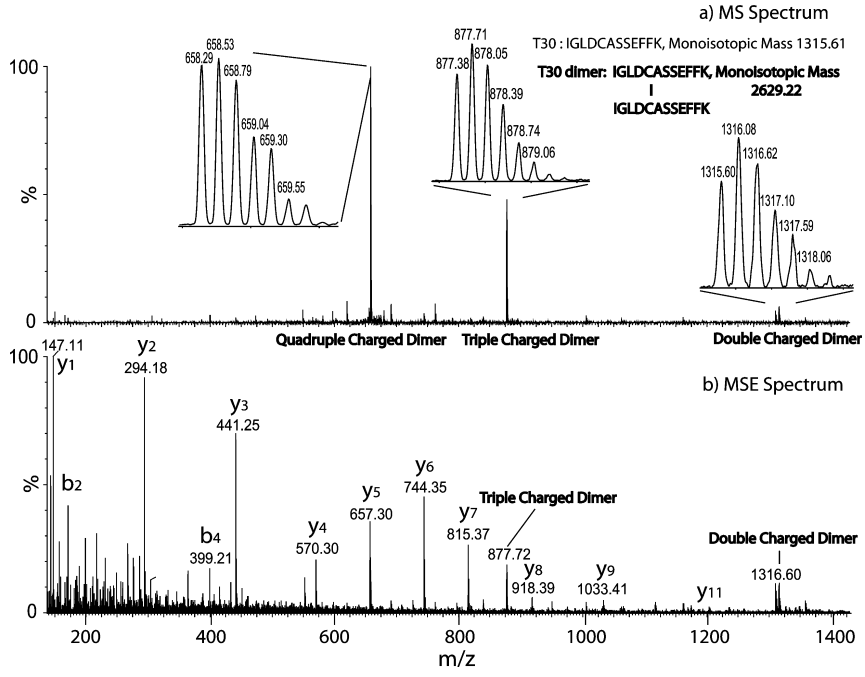

Figure 3. Low-energy and elevated-energy MS spectra of the peak eluting at 28.09 min: (a) MS spectrum, (b) MS^E spectrum. The monoisotopic molecular weight of 2629.2 determined by the +2, +3, and +4 ions in the MS spectrum and the y-series fragment ions of T30 in the MS^E spectra indicate that the peak is a disulfide-linked T30 dimer of enolase 1.

the protein mixture of interest, Silva et al.[21] demonstrated that relative quantification of proteins across 3−4 orders of magnitude in dynamic range was achieved for the method.

Since protein products contain a limited number of contaminating proteins and are much less complex than proteomics samples (e.g., tissue, cells, or biofluids), the peptides are generally well chromatographically resolved. Ion suppression of minor peptides by high-abundance ones plays a minimal role. This was useful for accurate quantification of low-level impurity proteins. The standard deviation (SD) values for quantitation of impurity proteins in enolase and ADH samples in three repetitive injections was below 10%.

The method described by Silva and colleagues[21] has limitations for quantification of homologues of target proteins, as they share

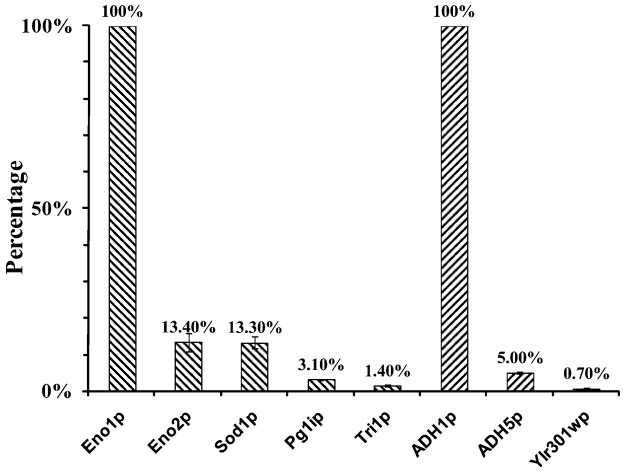

Figure 4. Relative concentration of impurity proteins found in enolase 1 or ADH1 samples. The concentration was normalized to the target protein (enolase 1 or ADH1).

common peptides, jointly contributing to the MS ion signal. This is the case for Eno2p protein contaminating the enolase sample or ADH5p in the ADH sample. Therefore, we compared the intensity of three unique tryptic peptides from each homologue. For example, the three most abundant unique tryptic peptides for Eno2p are TAGIQIVADDLTVTNP**AR**, AVDDFL**L**SLDGTANK, and LGANAILGVS**M**AAA**R**, while the corresponding tryptic peptides from Eno1p are TAGIQIVADDLTVTNPK, AVDDFLISLDGTANK, and LGANAILGVSLAASR (amino acid differences noted in bold). Because the peptides have similar sequences, we assume that each pair has nearly identical ionization efficiency. Therefore, their relative MS signal is illustrative of relative concentration between the proteins.

It has been reported that data independent LC-MS is an efficient and sensitive method for proteome analysis.[24] As a result of the parallel data acquisition of all isotopes and charge-states across the whole precursor ions chromatographic peak width, the precursor and fragment ion intensities exceed the signal acquired in typical DDA MS/MS analysis. Because of this feature, the data independent LC-MS is suitable for detection of low-abundance protein impurities in the sample as well as for analysis of substoichiometric PTMs of target proteins.

Detection and Quantitation of PTMs. Data independent LC-MS detects multiple sites of covalent modifications in analyzed protein digests. Four deamidation sites (N69, N108, N152, and N155; see Tables 2 and 4) were identified from Eno1p, while ADH1 (Jörnvall sequence)[29,30] contained eight modified amino acid sites located in seven tryptic peptides, including N-terminal acetylation, three oxidized methionines, and four deamidated asparagines. The modification types and sites, as well as retention times of identified modified peptides from ADH1, are listed in Table 5.

Figure 5 shows MS^E spectra of unmodified and M-oxidized peptides from ADH1. The b and y fragment ions directly confirm the modification site in the T12 peptide. Figure 6 illustrates a similar situation for ADH1 N-deamidation in the T22 peptide.

Retention time shift of peptides upon modifications can be used as an additional confirmation of protein modifications. The chromatographic retention of the N-terminal peptide T1 of ADH1 shifted from 16.37 to 25.07 min after N-terminal acetylation due to the increased hydrophobicity.[32] On the other hand, M-oxidation of a peptide reduces the hydrophobicity generally leading to reduction of retention time. For example, the peptide T7 shifted from 52.7 to 46.59 min after M75 oxidation, and T12 peptide retention decreased from 70.0 to 65.82 min upon M168 oxidation. The M-oxidation sites detected in our study were found to be susceptible to oxidation by an earlier report.[33]

Peptide N-deamidation results in formation of two different isoforms, isoaspartic acid, and aspartic acid. We observed two N-deamidated T22 peptides with N282 deamidation site, one eluting prior to and one after the unmodified peptide. The retention pattern is shown in Figure 6a. Because the modified peptides containing either isoaspartic acid or aspartic acid have an identical +0.98 Da mass difference from the unmodified T22 peptides, we were unable to differentiate the two deamidated isoforms by MS. On the basis of the literature[34−37] reports we assign the earlier eluting peak to the isoaspartic acid form as shown in Figure 6a. The y-series ions in the MS^E spectra (Figure 6c) clearly identify the N282 deamidation site in T22.

Since data independent LC-MS method provides for accurate MS intensities of each detected precursor in the chromatogram, it is able to estimate the relative stoichiometry of identified modifications. The relative quantitation of posttranslational modifications can be obtained from the relative ratio of MS signal intensities of modified and unmodified peptides providing that their MS ionization efficiencies are the same. We have tested the ionization efficiency of N-deaminated and M-oxidized peptides using synthetic peptides standards (see Supplemental Figure 1 and Supplemental Table 2, Supporting Information). We observed that N-deaminated peptides have essentially identical ionization efficiency as their unmodified counterparts. There was a measurable difference in ionization efficiency of selected M-oxidized and unmodified peptides; the former one provided ~10% greater MS signal. For the relative quantitation shown in Table 5 we assume equal ionization efficiency for all modifications.

Table 5 presents the estimated stoichiometry of modifications identified in the ADH sample. A majority (over 99%) of the N-terminal of ADH1 was acetylated. The oxidation level of methionines in the three identified oxidized peptides was lower, less than 2% for M75 and M168, and around 10% for M270. For peptide T21, about 10% M270 was oxidized, and nearly 97% N262 was deamidated (four isoforms as shown in Table 5). A high percentage of deamidation was also observed for the N94 site. The series of partially tryptic peptides (P1, P2, P3, P4, and P5; see Table 5) with N94 deamidation were identified, while no corresponding peptides without N94 deamidation were detected. This finding indicates that the unmodified counterparts were below the limit of detection. It is known that N sites with the -XNG-

(32) Xie, H.; Bandhakavi, S.; Roe, M. R.; Griffin, T. J. *J. Proteome Res.* **2007**, *6*, 2019–2026.
(33) Men, L.; Wang, Y. *J. Proteome Res.* **2007**, *6*, 216–225.
(34) Stephenson, R. C.; Clarke, S. *J. Biol. Chem.* **1989**, *264*, 6164–6170.
(35) Sadakane, Y.; Yamazaki, T.; Nakagomi, K.; Akizawa, T.; Fujii, N.; Tanimura, T.; Kaneda, M.; Hatanaka, Y. *J. Pharm. Biomed. Anal.* **2003**, *30*, 1825–1833.
(36) Johnson, B. A.; Shirokawa, J. M.; Hancock, W. S.; Spellman, M. W.; Basa, L. J.; Aswad, D. W. *J. Biol. Chem.* **1989**, *264*, 14262–14271.
(37) Harris, R. J.; Kabakoff, B.; Macchi, F. D.; Shen, F. J.; Kwong, M.; Andya, J. D.; Shire, S. J.; Bjork, N.; Totpal, K.; Chen, A. B. *J. Chromatogr. B, Biomed. Sci. Appl.* **2001**, *752*, 233–245.

Table 5. Modification Type, Site, and Stoichiometry of Identified Modified Peptides from ADH1

peptide	modification type	sequence[a] and modification site	cleavage[b]	MH+	RT (min)	stoichiometry ± SD[c] (%)
T1	N-terminal Acetylation	(−)***S***IPETQK(G)	fully	844.441	25.07	99.1 ± 0.01
T1	no modification	(−)SIPETQK(G)	fully	802.431	16.37	0.9 ± 0.01
T7	oxidation M75	(K)LPLVGGHEGAGVVVG***M***GENVK(G)	fully	2035.064	46.59	1.8 ± 0.30
T7	no modification	(K)LPLVGGHEGAGVVVGMGENVK(G)	fully	2019.069	52.70	98.2 ± 0.30
T12	oxidation M168	(K)SANL***M***AGHWVAISGAAGGLGSLAVQYAK(A)	fully	2716.388	65.82	1.7 ± 0.33
T12	no Modification	(K)SANLMAGHWVAISGAAGGLGSLAVQYAK(A)	fully	2700.393	70.01	98.3 ± 0.33
T21	deamidation N262 (isoD[a]) + oxidation M270	(R)A***N***GTTVLVG***M***PAGAK(C)	fully	1403.720	34.03	1.0 ± 0.01
T21	deamidation N262 (D) + oxidation M270	(R)A***N***GTTVLVG***M***PAGAK(C)	fully	1403.720	43.99	8.8 ± 1.35
T21	deamidation N262 (isoD)	(R)A***N***GTTVLVGMPAGAK(C)	fully	1387.725	43.97	84.1 ± 1.72
T21	deamidation N262 (D)	(R)A***N***GTTVLVGMPAGAK(C)	fully	1387.725	44.27	3.0 ± 0.39
T21	no modification	(R)ANGTTVLVGMPAGAK(C)	fully	1386.741	43.07	3.1 ± 0.01
T5	deamidation N31 (isoD)	(K)A***N***ELLINVK(Y)	fully	1014.583	43.54	1.4 ± 0.10
T5	deamidation N31 (D)	(K)A***N***ELLINVK(Y)	fully	1014.583	47.18	1.4 ± 0.19
T5	no modification	(K)ANELLINVK(Y)	fully	1013.599	45.66	97.2 ± 0.08
T22	deamidation N282 (isoD)	(K)C*C*SDVF***N***QVVK(S)	fully	1356.592	37.93	4.8 ± 0.01
T22	deamidation N282 (D)	(K)C*C*SDVF***N***QVVK(S)	fully	1356.592	44.95	3.3 ± 0.04
T22	no modification	(K)C*C*SDVFNQVVK(S)	fully	1355.608	43.07	91.9 ± 0.08
P1	deamidation N94	(K)WL***N***GSC*MAC*EYC*ELGNESNC*PHADLSGYTHDGSFQQY(A)	partially	4358.689	59.67	100[d]
P2	deamidation N94	(K)WL***N***GSC*MAC*EYC*ELGNESNC*PHADLSGYTH(D)	partially	3533.360	55.82	100
P3	deamidation N94	(K)WL***N***GSC*MAC*EYC*ELGNESNC*PHADLSGY(T)	partially	3295.253	58.80	100
P4	deamidation N94	(K)WL***N***GSC*MAC*EYC*ELGNESNC*PH(A)	partially	2688.988	52.51	100
P5	deamidation N94	(K)WL***N***GSC*M(A)	partially	868.333	42.63	100

[a] C*, carbamidomethyl C; amino acid (S, N or M) in italic and bold is with modification; each deamidated N results in two isofroms, isoaspartic acid (isoD) and aspartic acid (D). [b] Fully, fully tryptic; partially, partially tryptic. [c] In percentage (%), an average from three replicate analyses and calculated by intensity of the modified peptide/(intensity of the modified peptide + intensities of related unmodified and other modified peptides); SD, standard deviation. [d] Identified partially tryptic peptides with N94 deamidation from the longest peptide T10 (69 amino acids, MW 7601.4); no corresponding peptides without N94 deamination were identified.

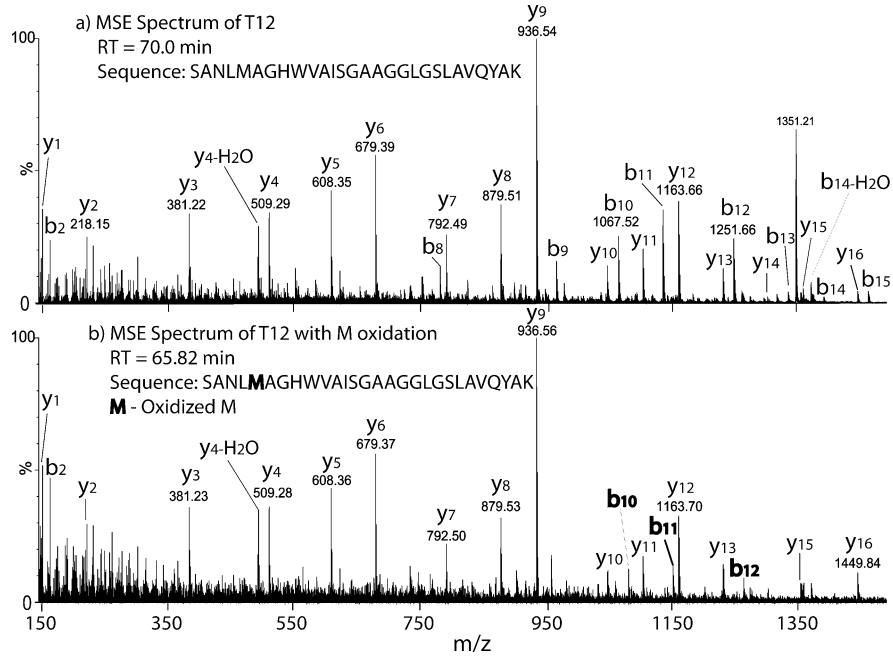

Figure 5. MS^E spectra of peptide T12 (**SANLMAGHWVAISGAAGGLGSLAVQYAK**) of ADH1 (ADH1p) before and after M-oxidation. (a) MS^E spectrum of T12, (b) MS^E spectrum of T12 with M-oxidation. b_{10}, b_{11}, and b_{12} (marked in bold) clearly prove the oxidation.

motif are very susceptible to deamidation.[38] The deamidation of N31 and N282 sites was found to be less than 10%.

CONCLUSIONS

The data independent LC-MS method using alternating low and elevated collision energy MS scanning is suitable for characterization of peptide maps. Database searching of MS (combination of peptide precursor and fragments) data revealed a low level of impurity proteins present in the target protein samples. Because the quantitative information (ion intensity) of each detected precursor was captured, we were able to quantify relative

(38) Robinson, N. E.; Robinson A. B. *Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins*; Athouse Press: Cave Junction, OR, 2004.

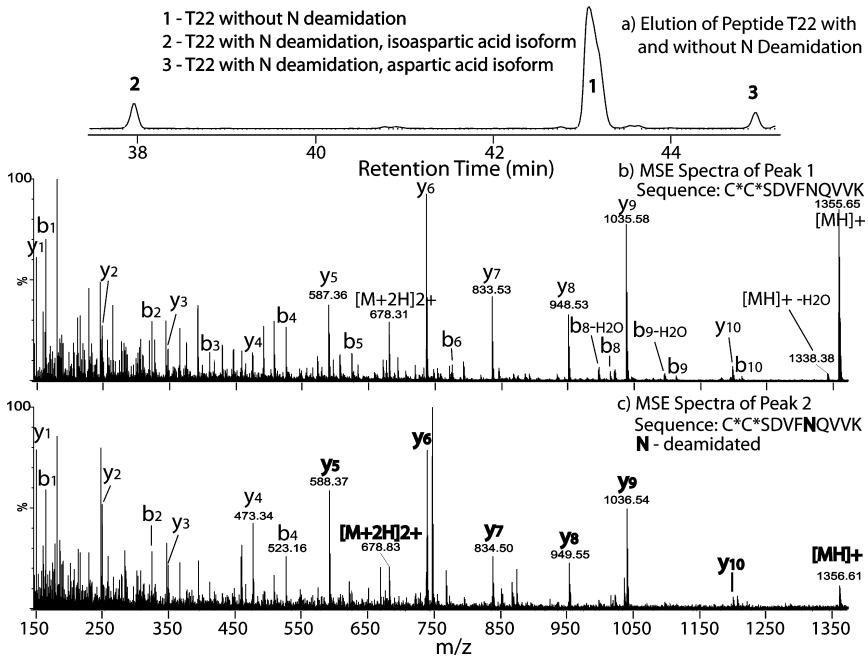

Figure 6. Elution pattern and MS^E spectra of peptide T22 (**C*C*SDVFNQVVK**) of ADH1 (ADH1p) before and after N-deamidation. (a) Elution pattern, (b) MS^E spectrum before N-deamidation, (c) MS^E spectrum after N-deamidation (isoaspartic acid isoform); y-series ions (marked in bold) starting from y_5 clearly prove the deamidation. **C*** is carbamidomethyl **C**.

abundances of the protein contaminants with ±10% SD. Contaminants at ~1% level were readily detectable. The method provided high sequence coverage and identifies all LC-eluted peptides within the limits of detection. Because of the unbiased acquisition nature it ensured sampling and identification of low-abundance LC-eluted components. High sequence coverage is essential for identification and quantitation of PTMs. Site-specific modifications at ~0.5% level were successfully identified and quantified. The PTM quantification repeatability was within several percent of SD, increasing to 10% or more for peptides near the limits of detection.

The protein identification, peptide assignment, and quantitation are achieved within a single LC-MS analysis. The unexpected peptides with unknown modification and sequences can be identified and validated manually from the acquired data. No sample reanalysis was necessary. High LC resolution of the peptides was useful for characterization of isobaric N-deamidation isoforms and in general for analysis of complex peptide maps.

In conclusion, the data independent LC-MS method is suitable for analysis of therapeutic proteins which needs to be well characterized prior to and during clinical trials and administration. We are applying the approach demonstrated here for detailed characterization of glycoproteins in monoclonal antibody and vaccine samples, including identification and quantification of N-linked glycosylations.

SUPPORTING INFORMATION AVAILABLE

Supplemental Tables 1 and 2 and Supplemental Figure 1. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review March 3, 2009. Accepted May 22, 2009.

AC900468J