Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 259 (2007) 161–173
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Assigning *in vivo* carbamylation and acetylation in human lens proteins using tandem mass spectrometry and database searching

Zee-Yong Park [a], Rovshan Sadygov [b], Judy M. Clark [c], John I. Clark [c], John R. Yates III [d,*]

[a] *Department of Life Science, Gwangju Institute of Science & Technology, 1 Oryong-Dong, Gwangju, South Korea*
[b] *ThermoElectron, 355 River Oaks Parkway, San Jose, CA 95134, United States*
[c] *Department of Biological Structure, Box 357420, HSB G-514, University of Washington, Seattle, WA 98195-7420, United States*
[d] *Department of Cell Biology, 10550 North Torrey Pines Road, SR11, The Scripps Research Institute, LaJolla, CA 92037, United States*

Received 19 June 2006; received in revised form 21 August 2006; accepted 21 August 2006
Available online 25 September 2006

This paper is dedicated to Professor Donald Hunt on the occasion of his 65th birthday.

Abstract

In this paper, we show that ion trap mass spectrometers can differentiate acetylation and carbamylation modifications based on database search results for a lens protein sample. These types of modifications are difficult to distinguish on ion trap instruments because of their lower resolution and mass accuracy. The results were corroborated by using accurate mass information derived from MALDI TOF MS analysis of eluted peptides from a duplicate capillary RPLC separation. Tandem mass spectra of lysine carbamylated peptides were further verified by manual assignments of fragment ions and by the presence of characteristic fragment ions of carbamylated peptides. It was also observed that carbamylated peptides show a strong neutral loss of the carbamyl group in collision induced dissociation (CID), a feature that can be prognostic for carbamylation. In a lens tissue sample of a 67-year-old patient, 12 *in vivo* carbamylation sites were detected on 7 different lens proteins and 4 lysine acetylation sites were detected on 3 different lens proteins. Among the 12 *in vivo* carbamylation sites, 9 are novel *in vivo* carbamylation modification sites. Notably, *in vivo* carbamylation of γS crystallin, βA4 crystallin, βB1 crystallin, and βB2 crystallin observed in this study have never been reported before.

Keywords: *In vivo* carbamylation; Acetylation; LC–MS/MS; Shotgun proteomics; SEQUEST

1. Introduction

A goal of proteomics has been the identification and relative quantification of expressed proteins. Mass spectrometric protein identification techniques utilizing accurate masses (MS) or fragmentation patterns (MS/MS) of peptides were developed and combined with protein/peptide separation techniques to identify individual components of complicated protein mixture samples [1–9]. *In vitro* and *in vivo* amino acid labeling techniques using stable isotopes were also developed and combined with mass spectrometric techniques to facilitate relative quantification of expressed proteins under different conditions [10–16]. By using these proteomics tools, many important biological systems are being investigated. A challenge for proteomics is to characterize biologically important post-translational modifications occurring on cellular proteins.

Post-translational modifications are known to affect the structure and function of proteins and many functions are tightly regulated by post-translational modifications [17,18]. There are obstacles to overcome in the study of post-translational modifications. Information regarding co- and post-translational modifications cannot be easily inferred from genome sequences and more than 200 different post-translational modifications have been reported. Only a handful of them can be easily characterized including phosphorylation, acetylation, ubiquitination, and sumoylation. Several mass spectrometric methods have been developed for large-scale identification of post-translational modifications in complicated protein mixtures [19–24]. These methods have mainly focused on one or two

* Corresponding author. Tel.: +1 858 784 8862; fax: +1 858 784 8883.
E-mail address: jyates@scripps.edu (J.R. Yates III).

1387-3806/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2006.08.013

types of post-translational modifications and typically involve multi-step sample preparations that may include chemical derivatization or peptide enrichment processes. Identification of post-translational modifications remains a significant challenge to better understand the regulation of biological processes.

Shotgun proteomics when combined with multi-dimensional capillary liquid chromatography and tandem mass spectrometry has been used to identify different types and sites of post-translational modifications [17,25,26]. A popular mass spectrometer for shotgun proteomics experiments is the ion trap because the robust performance and fast scan speed are well suited for liquid chromatography. A drawback to these low resolution and mass accuracy instruments is the ambiguity that can result in the identification of modifications with closely related molecular weights. Modifications with very small molecular weight differences, such as phosphorylation and sulfation require very high resolution instruments to differentiate. There are several post-translational modifications with only 1 Da mass difference, such as acetylation and carbamylation, deamidation of Asn and Gln, and trimethylation and acetylation. An inability to differentiate these modifications with high confidence leads to ambiguities in the identification of large-scale post-translational modifications. MacCoss et al. analyzed human lens tissue from a 4-year-old congenital cataract patient using shotgun proteomics and 13 sites of lysine acetylation or carbamylation were reported since the possibility of carbamylation could not be ruled out [17].

The presence of carbamylation was thought to be a result of urea decomposition, but Smith and co-workers recently identified three sites of *in vivo* carbamylation in γ crystallin and suggested that *in vivo* carbamylation is not a minor post-translational modification in lens tissue proteins and it might be related to the aging process of lens [27]. In the Smith et al. studies, the modified proteins were first separated from unmodified proteins by various LC techniques prior to enzymatic digestion. A high resolution scan on an ion trap mass spectrometer using the zoom scan function was performed after a regular capillary RPLC–MS/MS analysis [28]. This type of approach is time-consuming and does not lend itself to large-scale studies of post-translational modifications in complex protein samples. Thus, we investigated if lysine acetylation and carbamylation can be unambiguously differentiated using tandem mass spectrometry data obtained from ion trap mass spectrometer and a database search.

Here, we show that the cross-correlation scores from differential modification searches can be used to distinguish acetylation from carbamylation on lysine residues in a conventional ion trap mass spectrometer data without additional high mass accuracy analysis of peptides. In this study, lens tissue sample from a 67-year-old patient was analyzed to identify sites of acetylation and carbamylation using a new approach that would confirm the observations of the Smith and co-workers' study [27,28]. Both water soluble and insoluble lens proteins were prepared by using high concentrations of guanidine hydrochloride to avoid artifactual carbamylation [29,30].

2. Materials and methods

2.1. Preparation of protein extracts from human lens

A lens sample was obtained from a 67-year-old patient. The nucleus and cortex of the lens were not separated and the whole lens was homogenized with microtube pestle in 0.2 ml of 20 mM phosphate buffer (1 mM EGTA, pH 7.0) at 4 °C. Soluble and insoluble portions were separated by centrifugation at 10,000 × *g* for 30 min at 4 °C. In order to avoid chemical modification by urea, guanidine hydrochloride (GuHCl) was used as a denaturant throughout the experiment. The soluble portion was diluted with Tris–HCl buffer (pH 8.5) containing GuHCl to a final concentration of 6 M GuHCl, 100 mM Tris–HCl. The insoluble fraction was resuspended in 6 M GuHCl, 100 mM Tris–HCl (pH 8.5) buffer. Both fractions were reduced with 5 mM tris[2-carboxyethyl] phosphine (TCEP) for 30 min and alkylated with 25 mM iodoacetamide for 30 min.

2.2. Enzymatic digestion protocol

Reduced and alkylated protein samples were digested with Endoproteinase Lys-C (Roche Diagnostics) at a ratio of 100:1 (substrate:enzyme) at 37 °C for 12 h. The Lys-C digested sample was diluted four-fold with 100 mM Tris–HCl buffer, pH 8.5. Sequencing grade modified trypsin (Promega) was added at an enzyme-to-substrate ratio of 1:50 (w/w) and incubated for 12 h at 37 °C. After incubation, trypsin digestion was quenched with 90% formic acid to a final concentration of 4% and stored at −80 °C.

2.3. MicroLC–MS/MS analysis

A single phase microcapillary column was constructed with 100 μm i.d. fused silica capillary tubing pulled to a 5 μm i.d. tip by using a CO_2 laser puller (Sutter Instruments, P-2000). The capillary column was packed with 15 cm, 5 μm particle size. Polaris C18-A (Metachem) using a home-made high pressure column loader. Approximately, 1–5 μg of digested lens protein was directly loaded onto the capillary column. The buffer solutions used to separate the lens protein digests were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B). A 100 min gradient was used (0–5 min, 100% buffer A; 5–75 min, 0% buffer B to 55% buffer B; 75–100 min, 55–100% buffer B). Peptides eluted from the capillary column were electrosprayed into a LCQ Deca XP Plus ion trap mass spectrometer (ThermoFinnigan) with the application of a distal 2.4 kV spray voltage. A cycle of one full scan (400–1400 *m/z*) followed by three data-dependent MS/MS scans at a 35% normalized collision energy was repeated throughout the LC separation.

2.4. Analysis of MS/MS spectra

Tandem mass spectra were extracted from Xcalibur raw files using the program "ExtractMS". A filtering program, "2 to3", was then used to determine the charge states of multiply charged

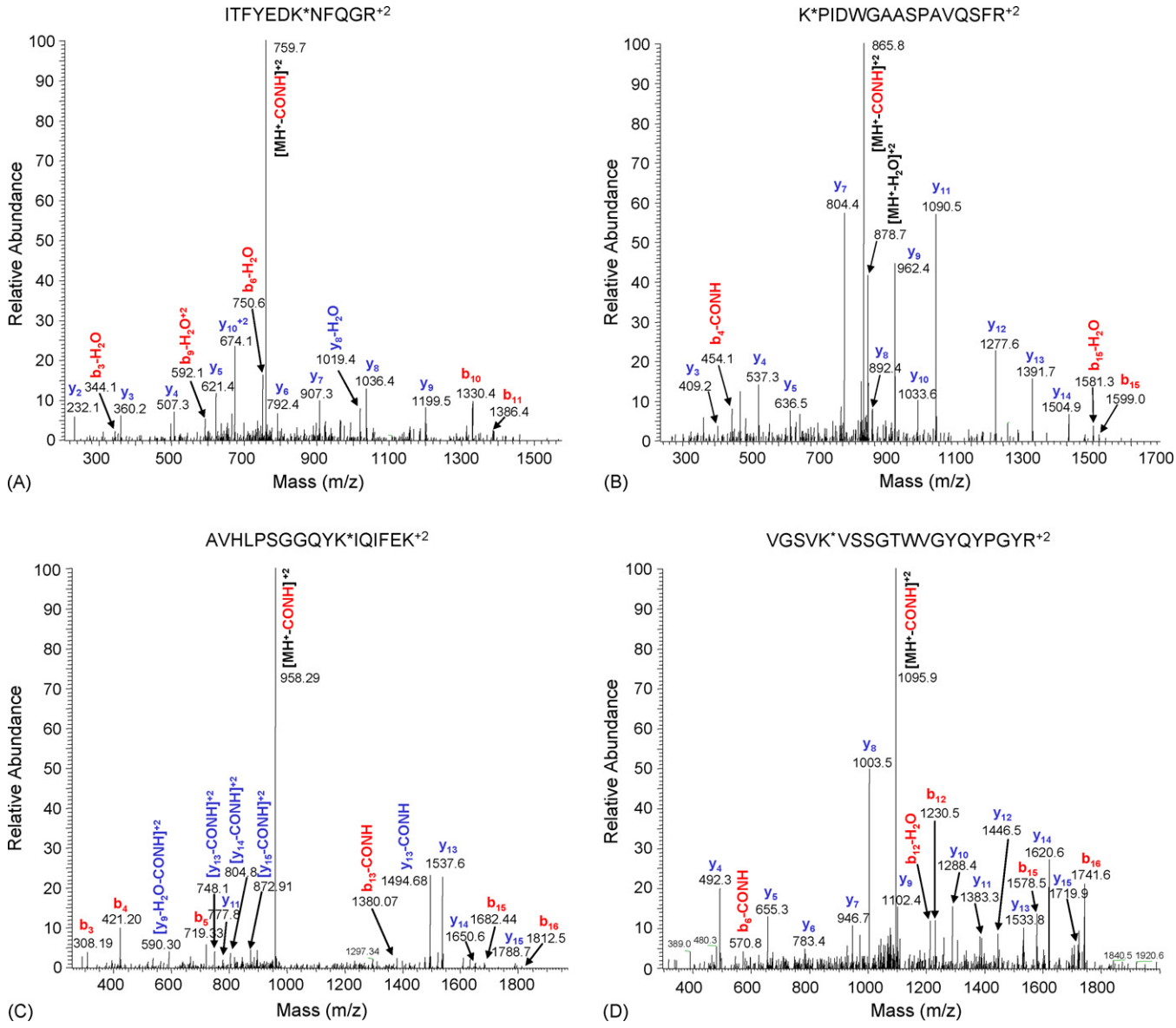

Fig. 1. MS/MS spectra of tryptic digest peptides with *in vivo* carbamylation modification site and manual assignments of fragment ions. Strong neutral loss ions of carbamyl group appeared in all four spectra. Asterisk (*) denotes modification sites.

precursor peptides and to remove poor quality spectra [31]. The final data file was searched against a protein database by a cluster version of "Pep_Probe" which provides a cross-correlation score of SEQUEST and a hypergeometric distribution based probability score for each peptide match [32]. Mass options used in the search were average mass for precursor ions and monoisotopic mass for fragment ions. The resulting output files were filtered using the program "DTASelect" [33]. A subset database was constructed using only the proteins identified and this subset database was mainly used to expedite differential modification searches. Modifications of carbamylation (+43) and acetylation (+42) were considered first and differential modifications of other mass shifts (+40, +41, +44, and +45) on lysine were also considered to assess XCorr score and probability score changes as a function of modification mass offset.

2.5. Capillary RPLC-off line-MALDI TOF MS analysis

To obtain accurate mass information on enzymatically digested peptides, a MALDI TOF mass spectrometer was used. The microLC separation used in the MS/MS analysis was duplicated using the same column under the same sample loading conditions. Instead of electrospraying into ion trap mass spectrometer, digested peptides eluted in the microLC separation were collected and directly spotted on the MALDI plate every 2 min so that the retention time of particular MS/MS spectra of microLC–MS/MS analysis could be linked to the particular spot on the MALDI sample plate. The MALDI sample plate was pre-coated with α-cyano-4-hydroxycinnamic acid. MALDI TOF mass analysis was performed by using a Micromass Tof-Spec 2E equipped with a pulsed nitrogen laser (337 nm) source. All mass spectra were taken in the reflected mode using delayed

extraction technology. Ion acceleration voltage was 20 kV and an extraction voltage was 16 kV. The delay time between laser pulse and pulsing in the extraction plate was 200 ns. Signals from at least 100 laser shots were averaged to increase signal to noise ratio. All mass spectra were externally calibrated using a matrix dimer signal ($m/z = 379.0930$) and bradykinin $[M+H]^+$ ion signal ($m/z = 1060.5692$). Average mass accuracy of less than 100 ppm was obtained in the mass range from 500 to 3000 Da.

3. Results

The digested lens sample was analyzed by LC/MS/MS and tandem mass spectra were collected. Despite the simplicity of the peptide separation technique, fairly good sequence coverage was obtained for the major proteins of the lens. We identified a total of 98 proteins from the two LC–MS/MS analyses. Nearly, all lens crystallins previously reported were detected and their sequence coverage was more than 55% except two minor lens proteins; γA and γB crystallins. Based on these identifications, a subset database consisting of these proteins was used to identify modified peptides with particular emphasis on acetylation and carbamylation.

A differential modification search using SEQUEST was evaluated to see if it is capable of identifying a lysine carbamylation modification in tandem mass spectral data generated from a conventional ion trap mass spectrometer. A differential modification search of carbamylation on lysine residues (K, +43) was performed and compared with the results obtained in a differential modification search of acetylated lysine (K, +42). Several tandem mass spectra were identified as carbamylated on lysine residues only in the differential modification search for lysine carbamylation. These tandem mass spectra yielded significantly higher XCorr values in the differential modification search of lysine carbamylation than of those in the lysine acetylation

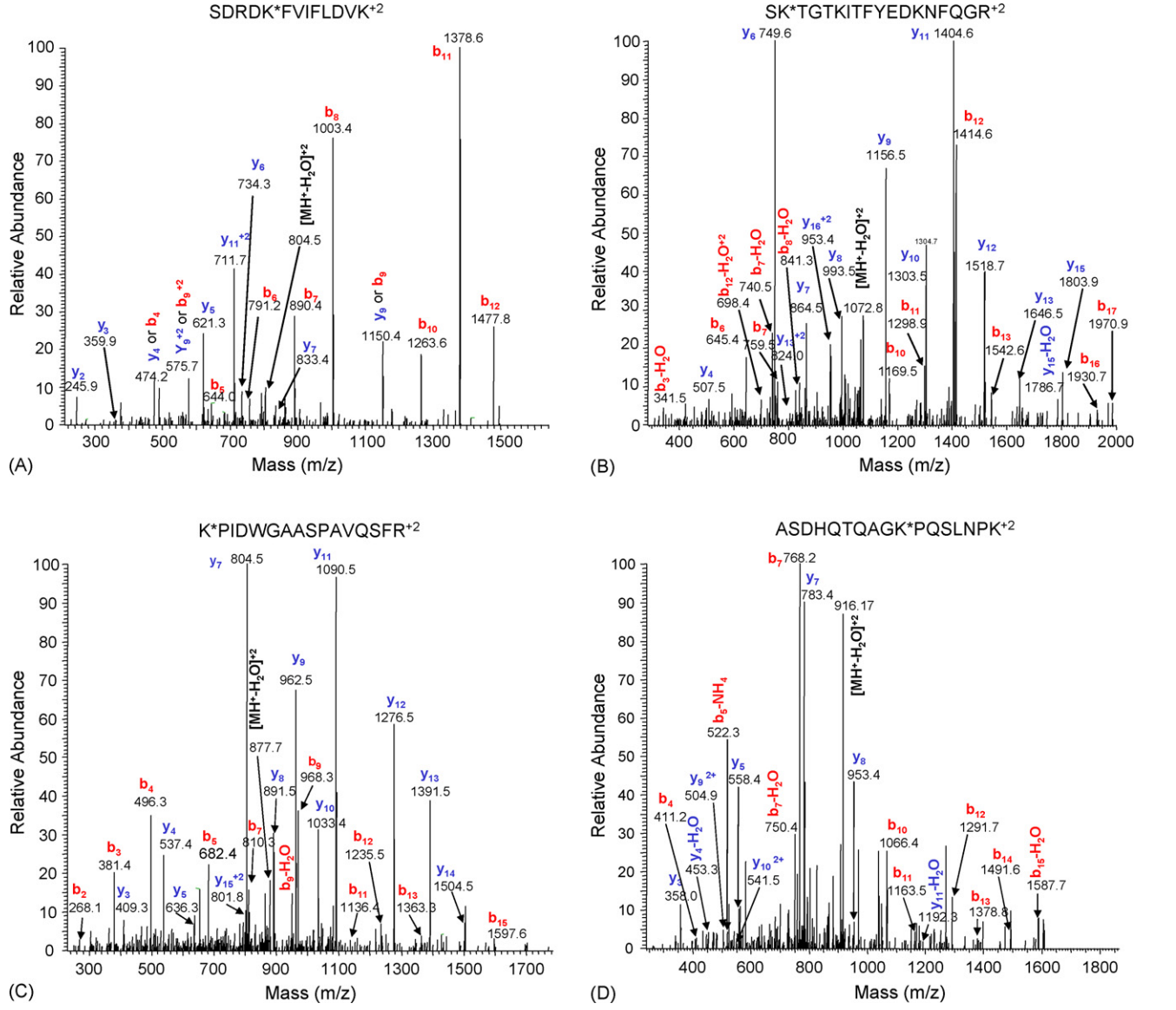

Fig. 2. MS/MS spectra of tryptic digest peptides with acetylation modification site and manual assignments of fragment ions. Neutral loss ions of acetyl group are missing in all four spectra. Asterisk (*) denotes modification sites.

Table 1
List of *in vivo* carbamylation and acetylation modification sites observed in water-soluble and insoluble lens proteins of 67-year- old patient lens tissue sample

Scan number[a]	Location[b]	Peptide sequence[c]	Neutral loss ion[d]	XCorr^{max} at K+43/K+42	Theoretial mass	Observed mass[e]
2130.2	γS K^{14}	ITFYED**K***NFQGR	Yes	K+43	1560.7 (+43)	1560.6 (−0.1)
2638.2	γS K^{95}	AVHLPSGGQY**K***IQIFEK	Yes	K+43	1958.0 (+43)	1958.0 (0.0)
2526.2	γS K^{159}	**K***PIDWGAASPAVQSFR	Yes	K+43	1772.9 (+43)	1771.8 (−1.1)
1734.2	γC G^{1}	**G***KITFYEDR	Yes	G+43	1171.6 (+43)	1171.7 (+0.1)
1716.2	γD G^{1}	**G***KITLYEDR	Yes	G+43	1137.6 (+43)	1137.5 (−0.1)
1007.2	αB K^{175}	EEKPAVTAAPK**K***	Yes	K+43	1311.7 (+43)	1311.6 (−0.1)
3239.3	$βA_4$ K^{48}	SL**K***VLSGAWVGFEHAGFQ GQQYILER	Yes	K+43	2963.5 (+43)	2963.9 (+0.4)
1472.2	$βB_1$ K^{50}	**K***AAELPPGNYR	Yes	K+43	1258.7 (+43)	1258.6 (−0.1)
3580.3	$βB_1$ K^{160}	ISLFEGANF**K***GNTIEIQGDDAPSLWVYGFSDR	Yes	K+43	3589.7 (+43)	N/A
2624.2	$βB_1$ K^{186}	VGS V**K***VSSGTWVGYQYPGYR	Yes	K+43	2233.1 (+43)	2233.3 (+0.2)
2838.3	$βB_1$ K^{234}	LRD**K***QWHLEGSFPVLATE PPK	Yes	K+43	2491.3 (+43)	2491.7 (+0.4)
1832.2	$βB_2K^{101}$	RTDSLSSLRPI**K***VDSQEHK	Yes	K+43	2239.2 (+43)	2239.0 (−0.2)
2984.2	αA K^{1}	SDRDK*FVIFLDVK	No	K+42	1623.9 (+42)	1623.9 (0.0)
1002.2	$βB_2$ K^{10}	ASDHQTQAGK*PQSLNPK	No	K+42	1848.9 (+42)	1849.0 (+0.1)
1840.2	γS K^{3}	SK*TGTKITFYEDKNFQGR	No	K+42	2162.1 (+42)	2162.5 (+0.4)
1839.3	γS K^{3}	SK*TGTKITFYEDKNFQGR	No	K+42	2162.1 (+42)	2162.5 (+0.4)
2563.2	γS K^{159}	K*PIDWGAASPAVQSFR	No	K+42	1771.9 (+42)	1771.8 (−0.1)
2383.3	γS K^{159}	K*PIDWGAASPAVQSFRR	No	K+42	1928.0 (+42)	1927.9 (−0.1)

[a] Scan number means scan number of MS/MS spectrum.
[b] Underline (_) in the second column indicates previously reported modification sites.
[c] Bold letter (**K***) in the third column denotes lysine *in vivo* carbamylation sites detected.
[d] Neutral loss ion means the presence of carbamyl group neural loss ion or acetyl group neutral loss ion.
[e] Observed mass indicates monoisotopic mass observed in the MALDI TOF MS analysis.

search. Hence, the criteria of yielding higher a XCorr value for lysine carbamylation *versus* lysine acetylation was used first to select peptide candidates for lysine carbamylation. Fig. 1 shows tandem mass spectra of four carbamylated peptides identified in the differential modification search. Interestingly, the base peak in both tandem mass spectra correspond to a neutral loss of the carbamyl group ($[M+H^+-CONH]^{+2}$) from the precursor ions. This is consistent with the characteristic MS/MS pattern previously reported for carbamylated peptides [27,28,34]. Other tandem mass spectra with higher XCorr values in the carbamylation search also showed a neutral loss of a carbamyl group even though they were not the base peak. In contrast, the tandem mass spectra which showed higher XCorr values for lysine acetylation (K, +42) did not show any evidence of a neutral loss of an acetyl group from the precursor ion (see, Fig. 2). Since there was no neutral loss ion present in their tandem mass spectra, intensities for b and y series ions were relatively higher than observed for the lysine carbamylated peptides.

In order to verify each modification of lysine modified peptides, capillary RPLC separation was repeated and accurate mass analysis of peptides was performed using a MALDI TOF mass spectrometer. Eluted peptides from capillary RPLC separation were spotted on a MALDI sample plate at 2 min intervals. Since a fairly reproducible peptide separation can be achieved with our capillary RPLC separation system (retention time deviation less than ±1.0 min), retention time information of MS/MS spectra was used to link which sample spot on a MALDI sample plate contained peptides of interest. All tandem mass spectra but the one shown in Fig. 1 corresponded to the exact mass of a carbamylated peptide. This example will be discussed in detail later in this section. Accurate mass analysis by MALDI TOF MS also confirmed all acetylation modifications of peptides shown in Fig. 2. The results from the off-line capillary RPLC MALDI TOF MS analysis are summarized in Table 1 and confirmed the initial database searching results.

Manual assignment of fragment ions in the tandem mass spectra clearly show why one modification scores higher than the other in the differential modification search (see Fig. 3). At the ±1.0 *m/z* fragment ion mass accuracy used by SEQUEST, the ^{14}Lys residue of gamma S crystallin can only be assigned as a carbamylated lysine (see Fig. 3 A and B). Several unassigned fragment ions in Fig. 3B were exactly 1 *m/z* higher than the theoretical fragment ions of the acetylated lysine containing peptide. If the fragment ion mass accuracy used in XCorr calculation was greater than ±1.0 *m/z*, then those unassigned fragment ions in Fig. 3 B would match the fragment ions of the acetylated lysine containing peptide and XCorr scores for the two different modification searches would be identical. Acetylation on the ^{70}Lys residue of alpha A crystallin was also manually assigned (Fig. 3C and D). A higher XCorr value was observed and more fragment ions were manually assigned for the acetylated lysine.

On the ^{159}Lys residue of γS crystallin, two different modifications were observed. Two tandem mass spectra were matched to the same peptide with different modifications (see Fig. 4). Fig. 4A shows a tandem mass spectrum that was assigned as a carbamylated lysine containing peptide and the tandem mass spectrum shown in Fig. 4B was identified with acetylated lysine. Mass spectra on the right focus on the region where the neutral loss of the carbamyl group should appear. A neutral loss of the carbamyl group is clearly missing in the MS/MS of the acetylated lysine containing peptide. Also it was noted that far more b and y ions were identified in the MS/MS of acetylated

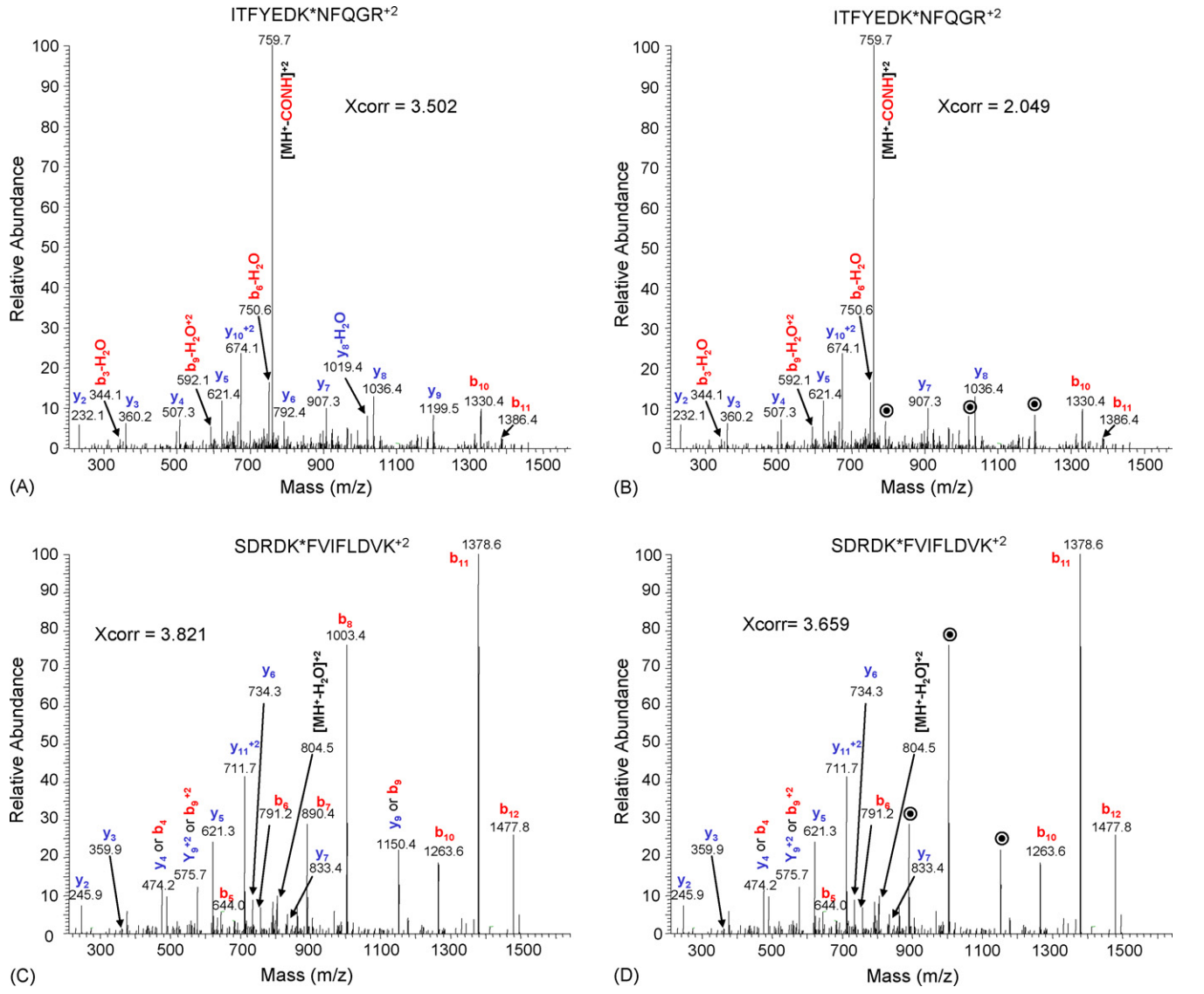

Fig. 3. Manual assignments of lysine modified tryptic digest peptides, ITFYEDK*NFQGR and SDRDK*FVIFLDVK with two different modification possibilities: (A) *in vivo* carbamylation of ITFYEDK*NFQGR, (B) acetylation of ITFYEDK*NFQGR, (C) acetylation of SDRDK*FVIFLDVK, and (D) *in vivo* carbamylation of SDRDK*FVIFLDVK. XCorr of corresponding differential modification searches were also shown in each MS/MS spectrum. Allowed mass error was better than ±1 *m/z*. Symbol (◉) denotes fragment ions of missing assignments in a false modification possibility.

lysine containing peptide. The MALDI TOF MS analysis of the sample spot which covers the retention time region of the above two tandem mass spectra is shown in Fig. 5. The monoisotopic mass in Fig. 5A matched exactly with the monoisotopic mass of the acetylated lysine containing peptide; however, the actual isotope distribution pattern is quite different from the expected theoretical isotope distribution pattern of the acetylated peptide shown in Fig. 5B. The theoretical composite isotope distribution for a 1:1 mixture of the acetylated and carbamylated peptide is shown in Fig. 5C. The similarity between the composite theoretical isotope pattern and the observed isotope pattern strongly suggests co-existence of the carbamylated and acetylated peptides in one MALDI sample spot. Selected ion chromatograms for these two differently modified peptides revealed that a slightly higher level of carbamylation was present on this lysine residue (see Fig. 6). As it was observed in previous studies, a carbamylated peptide elutes earlier than the acetylated version of the same peptide in a reversed phase LC separation [28].

To observe how XCorr changes as a function of the differential modification offset mass, a search for lysine modification was performed with the offset mass varied from 40 to 45. The resulting XCorr values were plotted to see if the XCorr value maximizes at the true offset mass. We also analyzed the hypergeometric scores (Pep_Probe) to see if the results were consistent with XCorr values. Both XCorr and the hypergeometric scores for tandem mass spectra shown in Fig. 1 are plotted in Fig. 7 and those for the tandem mass spectra shown in Fig. 2 are plotted in Fig. 8. Fig. 7 shows the search results for the carbamylated lysine containing peptides and the maximum XCorr values were always observed in the differential modification search of +43 on the lysine residue. Hypergeometric scores also followed a very similar trend to the XCorr values. The acetylated lysine contain-

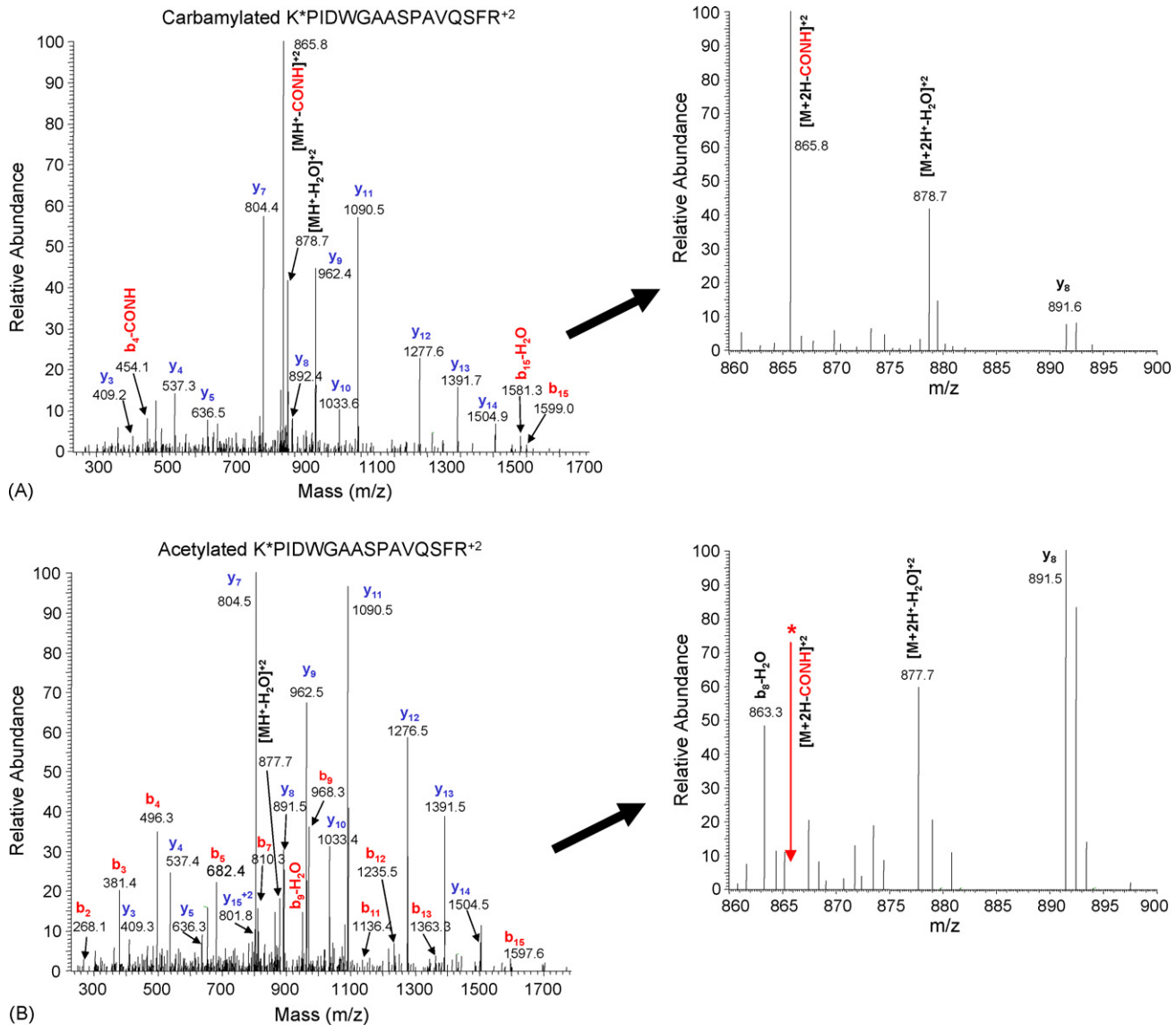

Fig. 4. MS/MS spectra of a tryptic digest peptide, KPIDWGAASPAVQSFR with two different modifications: (A) KPIDWGAASPAVQSFR with lysine *in vivo* carbamylation and (B) KPIDWGAASPAVQSFR with lysine acetylation. The inserts on the right of each spectrum show the presence and absence of carbamyl group neutral loss ion.

ing peptides are shown in Fig. 8 and all four peptides have their maximum XCorr values in a differential modification search with a +42 modification to lysine. Although the XCorr values of the tandem mass spectra of the peptide SDRDK*FVIFLDVK show only a small difference between the two different searches, MALDI TOF analysis of the fraction containing this peptide confirmed the acetylation modification (see Table 1).

In addition to lysine *in vivo* carbamylation sites, N-terminal glycine carbamylations were identified in this study. Two carbamylation sites of γC and D crystallin identified in the lysine differential modification search were later corrected as N-terminal glycine carbamylation in the manual assignments. Since both lysine residues are located next to N-terminal glycine, tandem mass spectra from the ion trap mass spectrometer did not show a sufficient number of fragment ions for a clear distinction. Thus, XCorr values for the two differential modification searches (N-terminal and lysine carbamylation) were not significantly different. Manual assignments of fragment ions for two different modification sites are shown in Fig. 9 and N-terminal glycine carbamylation and lysine carbamylation were differentiated by only two fragment ions;$[y_8\text{-}H_2O]^{+2}$ and $y_8{}^{+2}$. These two N-terminal carbamylation sites were previously reported as *in vivo* carbamylation sites [27].

Lastly, the ion chromatograms for the modified and unmodified peptides were compared to estimate levels of *in vivo* carbamylation for the different γ crystallins (see Fig. 10). γS crystallin appeared to have far lower *in vivo* carbamylation modification than other γ crystallins. *In vivo* carbamylation levels of γC and γD were comparable to those of unmodified ones.

4. Discussion

In vivo carbamylation and acetylation have a mass difference of only 1 Da. The unambiguous assignment of these modifica-

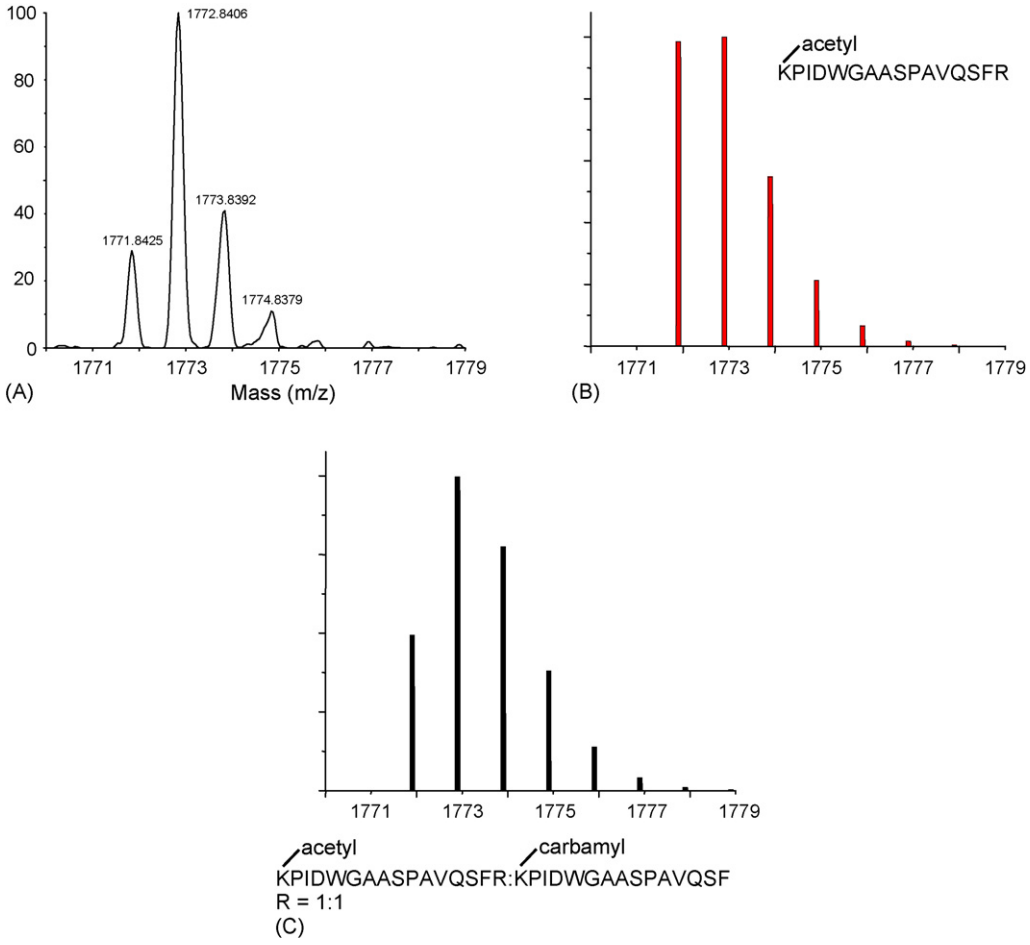

Fig. 5. Enlarged actual MALDI TOF mass spectrum of a fraction containing modified tryptic digest peptide, KPIDWGAASPAVQSFR (A) and theoretical mass spectra of acetylated KPIDWGAASPAVQSFR (B) and of acetylated and carbamylated KPIDWGAASPAVQSFR 1:1 mixture (C).

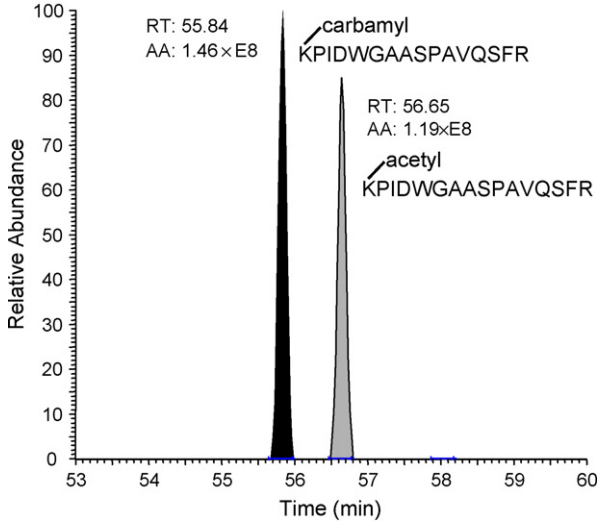

Fig. 6. Modification level comparison of *in vivo* carbamylation and acetylation. Selected ion chromatograms of *in vivo* carbamylated KPIDWGAASPAVQSFR and acetylated KPIDWGAASPAVQSFR are compared. Black colored region indicates ion current for 887.6 ± 0.2 and gray colored region indicates ion current for 886.86 ± 0.2. RT stands for retention time and AA stands for integrated area.

tions because of their close molecular weights is difficult using an ion trap mass spectrometer. Here, we have demonstrated that a simple XCorr value comparison in a differential modification search of SEQUEST can distinguish carbamylation from acetylation. In order to understand how XCorr values can distinguish two very similar post-translational modifications in mass, we need to look into the corresponding spectral modeling that is done for calculation of XCorr's. SEQUEST searches a protein database for candidate peptide sequences which fall within ±3 *m/z* of the molecular weight calculated from the precursor ion mass. Although this mass window of candidate peptide selection can be adjusted, it is typically set at 6 *m/z* for ion trap mass spectrometry data. Peptide candidates within the mass tolerance then undergo preliminary scoring to narrow the list of peptides to the 500 best candidates. Fragment ion masses from theoretical tandem mass spectra of the peptide candidates and the actual tandem mass spectrum are compared with a mass tolerance of 2 *m/z* (±1 *m/z*) in the preliminary scoring routine. Integer masses are used to speed up the process when fragment ions are compared. Since integer mass and 2 *m/z* mass windows are used in the preliminary scoring process, the resulting preliminary scores (Sp score) for two peptides with very similar modifications in mass, *i.e.*, 1 Da difference, tend to be the same. Cross-correlation scores for these two modifications can be different; however,

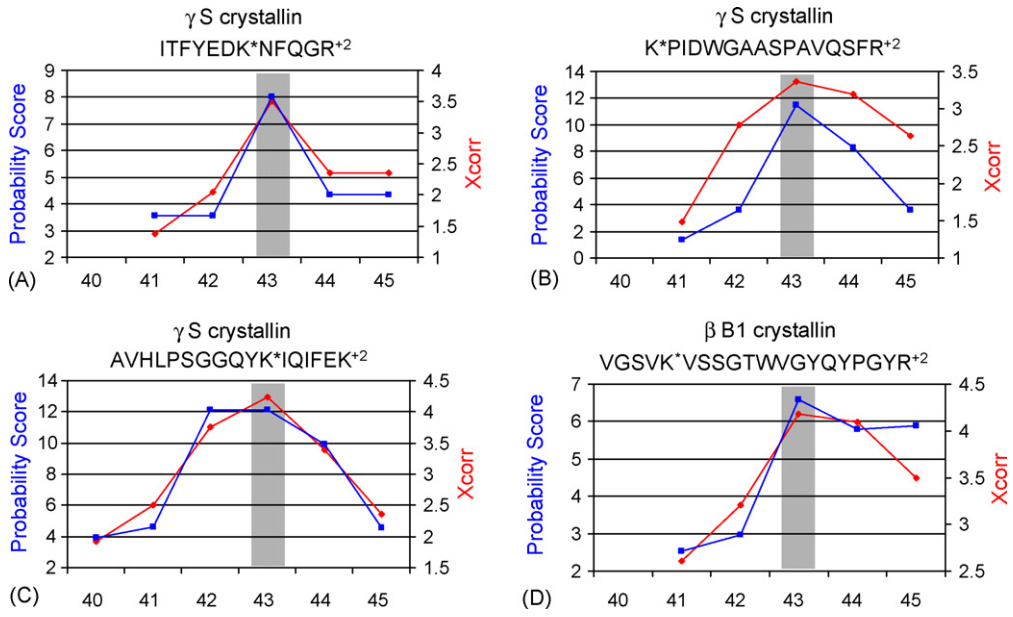

Fig. 7. Plot of effects of changing differential modification search option (lysine) on peptide identification algorithm scores of *in vivo* carbamylated lysine containing peptides. Shaded area indicates where maximum search scores are observed. Probability score was obtained from hypergeometry distribution of Pep_Probe.

because the effective mass accuracy of matching fragment ions to the elements of the experimental peak list is better than 1 *m/z*. If the theoretical mass of a fragment ion is different from the experimental mass by 1 *m/z* due to incorrect modification information, then this fragment ion will not be considered in the calculation of XCorr. The mass accuracy (monoisotopic mass) of the ion trap MS/MS scan is typically higher than that of the MS scan. Manual assignments of fragment ions in the MS/MS spectra to theoretical fragment ions of the corresponding peptide showed an average mass deviation of around 0.5 *m/z* in our study (data not shown). Therefore, it is plausible that two very similar modifications in mass will be differentiated by the masses of experimental fragment ions containing modification sites and it will be ultimately reflected in their XCorr values. The difference in the cross-correlation scores of two different post-translational modifications is also dependent upon how many fragment ions containing the modification site, especially y ions for tryptic digestion cases, are present in the tandem mass spectrum. XCorr calculations use a preset value for fragment ion mass accuracy and cannot be changed by ordinary users. On the contrary, fragment ion mass accuracy can be changed in Pep_Probe. The highest mass accuracy currently implemented

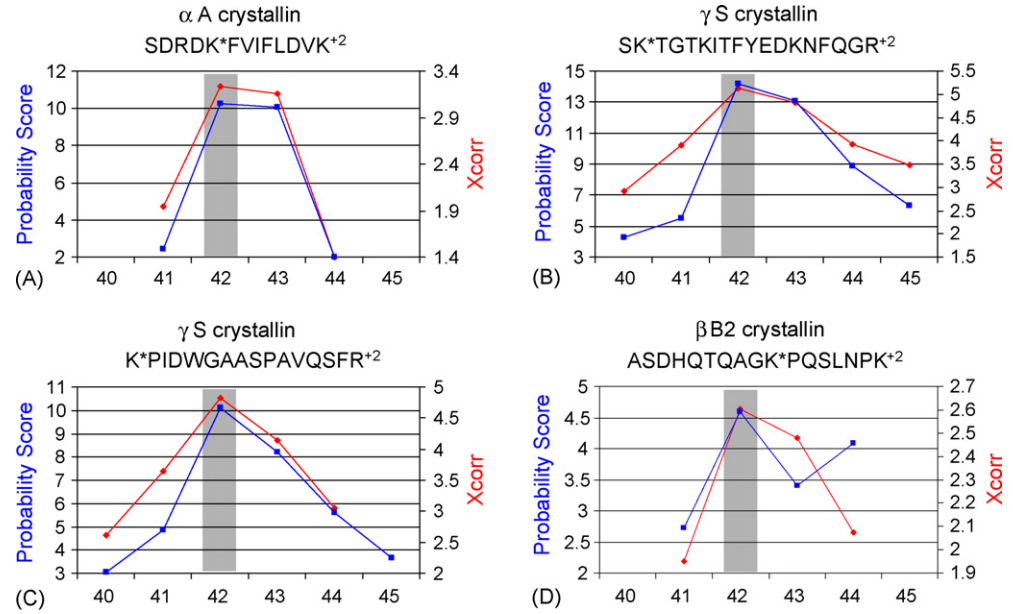

Fig. 8. Plot of effects of changing differential modification search option (lysine) on peptide identification algorithm scores of acetylated lysine containing peptides. Shaded area indicates where maximum search scores are observed. Probability score was obtained from hypergeometric distribution of Pep_Probe.

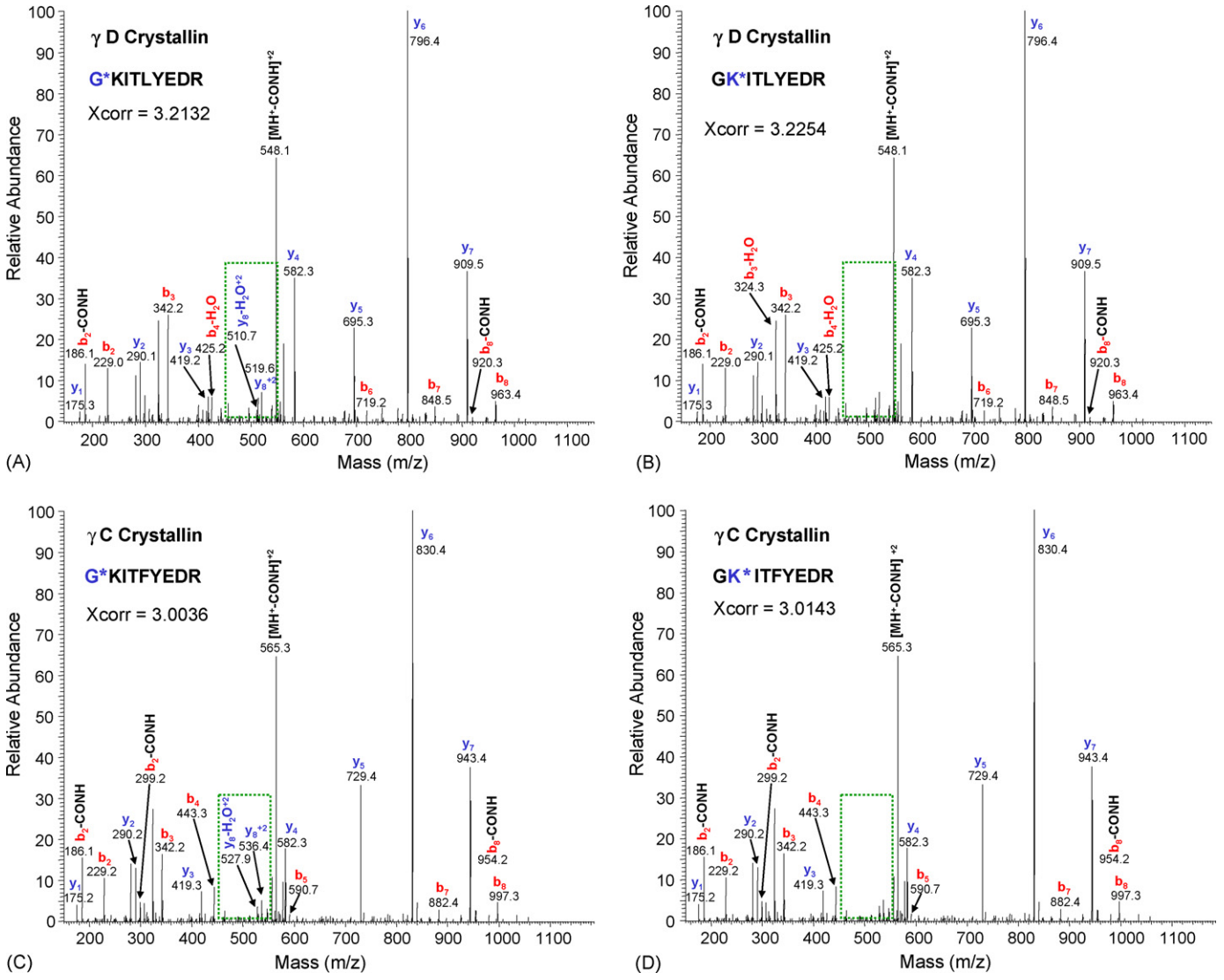

Fig. 9. Comparison of manual assignment results of MS/MS spectra for two different modification site possibilities: (A and B) *in vivo* carbamylation of N-terminal glycine *vs.* lysine of γD crystallin, (C and D) *in vivo* carbamylation of N-terminal glycine *vs.* lysine of γC crystallin. Dotted box shows the region where fragment ions matching N-terminal carbamylation of γD and γC appeared ($[y_8\text{-}H_2O]^{+2}$ and y_8^{+2}).

is 0.1 *m/z*. When the mass accuracy for the probability scoring system was increased from 1.0 to 0.1 *m/z*, the scores of correct identifications were clearly separated from those of incorrect identifications at 0.5 *m/z* mass accuracy.

Tandem mass spectrometric identification of *in vivo* carbamylation sites of proteins has been only recently reported by several research groups [27,28,34,35]. An internal lysine residue of human αB crystallin was the first *in vivo* carbamylation site identified using tandem mass spectrometry [28]. Recently, Smith and co-workers reported N-terminal *in vivo* carbamylation of γB, γC, and γD crystallins [27]. We identified a total of 16 lysine and N-terminal modification sites in soluble and insoluble lens proteins from a 67-year-old patient. *In vivo* carbamylation of the N-terminal amino acid of γD crystallin was confirmed in this study and one *in vivo* carbamylated peptide could be matched to the N-terminal region of γB or γC crystallin. N-terminal carbamylation of γB and γC crystallins could not be distinguished in this study because the tryptic peptides from the N-terminal region of γB and γC crystallins have exactly the same amino acid sequence. γB crystallin is a minor component of the γ crystallin family constituting only 4% of the protein family. Hence, the carbamylated peptide identified in this study is probably originated from γC crystallin. Proteolytic enzymes with different cleavage specificities might reveal if the N-terminal glycine of γB crystallin is modified by *in vivo* carbamylation. Two more previously reported lysine modification sites, carbamylation of βB1 K^{160} and acetylation of αA K^{70} were also detected [17]. Interestingly, three *in vivo* lysine carbamylation sites and two acetylation sites were detected on γS crystallin in this study. Although many different types of post-translational modifications have been reported on γS crystallin, *in vivo* carbamylation and acetylation of γS crystallin have never been reported [36–39]. In previous reports of carbamylation, those sites have often shown a comparable level of acetylation [28]. Only ^{159}K of γS crystallin was identified as a co-modification site in this study. Among the 14 lysine modification sites identified in this study, only 4 lysine

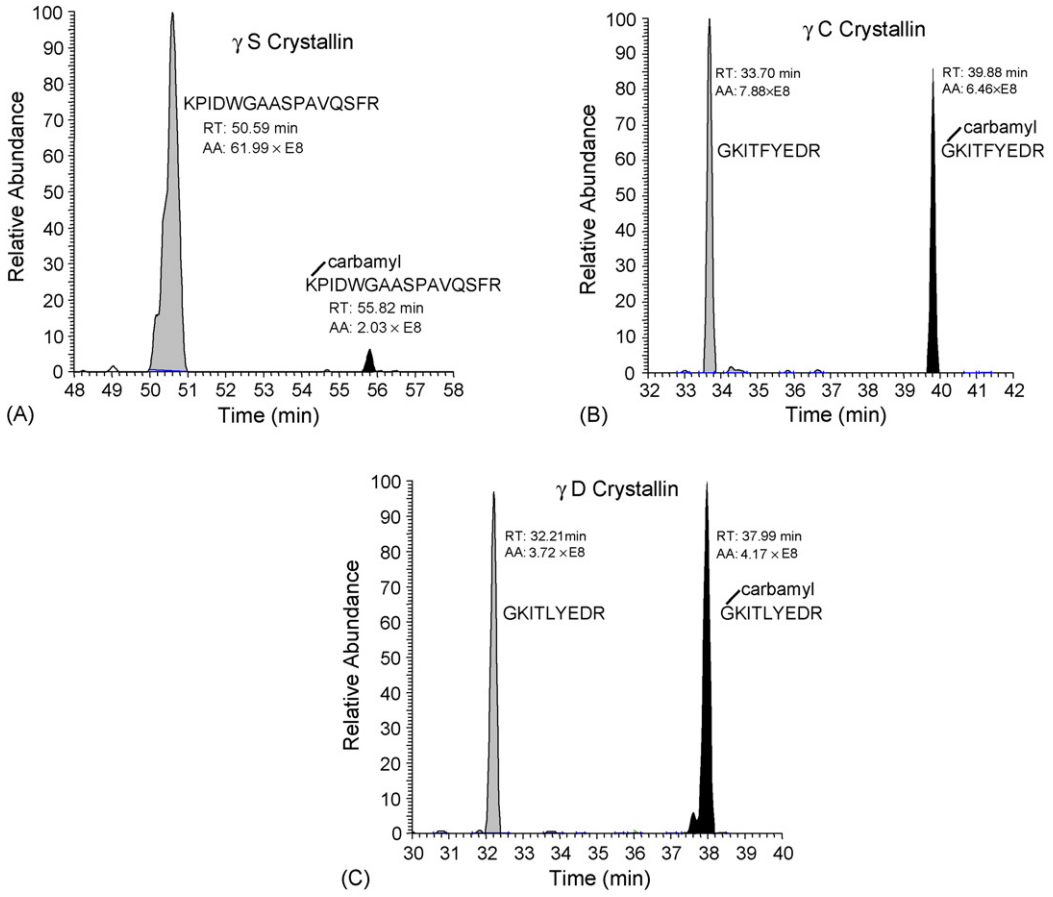

Fig. 10. Comparison of abundances of modified and unmodified peptides. Selected ion chromatograms were constructed and compared for *in vivo* carbamylated lysine containing tryptic digest peptide and unmodified lysine containing tryptic digest peptide of γS crystallin (A), for *in vivo* cabamylated N-terminal glycine containing tryptic digest peptide and unmodified N-terminal glycine containing tryptic digest peptide of γC crystallin (B), and for *in vivo* cabamylated N-terminal glycine containing tryptic digest peptide and unmodified N-terminal glycine containing tryptic digest peptide of γD crystallin (C). Gray shaded peak indicates ion current of unmodified peptides and black shaded peak indicates ion current of modified peptides.

residues were found to be acetylated. The relatively low level of acetylation in this lens sample might be a reason why fewer co-modification sites were observed.

Post-translational modifications of lens tissue proteins have been studied intensively to uncover key post-translational modifications involved in cataractogenesis [37–45]. Previous studies have mostly focused on well-separated single component lens proteins. Various protein separation techniques, such as reverse phase, size exclusion, and ion exchange chromatography were employed to separate individual component of lens protein mixtures and intact molecular weight of separated proteins measured by ESI–MS [46]. The types and level of post-translational modifications were easily determined by mass spectrometry because lens tissue proteins are mainly composed of a few classes of crystallins and several cytoskeleton proteins. Purified lens proteins were then separately subjected to enzymatic digestion and capillary RPLC tandem mass spectrometry analysis was used to identify modification sites. An advantage of the protein separation approach is that it enables relative quantitation of specific post-translational modification using mass spectrometric responses of modified and unmodified intact proteins. Although this kind of protein separation approach appears to be ideal for lens proteome studies, it has several limitations as well. Sample preparation procedures for protein separations is laborious and time-consuming and complete separation of proteins of different modification states cannot be always achieved. Chromatographic separation of proteins from old-age lens tissue samples have been a daunting task because a fairly large portion of the lens proteins remain unseparated even after extensive chromatographic separations [37,39]. Various types of aging-related post-translational modifications are significantly increased in old-age lens proteins, which make the proteins far more heterogeneous than young lens proteins. As a result, some post-translationally modified proteins of low abundances could be missed in an analysis. For example, carbamylated γS crystallin that appeared to be less than 4% of unmodified gamma S crystallin was never observed in previous studies (see Fig. 10). Moreover, accurate quantification of specific post-translational modification becomes difficult in proteins from old-age lens tissue as one class of lens protein could have fairly diverse post-translational modification status. On the contrary, shotgun proteomics approach which does not separate individual components of a protein mixture can identify various types of post-translational modifications from a sin-

gle analysis. Using this approach, MacCoss et al. identified 73 modification sites of 4 different post-translational modifications occurring in cataract lens samples from a single MuDPIT analysis [17]. The quantitative analysis information, *i.e.*, the ratio between modified proteins and non-modified proteins cannot be directly obtained in the shotgun proteomics; however, modification levels of specific amino acid residues can be estimated by comparing ion current signals of modified and unmodified peptides. Recent studies demonstrated that quantitative analysis obtained from this type of label free approach is almost as accurate as more elaborate isotope labeling approaches [47–49]. Quantitative analysis of post-translational modifications at the peptide level by shotgun proteomics is a reasonable approach for lens tissue samples because it could provide more comprehensive modification results for each modification site by having nearly all lens proteins in the sample subjected to modification analysis.

There are a few cases where one should be cautious in assigning *in vivo* carbamylation and acetylation sites using current database searching algorithms. As it was noted in the results section, the search score itself seldom fails to determine the exact location or identity of modifications. When other modifications, such as a modified N-terminal amino acid or a deamidation site are located nearby, insufficient fragment ions can make a clear distinction difficult. In such cases, manual assignments of tandem mass spectra are helpful in providing correct modification information.

5. Conclusion

The MS/MS fragmentation information was shown to be a reasonable method to distinguish between two modifications that are otherwise difficult to unambiguously differentiate by mass alone in ion trap mass spectrometers. Carbamylated peptides identified in this study all showed a significant neutral loss of the carbamyl group from the precursor ion, while none of the acetylated lysine containing peptides showed a neutral loss of an acetyl group. Previous studies showed that carbamylated peptides undergo a neutral loss of carbamyl group from parent ions or from fragment ions during the CID process and this study confirms those findings. By comparing the scores derived from differential modification searches and corroborating that information with high mass accuracy data, we have shown the ability to distinguish carbamylation from acetylation. This is possible because the ion trap mass spectrometer is better able to resolve the singly charge fragment ions in MS/MS mode as opposed to the multiply charged precursor ions observed in MS mode. In addition, this study confirms the presence of *in vivo* carbamylation on the lens proteins reported by Smith et al. and extends the presence of *in vivo* carbamylation to other lens proteins. The recently developed hybrid type ion trap mass spectrometers, such as the LTQ-Fourier transform MS and LTQ-Orbitrap can solve this problem as well, but not all researchers have access to these high performance instruments and a majority of shotgun proteomics data is routinely generated by using conventional ion trap mass spectrometers.

Acknowledgements

The authors gratefully acknowledge financial support from National Institute of Health grants RO1EY13288 and P41 RR11823-05.

References

[1] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 3349.
[2] J.K. Eng, A.L. McCormack, J.R. Yates 3rd, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[3] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Biochem. Biophys. Res. Commun. 195 (1993) 58.
[4] M. Mann, P. Hojrup, P. Roepstorff, Biol. Mass Spectrom. 22 (1993) 338.
[5] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.
[6] D.J. Pappin, P. Hojrup, A.J. Bleasby, Curr. Biol. 3 (1993) 327.
[7] J. Qin, D. Fenyo, Y. Zhao, W.W. Hall, D.M. Chao, C.J. Wilson, R.A. Young, B.T. Chait, Anal. Chem. 69 (1997) 3995.
[8] A. Shevchenko, O.N. Jensen, A.V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, M. Mann, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 14440.
[9] J.R. Yates 3rd., S. Speicher, P.R. Griffin, T. Hunkapiller, Anal. Biochem. 214 (1993) 397.
[10] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[11] O.A. Mirgorodskaya, Y.P. Kozmin, M.I. Titov, R. Korner, C.P. Sonksen, P. Roepstorff, Rapid Commun. Mass Spectrom. 14 (2000) 1226.
[12] M. Munchbach, M. Quadroni, G. Miotto, P. James, Anal. Chem. 72 (2000) 4047.
[13] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 6591.
[14] S.E. Ong, B. Blagoev, I. Kratchmarova, D.B. Kristensen, H. Steen, A. Pandey, M. Mann, Mol. Cell Proteomics 1 (2002) 376.
[15] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[16] H. Zhou, J.A. Ranish, J.D. Watts, R. Aebersold, Nat. Biotechnol. 20 (2002) 512.
[17] M.J. MacCoss, W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark, J.R. Yates 3rd, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 7900.
[18] M. Mann, O.N. Jensen, Nat. Biotechnol. 21 (2003) 255.
[19] H. Zhou, J.D. Watts, R. Aebersold, Nat. Biotechnol. 19 (2001) 375.
[20] H. Zhang, X.J. Li, D.B. Martin, R. Aebersold, Nat. Biotechnol. 21 (2003) 660.
[21] J.A. Wohlschlegel, E.S. Johnson, S.I. Reed, J.R. Yates 3rd, J. Biol. Chem. 279 (2004) 45662.
[22] G. Rosas-Acosta, W.K. Russell, A. Deyrieux, D.H. Russell, V.G. Wilson, Mol. Cell. Proteomics 4 (2005) 56.
[23] J. Peng, D. Schwartz, J.E. Elias, C.C. Thoreen, D. Cheng, G. Marsischky, J. Roelofs, D. Finley, S.P. Gygi, Nat. Biotechnol. 21 (2003) 921.
[24] Y. Oda, T. Nagasu, B.T. Chait, Nat. Biotechnol. 19 (2001) 379.
[25] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates 3rd, Nat. Biotechnol. 17 (1999) 676.
[26] C.C. Wu, M.J. MacCoss, K.E. Howell, J.R. Yates 3rd, Nat. Biotechnol. 21 (2003) 532.
[27] V.N. Lapko, D.L. Smith, J.B. Smith, Protein Sci. 12 (2003) 1762.
[28] V.N. Lapko, D.L. Smith, J.B. Smith, Protein Sci. 10 (2001) 1130.
[29] J. Lippincott, I. Apostol, Anal. Biochem. 267 (1999) 57.
[30] J. McCarthy, F. Hopwood, D. Oxley, M. Laver, A. Castagna, P.G. Righetti, K. Williams, B. Herbert, J. Proteome Res. 2 (2003) 239.
[31] R.G. Sadygov, J. Eng, E. Durr, A. Saraf, H. McDonald, M.J. MacCoss, J.R. Yates 3rd, J. Proteome Res. 1 (2002) 211.
[32] R.G. Sadygov, J.R. Yates 3rd, Anal. Chem. 75 (2003) 3792.
[33] D.L. Tabb, W.H. McDonald, J.R. Yates 3rd, J. Proteome Res. 1 (2002) 21.

[34] G. Van Driessche, I. Vandenberghe, F. Jacquemotte, B. Devreese, J.J. Van Beeumen, J. Mass Spectrom. 37 (2002) 858.
[35] C.C. Chao, S.L. Wu, W.M. Ching, Biochim. Biophys. Acta 1702 (2004) 145.
[36] V.N. Lapko, D.L. Smith, J.B. Smith, Biochemistry 41 (2002) 14645.
[37] V.N. Lapko, A.G. Purkiss, D.L. Smith, J.B. Smith, Biochemistry 41 (2002) 8638.
[38] S.R. Hanson, D.L. Smith, J.B. Smith, Exp. Eye Res. 67 (1998) 301.
[39] S.R. Hanson, A. Hasan, D.L. Smith, J.B. Smith, Exp. Eye Res. 71 (2000) 195.
[40] Z. Zhang, L.L. David, D.L. Smith, J.B. Smith, Exp. Eye Res. 73 (2001) 203.
[41] J.B. Smith, G.A. Shun-Shin, Y. Sun, L.R. Miesbauer, Z. Yang, Z. Yang, X. Zhou, J. Schwedler, D.L. Smith, J. Protein Chem. 14 (1995) 179.
[42] Z. Ma, S.R. Hanson, K.J. Lampi, L.L. David, D.L. Smith, J.B. Smith, Exp. Eye Res. 67 (1998) 21.
[43] A.L. Lund, J.B. Smith, D.L. Smith, Exp. Eye Res. 63 (1996) 661.
[44] V.N. Lapko, R.L. Cerny, D.L. Smith, J.B. Smith, Protein Sci. 14 (2005) 45.
[45] J. Feng, D.L. Smith, J.B. Smith, J. Biol. Chem. 275 (2000) 11585.
[46] Z. Zhang, D.L. Smith, J.B. Smith, Proteomics 1 (2001) 1001.
[47] A.D. Hegeman, A.C. Harms, M.R. Sussman, A.E. Bunner, J.F. Harper, J. Am. Soc. Mass Spectrom. 15 (2004) 647.
[48] R.E. Higgs, M.D. Knierman, V. Gelfanova, J.P. Butler, J.E. Hale, J. Proteome Res. 4 (2005) 1442.
[49] W. Wang, H. Zhou, H. Lin, S. Roy, T.A. Shaler, L.R. Hill, S. Norton, P. Kumar, M. Anderle, C.H. Becker, Anal. Chem. 75 (2003) 4818.