Novel Bioinformatics Methods for Identification and Quantification of

Deamidation in Shotgun Proteomics Experiments

By

Surendra Dasari

A DISSERTATION

Presented to the Department of Medical Informatics and Clinical

Epidemiology and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

May 2008

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Surendra Dasari

has been approved

_____
Mentor

_____
Member

_____
Member

_____
Member

_____
Member

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

Reliable identification of deamidation remains challenging when using low-resolution mass spectrometry due to the MS/MS mis-triggers on first isotopic peak of unmodified peptide and limited mass resolution. We developed a bioinformatics method that utilizes differential chromatographic behavior and corrected peptide masses of the amidated and deamidated peptides to validate deamidation identifications. The method has been automated to facilitate rapid validation of deamidation identifications in large-scale proteomics experiments. Accurate quantification of deamidation is necessary for reliable protein pharmaceutical shelf life measurements, minimizing deamidation in sample processing/purification methods, and differentiating between healthy and diseased tissue in protein aggregation diseases. Deamidation quantification is complicated by co-elution of modified and unmodified peptide forms, interference from other peptides with similar elution times, and poor chromatography peak shapes. We developed a robust mathematical quantification technique that uses Gaussian isotopic envelope modeling with peptide mixture models. This method estimates the abundance of deamidation by comparing the predicted isotopic envelope of deamidated and amidated peptide forms to the experimentally measured isotopic envelope. The technique has undergone extensive manual validation during development and a semi-automated graphical user interface has been designed to estimate the abundance of deamidations in large-scale proteomics experiments. We have applied these methods to estimate deamidation abundances in an aged series of human lens tissues. Increased deamidation abundance at three sites in γS lens crystallin is correlated with increasing age, loss of protein solubility, and changes in 3-dimensional protein structure.

1. BACKGROUND

1.1 Genomics: Genomics is the large-scale study of genomes (DNA blueprints) of organisms. Advances in DNA sequencing techniques have created the possibilities of complete genome determination and even studies of genome-wide gene expression changes (functional genomics). It was anticipated that there would be upwards of 100, 000 human genes before the human genome project was started. Surprisingly, only about 25,000 genes were found[1]. Alternative splicing, RNA editing, and post-translation modifications have been proposed as mechanisms to increase human protein complexity. The static information contained in genomic sequences tells only a portion of the story. Thus, a new field of science called proteomics was developed to directly study these gene products (proteins) which carry out all physiological functions[2, 3].

1.2 Proteomics: Proteomics is the large-scale study of the full set of proteins (present in cells or tissues or biological fluids) and their expression changes during various conditions[3, 4]. An organism's proteome is more complex than its genome due to its dynamic biochemical interactions with genome and environment. The proteomics field is extremely active with many areas of research world-wide: protein sequencing, protein-protein interaction, protein structure, protein function, protein quantification, and protein lifecycle[5] to name a few. Many traditional biochemical techniques have been adapted for use in these proteomic studies. For example, Edman degradation was developed to sequence proteins and the yeast two-hybrid system is used to study protein-protein interactions[5]. The challenge of proteomic studies have resulted in many new biological technical developments as well. The recent availability of genomic sequences for many organisms has been an essential component of high-throughout (shotgun) mass

spectrometry based proteomics. These shotgun techniques allow analysis of the proteins in complex biological samples like cell lysates, serum, and many tissues including human lens.

1.3 Shotgun Proteomics: Tandem mass spectrometry plays a major role in shotgun proteomics[4]. Protein sample complexity is reduced using fractionating techniques like gel-based or liquid-based electrophoresis (Figure 1A). Proteins in the fractions are digested into peptides using proteases (Figure 1B). Digested peptides are time separated using reverse-phase liquid chromatography (RP-HPLC), ionized using electro spray ionization (ESI), and a mass spectrum (MS) of their corresponding masses is acquired (Figure 1C). A peptide ion is selected from the resulting MS spectrum (Figure 1D), fragmented using collision-induced dissociation, and a mass spectrum of its constituents is acquired (tandem MS/MS). One of the most successful applications of bioinformatics in proteomics is determining which proteins are present in a sample. The proteins are inferred from peptide amino acid sequences determined from tandem MS/MS scans (Figure 1E). However, simple lists of protein identifications only give you a partial understanding of biological processes.

1.4 Posttranslational Modification (PTM) of Proteins: Many proteins are posttranslationally modified *in vivo*. Protein PTMs play an important role in several normal physiological processes. For example, phosphorylation is known to play a key role in protein signaling[6], acetylation is known to regulate protein expression[7], ubiquination is known to play an important role in protein degradation,[6] and deamidation is known to time the function of a protein[8]. Protein modifications are also known to play

an important role in several pathophysiological processes. For example, accumulation of deamidation in tissues is linked to aging[9], O-glycosylation of intracellular proteins is linked to type II diabetes[10], and citrullination of arginine is linked to rheumatoid arthritis[11] and multiple sclerosis[12]. Thus, identification and quantification of PTMs is vital to understand their role in both normal and patho- physiological processes.

1.5 Human Lens: One tissue where PTMs are of particular importance is the extremely long lived human ocular lens tissue. The human lens tissue predominantly consists of eleven highly soluble proteins known as crystallins. The crystallins in mature lens fiber cells are organized at very high concentrations to maintain lens transparency for decades. The proteins in the fiber cells do not turnover and, as the lens grows, new proteins accumulate on the surface with mature proteins at the core of the lens (nucleus). Given the age of the lens, the mature proteins present in the lens nucleus may accumulate several posttranslational modifications (PTMs). Age-related accumulation of modifications in crystallin proteins contributes to loss of their structure and solubility, which may lead to formation of aggregates[9, 13, 14]. Age-related accumulation of modifications in lens is linked to loss of its functions like, gradual yellowing of lens with age, hardening of lens (loss of focus) at middle age, and formation of nuclear cataracts at an advanced age[15]. Cataracts is one of the leading causes of blindness worldwide[15]. Characterization of PTMs that contributes to age-related malfunction of lens is essential to develop appropriate preventive and therapeutic strategies.

1.6 PTM Characterization using Shotgun Proteomics: A posttranslational modification of an amino acid changes its mass, and the obvious instrument to measure

mass changes is a mass spectrometer. This is demonstrated using an example methionine containing peptide (APVIHQEMIGGLR). When methionine residue in the peptide is oxidized (APVIHQEM+16IGGLR) then all the fragment ions in tandem MS/MS that contain oxidized methionine will be shifted by +16 Daltons. This mass shift information can be used to identify PTMs present in peptides. A typical tandem mass spectrometry experiment produces ~ 1-2 million tandem MS/MS spectra due to recent advances in instrumentation. The size of such datasets precludes the possibility of using manual interpretation to determine peptide sequences and possible PTMs present in tandem MS/MS. Hence, several algorithms have been developed that computationally match the acquired tandem MS/MS spectra against known proteins and their corresponding peptides to identify the peptides, PTMs and proteins present in the sample.

1.7 Algorithmic Approaches for Peptide and PTM Identification: Search algorithms[16-25] have been developed to match tandem mass spectra to peptide sequences in protein databases. These search algorithms fall in four categories: database searching[16, 18, 24], partial tag based matching[3 19, 26, 27], error tolerant de novo sequence alignment[21 17, 25, 28] and error-tolerant spectral alignment[22, 29, 30]. Database search programs like SEQUEST[18] and X! Tandem[16] match experimental spectra to theoretical spectra generated from an *in silico* enzymatic digest of the protein database. Partial tag based search engines use the fragment ion information present in an experimental tandem mass spectrum to derive partial 3-4 amino acid sequence tags. The derived partial tags and peptide mass information are used to match potential peptide and protein sequences in the database. Error-tolerant database search engines derive possible peptide sequences present in the tandem mass spectrum using de novo sequencers[28, 31-33]. The derived de

novo sequences often contain errors due to missing fragment ions in the mass spectra, and presence of isobaric and posttranslationally modified amino acids in the peptide. Search engines match such sequences to peptides in the protein database utilizing error-tolerant methods[17, 21, 23, 25]. Any mass difference (ΔM) between the theoretical mass of the matched peptide sequence and its measured mass is used to detect possible posttranslational modifications (PTMs) in the peptide. Error-tolerant spectral alignment algorithms[22, 29, 30] perform a gapped alignment between the spectrum masses and the prefix masses of candidate peptide sequences in the database. The mass gaps in the alignment are interpreted as modifications present in the peptide sequence. These search algorithms have been successfully applied to biological samples to detect PTMs on a large scale[22, 34, 35].

1.8 Deamidation as Major PTM in Aging Human Lens Tissue: Several proteomic studies have identified many posttranslational modifications in human lens including acetylation[34-39], carbamylation[40], methylation[35, 39-41], deamidation[34, 35, 42-51], oxidation[34, 35, 37, 50-53], phosphorylation[35, 37, 53, 54], proteolysis[13, 49-51, 53, 55], and UV filter adducts[56-58]. The limited number of crystallin proteins and their well documented PTMs makes the lens an ideal mixture for development of shotgun proteomic analysis techniques to characterize PTMs in complex samples.

Recently, Wilmarth *et. al.*[35] performed a most comprehensive shotgun proteomic analysis of an aged lens sample to find novel lens PTMs and compare lens proteins by solubility to identify PTMs that were linked to loss of solubility. In that study, a 93-year old cataractous lens sample was homogenized and water soluble and water insoluble

portions of the lens were separated using centrifugation. Both fractions were analyzed using 2-DLC based mass spectrometry. Mass spectra were acquired using a LCQ Classic ion trap instrument and a total of 151 *in vivo* PTM sites were identified in major lens proteins. Relative abundances of detected PTMs were estimated by counting the number of tandem MS/MS spectra that contained the PTM site (spectral counting). Deamidation emerged as the most abundant PTM in aged lens and its frequency almost doubled in the water-insoluble fraction[35]. This supported the theory that accumulation of deamidation in aged tissues may denature proteins or disrupt favorable protein-protein interactions and result in loss of protein solubility[35, 44, 50, 51, 59].

1.7 Significance of Deamidation: Deamidation of Asparagine (N) and Glutamine (Q) residues can occur at physiological pH via spontaneous, non-enzymatic reactions that converts the primary amines (N and Q) to acids (D and E)[8]. Deamidation of a peptide leads to a change in its mass (0.98 Daltons) and the pI. Several succinimide intermediates can be formed during Aspargine deamidation, which then undergo isomerization and/or racemization, resulting in several deamidation end products[8, 60]. Previous *in vivo* and *in vitro* deamidation studies have shown evidence for the formation of β−D (iso-aspartate) as an end product of Asparagine deamidation[8, 60, 61], and an increase in the abundance of D-Aspartic acid has been observed in aged tissues[62]. However, there is no *in vivo* evidence for formation of β-E during Glutamine deamidation.

Deamidation, irrespective of the isomeric nature of the end product, introduces a charged residue at physiological pH that may alter protein structure, and deamidation has been suggested as one mechanism to time protein function[8]. Proteins that must function

for long periods of time, such as lens crystallins, may accumulate deamidations that can result in loss of protein structure, solubility, and function[9, 63, 64] . Deamidation has been linked to protein aggregation, spontaneous peptide bond cleavage, and inactivation of protein pharmaceuticals[42, 65-68]. Accumulation of iso-aspartate in neuronal tissues and spontaneous peptide blond cleavage of neuro- peptides and/or proteins has been linked to neuro-degenerative diseases like Alzheimer's[69-72]. Enhanced T-cell activation of deamidated gluten proteins in the intestine has been linked to the celiac disease (CD)[73, 74]. Deamidation of food proteins is a particular concern for the food industry as it releases ammonia (substrate for Maillard browning reaction) which forms aroma compounds and pigments[75]. Detection of deamidation sites in specific sequence motifs are often used to identify N-linked glycosylation sites in proteomic studies[76]. Deamidation can also occur during β-elimination procedures in phosphorylation studies[77]. Thus, reliable methods to detect sites of deamidation would be beneficial for many proteomic applications beyond lens studies.

1.8 Quantification of Deamidation: Recent technological advances have increased mass spectrometer sensitivity by an order of magnitude (picomol range). It is routine to detect proteins and PTMs present in a biological sample from very high (μmol) to low (picomol) abundance levels. The biological relevance of a PTM often depends on its abundance levels. This is especially true for deamidation, which is a spontaneous process with half-lives of amide to acid conversion anywhere between 1 hour to 100 years depending on the location of corresponding amide residue in the protein[8, 78]. Protein damage could occur by extensive deamidation at one site or by accumulated lower levels of deamidation at several sites. Hence, accurate quantification is necessary to correlate

the extent of deamidation to its effects on structure and function of proteins[8]. Precise quantification of deamidation is also necessary for reliable protein shelf life measurements[79, 80], evaluating separation/purification methods to minimize deamidation, and also distinguishing diseased tissue (protein aggregation diseases) from healthy tissue[8, 45].

2. INTRODUCTION

Large-scale proteomics experiments generate vast amounts of information about peptide, protein and PTM's present in biological samples. Reliable tools to detect sites of deamidation and to quantify deamidated peptides are needed in large-scale proteomic applications.

2.1 Reliable Deamidation Identification: Deamidation can be identified using traditional database search tools, such as SEQUEST[18, 81] or X! Tandem[16], configured to search for variable modifications of 0.984 Dalton on N and Q residues. Since parent ion mass tolerances in these types of searches would typically be greater than 0.984 Da, the reliable detection of the deamidation site will depend on small mass shifts in fragment ions caused by deamidation. As the fragment ion spectrum of a deamidated peptide closely resembles that of its corresponding unmodified form, both peptide forms may have nearly identical search scores and be difficult to distinguish. Enhanced cleavage C-terminal to aspartic acid may be observed following asparagine deamidation, but only for peptide charge states where there are no mobile protons[82-84]. Since programs usually report only the interpretation with the highest search score, inconsistent identification of the correct peptide form may occur, particularly as spectral quality diminishes.

Error tolerant[17, 21, 23, 25] and tag-based[19, 20, 26] search engines are capable of deamidation detection using partial sequences (*de novo* sequences or tags) derived from the experimental tandem mass spectra. These strategies rely on reliable detection of the 0.984 Dalton mass shifts caused by deamidation in experimental MS/MS spectra. However, a mis-triggered MS/MS scan on the first isotopic peak of an unmodified peptide also introduces a similar (1.003 Da) mass shift and true deamidation can be difficult to determine in the presence of the isotopic peaks from the unmodified peptide. Thus, these tools suffer from many of the same pitfalls as traditional database search programs.

Deamidation introduces two changes in physical properties of a peptide: a shift in mass (0.984 Daltons) and pI relative to the unmodified peptides. Most of the current search programs rely on the small mass shift to detect deamidation in a peptide. Modifications with small mass differences, such as deamidation, are usually excluded from these algorithms because of naturally occurring isotopic peaks or limited mass accuracy of some instruments. Hence, detection of deamidation on low-resolution mass spectrometers, like ion traps, is difficult. Deamidated forms of peptides have been reported to elute later than their corresponding unmodified forms when using shallow reverse-phase acetonitrile gradients[42, 85, 86]. The differential RP-HPLC elution behavior is independent of mass resolution of the instrument used for detecting the peptide sequence. This additional experimental evidence can be used to validate the deamidated peptide identifications in low resolution mass spectrometry based shotgun experiments. This strategy has been used by one search program[87] to decrease the possible posttranslational

modifications to be considered for a specific ΔM between matched peptide and its measured mass. However, retention time shift (RTS) between deamidated peptides and their unmodified forms in two-dimensional liquid chromatography (2-DLC) separations has never been quantified nor utilized. The use of RTS to reliably identify deamidations on a large scale has also never been demonstrated.

High resolution data acquisition modes (zoom scans) available on low resolution ion traps have also been successfully used to study deamidated peptides[78], but reduce the duty factor too severely in shotgun proteomic experiments where maximizing the numbers of MS/MS scans is important. Incorporating additional experimental data to validate deamidation identifications can be automated and large numbers of deamidation sites in shotgun experiments can be successfully detected.

2.2 Deamidation Quantification: Protein and peptide quantification methods can be broadly categorized into two classes: label-free quantification and stable isotope label based quantification[88]. Proteins and their corresponding deamidations in shotgun experiments can be quantified using label-free quantification techniques such as spectral counting[89, 90] or extracted ion chromatogram integrations (XIC). It has been shown that MS/MS spectral count is proportional to analyte concentration in complex mixtures analyzed by 2-DLC[91]. The spectral counting method estimates the ratio of overall protein deamidation by counting the total number of MS/MS spectra assigned to deamidated peptide forms versus total number of MS/MS spectra matched to the protein[34, 35]. This technique can be extended to quantify modifications at the peptide level in some cases,

10

but often fails when spectral counts are too small (due to low peptide abundance or low deamidation abundance).

A more traditional label-free quantification method is integrating ion currents (XIC) of deamidated peptides and their corresponding unmodified forms[92, 93]. However, there are several factors such as co-elution of peptide forms that have overlapping m/z signals with similar elution times, and poor chromatography peak shapes that complicate XIC integrations. Unmodified and deamidated peptides may also have different ionization efficiencies or undergo selective losses during sample processing.

Many of these complications are eliminated in stable isotopic labeling techniques used to compare two different samples[94-96]. Using these techniques, peptides in two different samples are labeled with heavy and light isoform tags (SILAC[96], ICAT[97], and O^{16}/O^{18}) respectively. Labeled samples are multiplexed and analyzed by mass spectrometry. The MS ion current of heavy and light labeled peptide species are used to perform relative quantification of protein present in the samples. Stable isotopic labeling can not be used for deamidation quantification due to software and experimental limitations. Existing stable isotopic label based quantification software assumes complete co-elution of labeled peptides and complete separation (>5 Daltons) of corresponding MS ion current signals. However, unmodified and deamidated peptides partially co-elute and their corresponding MS ion current signals are partially separated (by 1.0 Dalton). SILAC can only be used on cultured cells, not on tissue sample. ICAT only labels peptides that contain Cysteine and would not be sensitive to most sites of deamidation in crystallins. O^{16}/O^{18} causes small mass shifts of labeled and unlabeled peptides and would be far too

complex to analyze if deamidations were present. Also, many chemical labeling protocols could introduce artificial deamidation during sample processing.

2.3 Instrument Requirements for Deamidation Quantification: Some complications of XIC integration can be eliminated or reduced by determining ion currents from extracted mass spectra rather than using chromatogram peak integrations. Individual isotopic distributions of unmodified and deamidated forms of a peptide present in a hypothetical extracted mass spectrum are shown in Fig. 2A-I and Fig. 2A-II, respectively. The isotopic envelopes of both forms are separated in M/Z space by 1 Dalton or less than 1 Dalton depending on the charge state of the peptide. Hence, both distributions overlap strongly in combined mass spectra acquired by a mass spectrometer (shown in Fig. 2A-III). A robust mass spectrum (MS) based deamidation quantification method has to rely on the theoretical peptide isotopic envelopes of unmodified and deamidated forms to deconvolute the combined mass spectrum (Fig. 2A-III) into corresponding individual signals (unmodified –Fig. 2A-I and deamidated –Fig. 2A-II). The ion current in the deconvoluted signals is used to quantify deamidations. The method will be called isotopic envelope modeling (IEM).

Accuracy of the IEM method and sensitivity of peptide detection are two key factors in MS based deamidation quantification. The resolution of the mass spectrometer directly affects the accuracy of IEM and inversely affects the sensitivity of peptide detection. Theoretical unmodified, deamidated, and combined mass spectra of a peptide acquired on a high resolution (QTOF) and low resolution (linear ion trap) mass spectrometer are shown in Fig. 2A and Fig. 2B, respectively. The baseline in the

12

combined mass spectrum increases and the peak separation decreases at lower mass resolution (compare Fig. 2A-III and Fig. 2B-III). Lower resolution also decreases signal to noise, and these factors make it hard to employ the IEM technique for deamidation quantification when using low resolution mass spectrometers. Recently, the 19 mDa mass difference between deamidation (0.984 Da) and isotopic peaks (1.003 Da) has been used in ultra high resolution Fourier transform (FT) instruments to unambiguously detect deamidation[98, 99]. MS based deamidation quantification can be performed without using the deconvolution process (IEM) when using such ultra high resolution mass spectral data. However, all FT instruments employ long scan times to acquire ultra high resolution data on a peptide, thereby lowering its sensitivity of peptide detection. The QTOF mass spectrometer strikes a good balance between mass accuracy and peptide detection sensitivity making it an ideal instrument for MS based deamidation quantification.

2.3 Bioinformatics Tools to Assist Deamidation Quantification: Deamidated and unmodified peptides are detected in multiple peptide or protein fractions of a typical shotgun experiment. This is because individual peptides often elute across several first dimension ion exchange fractions, resulting in the same peptide appearing in multiple reverse phase runs when mass spectra are acquired. Deamidation peptide quantities have to be calculated by integrating ion currents of corresponding forms across all analyzed fractions. This requires full integration of peptide identification (SEQUEST) information into the deamidation quantification method (IEM). Currently, there is no deamidation quantification software that can handle such multi-dimensional datasets in a combined, elegant user interface.

13

In this thesis, we present two novel bioinformatics methods for both robust identification and quantification of deamidations. The methods were prototyped, tested, and developed into computer tools that aid in rapid identification and quantification of deamidations present in large-scale shotgun proteomics experiments.

3. NOVEL BIOINFORMATICS METHODS TO IDENTIFY DEAMIDATION IN SHOTGUN PROTEOMICS STUDIES.

3.1 INTRODUCTION

In order to identify as many peptides in complex mixtures, shotgun proteomics experiments typically try to maximize the number of tandem mass spectrometry scans. Ion traps have the highest sensitivity and fastest scan speeds, but the low resolution mass measurements make detecting a small mass shift (0.984 Da) PTM like deamidation very difficult. Using synthetic peptide mixtures of unmodified and deamidated peptide forms, we show that bioinformatics peptide identification software cannot reliably detect deamidation without use of additional experimental evidence even when peptide forms are completely chromatographically resolved. A novel bioinformatics tool was developed to automatically use this additional evidence to correct SEQUEST results and identify truly deamidated peptides.

Deamidated forms of peptides are known to elute later than their non-deamidated forms when using shallow reverse-phase acetonitrile gradients[42, 85, 86], and this retention time difference can be used as a deamidation signature. Lens tissue was chosen for this study, since it contains a particularly rich source of known deamidated peptides[22, 34, 35, 100, 101]. Synthetic peptides were used to find chromatography conditions where unmodified and deamidated peptides co-eluted during first dimension low pH strong cation separation (SCX), and were completely resolved during second dimension reverse phase (RP). Peptides from a 70-year old human lens were used as training data to determine reverse-phase retention time shift (RTS) between unmodified and deamidated forms of peptides. Averaging survey scans acquired within chromatogram peaks of each peptide

form resulted in more accurate measurements of the difference in mass (ΔM) between the unmodified and deamidated peptides. The RTS and ΔM data were then used to filter SEQUEST search results, resulting in a dramatically improved accuracy and sensitivity in detection of deamidation. Using RTS and ΔM constraints resulted in >93% deamidation detection accuracy. This approach was automated and a total of 72 crystallin deamidation sites, 18 of which were not previously reported in human lens tissue, were detected in the first large scale deamidation site mapping of aged human lens tissue.

3.2 EXPERIMENTAL

3.2.1 Sample Processing: Synthetic peptides corresponding to human βB1 crystallin (SwissProt P05813) residues 150-159 (ISLFEGA<u>N</u>FK), residues 202-213 (GY<u>Q</u>YLLEPGDFR), and 214-229 (HW<u>N</u>EWGAF<u>Q</u>PQM<u>Q</u>SLR), and their deamidated forms (ISLFEGA<u>D</u>FK, GY<u>E</u>YLLEPGDFR, HW<u>D</u>EWGAFQPQMQSLR, HWNEWGAF<u>E</u>PQMQSLR, and HWNEWGAFQPQM<u>E</u>SLR, respectively) were obtained from Sigma Genosys (The Woodlands, Texas) and diluted to 1 pmol/μl in 5% formic acid. Thirty micrograms of each peptide were injected either independently or simultaneously for SCX separations using conditions described below. Ten picomoles of each peptide were also analyzed by RP LC-MS, either alone or by simultaneously injecting equal mixtures of unmodified and deamidated forms of the same peptide.

Lenses from 3-day, 2-year, 18-year, 35-year, 70-year old human donors with healthy vision and a 93-year old human donor with type-III nuclear cataracts (Pirie scale) were obtained from the Lions Eye Bank of Oregon with Oregon Health & Science University's IRB approval. To assess the accuracy of deamidation detection using

16

SEQUEST with ion trap data, we analyzed a 3-day old lens (where little deamidation was expected) under identical conditions to the more extensively deamidated aged lens samples [35]. Lens proteins were separated into water-soluble and water-insoluble fractions by centrifugation, as previously described[35]. The amount of water-insoluble material from the 3-day old lens was negligible and not analyzed. Following a BCA protein assay (Pierce Biotechnology, Inc. Rockville, IL), 2.5 mg portions of proteins were reduced and alkylated in the presence of urea, digested overnight with trypsin, digested peptides solid phase extracted, and peptides separated by SCX chromatography.

The SCX separation at low pH was performed using a 100 x 2.1 mm polysulfoethyl A column (The Nest Group, Inc. Southborough, MA, USA). Mobile phase A contained 10mM sodium phosphate (pH 3.0) and 25% acetonitrile. Mobile phase B was identical except that it contained 350 mM KCl. After 5 min to load and wash, the separation was performed with a 45 min linear gradient from 0% to 50% B followed by a 20 min linear gradient from 50% to 100% B using a 0.2 ml/min flow rate. One minute fractions were collected.

3.2.2 LC-MS/MS Analysis: Both synthetic peptides and 10% portions of SCX fractions of lens digests were similarly analyzed by LC-MS using an Agilent 1100 series capillary LC system and an LCQ Classic ion trap mass spectrometer (ThermoFinnigan, San Jose, CA, USA). Samples were applied at 20 µl/min to a trap cartridge (Michrom Bioresources, Inc., Auburn, CA), and then switched onto a 0.5 X 250 mm Zorbax SB-C18 column (Agilent Technologies, Palo Alto, CA) using a mobile phase containing 0.2% acetic acid. The gradient used 7-35% acetonitrile over 90 min at a 10 µl/min flow

rate. Survey and MS/MS scans were collected in centroid mode. Data dependent collection of MS/MS spectra used the dynamic exclusion feature of the instruments control software (exclusion mass width of 3.0 Da, repeat count of 1, exclusion list size of 25 ions, and exclusion duration of 3 min) to obtain MS/MS spectra of the three most abundant parent ions following each survey scan. Due to exclusion list saturation, the effective exclusion time was approximately 90 s for the lens samples. The exclusion duration was shortened to 30 s for the synthetic peptides to increase the number of MS/MS scans for analysis. Since the mass shift due to deamidation was small, repeat counts greater than 1 or dynamic exclusion duration times greater than the RTS of the unmodified and deamidated forms of a peptide may prevent MS/MS scans of deamidated peptides from occurring. Thus, it was vital that the effective exclusion duration, exclusion list size, and repeat count settings were appropriate for deamidation detection. DTA files were generated using Bioworks Browser software (version 3.1 SR1, ThermoFinnigan) with a molecular weight range of 400 to 4000 Da, a minimum of 25 ions, and a low TIC threshold of 500. Charge state analysis (ZSA algorithm, ThermoFinnigan) was performed following DTA creation; however, the CombIon feature (ThermoFinnigan) was not used to avoid the averaging of MS/MS spectra of deamidated and unmodified peptides. A total of 664,497 tandem mass spectra were generated from all LC-MS/MS experiments.

3.2.3 Peptide and PTM Identification: Peptides were identified by searching tandem mass spectra against a human subset of the Swiss-Prot protein database (version 46.6, 12655 entries) using SEQUEST[18] (version 27, rev. 12, ThermoFinnigan). Searches were configured to use monoisotopic masses, the parent ion mass tolerance was 2.5 Da, a

static mass of 57.02 Da was specified for cysteine residues due to alkylation, and tryptic cleavage was specified. SEQUEST was also configured to look for the following variable modifications in peptides: deamidation (+0.984 Da on N or Q); oxidation (+15.995 Da on M or W); and acetylation (+42.011 Da) on the N-terminus of peptides. The tryptic search results were used to produce a subset database of human lens proteins using Contrast[102]. This database contained all proteins (228 total) identified by three or more peptides exceeding DTASelect[102] (version 1.9, The Scripts Research Institute) default thresholds in any of the lens fractions. SEQUEST searches were repeated against the human lens database without enzymatic cleavage specificity keeping all other parameters the same as in the tryptic searches. A list of identified peptides, proteins, and potential modifications was generated using DTASelect configured to use Xcorr thresholds of 1.8, 2.5, and 3.5 for 1+, 2+ and 3+ parent ions, respectively, and to select full or half-tryptic peptide termini. A minimum DeltaCN value of 0.0 instead of the DTASelect default of 0.08 was used so that deamidations or peptides having multiple potential modification sites would not be filtered out. Peptides identified by SEQUEST containing modified amides that had a 1+ ion m/z value of less than 2000 were considered for further retention time analysis described below.

3.2.4 Manual Deamidation RTS Analysis: Raw data files were converted into mzXML[103] format files and extracted ion chromatograms (XIC) of peptides were visualized using a locally developed program called RAPID[104]. A manual procedure was employed to calculate the experimental RTS and ΔM values between the deamidated peptides and their corresponding unmodified forms (Figure 3). Theoretical peptide

masses of the deamidated and unmodified peptides were tabulated from SEQUEST results and were used to generate XICs for singly-charged ions in each RP run.

Manual chromatographic peak assignments in the 3-day soluble, 70-year soluble and 70-year insoluble reverse-phase runs were based on confirming MS/MS scans located within each chromatogram peak, the relative elution order of unmodified and deamidated peptides, and survey scan masses averaged across each peak. If peaks for both the unmodified peptide and putative deamidated peptide could be distinguished from background noise, the retention times of the two peaks were tabulated. Experimental peptide masses were determined by averaging MS survey scans within each chromatogram peak, and picking the most intense centroided mass values within -1.0 to +4.0 Dalton intervals around the theoretical masses of each peptide form. The ΔM values were determined using these averaged experimental masses rather than from the parent ion masses listed in the respective DTA files. Putative deamidated peptide assignment was made if the chromatogram peak eluted later than the unmodified reference peak (an RTS between 1 and 7 min) and the ΔM value was nominally 1 Da greater (ΔM between 0.6 and 1.5 Da).

3.2.5 Automated Deamidation RTS Analysis: An algorithm that automated the manual procedure was also developed and implemented. The program (AutoDeamidValidator) accepts mzXML[103] and OpenSea[34] formatted XML files for raw data and MS/MS results, respectively. Scripts were written to convert SEQUEST output into OpenSea XML files. Peptide identifications that had deamidation modification sites and 1+ ions with m/z values less than 2000 were extracted from the MS/MS results.

Corresponding unmodified peptide identifications present in the same reverse-phase run were also located. The program then generated XICs which were smoothed using a Savitzy-Golay filter[105] of a 3rd degree polynomial with a window width of 9 seconds in 2 cycles to remove noise, and the resulting XICs were background subtracted. Peaks in the XIC were located by finding slope variations in the first derivative of the smoothed XIC. The RTS and ΔM values between unmodified peptides and putative deamidated peptides were then automatically calculated using the same steps as in the manual procedure (Figure 3).

3.3 RESULTS

3.3.1 Synthetic Peptides: Since co-elution of unmodified and deamidated peptide forms would likely confound search programs, we used shallow acetonitrile RP gradients that allowed resolution of unmodified and deamidated forms of the synthetic peptide pairs listed in Table 1. In all cases (2 asparagines and 3 glutamines), the deamidated form of the peptides had RP elution times that were approximately 3.0 min later than their corresponding unmodified form, regardless of amide or location within the peptide. Figure 4 shows XICs during RP chromatography for two of the synthetic peptide pairs listed in Table 1. An asparagine/aspartic acid pair (ISLFEGA<u>N</u>LK/ISLFEGA<u>D</u>LK) is shown in Fig. 4a, and a glutamine/glutamic acid pair (GY<u>Q</u>YLLEPGDFR/GY<u>E</u>YLLEPGDFR) is shown in Fig. 4b. MS/MS spectra acquired during the separations were searched using SEQUEST, configured with deamidation as a variable modification, to see if reliable deamidation identification could be achieved. Figure 4 shows the retention times of the MS/MS scans where unmodified (arrows) and deamidated peptides (triangles) were detected. A total of 22 MS/MS spectra were

identified as deamidated in the two separations: 15 were correctly identified during elution of the deamidated peak (the peptide containing the acidic residue) and 7 were incorrectly identified during elution of the unmodified peak. Only 68% of the reported deamidations were actually associated with the acid-containing peptides; a clear demonstration of the difficulty in identifying deamidation.

Complex peptide mixtures often require multiple chromatographic separations and retention times for the peptide pairs during SCX separations are also listed in Table 1. The average RTS during SCX was determined as -0.12 min and the full width at half maximum of each peak during SCX separation was approximately 2 min. At pH 3.0, E and D residues are largely protonated and deamidated peptides do not readily resolve from their corresponding amide forms. Thus, under the 2-DLC conditions commonly used to separate complex mixtures of peptides, at least partial co-elution of unmodified and deamidated peptide forms would occur, even when collecting 30 or more fractions prior to performing the second dimension reverse phase separation. Therefore RTS information from 2nd dimension RP separations can be used to identify deamidation in shotgun proteomics experiments such as these.

3.3.2 Complex Mixtures: To test whether mixtures of unmodified and deamidated peptides could be separated in more complex samples, RTS values were manually determined for a large number of potentially deamidated peptides in 3 lens protein digests separated by 2-DLC (3-day soluble, 70-year soluble, and 70-year insoluble).   SEQUEST search results were filtered for peptides that had reported deamidated N or Q residues and that had a 1+ ion with m/z values less than 2000. The

manual procedure described in the Experimental section and outlined in Fig. 3 was used to compile RTS and ΔM values for the deamidation candidates. A majority of the deamidated peaks had an RTS between 2.0 and 5.0 min relative to their corresponding unmodified peptide forms (Fig. 5). As an example, the XICs of βB1 peptide ISLFEGA<u>N</u>FK and its deamidated form from SCX fraction 29 of the water-insoluble 70-year old lens protein are shown in the insert to Fig. 5 where an RTS of 3.1 min was observed.

Parent ion masses of the peptides were manually determined using the procedure in Figure 3 and detailed in the Experimental section. The mass differences (ΔM) between the deamidated peptides and their corresponding unmodified forms were calculated using either manually determined parent ion masses or peptide masses from the corresponding DTA files. The distribution of manually determined delta masses and the distribution of delta masses calculated from MS survey scans (the masses listed in DTA files) are shown in Figure 6A. For deamidated peptides that had an RTS between 2.0 and 5.0 min, the corrected ΔM was between 0.7 and 1.4 Da. However, uncorrected delta masses had a stochastic pattern (Figure 6B) where the expected ΔM of 0.984 was difficult to discern. This implies that peptide masses determined using averaged MS survey scans, even when using lower resolution mass spectrometers like an ion trap, can provide sufficient mass accuracy for deamidation detection.

Another check of the mass correction method was possible from the deamidation identifications that did not have an associated RTS, which were assumed to be incorrect SEQUEST modification identifications since the MS/MS scans were within the

unmodified peptide reference peak. The differences between theoretical unmodified peptide masses and MS-survey-scan-derived parent ion masses are shown as the solid histogram in Figure 7. In theory, the distribution should have been centered on 0.0 Da. However, a prominent peak was observed at 1.0 Da due to MS/MS scans mis-triggered on the first isotopic peaks of the unmodified peptides. Those peptides were more likely mis-interpreted by SEQUEST as deamidations. The distribution for the same incorrect identifications using ΔM values derived from manually determined parent ion masses was centered on 0.0 Da as can be seen from the dotted histogram in Fig. 7.

3.3.3 Deamidation Detection Accuracy: Deamidation detection accuracy rates for all samples used in this study were calculated by assuming that SEQUEST deamidation identifications were incorrect if they did not have an associated RTS between 2 and 5 min and a corrected ΔM shift between 0.7 and 1.4 Da. In synthetic peptide runs, only 68% of the deamidations reported by SEQUEST were correct using these criteria. In the lens samples, peptides were observed in several SCX fractions and in several different peptide forms, contributing to a high degree of MS/MS redundancy. There were 930 MS/MS spectra from the water-insoluble fraction of the 70-year old lens matched to peptides with candidate deamidation sites: 478 spectra of unmodified peptides and 452 spectra of deamidated peptides. For all unmodified and deamidated peptide pairs in all SCX fractions, RTS and ΔM values were manually tabulated. Only 58% of the SEQUEST deamidation identifications were found to be correct using the above criteria. The deamidation detection accuracy rate was 38% for 335 redundant deamidation identifications reported by SEQUEST in the water-soluble fraction of the same lens

sample. In the control 3-day old sample, where the extent of deamidation was expected to be minimal, only 4% of the 179 spectra with reported deamidation sites were correct.

The fragmentation pattern of unmodified peptides and their deamidated forms are frequently very similar and result in nearly identical Xcorr values that can generate unexpectedly small DeltaCN values when differential deamidation searches are specified. Figure 8 shows the distributions of DeltaCN values from SEQUEST searches with and without deamidations specified as differential modifications for the subset of MS/MS scans associated with deamidated peptides that passed our manual deamidation validation criteria. Without specifying deamidation in the search, 98% of the MS/MS results still correctly identified peptide sequences with a DeltaCN value greater than 0.1 (a common threshold value). In contrast, 62% of the same MS/MS spectra had DeltaCN values less than 0.1 (and would be discarded) when deamidation was specified in the searches despite the fact that all of the peptides were actually deamidated.

3.3.4 Validation Criteria for Automated Analysis: In order to speed up the analyses of large data sets and allow analyses of several lenses, a Java program that automates the procedure was written. A first step in automating the validation procedure was to establish empirical ranges for the RTS values and the ΔM values derived from averaging survey scans. Utilizing the observed RTS and ΔM values from the manual analysis for true deamidated peptides present in the 70-year old lens sample, an RTS range of 2.0 to 5.0 min and ΔM range of 0.7 to 1.4 Da were selected as the criteria for correct deamidation identification in the automated analysis. Only 6.5% of the manually validated true deamidation identifications fell outside the selected ranges.

3.3.5 Deamidation in Aged Lenses: A total of 72 deamidation sites in lens crystallin proteins were detected by the automated analysis of the six lenses ranging in age from 3 days old to 93 years old. However, some of identified deamidation sites were located in peptides with multiple potential deamidation sites, each site being identified in at least one MS/MS scan. Due to insufficient evidence to localize the deamidation site to a specific residue, such ambiguous identifications were collapsed into single entries. This resulted in identification of the 57 unique deamidated peptides listed in Table 2. The 16 new deamidated peptides found in this study are highlighted in Table 2 in bold.

MS/MS spectral counting is a semi-quantitative technique that has been used to estimate the abundances of proteins and posttranslational modifications[35, 89]. Figure 9 shows the total MS/MS counts of valid deamidation identifications in lens crystallin proteins in water-soluble and water-insoluble fractions across the series of aged lens samples. Spectral counts were not normalized between lens samples since the same amount of protein was digested from each lens fraction, and the 2D-LC mass spectrometry analyses were done identically.

3.4 DISCUSSION

A robust method for reliable deamidation detection in shotgun proteomic experiments was developed for low mass resolution instruments. The method utilizes the differential reverse-phase chromatographic behavior of deamidated peptides and their corresponding unmodified forms and computes more accurate peptide masses from survey scans averaged during peptide elution. In order to make the method applicable for

high-throughput use, a Java program that automated the manual procedure was also developed. A series of aged human lens samples were analyzed using the automated procedure and several previously unreported deamidation sites were detected.

3.4.1 Chromatographic Behavior of Deamidated Peptides: Synthetic peptide pairs (unmodified peptide and corresponding deamidated form) were used to establish that unmodified and deamidated peptides could be resolved using shallow acetonitrile RP gradients. The synthetic peptide pairs were also used to show that both peptide forms largely co-eluted during common SCX separation conditions. Thus the differential RP elution behavior could be applied to separations of complex peptide mixtures where both SCX and RP are necessary since the unmodified peptides and their deamidated forms would both be present in the same RP runs.

Complex digests from a series of human lenses were separated by 2-DLC using shallow reverse-phase gradients, and the MS/MS spectra were searched with SEQUEST to identify candidate unmodified and deamidated peptides. In a subset of the data, the identifications were manually assigned as correct or incorrect using retention time information and improved mass measurements obtained by averaging survey-scan spectra. The results in Fig. 5, where the peak in the RTS distribution was 3 minutes, confirmed that in vivo deamidated peptides exhibited similar RP elution time shifts to those of the synthetic peptides summarized in Table 1.

3.4.2 Isomerization during Deamidation: A small fraction of the lens peptides had RTS values of less than 2 minutes but still had a corrected ΔM between 0.7 and 1.4

Da, indicating deamidation. These decreased RTS values could be explained by formation of isoaspartate during deamidation[8, 69, 106]. Previous studies have reported that isoaspartate containing peptides may elute earlier in reverse-phase separations than corresponding aspartate containing peptides[86, 107], and this could have resulted in the reduced RTS values. In vitro studies also suggest that isomerized forms of deamidated peptides containing isoaspartate may be more abundant than forms containing normal aspartate[61, 108]. These changes could potentially impact the automated algorithm for deamidation detection. However, in the lens samples, multiple deamidated peaks of asparagine-containing peptides were relatively rare, as can be seen in the inset to Fig. 5. Since there will always be a non-isomerized deamidated peak when asparagine deamidation occurs, as long as an MS/MS scan is acquired within the non-isomerized peak, the peptide will still likely be correctly identified as deamidated. Figure 10 shows a hypothetical aspargine deamidation with different isomerized deamidated peptide elution times[86]. The site is correctly identified as deamidated in all three of the depicted cases, provided that MS/MS scans were acquired during the elution peaks. Furthermore, a majority of the lens protein deamidation sites discovered in this study are located at glutamine residues, which do not readily undergo isomerization during deamidation.

3.4.3 Deamidation Detection from MS/MS Spectra: Deamidation detection accuracy rates with SEQUEST were only 68% even for the very favorable situation of synthetic peptide pairs. In the lens samples that were manually analyzed, the SEQUEST deamidation detection accuracy rate decreased dramatically as the abundance of true deamidations in the sample decreased. This was probably because a larger percentage of

the SEQUEST identified deamidations were the result of MS/MS scans mis-triggered on isotopic peaks of more highly abundant unmodified peptides.

Deamidation also illustrates the potential loss of sensitivity when delta score thresholds are used in modification searches. For the deamidated peptides in Fig. 8, only 38% would pass a DeltaCN threshold of 0.1 if deamidation was specified as a variable modification. Thus DeltaCN, normally a strong predictor of correct SEQUEST results, can seriously compromise modification search results. We emphasize that Xcorr and DeltaCn are still quite effective in determining correct amino acid sequences of peptides, but function far less effectively for reliable deamidation detection.

High mass accuracy instruments may also have difficulty detecting deamidation from MS/MS spectra if unmodified and deamidated peptides are not chromatographically separated. This is especially true due to the common use of nanospray chromatography with shorter lower resolution columns. Parent and fragment ion tolerances specified in database searches are often much smaller for instruments with TOF or FT analyzers compared to ion traps, and mixtures of ions differing by 1 Da may complicate analysis using search algorithms. Higher resolution chromatography and incorporation of peptide retention time shifts, as used in this study, can be used to improve deamidation measurements with any mass spectrometer.

3.4.4 Automated Algorithm Performance: The specificities and sensitivities of the automated program were calculated by comparison of the manual results for the same lens data to asses the performance of the implementation. Any deamidation

identifications that passed the above-mentioned criteria by the automated analysis but not by manual validation were considered incorrect identifications (false positives, FP), and manually validated correct identifications that failed the automated analysis were considered as false negatives (FN). Deamidation identifications that passed both manual and automated validation were considered as true positives (TP), and identifications that failed both were classified as true negatives (TN). The specificities [TN/(FP+TN)] of the automated deamidation validation were 93%, 95%, and 99% in the 70-year water-insoluble fraction, the 70-year water-soluble fraction, and the soluble 3-day old lens fraction, respectively. The sensitivities [TP/(TP+FN)] of automated deamidation validation were 95%, 91% and 75% in 70-year old water-insoluble fraction, 70-year old water-soluble fraction and soluble 3-day old lens fraction, respectively. The relatively lower sensitivity for the 3-day old sample was due to lower sample complexity, which results in more MS/MS scans triggered outside of the chromatographic peak regions. The high specificity and sensitivity of the automated procedure demonstrates that the method was reliable and could be used as a post-processing tool to identify true deamidated peptides.

3.4.5 Deamidation in Human Lens: In this study, 72 deamidation sites could be identified in a series of lenses of increasing age; 54 had been previously reported and 18 were identified here for the first time (Bolded sites in Table 2). There have been a total of 81 deamidation sites reported in human lens to date (see Ref. 22 and Supplemental materials therein), and this work was able to confirm 67% of all known deamidation sites. Many of the reported sites that were not observed occurred in tryptic peptides that had a 1+ ion mass greater than 2000 Da and would have been excluded from our analysis.

Quantification of deamidation is of far greater significance than identifying deamidation sites, and we were able to use the RTS and ΔM constraints to more accurately tally the MS/MS counts associated with unmodified and deamidated peptide forms. The spectral counts were used to provide estimates of the relative abundance of each peptide form and are shown in Fig. 9. The deamidated peptide MS/MS counts in the water-insoluble fractions of lens proteins showed a pronounced increase with age. The spectral counts of deamidated peptides in the water-soluble fractions also increased with age, but remained relatively constant after middle age. Deamidation has emerged as the major lens post-translational modification associated with crystallin insolubilization, and additional ion current integration are underway to understand the progression of deamidation with age and condition of the lens.

4. RELATIVE DEAMIDATION QUANTIFICATION BY PEPTIDE MIXTURE MODELING OF ISOTOPIC ENVELOPE INTENSITY DISTRIBUTIONS.

4.1 INTRODUCTION

A robust computational relative deamidation quantification method was developed that modeled experimental peptide isotopic distribution intensity as a mixture of predicted peptide isotopic envelope models (IEM). Changes in extent of deamidation as a function of lens age or solubility were needed to study the role deamidation may play in age-related nuclear cataract. The overlap in mass of deamidated peptides with the isotopic peaks of the unmodified peptide makes conventional mass extracted ion chromatograms (the usual mass spectrometry quantification method) problematic unless peptide forms are completely chromatographically resolved. Poor chromatography and isomeric deamidation products (for asparagine) with shifted retention times occur frequently enough that chromatographic separation can not always be accomplished when complex mixtures are separated. Therefore, relative ion currents were extracted from parent ion mass measurements averaged over ranges of retention times rather than from chromatogram ion currents averaged over mass ranges. Equivalence of the two methods was demonstrated, and the new method was extremely robust with respect to poor chromatography peak shapes and presence of interfering peptides.

A series of known mixtures of synthetic unmodified peptides and their corresponding deamidated forms (1:0, 1:2, 2:1, 4:1, 10:1 and 20:1) were analyzed with LC/MS and experimental deamidation ratios were determined using the IEM method. A high correlation (R^2=0.96) observed between experimental and known ratios proves the

accuracy of IEM method. The IEM method was programmed into a semi-automated GUI (Mass Ion Chromatogram Integration Tool, a.k.a MICIT) and a workflow for efficient quantification of peptides in large-scale proteomics datasets was developed. An aged series of normal (3-day, 2-year, 22-year, 35-year, and 70-year) and cataractous (93-year) human lenses were analyzed using two dimensional liquid chromatography tandem mass spectrometry and a total of 40 novel deamidation sites were detected. As a proof of principle, MICIT was used to determine deamidation quantities of three γS peptides ([N14-NQ16], N53, and [Q63-Q70]) in an aged series of normal lenses (3-day, 2-year, 35-year, and 70-year). N53 and [Q63-Q70] peptides showed age-related accumulation of deamidation in the water insoluble portion of lens samples. We have also measured the deamidation rate of N143 in γS in aged normal (70-year) and cataractous (93-year) lens samples. Age-related accumulation of deamidation at N143 was observed, which is linked to severe loss of γS crystallin structure[42, 45].

4.2 EXPERIMENTAL

4.2.1 Sample Processing: Synthetic peptides corresponding to human βB1 crystallin (SwissProt P05813) residues 150-159 (ISLFEGANFK), residues 202-213 (GYQYLLEPGDFR), and 214-229 (HWNEWGAFQPQMQSLR), and their deamidated forms (ISLFEGADFK, GYEYLLEPGDFR, HWDEWGAFQPQMQSLR, HWNEWGAFEPQMQSLR, and HWNEWGAFQPQMESLR, respectively) were obtained from Sigma Genosys (The Woodlands, Texas) and diluted to 10 pmol/μl in 5% formic acid. Ten picomoles of each peptide were analyzed by RP LC-MS, either alone or by simultaneously injecting equal mixtures of unmodified and deamidated forms of the same peptide. Unmodified and their corresponding deamidated forms of synthetic

peptides were also mixed in various ratios (1:0, 1:2, 2:1, 4:1, 10:1 and 20:1) and 60 picomoles of each mixture was analyzed by RP LC-MS.

Lenses from 3-day, 2-year, 22-year, 35-year, 70-year old human donors with healthy vision and a 93-year old human donor with type-III nuclear cataracts (Pirie scale) were obtained from the Lion Eye Bank of Oregon with Oregon Health & Science University IRB approval. To asses the accuracy of deamidation quantification, we analyzed the 3-day old lens (where little or no deamidation is expected) under identical conditions to more extensively deamidated aged lens samples[35]. Lens proteins were separated into water-soluble and water-insoluble fractions by centrifugation. The amount of water-insoluble portion of the 3-day old lens sample was negligible and not analyzed. Following BCA protein assay (Pierce Biotechnology, Inc. Rockvillie, IL), 2.5 mg portions of proteins were reduced and alkylated in the presence of urea, digested overnight with trypsin, digested peptides were solid phase extracted, and peptides were separated by SCX chromatography as described previously[35].

4.2.2 LC-MS/MS Analysis: Synthetic peptides, peptide mixtures, and 10% portions of SCX fractions of lens digests were similarly analyzed by LC-MS using a Waters capillary LC system and a Q-TOF2 mass spectrometer (Waters, Milford, MA). Samples were applied at 15 μl/min to a Nanoease C18 trap cartridge (Waters, Milford, MA), and then switched onto a 0.15 x 150 mm Nanoease Symmetry C18 column (Waters, Milford, MA) using a mobile phase containing 0.1% formic acid. The gradient used 9.4-37.5% acetonitrile over 95 min at a 0.4 μl/min flow rate. One second survey and MS/MS scans were collected in profile mode. Data dependent MS/MS collection used

charge state dependent dynamic exclusion feature (exclusion mass window width 3.0 Da, repeat count of 1, exclusion list size 80 ions, and exclusion duration of 10 sec) to obtain MS/MS spectral of the five most abundant parent ions following each survey scan. Since the mass shift associated with deamidation was small, repeat counts greater than 1 or exclusion duration greater than RTS of the unmodified and deamidated forms of a peptide would have prevented MS/MS scans of deamidated forms from occurring. DTA files were generated using ProteinLynx software (version 2.1, Waters, Milford, MA) with a *m/z* range of 50 to 2000 Daltons. Profile peaks in MS/MS scans were smoothed with a savitzky-golay filter (2 iterations with 3 channels) and centroided to generate DTA files. A total of 133,267 tandem mass spectra were acquired from all LC-MS/MS experiments.

4.2.3 Peptide and PTM Identification: Peptides were identified by searching the tandem mass spectra against a human lens database described in chapter 3.1 using SEQUEST[18]. Searches were configured to use monoisotopic masses, the parent ion mass tolerance was 0.5 Da, a static mass of 57.02 was specified for cysteine residues due to alkylation, and no enzyme cleavage was specified. SEQUEST was also configured to look for the following variable modifications in peptides: deamidation (+0.984 Da on N or Q); Oxidation (+15.995 on M or W); and acetylation (+42.011 Da) on the N-terminus of peptides. A list of peptides, proteins, and potential posttranslational modifications was generated using DTASelect[102] configured to use Xcorr thresholds of 1.8, 2.5, 3.5, and 3.5 for +1, +2, +3, and +4 parent ions, respectively, to use a minimum deltaCN of 0.0 instead of default 0.08, and to select full or half-tryptic peptide termini. All crystallin peptides identified by SEQUEST containing modified amides were considered for deamidation quantification analysis.

4.2.4 Deamidation Quantification with Isotopic Envelope Modeling: An alternative quantification method was developed that modeled the parent ion isotopic envelope as a mixture of the unmodified and its corresponding deamidated peptide forms using a Gaussian instrument response function. In this model, the total calculated ion current intensity Y at m/z value x is given by:

$$(1) \quad Y(x) = \sum_{j=1}^{p} \sum_{i=1}^{7} I_j \frac{f_{ij}}{\sigma_{ij}\sqrt{2\pi}} e^{\left(-\frac{(x-\overline{x_{ij}})^2}{2\sigma_{ij}^2}\right)} + B(x)$$

where I_j are intensity factors proportional to the total ion current of peptide j, p is the total number of peptide forms included in the model, f_{ij} are the predicted fractional relative intensities of the i isotopic peaks of peptide j, $\overline{X_{ij}}$ are the positions of the ith isotopic peak of peptide j, and σ_{ij} are the widths of the ith isotopic peak of peptide j, $B(x)$ is the background intensity contribution from the mass spectrometer which is modeled as either a quadratic or cubic polynomial.

To compensate for the small difference between measured and theoretical masses, the peak position parameter $\overline{X_{ij}}$ is given by:

$$(2) \quad \overline{x_{ij}} = M_{ij} + \sum_{k=0}^{1,2,or3} C_k \left(M_{ij} - x_o\right)^k$$

where C_k are polynomial m/z calibration correction coefficients, M_{ij} are the charge-to-mass values of the ith isotopic peak of peptide j, and x_o is the beginning m/z value of

the fitting interval. The charge-to-mass values of the ith isotopic peak of peptide j are calculated according to Eqn. 3 from the molecular weights, MW_{ij}, and peptide charge z.

$$(3) \quad M_{ij} = \left(\frac{MW_{ij} + z \times 1.007825}{z} \right)$$

The peak width parameter σ_{ij} is given by:

$$(4) \quad \sigma_{ij} = \sum_{k=0}^{k=0 or 1} S_k \left(M_{ij} - x_o \right)^k$$

where S_k are polynomial coefficients to calculate σ_{ij} at position M_{ij}, and x_o is defined as in Eqn. 2. The order of $\overline{X_{ij}}$ and σ_{ij} functions are dependent on the choice of mass spectrometer. Higher order polynomial forms of $\overline{X_{ij}}$ and σ_{ij} are often required for low-resolution instruments, such as ion traps, where distortions from space charging are more common. Simpler, lower order polynomial functional forms could be used with high resolution Q-TOF instruments to reduce the number of degrees of freedom in the fitting process. The intensities of unmodified and deamidated peptides are varied until the best agreement with the total ion current associated with the measured isotopic envelope is obtained.

4.2.5 Semi-Automated Deamidation Quantification Analysis: A semi-automated algorithm for large-scale deamidation quantification in complex samples was also developed. The algorithm accepts native instrument files from Q-TOF2 (Waters, Milford, MA) instrument for raw data and OpenSea[21] formatted XML files for MS/MS results. Scripts were written to convert SEQUEST output to OpenSea XML files.

Peptide identifications containing deamidation modification sites and their corresponding chromatography coordinates (SCX fractions and RP retention times) were extracted from the MS/MS results. Peptide identifications are grouped according to their corresponding protein sequence, unmodified peptide sequence, observed SCX fractions, and charge states. The flowchart in Figure 11 illustrates the core automation procedure employed to quantify the unmodified and various deamidated forms of a peptide using the IEM method. A mass ion chromatogram integration tool (MICIT) was developed by implementing the procedure shown in Figure 11 inside a graphical user interface (GUI). The MICIT software calculates peptide masses used to generate compound ion chromatograms, default retention time ranges for generating averaged mass spectra, and automatically generates initial IEM guesses. The user can preview and modify any of these parameters, add additional modified peptide forms to the IEM, and add an interfering peptide isotopic distribution to the IEM. The MICIT also has several information rich display features: multiple chromatograms of different peptide charge states are displayed in one panel, the retention time of peptide MS/MS identifications are highlighted on the corresponding ion chromatograms, the experimental mass chromatograms spectra and fitted IEMs are plotted together in a single frame, and other peptide identifications present in a selected retention time range can be displayed. The MICIT was written in a combination of java and c++ programming languages. All function minimizations were performed using the minimization toolkit Minuit (version 1.7.6, CERN, Geneva).

4.2.5 MICIT Workflow: A typical workflow for the MICIT is illustrated in appendix 1 in a story board format. The user starts the tool by loading SEQUEST results

and raw data from a sample. MICIT reads and displays the protein identifications present in the sample (appendix 1.1). The user selects a protein of interest and expands the node to display its corresponding peptide identifications (appendix 1.2). The peptide identifications are annotated with their corresponding sequence, plus one mass (MH+), and global retention time bounds (determined from the maximum and minimum retention times of its MS2 identifications). The user selects a peptide of interest to display all SCX fractions in which one of its corresponding forms has been detected (appendix 1.3). Selecting a SCX fraction results in generation of compound ion chromatograms (CICs) using all detected forms of a peptide (unmodified and deamidated). Since multiple charge states of a peptide could be detected in a single SCX fraction, separate CICs are generated for all detected charge states and displayed in a single panel (appendix 1.3). The user proceeds to select retention time integration bounds for the peptide using its elution profile shown by its CICs (appendix 1.4). After this step, the user expands the SCX fraction and selects a detected charge state to generate a corresponding averaged experimental mass ion chromatogram (MIC) for the peptide (appendix 1.4). The generated MIC is displayed in a panel below the CICs (appendix 1.5). After adjustment of the number of peptide forms or addition of interfering peptides, MICIT generates initial guesses for the theoretical isotopic envelope model (IEM) that matches the MIC (appendix 1.6). In the next step, IEM parameters are optimized to fit the MIC (appendix 1.7). The optimized IEM is displayed on top of MIC to show the goodness of the fit and its corresponding model parameters are also displayed (appendix 1.8). After this step, the user has a choice to save the results of the integration before proceeding to the next available charge state, SCX fraction, peptide, or protein. The results are saved in an XML formatted document for post processing and results viewing.

4.3 RESULTS

4.3.1 Synthetic Peptides: Shallow reverse phase (RP) acetonitrile gradients described in chapter 3.1 (for capillary flows) were employed to resolve the unmodified and deamidated synthetic peptide forms by RP when using nano flows. Figure 12 shows the XICs during RP chromatography of two synthetic peptide pairs. An asparagine/aspartic acid pair (HWNEWGAFEPQMQSR/HWDEWGAFEPQMQSR) is shown in Figure 12a and a glutamine/glutamic acid pair (GYQYLLEPGDFR/GYEYLLEPGDFR) in Figure 12b. Deamidated peptide forms eluted approximately 3.0 mins later than their corresponding unmodified forms.

To test the accuracy of the IEM quantification method, unmodified and deamidated forms of an asparagine/aspartic acid pair (ISLFEGANFK/ISLFEGADFK) and glutamine/glutamic acid pair (GYQYLLEPGDFR/GYEYLLEPGDFR) were mixed in various amide:acid ratios (1:0, 1:2, 2:1, 4:1, 10:1 and 20:1), analyzed by LC-MS using shallow RP acetonitrile gradients, and experimentally measured ratios were calculated using the IEM method. A scatter plot between the experimental and known ratios is shown in Figure 13a. A high correlation (R^2=0.96) between the two quantities was observed both at high and low deamidation abundances. Extracted ion chromatogram (XIC) integration is a traditional peptide quantification technique, and synthetic peptide mixture ratios were also calculated using the XIC method with Masslynx (Waters, Milford, MA) software as a control. A very high correlation (R^2=0.98) between the ratios determined by XIC and IEM methods was observed as shown in Figure 13b demonstrating the compatibility of the two methods.

4.3.2 Complex Mixtures: To demonstrate the accuracy of the IEM technique in complex sample analysis, a single asparagine containing peptide WNTWSSSYR (βB1, 124-132) was chosen from a representative SCX fraction of a 93-year old soluble lens digest. A compound XIC was obtained for both unmodified and deamidated forms of the peptide (Fig. 14a) and areas under the peaks were integrated using Masslynx software (Waters, Milford, MA). A combined mass chromatogram spectrum was obtained for both peptide forms by averaging MS scans acquired during the peptide elution (determined from the chromatogram peaks). The resulting isotopic envelope mixture was modeled using the predicted unmodified peptide isotopic distribution, the deamidated isotopic distribution, and a cubic polynomial background (Fig. 14b.). The total ion current for this peptide was computed by summing intensities of all peptide forms. Percent deamidation, for both IEM and XIC integration methods, was computed by normalizing the obtained ion currents of each peptide form by the total ion current. The percent deamidation obtained by XIC integration (60%) agreed well with the value computed by the IEM method (58%).

Co-eluting peptide forms complicate ion chromatograms in complex samples and make accurate quantification difficult. A compound ion chromatogram for various peptide forms (unmodified, singly deamidated, doubly deamidated) of ITIYDQENFQFK (αA1, 33-44) from a representative SCX fraction of 93-year old lens insoluble digest is shown in Fig. 15a. The averaged experimental chromatogram suggested the presence of multiple peptide forms. Since the chromatogram peaks were not baseline resolved, traditional quantification with XIC integration would be error-prone. However, the IEM

41

method handles such situations by performing the integration in m/z space. The MS scans during 34.0 to 41.0 min. were averaged to obtain an experimental isotopic envelope for the peptide forms (Fig. 15b). The experimental isotopic envelope was modeled using predicted isotopic envelopes of unmodified, singly deamidated, doubly deamidated peptide forms, a cubic polynomial background, and a mixture of all the afore-mentioned components. Individual component fits are shown in Figure 15b and composite fit is shown in Figure 15c. Singly and doubly deamidated peptide ion currents were normalized using the total ion current explained by all the peptide forms and shown in Figure 15b.

4.3.3 Spectral Counting Accuracy: Eight single-amide containing peptides from 3-day old soluble, 93-year old soluble, and 93-year old insoluble lens samples were chosen for comparison. The amounts of the unmodified and deamidated forms of the peptide in each SCX fraction of the sample were determined by integrating the corresponding peptide extracted ion currents (XIC) and by summing the corresponding MS/MS spectral identifications (spectral counting). The amounts of each peptide form were summed across multiple SCX fractions to determine the total amount of unmodified and deamidated forms present in the sample. Figure 16 shows the correlation between deamidation ratios determined by XIC integration and by spectral counting. A good correlation ($R^2=0.92$) between the two methods indicates that spectral counting can be an alternative semi-quantitative method for the study of post-translational modification abundance changes.

4.3.4 Deamidation in Aged Lenses: A total of 124 deamidation sites in lens crytallin proteins were detected by the LC-MS/MS analysis of six lenses ranging from 3 days old to 93 years old. Deamidation sites found in this study were cross-referenced with existing literature[13, 34, 35, 42, 44, 45, 50, 51, 55, 109-117] and the previous low-resolution study (chapter 3). The Q-TOF data resulted in identification of an additional 40 novel deamidation sites, which are highlighted in Table 3 in bold.

4.3.5 Determining γS Crystallin Peptide Deamidation Rates Using MICIT: Deamidation and corresponding unmodified peptide identifications of γS crystallin from the lens sample were loaded into MICIT and analyzed. All detected SCX fractions and charge states of an unmodified peptide and its corresponding deamidated forms (multiple forms, if present) were determined and compound ion chromatograms are generated accordingly. Peptide retention time integration boundaries for each SCX fraction were determined using a 2.5% baseline threshold in the corresponding compound ion chromatograms. An averaged mass chromatogram was generated for each of the detected charge states and SCX fractions of the peptide. Peptide mass chromatograms with poor signal-to-noise ratio (less than 2.5% of baseline intensity in corresponding averaged mass chromatograms) were not considered for integrations. Unmodified and deamidated ion currents of peptides present in an averaged mass chromatogram were integrated using the IEM method. Peptide ion currents from all SCX fractions were grouped and summed according to peptide form (unmodified, single deamidated, etc.) and detected charge state. Relative percent deamidation rates were calculated for all peptide forms (unmodified and deamidated forms) for the same charge state. The overall percent

deamidation rate for a site (or sites: in case of multiple deamidations) was computed by averaging percent deamidations over all observed peptide charge states.

As a proof of concept, percent deamidation rates of the following deamidation sites in γS crystallin in both water soluble and water insoluble portions of 3-day, 2-year, 35-year, and 70-year normal lenses: [N14-Q16] (ITFYEDK<u>N</u>F<u>Q</u>GR), [Q63-Q70] (MYILP<u>Q</u>GEYPEY<u>Q</u>R) and N53 (VEGGTWAVYERP<u>N</u>FAGY) were computed using the above-mentioned protocol. Both single and double deamidation rates were computed for peptides with two potential deamidation sites. Computed single deamidation rates of these γS crystallin sites in water soluble and water insoluble portions of a series of aged lens samples are shown in Figures 17a and 17b, respectively. The double deamidation rates for [N14-Q16] and [Q63-Q70] were negligible (maximum double deamidation rate < 3.5%) and not shown.

Asparagine[143] (N143) of γS crystallin is known to be deamidation resistant due to its location in a highly ordered β-hairpin loop[118]. Some earlier studies have also detected deamidation of N143 in cataractous human lens samples[42, 45]. No appreciable deamidation was detected at N143 in 3-day, 2-year, and 35-year old samples. However, percent deamidation rates of N143 in water soluble and water insoluble portions of 70-year normal and 93-year old cataractous human lens samples were computed using the above-mentioned protocol and shown in Figure 18.

4.4 DISCUSSION

  A robust computational method for reliable quantification of deamidation in shotgun proteomic experiments was developed. The method quantifies deamidation by comparing experimental isotopic ion currents to theoretical ion currents constructed from Gaussian-folded predicted isotopic envelopes of the constituent unmodified and deamidated peptides. The accuracy of the method was tested using known mixtures of unmodified and deamidated synthetic peptides. In order to make the method applicable for high-throughput use, a semi-automated graphical user interface (MICIT) was developed. A series of aged human lens samples were analyzed using a high-resolution Q-TOF mass spectrometer. As a proof of principle, age-related deamidation rates of several sites in γS lens crystallin protein were determined using the MICIT.

  4.4.1 Accuracy of IEM Method and Utility of Automation: The isotopic envelope modeling (IEM) was tested using a wide dynamic range of known unmodified and deamidated synthetic peptide mixtures. The IEM method accurately determined both high and low deamidation ratios. An equivalence between traditional XIC integration and IEM methods was demonstrated. Traditional XIC integration methods were not as sensitive in measuring low abundance deamidation ratios as the IEM method which was highly accurate in determining peptide deamidation ratios in both simple (synthetic peptide mixtures) and complex mixtures (human lens digests).

  The mass ion chromatogram integration tool (MICIT) is a semi-automated GUI that performs deamidation quantification using the IEM method. The tool allows the user to interactively drive the quantification process using any search engine results. The tool consolidates disparate information sources such as protein identifications, peptide

identifications, their corresponding MS scans, and multi-dimensional chromatographic coordinates (SCX fractions and RP retention times) in a single user interface. Contemporary multi-dimensional liquid chromatography tandem mass spectrometry experiments often contain large numbers of protein and peptide identifications present in multiple fractions. Peptide quantification in such complex datasets would become cumbersome without a tool like MICIT. The tool could be easily altered for use with other types of PTM (phosphorylation, acetylation, methylation etc.) quantification besides deamidation.

4.4.2 Advantages of IEM over XIC integration: The XIC integration technique is not reliable in quantifying deamidations present in low stoichiometric levels due to the difficulties in detecting ion chromatogram peaks of peptides (see lower quadrant of scatter plot in Fig. 13b). Co-eluting peptides in complex mixtures often complicate the chromatograms making it difficult to determine the peak retention time boundaries of each peptide (Fig 15a). Hence, it is difficult and error-prone to use traditional XIC integration technique to quantify peptides present in such complex mixtures. The IEM method bypasses such complications by inverting the problem by performing peptide ion current integrations from space in m/z space rather than retention time.

4.4.3 Alternative Quantification Techniques: Peptide PTM abundances derived from spectral counting and XIC integration methods showed a high correlation. The reproducibility of the spectral counting technique relies on total number of redundant MS2 spectra matched to a peptide of interest. For example, a total number of 17 redundant MS2 spectra (on average) were matched to each of the peptides used to derive

the correlation plot between spectral counting and the XIC integration technique. In general, high numbers of MS2 spectra cannot be matched to low abundant peptides present in complex samples, which limits the utility of spectral counting. Hence, spectral counting may not be appropriate for accurately quantifying PTM abundances in many peptides. We want to acknowledge the fact that spectral counting is still a reliable semi-quantitative technique to measure overall PTM abundances in proteins due to high redundancy of peptides matching to a protein.

4.4.4 Deamidation in γS Human Lens Crystallin: γS crystallin is an abundant, highly structured, highly soluble, and stable lens protein. Structural changes in γS has been correlated with its insolubility[42, 45]. γS protein consists of two N-terminal domains and two C-terminal domains connected with a peptide linker. The water soluble and water insoluble deamidation rates of three N-terminal domain peptides ([N14-Q16], [Q63-Q70], Q53) and one C-terminal domain peptide (N143) from γS were computed for a series of human lens samples with increasing age.

Water soluble and water insoluble portions of lens tissue did not show significant differences in age-related accumulation of deamidation at [N14-Q16] (Fig. 17.). [N14-N16] peptide, being located close to the N-terminus of γS, could be solvent accessible from an early age. Hence, age-related accumulation of deamidation at this site may have little biological significance. Peptides [Q63-Q70] and Q53 of γS (situated in the 2nd N-terminal domain) showed a higher rate of age-related accumulation of deamidation in the water insoluble fraction, than the corresponding water soluble fraction, of human lens (Fig. 17.). Hence, age-related accumulation of deamidation in these two peptides is linked

47

to loss of γS structure and solubility. It is interesting to note that these peptides did not show accumulation of deamidaiton at an early age (like [N14-Q16]). This suggests that they are located in hydrophobic core of γS, which becomes more solvent exposed due to gradual changes in protein conformation. However, the cause of such structural losses is unknown. It is possible that age-related oxidative stress or deamidation via succinimide formation could initiate the structural damage, which may accelerate deamidation in other parts of the protein. However, further structural and *in vitro* studies must be performed to test the above hypothesis.

N143 of γS (situated in first C-terminal domain) is thought to be highly deamidation resistant. Previous structural studies have localized the site to a highly ordered and folded β-hairpin loop, which plays a key role in maintaining the tertiary βγ fold[118]. Deamidation at this site has been linked to destabilization of the γS domain, leading to loss of protein solubility and changes in structure[42, 45]. Supporting the above hypotheses, we have detected high N143 deamidation abundances in water insoluble portions of aged normal (70-year) and cataractous (93-year) human lens samples. No other lens samples analyzed in this study showed deamidation at N143. The extent of N143 deamidation observed in water insoluble and water soluble portions of cataractous (93-year) human lens sample agrees well with previous observations[44, 45]. Also consistent with previous observations[45], the water insoluble portion of cataractous human lens (93-year) showed a higher percentage of N143 deamidation than normal lens (70-year). Additional cataractous lenses need to be analyzed to see if γS N143 deamidation is indeed associated with catractogenesis.

5. CONCLUSIONS, FUTURE DIRECTIONS & GLOBAL PRESPECTIVES

5.1 CONCLUSIONS

Traditional search programs have very high false discovery rates (30%) to be used to identify deamidations when using low resolution mass spectrometry. A novel bioinformatics approach was developed that incorporates MS and LC information to control deamidation FDRs from search engines enabling them to be used for deamidation detection. The method was automated for high-throughput validation of deamidated peptides present in large-scale proteomics datasets. Large scale processing of human lenses allowed age-dependent deamidation changes to be globally measured for the first time.

Quantification of deamidation using existing software is infeasible due to mixed LC and MS signals. Novel deamidation quantification software (MICIT) that can accurately deconvolute mixed MS signals to quantify site-specific deamidations present in large-scale proteomics datasets was developed. Large-scale processing of human lenses using the tool will allow age-dependent site-specific deamidation changes to be measured for the first time.

5.2 FUTURE DIRECTIONS

All post-translational modifications change the amino acid composition of the peptide sequence causing the modified peptide behave differently during cation exchange (SCX) and/or reverse-phase (RP) chromatography. Such chromatography shift information can be used to validate other types of PTMs present in complex samples. IEM method can also be extended to quantify other PTMs like acetylation and

phosphorylation, which have significance in other biological processes like regulation and signaling.

Although the semi-automated GUI (MICIT) has been developed for quantification of deamidated peptides, it was programmed following the reusability principles of software engineering. Hence, MICIT could be adapted to perform quantification of metabolically labeled, chemically labeled, enzymatically labeled (with mass shifted isotopic distributions), or label-free peptides (with mass shifted isotopic distributions). The tool could process data from other mass spectrometers using java-based plug-ins. More extensive data analysis (statistical) and reporting functions could be added. The MICIT workflow could also be more automated by implementing a template based (peptide, chromatography) coordinate system. The system could calculate retention time integration windows for a peptide based on training data. The initial IEM parameters could also be based on training data for the peptides in consideration. New experiments would be analyzed using the predetermined template (from prior analysis) to perform integrations without user intervention.

Complex bioinformatics tools (like MICIT) are necessary for large-scale analysis of proteomics experiments designed to answer complicated biological questions. We plan to use the tool to determine deamidation rates at several lens crystallin sites in a series of aged normal and cataractous human lens samples. We also plan to generate a detailed map for all lens crystallins that illustrates potential deamidation hotspots. Development of such a map would be invaluable in understanding the changes that occur in aged lens that ultimately lead to age related nuclear cataracts.

5.3 GLOBAL PRESPECTIVES

Posttranslational modifications (PTMs) of proteins are known to play a role in many patho-physiological processes. For example, accumulation of deamidation in tissues is linked to aging[9], O-glycosylation of intracellular proteins is linked to type II diabetes[10], and citrullination of arginine is linked to rheumatoid arthritis[11] and multiple sclerosis[12]. This highlights the need for development of PTM characterization tools to monitor disease progress. A generalized mass spectrometry-based framework is proposed that will allow characterization of PTMs that play a role in disease (Figure 19). According to the workflow, control and disease samples are subjected to individual proteomics mass spectrometric experiments. Proteins, peptides and PTMs present in both samples are identified using search engines like SEQUEST or InsPect. PTMs identified in both samples are first validated and then quantified using the bioinformatics concepts developed in chapter #3 and #4 of this thesis. PTMs that are differentially expressed between control and disease samples are determined from the quantification results. Differentially expressed PTMs can be used to derive biological knowledge pertaining to the disease. This information can be used to develop appropriate diagnostic, preventive or therapeutic strategies for the disease. This speaks for the importance of bioinformatics tools and workflows in translation of molecular level knowledge into the clinical and diagnostic realm, further highlighting the integration of informatics into the translational life cycle.

6. BIBLIOGRAPHY

1.   Stein, L. D., Human genome: end of the beginning. *Nature* **2004,** 431, (7011), 915-6.
2.   Kanehisa, M., *Post-genome informatics*. Oxford University Press: Oxford New York, 2000; p ix, 148 p.
3.   Liebler, D. C., *Introduction to proteomics : tools for the new biology*. Humana Press: Totowa, NJ, 2002; p ix, 198 p.
4.   Aebersold, R.; Goodlett, D. R., Mass spectrometry in proteomics. *Chem Rev* **2001,** 101, (2), 269-95.
5.   Wilkins, M. R., *Proteome research : new frontiers in functional genomics*. Springer-Verlag: New York, 1997; p xviii, 243 p.
6.   Hunter, T., The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell* **2007,** 28, (5), 730-8.
7.   Smith, C. L., A shifting paradigm: histone deacetylases and transcriptional activation. *Bioessays* **2008,** 30, (1), 15-24.
8.   Robinson., N. E.; Robinson., A. B., *Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins*. Althouse Press: 2004.
9.   McKerrow, J. H., Non-enzymatic, post-translational, amino acid modifications in ageing. A brief review. *Mech Ageing Dev* **1979,** 10, (6), 371-7.
10.   Lim, J. M.; Sherling, D.; Teo, C. F.; Hausman, D. B.; Lin, D.; Wells, L., Defining the Regulated Secreted Proteome of Rodent Adipocytes upon the Induction of Insulin Resistance. *J Proteome Res* **2008,** 7, (3), 1251-1263.
11.   Yamada, R.; Suzuki, A.; Chang, X.; Yamamoto, K., Citrullinated proteins in rheumatoid arthritis. *Front Biosci* **2005,** 10, 54-64.
12.   Harauz, G.; Musse, A. A., A tale of two citrullines--structural and functional aspects of myelin basic protein deimination in health and disease. *Neurochem Res* **2007,** 32, (2), 137-58.
13.   Zhang, Z.; Smith, D. L.; Smith, J. B., Human beta-crystallins modified by backbone cleavage, deamidation and oxidation are prone to associate. *Exp Eye Res* **2003,** 77, (3), 259-72.
14.   Slingsby, C.; Clout, N. J., Structure of the crystallins. *Eye* **1999,** 13 (Pt 3b), 395-402.
15.   McGinty, S. J.; Truscott, R. J., Presbyopia: the first stage of nuclear cataract? *Ophthalmic Res* **2006,** 38, (3), 137-48.
16.   Criag R, B. R., TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004,** 20, (9), 908-919.
17.   Han, Y.; Ma, B.; Zhang, K., SPIDER: software for protein identification from sequence tags with de novo sequencing error. *Proc IEEE Comput Syst Bioinform Conf* **2004**, 206-15.
18.   Eng, J. K.; McCormack, A. L.; Yates, J. R., 3rd, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in protein database. *J Am. Soc. Mass Spectrom* **1994**, (5), 976-989.
19.   Tabb, D. L.; Saraf, A.; Yates, J. R., 3rd, GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* **2003,** 75, (23), 6415-21.

20.   Hansen, B. T.; Jones, J. A.; Mason, D. E.; Liebler, D. C., SALSA: a pattern recognition algorithm to detect electrophile-adducted peptides by automated evaluation of CID spectra in LC-MS-MS analyses. *Anal Chem* **2001,** 73, (8), 1676-83.

21.   Searle, B. C.; Dasari, S.; Turner, M.; Reddy, A. P.; Choi, D.; Wilmarth, P. A.; McCormack, A. L.; David, L. L.; Nagalla, S. R., High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. *Anal Chem* **2004,** 76, (8), 2220-30.

22.   Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A., Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol* **2005,** 23, (12), 1562-7.

23.   Taylor, J. A.; Johnson, R. S., Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal Chem* **2001,** 73, (11), 2594-604.

24.   Havilio, M.; Wool, A., Large-scale unrestricted identification of post-translation modifications using tandem mass spectrometry. *Anal Chem* **2007,** 79, (4), 1362-8.

25.   Hernandez, P.; Gras, R.; Frey, J.; Appel, R. D., Popitam: towards new heuristic strategies to improve protein identification from tandem mass spectrometry data. *Proteomics* **2003,** 3, (6), 870-8.

26.   Tanner, S.; Shu, H.; Frank, A.; Wang, L. C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V., InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem* **2005,** 77, (14), 4626-39.

27.   Clauser, K. R.; Baker, P.; Burlingame, A. L., Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* **1999,** 71, (14), 2871-82.

28.   Taylor, J. A.; Johnson, R. S., Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **1997,** 11, (9), 1067-75.

29.   Pevzner, P. A.; Dancik, V.; Tang, C. L., Mutation-tolerant protein identification by mass spectrometry. *J Comput Biol* **2000,** 7, (6), 777-87.

30.   Pevzner, P. A.; Mulyukov, Z.; Dancik, V.; Tang, C. L., Efficiency of database search for identification of mutated and modified proteins via mass spectrometry. *Genome Res* **2001,** 11, (2), 290-9.

31.   Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G., PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2003,** 17, (20), 2337-42.

32.   Frank, A.; Pevzner, P., PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **2005,** 77, (4), 964-73.

33.   Grossmann, J.; Roos, F. F.; Cieliebak, M.; Liptak, Z.; Mathis, L. K.; Muller, M.; Gruissem, W.; Baginsky, S., AUDENS: a tool for automated peptide de novo sequencing. *J Proteome Res* **2005,** 4, (5), 1768-74.

34.   Searle, B. C.; Dasari, S.; Wilmarth, P. A.; Turner, M.; Reddy, A. P.; David, L. L.; Nagalla, S. R., Identification of protein modifications using MS/MS de novo sequencing and the OpenSea alignment algorithm. *J Proteome Res* **2005,** 4, (2), 546-54.

35.   Wilmarth, P. A.; Tanner, S.; Dasari, S.; Nagalla, S. R.; Riviere, M. A.; Bafna, V.; Pevzner, P. A.; David, L. L., Age-related changes in human crystallins determined from comparative analysis of post-translational modifications in young and aged lens: does deamidation contribute to crystallin insolubility? *J Proteome Res* **2006,** 5, (10), 2554-66.

36.   Lapko, V. N.; Smith, D. L.; Smith, J. B., Expression of betaA2-crystallin in human lenses. *Exp Eye Res* **2003,** 77, (3), 383-5.

37.   MacCoss, M. J.; McDonald, W. H.; Saraf, A.; Sadygov, R.; Clark, J. M.; Tasto, J. J.; Gould, K. L.; Wolters, D.; Washburn, M.; Weiss, A.; Clark, J. I.; Yates, J. R., 3rd, Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A* **2002,** 99, (12), 7900-5.

38.   Lin, P. P.; Barry, R. C.; Smith, D. L.; Smith, J. B., In vivo acetylation identified at lysine 70 of human lens alphaA-crystallin. *Protein Sci* **1998,** 7, (6), 1451-7.

39.   Lapko, V. N.; Smith, D. L.; Smith, J. B., In vivo carbamylation and acetylation of water-soluble human lens alphaB-crystallin lysine 92. *Protein Sci* **2001,** 10, (6), 1130-6.

40.   Lapko, V. N.; Smith, D. L.; Smith, J. B., Methylation and carbamylation of human gamma-crystallins. *Protein Sci* **2003,** 12, (8), 1762-74.

41.   Liu, X.; Li, S., Carbamylation of human lens gamma-crystallins: relevance to cataract formation. *Yan Ke Xue Bao* **1993,** 9, (3), 136-42, 157.

42.   Takemoto, L., Deamidation of Asn-143 of gamma S crystallin from protein aggregates of the human lens. *Curr Eye Res* **2001,** 22, (2), 148-53.

43.   Takemoto, L.; Fujii, N.; Boyle, D., Mechanism of asparagine deamidation during human senile cataractogenesis. *Exp Eye Res* **2001,** 72, (5), 559-63.

44.   Lapko, V. N.; Purkiss, A. G.; Smith, D. L.; Smith, J. B., Deamidation in human gamma S-crystallin from cataractous lenses is influenced by surface exposure. *Biochemistry* **2002,** 41, (27), 8638-48.

45.   Takemoto, L.; Boyle, D., Increased deamidation of asparagine during human senile cataractogenesis. *Mol Vis* **2000,** 6, 164-8.

46.   Schey, K. L.; Little, M.; Fowler, J. G.; Crouch, R. K., Characterization of human lens major intrinsic protein structure. *Invest Ophthalmol Vis Sci* **2000,** 41, (1), 175-82.

47.   Hanson, S. R.; Smith, D. L.; Smith, J. B., Deamidation and disulfide bonding in human lens gamma-crystallins. *Exp Eye Res* **1998,** 67, (3), 301-12.

48.   Lampi, K. J.; Amyx, K. K.; Ahmann, P.; Steel, E. A., Deamidation in human lens betaB2-crystallin destabilizes the dimer. *Biochemistry* **2006,** 45, (10), 3146-53.

49.   Ma, Z.; Hanson, S. R.; Lampi, K. J.; David, L. L.; Smith, D. L.; Smith, J. B., Age-related changes in human lens crystallins identified by HPLC and mass spectrometry. *Exp Eye Res* **1998,** 67, (1), 21-30.

50.   Hanson, S. R.; Hasan, A.; Smith, D. L.; Smith, J. B., The major in vivo modifications of the human water-insoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage. *Exp Eye Res* **2000,** 71, (2), 195-207.

51.   Lund, A. L.; Smith, J. B.; Smith, D. L., Modifications of the water-insoluble human lens alpha-crystallins. *Exp Eye Res* **1996,** 63, (6), 661-72.

52.   Takemoto, L.; Horwitz, J.; Emmons, T., Oxidation of the N-terminal methionine of lens alpha-A crystallin. *Curr Eye Res* **1992,** 11, (7), 651-5.

53.   Kamei, A.; Hamaguchi, T.; Matsuura, N.; Iwase, H.; Masuda, K., Post-translational modification of alphaB-crystallin of normal human lens. *Biol Pharm Bull* **2000,** 23, (2), 226-30.

54.   Kamei, A.; Hamaguchi, T.; Matsuura, N.; Masuda, K., Does post-translational modification influence chaperone-like activity of alpha-crystallin? I. Study on phosphorylation. *Biol Pharm Bull* **2001,** 24, (1), 96-9.

55.   Srivastava, O. P.; Srivastava, K., Existence of deamidated alphaB-crystallin fragments in normal and cataractous human lenses. *Mol Vis* **2003,** 9, 110-8.

56.   Parker, N. R.; Korlimbinis, A.; Jamie, J. F.; Davies, M. J.; Truscott, R. J., Reversible binding of kynurenine to lens proteins: potential protection by glutathione in young lenses. *Invest Ophthalmol Vis Sci* **2007,** 48, (8), 3705-13.

57.   Aquilina, J. A.; Truscott, R. J., Identifying sites of attachment of UV filters to proteins in older human lenses. *Biochim Biophys Acta* **2002,** 1596, (1), 6-15.

58.   Aquilina, J. A.; Carver, J. A.; Truscott, R. J., Oxidation products of 3-hydroxykynurenine bind to lens proteins: relevance for nuclear cataract. *Exp Eye Res* **1997,** 64, (5), 727-35.

59.   Flaugh, S. L.; Mills, I. A.; King, J., Glutamine deamidation destabilizes human gammaD-crystallin and lowers the kinetic barrier to unfolding. *J Biol Chem* **2006,** 281, (41), 30782-93.

60.   Terrence Geiger, S. C., Deamidation, Isomerization, and Racemization at Asparaginyl and Aspartyl Residues in Peptides. *The Journal of Biological Chemistry* **1986,** 262, (2), 785-794.

61.   Fujii, N., D-amino acid in elderly tissues. *Biol Pharm Bull* **2005,** 28, (9), 1585-9.

62.   McCudden, C. R.; Kraus, V. B., Biochemistry of amino acid racemization and clinical application to musculoskeletal disease. *Clin Biochem* **2006**.

63.   Kim, Y. H.; Kapfer, D. M.; Boekhorst, J.; Lubsen, N. H.; Bachinger, H. P.; Shearer, T. R.; David, L. L.; Feix, J. B.; Lampi, K. J., Deamidation, but not truncation, decreases the urea stability of a lens structural protein, betaB1-crystallin. *Biochemistry* **2002,** 41, (47), 14076-84.

64.   Stadtman, E. R., Protein modification in aging. *J Gerontol* **1988,** 43, (5), B112-20.

65.   Joshi, A. B.; Sawai, M.; Kearney, W. R.; Kirsch, L. E., Studies on the mechanism of aspartic acid cleavage and glutamine deamidation in the acidic degradation of glucagon. *J Pharm Sci* **2005,** 94, (9), 1912-27.

66.   Manning, M. C.; Patel, K.; Borchardt, R. T., Stability of protein pharmaceuticals. *Pharm Res* **1989,** 6, (11), 903-18.

67.   Brange, J.; Langkjoer, L., Insulin structure and stability. *Pharm Biotechnol* **1993,** 5, 315-50.

68.   Huang, L.; Lu, J.; Wroblewski, V. J.; Beals, J. M.; Riggin, R. M., In vivo deamidation characterization of monoclonal antibody by LC/MS/MS. *Anal Chem* **2005,** 77, (5), 1432-9.

69.   Reissner, K. J.; Aswad, D. W., Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals? *Cell Mol Life Sci* **2003,** 60, (7), 1281-95.

70.   Watanabe, A.; Hong, W. K.; Dohmae, N.; Takio, K.; Morishima-Kawashima, M.; Ihara, Y., Molecular aging of tau: disulfide-independent aggregation and non-enzymatic degradation in vitro and in vivo. *J Neurochem* **2004,** 90, (6), 1302-11.

71.   Shimizu, T.; Matsuoka, Y.; Shirasawa, T., Biological significance of isoaspartate and its repair system. *Biol Pharm Bull* **2005,** 28, (9), 1590-6.

72.   Watanabe, A.; Takio, K.; Ihara, Y., Deamidation and isoaspartate formation in smeared tau in paired helical filaments. Unusual properties of the microtubule-binding domain of tau. *J Biol Chem* **1999,** 274, (11), 7368-78.

73.   van de Wal, Y.; Kooy, Y.; van Veelen, P.; Pena, S.; Mearin, L.; Papadopoulos, G.; Koning, F., Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J Immunol* **1998,** 161, (4), 1585-8.

74.   Skovbjerg, H.; Koch, C.; Anthonsen, D.; Sjostrom, H., Deamidation and cross-linking of gliadin peptides by transglutaminases and the relation to celiac disease. *Biochim Biophys Acta* **2004,** 1690, (3), 220-30.

75.   Riha, W. E., 3rd; Izzo, H. V.; Zhang, J.; Ho, C. T., Nonenzymatic deamidation of food proteins. *Crit Rev Food Sci Nutr* **1996,** 36, (3), 225-55.

76.   Atwood, J. A., 3rd; Sahoo, S. S.; Alvarez-Manilla, G.; Weatherly, D. B.; Kolli, K.; Orlando, R.; York, W. S., Simple modification of a protein database for mass spectral identification of N-linked glycopeptides. *Rapid Commun Mass Spectrom* **2005,** 19, (21), 3002-6.

77.   Karty, J. A.; Reilly, J. P., Deamidation as a consequence of beta-elimination of phosphopeptides. *Anal Chem* **2005,** 77, (14), 4673-6.

78.   Robinson, N. E.; Robinson, A. B., Molecular clocks. *Proc Natl Acad Sci U S A* **2001,** 98, (3), 944-9.

79.   Wakankar, A. A.; Borchardt, R. T., Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization. *J Pharm Sci* **2006,** 95, (11), 2321-36.

80.   Zomber, G.; Reuveny, S.; Garti, N.; Shafferman, A.; Elhanany, E., Effects of spontaneous deamidation on the cytotoxic activity of the Bacillus anthracis protective antigen. *J Biol Chem* **2005,** 280, (48), 39897-906.

81.   Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D., Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* **1995,** 67, (8), 1426-36.

82.   Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000,** 35, (12), 1399-406.

83.   Gu, C.; Tsaprailis, G.; Breci, L.; Wysocki, V. H., Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal Chem* **2000,** 72, (23), 5804-13.

84.   Kapp, E. A.; Schutz, F.; Reid, G. E.; Eddes, J. S.; Moritz, R. L.; O'Hair, R. A.; Speed, T. P.; Simpson, R. J., Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. *Anal Chem* **2003,** 75, (22), 6251-64.

85.   Chelius, D.; Rehder, D. S.; Bondarenko, P. V., Identification and characterization of deamidation sites in the conserved regions of human immunoglobulin gamma antibodies. *Anal Chem* **2005,** 77, (18), 6004-11.

86.   Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G., Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* **2006,** 78, (18), 6645-50.

87.   Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A., ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics* **2006,** 5, (5), 935-48.

88.   Mueller, L. N.; Brusniak, M. Y.; Mani, D. R.; Aebersold, R., An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res* **2008,** 7, (1), 51-61.

89.   Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G., Comparison of label-free methods for

quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **2005,** 4, (10), 1487-502.

90.   Zybailov, B. L.; Florens, L.; Washburn, M. P., Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol Biosyst* **2007,** 3, (5), 354-60.

91.   Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **2004,** 76, (14), 4193-201.

92.   MacCoss, M. J.; Wu, C. C.; Liu, H.; Sadygov, R.; Yates, J. R., 3rd, A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal Chem* **2003,** 75, (24), 6912-21.

93.   Li, X. J.; Zhang, H.; Ranish, J. A.; Aebersold, R., Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal Chem* **2003,** 75, (23), 6648-57.

94.   Pan, C.; Kora, G.; McDonald, W. H.; Tabb, D. L.; VerBerkmoes, N. C.; Hurst, G. B.; Pelletier, D. A.; Samatova, N. F.; Hettich, R. L., ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal Chem* **2006,** 78, (20), 7121-31.

95.   Halligan, B. D.; Slyper, R. Y.; Twigger, S. N.; Hicks, W.; Olivier, M.; Greene, A. S., ZoomQuant: an application for the quantitation of stable isotope labeled peptides. *J Am Soc Mass Spectrom* **2005,** 16, (3), 302-6.

96.   Schulze, W. X.; Mann, M., A novel proteomic screen for peptide-protein interactions. *J Biol Chem* **2004,** 279, (11), 10756-64.

97.   Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **1999,** 17, (10), 994-9.

98.   Robinson, N. E.; Lampi, K. J.; McIver, R. T.; Williams, R. H.; Muster, W. C.; Kruppa, G.; Robinson, A. B., Quantitative measurement of deamidation in lens betaB2-crystallin and peptides by direct electrospray injection and fragmentation in a Fourier transform mass spectrometer. *Mol Vis* **2005,** 11, 1211-9.

99.   Robinson, N. E.; Zabrouskov, V.; Zhang, J.; Lampi, K. J.; Robinson, A. B., Measurement of deamidation of intact proteins by isotopic envelope and mass defect with ion cyclotron resonance Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom* **2006,** 20, (23), 3535-41.

100.   Bloemendal, H.; de Jong, W.; Jaenicke, R.; Lubsen, N. H.; Slingsby, C.; Tardieu, A., Ageing and vision: structure, stability and function of lens crystallins. *Prog Biophys Mol Biol* **2004,** 86, (3), 407-85.

101.   Hoehenwarter, W.; Klose, J.; Jungblut, P. R., Eye lens proteomics. *Amino Acids* **2006,** 30, (4), 369-89.

102.   Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd, DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **2002,** 1, (1), 21-6.

103.   Pedrioli, P. G.; Eng, J. K.; Hubley, R.; Vogelzang, M.; Deutsch, E. W.; Raught, B.; Pratt, B.; Nilsson, E.; Angeletti, R. H.; Apweiler, R.; Cheung, K.; Costello, C. E.; Hermjakob, H.; Huang, S.; Julian, R. K.; Kapp, E.; McComb, M. E.; Oliver, S. G.; Omenn, G.; Paton, N. W.; Simpson, R.; Smith, R.; Taylor, C. F.; Zhu, W.; Aebersold, R., A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol* **2004,** 22, (11), 1459-66.

104.   Rustvold, D. L., Wilmarth, P.A., Dasari, S., and David, L.L. In *RAPID: Coordinated Visualization of Multiple Charge States from Individual or Isotopically Labeled Peptides in Liquid Chromatography Mass Spectrometry Experiments*, 54th American Society of Mass Spectrometry Conference on Mass Spectrometry, Seattle, WA, 2006; Seattle, WA, 2006.

105.   Savitzky, A., Golay, M.J.E, Smoothing and Differentiation of Data by Simiplified Least Squares Procedure. *Anal Chem* **1964,** 36, (8), 1627-1639.

106.   Clarke, S., Aging as war between chemical and biochemical processes: protein methylation and the recognition of age-damaged proteins for repair. *Ageing Res Rev* **2003,** 2, (3), 263-85.

107.   Ball, L. E.; Garland, D. L.; Crouch, R. K.; Schey, K. L., Post-translational modifications of aquaporin 0 (AQP0) in the normal human lens: spatial and temporal occurrence. *Biochemistry* **2004,** 43, (30), 9856-65.

108.   Carter, D. A.; McFadden, P. N., Determination of beta-isomerized aspartic acid as the corresponding alcohol. *J Protein Chem* **1994,** 13, (1), 97-106.

109.   Lampi, K. J.; Ma, Z.; Shih, M.; Shearer, T. R.; Smith, J. B.; Smith, D. L.; David, L. L., Sequence analysis of betaA3, betaB3, and betaA4 crystallins completes the identification of the major proteins in young human lens. *J Biol Chem* **1997,** 272, (4), 2268-75.

110.   Miesbauer, L. R.; Zhou, X.; Yang, Z.; Yang, Z.; Sun, Y.; Smith, D. L.; Smith, J. B., Post-translational modifications of water-soluble human lens crystallins from young adults. *J Biol Chem* **1994,** 269, (17), 12494-502.

111.   Takemoto, L., Increased deamidation of asparagine-101 from alpha-A crystallin in the high molecular weight aggregate of the normal human lens. *Exp Eye Res* **1999,** 68, (5), 641-5.

112.   Takemoto, L.; Boyle, D., Deamidation of alpha-A crystallin from nuclei of cataractous and normal human lenses. *Mol Vis* **1999,** 5, 2.

113.   Takemoto, L. J., Quantitation of asparagine-101 deamidation from alpha-A crystallin during aging of the human lens. *Curr Eye Res* **1998,** 17, (3), 247-50.

114.   Kramps, J. A.; de Jong, W. W.; Wollensak, J.; Hoenders, H. J., The polypeptide chains of alpha-crystallin from old human eye lenses. *Biochim Biophys Acta* **1978,** 533, (2), 487-95.

115.   Harrington, V.; McCall, S.; Huynh, S.; Srivastava, K.; Srivastava, O. P., Crystallins in water soluble-high molecular weight protein fractions and water insoluble protein fractions in aging and cataractous human lenses. *Mol Vis* **2004,** 10, 476-89.

116.   Zhang, Z.; David, L. L.; Smith, D. L.; Smith, J. B., Resistance of human betaB2-crystallin to in vivo modification. *Exp Eye Res* **2001,** 73, (2), 203-11.

117.   Harms, M. J.; Wilmarth, P. A.; Kapfer, D. M.; Steel, E. A.; David, L. L.; Bachinger, H. P.; Lampi, K. J., Laser light-scattering evidence for an altered association of beta B1-crystallin deamidated in the connecting peptide. *Protein Sci* **2004,** 13, (3), 678-86.

118.   Purkiss, A. G.; Bateman, O. A.; Goodfellow, J. M.; Lubsen, N. H.; Slingsby, C., The X-ray crystal structure of human gamma S-crystallin C-terminal domain. *J Biol Chem* **2002,** 277, (6), 4199-205.

7. TABLES

Table 1: SCX and RP retention times for synthetic peptides.

Peptide		SCX [a] Retention Time (min)			RP [b] Retention Time (min)		
Unmodified [c]	Deamidated [d]	Unmodified	Deamidated	Δ [e]	Unmodified	Deamidated	Δ [e]
ISLFEGA**N**FK	ISLFEGA**D**FK	30.4	30.1	-0.3	56.15	59.26	3.11
GY**Q**YLLEPGDFR	GY**E**YLLEPGDFR	32.6	30.5	-2.1	60.40	63.35	2.95
HW**N**EWGAFQPQMQSLR	HW**D**EWGAFQPQMQSLR	50.2	48.8	-1.4	59.93	63.09	3.16
HWNEWGAF**Q**PQMQSLR	HWNEWGAF**E**PQMQSLR	50.2	52.0	1.8	59.93	62.97	3.04
HWNEWGAFQPQM**Q**SLR	HWNEWGAFQPQM**E**SLR	50.2	51.6	1.4	59.93	62.74	2.81

[a]Strong cation exchange chromatography.
[b]Reverse-phase chromatography.
[c]Amide-containing synthetic βB1 peptide (deamidation site in bold).
[d]Acid-containing synthetic βB1 peptide (deamidation site in bold).
[e]Retention time difference between deamidated and unmodified peptide.

Table 2: Deamidation sites detected in lens crystallin proteins using low resolution mass spectrometry (3-D ion trap) of 3-day, 2-year, 18-year, 35-year, 70-year old human normal lens, and a 93-year old human cataractous lens samples that passed the automated deamidation validation procedure outlined in Fig. 3.

Protein	Deamidation Sites [a]					
	3-day	2-year	18-year	35-year	70-year	93-year Catactous
αA	Q90, Q147		Q6, Q90, N101, Q147	Q6, Q90, [N123-Q126], Q147	Q6, Q90, Q50, [N123-Q126], Q147	Q6, Q90, [N123-Q126], Q147
αB				N78	**Q26**, N78	**Q26**, N78, Q108
βBA1/A3	N103, [Q203-Q206 -Q208]	[Q38-N40-Q42], N103, N133	[Q38-N40-Q42], N103, N120, N133	[Q38-N40-Q42], N103, N120, N133, [Q164-Q172], [Q203-Q206-Q208]	[Q38-N40-Q42], N103, N120, N133, [Q164-Q172], [Q203-Q206-Q208]	[Q38-N40-Q42], N103, N120, N133, **Q149**, [Q164-Q172], [Q203-Q206-Q208]
βA4	Q62	[Q111-N113]	N82	Q22, [Q111-N113]	Q22, [Q111-N113]	Q22, N82, [Q111-N113]
βB1		**N81**, N157	N157	[N67-Q69], [Q105-N107], N124, N157	[N67-Q69], **N81**, [Q105-N107], N124, N157, Q196	[N67-Q69], **N81**, [Q105-N107], N124, N157, Q196, Q204, N216, **Q235**

	3-day old	2-year old	22-year old	35-year old	70-year old	93-year old Cataractous
βB2	[**Q193-Q196**]	[N113-N115]	[N113-N115], **Q137**, [**Q146-Q154**], [Q182-Q184]	[N113-N115], **Q137**, [**Q146-Q154**]	[N113-N115], **Q137**, [**Q146-Q154**], [Q182-Q184], [**Q193-Q196**]	[N113-N115], **Q137**, [**Q146-Q154**], [**Q193-Q196**]
βB3		N155		N155	**Q78**	N155
γS	Q170	[N14-Q16]	[N14-Q16], Q92	[N14-Q16], [Q63-Q70], Q92, **Q96**, Q106, Q120	[N14-Q16], [Q63-Q70], Q92, **Q96**, Q120, Q170	[N14-Q16], N53, [Q63-Q70], Q92, **Q96**, Q120, **Q148**, Q170
γB		**N161**			**N161**	
γC		Q26, Q83	Q83, Q112	Q83, **Q51**, Q112	Q83, Q112, [**Q142-Q148**], **Q154**	Q66, Q83
γD			N160		Q12, **Q67**, N160	Q12, **Q154**, N160

aSites highlighted in bold are reported here for the first time. Ambiguous, multiple deamidation sites located in the same peptide are denoted using square brackets.

Table 3: Deamidation sites detected in lens crystallin proteins using high resolution mass spectrometry (Q-TOF) of 3-day, 2-year, 22-year, 35-year, 70-year old human normal lens, and a 93-year old human cataractous lens samples.

Protein	Deamidation Sites [a]					
	3-day old	**2-year old**	**22-year old**	**35-year old**	**70-year old**	**93-year old Cataractous**
βA1/βA3	**Q7**, **Q19**, **N21**, N54, N62, N120, N133, Q164, Q180, Q203, Q206, Q208	**Q7**, **Q8**, **Q19**, **N21**, Q38, N40, N54, N120, N133, Q164, Q172, Q180, Q203, Q206	**N21**, Q38, N40, Q42, N120, N133, Q164, Q172, Q180, Q203, Q206	**N21**, Q38, N40, Q42, N54, N62, N120, N133, **Q138**, Q149, N156, N155, Q164, Q172, Q180, Q203, Q206, Q208	**Q19**, Q38, N40, Q42, N54, N62, N120, N133, Q149, N155, N156, Q164, Q172, Q180, Q203, Q206, Q208	**N21**, Q38, N40, Q42, N54, N62, N120, N133, **Q138**, Q149, N155, N156, Q164, Q172, Q180, Q203, Q206, Q208
βA2	**N101**			**Q66**	**Q66**	**Q66**
βA4	Q22, N82, **Q160**, **Q186**	Q22, N82, Q111, **Q160**, **Q186**	Q22, N82, Q111, N113, **Q160**, **Q186**	Q22, N82, **N100**, Q111, N113, **Q186**	Q22, Q62, N82, N113, **Q160**, **Q186**	Q22, Q62, Q64, N82, N113, **Q160**, **Q186**, **Q188**
βB1	N67, Q69, N81, N124, N157, Q196, Q204, Q222, Q224, Q226, Q235	N67, Q69, N81, N124, N157, N161, Q196, Q204, Q235	N67, Q69, N81, N107, N157, N161, Q196, Q204, N216, Q222, Q224, Q235	N67, Q69, N81, N107, N124, Q146, N157, N161, Q166, Q196, Q204, N216, Q222, Q224, Q226, Q235	N67,Q69, N81, Q105, N107, N124, N157, N161, Q166, Q196, Q204, N216, Q222, Q224, Q226, Q235	N67,Q69, N81, Q105, N107, N124, Q146, N157, N161, Q166, Q196 Q204, N216, Q222, Q224, Q226, Q235
βB2	**N25**, **Q27**, **N34**, **N39**, **Q54**, **N65**, **Q104**, N113, N115, Q137, Q146, Q154, **Q162**, Q182	**Q27**, **N34**, **N39**, **Q54**, **Q63**, **Q70**, **Q104**, N113, N115, Q137, Q146, Q154, **Q162**, Q182, Q184	**N25**, **N34**, **Q63**, **Q54**, **N65**, **Q70**, **Q104**, N113, N115, Q137, Q146, Q154, Q182, Q184	**Q27**, **N34**, **N39**, **Q63**, **Q70**, **Q104**, N113, N115, Q137, Q146, Q154, Q162, Q182, Q184	**N25**, **Q27**, **Q63**, **N65**, **Q70**, **Q104**, N113, N115, Q137, Q146, Q154, Q162, Q182, Q184	**N25**, **Q27**, **N34**, **N39**, **Q54**, **N65**, **Q70**, **Q104**, N113, N115, Q137, Q146, Q154, Q162, Q182, Q184
βB3	N155	**N33**, **Q60**, N155	**Q60**, N155	N155	**Q60**, N155	**Q60**, N155
βS	N14, Q16, N53, Q106, Q170	N14, Q16, N53, Q63, Q70, Q106, Q120, Q170	N14, Q16, N53, Q63, Q70, Q120, Q170	N14, Q16, N53, Q63, Q70, Q106, Q120, Q170	N14, Q16, N53, Q63, Q70, Q106, Q120, N143, Q170	N14, Q16, N53, Q63, Q70, Q106, Q120, N143, Q170
γA	**N17**, **Q51**					

60

γB	Q66, **Q67**	Q66, **Q67**	**Q26**, Q66, **Q67**, **Q101**	**Q26**, Q66, **Q67**	**Q26**, Q66, **Q67**	**N24**, **Q26**, **N49**, Q66, **Q67**
γC	N24, Q26, Q51, **Q53**, Q66, Q67, Q112, Q148, Q154	N24, Q26, N49, Q51, **Q53**, **Q54**, Q66, Q67, Q83, Q112	N24, Q26, N49, Q51, **Q53**, **Q54**, Q66, Q67, Q83, Q112	N24, Q26, N49, Q51, **Q53**, **Q54**, Q66, Q67, Q83, Q112, Q154	N24, Q51, N49, **Q53**, **Q54**, Q66, Q67, Q83, Q112, Q154	N24, Q26, N49, Q51, **Q53**, **Q54**, Q66, Q67, Q83, Q112, N137, Q142
γD	N24, **Q66**, Q67	N24	N24, **Q66**	N24, N49, **Q66**	Q12, N24, **Q26**, N49, **Q54**, **Q66**, Q67, Q154	Q12, N24, **Q26**, **Q47**, N49, **Q54**, **Q66**, Q67, **N137**, **Q142**
αA	Q6, Q90, Q147	Q6, Q90, N101, Q104, Q126, Q147,	Q6, Q90, N101, Q126	Q6, Q90, N101, Q126, Q147	Q6, Q90, N123, Q126, Q147	Q6, Q50, Q90, N123, Q126, Q147
αB		N146	N146	N146	N78, N146	Q26, N146

[a]Sites highlighted in bold are reported here for the first time.

8. Figures
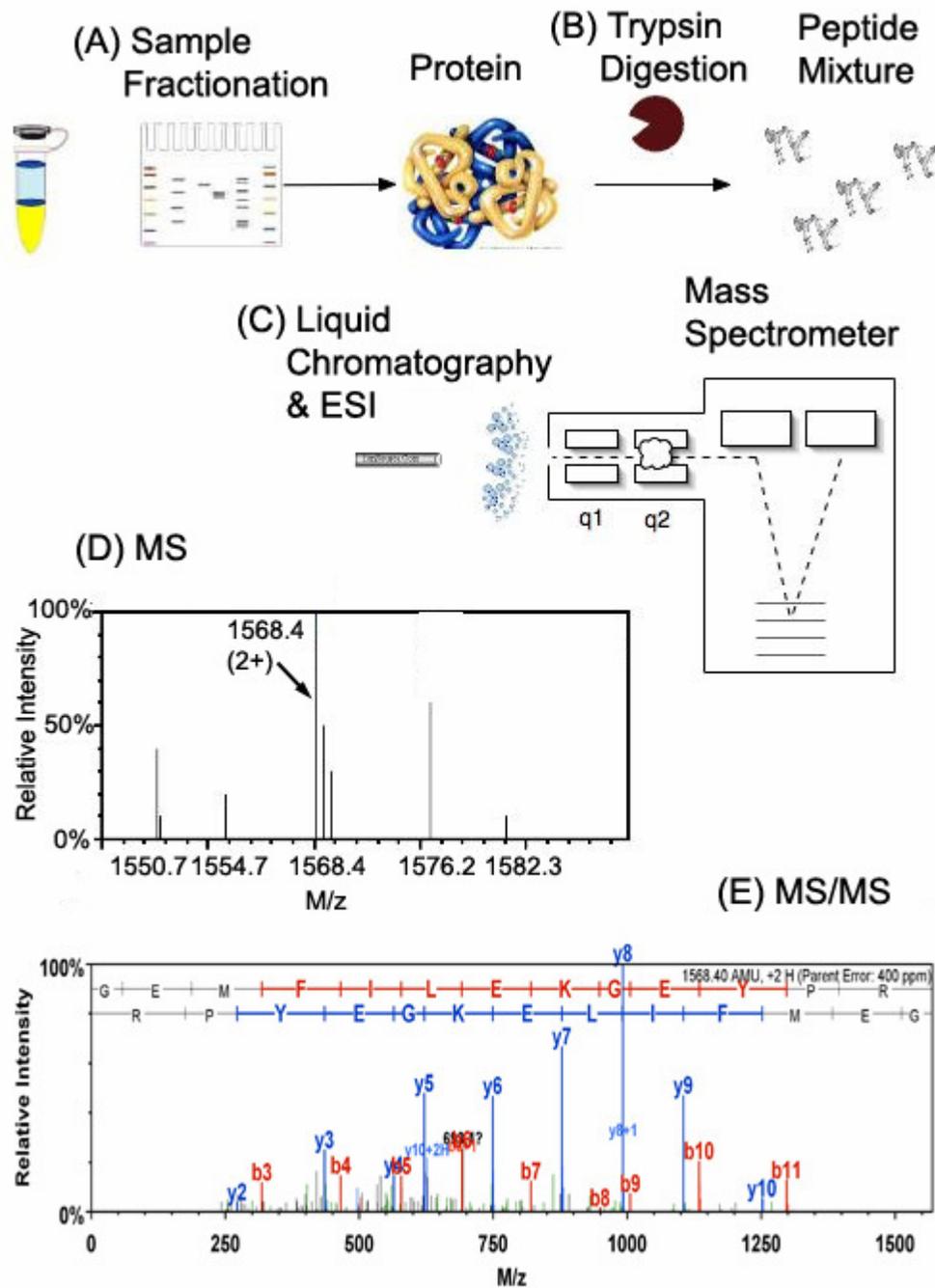
Figure 1. Tandem Mass Spectrometry based Proteomics: (A) Proteins in a sample are fractionated using either gel-based or liquid-based techniques. (B) Proteins in fractions are digested into peptides using proteases. (C) Resulting peptide mixtures are time separated using reverse-phase liquid chromatography (LC), ionized using electro spray

ionization (ESI), and analyzed by mass spectrometer to measure their masses (MS spectrum). (D) A peptide ion selected from a MS spectrum, fragmented using collision-induced dissociation, and a mass spectrum of its constituents will be acquired (tandem MS/MS). (E) The acquired tandem mass spectrum is searched against known proteins and their corresponding peptides to identify the peptide sequence present in the mass spectrum.
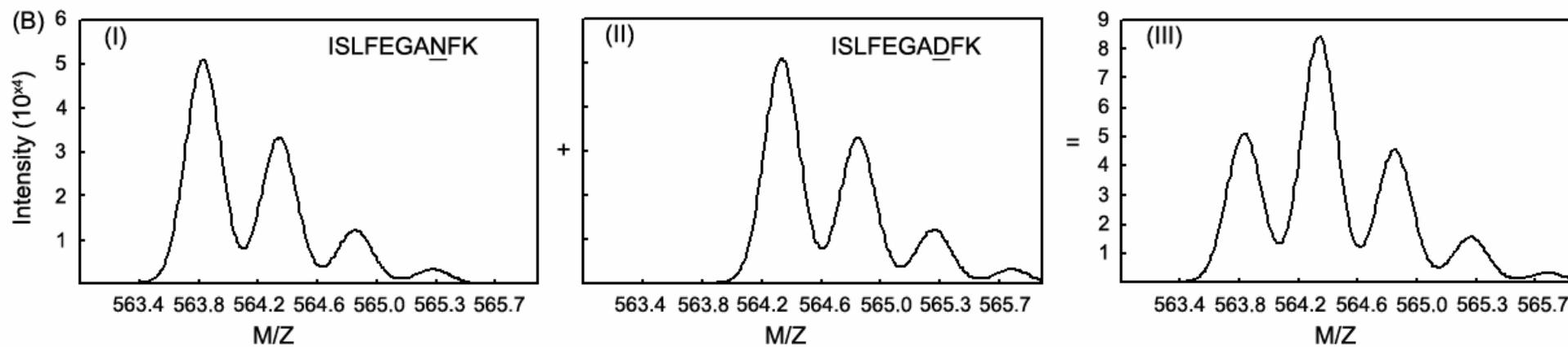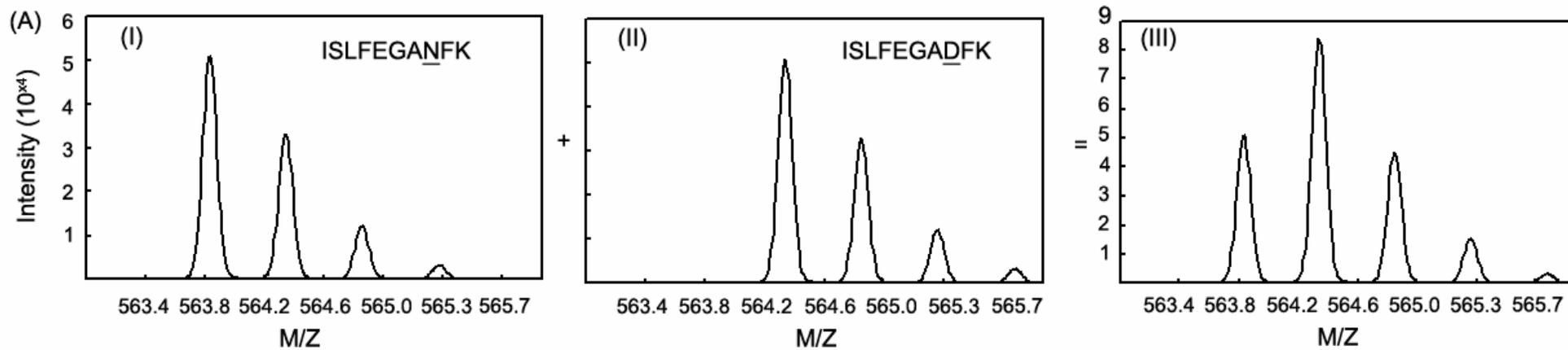
Figure 2. Deconvolution of Unmodified and Deamidated Peptide Isotopic Envelopes: Individual peptide isotopic envelopes of unmodified (I) and deamidated peptide (II) forms are simulated for a high-resolution time-of-flight mass spectrometer (A) and low resolution linear ion trap mass spectrometer (B). The combined mass chromatogram of both forms as measured by corresponding mass spectrometers are shown in figure A-III and B-III respectively. The isotopic envelope modeling method uses theoretical peptide isotopic envelopes to decompose combined mass chromatogram (III) into corresponding individual signals (unmodified –I and deamidated –II). The accuracy of decomposition process depends on the resolution of the mass spectrometer used to acquire the combined mass chromatogram.

For each peptide containing a deamidated N or Q reported by SEQUEST

       Generate XICs using theoretical singly charged peptide masses of the unmodified and deamidated peptide forms for all RP runs where that peptide was reported.

       Select chromatogram peaks associated with unmodified or deamidated forms of the peptide using the RTs of MS/MS identifications from SEQUEST.

       Calculate the experimental masses of the selected chromatogram peaks using the peptideMassCorrection procedure.

       Assign the unmodified reference peak to the most intense chromatogram peak having a monoisotopic mass within +/- 0.4 Da of the unmodified theoretical mass.

       For each MS/MS scan with a reported N or Q deamidation site

              Determine the RT corresponding to the apex of the chromatogram peak containing the MS/MS scan

              If the peak containing the MS/MS scan is not the reference peak then

                     Calculate the RTS between the apex RT values of the MS/MS containing peak and the reference peak.

                     Obtain corrected peptide masses for the MS/MS containing peak using peptideMassCorrection procedure.

                     Using the corrected masses, calculate ΔM between the reference peak peptide and the MS/MS containing peak peptide.
              End
       End
End

Procedure peptideMassCorrection(chromatogramPeak)

       Identify all the MS scans collected within the chromatogram peak and average them.

       Determine the most intense averaged centroided mass value in a -1.0 to +4.0 Da interval relative to the theoretical peptide mass.
End Procedure

Figure 3. Manual Deamidated Peptide Validation Procedure for LCQ: Pseudo code of manual procedure used to determine the retention time shifts (RTS) and experimental mass differences (ΔM) caused by deamidation.
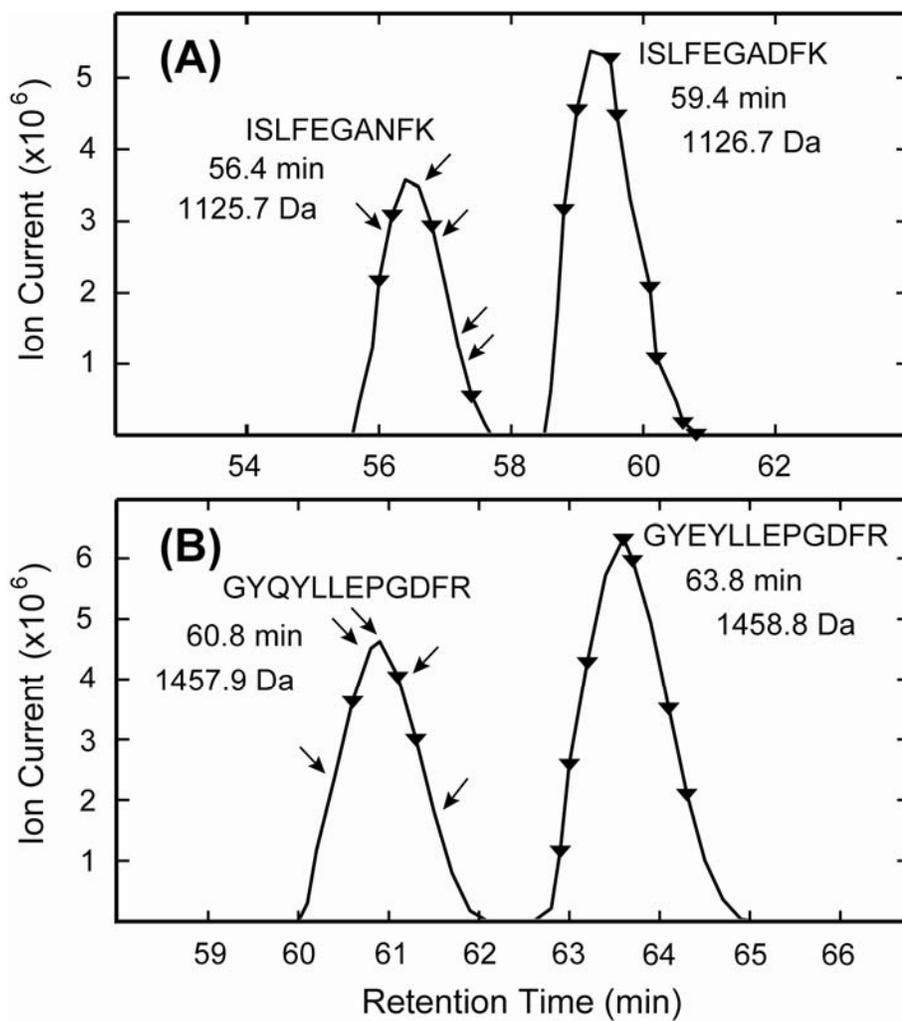
Figure 4. Synthetic Peptide Chromatograms on LCQ: Extracted ion chromatograms of 1+ charge states of (A) synthetic deamidated ISLFEGA<u>D</u>FK and unmodified (amidated) ISLFEGA<u>N</u>FK peptide mixtures and (B) GY<u>E</u>YLLEPGDFR and GY<u>Q</u>YLLEPGDFR peptide mixtures that were resolved by reverse-phase chromatography. The deamidated form of each peptide eluted approximately 3.0 min later than its corresponding amidated form. Experimental masses were determined by averaging the MS scans during each observed chromatography peak. MS/MS spectra were searched with SEQUEST and the unmodified (arrows) and deamidated peptide identifications (filled ∇) are indicated.

67
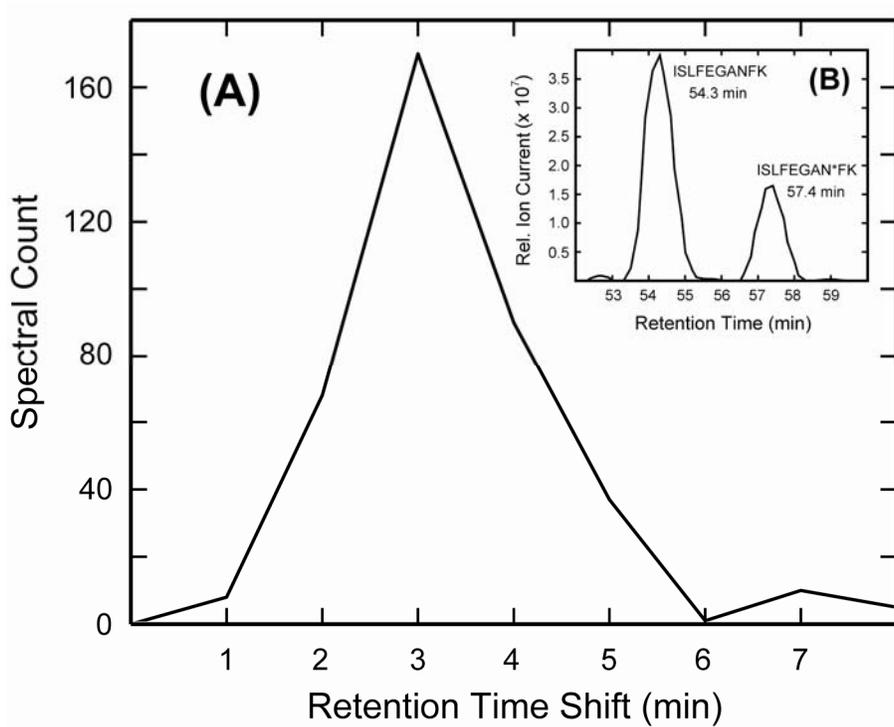
Figure 5. Retention Time Shift of Deamidated Peptides in 70-year Old Human Lens Digests: The RTS values between deamidation identifications and their corresponding unmodified peptide forms were manually calculated for the 70-year old normal lens tissue sample. The distribution of the RTS values showed that a majority of deamidated peptides had RTS values between 2 and 5 min relative to their corresponding unmodified forms. The insert shows the chromatograms associated with βB1 peptide ISLFEGANFK from the 70-year old water-insoluble lens fraction and is very similar to Fig. 4(A).
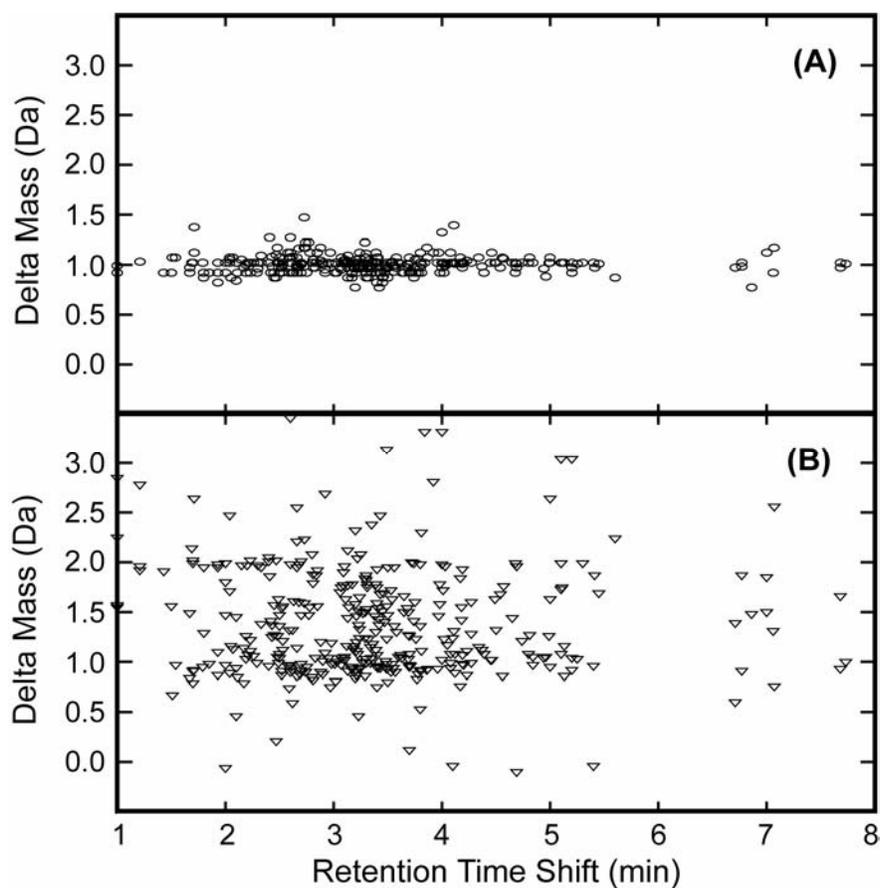
Figure 6. **Delta Mass vs. Retention Time Shift Scatter Plots:** Peptide mass differences (ΔM) were obtained by averaging MS survey scans during corresponding chromatogram peaks (o), or using peptide masses from the corresponding DTA files (∇). Scatter plots of ΔM vs. the averaged ΔM values in panel A are tightly clustered around the expected mass shift associated with deamidation (0.984 Da), whereas the ΔM values from DTA files in panel B are more stochastically distributed. RTS for peptides from the 70-year old lens are shown.
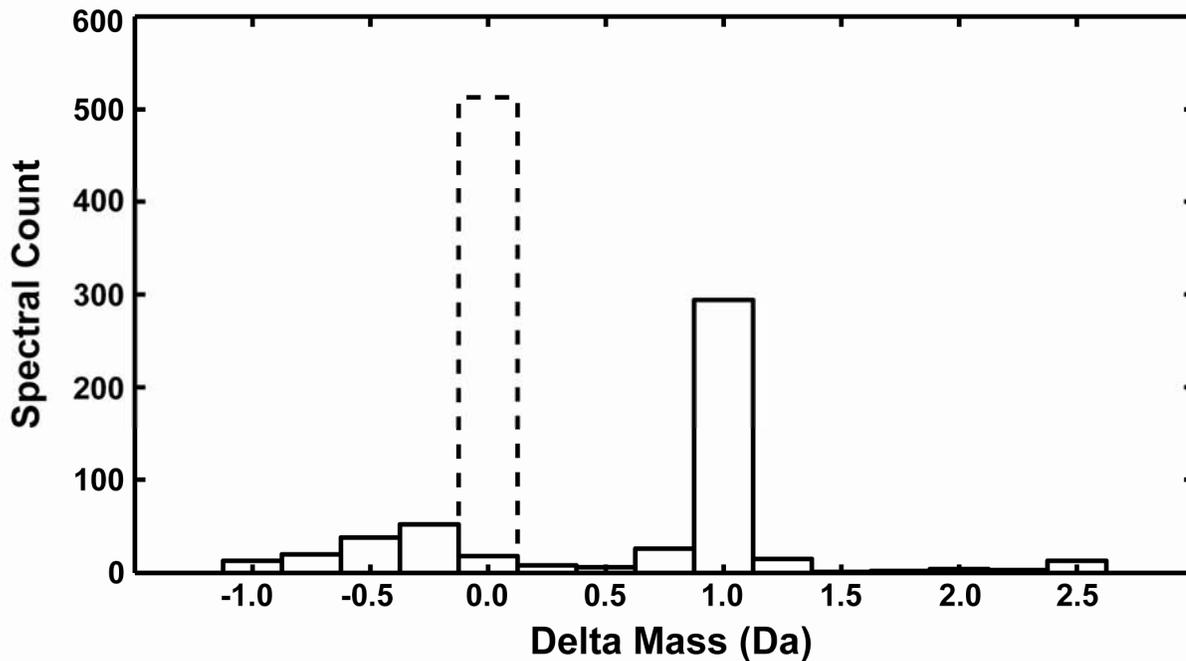
Figure 7. Delta Mass Histograms of False Positive Deamidation Identifications: The solid histograms correspond to the mass differences between parent ion masses written to DTA files and the theoretical masses of the unmodified peptides that were mis-identified as deamidated peptides. The prominent peak at 1.0 Da suggested that survey scans were frequently triggered on isotopic rather than monoisotopic peaks. The dotted histogram of the mass differences between corrected experimental parent ion masses and theoretical masses was centered on the expected value of 0.0 Da.
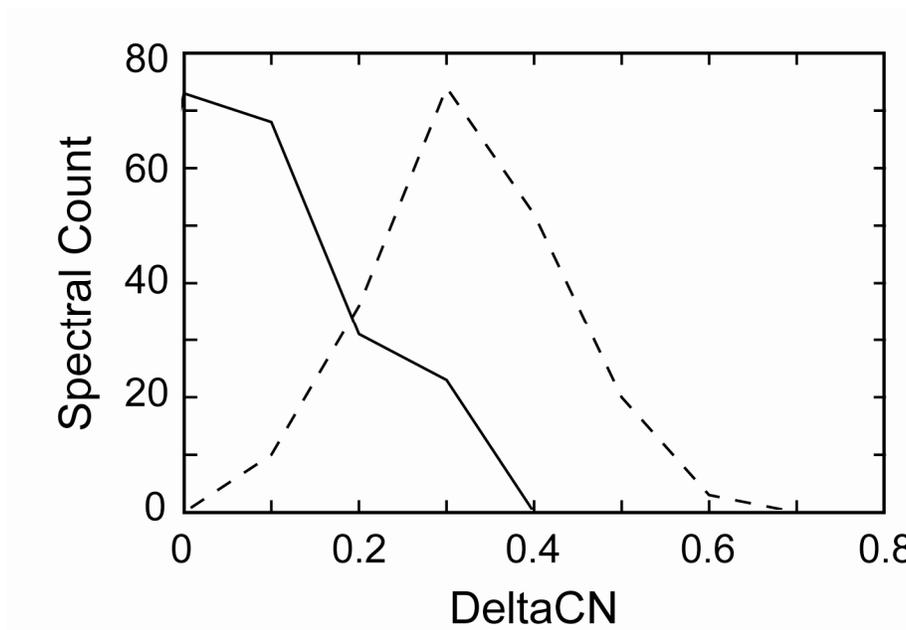
Figure 8. DeltaCN Distributions of Deamidation and Non-deamidation Searches: Distributions of DeltaCN values for SEQUEST searches with deamidations as variable modifications (solid line distribution), and for searches without differential modifications (broken line distribution). The MS/MS spectra producing each distribution are the identical set of deamidated peptide scans that passed the manual validation method. Using DeltaCN thresholds to filter deamidation results caused a significant decrease in sensitivity.
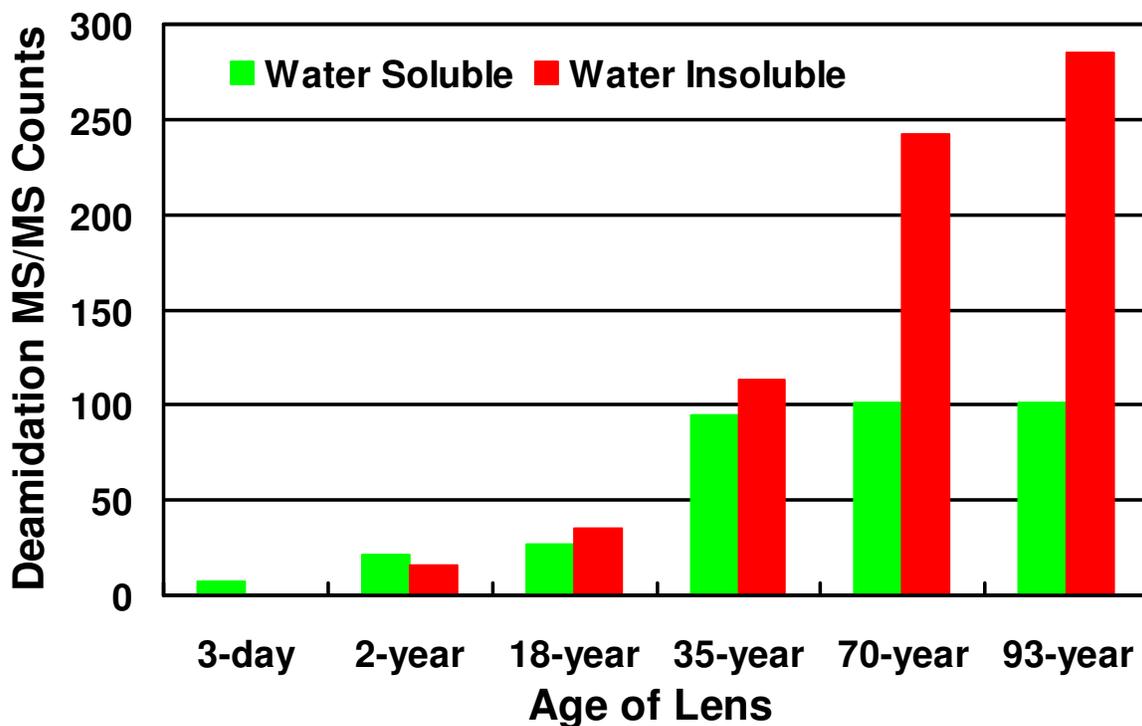
Figure 9. Deamidated Peptide MS/MS Counts for an Aging Series of Lens Samples: Total validated spectral (MS/MS) counts of deamidated peptides from lens crystallin proteins are shown for water-soluble and water-insoluble fractions from the six lens samples. Deamidation increases with age from 3-days to 93-years of age, particularly in the water-insoluble fractions.
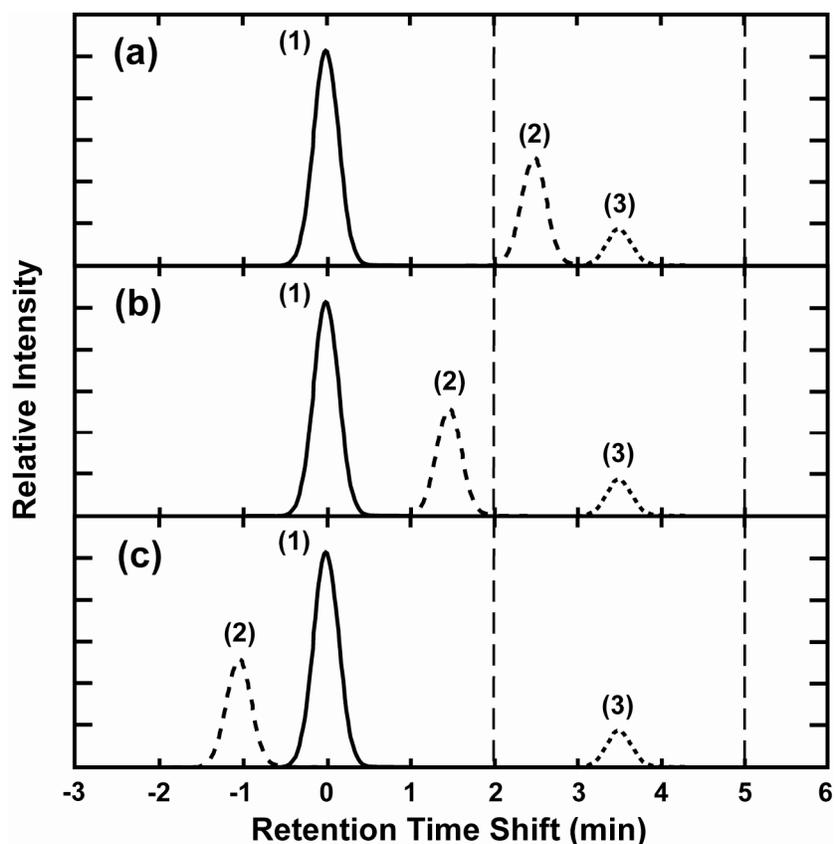
Case	Calculated Values	Test Results		Peptide Status
(a)	$\Delta M(1,2) = 1.0$ Da	TRUE	TRUE	deamidated
	$\Delta RT(1,2) = 2.5$ m	TRUE		
	$\Delta M(1,3) = 1.0$ Da	TRUE	TRUE	
	$\Delta RT(1,3) = 3.5$m	TRUE		
(b)	$\Delta M(1,2) = 1.0$ Da	TRUE	FALSE	deamidated
	$\Delta RT(1,2) = 1.5$ m	FALSE		
	$\Delta M(1,3) = 1.0$ Da	TRUE	TRUE	
	$\Delta RT(1,3) = 3.5$m	TRUE		
(c)	$\Delta M(1,2) = 1.0$ Da	TRUE	FALSE	deamidated
	$\Delta RT(1,2) = -1.0$ m	FALSE		
	$\Delta M(1,3) = 1.0$ Da	TRUE	TRUE	
	$\Delta RT(1,3) = 3.5$m	TRUE		

Figure 10. Effect of Isomerization on Reliability of Deamidation Identifications: The method outlined in this work is largely unaffected by the presence of iso-aspartate containing peptides. In the above hypothetical chromatograms, peak (1) is the unmodified peptide, peak (2) is the peptide containing iso-aspartate, and peak (3) is the peptide containing aspartate. The three cases show the variable reverse phase behavior of peptides containing iso-aspartate (Krokhin et al, Anal Chem 2006, 78(18), 6645-50) and

the table shows that the peptide would be correctly identified as being deamidated in all cases. The special case where the peptide containing iso-aspartate (2) co-elutes with the peptide containing asparagine (1) could cause an increase the mass of the unmodified peptide form and result in the peptide no longer passing the ΔM test.

For each peptide containing N or Q reported by SEQUEST in a SCX fraction

numberOfPeptideForms = count(unModifiedForm+number of deamidations in the peptide)

baseMass = m/z(unmodified form of peptide)

basePeptide = unmodified form of peptide

For each charge state detected for the peptide

generateIonChromatogram(baseMass, z)
End

Display ion chromatograms for all charge states, wait for user to select integration retention time bounds

For each charge state (z) detected for the peptide

massChromatogram = generateMassChromatogram (retentionTimeBounds, baseMass, z)

IEM = guessAnIEM(basePeptide, z, massChromatogram)

optimizedIEM = optimizeGuessedIEM(IEM, massChromatogram)

Generate predicted chromatogram using optimizedIEM and superimpose on massChromatogram
End
End

Procedure generateIonChromatogram(baseMass, z)

lowMass = baseMass; highMass = baseMass+numberOfPeptideForms/z;

Generate a compound ion chromatogram (CIC) with masses between lowMass and highMass and smooth it by applying Savitzky-Golay filter (3rd degree, 9 seconds window width, 2 cycles)
End Procedure

Procedure generateMassChromatogram(retentionTimeBounds,baseMass, z)

Average MS scans between retentionTimeBounds

lowMass = baseMass–5.0; highMass = baseMass+numberOfPeptideForms/z+5.0;

Display averaged mass chromatogram between lowMass and highMass

Return generated mass chromatogram
End Procedure

Procedure guessAnIEM(basePeptide, z, massChromatogram)

Generate predicted isotopic envelopes (PIEs) for all peptide forms (7 isotopic peaks each)

Figure 11. IEM Flowchart: Pseudo-code of algorithm for deamidation quantification using isotopic envelope modeling method.

Set fittingInterval as lowestMassInAllPIEs-2.0 and highestMassInAllPIEs+2.0

Smooth peaks in massChromatogram by applying Savitzky-Golay filter (3rd degree, 0.09 Da window width, 2 cycles) and find peaks above 5% of the baseline intensity

Match predicted isotopic peak positions to massChromatogram using centroided mass values

Determine peak position parameter ($\overline{X}$) by solving equation 2 using top 2 most intense matched peaks of predicted and measured isotopic envelope

Determine peak width parameter (σ) by solving equation 4 using top 2 most intense matched peaks in measured isotopic envelope

For each peptide form (j)

    Determine monoisotopic peak intensity from massChromatogram

    Subtract intensity contribution due to the isotopic peaks of prior forms and set the resultant intensity divided by 4 as I_j

End

determineBackground(massChromatogram, fittingInterval)

Generate a composite IEM function (shown in eq. 1) for all peptide forms using initial parameters

Return generated composite IEM

End Procedure

Procedure determineBackground(massChromatogram, fittingInterval)

    Slice the massChromatogram between fittingInterval into 0.5 Dalton wide bins

    Collect lowest intensity valley in each bin and fit a cubic polynomial through the set

End Procedure

Procedure optimizeGuessedIEM(IEM, massChromaotogram)

    Formulate a least squares function between IEM and raw massChromatogram within fittingInterval

    Fix higher order background parameters of the IEM

    Minimize the least squares function

    Get optimal parameters and reseed the IEM

    Release all the parameters and repeat the minimization procedure

    Return optimized IEM

End Procedure

Figure 11. IEM flowchart: Pseudo-code of algorithm for deamidation quantification using isotopic envelope modeling method.
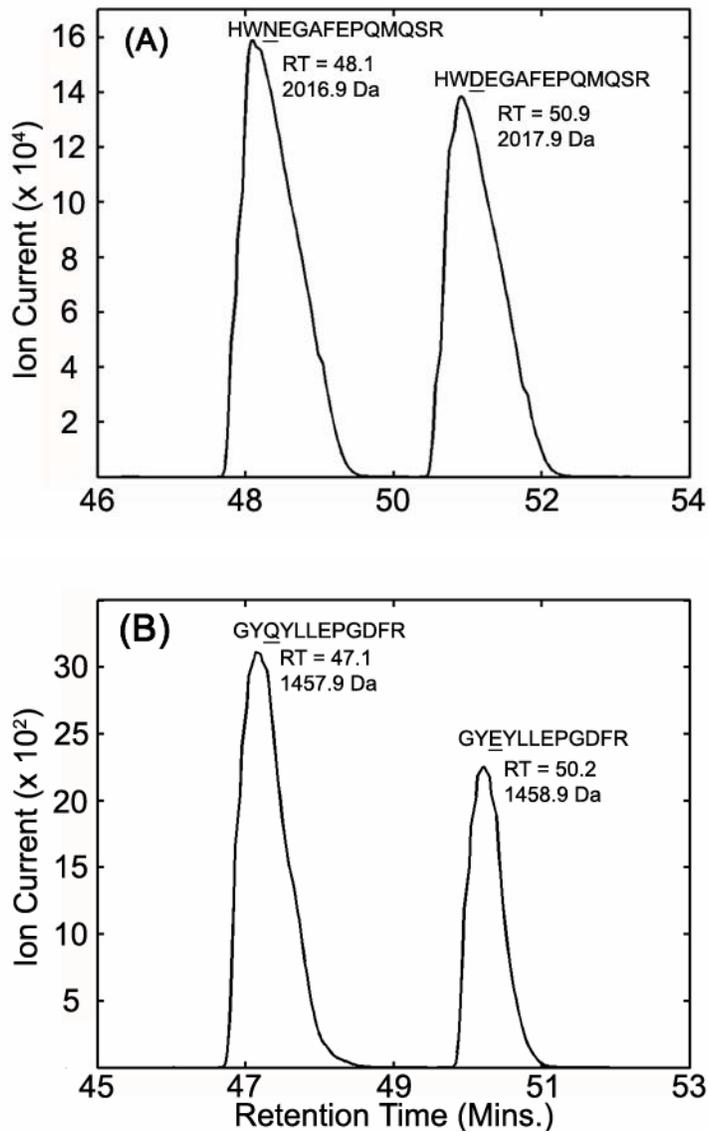
Figure 12. Synthetic Peptide Chromatograms on QTOF: Extracted ion chromatograms of (A) synthetic deamidated HW<u>D</u>EGAFEPQMQSR and unmodified (amidated) HW<u>N</u>EGAFEPQMQSR peptide mixtures and (B) GY<u>E</u>YLLEPGDFR and GY<u>Q</u>YLLEPGDFR peptide mixtures that were resolved by reverse-phase chromatography on a QTOF equipped with nanospray. The deamidated peptide eluted at 2.8 and 3.1 mins later than its corresponding amidated form.
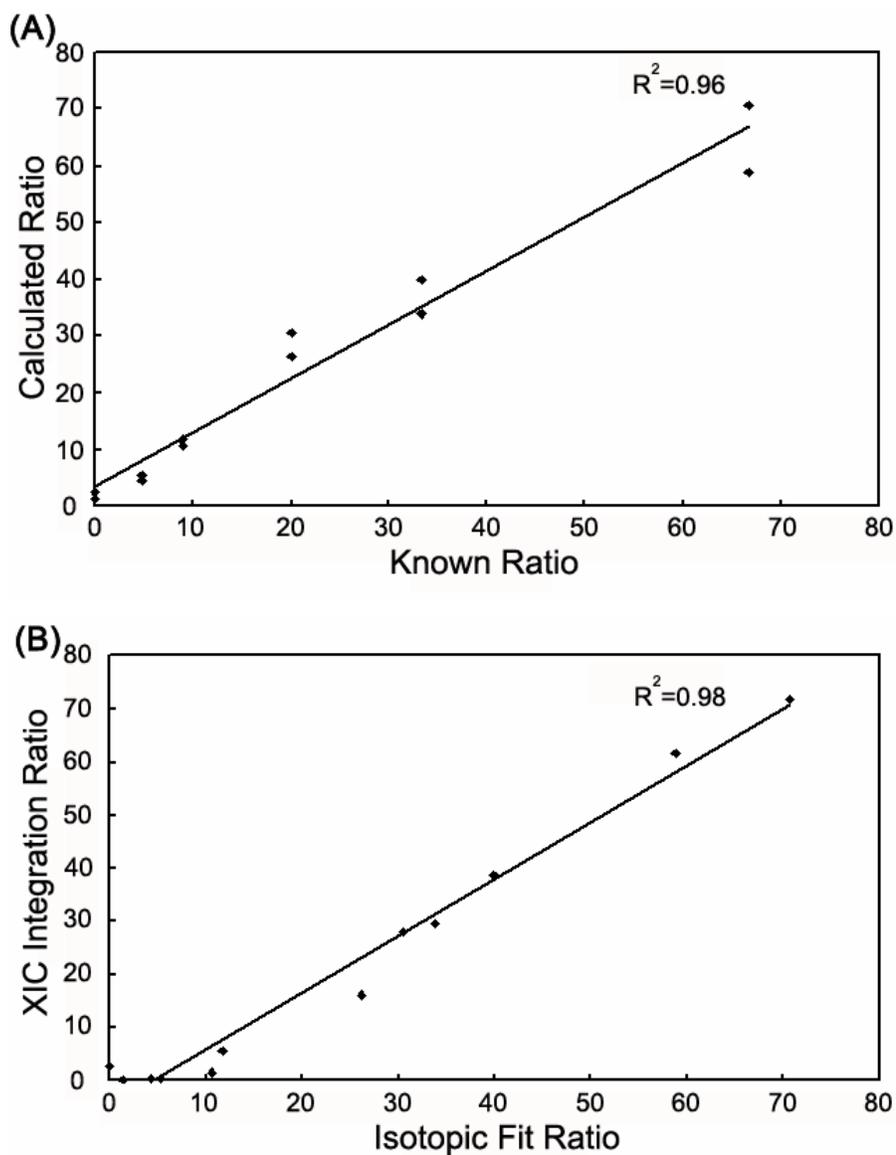
Figure 13. Equivalence of XIC and IEM Methods: Known ratios of amide:acid synthetic peptide mixtures were analyzed on LC-MS. Experimentally measured ratios were determined using isotopic envelope modeling (IEM) method and plotted against know ratios as shown in (A). Excellent correlation is observed between the known and calculated ratios at both high and low deamidation ratios. Deamidation ratios measured using IEM and extracted ion chromatogram extraction (XIC) methods are compared in (B). A high correlation is observed between the traditional XIC and IEM methods.
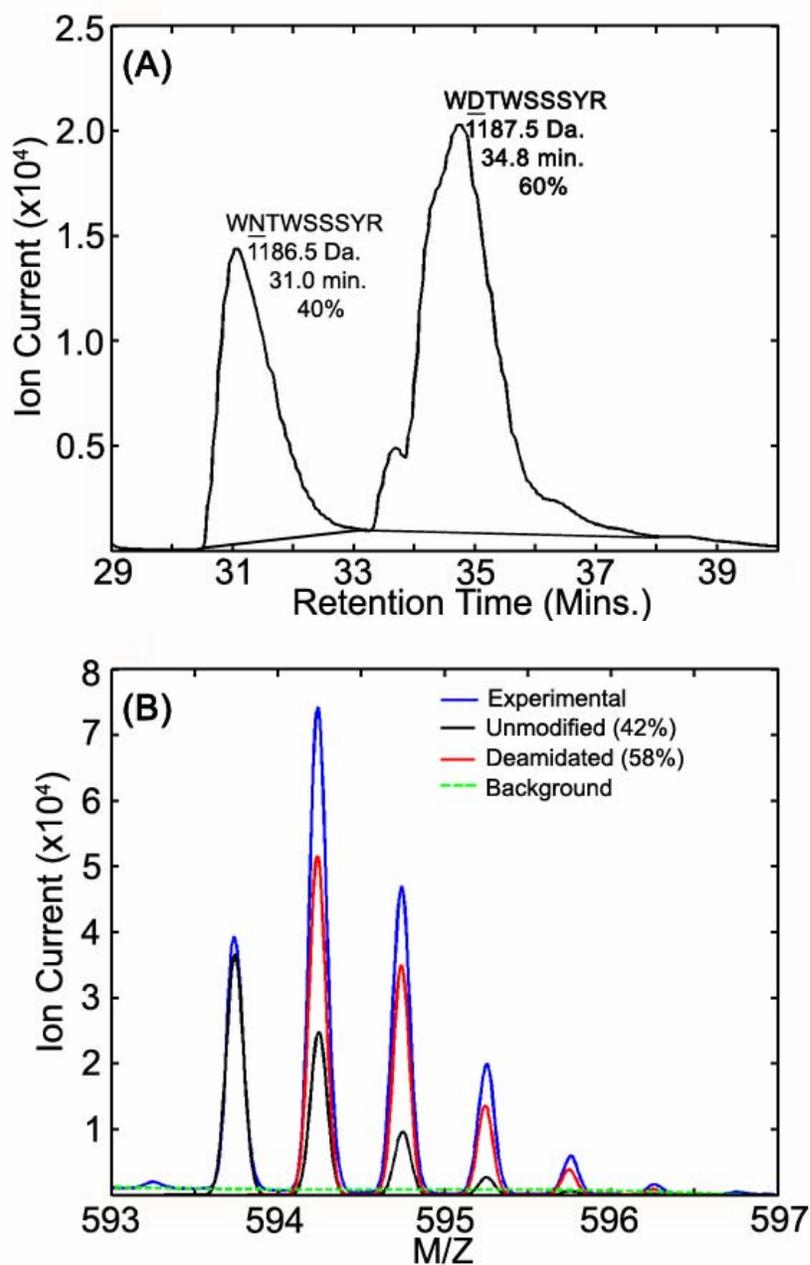
Figure 14: Comparison of IEM and XIC Integration Methods Using Peptides in Complex Mixtures: Compound ion trace of WNTWSSSYR (βB1 crystallin, 124-132) and its deamidated form is shown in Fig. 14a. Solid lines in 14a mark the peak areas used for integration. The combined m/z spectrum (shown in Fig. 14b) was obtained by averaging MS scans from 29.8 to 37.5 minutes. The resulting IEM fits for the unmodified peptide and the deamidated peptide isotopic envelopes are shown in Fig. 14b. The XIC integration and IEM results agreed very well (shown both in A and B as percentages).
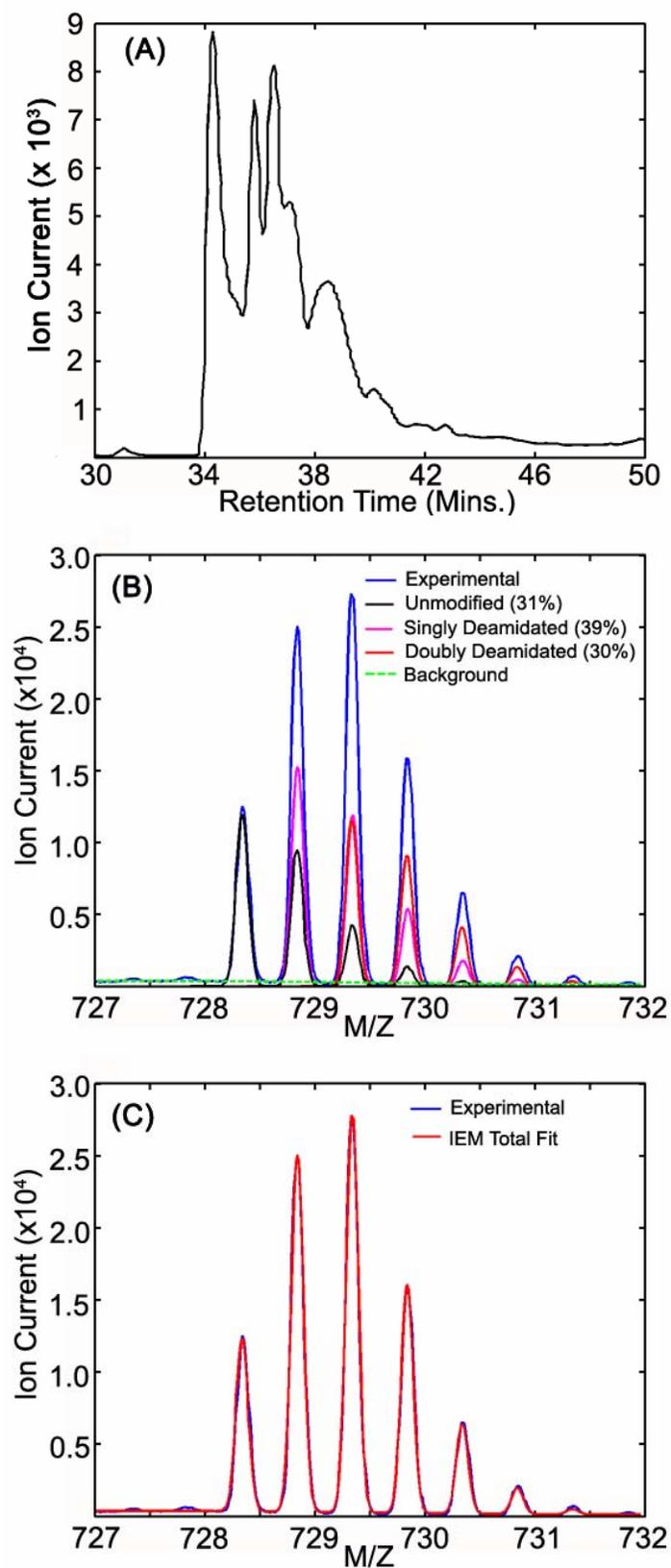
Figure 15. Advantages of IEM over XIC integration: Compound ion trace for peptide

ITIYDQENFQGK (αA1 33-44) from a 93-year old insoluble SCX fraction is shown in

Fig. 15a. Individual IEM fits for three peptide forms (unmodified, singly deamidated, and doubly deamidated) and their corresponding relative areas are shown in Fig. 15b. Composite IEM fit is overlaid on modeled experimental averaged mass chromatogram to show the overall goodness of fit in Fig. 15c.
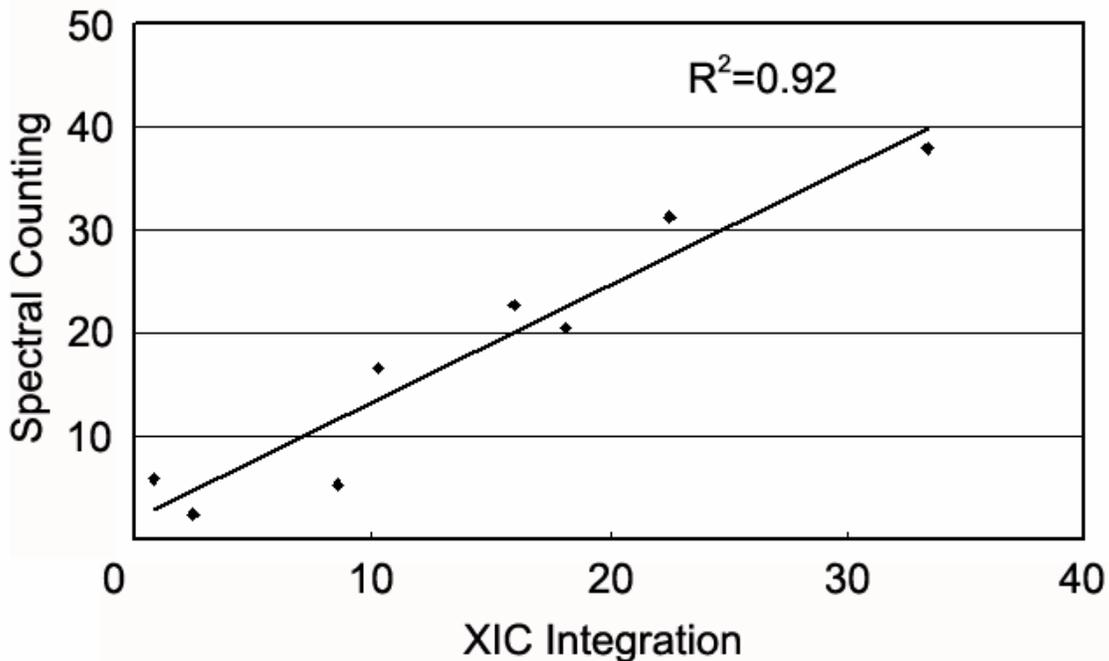
Figure 16. Spectral Counting vs. XIC Integration: Scatter plot of deamidation ratios determined by extracted ion chromatogram integration (XIC integration) and spectral counting of eight single amide containing peptides from 3 different samples.
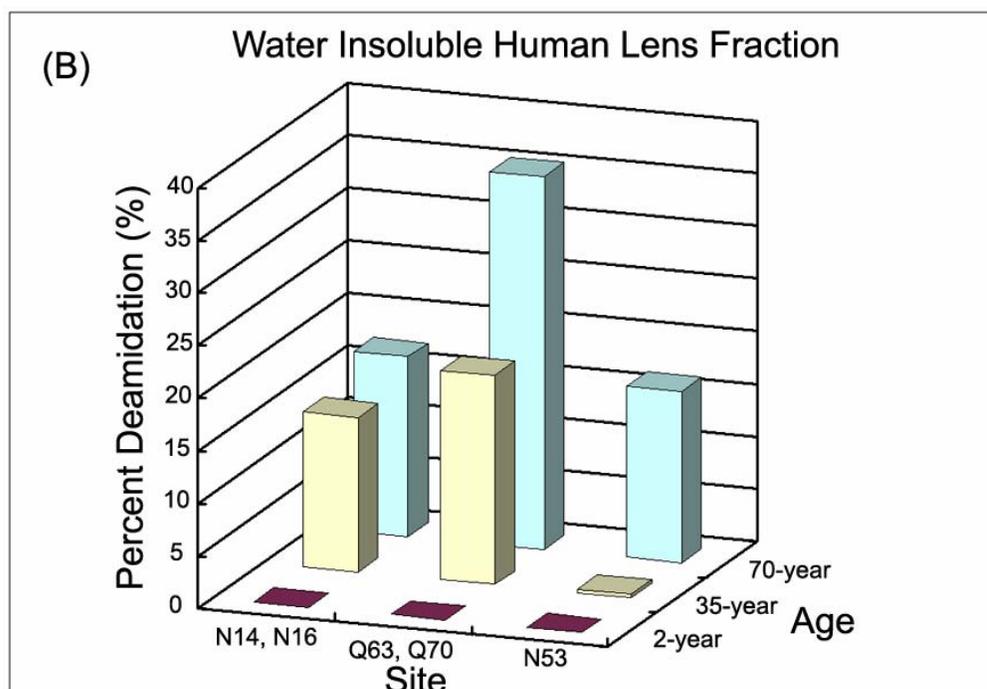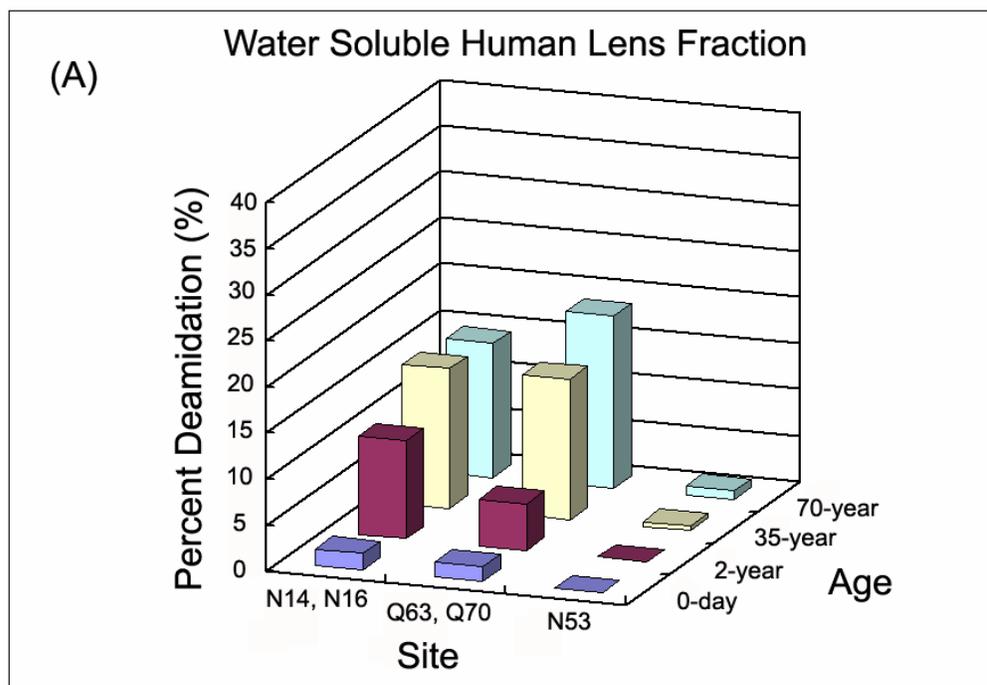
Figure 17. γS Crystallin Peptide in vivo Single Deamidation Rates: Single deamidation rates (%) of peptides from γS lens crystallin are calculated for water soluble (A) and water insoluble (B) fractions of an aged series of human lens samples. Computed rates in both water soluble (A) and water insoluble fraction (B) are plotted on same scale. [Q63, Q70] and N53 deamidation sites showed a pronounced deamidation increase in

water insoluble fractions of human lens compared to corresponding water soluble fractions.
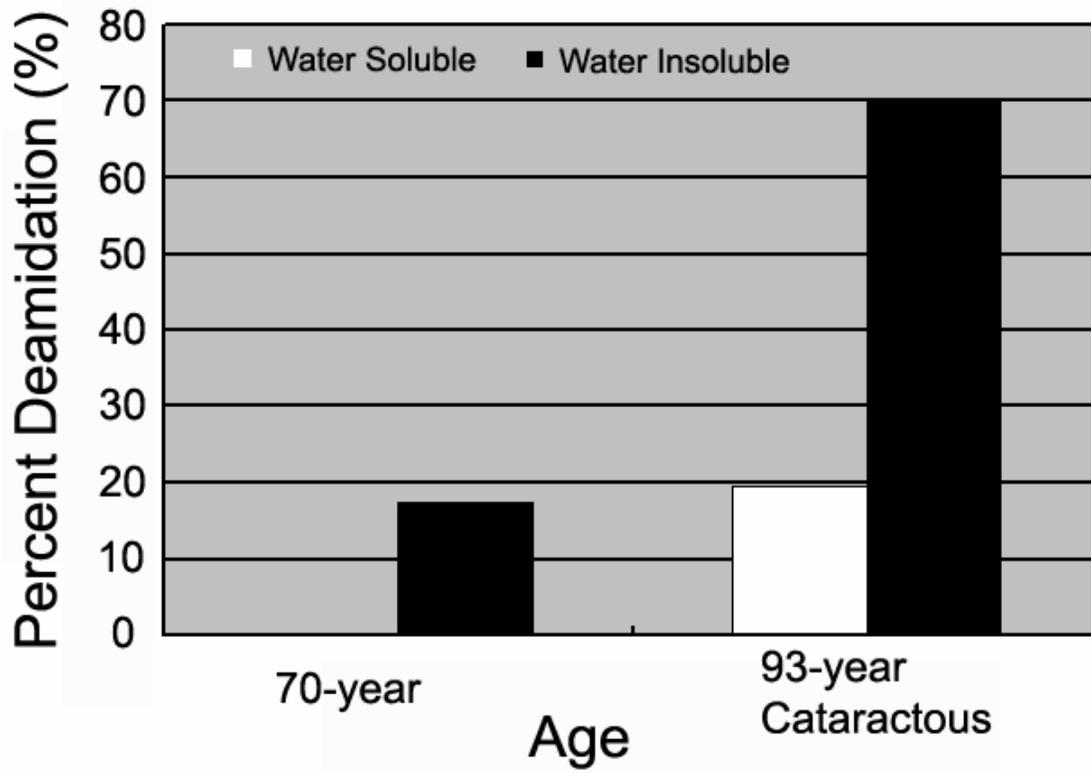
Figure 18. Deamidation of N143 site in γS Crystallin: Single deamidation rates of N143 site in γS lens crystallin are calculated for water soluble and water insoluble portions of a normal (70-year) and cataractous (93-year) lens samples. Insoluble portions of human lens samples showed an increase in deamidation at N143 site of γS crystallin. No deamidation could be detected for the same site in 3-day, 2-year, and 35-year samples.
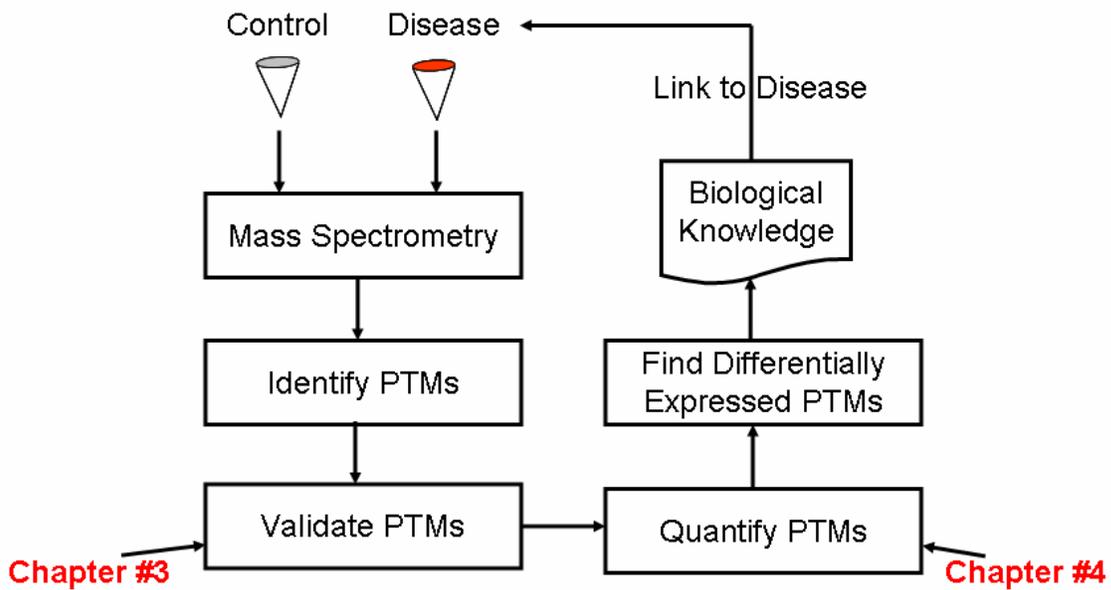
Figure 19. A Generalized Mass Spectrometry Based workflow for Characterization of PTMs that may Play a Role in Disease: The bioinformatics contributions to this workflow from this thesis are highlighted in red font

Appendix 1. MICIT Workflow Story Boards

MICIT (Panel 1)

Appendix 1.1. Loading Results: Panel 9.1 shows the MICIT after protein and peptide identifications are loaded. The protein list shows all protein identifications present in the sample in an expandable tree table.

MICIT (Panel 2)

Appendix 1.2. Viewing Protein and Peptide Identifications: A protein identification node, when expanded, displays all of its corresponding peptide identifications (Peptide List) as shown above. Plus one mass (MH+) and global retention time bounds (determined from MS2 identifications of the peptide) of each peptide are also displayed.

MICIT (Panel 3)
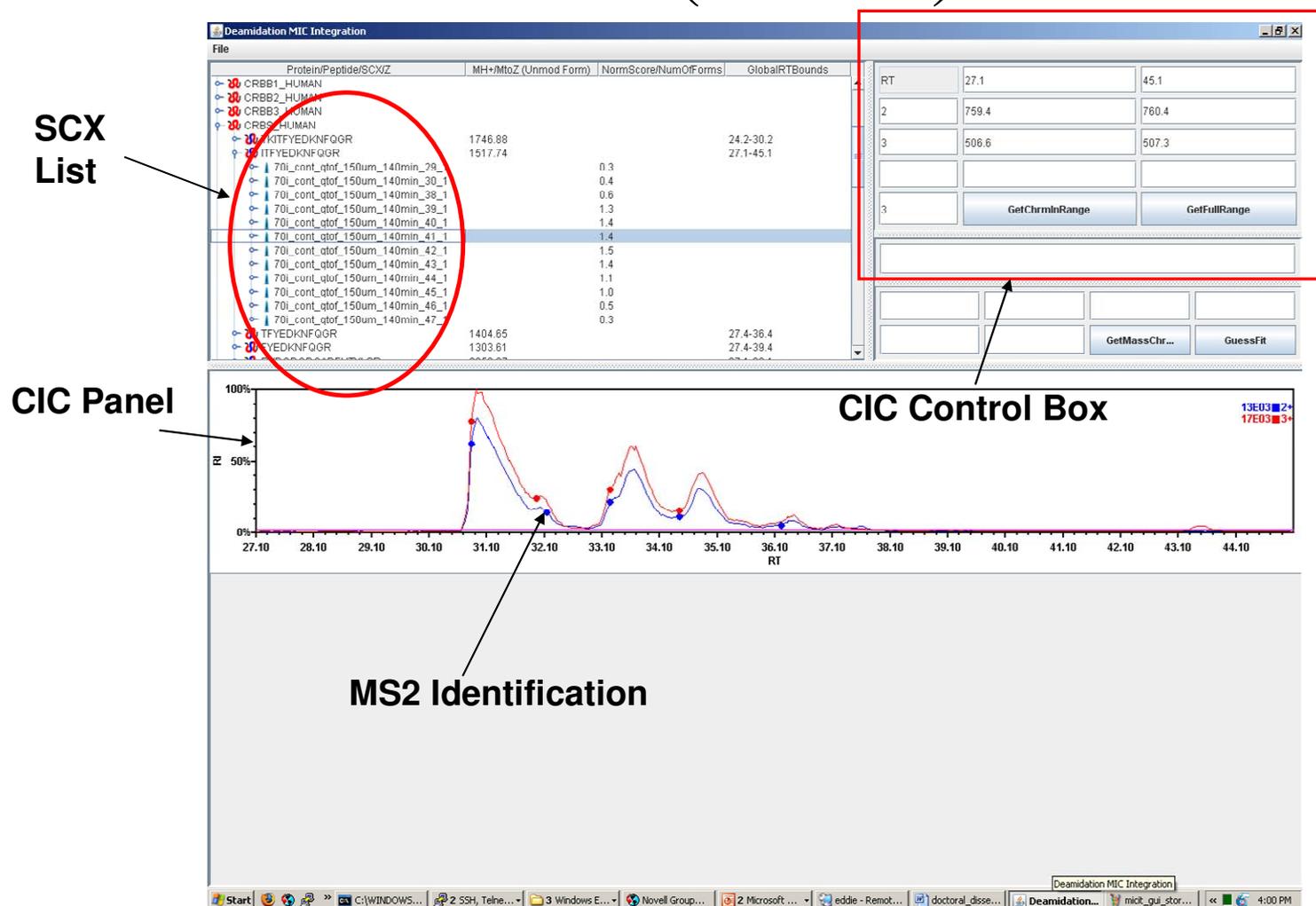
Appendix 1.3. Extracting Peptide Ion Chromatograms: A peptide identification node, when expanded, displays all SCX fractions (SCX List) in which corresponding peptide forms (unmodified and deamidated) have been detected. When a SCX fraction is selected, a compound ion chromatogram (CIC) for all peptide forms present in a particular SCX fraction is generated and displayed in CIC panel. Separate CICs are generated for all detected charge states in a SCX fraction and displayed using different colors (red color for 3+ and green color for 2+ as shown in above example). All MS2 identifications of selected peptide (unmodified and deamidated) are overlaid on corresponding CICs. The CIC control box allows control of CIC generation.

MICIT (Panel 4)

Appendix 1.4. Generating Mass Ion Chromatograms (MICs): Retention time integration bounds (RT Integration Bounds) for a SCX fraction are selected using a left button depressed mouse drag on corresponding combined chromatogram. Selected SCX fraction is expanded to display all detected charge states of the peptide (Z State List). Selection of a specific charge state would result in generation of an averaged mass ion chromatogram (between selected integration bounds).

MICIT (Panel 5)
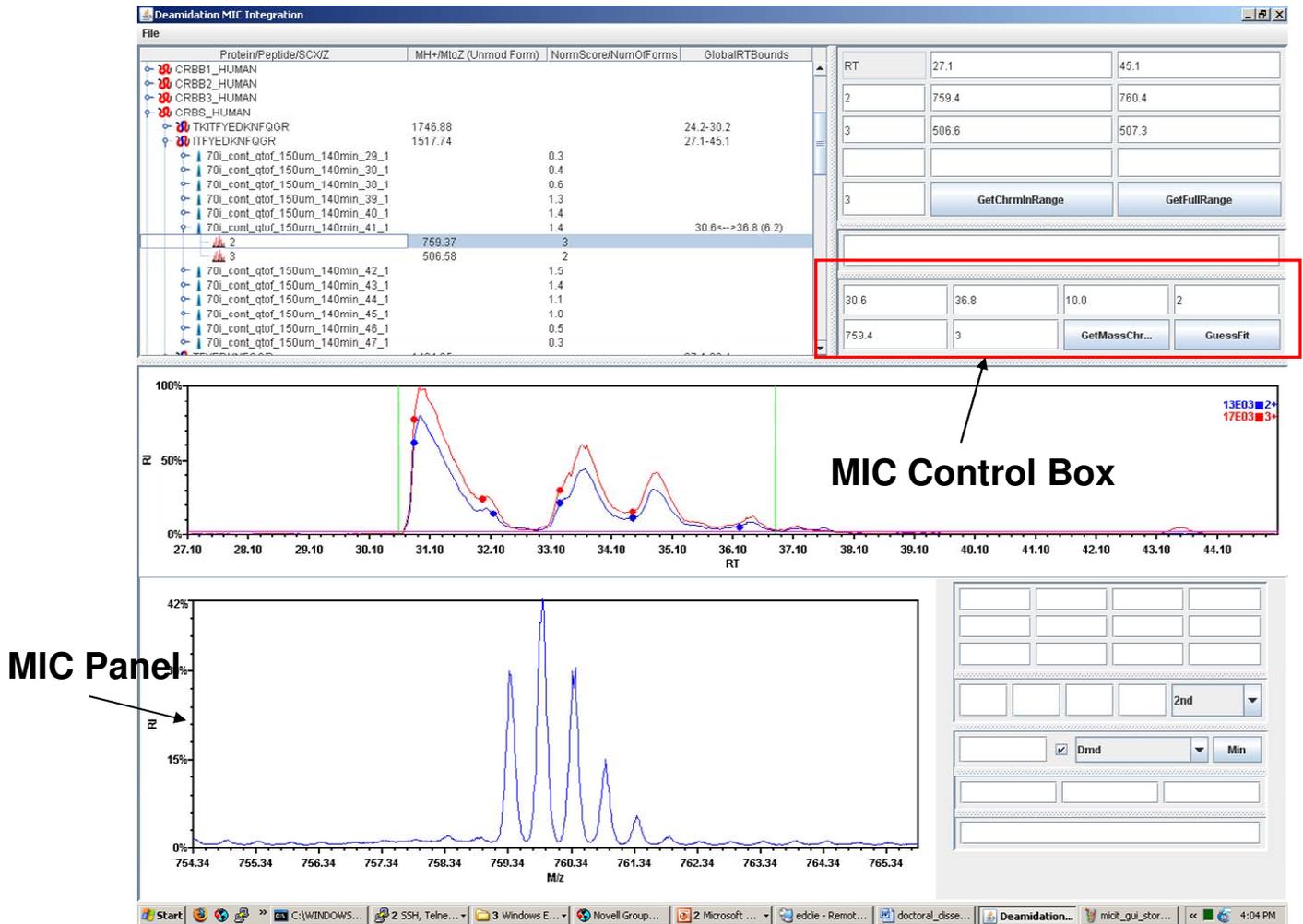
Appendix 1.5. Viewing Mass Ion Chromatogram: Averaged experimental mass ion chromatogram (MIC) for a peptide is displayed in MIC panel below combined ion chromatogram. The view port for MIC is automatically chosen to display ± 5.0 Dalton around the isotopic distribution of peptide of interest. Control of the MIC panel is available using the MIC control box.

MICIT (Panel 6)

Appendix 1.6. Initializing an Isotopic Envelope Model (IEM): After generation of an averaged experimental mass ion chromatogram (MIC), user produces an initial theoretical isotopic envelope model (IEM) by clicking on "GuessFit" button in MIC control panel. Generated IEM is overlaid on MIC in red color. The parameters of the guessed IEM (peptide intensities, peak shape, and background) are displayed in a control box next to MIC panel and are user adjustable.

MICIT (Panel 7)
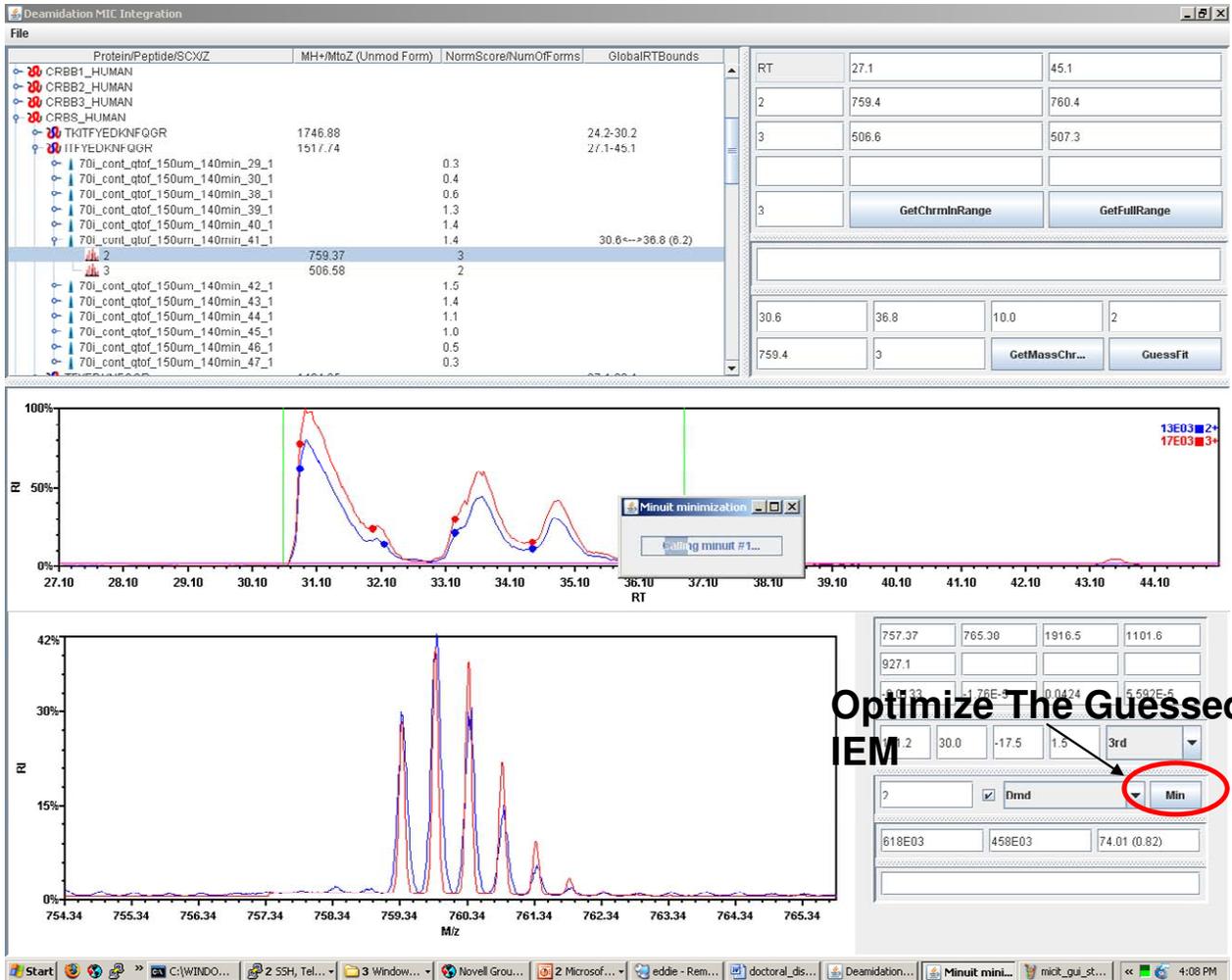
Appendix 1.7. Optimizing the Guessed IEM: The initial IEM is optimized by clicking on the "Min" button as shown in above panel. The optimization is carried out by Minuit function minimization package (version 1.7, CERN, Geneva, Switzerland) using a least squares approach. User is updated in real time about the status of the optimization procedure.

MICIT (Panel 8)
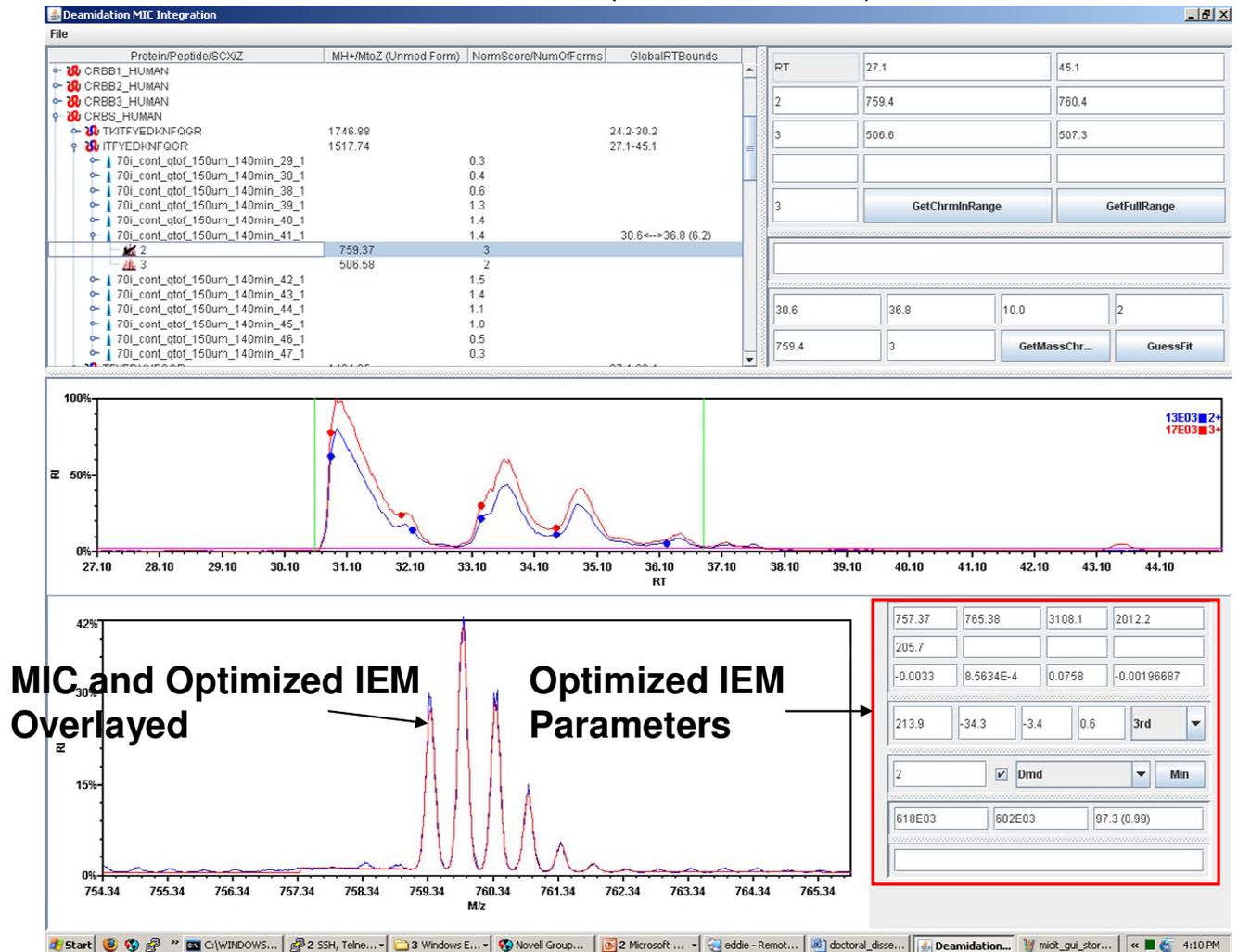
Appendix 1.8. MIC and Optimized IEM display: Optimized IEM is overlaid on top of the averaged experimental mass ion chromatogram (MIC) in the MIC panel. The guessed IEM parameters are updated with optimized IEM parameters. The optimized peptide intensity parameters are used to quantify different forms of the peptides present in the IEM. The result of the optimization can be saved to a XML formatted file in real time.