QD
96
.A2
W55
2008

ANALYSIS OF HUMAN SERUM PROTEIN EXTRACTS BY MATRIX ASSISTED LASER DESORPTION IONIZATION AND ELECTROSPRAY IONIZATION QUADRUPOLE-QUADRUPOLE TIME-OF-FLIGHT TANDEM MASS SPECTROMETRY

by

Gwillym Declan Williams, Bachelor of Science, Ryerson University, 2005

A thesis presented to Ryerson University

in partial fulfillment of the requirements for the degree of Master of Science

in the Program of Molecular Science

Toronto, Ontario, Canada, 2008.

© Gwillym Declan Williams 2008.

PROPERTY OF
RYERSON UNIVERSITY LIBRARY

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. I authorize Ryerson University to lend this thesis or dissertation to other institutions or individuals for the purpose of scholarly research.

Gwillym Declan Williams

I further authorize Ryerson University to reproduce this thesis or dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Gwillym Declan Williams

Analysis of Human Serum Protein Extracts by Matrix Assisted Laser Desorption Ionization and Electrospray Ionization Quadrupole-Quadrupole Time-of-Flight Tandem Mass Spectrometry.

Gwillym Declan Williams.

M.Sc. candidate in Molecular Science, Ryerson University, 2008.

ABSTRACT

Serum is a complex heterogeneous matrix containing endogenous proteins with potential diagnostic value. No previously reported methods have allowed the detection of all its constituent proteins from a single sample. In the work described in this thesis, normal human serum was separated by reversed-phase or dye affinity chromatography, ultrafiltration and by extraction under various pH and salt concentrations in aqueous and organic solutions. The serum extracts were analyzed by SDS-PAGE, MALDI and LC-ESI-Qq-TOF MS/MS without prior enzymatic digestion. Peptides observed as multiple isotopic peaks in MS mode were correlated with known serum proteins by ProteinPilot and X!Tandem analysis of their fragmentation spectra using the no enzyme [X|X] condition. Fragment ion masses within 0.04 Da of those predicted yielded high confidence identifications from both algorithms. Angiotensin I and [Glu1] fibrinopeptide B were clearly detected in serum spiked with as little as 250 and 100 femtomoles respectively. Post-translational modifications of serum peptides were identified using ProteinPilot.

ACKNOWLEDGEMENTS

I wish to thank my thesis supervisor, John Marshall, for his contribution to this work. John's energy, guidance, and patience throughout my graduate studies were invaluable to me. I consider myself privileged to have worked with him.

I gratefully acknowledge the contribution of the Ontario Cancer Biomarker Network (OCBN) for operating a grant to John Marshall.

This work was founded on the accumulated study of generations of researchers. An innumerable number of contributors have refined the analysis of biomolecules to its current sophisticated state. I am deeply thankful for the opportunity to apply the technology and information their work has provided.

TABLE OF CONTENTS

AUTHOR'S DECLARATION iii

ABSTRACT v

ACKNOWLEDGEMENTS vii

LIST OF TABLES x

LIST OF FIGURES xi

LIST OF APPENDICES xii

LIST OF ABBREVIATIONS xiii

INTRODUCTION 1
- Tandem Mass Spectrometry 2
- Ionization Technologies for the Mass Spectrometry of Non-Volatile Compounds 3
- Blood Protein Analysis by Mass Spectrometry 4
- Quadrupole-Quadrupole Time-of-Flight 5
- ProteinPilot and X!Tandem 6
- Polypeptide Ion Fragmentation 6
- Mass Spectrometry for Polypeptide Quantification 8
- Trypsin Digestion and MS/MS 10
- In and Ex -Vivo Blood Protein Degradation 11
- Separatory Techniques for the Polypeptides of Blood 11
- Serum Preparation and Analysis by Qq-TOF Based on Precipitation and Extraction . 12

EXPERIMENTAL 14
- Reagents and Materials 14
- Serum Fractionation by Chromatography,Ultrafiltration,Precipitation and Extraction 14
- Protease Inhibitor Treatment 15
- Dot Blot Protein Assay 16
- SDS-PAGE 17
- MALDI-Qq-TOF 19
- LC-ESI-Qq-TOF 19
- MS/MS Data Analysis Using ProteinPilot, X!Tandem and Relational Databases 20
- Internal Standards in Serum for MALDI Analysis 21

RESULTS 23
- Serum Peptide Extraction 23
- Acidic and Basic Extraction 24
- SDS-PAGE of Protein Extracts 24
- LC-ESI-MS 26
- Peptides Identified by X!Tandem and ProteinPilot from ESI-Qq-TOF Spectra 28

Peptides in the Acetonitrile Supernatant 30
Reverse Library Identifications 38
The Effect of X!Tandem Search Parameters on Spectral Assignment 39
Evaluation of Cleavage Sites in ProteinPilot Results 43
Modifications of MS/MS Analytes Detected by ProteinPilot and X!Tandem 44
Detection of Internal Standards 44
Effects of Protease Inhibitors on MALDI-MS 46

DISCUSSION 50
Extraction Parameters Influence Polypeptide Composition 50
The Effect of Salt on Polypeptide Solubility 51
The Effect of Solvents on Polypeptide Solubility 53
Surfactants and Non-Polar Solvents 54
Diagnostic Paradigms of Peptide Identification and Pattern Profiling Can be Unified 54
Combined ESI MS/MS and MALDI MS for the Characterization of Serum Peptides 56
X!Tandem and ProteinPilot Interpretation of Qq-TOF Tandem Mass Spectra 57
Apolipoprotein 58
Serum Protease inhibitors 59
Serum Proteins Identified from the Reverse Library 61
Evaluation of Reverse Search Results 61
MS/MS Search Parameters Influence Peptide Results 63
Modified Serum Peptides 64
MALDI-Qq-TOF Analysis of Serum Treated with Protease Inhibitors 67
Tryptic Peptides in Serum Extracts 70
Internal Standards in Untreated Serum Detected by MALDI-MS 71

CONCLUSIONS 73

APPENDIX 1: Supplemental Figures and Tables 75

BIBLIOGRAPHY 78

LIST OF TABLES

Table 1: Protein concentrations in serum preparations.....23

Table 2: Selected peptides of serum identified by X!Tandem.....31

Table 3: Selected peptides Identified by ProteinPilot with 99% confidence.....34

Table 4: X!Tandem results from ±100 PPM and +3-1 Da parent mass tolerance.....42

Table 5: ProteinPilot results compared by peptide cleavage site.....43

LIST OF FIGURES

Figure 1: Roepstorff–Fohlmann–Biemann nomenclature and fragment masses.....7

Figure 2: Serum polypeptide extraction and analysis scheme.....13

Figure 3: Protein assays and ImageJ intensity plots of serum protein concentration 18

Figure 4: Extracts of acetone and acetonitrile serum pellets separated by SDS-PAGE...25

Figure 5: Total ESI-Qq-TOF ion currents of reversed-phase C18 eluates.....27

Figure 6: Apolipoprotein peptides analyzed by nanoflow LC-ESI-MS/MS.....36

Figure 7: ESI-Qq-TOF tandem mass spectrum an albumin peptide.....37

Figure 8: ESI-Qq-TOF tandem mass spectrum of a fibrinogen alpha phosphopeptide....38

Figure 9: Internal standards in serum analyzed by MALDI-Qq-TOF-MS.....45

Figure 10: MALDI-MS of serum treated with protease inhibitors for 72 hours.....47

Figure 11: MALDI MS with ions corresponding to peptides identified by ESI.....48

Figure 12: Serum peptides identified by both MALDI and ESI-Qq-TOF MS/MS.....49

LIST OF APPENDICES

Appendix 1: Supplemental Figures and Tables.....75

LIST OF ABBREVIATIONS

AEBSF; 4-(2-Aminoethyl) benzenesulfonyl fluoride, BLAST; basic local alignment search tool, CBBR; coomassie brilliant blue R-250, CHCA; α-cyano-4-hydroxycinnamic acid, CID; Collisionally-induced dissociation, Da; Dalton, DEAE; diethylaminoethyl, ESI; electrospray ionization, FTICR; Fourier transform ion cyclotron resonance, FWHM; full width at half maximum, GPM; Global Proteome Machine, HUPO; Human Proteome Organization, ICAT; isotope-coded affinity tags, ITIH4; inter-alpha (globulin) inhibitor H4, MALDI; matrix-assisted laser desorption/ionization, MS; mass spectrometry, MRM; multiple reaction monitoring, MS/MS; tandem mass spectrometry, PBS; phosphate buffered saline, NCBI; National Center for Biotechnology Information, PPM; parts per million, PPP; Plasma Proteome Project, PMSF; phenylmethanesulfonyl fluoride, Qq-TOF; quadrupole-quadrupole time-of-flight, RPM; rotations per minute, Th; Thomson, Tris; tri(hydroxymethyl) aminomethane, SDS-PAGE; sodium dodecyl sulfate polyacrylamide electrophoresis, SILAC; stable isotope labeling by amino acids in cell culture, TFE; trifluoroethanol.

INTRODUCTION

Polypeptides can be rapidly analyzed by tandem mass spectrometry (MS/MS) following ionization by electrospray (ESI) or matrix-assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988; Fenn et al., 1989). For complex protein mixtures, electrophoresis, chromatography, affinity purification or a combination of these separatory techniques are typically employed prior to enzymatic digestion and MS/MS.

Peptides are identified from their tandem mass spectra using heuristic or probabilistic algorithms with *in silico* sequence libraries (Kapp et al., 2005). From known protein sequences and those derived from genetic data, correlation algorithms generate lists of masses corresponding to peptide ions and their complete fragmentation series. The search algorithms then compare predicted masses to experimental MS/MS data, reporting sequence data and the statistical relevance of the identifications. The distinct proteins from which the ions originate are thus determined.

Blood serum and plasma are challenging matrices for protein analysis by MS/MS. They contain known disease biomarkers and many biomarker candidates (Diamandis, 2007; Hortin, 2006; Leman et al., 2008; Han et al., 2008; Plumer et al., 2008) but the number of protein analytes they contain and chemical interference to separation and ionization hinder accurate and sensitive mass spectrometry. In addition to proteins and peptides serum and plasma contain lipids, nucleic acids, inorganic and small organic compounds. Extensive biological modification of serum polypeptides increases the complexity of their tandem mass spectra, necessitating the incorporation of modification-specific mass correlation into spectral analysis algorithms. Post-translational

modifications, many of which are dependent on physiological state, can abrogate signal and prevent high confidence peptide identification.

Tandem Mass Spectrometry

Structural complexity and diversity among biopolymers have spurred the creation of specialized mass spectrometry for their analysis, one of the most prevalent being MS/MS. In the absence of extremely accurate mass measurement, the identity of a heteropolymer, a peptide for instance, may be ambiguous because many monomeric mass combinations could result in its observed mass.

MS/MS begins with the isolation of a precursor from an ion mixture. The selected ion is then fragmented to produce product ions by one of a number of mechanisms. Collisionally-induced dissociation (CID), in which gas molecules are collided with the precursor ion is a common approach. Following dissociation a mass spectrum of the fragments is collected (McLafferty and Bockhoff, 1978).

With the exception of leucine and isoleucine, the mass of each amino acid is unique, so each fragment ion has a limited number of potential sequences. By reducing the number of amino acid combinations matching the ions of an observed fragmentation spectrum, MS/MS allows the determination of the entire peptide sequence.

Tandem mass spectra of polypeptides can be interrogated for sequence information manually or using automated algorithms, typically with volumes of genomic data for reference.

Ionization Technologies for the Mass Spectrometry of Non-Volatile Compounds

The low volatility of polypeptides has been a barrier to their analysis by MS, which separates gas phase ions based on their mass-to-charge ratios. Two technologies employed to circumvent this problem are MALDI and ESI.

Laser desorption was developed from the concept that a sufficiently rapid transfer of energy to a solid analyte will cause its sublimation before it can decompose (Fenn et al., 1989). High energy particles such as atoms and ions can be used for desorption but electromagnetic radiation yields superior efficiency for large analytes (Fenn et al., 1989). Photoionization can be achieved on a metal surface alone but MALDI employs an energy-absorbing matrix to improve ionization. Though no mechanistic model of desorption/ionization explains all MALDI phenomena, it is accepted that ionization results from interaction between matrix and polypeptide protons (Chang et al., 2007). Matrices such as α-cyano-4-hydroxycinnamic acid (CHCA) are commonly used because these tend not to form protein-matrix adducts which alter analyte mass (Beavis et al., 1992). The desorbing laser is pulsed creating discrete packets of mainly singly charged ions.

MALDI is not as sensitive to salt or detergents as ESI and can be used for proteins ranging in mass from the smallest to those much larger than 10 kDa, depending on the matrix used (Karas and Hillenkamp, 1988; Domon and Aebersold, 2006). The transient nature of the MALDI signal is well suited to TOF instruments which cannot trap and accumulate selected ions. Because laser desorption does not provide temporal resolution of the analytes, sample complexity must be low and offline separation is therefore essential for complex samples.

Electrospray ionization a direct interface between chromatographic systems and the mass analyzer and provides a continuous signal. ESI incorporates a metal micro or nanobore capillary tip through which the dissolved analyte is passed. The tip is maintained at high electric potential in relation to the inlet of the mass spectrometer. This potential, which is maintained in the kV range, charges liquid droplets emerging from the tip and directs them to the inlet. High temperature and a jacket of inert drying gas reduce droplet diameter through evaporation. Decreasing droplet diameter increases charge density and destabilizes the liquid particles, causing their dissociation through the process of Coulomb fission (Cole, 2000). A succession of these collapses eventually liberates the analyte to the gas phase, with ionization caused by the initial droplet charge (Fenn et al., 1989; Cole, 2000). Electrospray yields multiply charged ions. The charge of each analyte must therefore be determined by the relative masses of its isotopic envelope, requiring mass resolution greater than one Thomson.

The contrasting properties of ESI and MALDI sources explain why they are often used in concert and many instrument configurations have been adapted to both.

Blood Protein Analysis by Mass Spectrometry

Early examinations of blood using MALDI-Time-of-flight (TOF) demonstrated the simultaneous detection of multiple known blood polypeptides in a single sample (Oleschuck et al., 2000). Studies of peptide and protein distribution in serum using MALDI-TOF (Hortin, 2006; Tammen et al., 2005; Marshall et al., 2004; Richter et al., 1999; Merrell et al., 2004; Yi et al., 2007), ESI-quadrupole TOF (Q-TOF) (Yi et al., 2007), ESI-Ion trap (Zheng et al., 2006; Chan et al., 2004; Tian et al., 2007; Adkins et al.,

2002; Tirumalai et al., 2003; Shen et al., 2005; Govoukhina et al., 2006) and Fourier transform ion cyclotron resonance (FTICR) MS (Palmblad et al., 2000) have been reported.

Quadrupole-Quadrupole Time-of-Flight

Hybrid instruments incorporating two successive quadrupoles followed by an orthogonally oriented time-of-flight mass analyzer were shown to be well suited for peptide analysis, requiring minimal sample preparation and yielding high quality spectra (Verhaert et al., 2001). Qq-TOF-MS/MS operation entails precursor ion isolation in the first quadrupole, fragmentation in the second and mass-to-charge measurement in the flight tube (Morris et al., 1996). The TOF tube can include an ion detector at each end and an ion mirror or reflectron, which reflects ions to the detector adjacent to the tube entrance. Both the orthogonal orientation and the reflectron improve resolution of the time-of-flight analyzer considerably.

Qq-TOF is compatible with both MALDI and ESI sources and provides mass resolution around 10000 full width at half maximum (FWHM) peak height and low PPM mass accuracy (Shevchenko et al., 1997; Loboda et al., 2000). Qq-TOF can attain sub-femtomole and low femtomole sensitivity in single MS and MS/MS modes respectively (Loboda et al., 2000). The high confidence spectral assignment accompanying high quality Qq-TOF data make it ideal for the study of non-tryptic serum peptides.

ProteinPilot and X!Tandem

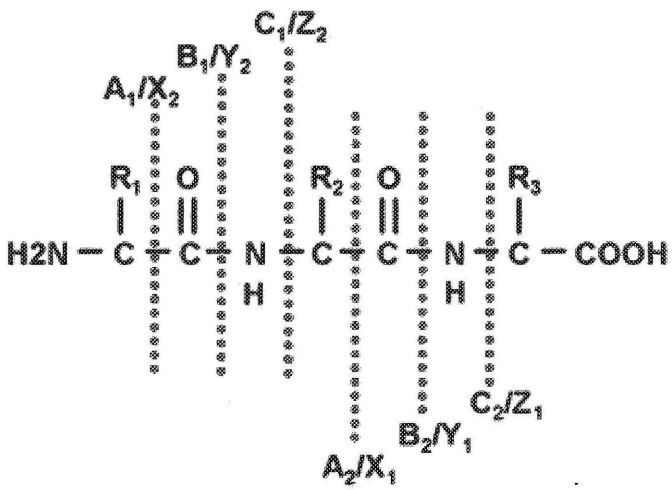
A number of MS/MS peptide correlation programs are available (Kapp et al., 2005). Many such programs are compatible with the FASTA file format developed for nucleic acid and protein sequence alignment (Pearson and Lipman, 1988). X!Tandem is an open source heuristic algorithm for the interrogation of proteomic MS/MS data against FASTA libraries (Craig and Beavis, 2004), available online from the Global Proteome Machine (GPM) Organization at *www.thegpm.org*. Rather than calculating amino acid combinations that fit an observed m/z, X!Tandem compares observed ions to calculated m/z values for every possible fragment sequence in the library. Thus genomic context is inherent in X!Tandem results.

ProteinPilot is a proprietary application from Applied Biosystems/MDS Sciex which employs the Paragon and Pro Group algorithms for spectral peak assignment and screening of the top peptide hypotheses respectively (Shilov et al., 2007). The ProteinPilot search strategy combines *de novo* sequence determination and FASTA library fragment comparison to exclude unlikely candidate peptides from scoring. Other novel discriminatory routines incorporated into the ProteinPilot software are described in publication (Shilov et al., 2007).

Polypeptide Ion Fragmentation

During MS/MS the most common peptide bond fragmentation mechanism results in cleavage of the N-terminal carbonyl carbon and the C-terminal nitrogen, producing one of two ions respectively termed the b and y ions in the standard Roepstorff–

Fohlmann–Biemann nomenclature. Two more peptide bond fragmentations occur, with accompanying differences in fragment mass (Steen and Mann, 2004). These fragmentations are shown in figure 1. The a/x and c/z series result from cleavage immediately N-terminal and C-terminal to the peptide bond carbonyl carbon respectively (Fig 1). Many spectra assigned to peptides contain peaks that cannot be correlated to a, b or y series ion. These peaks can correspond to chemically modified ions or b and y series ions beyond the mass accuracy tolerance defined in the search parameters. Fragment ions from series not discriminated by the algorithm also may be present in the spectra. Search algorithms



Ion	Relative Mass (Da)
A	B - 27.9949/C - 43.0098
B	A + 27.9949/C - 15.0149
C	B + 15.0149/A + 43.0098
X	Y + 27.9949/Z + 43.0098
Y	X - 27.9949/Z + 15.0149
Z	Y - 15.0149/X - 43.0098

Figure 1. Roepstorff–Fohlmann–Biemann nomenclature and relative masses of peptide bond fragmentations in tandem mass spectrometry. The bonds broken are indicated by dotted lines.

often compare spectral data to predicted b and y series without scrutiny of alternate ion series for tryptic peptides ending in charged, basic residues. The contributions of the a/x and c/z series to non-tryptic peptide fragmentation are undetermined. The X!Tandem

algorithm can be set to identify all three ion series. ESI and MALDI-Qq-TOF data from serum preparations was searched separately for b/y and for all ions to evaluate the contribution of a, c, x and z series.

Mass Spectrometry for Polypeptide Quantification

Quantitative MS-based approaches for complex protein mixtures include the measurement of peak intensity or peak area, addition of internal standards, multiple reaction monitoring, and isobaric labeling (Andersen et al., 2003; Gygi et al., 1999; Malik et al., 2005; Chelius and Bondarenko, 2002; Cutillas et al., 2005; Domon and Aebersold, 2006; Wu et al., 2006).

Peak area can be used to measure peptide abundance with femtomolar sensitivity and accuracy can be improved with a correction factor calculated from internal standard measurement (Chelius and Bondarenko, 2002). The dynamic intensity of MALDI signals and LC ion suppression can introduce error into peak measurement, necessitating more complex quantitative methods (Fenyö and Beavis, 2008).

The combination of ESI. ion-trap and MS/MS allows the study of experimentally defined ions by multiple reaction monitoring (MRM). In this system precursor ions are isolated and dissociated with a single fragment ion identified (Domon and Aebersold, 2006).

Isotope-coded affinity tags (ICAT) allow comparison of analyte concentrations between sample groups (Gygi et al., 1999). The ICAT system involves labeling two or more protein samples at cysteine residues, using chemically identical tags with unique isotope-dependent masses for each sample. The tag incorporates biotin for affinity

capture allowing enrichment of the labeled compounds. Following labeling, the samples are combined and analyzed by MS. The analytes are distinguished by their unique isotope-coded masses and quantified by their relative peak characteristics. Isobaric labeling systems suffer decreased accuracy with samples as complex as serum (Griffin et al., 2001; Wu et al., 2006). In such matrices, stably expressed proteins can conceal those differing between samples. This is particularly limiting to automated LC-ESI-MS/MS experiments, but can be controlled by manual peak selection with MALDI-Qq-TOF (Griffin et al., 2001). ICAT is semiquantitative and dependent on the presence of cysteine residues. In contrast, isobaric tags for relative and absolute quantification (iTRAQ) label lysine residues and N-termini, and are therefore compatible with both non-tryptic and trypsin digested peptides. Despite improved sensitivity and broader peptide compatibility compared to ICAT, iTRAQ has similar limitations to cysteine tagging (DeSousa et al., 2005, Wu et al., 2006). The cost and lengthy preparation associated with isotope labeling are also problematic.

Stable isotope labeling by amino acids in cell culture (SILAC) is another quantitative method that leverages the difference in isotopic masses for MS analysis (Ong et al., 2002). In SILAC, cells are grown under two conditions in media supplemented with amino acids containing heavy or light isotopes and MS analysis is performed as in ICAT and iTRAQ. While useful for many proteomic applications, this technique is not compatible with serum analysis because labeling takes place during translation of RNA into protein.

Internal standard measurement is a simple, rapid and proven quantitative method shown to be applicable to MS/MS of tissue samples including serum (Cuttilas et al.,

2005, Findeisen et al., 2007, 2008). Heights and areas of internal standard peaks identified by their known m/z are related to their known concentration over a linear range. The standard signal can be compared to peptide ion signals for quantification. Sensitivity depends on the instrument and must be measured for each matrix.

Trypsin Digestion and MS/MS

Trypsin is a serine protease that cleaves specifically C-terminal to arginine and lysine residues unless these are followed by proline (Olsen et al., 2004). Trypsin digestion is commonly employed prior to MS analysis to fragment proteins into smaller, more easily ionized peptides. Provided that complete digestion is achieved, this treatment also simplifies the searching of MS/MS data against sequence libraries as fully tryptic sequences represent a small number of all possible fragments. While trypsin cleavage is highly specific, digestion efficiency is dependent on digestion conditions, substrate stability and local tertiary structure even after denaturation (Olsen et al., 2004, Strader et al., 2006, Fenyö and Beavis, 2008). Searching for the spectra of tryptic peptides alone might cause one to reject valid MS/MS data. It is reasonable to assume that searching a tryptic dataset for peptides with any amino acid at the C-terminus and N-terminus will yield the same tryptic peptide identifications and also any non-tryptic or semi-tryptic peptides but this has not been demonstrated.

In and Ex -Vivo Blood Protein Degradation

Serum contains the products of complex degradative cascades, some of which result from sample handling, which eventually reach a steady state (Marshall et al., 2003, Villanueva et al., 2006, Koomen et al., 2005). This process is mediated by endogenous proteases, a group of enzymes involved in biological functions including coagulation cascade in blood, inflammation and the complement system (Reid, 1983, Davie et al., 1979). Peptides generated during biological responses are compelling diagnostic targets but their distribution in blood from healthy and diseased populations remain poorly understood. Pronounced increases in specific serum peptides resulting from inflammatory processes can be monitored by MS (Marshall et al., 2003). Coagulation increases peptide concentration in serum while plasma, which is not coagulated during production, has relatively lower peptide abundance (Yi et al., 2007). Low abundance serum peptide biomarkers may be obscured or destroyed during sample collection and preparation.

Separatory Techniques for the Polypeptides of Blood

Serum polypeptides are commonly separated by sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE), solid phase extraction or chromatography prior to analysis. Two dimensional SDS-PAGE provides relative purity for MS analysis but excision and extraction of the analytes is inefficient and make large scale analysis impractical. One and two dimensional LC-ion trap MS/MS have been applied to serum analysis resulting in the identification of over 800 proteins from multiple runs (Shen et al., 2004). Such studies show that powerful separatory techniques capable of resolving

thousands of distinct proteins across a wide mass range and pg/ml to mg/ml concentration range cannot produce more than a small fraction of the predicted number of serum proteins. It is possible that no single sample preparation technique will enable the detection of all serum proteins by mass spectrometry. Sequential separation combining chromatography and other methods may enhance resolution of low abundance and physiologically significant serum components.

Batch chromatography with C18 reversed-phase is exclusively required prior to MALDI or ESI analysis because it removes salt that impedes ionization and instrument performance. When used alone, C18 microcolumns are non-selective and therefore provide robust detection of many known serum analytes, but less abundant species require further separation (Tiss et al., 2007). Solid phase extraction reagents with selectivity for plasma peptides have been demonstrated (Li et al., 2007). Serum polypeptides can also be precipitated with salts or organic solvents, with some remaining in solution. Though the supernatant can be rapidly analyzed by MS/MS, it is depleted in protein analytes relative to the pellet (Williams et al., 2006, Tucholska et al., 2007, Chertov et al., 2004). Organic precipitates of serum protein and peptides can be partially or totally dissolved in aqueous solvent mixtures. Partial extraction may differentially dissolve and concentrate specific proteins, enabling their rapid determination by mass spectrometry.

Serum Preparation and Analysis by Qq-TOF Based on Precipitation and Extraction

The objectives of this thesis were to separate serum peptides by precipitation and extraction and identify them with high confidence. A simple preparatory technique for

the separation of serum polypeptides that is compatible with both MALDI and LC-ESI-Qq-TOF-MS/MS is presented here (fig 2). The Qq-TOF mass analyzer detected peptides from known serum proteins with high resolution and signal-to-noise ratios. Tandem mass spectra were subjected to peptide searches unrestricted in sequence using the no enzyme condition. Multiple peptides from known serum proteins were assigned by X!Tandem and ProteinPilot from data of all extractions. This method apparently identified post-translational modifications in serum peptide analytes. The role of native serum proteases in serum peptide mass distribution was examined by MALDI-Qq-TOF using a group of protease inhibitors. The sensitivity of Qq-TOF for serum analysis was assessed using internal standards in the femtomole range.

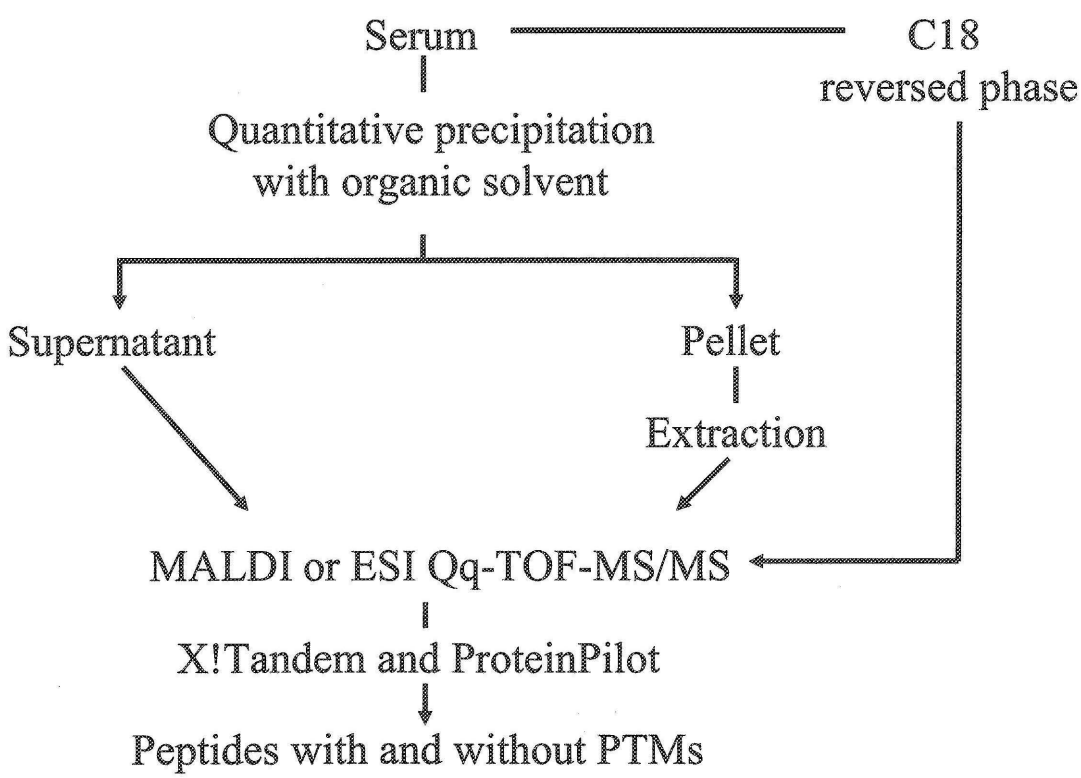


Figure 2. Serum polypeptide extraction and analysis scheme employed in this study.

EXPERIMENTAL

Reagents and Materials

All solvents used were of HPLC grade. Acetone, acetonitrile, 50% glutaraldehyde, methanol, water and reagent grade HCl were obtained from Caledon (Georgetown, ON, Canada). NaCl, NaOH and PBS tablets, and all reagents for the preparation of SDS-PAGE gels were from Bioshop (Burlington, ON, Canada). Angiotensin 1, formic acid, [Glu^1] fibrinopeptide B, trifluoroethanol (TFE) and protein molecular weight standards for SDS-PAGE were obtained from Sigma-Aldrich (St. Louis, MO, USA). Silver nitrate from BDH was obtained from VWR (West Chester, PA, USA). Normal human serum was obtained from Gemini Bio-Products (West Sacramento, CA, USA). Size exclusion filters and C18 reversed-phase microcolumns were from Millipore (Billerica, MA, USA) DEAE-Blue and Cibacron Blue were from BioRad (Hercules CA, USA).

Serum Fractionation by Chromatography, Ultrafiltration, Precipitation and Extraction

The general sample preparation strategy for the separation and enrichment of serum peptides is illustrated in figure 2. Serum was precipitated on ice in five volumes of acetonitrile or acetone with or without NaOH or acid and the precipitate was concentrated by centrifugation at 14000 RPM. The supernatants were dried in a vacuum concentrator (Labcono, Kansas City, MO, USA) prior to analysis. Acetonitrile pellets were mixed

with four volumes of solvent and the remaining solid isolated by centrifugation. The extraction solutions included water, acetonitrile and methanol with or without formic acid, TFE, NaCl, or NaOH. For MS/MS analysis, formic acid and TFE concentrations in the extracts were 5% (v/v). Serum precipitated with acetonitrile, methanol, ethanol or isopropanol was extracted in phosphate buffered saline (PBS). Extract protein concentration was determined by colorimetric protein assay (Ghosh et al., 1987). Supernatants from precipitation and extractions were dried, resuspended in 0.1% formic acid and purified over C18 reversed-phase prior to MALDI or LC-ESI-Qq-TOF-MS/MS.

Dye-affinity partition chromatography was performed in a preparative batch (Marshall et al., 2003, Williams et al., 2006). Serum was diluted in five volumes of loading buffer prior to application. Cibacron Blue and DEAE Blue columns at least 100 microlitres in volume were equilibrated and loaded then washed in five to ten column volumes of 20mM Tris with 40mM NaCl at pH 8.0 or with PBS pH 7.3. The columns were at least 10 times the volume of the serum. Bound protein was eluted in 1 M NaCl. For ultrafiltration, serum was diluted in 20% (v/v) acetonitrile with 100 mM ammonium bicarbonate at pH 8.3. The protein solutions were applied to wet 50 kDa molecular mass filters and centrifuged at 14000 RPM.

Protease Inhibitor Treatment

Serum was divided into 1.5 ml volumes and treated with a range of protease inhibitors, then reacted by incubation at room temperature for 72 hours. Untreated aliquots were subjected to the same incubation conditions or stored at -20°C for the incubation period. The concentrations of protease inhibitors were 250 microgram/ml 4-

(2-Aminoethyl) benzenesulfonyl fluoride (AEBSF), 50 microgram/ml bestatin, 1.0 mg/ml benzamidine, 1.0 mg/ml and saturated phenylmethanesulfonyl fluoride (PMSF), 2 mg/ml potato tuber carboxypeptidase inhibitor, 100 microgram/ml soybean trypsin inhibitor, and 3.75 microlitre/ml of a commercial protease inhibitor mixture (Sigma-Aldrich, St. Louis, MO, USA). The protease inhibitor mixture contained AEBSF, aprotinin, bestatin, E-64, leupeptin, pepstatinA. Volumes of 200 μL were withdrawn at regular intervals, acidified with formic acid and purified by C18 reversed-phase and spotted onto a MALDI target plate, prior to addition of the matrix material CHCA.

Dot Blot Protein Assay

Total protein concentrations in normal human serum and serum extracts were measured using a standardized colorimetric assay with densitometry (Ghosh et al., 1987). Samples were dissolved in 0.1% NaOH containing 1% SDS (w/v) or in 1% SDS (w/v) and 1 microlitre volumes were blotted in duplicate or triplicate onto 15.0 cm diameter Whatman # 1 filter papers (Whatman International, Maidstone, UK). A dilution series of bovine serum albumin in 1% SDS from 0.1 to 15 μg/μL was also spotted in 1 microlitre volumes onto each paper to provide a standard curve. The blots were allowed to dry completely before staining with 0.2% (w/v) Coomassie brilliant blue R-250 (CBBR) in 50% methanol. Assays were dehydrated in 99% methanol and de-stained in a 2:6:1 solution of methanol, MilliQ water and acetic acid respectively and air dried. Bovine serum albumin (minimum 98%) was obtained from Sigma-Aldrich.

Digital images of all dot blot assays were acquired with an Epson Expression 1680 flatbed scanner using the manufacturer's software. Images were recorded in RBG

colour mode at a resolution of 600 dpi and stored in high resolution format (.tif file extension).

To eliminate noise and simplify the extraction of data, the images were processed prior to analysis. Adobe Photoshop 7.0 was used to invert the image colour scale, making the background appear uniformly black and the Coomassie blue-stained blots to appear varying shades of yellow to white with intensity proportional to their protein concentrations. Following colour inversion, color intensity data was extracted from the images using ImageJ version 1.38x, an open source Java based image processing application publicly available from the National Institute of Health, USA at http://rsb.info.nih.gov/ij/. To minimize bias in the data, segmented lines from which intensity data was collected had no more than one point of inflection. For each separate assay paper, a standard curve was generated from the relative intensities of BSA standards. The R^2 values for the standard curves all exceeded 0.97. For each blot, two single points representing Gray Value maxima were used to calculate the protein concentration using the equation of the standard curve (Fig 3).

SDS-PAGE

Prior to separation by SDS-PAGE, samples were dried in a vacuum concentrator and resuspended to equal volumes in buffer containing tricine and DTT, then boiled for 5 minutes. The samples were applied to gels containing 4% acrylamide at 30 V then separated at a constant potential of 90 V on 17% acrylamide gels (Schagger and von Jagow, 1987). Polypeptides were visualized using CBBR in 40% methanol with 10% acetic acid or by diamine silver stain (Dionne et al., 1994).

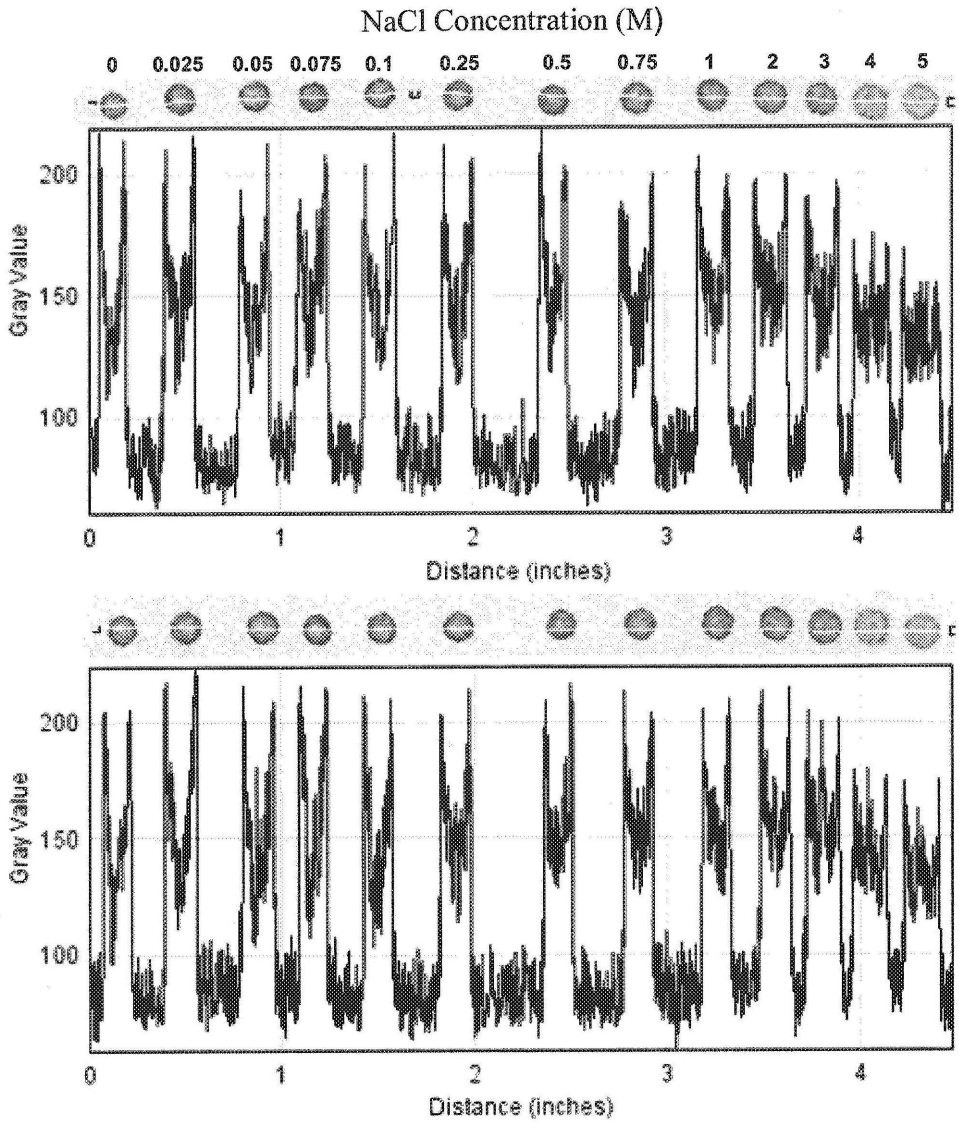


Figure 3. Protein assays and ImageJ intensity plots of serum protein concentration from acetonitrile pellets at increasing extract salt concentration. The lines illustrated in the plots transverse the assay spots.

MALDI-Qq-TOF

Serum extracts were purified using C18 reversed-phase microcolumns then spotted onto stainless steel target plates (Applied Biosystems/Sciex) to diameters not exceeding 3 millimeters. The matrix material CHCA (Agilent, Santa Clara, CA, USA) was manually applied. MALDI MS and MS/MS were manually collected on a QSTAR Elite Qq-TOF in the positive ion mode, between m/z 0-4400 (AB/Sciex, Concord, ON, Canada). Three randomly selected points on each dried sample spot were irradiated until ion current declined. MALDI-MS/MS were produced using CID voltages not exceeding 115 eV. Instrument calibration was performed using the instructions of the instrument manufacturer.

LC-ESI-Qq-TOF

Serum preparations were purified by C18 reversed-phase then separated by HPLC. Polypeptides were introduced onto the column via a 1 microlitre fixed volume injection loop and separated on a 150 mm 75 micrometer inner diameter capillary column packed with C18 reversed-phase beads of 3 micrometer diameter and 100 Å pore size (LC Packings, Amsterdam, The Netherlands). The column was operated over a 90 minute gradient from 5 to 65% acetonitrile at a flow rate of 200 nanoliters per minute via a Proxion nano-LC capillary pump (Amsterdam, The Netherlands). The solutions used for the binary mobile phase system were 94.5% acetonitrile with 5% water and 94.5% water with 5% acetonitrile. Both contained 0.1% acetic acid (v/v). Each chromatographic separation was followed by an extensive wash in 65% acetonitrile. A 10 micrometer inner diameter PicoTip SilicaTip emitter was used for electrospray (New

Objective, Woburn, MA, USA). The tip was mounted into a NanoES ion source (Protana Engineering A/S, Odense, Denmark) and an ESI voltage between 1800 and 3000 was applied. The Qq-TOF instrument was operated in the positive ion mode with a curtain gas setting of 20, a declustering potential of 65 V and a 265 V focusing potential. Nitrogen from a Parker Balston LCMS-5000 Tri-Gas Generator (Haverhill, MA, USA) was used as a collision gas with the collision-activated dissociation parameter set to 5. MS scans were acquired for two seconds followed by three product ion scans. Switching criteria used were for ions between m/z 400 and 1500 Th with charge from 2 to 5 and a minimum of 10 counts.

MS/MS Data Analysis Using ProteinPilot, X!Tandem and Relational Databases

LC-ESI-Qq-TOF MS/MS data were searched both by X!Tandem and ProteinPilot against a forward human Reference Sequence (RefSeq) FASTA library downloaded from The National Center for Biotechnology Information (NCBI) in 2007 as well as the combined forward/reverse library (Pruitt et al., 2007). MALDI Qq-TOF MS/MS results were interrogated using X!Tandem with the same libraries. X!Tandem (Version 2006.06.01.2) analyses were conducted with two variables in search conditions. Parent mass accuracy tolerance settings of +3-1 Dalton (Da) and $\pm$100 parts per million (PPM) were used. Spectral features were compared to the theoretical b and y fragmentation series and all ion series; a, b, c, x, y, and z. Conserved search conditions included fragment mass error of 0.4 Da, no enzyme [X|X], no noise suppression, a maximum of 25 missed cleavages, and all possible modifications allowed, excluding mass shifts due to

isobaric labeling. The identical data set was searched using ProteinPilot Version 2.0.1 (Applied Biosystems, Foster City, CA, USA).

All ProteinPilot analyses were performed using the thorough ID search effort, phosphorylation emphasis setting and a protein detection threshold of 0.47 (66.0% confidence). The Mascot algorithm can be used through ProteinPilot but this feature was not enabled.

X!Tandem and ProteinPilot result files were parsed into single Microsoft database (.mdb file extension) format files to facilitate evaluation of the results. The database files retained each peptide's sequence, precursor protein, mass, charge, deviation from predicted mass (delta), probability scores and corresponding sample among other characteristics. These were used to filter and rank the MS/MS data with Microsoft Access. Sequence information provided by ProteinPilot forward and forward/reverse library searching was used to assess the selectivity of the extraction and analysis protocols for non-tryptic sequences. This was facilitated by Structured Query Language (SQL), with the results integrated into Microsoft database reports. Peptides with arginine or lysine residues at their C-termini or with these amino acids immediately N-terminal to their first residue were designated semi-tryptic. Peptides with both of these features were designated fully-tryptic and those with neither were designated non-tryptic.

Internal Standards in Serum for MALDI Analysis

[Glu^1] fibrinopeptide B and Angiotensin 1 were diluted in water and applied to 100 microlitre volumes of normal human serum to final amounts of 100 femtomoles to 10

picomole per microlitre. The spiked sera were prepared by C18 reversed-phase and spotted onto a metal target plate with CHCA for MALDI analysis.

RESULTS

Serum Peptide Extraction

Acetonitrile and acetone pellets were both water soluble, and could almost be completely resuspended by vigorous agitation in an excess of water. Protein concentrations of replicate extracts using the same solvents were between 0.1 and 8 mg/ml with an average standard deviation around 15% following brief incubation and minimal disruption of the pellet. Protein concentration was dependent on salt and acetonitrile concentration. From 20 to 40% acetonitrile, protein content was highest but approached the lower detection limit at concentrations around 65% (Table 1).

Sample Preparation	Protein Concentration (mg/mL)	Standard Deviation
Acetonitrile Supernatant	0.26 mg/mL (Williams *et al.* 2006)	
Aqueous Extract	0.89	0.11
100 mM NaCl	1.5	0.21
600mM NaCl	1.6	0.23
750 mM NaCl	0.96	0.13
1 mM HCl	0.76	0.089
10 mM HCl	1.2	0.14
500 mM HCl	0.12	0.014
1 mM NaOH	0.75	0.088
10 mM NaOH	1.9	0.23
500 mM NaOH	1.3	0.16
20% Acetonitrile	7.7	1.1
40% Acetonitrile	7.2	1.0
60% Acetonitrile	0.42	0.060

Table 1. Protein concentrations among extracts of acetonitrile serum precipitates measured using a colorimetric assay with Coomassie blue protein stain. Each assay was

standardized with bovine serum albumin in 1% SDS (w/v). For each concentration value, four intensity measurements of two replicates were averaged.

Acidic and Basic Extraction

There was no clear relationship between protein concentration and HCl content in the aqueous extracts. Total protein was comparable in extracts containing equivalent amounts of formic or acetic acid. Protein concentration increased with NaOH concentration over a range of 3 orders of magnitude. Base-catalyzed hydrolysis might create soluble peptides from larger less soluble sequences but this effect on protein concentration was minimized by the low extraction temperature. The large range of isoelectric points within the serum proteome is exploited in the isoelectric focusing dimension during two-dimensional electrophoresis. Many different proteins from the acetonitrile pellet likely enter solution over a wide pH range irrespective of their isoelectric points.

SDS-PAGE of Protein Extracts

Representative extracts were compared by SDS-PAGE and differed in their relative concentrations of high, mid and low molecular weight polypeptides (Fig 4). For acetone and acetonitrile pellets, the mass distribution profile from each extraction condition was comparable. Acetonitrile pellet extracts from 60% acetonitrile apparently contained a greater number of low molecular weight bands than equal volumes of the acetone pellet extracts. Extraction in 600mM NaCl gave more high molecular weight

bands using acetonitrile precipitation (Fig 4). Aqueous, 1 mM HCl and 5 mM NaOH extracts had similar profiles by SDS-PAGE with concentrated but poorly resolved low and high molecular mass fractions. The 60 % acetonitrile extract was well resolved in the region below 14 kD. In 600 mM NaCl extracts, silver staining of the low molecular mass region was vanishingly faint, though total protein in this sample was decreased by the effect of salt on polypeptide solubility after drying. Considering the decrease in stain intensity with decreasing molecular mass, all extracts from acetonitrile pellets were rich in low molecular mass analytes.

Figure 4.

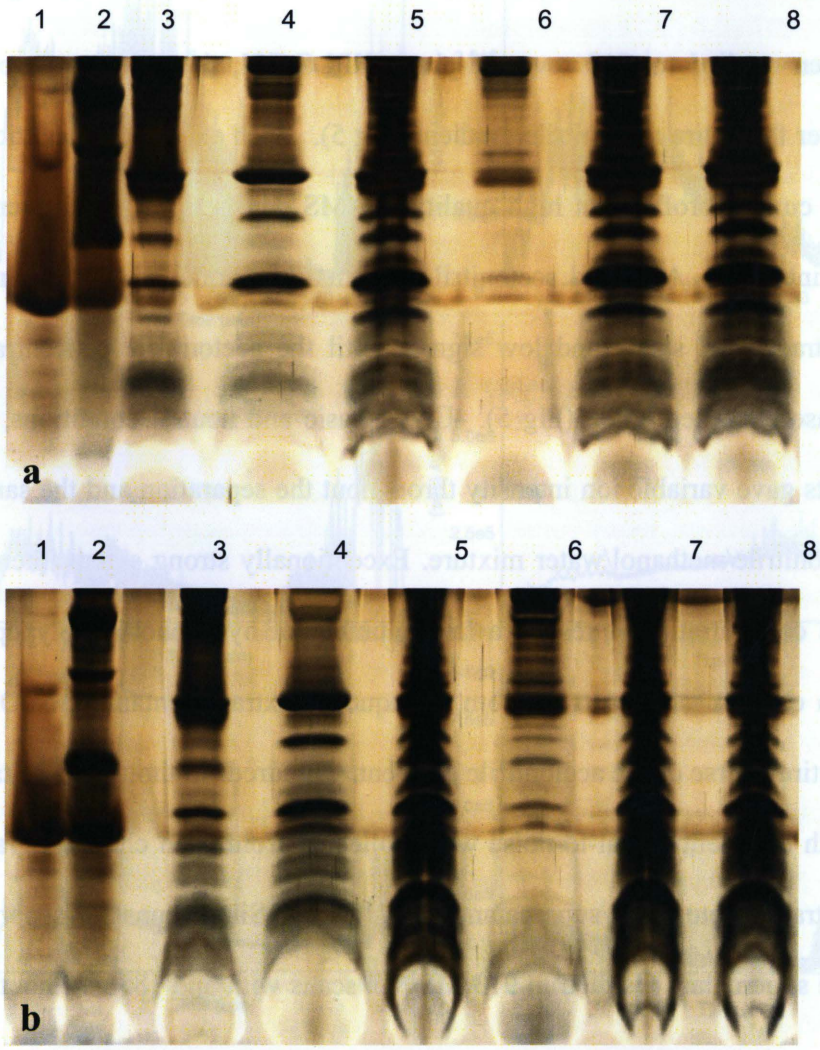


Figure 4. Extracts of acetone and acetonitrile serum pellets separated by 17% acrylamide SDS-PAGE and visualized by diamine silver stain. (a) Acetone pellets (b) acetonitrile pellets. Lane 1: 3.75μg lysozyme; Lane 2: molecular weight markers 14, 97 kDa; Lane 3: 0.1μl NHS; 4: 30μl 60% acetonitrile supernatant; Lane 5: H2O supernatant; Lane 6: 600 mM NaCl; Lane 7: 1mM formic acid supernatant; Lane 8: 5mM NaOH.

LC-ESI-MS

When applied to C18 reversed-phase HPLC-ESI, the extracts gave strong ion currents over the entire acetonitrile gradient (Fig 5). Each extraction condition yielded a unique ion current profile, but high quality MS/MS data was obtained over a range of parent ion intensity. Acidified acetonitrile supernatants, aqueous extracts and aqueous-organic extracts had stable and low signals until the acetonitrile concentration of the mobile phase exceeded 45 % (Fig 5). Under basic and neutral conditions, acetonitrile supernatants gave variable ion intensity throughout the separation and the same was true of the acetonitrile/methanol/water mixture. Exceptionally strong signal intensity at high acetonitrile concentrations attest to a large number of hydrophobic polypeptides being retained on column. Ion current from the aqueous extract containing NaOH was high over the entire course of the acetonitrile gradient. Hundreds to thousands of tandem mass spectra with excellent signal-to-noise were collected from each chromatographic run of the ultrafiltrate, acetonitrile supernatant and extracts. Single runs for separation of the acetonitrile supernatant resulted in 913 MS/MS scans with and 1836 without TFE. The

methanol/acetonitrile/water extract had only 562 tandem spectra and 573 when acidified with TFE. The aqueous extract had 203 MS/MS scans by ESI-Qq-TOF.

Figure 5.

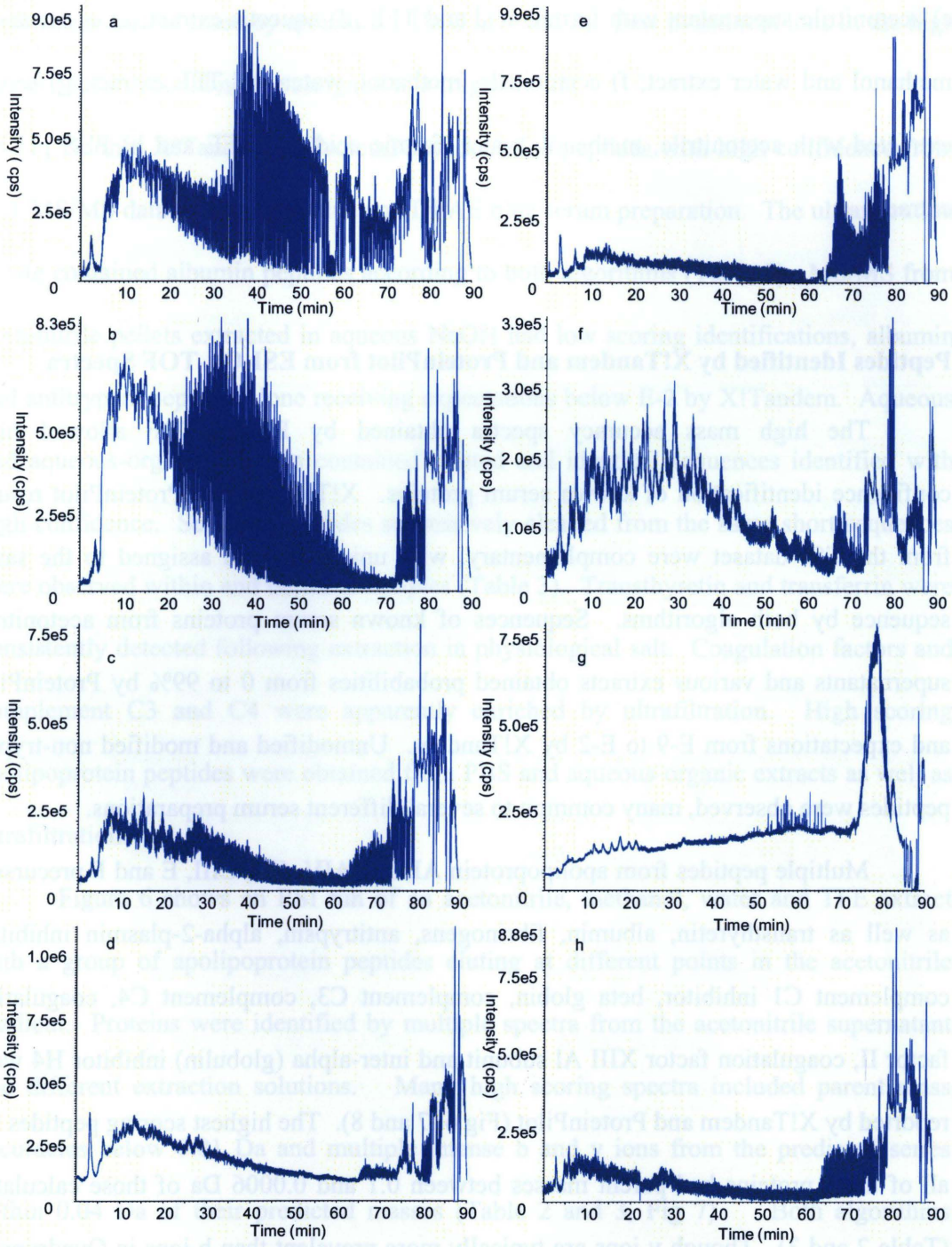


Figure 5. Total ESI-Qq-TOF ion currents of reversed-phase C18 eluates over ninety minute 0-65% acetonitrile gradients. Serum was precipitated with acetonitrile and the samples shown are a) the acetonitrile supernatant; b) acetonitrile supernatant with NaOH; c) acetonitrile supernatant with formic acid and TFE; d) aqueous extract; e) acetonitrile, methanol and water extract; f) acetonitrile, methanol, water and TFE extract; g) serum extracted with acetonitrile, methanol, water, formic acid and TFE and h) PBS pH 8.8 extract.

Peptides Identified by X!Tandem and ProteinPilot from ESI-Qq-TOF Spectra

The high mass accuracy spectra obtained by ESI-Qq-TOF allowed high confidence identification of known serum proteins. X!Tandem and ProteinPilot results from the ESI dataset were complementary, with unique spectra assigned to the same sequence by both algorithms. Sequences of known serum proteins from acetonitrile supernatants and various extracts obtained probabilities from 0 to 99% by ProteinPilot and expectations from E-9 to E-2 by X!Tandem. Unmodified and modified non-tryptic peptides were observed, many common to several different serum preparations.

Multiple peptides from apolipoprotein AI, AII AIV, CII, CIII, E and F precursors as well as transthyretin, albumin, fibrinogens, antitrypsin, alpha-2-plasmin inhibitor, complement C1 inhibitor, beta globin, complement C3, complement C4, coagulation factor II, coagulation factor XIII Al subunit and inter-alpha (globulin) inhibitor H4 were reported by X!Tandem and ProteinPilot (Fig 6, 7 and 8). The highest scoring peptides for all of these proteins had parent masses between 0.1 and 0.0006 Da of those calculated (Table 2 and 3). Though y ions are typically more prevalent than b ions in Quadrupole-

TOF instruments, the relative number of these two series was generally comparable in the MS/MS search results (Fig 7 and 8) (Steen and Mann, 2004). A small number of fragment ions approached the set X!Tandem mass accuracy limit of 0.4 Da, most varying from the predicted mass by around 0.1 Da (Fig 7 and 8). Few prominent ions in the high scoring spectra could not be assigned to a fragment ion.

Neither X!Tandem or ProteinPilot identified peptides with high confidence from ESI-MS/MS data of Cibacron blue and DEAE blue serum preparation. The ultrafiltration eluate contained albumin peptides according to both algorithms (Table 3). MS/MS from acetonitrile pellets extracted in aqueous NaOH had low scoring identifications, albumin and antitrypsin peptides alone receiving expectations below E-2 by X!Tandem. Aqueous and aqueous-organic extracts contained related and identical sequences identified with high confidence. Series of peptides successively cleaved from the same short sequences were observed within and between samples (Table 3). Transthyretin and transferrin were consistently detected following extraction in physiological salt. Coagulation factors and complement C3 and C4 were apparently enriched by ultrafiltration. High scoring apolipoprotein peptides were obtained from PBS and aqueous-organic extracts as well as ultrafiltration.

Figure 6 shows an ESI run of an acetonitrile, methanol, water and TFE extract with a group of apolipoprotein peptides eluting at different points in the acetonitrile gradient. Proteins were identified by multiple spectra from the acetonitrile supernatant and different extraction solutions. Many high scoring spectra included parent mass accuracies below 0.01 Da and multiple intense b and y ions from the predicted series within 0.04 Da of their predicted masses (Table 2 and 3, Fig 7). Both algorithms

identified peptides from the same set of known serum proteins. Search results were examined manually and sequences containing strings of alanine, glycine or proline were discarded. Assignments incorporating continuous runs of a single amino acid typically had X!Tandem expectation values on the order of E-2.

Peptides in the Acetonitrile Supernatant

According to the search results for the ESI-Qq-TOF data, peptides from many serum proteins remained in solution following acetonitrile precipitation. Apolipoprotein peptides were identified with high confidence in acidified acetonitrile supernatant but had low scores in the acetonitrile supernatant alone or with NaOH. Acetonitrile precipitation including NaOH generally yielded poor quality spectra relative to other supernatants, but abundant serum proteins such as albumin and serine (or cysteine) protease inhibitor (antitrypsin) were found in this case (Table 2 and 3). Several peptides from complement C4 were observed with confidence below 90% in this sample pool. Peptides from albumin and inter-alpha (globulin) inhibitor H4 (ITIH4) were also routinely detected in the acetonitrile supernatants. Focusing the mass range analyzed between m/z 800 and 2500 may improve resolution, signal and mass accuracy for these analytes.

Table 2.

Protein/ Accession	M+H	z	Delta	Expect	Peptide Sequence	Sample
transthyretin						
4507725	2041.049	2	-0.044	2.9E-10	ALLSPYSYSTTAVVTNPKE	H2O
	1970.012	2	-0.111	1.5E-09	LLSPYSYSTTAVVTNPKE	AMH
	1273.631	2	-0.032	0.000071	ALLSPYSYSTTA	MPBS
	3443.745	4	-0.044	0.00079	TANDSGPRRYTIAALLSPYSYSTTAVVTNPKE	IPBS
	1743.844	2	-0.054	6E-09	SPYSYSTTAVVTNPKE	MPBS
	1396.6959	2	-0.0009	2.9E-08	SYSTTAVVTNPKE	PBS
albumin precursor						
4502027	1654.786	2	-0.012	5.4E-09	ALEVDETYVPKEFN	AS
	1741.818	2	0.022	7.9E-09	SALEVDETYVPKEFN	AS
	1888.886	2	0.025	6.5E-08	FSALEVDETYVPKEFN	ANS
	1525.767	2	-0.036	5.9E-06	PLVEEPQNLIKQN	MPBS
	1410.739	2	0.016	0.00029	PLVEEPQNLIKQ	AS
	1955.039	3	-0.047	0.000018	FEQLGEYKFQNALLVR	EPBS
	1344.669	2	0.016	0.0038	PADLPSLAADFVE	ANS
alpha-2-HS-glycoprotein						
4502005	1858.912	3	-0.131	3.5E-07	FMGVVSLGSPSGEVSHPR	AMHFE
	2081.024	3	-0.022	8.3E-07	HTFMGVVSLGSPSGEVSHPR	YM50
	2072.0598	3	-0.003	3.9E-06	FMGVVSLGSPSGEVSHPRKT	YM50
beta globin						
4504349	3274.743	4	0.954	0.000011	VHLTPEEKSAVTALWGKVNVDEVGGEALGRL	IPBS
	1912.029	3	-0.016	0.00013	WGKVNVDEVGGEALGRLL	IPBS
	1869.0134	4	-0.0006	0.012	YQKVVAGVANALAHKYH	AFES
apolipoprotein A-I preproprotein						
4557321	1303.665	2	-0.025	7.8E-06	REQLGPVTQEF	IPBS
	1259.583	2	1.019	8.4E-06	QEKLSPLGEEM	PBS
	1285.668	2	-0.032	0.000025	KVSFLSALEEY	IPBS
	2099.139	3	0.093	0.000019	KVSFLSALEEYTKKLNTQ	AMHE
apolipoprotein A-II preproprotein						
4502149	1553.785	2	-0.0094	8.3E-07	LSYFVELGTQPATQ	PBS
	1353.6689	2	-0.0053	1.9E-06	YFVELGTQPATQ	PBS
	1190.606	2	0.039	0.00055	FVELGTQPATQ	AMHE
	1062.547	2	0.02	0.00063	FVELGTQPAT	PBS
apolipoprotein A-IV precursor						
71773110	1771.85	2	-0.107	4.4E-08	SLAELGGHLDQQVEEF	AMH
	1495.739	2	-0.012	0.000079	SLAELGGHLDQQVE	PBS
apolipoprotein C-II precursor						
32130518	2175.017	2	-0.024	3.7E-08	DEMPSPTFLTQVKESLSSY	IPBS
	2885.351	3	-0.038	2.6E-07	TQQPQQDEMPSPTFLTQVKESLSSY	IPBS
	2885.351	3	-0.126	2.8E-06	TQQPQQDEMPSPTFLTQVKESLSSY	EPBS
apolipoprotein C-III precursor						
4557323	1810.751	2	-0.114	1.1E-07	SEAEDASLLSFMQGYM	AMHE
	1443.631	2	-0.011	1.4E-06	SEAEDASLLSFMQ	PBS
	1688.845	2	-0.085	6.1E-06	KDALSSVQESQVAQQA	AMH
	1560.75	2	-0.096	0.000013	DALSSVQESQVAQQA	AMHE
apolipoprotein E precursor						
4557325	1597.796	2	0.0039	0.00061	KVEQAVETEPEPEL	AFS
apolipoprotein F precursor						
4502165	1249.595	2	-0.055	0.000015	YDLDPGAGSLEI	EPBS
	1086.532	2	-0.01	0.0018	DLDPGAGSLEI	PBS

Table 2 continued.

coagulation factor II precursor						
4503635	1807.7469	2	-0.0046	1.6E-06	EETGDGLDEDSDRAIEG	PBS
	1549.662	2	-0.082	0.000019	TGDGLDEDSDRAIEG	AMH
coagulation factor XIII A1 subunit precursor						
119395709	2238.1001	2	-0.0032	4E-09	NNSNAAEDDLPTVELQGVVPR	YM50
complement component 3 precursor						
115298678	1855.856	3	0.09	0.00035	SEETKENEGFTVTAEGK	AMHE
	1727.761	2	-0.098	1.2E-06	SEETKENEGFTVTAEG	AMHE
	1640.729	2	-0.0058	7.3E-06	EETKENEGFTVTAEG	PBS
	1617.904	2	-0.037	0.000014	SPMYSIITPNILRL	IPBS
complement component 4 preproprotein						
67190748	2551.1774	3	-0.0059	0.00037	TLEIPGNSDPNMIPDGDFNSYVR	YM50
	2148.939	2	-0.142	1.9E-07	TLEIPGNSDPNMIPDGDFNS	AMHE
	2045.912	2	-0.128	8.8E-07	TLEIPGNSDPNMIPDGDFN	AMH
	1762.923	2	-0.018	7.2E-06	GLEEELQFSLGSKINV	YM50
	1951.9652	2	-0.0057	0.000089	DDPDAPLQPVTPLQLFEG	YM50
complement component 1 inhibitor precursor						
73858568	1433.804	2	-0.024	0.000038	TLLVFEVQQPFL	AES
	1253.657	2	-0.021	0.0027	VFEVQQPFLF	AFES
complement factor B preproprotein						
67782358	1637.704	2	-0.068	0.000074	IEGVDAEDGHGPGEQQ	EPBS
fibrinogen, alpha polypeptide isoform alpha preproprotein						
11761629	1389.521	2	-0.108	3E-09	DSGEGDFLAEGGGV	AMHE
	1542.594	2	-0.038	0.000003	DEAGSEADHEGTHST	H2O
	1354.515	2	-0.047	0.00074	DEAGSEADHEGTH	H2O
	1298.525	2	0.055	0.00097	AGSEADHEGTHST	AMHE
	1227.488	2	0.039	0.007	GSEADHEGTHST	AMHE
fibrinogen, beta chain preproprotein						
70906435	1325.528	2	-0.094	0.0016	QGVNDNEEGFFS	AMHE
	1238.496	2	0.037	0.00027	QGVNDNEEGFF	PBS
inter-alpha (globulin) inhibitor H4						
31542984	2010.971	2	-0.101	4.2E-08	QLGLPGPPDVPDHAAYHPF	EPBS
	1863.903	2	-0.15	2.6E-07	QLGLPGPPDVPDHAAYHP	AMHFE
	1616.75	2	0.023	7.9E-09	PGPPDVPDHAAYHPF	AS
	1466.7278	2	-0.0038	0.00091	QLGLPGPPDVPDHAA	PBS
	1395.691	2	0.014	0.000031	QLGLPGPPDVPDHA	H2O
	1324.654	2	-0.083	3.8E-06	QLGLPGPPDVPDH	AMH
serine (or cysteine) proteinase inhibitor clade A (antitrypsin)						
50363217	1274.55	2	-0.041	1.6E-06	EDPQGDAAQKTD	MPBS
	1320.724	2	-0.046	0.000077	PVSIATAFAMLSL	H2ON
	1377.745	2	-0.025	0.000012	PVSIATAFAMLSLG	H2ON
	2262.163	3	-0.053	0.00027	MIEQNTKSPLFMGKVVNPTQ	MPBS
	1608.813	2	-0.036	0.00032	LMIEQNTKSPLFMG	IPBS
transferrin						
4557871	2961.467	3	0.968	0.000017	AYLAPNNLKPVVAEFYGSKEDPQTFY	IPBS
	1645.895	2	-0.028	2.5E-07	AYLAPNNLKPVVAEF	IPBS
	1574.858	2	-0.027	0.00011	YLAPNNLKPVVAEF	IPBS
	1411.795	2	-0.021	0.000075	LAPNNLKPVVAEF	IPBS
	1760.922	2	-0.07	0.0018	DAYLAPNNLKPVVAEF	EPBS

Table 2. Selected peptides identified by X!Tandem from spectra of serum acetonitrile supernatants and pellet extracts. The column titles are abbreviated as follows: M+H: observed precursor ion m/z, z: predicted precursor ion charge, Delta: difference between the calculated peptide mass and the observed mass, Expect: expectation value of the peptide. AFS: acetonitrile supernatant with formic acid, AFES: acetonitrile supernatant with formic acid and TFE, AMH: acetonitrile and methanol and water extract, AMHE: acetonitrile and methanol and water extract with TFE, AMHFE: acetonitrile and methanol and water extract with formic acid and TFE, ANS: acetonitrile supernatant with NaOH, AS: acetonitrile supernatant, EPBS: PBS extract containing TFE, H_2O: aqueous extract, IPBS: PBS extract from isopropanol pellet, MPBS: PBS extract from methanol pellet, PBS: PBS extract, YM50: ultrafiltrate from YM50 size-exclusion cartridges.

Table 3. Peptides Identified by ProteinPilot with 99% confidence arranged by protein. Sequence Coverage P>95%: percentage of a given protein sequence identified from peptide spectra exceeding 95% probability, m/z: ion mass divided by the predicted ion charge, z: predicted ion charge, Delta: difference between the calculated peptide mass and the observed mass. Modifications are identified on the left and their locations indicated by the amino acid residue number following the @ symbol. Sample identification follows the same conventions as in Table 2, with the addition of the abbreviation FE meaning aqueous extract of the acetonitrile pellet containing formic acid and TFE.

Table 3.

Protein/ Total Score	Sequence Coverage P > 95%	m/z	z	Delta	Peptide Score	Peptide Sequence	Modifications	Sample
albumin								
94.3	48	505.57	3	-0.082	16	ACLLPKLDELRDE		H2ON
		689.11	4	-0.030	24	DAHKSEVAHRFKDLGEENFKALVL		IPBS
		904.91	2	-0.010	23	EVTEFAKTCVADESAEN	Cys->Dha(C)@9	ANS
		737.32	2	-0.070	19	FEQLGEYKFQNA		EPBS
		568.02	4	-0.033	18	IAFAQYLQQCPFEDHVKL	Cysteinyl(C)@10	IPBS
		941.12	3	-0.049	17	IAFAQYLQQCPFEDHVKLVNEVTE		IPBS
		649.39	2	-0.021	17	KKLVAASQAALGL	Formyl@N-term	AFS
		571.35	2	-0.009	16	KLVAASQAALGL		AFES
		780.38	2	-0.037	17	PADLPSLAADFVESK		H2ON
		545.97	3	-0.006	18	PNLPRLVRPEVDVM	Deamidated(N)@2	H2ON
alpha-1 antiproteinase(antitrypsin)								
54.5	36	688.30	2	-0.010	20	EDPQGDAAQKTDT		ANS
		512.03	4	-0.084	18	GLKLVDKFLEDVKKLYH		H2ON
		626.55	4	-0.180	23	LMIEQNTKSPLFMGKVVNPTQK		AMHE
		628.33	3	-0.086	20	TYDLKSVLGQLGITKVF	Deamidated(Q)@10	H2ON
alpha-2-plasmin inhibitor								
10.8	18	816.41	2	-0.009	20	MSLSSFSVNRPFLF		AFES
		835.68	4	-0.053	20	IFEDTTGLPLFVGSVRNPNPSAPRELKEQQ		AFES
		824.41	2	-0.002	21	MSLSSFSVNRPFLF	Oxidation(M)@1	AFES
apolipoprotein A-I								
135.0	72	626.99	3	-0.021	16	AELQEGARQKLHELQE		PBS
		553.05	4	-0.022	23	ATEHLSTLSEKAKPALEDLR		YM50
		770.64	4	0.003	17	ELQEGARQKLHELQEKLSPLGEEMRD	Oxidation(M)@24	AFES
		700.38	3	-0.016	26	KVSFLSALEEYTKKLNTQ		AFS
		806.85	2	-0.094	19	LDNWDSVTSTFSKL		EPBS
		565.29	2	-0.002	16	QEKLSPLGEE		PBS
		623.32	4	-0.032	24	QKLHELQEKLSPLGEEMRDRA	Gln->pyro-Glu@N-term	YM50
		762.87	2	-0.061	21	SALEEYTKKLNTQ		EPBS
		770.66	4	-0.006	22	VDVLKDSGRDYVSQFEGSALGKQLNLKL		IPBS
apolipoprotein A-II								
22.8	77	643.79	2	-0.058	15	FQTVTDYGKDL		EPBS
		595.79	2	-0.041	15	FVELGTQPATQ		FE
		777.34	2	-0.114	17	LSYFVELGTQPATQ		AMHE
		618.65	3	-0.009	17	MEKVKSPELQAEAKSY	Oxidation(M)@1	IPBS
apolipoprotein A-IV								
135.0	54	817.76	3	-0.006	21	GRLTPYADEFKVKIDQTVEEL		YM50
		818.40	2	-0.009	18	QDTQEKLNHQLEGL	Gln->pyro-Glu@N-term	PBS
		512.93	3	-0.007	15	RVEPYGENFNKAL		PBS
		674.09	4	-0.007	22	RVEPYGENFNKALVQQMEQLRQ	Oxidation(M)@17	YM50
		559.29	4	-0.006	20	VLRENADSLQASLRPHADEL		AFS
apolipoprotein C-II								
9.7	59	1088.00	2	-0.021	41	DEMPSPTFLTQVKESLSSY	Oxidation(M)@3	IPBS
		855.42	3	0.004	54	STAAMSTYTGIFTDQVLSVLKGEE	Oxidation(M)@5	YM50
		962.44	3	-0.032	38	TQQPQQDEMPSPTFLTQVKESLSSY	Oxidation(M)@9	IPBS
		1088.00	2	-0.021	26	DEMPSPTFLTQVKESLSSY	Oxidation(M)@3	IPBS
apolipoprotein C-III								
47.6	80	723.35	2	-0.025	22	ALSSVQESQVAQQA		PBS
		584.74	2	-0.062	17	SEAEDASLLSF		EPBS
		722.31	2	-0.011	17	SEAEDASLLSFMQ	Oxidation(M)@12	PBS
		709.33	2	-0.017	16	WVTDGFSSLKDY		IPBS
apolipoprotein E								
26.2	34	583.63	3	-0.017	19	KVQAAVGTSAAPVPSDNH		IPBS
		839.08	3	-0.019	20	QWAGLVEKVQAAVGTSAAPVPSDNH	Gln->pyro-Glu@N-term	YM50
		486.48	4	-0.117	18	TRDRLDEVKEQVAEVR		AMH
		735.35	2	-0.004	15	VEQAVETEPEPEL		AFES
apolipoprotein F								
4.2	4	543.76	2	-0.010	17	DLDPGAGSLEI		PBS
		625.27	2	-0.059	18	YDLDPGAGSLEI		EPBS
coagulation factor II								
19.7	9	834.39	2	-0.049	15	DKTERELLESYIDG		H2O
		638.79	2	0.003	14	GLDEDSDRAIEG		PBS
		718.35	3	-0.027	23	SLEDKTERELLESYIDGR		YM50
		775.29	2	-0.089	21	TGDGLDEDSDRAIEG		AMH
coagulation factor XIII A1 subunit								
2.1	3	1119.55	2	-0.008	22	NNSNAAEDDLPTVELQGVVPR		YM50
		747.05	3	0.039	14	NNSNAAEDDLPTVELQGVVPR	Deamidated(N)@4	YM50

Table 3 continued.

complement factor B								
2.3	2	819.32	2	-0.072	19	IEGVDAEDGHGPGEQQ		EPBS
complement C3								
27.0	4	820.86	2	-0.007	19	EETKENEGFTVTAEG		PBS
		593.33	3	-0.010	16	SKITHRIHWESASLL		YM50
		484.27	4	-0.030	18	SKITHRIHWESASLLR		YM50
		809.45	2	-0.019	17	SPMYSIITPNILRL		IPBS
		506.03	4	0.006	24	SSKITHRIHWESASLLR		YM50
		483.92	3	-0.016	14	THRIHWESASLL		YM50
complement C4								
42.0	7	976.49	2	-0.001	17	DDPDAPLQPVTPLQLFEG		YM50
		881.96	2	-0.015	22	GLEEELQFSLGSKINV		YM50
		1031.45	2	-0.017	22	TLEIPGNSDPNMIPDGDFN	Oxidation(M)@12	YM50
		851.06	3	-0.005	20	TLEIPGNSDPNMIPDGDFNSYVR		YM50
complement C1 inhibitor								
9.8	7	717.39	2	-0.023	17	TLLVFEVQQPFL		AES
		627.32	2	-0.021	15	VFEVQQPFLF		AFES
fibrinogen alpha								
21.5	10	677.74	2	-0.034	20	DEAGSEADHEGTH		FE
		771.78	2	-0.050	22	DEAGSEADHEGTHST		H2O
		695.22	2	-0.080	21	DSGEGDFLAEGGGV	Phospho(S)@2	AMHE
		704.84	4	0.014	22	GSESGIFTNTKESSSHHPGIAEFPSRG		YM50
		1062.99	2	0.022	22	TFPGFFSPMLGEFVSETES	Oxidation(M)@9	YM50
fibrinogen beta								
4.3	2	619.75	2	-0.004	14	QGVNDNEEGFF	Gln->pyro-Glu@N-term	PBS
		663.22	2	-0.087	16	QGVNDNEEGFFS	Gln->pyro-Glu@N-term	AMHE
		619.72	2	-0.063	14	QGVNDNEEGFF	Gln->pyro-Glu@N-term	EPBS
inter-alpha (globulin) inhibitor H4								
53.4	7	507.24	3	-0.001	14	GPPDVPDHAAYHPF		YM50
		646.59	4	0.005	24	NVHSGSTFFKYYLQGAKIPKPEA	Deamidated(N)@1	YM50
		789.91	4	-0.009	21	NVHSGSTFFKYYLQGAKIPKPEASFSPR		YM50
		681.85	4	-0.028	23	PGVLSSRQLGLPGPPDVPDHAAYHPF		YM50
plasminogen								
2.3	2	761.357	2	-0.057648	16	EPLDDYVNTQGASL		EPBS
transferrin								
34.3	8	823.41	2	-0.075	20	AYLAPNNLKPVVAEF		EPBS
		933.46	3	-0.039	18	AYLAPNNLKPVVAEFYGSKEDPQTF		IPBS
		987.82	3	-0.032	23	AYLAPNNLKPVVAEFYGSKEDPQTFY		IPBS
		880.93	2	-0.069	21	DAYLAPNNLKPVVAEF		EPBS
		787.92	2	-0.025	15	YLAPNNLKPVVAEF		IPBS
		752.05	3	-0.017	23	YLAPNNLKPVVAEFYGSKED		IPBS
transthyretin								
33.4	65	650.97	3	-0.041	19	GPTGTGESKCPLMVKVLD	Cysteinyl(C)@10	MPBS
		985.50	2	-0.014	22	LLSPYSYSTTAVVTNPKE		PBS
		892.78	3	-0.015	19	TANDSGPRRYTIAALLSPYSYSTTA		IPBS
		637.29	2	-0.057	15	ALLSPYSYSTTA		EPBS

Figure 6.

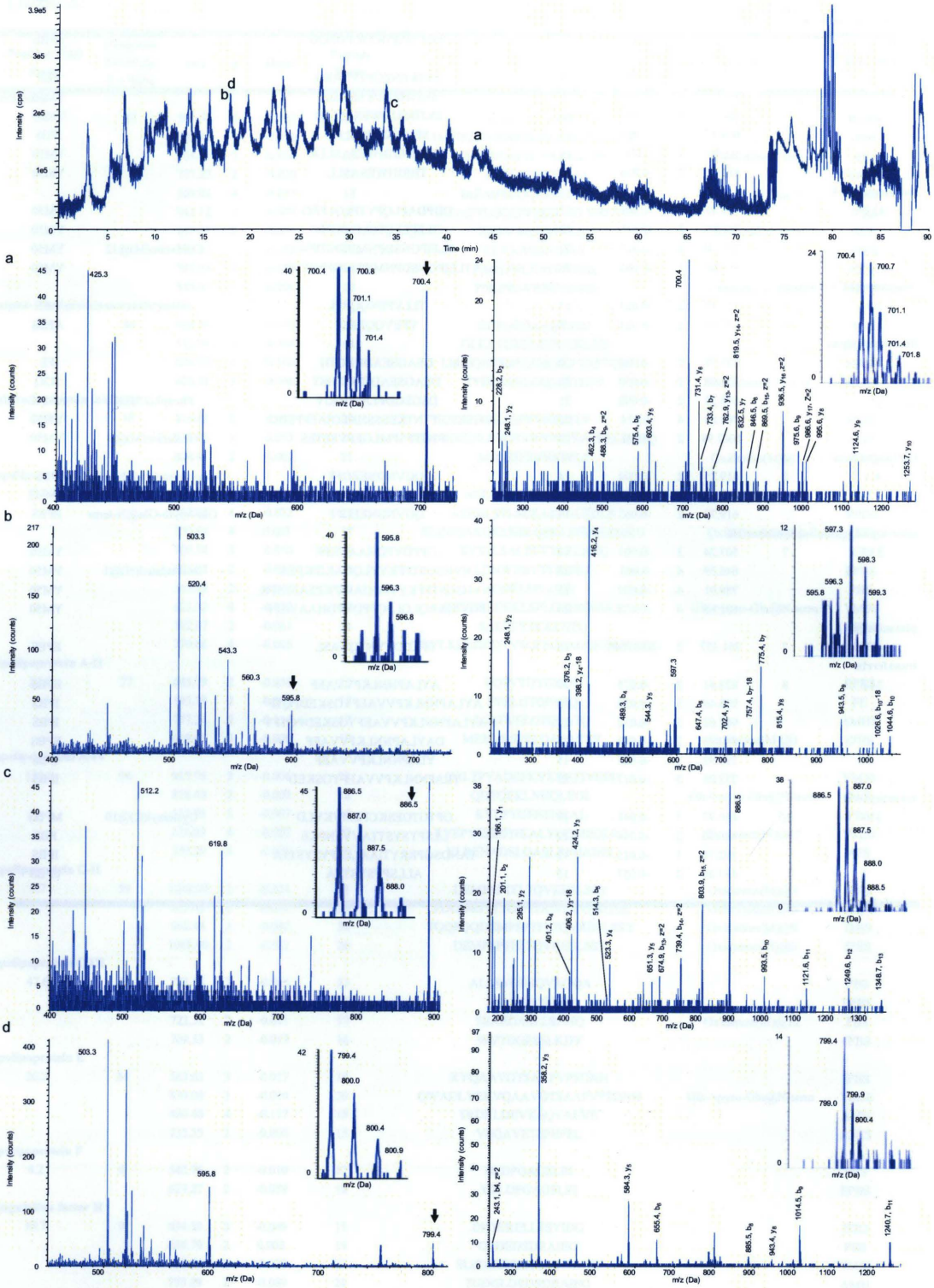


Figure 6. Apolipoprotein peptides from acetonitrile serum pellets extracted in acetonitrile, methanol and water with 5% TFE (v/v) analyzed by nanoflow LC-ESI-MS/MS. The top panel shows the total ion current (TIC) trace with the times of the peptide spectra marked by their corresponding letters. (a-d) The spectra on the left are MS scans and those on the right are the corresponding MS/MS. (a) KVSFLSALEEYTKKLNTQ from Apolipoprotein AI eluted at 44.0 minutes into the acetonitrile gradient, M+H 2099.1, z = 3, Delta 0.093 Da. (b) FVELGTQPATQ from apolipoprotein AII eluted at 17.6 minutes, M+H 1190.6, z = 2, Delta 0.039 Da. (c) SLAELGGHLDQQVEEF from apolipoprotein AIV eluted at 35.2 minutes, M+H 1771.9, z = 2, Delta 0.065 Da (d) KVEQAVETEPEPEL from apolipoprotein E eluted at 17.7 minutes, M+H 1597.8, z = 2, Delta 0.045. The parent ions are indicated by arrows and isotopic mass envelopes are inset in the MS and MS/MS panels.

Figure 7.

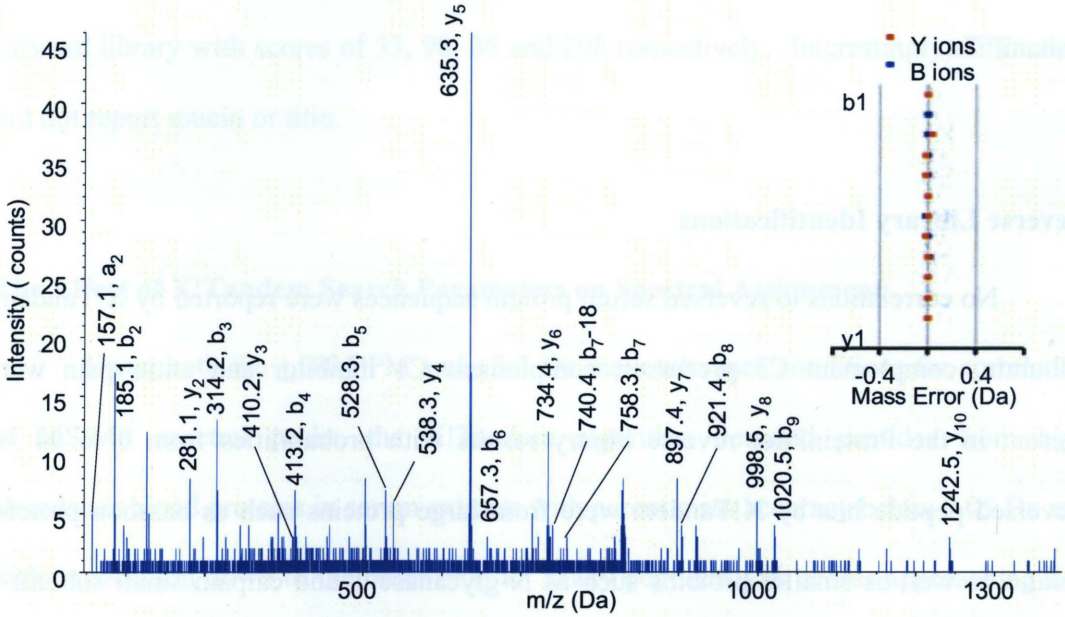


Figure 7. ESI-Qq-TOF tandem mass spectrum of the doubly charged albumin peptide ALEVDETYVPKEFN identified by X!Tandem. The peaks corresponding to a, b and y series fragments of a precursor ion with a mass of 1654.8 are labeled. Inset is the X!Tandem fragment mass accuracy diagram.

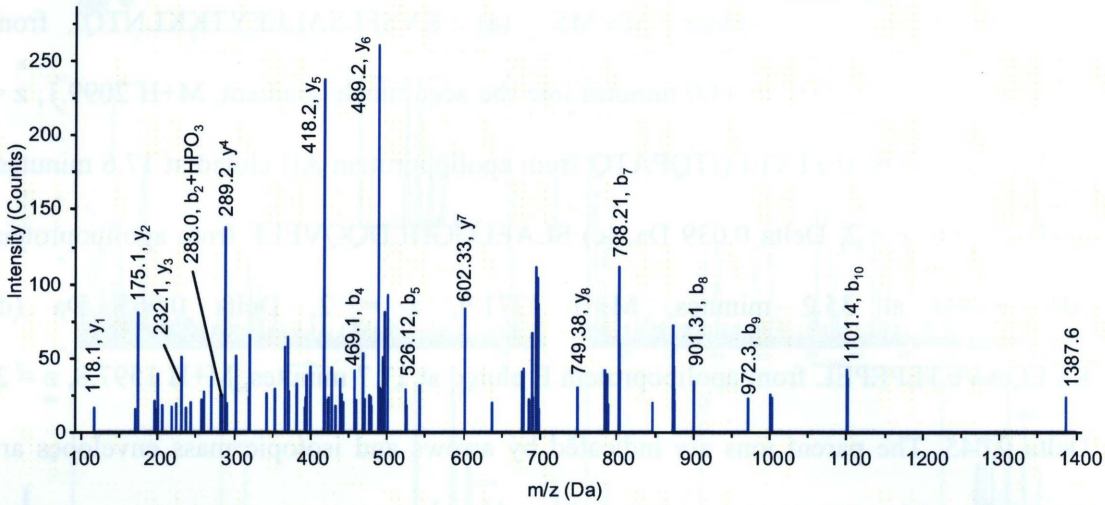


Figure 8. ESI-Qq-TOF tandem mass spectrum of the phosphopeptide $D_PSGEGDFLAEGGGV$ from fibrinogen alpha with b and y series ions matched by ProteinPilot.

Reverse Library Identifications

No correlations to reversed serum protein sequences were reported by X!Tandem. Albumin, complement C3 precursor, complement C1 inhibitor and antitrypsin were present in the ProteinPilot reverse library results with probabilities from 0 to 94 %. Reversed peptide hits by X!Tandem were from large proteins such as bassoon protein, keratin as well as smaller proteins such as N-glycanase 1 and calpain small subunit 1 (Table A1). The majority of these reverse hits were from serum extracted under basic

aqueous extraction conditions while the supernatant from precipitation with acetonitrile containing NaOH had few reverse hits.

ProteinPilot reported a total of 119477 unique peptides from the combined forward and forward/reverse library searches, with 41573 reversed and 83587 forward sequences. Forward and reverse libraries shared 5683 identifications by ProteinPilot. The vast majority of these shared sequences were four or five residues in length with the exception of longer repetitive sequences such as QQQQQQLQL and AGGGGGGGG. From 77904 sequences exclusive to the forward library 3510 proteins were identified with some redundancy. The number of forward proteins was dependent on the library used: 3247 were identified from forward and 1962 from forward/reverse searching.

The highest scoring reverse library peptides by ProteinPilot were from mucin and titin. Mucin 2, 5B, 16, 17 and 19 were identified from the reverse library with total protein scores between 23 and 165 while mucin 2, 16, 17 and 19 were identified from the forward library with scores of 33, 98, 35 and 208 respectively. Interestingly, X!Tandem did not report mucin or titin.

The Effect of X!Tandem Search Parameters on Spectral Assignment

Using both the $\pm$100 PPM and +3-1 Da mass tolerance conditions for correlation of MS/MS spectral peaks, the X!Tandem algorithm detected peptides from high abundance blood proteins in serum extracts with parent mass accuracy below 0.05 Da and with expectation down to the order of E-9. Many peptides with scores approaching E-0 belonged to the same proteins as those identified with high confidence. Identical spectra were correlated to the same peptides using both $\pm$100 PPM and +3-1 Da conditions, with

minor variations in expectation values and in the number of spectral peaks assigned (Table 4). X!Tandem scored identical spectra differently under different parent mass tolerance settings. Neither the ±100 PPM nor +3-1 Da setting produced consistently lower expectation for the same peptide identifications. A small minority of spectra were assigned identical expectation values when searched under both conditions. Typically, peptides from the same precursor ion yielded expectations within one order of magnitude under the two conditions. X!Tandem matched more peptides to spectra under low parent mass tolerance than high mass tolerance. For instance, searching against the same forward/reverse library under the all ion condition yielded 150 redundant sequences by +3-1 Da and 430 by ±100 PPM. From the same results, 26 and 168 reverse library hits were obtained under the Dalton and the PPM conditions respectively. Assuming false positive identification rates are equal to the reverse library identification rate, the lower parent mass tolerance setting yielded a disproportionately large number of false identifications. Though reverse library reporting only approximates an actual error rate, it is clear that the ±100 PPM setting decreases X!Tandem selectivity while increasing sensitivity.

The same peptide correlations and b/y ion assignments were obtained when the data was searched for all ion series and for the b and y series ions alone, but the b and y ion condition consistently reported them with higher confidence. Some low scoring peptides from the b/y series searches were absent from the all-ion searches. Poor scoring peptides from serum proteins identified with high confidence were rejected by searching for all ion series.

Library size affected scoring and sensitivity. Searching against the forward/reverse library gave identical spectra higher probability compared to searches against the forward library alone. For each sample condition, few or no reverse library identifications were obtained from forward/reverse X!Tandem or ProteinPilot searches. Reverse hits were most prevalent in acetonitrile supernatant data. X!Tandem assigned expectation values on the order of E-4 to repetitive glycine or alanine-rich peptides, mostly pentapeptides with m/z around 400. Searching against a combined forward and reverse library yielded fewer forward peptide identifications than searching under the same parameters against a forward library. Identical spectra were consistently assigned lower expectation values when searched against a forward library compared to the forward/reverse library. Some entries with expectation values in the order of E-4 or lower were ambiguous short sequences and many at E-2 were apparently false positives, containing repeats of the same amino acids or belonging to proteins not known to be in blood.

Spectra were manually inspected to ensure that high scoring ions had high signal-to-noise ratios. LC-ESI-Qq-TOF-MS/MS spectra with expectation values on the order of E-5 typically contained several ions well within 1 Da of the predicted b and y series, but also many assigned ions with intensities approaching baseline. Omitting peaks less than four times the baseline would increase expectation values but the majority of ions identified by X!Tandem were well above this threshold. High scoring peptides from well established blood components illustrate the reliability of peptide identifications from the Qq-TOF data of serum extracts (Table 3).

Protein\ Peptide Sequence	Calculated M+H (Da)	Delta (Da)	Expect (+3-1Da)	Expect (±100PPM)	Sample
albumin					
SALEVDETYVPKEFN	1741.8	0.022	9.5E-9	8.0E-9	**Supernatant**
FSALEVDETYVPKEFN	1888.9	0.025	3.7E-8	6.0E-8	**Super+NaOH**
PLVEEPQNLIKQN	1525.8	-0.0004	8.6E-5	4.4E-5	**Supernatant**
apolipoprotein A-I					
KVSFLSALEEYTKKLNTQ	2099.1	0.093	3.8E-5	1.9E-5	**AMHE**
		0.018	0.0034	0.0009	**Super+Formic+TFE**
apolipoprotein A-II					
FVELGTQPATQ	1190.6	0.039	9.4E-4	5.5E-4	**AMHE**
FVELGTQPAT	1062.5	0.02	0.0022	6.3E-4	**PBS**
VELGTQPATQ	1043.5	0.037	0.0067	0.0076	**AMHE**
apolipoprotein A-IV					
SLAELGGHLDQQVEEF	1771.9	0.02	4.1E-7	4.5E-7	**H_2O**
		0.065	2.0E-6	1.9E-6	**AMHE**
HLDQQVEEF	1144.5	0.042	0.091	0.051	**AMHE**
apolipoprotein E					
KVEQAVETEPEPEL	1597.8	0.045	0.019	0.0093	**AMHE**
fibrinogen alpha					
DSGEGDFLAEGGGV	1309.6	0.025	8.1E-9	2.2E-8	**PBS**
		0.015	9.1E-7	1.3E-6	**H_2O**
DEAGSEADHEGTHST	1542.6	0.0093	0.041	0.011	**Supernatant + TFE**
fibrinogen beta					
QGVNDNEEGFF	1238.5	0.037(PPM)	-	0.0002	**PBS**
		0.014(PPM)	-	0.0085	**H_2O**
		0.047(PPM)	-	0.013	**AMHE**
inter-alpha (globulin) inhibitor H4					
PGPPDVPDHAAYHPF	1616.8	0.023	1.2E-8	8.2E-9	**Supernatant**
QLGLPGPPDVPDHA	1395.7	0.014	1.4E-5	3.1E-5	**H_2O**
QLGLPGPPDVPDHAAYHPF	2011.0	0.062	1.5E-5	3.4E-6	**AMHE**

Table 4. Serum extract peptides assigned lowest expectation values by X!Tandem searched under ±100 PPM and +3-1 Da parent mass tolerances using a forward library. LC-ESI-MS/MS results correlated to the same sequences under the two conditions. Abbreviations used for the sample conditions are the same as for Table 2 and 3 with the exception of the following. Supernatant: Acetonitrile supernatant, Super+NaOH: Acetonitrile supernatant containing NaOH, Super+Formic+TFE: Acetonitrile supernatant containing formic acid and TFE.

Evaluation of Cleavage Sites in ProteinPilot Results

High scoring proteins were sorted according to their non-redundant peptides as determined by ProteinPilot and the sequences compared for trypsin cleavage (Table 5). Nearly 76% of the 1162 peptides in this group were non-tryptic while 22% were semi-tryptic. Less than 3% of the total number was fully tryptic.

Accession	Protein	Total	Tryptic	Semi-tryptic	Non-tryptic
4502027	albumin precursor	157	12	45	100
115583663	alpha-2-plasmin inhibitor	40	2	7	31
4502149	apolipoprotein A-II	17	0	3	14
71773110	apolipoprotein A-IV	86	2	29	55
4502157	apolipoprotein C-I	13	0	2	11
32130518	apolipoprotein C-II	7	0	2	5
4557323	apolipoprotein C-III	21	0	3	18
4557325	apolipoprotein E	25	1	10	14
4502165	apolipoprotein F	13	0	2	11
4503635	coagulation factor II	33	1	8	24
119395709	coagulation factor XIII A1	28	0	7	21
73858570	complement component 1 inhibitor	41	0	6	35
115298678	complement component 3	133	3	28	102
67190748	complement component 4	155	4	35	116
67782358	complement factor B	25	0	5	20
4503689	fibrinogen alpha	74	0	17	57
70906435	fibrinogen beta	22	0	5	17
31542984	inter-alpha (globulin) inhibitor H4	106	3	19	84
4505881	plasminogen	29	0	3	26
50363221	serine (or cysteine) proteinase inhibitor (antitrypsin)	61	0	6	55
4557871	transferrin	58	1	9	48
4507725	transthyretin	18	0	1	17
Total		1162	29	252	881

Table 5. Tryptic and non-tryptic peptides identified by ProteinPilot from forward libraries in both forward and forward/reverse searches. All entries for known serum proteins identified with high confidence were designated fully tryptic, semi-tryptic or non-tryptic.

Modifications of MS/MS Analytes Detected by ProteinPilot and X!Tandem

The Paragon algorithm identified cysteinylation, deamidation, dehydration, formylation, methylation, oxidation, and phosphorylation of peptides with confidence exceeding 90% (Table 3). Many peptides were identified with multiple modifications. For example the albumin sequence PLVEEPQNLIKQN was identified in nine result files with four deamidations at 99% probability and scores ranging from 16 to 21. ProteinPilot revealed trends in oxidation among peptides. For example, the complement C4 sequence TLEIPGNSDPNMIPDGDFN was reported to be oxidized specifically at the internal methionine as well as the tenth proline residue but at no other sites including the two remaining prolines. These two oxidations were observed in acetonitrile-methanol-water ultrafiltrate and PBS extracts and the unmodified peptide was also present. Paragon identified the fibrinogen alpha phosphopeptide $D_PSGEGDFLAEGGGV$ in the extract containing acetonitrile, methanol, water and TFE (Table 3, Fig 8).

Detection of Internal Standards

Using MALDI-Qq-TOF-MS, [Glu^1] fibrinopeptide B and Angiotensin 1 were detected in spiked serum at nanomolar concentration, or in amounts in the femtomole range (Fig 9). These appeared as ions with well resolved isotopic envelopes among a series of serum analyte peaks. At the lowest concentration tested both standards were above the limit of detection although angiotensin 1 was barely resolved from the baseline. The practical detection limit for angiotensin was somewhere between 100 and 250 femtomole per microlitre. Internal standards did not appear to depress ionization of the

endogenous peptides as an untreated control yielded ions with the same masses at comparable intensities.

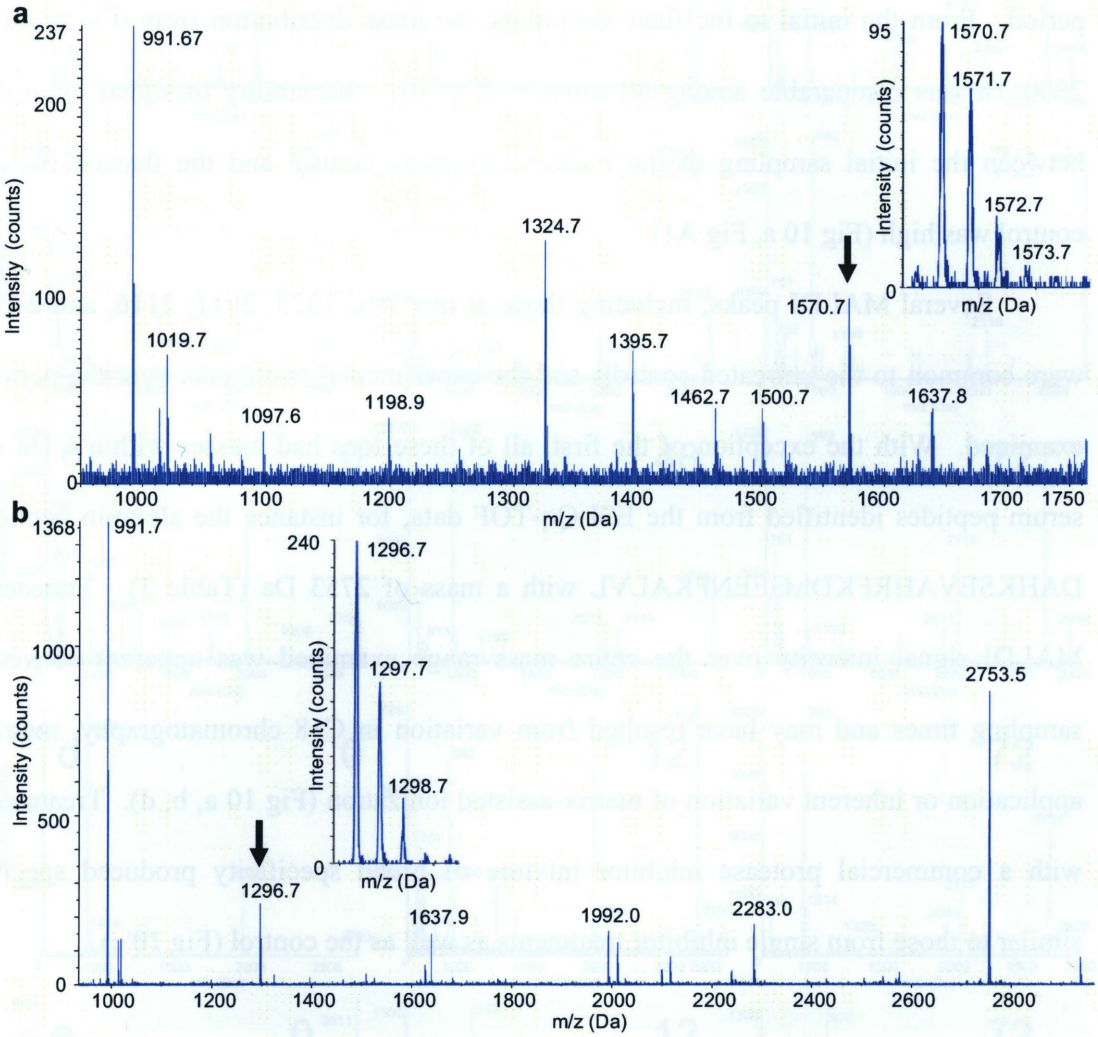


Figure 9. Internal standards in serum prepared by C18 reversed-phase and analyzed by MALDI-Qq-TOF-MS/MS. The instrument detected 250 femtomoles of angiotensin 1 (a) and 100 femtomoles of [Glu^1] fibrinopeptide B (b) well above baseline. The arrows indicate the molecular ion of each standard.

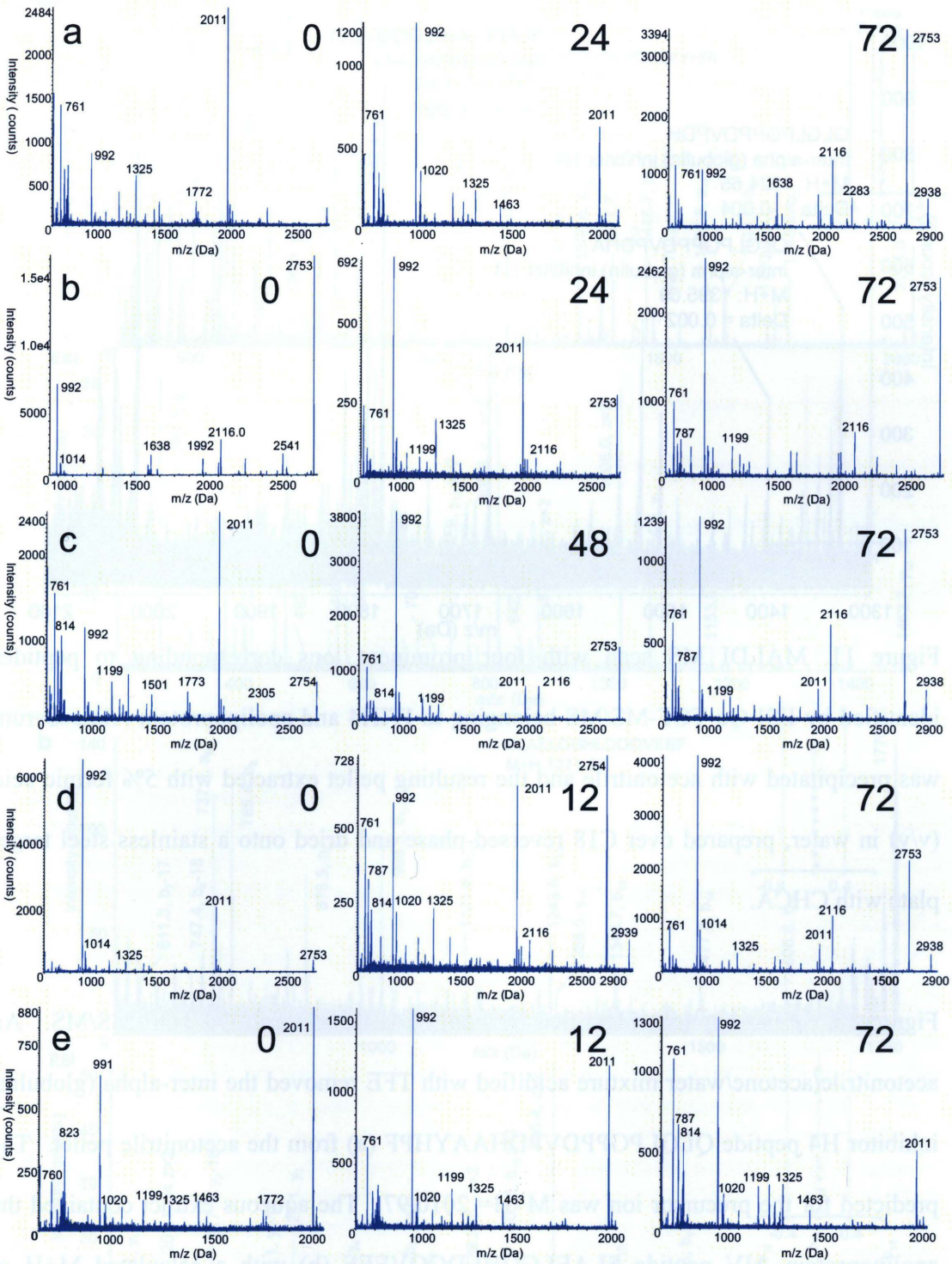
Effects of Protease Inhibitors on MALDI-MS

There was no clear pattern in the MALDI-MS spectra of serum incubated at ambient temperature with small molecule and protein protease inhibitors over a 72 hour period. From the initial to the final samplings, the mass distribution from 0 to around 2800 Th was comparable among all samples (Fig 10). Variability in signal intensity between the initial sampling of the room temperature control and the thawed frozen control was high (Fig 10 a, Fig A1).

Several MALDI peaks, including those at m/z 991, 1325, 2011, 2116, and 2753 were common to the untreated controls and the experimental treatments over the period examined. With the exception of the first, all of these ions had masses within 1 Da of serum peptides identified from the ESI-Qq-TOF data, for instance the albumin peptide DAHKSEVAHRFKDLGEENFKALVL with a mass of 2753 Da (Table 3). Transient MALDI signal intensity over the entire mass range examined was apparent between sampling times and may have resulted from variation in C18 chromatography, matrix application or inherent variation of matrix-assisted ionization (Fig 10 a, b, d). Treatment with a commercial protease inhibitor mixture of broad specificity produced spectra similar to those from single inhibitor treatments as well as the control (Fig 10 e).

Figure 10. MALDI-Qq-TOF MS time series of serum treated with protease inhibitors and stored at room temperature for a 72 hour period. Serum was left untreated (a) or was treated with (b) AEBSF, (c) bestatin, (d) benzamidine or (e) a commercial protease inhibitor mixture and prepared over C18 reversed-phase. The sampling times, in hours, are indicated in the top right hand corner of each spectrum.

Figure 10.



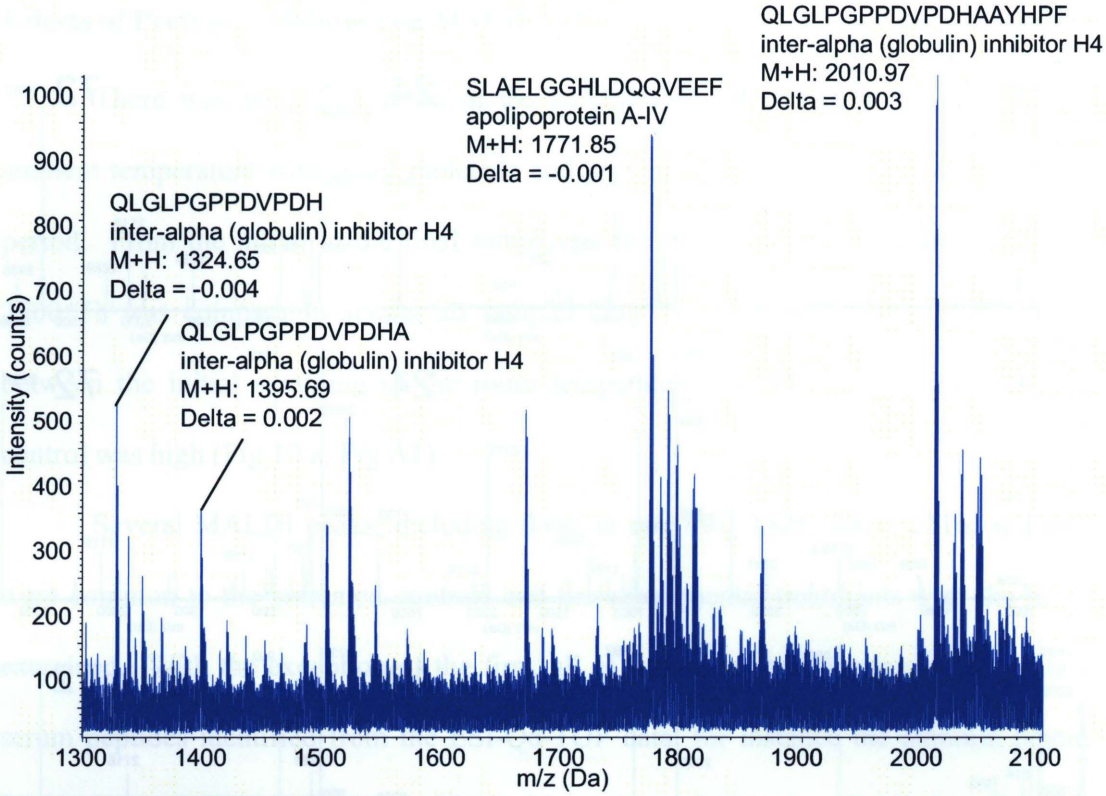
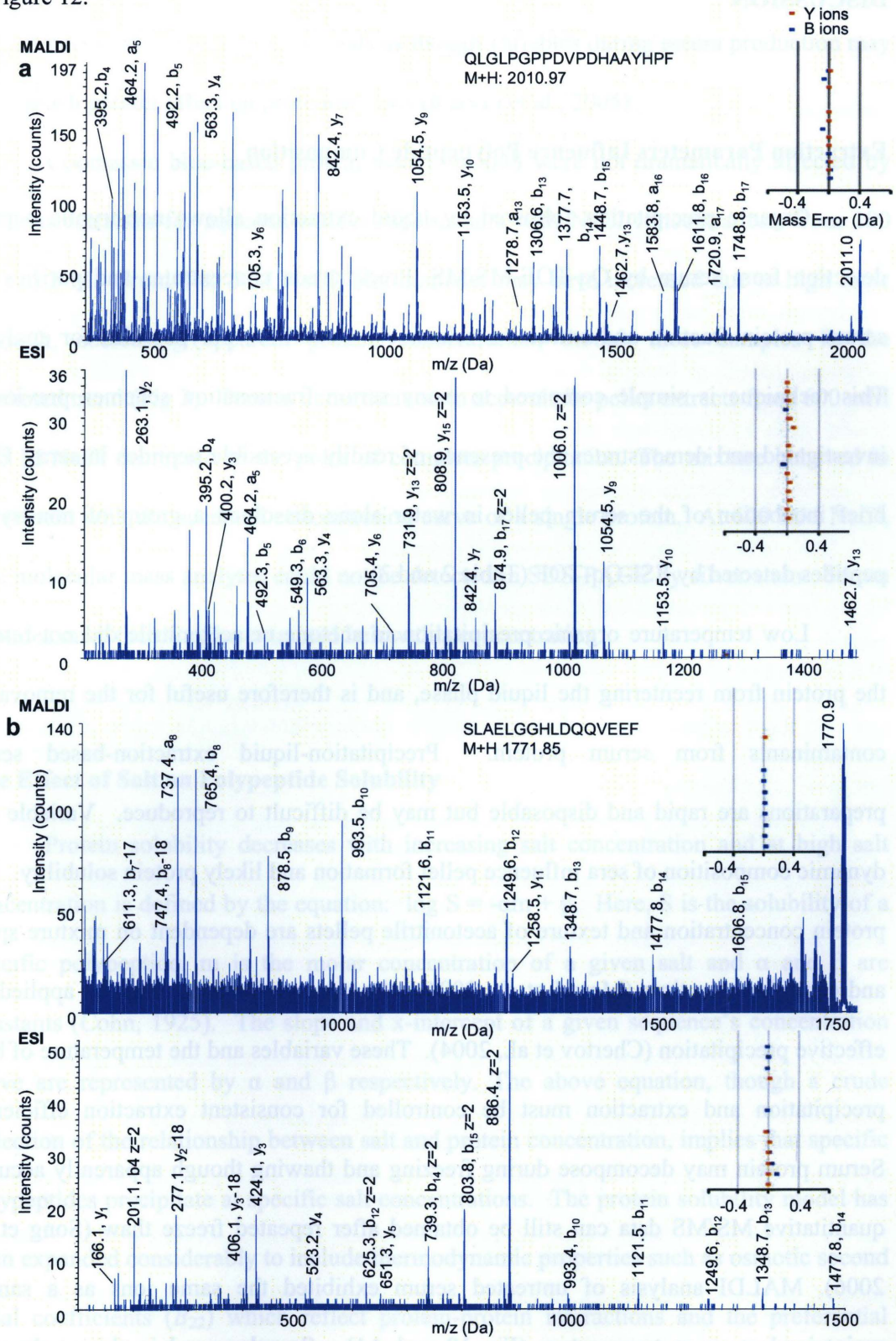


Figure 11. MALDI MS scan with four prominent ions corresponding to peptides identified by ESI-Qq-TOF-MS/MS belonging to ITIH4 and apolipoprotein AIV. Serum was precipitated with acetonitrile and the resulting pellet extracted with 5% formic acid (v/v) in water, prepared over C18 reversed-phase and dried onto a stainless steel target plate with CHCA.

Figure 12. Serum peptides identified by both MALDI and ESI-Qq-TOF MS/MS. An acetonitrile/acetone/water mixture acidified with TFE removed the inter-alpha (globulin) inhibitor H4 peptide QLGLPGPPDVPDHAAYHPF (a) from the acetonitrile pellet. The predicted for the precursor ion was M+H= 2010.97. The aqueous extract contained the apolipoprotein AIV peptide SLAELGGHLDQQVEEF (b) with a calculated M+H of 1771.85.

Figure 12.


MALDI
a
QLGLPGPPDVPDHAAYHPF
M+H: 2010.97
Y ions
B ions
Mass Error (Da)
Intensity (counts)
m/z (Da)
ESI
MALDI
b
SLAELGGHLDQQVEEF
M+H 1771.85
ESI

DISCUSSION

Extraction Parameters Influence Polypeptide Composition

Organic precipitation followed by liquid extraction allows non-tryptic peptide detection from serum by Qq-TOF-MS/MS. Precipitation concentrates and purifies the serum protein fraction and extraction selectively solubilizes polypeptides for analysis. This technique is simple compared to many serum fractionation schemes previously investigated and demonstrates the presence of readily accessible peptides in sera. Even brief incubation of the serum pellet in water alone dissolved a group of non-tryptic peptides detected by ESI-Qq-TOF (Table 2 and 3).

Low temperature organic precipitation in acetone or acetonitrile did not restrict the protein from reentering the liquid phase, and is therefore useful for the removal of contaminants from serum protein. Precipitation-liquid extraction-based serum preparations are rapid and disposable but may be difficult to reproduce. Variable and dynamic composition of sera influence pellet formation and likely protein solubility. The protein concentration and texture of acetonitrile pellets are dependent on mixture speed and centrifugation time (Merrel et al., 2004). An excess of solvent must be applied for effective precipitation (Chertov et al., 2004). These variables and the temperature of both precipitation and extraction must be controlled for consistent extraction efficiency. Serum protein may decompose during freezing and thawing though apparently accurate quantitative MS/MS data can still be obtained after repeated freeze thaw (Song et al., 2006). MALDI analysis of untreated serum exhibited the same ions as a sample maintained at room temperature (Fig 10 and A1). Samples used in this study were

thawed, aliquoted and refrozen once prior to use and disposal. Clinical serum samples are not necessarily frozen prior to analysis though variables during serum production may have much greater effect on proteomic data (Karsan et al., 2005).

Coomassie blue-based protein assays in SDS were not dramatically affected by salt concentration for aqueous extract supernatants with NaCl concentrations from 100 mM to 2 M. From 3 to 5 M NaCl, blot diameter may have increased due to high ionic strength, decreasing apparent protein concentration, though these samples fit the predicted trend (Fig 3). Protein concentration in acetonitrile pellet extracts from 600 mM to 5 M NaCl decreased to the point of complete precipitation. The mixture behaved in agreement with the accepted concentration curve of a single protein. At 600 mM NaCl, low molecular mass analytes could not be detected in SDS-PAGE by silver stain. Serum peptides must therefore be extracted below this concentration.

The Effect of Salt on Polypeptide Solubility

Protein solubility decreases with increasing salt concentration and at high salt concentration is defined by the equation: $\log S = -\alpha m + \beta$. Here, S is the solubility of a specific polypeptide, m is the molar concentration of a given salt and α and β are constants (Cohn, 1925). The slope and x-intercept of a given sequence's concentration curve are represented by α and β respectively. The above equation, though a crude reflection of the relationship between salt and protein concentration, implies that specific polypeptides precipitate at specific salt concentrations. The protein solubility model has been expanded considerably to include thermodynamic properties such as osmotic second virial coefficients (B_{22}) which reflect protein-protein interactions and the preferential

binding parameter ($\Gamma_{23}^{(m)}$) a limit which describes protein-solvent-cosolvent interactions (Ruckenstein and Shulgin, 2006, Neal et al., 1998). Two notable equations incorporating these parameters separately were derived by Ruckenstein and Shulgin, demonstrating relationships between B_{22} and concentration at low solubility and between $\Gamma_{23}^{(m)}$ and phase change at low cosolvent concentration (Ruckenstein and Shulgin, 2006). Using measured B_{22} and $\Gamma_{23}^{(m)}$ values for lysozyme, predicted "salting in" and "salting out" conditions for specific proteins in aqueous NaCl and $MgCl_2$ solutions were shown to be in agreement with experimental results. Furthermore, at low salt concentration, the curve of the solubility versus mole fraction function can be predicted (Ruckenstein and Shulgin, 2006). It is feasible that an extraction system designed with the use of such equations would solubilize a set of peptides from organic pellets while confining others to the solid phase. Serum polypeptides do not differ enough in their solubility curves to exclude the majority of the protein fraction from aqueous NaCl extracts while retaining certain analytes (Table 1, Fig 4 lane 6).

While peptide or protein enrichment based on salt conditions may offer low selectivity, it should help exclude erroneous peptide or protein identifications from MS/MS data if a solubility cutoff based on sample extraction condition is applied to the FASTA protein library searched. Experimental validation of this cutoff and of concentration dependent solubility for a group of known serum proteins may enable the evaluation of MS/MS peptide results with increased scrutiny. It should be noted that the equations mentioned have been applied to the study of pure proteins with molecular masses exceeding 14kD, while the extent to which complex mixtures such as serum deviate from current models is unknown. Protein solubility in solutions containing

multiple solvents, such as the aqueous extracts with acetonitrile, methanol and TFE employed in this study, requires further characterization.

The Effect of Solvents on Polypeptide Solubility

Organic cosolvents decrease protein concentration in aqueous solutions by decreasing the dielectric constant of the system and displacing water from the protein surface (Arakawa and Timasheff, 1985). This principle is applied in chromatography, where polar C18 reversed-phase columns are washed with acetonitrile to elute bound protein. For acetone and acetonitrile pellets extracted in aqueous solutions of increasing acetonitrile concentration, protein concentration followed the accepted trend in solubility. Biphasic liquid extraction is an effective technique for polypeptide separation and can be highly selective. Hydrophobic solvents can separate polar and non-polar polypeptides in mixtures for direct MALDI-MS analysis (Kjellström and Jensen, 2003). Both low and high molecular weight proteins were extracted with butanol as a precipitating solvent and these were comparable in distribution to ethanol supernatants by SDS-PAGE (Tucholska et al., 2007). A precipitation to selectively remove non-polar peptides, using ethyl acetate or butanol for instance, could improve resolution in subsequent separations.

The preferential binding parameter is more useful than B_{22} for the prediction of protein solubility in organic-aqueous solutions, with the organic cosolvents of focus being urea, glycerol, glycine or polyethylene glycol. The atypical solubility behavior of beta-lactoglobulin in glycine-buffered aqueous solutions of varying pH suggests protein mixtures like serum can be preferentially extracted (Arakawa and Timasheff, 1987). The organic solvents employed in the present study have negligible buffering capacity but

considerable impact on the behavior of water molecules. This could account for the relative equivalency in the number and type of peptides observed in acetonitrile supernatants with and without NaOH or acid (Table 2 and 3). An analyte that elutes by reversed-phase or ion exchange chromatography at a given acetonitrile or salt concentration respectively, could be stripped from an organic pellet under similar conditions. This method permitted the determination of known serum proteins by MALDI-Qq-TOF (Fig 11), and may improve resolution and run time of LC-ESI-MS/MS.

Surfactants and Non-Polar Solvents

The solutions used for liquid extraction in this study were free of surfactants, which impair peptide analysis by inhibiting trypsin digestion or ionization (Zhang and Li, 2004, Steen and Mann, 2004). Surfactants are particularly useful for solubilizing hydrophobic compounds though other conditions can be controlled to extract these. Peptides from apolipoproteins, which circulate with extensive lipid modification, were soluble in aqueous and organic solutions. Peptides presented in this serum study are not particularly hydrophobic (Table 2 and 3). Non-polar solvents can be used in either precipitation or extraction to partition non-polar species. For mass spectrometric work on hydrophobic proteins, non-polar solvents are preferable to surfactants.

Diagnostic Paradigms of Peptide Identification and Pattern Profiling Can be Unified

The mass distribution and intensity of serum peptide ions observed by MS were suggested to be applicable to disease diagnosis with the use of mathematical

interpretation (Petricoin et al., 2002, Ornstein et al., 2004). These MS profiling diagnostics were shown to be biased by artifacts of sample handling and unreliable for differentiating disease from normal states (Diamandis, 2003 and 2006). *Ex vivo* endo- and exopeptidase activity forms peptide patterns in serum (Marshall et al., 2003, Koomen et al., 2005). Patterns of relative ion intensity are difficult to reproduce by MALDI, and this variability must be addressed before a discriminatory method based on this principle can be successful. Accurate ratiometric measurement of peptide abundance depends on a reliable control signal in the MS data, a feature absent from many previous studies. A known analyte whose signal intensity can be controlled within a known range would represent such a control.

Though ion intensity patterns may elude reliable interpretation, accurate m/z values can be correlated to specific biomolecules with defined statistical certainty, provided they exceed a minimum signal-to-noise ratio. As more peptides are identified from blood fluids, a data analysis strategy incorporating MS/MS search and pattern recognition functions might reveal new patterns in peptide or protein distribution among diseased populations. A proteomic or peptidomic serum profile including a group of known peptides may constitute a diagnostic tool particularly if combining detection or quantification of a group of low significance biomarkers in a single sample discriminates physiological states. Diagnostics based on MS patterns can include the simultaneous determination of ion identity. Interestingly, such studies identify many peptides with sequences partially matching those in Tables 2 and 3 as being responsible for cancer dependent patterns although relative ion intensity in the absence of an added internal standard was used as the criterion for measuring their abundance (Villanueva et al.,

2006). The great diversity and apparent transience of peptides from known serum proteins observed here and in other studies may explain why poor reproducibility continues to plague MS profiling diagnostics. Serum peptides extracted in aqueous and aqueous/organic conditions were consistently detected by ESI-Qq-TOF, suggesting they can be reliably quantified by multiple reaction monitoring (Domon and Aebersold, 2006).

Combined ESI MS/MS and MALDI MS for the Characterization of Serum Peptides

Peptides observed by LC-ESI were compared to those apparent in the MALDI-Qq-TOF spectra. Many of the peptides detectable as +1 ions in the MALDI profile could be identified by LC-ESI-Qq-TOF. This result suggests that many peptides can be detected and identified by ESI MS/MS that cannot be readily identified by MALDI MS/MS. MALDI peptide ions had comparable parent mass accuracy to those observed by ESI, but relatively few peptides were identified from MALDI MS/MS by either algorithm used. Since the peptides we observed are not tryptic and many do not contain basic amino acid residues, these may not ionize or fragment as efficiently by MALDI compared to ESI-MS/MS.

MALDI MS peaks can be attributed to specific peptides by comparing their masses to those observed in separate sequencing experiments, provided mass accuracy within 1Da (Richter et al., 1999). The high throughput of ESI and its compatibility with MS/MS search algorithms make it ideal for the collection of reference data. The serum extract dataset included peptide ions common to MALDI MS and ESI data (Fig 10 and 11). Many of these were detected in similar sample preparations by both methods, for instance the same ITIH4 peptide ions were present in aqueous extracts (Table 2 and 3,

Fig 11). High precursor ion accuracy obtained using both ion sources and high product ion mass accuracy of the ESI data increase confidence of the identifications (Fig 11). From the MALDI-MS/MS dataset, X!Tandem identified apolipoprotein AIV and ITIH4 peptides with parent ions of m/z 1771 and 2010 respectively (Fig 12) confirming their presence in MALDI MS scans. Automated screening of MALDI MS data against ESI data has been developed but an extensive library specific to blood peptides is needed (Zimmerman et al., 2005). Construction of an extensive ESI-MS/MS database of serum peptides and its application to MALDI-MS is feasible using Qq-TOF and SQL databases. Analyte primary structure, sample preparation, and ionization conditions all contribute specific ionization characteristics, with consequences for sequence coverage in MS analysis. An MS/MS library consisting of spectra from overlapping sequences of the entire proteome would likely require all proteins to be isolated to near purity for analysis. Alternatively, a spectral database built from MS/MS of serum extracts, ultrafiltrates and chromatography could be assembled relatively quickly but would be far less complete. Data processing speed is determined in part by the size of the FASTA library being screened. An abbreviated library including only biomarker proteins or MS/MS spectra of their commonly observed peptides would be many times smaller than the RefSeq database (Pruitt et al., 2007).

X!Tandem and ProteinPilot Interpretation of Qq-TOF Tandem Mass Spectra

Peptides from known blood proteins were identified with confidence from E-9 to E-2 or as high as 99% by X!Tandem and ProteinPilot, respectively. Spectra scoring above E-2 were rejected, as they were likely spurious and their contribution to the overall

confidence of protein identification was negligible. Some of these sequences were also identified with high probability and are shown in Tables 2 and 3. Low and high scoring identifications from a common sequence can strengthen confidence in a single protein match, a feature exploited by the Pro Group algorithm (Shilov et al., 2007). ProteinPilot results for individual peptides discussed below had minimum probabilities of 95%, though Pro Group included peptides below this threshold in the protein scores.

A variety of cleavage sites among peptides from common sequences suggest a complex degradative process precedes mass analysis, one that can be monitored by Qq-TOF. Multiple related peptides from the same short spans of their precursor proteins, differing in length by single and multiple amino acid residues, were obtained from single extractions (Tables 2 and 3). These are consistent with an initial cleavage of peptides from whole proteins followed by amino- and carboxypeptidase activity on these (Marshall et al., 2003). Such patterns are heavily influenced by sample preparation conditions (Yi et al., 2007). Tryptic peptides apparently are differentially susceptible to decomposition prior to LC-MS analysis and the same may be true of any peptide (Alves et al., 2008). A serum extract sample may therefore contain peptides that cannot be definitively associated with a protease.

Apolipoprotein

Non-redundant sequences from seven apolipoprotein isoforms were detected by LC-ESI-Qq-TOF in all samples except the dye affinity column eluate. Apolipoproteins have relatively low serum concentration by mass, but their molar serum concentration is high (Hortin, 2006). The apolipoproteins reported by both algorithms are between 9 and

46 kDa in mass. Compared to albumin, with a mass of 67 kDa, short apolipoprotein sequences are less likely to be falsely identified as their set of theoretical fragments is relatively small. Apolipoprotein AII has been shown to be over expressed in prostate cancer tissue and serum concentrations of apolipoprotein CII and CIII differ between normal and pancreatic cancer populations (Malik et al., 2005, Chen et al., 2007). Supernatants from serum precipitated in organic solvent containing TFE were suggested to contain cancer-dependent concentrations of apolipoprotein AII by SELDI-TOF and LC-MS/MS (Chertov et al., 2004). Interestingly, this species is apparently downregulated in a B cell lymphoma mouse model, though detailed explanation of this result was lacking.

Serum Protease inhibitors

Protease inhibitors were detected by MS/MS in acetonitrile supernatants and ultrafiltration eluates as well as aqueous and organic extracts (Table 2 and 3). ITIH4, Complement C1 inhibitor and serine (or cysteine) proteinase inhibitor were all dissolved under multiple extraction conditions.

Plasma concentrations of inter-alpha inhibitors range from 0.6 to 1.2 mg/ml and this glycoprotein family has been studied by SELDI and LC-Q-TOF MS/MS (Josic et al., 2006). ITIH4, the fourth inter-alpha inhibitor heavy chain to be discovered, has a unique 300 residue C-terminal sequence and lacks the C-terminal covalently bound bikunin moiety that contributes to the inhibitory properties of this group (Saguchi et al., 1995). All ITIH4 peptides detected in the serum extracts were within this 300 residue region, from either the 3140 Da sequence NFRPGVLSSRQLGLPGPPDVPDHAAYHPF from

position 659 to 687 or the 3155 Da sequence NVHSGSTFFKYYLQGAKIPKPEASFSPR from to 617 to 644 of the 930 amino acid sequence matched by BLAST (Table 2 and 3) (Altschul et al., 1997). Bonding between inter-alpha inhibitor heavy chains and bikunin as well as the glycosaminoglycan hyaluronan increases their mass and alters their stability and this may explain their absence in the MS/MS data. Inter-alpha heavy chains can be dissociated from glycosaminoglycan by alkaline treatment such as 50 mM NaOH at 23°C (Josic et al., 2006). Factors other than this post-translational modification could have prevented the detection of other ITIH species because serum precipitated or extracted in the presence of NaOH did not yield hits for members of this group, although low temperature was used and the detrimental effect of NaOH on MS/MS is noted below.

Inter-alpha inhibitors have been implicated in sepsis and cancer progression (Lim et al., 2003, Himmelfarb et al., 2004, Fung et al., 2005, Song et al., 2006). Peptides from the C-terminal region of ITIH4 have been measured in serum from normal, cancer and diabetes populations by immunoaffinity SELDI-TOF, and were found to be variable (Fung et al., 2005, Song et al., 2006). Many of the same ITIH4 peptides were in the serum extracts studied here. Peptides at m/z 3157, 2270, 2626, 2723, 3027 and 3140 were identified by ProteinPilot with 99% confidence, with and without modification (Table 3). These had the same predicted sequences as reported previously, and were contained within the two regions mentioned above (Song et al., 2006). X!Tandem identified the same sequences at m/z 3157 and 2270, with expectation values above E-4. According to both algorithms, serum extracts contained additional ITIH4 peptides from those described as potential biomarkers (Table 2 and 3). The diversity of peptides observed from two short portions of the ITIH4 sequence suggests the kallikrien-mediated

proteolysis of the molecule is followed by non-specific cleavage (Nishimura et al., 1995). ITIH4 and antitrypsin peptides produced exceptional spectra and could be used to evaluate MS/MS performance in serum analysis (Fig 12, Table 2 and 3). To test the clinical utility of protease-derived patterns in serum, measurement of serum protease inhibitor concentration or a protease-to-inhibitor ratio should be possible using Qq-TOF.

Serum Proteins Identified from the Reverse Library

Complement C1 inhibitor and serine (or cysteine) proteinase inhibitor had forward and reverse library hits but their forward identifications were not rejected on this basis. In the case of Complement C1 inhibitor, of all reverse library associations only one spectrum scored higher than 25% probability, a quadruply-modified 34 residue sequence with only 60% confidence. The length of this sequence alone casts doubt on its validity. By comparison, the forward library matches for this protein included 11 peptides with scores exceeding 90% and 7% sequence coverage above 95% confidence. One reversed antitrypsin peptide with four modifications scored 76% confidence, the remainder scored well below 50%. These cases exemplify the independence of true positives on reverse library results. A probability threshold of E-2 established by manual inspection of forward library results effectively screened the data.

Evaluation of Reverse Search Results

ProteinPilot identifications of mucin peptides were likely false and demonstrate that accuracy of probability-based identification depends on sequence length as well as

library length (States et al., 2006). Peptides from mucin isoforms were identified by ProteinPilot from both the forward and reverse library. The isoforms identified vary in length from 4493 residues for mucin 17 to 14507 residues for mucin 16. As with albumin, the lengths of these sequences contribute to false identification of these unlikely peptides in both the forward and reverse libraries. All forward and reverse mucin 19 sequences with scores from 90 to 99% were high in threonine, glycine, and serine residues. These were distributed throughout the sequences with the exception of pairs of threonine residues. More than twice the number of mucin 19 peptides was reported from the forward compared to the reverse library within this confidence range, suggesting that false positive rate estimation based on reverse library identifications does not accurately reflect the true rate. Criteria for accepting peptide and protein assignments can be established from forward only searches as an alternative to determining a false positive rate by reverse library searching. Correct sequences from ProteinPilot and X!Tandem results were therefore manually differentiated from false positives on the basis of protein identity, spectral features and sequence characteristics.

More than half of the reverse peptides found in the extract data by X!Tandem under two parent mass accuracy settings were from MS/MS of basic aqueous extractions, while neutral and acidified extracts gave few reverse hits. With few exceptions, forward hits from this sample pool had poor scores (Table 2). These findings indicate that sodium hydroxide in the extracts had a negative effect on spectral quality (Table A1).

MS/MS Search Parameters Influence Peptide Results

ProteinPilot identified more forward library sequences when an identical dataset was searched against forward and forward reverse human FASTA libraries. X!Tandem assigned better scores using a forward only library. It is conceivable that the increased processing burden of screening a double library causes both algorithms to function below their optimal performance. Consequently, experimental and control searches should be conducted separately. The drawback to such a strategy is that it eliminates competition between the target and control libraries for the assignment of spectra, creating bias in forward identifications (Elias and Gygi, 2007). Though sequence reversal is not a random process, reverse library results are generally accepted as allowing the calculation of accurate false positive discovery rates (Kapp et al., 2005). Randomized lists identical in size to the forward list maintaining the representative protein sequence lengths and number of each amino acid can also be used (Ramos-Fernandez et al., 2008). Though randomized and reversed libraries have been claimed to provide similar estimated false positive rates a standard decoy design has yet to be implemented (Qian et al., 2005).

The stringencies of the two X!Tandem parent mass tolerances used here depend on analyte mass. For a 1 kDa peptide, 100 PPM represents a mass of 0.1 Da. A setting of $\pm$100 PPM becomes less stringent than the +3-1 Da condition for sequences exceeding 10 kDa, though this is beyond the mass range studied and precursor ions identified with this level of mass accuracy would be manually rejected. Accuracy provided by the mass spectrometer is directly linked to search algorithm performance (Brosch et al., 2008). As the range including the true mass of a given ion decreases so does the number of possible sequences that fit this range. Consequently fewer candidate sequences must be evaluated

by the algorithm. Mass stringency settings in search algorithms obey this pattern. At low stringency, specificity improves and sensitivity decreases whereas high stringency increases both false and correct identifications (Brosch et al., 2008). This concept is supported by results of the X!Tandem parent mass tolerance experiment presented here. Under the strict ±100 PPM condition, X!Tandem identified more proteins and more false positives than under the +3-1 Da condition. An exceptionally high number of reverse sequence identifications were observed compared to a previous study using the ±100 PPM condition (Brosch et al., 2008). This was likely due to other operational variables such as the use of non-tryptic data and a reverse library as opposed to a tryptic dataset with a randomized library.

When X!Tandem results from b/y and all-ion series were compared, the advantage of detecting the a/x and c/z ion series was shown to be offset by an overall decrease in expectation values. Using the same raw data, ProteinPilot built a complementary peptide list, reporting b and y ions with and without a number of biological and chemical modifications. Based on these observations b/y ions are sufficient for correct spectral interpretation.

Modified Serum Peptides

ProteinPilot reported a variety of chemical modifications which were independent of probability scores (Table 3). While some of these modifications are likely artifacts of sample handling and ionization, physiological modifications such as phosphorylation may be of clinical use. The phosphorylated fibrinogen alpha sequence $D_PSGEGDFLAEGGGV$ identified by Paragon and shown in figure 8 has biological

consequences. Fibrinogen alpha and fibrinopeptide A are known to circulate and phosphorylation at this residue is well documented (Adkins et al., 2002, Chan et al., 2004, Marshall et al., 2004). Fibrinopeptide A moderates the acute phase reaction and fibrinopeptide A phosphorylation increases during this inflammatory process (Henschen-Edman, 2001, Masuda and Sugiyama, 2001, Seydewitz and Witt, 1985). A fibrinopeptide A phosphopeptide has been observed in serum by MALDI-TOF previously (Williams et al., 2006). Fibrinopeptides released into circulation during myocardial infarction have been studied by LC-ESI-ion trap (Marshall et al., 2003). Other participants in the acute phase reaction have been identified as non-specific cancer markers by MS (Juan et al., 2004).

X!Tandem labeled peaks in some spectra as b and y series ions with modifications, though expectation values were not greatly affected by these assignments. Multiple peaks with m/z corresponding to 17 or 18 Da shifts, due to hydrolytic cleavage of the peptide C-terminus, were assigned with mass accuracy around 0.1 Da. In general, searching for all modifications allowed by X!Tandem provided no additional information about the proteins extracted or physiological modification.

Modifications are often found by comparing observed ion mass to the sum of a predicted mass and the mass of a known modification. An alternate approach to locating modifications involves probing spectra for fragments with mass shifts of any magnitude then identifying the modification responsible. MS-Alignment is a naïve searching program which aligns multiple spectra to elucidate which fragment, and ideally which residue, has a different mass from the unmodified version (Tsur et al., 2005). Such a search strategy may not be feasible for screening non-tryptic peptides against a whole

species library, but would certainly require many repetitions to build a reliable MS/MS database. The number of possible sequence-modification candidates from an undigested proteomic sample might increase search space to an unmanageable degree. That a number of related and identical sequences were observed from the extracts is favorable for the MS-Alignment approach to serum peptides modification analysis (Table 2 and 3).

Post-translational modification is essential for the regular function of many proteins. Phosphorylation is a complex process important in signalling and is strictly regulated. Uncontrolled kinase activity is associated with cancer (Blume-Jensen and Hunter, 2001). Phosphorylation can be monitored using several technologies but MS based systems alone do not require complete phosphoprotein purification (Mann et al., 2002).

A number of biological properties challenge the detection and quantification of protein phosphorylation. High basal phosphorylation, around a third of all eukaryotic proteins, is a barrier to phosphoproteomic mining of disease samples by MS (Mann et al., 2002). Peptides in the serum extracts may be dephosphorylated by phosphatases after sampling so inhibitors must be added during blood collection. Genomic and proteomic data mining for phosphorylation sites combined with phosphorylation targeted Western blot can identify candidate biomarkers. Screening candidate sites using MS adds high throughput and resolution. The amino acids serine, threonine and tyrosine are phosphorylated at levels differing over four orders of magnitude suggesting their signalling strengths may be inversely proportional to their concentrations. If minute changes in tyrosine phosphorylation have dramatic physiological consequences, the sensitivity of MS may make it uniquely suited to phosphorylation-based diagnostics.

MALDI-Qq-TOF Analysis of Serum Treated with Protease Inhibitors

In vivo, peptides are created and removed from circulation through a cascade involving numerous proteases of varying specificity. The clotting cascade is an example. Truncation of proteins produces peptides which are then degraded into successively smaller fragments. Renal filtration removes molecules smaller than 45 kDa and is maximal for those below 5 kDa (Anderson and Anderson, 2002, Werle and Bernkop-Schnürch, 2006). The exopeptidases, divided into amino- and carboxypeptidases, therefore regulate peptide-mediated activities and plasma homeostasis.

The effect of proteases on peptide formation following serum preparation was examined by MALDI-Qq-TOF. A number of ions with m/z between 700 and 2800 were common to spectra from the controls, from the point inhibitors were added and the time points sampled. Signal intensity of the common ions varied by three orders of magnitude and was independent of incubation time and the inhibitor used (Fig 10). This suggests inherent variability in MALDI peak intensity and m/z distribution can confound meaningful interpretation of MS serum peptide profiles. Over the period studied, serum protein degradation appeared to reach a steady state regardless of protease inhibition. Progressive cleavage produces many low molecular mass polypeptides with a corresponding decrease in larger species. This would be observed by MS as a shift in ion distribution with increasing numbers of low m/z ions, however at 72 hours the control had more peaks above m/z 2500 than at the initial sampling and their signals were among the highest.

A battery of eight protease inhibitor conditions failed to maintain the distribution of serum peptides over a 72 hour period more than untreated controls at room

temperature and -20°C (Fig 10). This finding is in agreement with the observation that proteolysis begins at collection and has progressed considerably by the time serum is made. Addition of metal chelating protease inhibitors during blood collection modulates ex-vivo degradation, which suggests that collection is a critical point in the control of diagnostic blood peptide characteristics (Yi et al., 2007).

In 2002 the Human Proteome Organization (HUPO) began the Plasma Proteome Project (PPP), a large scale international effort to describe the polypeptides of blood (Omenn et al., 2005). The study employed MS/MS with multiple separatory techniques to characterize pooled blood specimens. Plasma augmented with EDTA or citrate or heparin and unadulterated serum were prepared from the PPP blood samples. Almost half of the serum peptides were not found in the plasma samples, providing further evidence that protease inhibition cannot control peptide formation during serum preparation (Tammen et al., 2005).

The serine protease inhibitors PMSF and AEBSF affect circulatory enzymes such as kallikrein, plasmin and thrombin and therefore limit fibrinopeptide formation. Fibrinopeptide ions at m/z 2011 appeared in PMSF samples at the first sampling and in the AEBSF sample after 8 hours. It is likely that this fragment was present in the latter sample but did not ionize as AEBSF is an irreversible inhibitor should stop protease activity permanently (Fig 10 b). In the presence of water, PMSF has a half life around one hour at physiological pH and to account for its diminishing efficacy, it was used at two concentrations well above the minimum inhibition. These conditions apparently did not impact ionization as serum supersaturated with PMSF produced intense ions by MALDI-MS (not shown). AEBSF has similar selectivity to PMSF but is more stable in

aqueous solution and MALDI profiles were comparable in the presence of these two reagents. Serine proteases clearly create peptide patterns prior to precipitation and extraction.

Equilibrium between endopeptidase and exopeptidase catalysis controls the abundance of endopeptidase generated peptides, with consequences for quantitative analysis based on peak measurement. The effect of exopeptidases on MALDI-MS of serum was therefore examined. Bestatin is an inhibitory transition state analog of aminopeptidase (Taylor, 1993). Potato carboxypeptidase inhibitor inhibits carboxypeptidase A. If exopeptidases act on serum peptides over the course of the three day incubation, these two inhibitors might have decreased baseline signal, particularly in the low molecular mass region. Compared to the control however, neither of these resulted in lower signal-to-noise ratio, with the most intense peaks being 75 times the baseline intensity on average (Fig 10 d).

Serum peptides generated by proteases with specificity differing from those used for digestion pose a challenge to search algorithms. In fact non-tryptic assignments are used in the evaluation of search algorithm performance to assess false-positive rates (Kapp et al., 2005). The ESI and MALDI-Qq-TOF data shown in this study had sufficient resolution and mass accuracy for the confident assignment of non-tryptic peptides. An increased computational burden associated with non-tryptic data must be recognized, though increasing computing power addresses this issue.

Tryptic Peptides in Serum Extracts

ProteinPilot results from forward and forward/reverse library comparisons were screened for tryptic cleavages using an SQL database. This method assigned peptides as non-tryptic, semi-tryptic or fully tryptic. According to the selected enzymatic cleavage criteria tryptic peptides were almost absent from the set examined as was expected due to the omission of trypsin (Table 5). Albumin had the highest number of tryptic peptides, around 8%, while 13 of the 22 proteins investigated had none. The peptides examined were filtered by their unique sequences to produce a non-redundant list. Whether the inclusion of a peptide score cut-off would affect the ratio of tryptic to non-tryptic peptides remains unknown though the presence of apparently tryptic peptides is of little consequence when high probability and manual inspection determine the validity of identifications. Semi-tryptic sequences represented around one fifth of those examined and this is more likely due to the proteomic abundance of arginine and lysine than to trypsin. Enzyme-specific searching is not required for the analysis of large proteomic datasets provided high accuracy MS/MS results are used.

The high selectivity of trypsin cleavage is well documented and adds confidence to peptide assignments (Olsen et al., 2004). Where more detailed information regarding *in situ* protein stability is required, non-tryptic peptides with mass accuracy exceeding 0.1 Da confirmed by equally accurate MS/MS experiments should be used. If a diagnostic peptidomic signature exists in serum, it may be obscured by enzymatic digestion prior to MS/MS.

Though peptide degradation in complex samples can be monitored by MS/MS, the processes responsible are incompletely described. The use and improvement of non-

tryptic identification should help address this paucity of information. Separation of tryptic from non- or semi-tryptic data has been used for strengthening protein identifications (Alves et al., 2008). This showed that non-tryptic peptides above an established Mascot probability threshold to be reliable. A machine-learning component enabled the relationships between residue identity, location and peptide decay to be studied and commonalities between the sequence information and chymotryptic specificity were described. A similar strategy for extended serum peptide analysis, examining both known and learned cleavages, should allow all degradative processes to be interpreted by MS/MS.

Internal Standards in Untreated Serum Detected by MALDI-MS

Polymeric biological analytes can be quantified by internally or externally standardized mass spectrometry, a practice common in small molecule analysis. Here MALDI-Qq-TOF was shown to distinguish two molecules added to the equivalent of one microlitre of serum from peptide analytes in a single scan. Isotopic resolution suggests that these internal standards can be identified using LC-ESI-Qq-TOF.

Triple quadrupole instruments have traditionally offered lower sensitivity than TOF, and this accounts for the prevalence of such instruments for exploratory blood protein research. The sample volumes used as well as the observed detection limits of the internal standards suggest low abundance peptides can be quantified by MALDI-Qq-TOF, with the advantages of excellent resolution and accuracy (Fig 11). Tuning the MS mass window increases sensitivity and might permit the quantification of the least

abundant sequences, provided the standards approach the mass of the polypeptides under investigation.

[Glu^1] fibrinopeptide B and Angiotensin 1, which are similar in mass and structure to the peptides observed, gave peak heights well above baseline at concentrations equivalent to moderately abundant serum proteins (Anderson and Anderson, 2002). Exogenous peptides are ideal for standardizing peptide samples because they have similar ionization properties and can be distinguished from sample components by MS/MS. This study demonstrates that Qq-TOF can be employed in peptide quantification at a physiologically relevant scale.

CONCLUSIONS

In this study, organic precipitation followed by liquid extraction and mass spectrometry were applied to simultaneously monitor peptides from a number of known serum proteins with high confidence without prior trypsin digestion. Simple and rapid extractions were complementary for the enrichment of non-tryptic peptides with good signal-to-noise by MALDI and ESI-Qq-TOF. Precipitation and extraction exhibit low selectivity and are effective separatory techniques for fractionating complex samples which can be used directly or prior to further serum separation.

ESI-MS/MS provided spectra that achieved high confidence peptide assignments by X!Tandem and ProteinPilot. When paired, the two algorithms were in agreement regarding the identity of high quality MS/MS, but ProteinPilot identified additional peptides at all levels of confidence. Both were susceptible to bias due to FASTA library size and forward-only searching improved scores of manually verified sequences. To discriminate serum peptides, X!Tandem should be operated in the b/y ion mode as opposed to the all-ion mode.

X!Tandem interpreted only a small number of MALDI tandem mass spectra as belonging to peptides. Peptides could be identified in MALDI-MS peaks by matching their highly accurate masses with ESI-MS/MS results.

Decomposition affected serum analytes prior to organic precipitation as protease inhibitors with varying specificity did not moderate serum peptide profiles by MALDI. The most important period for preserving diagnostically relevant peptide information is likely around the time of blood collection. Measures must be taken to prevent clotting and other protease-driven reactions from obscuring endogenous peptides. Metal

chelating compounds such as EDTA and clotting inhibitors such as heparin have been used for preserving blood proteins (Yi et al., 2007). The set of proteins from non-tryptic peptides observed in the serum extracts represent a small fraction of the number known to be in serum. The mass accuracy and resolution of Qq-TOF in both single and tandem MS modes permitted the study of a representative group of serum peptides.

APPENDIX 1: Supplemental Figures and Tables

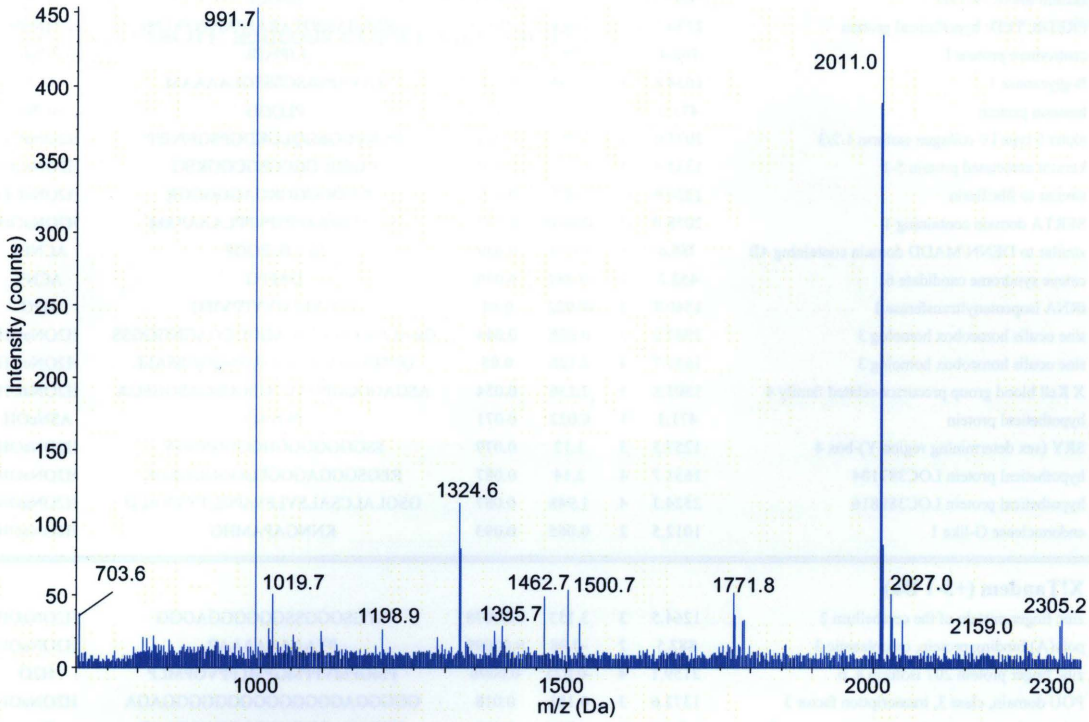


Figure A1. MALDI-MS of serum thawed after being frozen for 72 hours. Serum was aliquoted from the same pool treated with protease inhibitors and immediately placed at -20°C. Upon thawing, a 200 µl volume was withdrawn, purified by C18 reversed-phase chromatography and the eluate spotted onto a metal target plate. The matrix material used was CHCA. The range of 0 to 2350 Th, which included the only peaks with intensity above 30 counts is shown.

Reversed Protein	M+H	z	Delta (Da)	Expect	Sequence	Sample
X!Tandem (+100 PPM)						
keratin 6C, 13a, 15	408.2	1	0.025	0.00094	AGGGF	H2OF
zinc finger protein 207 isoform a	2159.1	4	-0.025	0.002	PMGPMVPPMGPMLPPVGPMLP	H2O
keratin 6A/6C/9/13B	408.2	1	0.023	0.0072	AGGGF	ACN
PREDICTED: hypothetical protein	2130.1	3	2.026	0.012	GRSAGASSQTARASPAAVGTTPE	H2ONaOH
centromere protein J	702.4	1	0.067	0.017	LIPFGR	ACNNaOH
N-glycanase 1	1634.8	3	1.144	0.018	EAVAPSASGSSSGLAAAAM	ACN
bassoon protein	471.3	1	0.022	0.02	PLQGG	ACNT
alpha 5 type IV collagen isoform 1/2/3	2075.0	4	1.002	0.024	DGKVGQEGQLGKDGPSGPPGPP	H2ONaOH
keratin associated protein 5-1	1333.4	3	-0.117	0.026	CGSGCGGCGSGCGGKSG	H2ONaOH
similar to fibrillarin	1230.6	3	1.095	0.029	GGGDGGGGKGAGGGGGR	H2ONaOH
SERTA domain containing 1	2058.0	3	-0.0038	0.032	NDAVSPAAPPSPVPPLAAAPAM	H2ONaOH
similar to DENN/MADD domain containing 4B	780.4	1	0.036	0.034	SLVPGPGGP	ACNFT
cateye syndrome candidate 6	453.2	1	-0.032	0.036	QAPPG	ACNT
tRNA isopentenyltransferase 1	1540.8	2	-0.022	0.04	VDSVELGYVPPVIPG	H2O
sine oculis homeobox homolog 3	1907.6	3	0.836	0.044	GGAGNGGGSGGGAGGGGGAGNGGGSS	H2ONaOH
sine oculis homeobox homolog 3	1631.7	4	2.126	0.05	GGSGGGAGGGGGAGNGGGSSALL	H2ONaOH
X Kell blood group precursor-related family 4	1803.8	3	2.136	0.054	ASGAGGGGPGAVGGSGGSGASGGSGA	H2ONaOH
hypothetical protein	471.3	1	0.022	0.071	PLNAG	ASNaOH
SRY (sex determining region Y)-box 4	1257.5	3	1.12	0.079	SSGGGGGGGHGGGGSGGV	H2ONaOH
hypothetical protein LOC387104	1631.7	4	2.14	0.087	REGSGGGAGGGEGGGGGGEGL	H2ONaOH
hypothetical protein LOC389816	2524.3	4	1.948	0.087	GSGLALCSALSVLFSAPGLTYVVALD	H2ONaOH
endonuclease G-like 1	1012.5	2	0.065	0.093	KNNGAPAMHG	H2ONaOH
X!Tandem (+3-1 Da)						
zinc finger protein of the cerebellum 2	1264.5	3	3.153	0.00078	ATGSGGGSSGGGGGGAGGG	H2ONaOH
poly(A) binding protein, cytoplasmic 3	882.5	2	-0.98	0.00099	PTAAAAAAAAP	H2ONaOH
zinc finger protein 207 isoform a, b	2159.1	4	-0.025	0.0096	PMGPMVPPMGPMLPPVGPMLP	H2O
POU domain, class 3, transcription factor 3	1372.6	3	0.181	0.018	GGGGGAGGGGGGGGGGGGGGAGA	H2ONaOH
voltage-dependent calcium channel gamma-8	1515.7	3	4.061	0.019	DREAGAGGGGGGGGGAGGAAG	H2O
POU domain, class 3, transcription factor 3	1730.7	4	-0.737	0.02	GPQMGGGGGGAGGGGGGGGGGGGGA	H2ONaOH
zinc finger protein 281	1339.5	3	0.204	0.033	SGGGGGGGSGGGGSGSGGSS	H2ONaOH
AT rich interactive domain 1B1/2/3	1303.5	2	2.182	0.036	GGGGGGGGSGGGGGGGGGGGG	H2ONaOH
MPN domain containing	1836.8	3	3.157	0.041	CGGAGAGGGGGGGSVSSGGGGSRGGG	H2ONaOH
retinoic acid induced 2	1126.7	2	1.915	0.042	PVPVPLPVIVP	ACNT
POU domain, class 3, transcription factor 3	1647.7	3	4.152	0.047	STVAASGPQMGGGGGGAGGGGG	H2ONaOH
calpain, small subunit 1	1287.5	3	1.206	0.053	GGGGGGGGGGGGGGGGGGGGAG	H2ONaOH
hypothetical protein	2130.1	3	2.026	0.074	GRSAGASSQTARASPAAVGTTPE	H2ONaOH
hypothetical protein	1547.7	2	3.47	0.081	EPDEVAVPVFGQSSS	H2ONaOH
AT rich interactive domain 1B 1/2/3	1744.7	3	3.217	0.084	GGAGAGGGGGGGGSGGGGGGGGGGGGAG	H2ONaOH
hypothetical protein LOC442132	1289.5	3	-0.793	0.086	DAGGAGGDAGGAGGDAGG	H2ONaOH
essential meiotic endonuclease 1 homolog 1	2279.3	3	3.808	0.087	QLAGLLQGGGEMQLLVPDLVVI	H2O
N-glycanase 1	1634.8	3	1.144	0.088	EAVAPSASGSSSGLAAAAM	ACN
hypothetical protein LOC442132	1625.6	3	-0.752	0.09	AEADAGGAGADAAGAGAAGE	H2ONaOH

Table A1. Non-redundant reverse library sequences matched by an X!Tandem search of acetonitrile pellet extracts against a combined forward/reverse human FASTA library. M+H: calculated mass. Abbreviations of sample conditions are as follows. ACN:

acetonitrile supernatant, ACNT: acetonitrile supernatant with TFE, ACNFT: acetonitrile supernatant with formic acid and TFE H2O: aqueous extract, H2OF: aqueous formic acid extract, H2OT: aqueous extract with TFE, H2OFE aqueous extract with formic acid and TFE, H2ONaOH: aqueous extract with NaOH.

BIBLIOGRAPHY

Adkins, J.N., Varnum, S.M., Auberry, K.J., Moore, R.J., Angell, N.H., Smith, R.D., Springer, D.L. and Pounds, J.G. 2002. Toward a human blood serum proteome: Analysis by multidimensional separation coupled with mass spectrometry. Molecular and Cellular Proteomics 1: 947-955.

Altschul, S. F., Madden,T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

Alves, P., Arnold, R.J., Clemmer, D.E., Li, Y., Reilly, J.P. Sheng, Q., Tang, H., Xun, Z., Zeng, R and Radivojac, P. 2008. Fast and accurate identification of semi-tryptic peptides in shotgun proteomics. Bioinformatics 24(1): 102-109.

Arakawa, T. and Timasheff, S.N. 1985. Theory of protein solubility. Methods in Enzymology 114: 49-77.

Arakawa, T. and Timasheff, S.N. 1987. Abnormal solubility behavior of beta-lactoglobulin: Salting-in by glycine and NaCl. Biochemistry 26(16): 5147-53.

Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A. and Mann, M. 2003. Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426: 570-574.

Anderson, L.N. and Anderson, N.G. 2002. The human plasma proteome: History, character and diagnostic prospects. Molecular and Cellular Proteomics 1:845-867.

Beavis, R.C., Chaudhary, T. and Chait, B.T. 1992. α-Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. Organic Mass Spectrometry 27: 156-158.

Blume-Jensen, P. and Hunter, T. 2001. Oncogenic kinase signalling. Nature 411: 355-365.

Brosch, M., Swamy, S., Hubbard, T., and Choudhary, J. 2008. Comparison of Mascot and X!Tandem performance for low and high accuracy mass spectrometry and the development of an adjusted Mascot threshold. Molecular and Cellular Proteomics 7(5): 962-970.

Chan, K.C., Lucas, D.A., Hise, D., Schaefer, C.F., Xiao, Z., Janini, G.M., Buetow, K.H., Issaq, H.J., Veenstra, T.D. and Conrads, T.P. 2004. Analysis of the human serum proteome. Clinical Proteomics 1: 101-225.

Chang, W.C., Huang, L.C.L., Wang, Y., Peng, W., Chang, H.C., Hsu, N.Y., Yang, W.B. and Chen, C.H. 2007. Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. Analytica Chimica Acta 582: 1-9.

Chelius, D. and Bondarenko, P.V. 2002. Quantitative profiling of proteins in complex mixtures using chromatography and mass spectrometry. Journal of Proteome Research 1: 317-323.

Chen, J., Anderson, M., Misek, D.E., Simeone, D.M. and Lubman, D.M. 2007. Characterization of apolipoprotein and apolipoprotein precursors in pancreatic cancer serum samples via two-dimentional liquid chromatography and mass spectrometry. Journal of Chromatography A 1162: 117-125.

Chertov, O., Biragyn, A., Kwak, L.W., Simpson, J.T., Boronina, T., Hoang, V.M., Prieto, D.A., Conrads, T.P., Veenstra, T.D. and Fisher, R.J. 2004. Organic solvent extraction of proteins and peptides from serum as an effective sample preparation for detection and identification of biomarkers by mass spectrometry. Proteomics 4: 1195-1203.

Craig, R. and Beavis, R.C. 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20(9): 1466-1467.

Cohn, E.J. 1925. The physical chemistry of the proteins. Physiological Reviews 5(3): 349-437.

Cole, R.B. 2000. Some tenets pertaining to electrospray ionization mass spectrometry. Journal of Mass Spectrometry 35: 763-772.

Cutillas, P.R., Geering, B., Waterfield, M.D. and Vanhaesebroeck, B. 2005. Quantification of gel-separated proteins and their phosphorylation sites by LC-MS using unlabeled internal standards. Molecular and Cellular Proteomics 4.8: 1038-1051.

Davie, E.W., Fujikawa, K., Kurachi, K and Kisiel, W. 1979. The role of serine proteases in the blood coagulation cascade. Advancements in Enzymology and Related Areas of Molecular Biology 48: 277-318.

DeSouza, L., Diehl, G., Rodrigues, M.J., Gou, J., Romaschin, A.D., Colgan, T.J. and Siu, K.W.M. 2005. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. Journal of Proteome Research 4(2): 377-386.

Domon, B. and Aebersold, R. 2006. Mass spectrometry and protein analysis. Science 312: 212-217.

Diamandis, E.P. 2003. Proteomic patterns in biological fluids: Do they represent the future of cancer diagnostics? Clinical Chemistry 49(8): 1272-1278.

Diamandis, E.P. 2006. Serum proteomic profiling by matrix-assisted laser desorption ionization time-of-flight mass spectrometry for cancer diagnosis: Next steps. Cancer Research 66(11): 5540-5541.

Diamandis, E.P. 2007. Oncopeptidomics: A useful approach for cancer diagnosis? Clinical Chemistry 53(6): 1004-1006.

Dionne, R., Forest, J.C., Moutquin, J.M., De Grandpre, P. and Masse, J. 1994. Electrophoretic method for separating small peptides in serum without extraction of macromolecules: Application to the detection of preeclampsia. Clinical Biochemistry 27(2) : 99-103.

Elias, J.E. and Gygi, S.P. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods 4(3): 207-214.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. 1989. Electrospray ionization for the mass spectrometry of large biomolecules. Science 246: 64-71.

Fenyö, D. and Beavis, R.C. 2008. Informatics development: Challenges and solutions for MALDI mass spectrometry. Mass Spectrometry Reviews 27: 1-19.

Findeisen, P., Post, S., Wenz, F and Neumaier, M. 2007. Addition of exogenous reporter peptides to serum samples before mass spectrometry-based protease profiling provides advantages over profiling of endogenous peptides. Clinical Chemistry 53 (10): 1864-1866.

Findeisen, P. Peccrtrlla, T., Post, S. Wenz, F. and Neumaier, M. 2008. Spiking of serum specimens with exogenous reporter peptides for mass spectrometry based protease profiling as diagnostic tool. Rapid Communications in Mass Spectrometry 22: 1223-1229.

Fung, E.T., Yip, T., Lomas, L., Wang, Z., Yip, C., Meng, X., Lin, S., Zhang, F., Zhang, Z., Chan, D.W. and Weinberger, S.R. 2005. Classification of cancer types by measuring variants of host response proteins using SELDI serum analysis. International Journal of Cancer 115: 783-789.

Han, K.Q., Huang, G., Gao, C.F., Wang, X.L., Ma, B., Sun, L.Q. and Wei Z.J. 2008. Identification of lung cancer patients by serum protein profiling using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. American Journal of Clinical Oncology 31(2): 133-139.

Henschen-Edman, A.E. 2001. Fibrinogen non-inherited heterogeneity and its relationship to function in health and disease. Annals New York Academy of Sciences 936: 580-593.

Himmelfarb, M., Klopocki, E., Grube, S., Staub, E., Klaman, I., Hinzmann, B., Kristiansen, G, Rosenthal, A., Dürst, M. and Dahl, E. 2004. ITIH5, a novel member of

the inter-alpha-inhibitor heavy chain family is downregulated in breast cancer. Cancer Letters 204: 69-77.

Hortin, G. 2006. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. Clinical Chemistry 52(7): 1223-1237.

Ghosh, S., Gepstein, S., Heikkila, J.J. and Dumbroff, E.B. 1987. Use of a scanning densitometer or an ELISA plate reader for measurement of nanogram amounts of protein in crude extracts from biological tissues. Analytical Biochemistry 169: 227-233.

Govorukhina, N.I., Reijmers, T.H., Nyangoma, S.O., van der Zee, A.G.J., Jansen, R.C., Bischoff, R. 2006. Analysis of human serum by liquid chromatography–mass spectrometry: Improved sample preparation and data analysis. Journal of Chromatography A 1120: 142-150.

Griffin, T.J., Gygi, S.P., Rist, B., Aebersold, R., Loboda, A., Jilkine, A., Ens, W. and Standing, K.G. 2001. Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer. Analytical Chemistry 73(5): 978-986.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nature Biotechnology 17(10): 994-999.

Juan, H., Chen, J., Hsu, W., Huang, S., Chen, S., Lin, J.Y., Chang, Y., Chiang, C., Wen, L., Chan, D., Liu, Y. and Chen, Y. 2004. Identification of tumor-associated plasma biomarkers using proteomic techniques: From mouse to human. Proteomics 4: 2766-2775.

Josic, D., Brown, M.K., Huang, Y., Rucevic, M., Clifton, J.G. and Hixson, D.C. 2006. Proteomic characterization of inter-alpha inhibitor proteins from human plasma. Proteomics 6: 2874-2885.

Kapp, E.A., Schütz, F., Connolly, L.M., Chakel, J.A., Meza, J.E., Miller, C.A., Fenyo, D., Eng, J.K., Adkins, J.N., Omenn, G.S. and Simpson, R.J. 2005. An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. Proteomics 5: 3475-3490.

Karas, M. and Hillenkamp, F. 1988. Laser desorption ionization of proteins with mass exceeding 10000 Daltons. Analytical Chemistry 60: 2299-2301.

Karsan, A., Eigl, B.J., Flibotte, S., Gelmon, K., Switzer, P., Hassell, P., Harrison, D., Law, J., Hayes, M., Stillwell, M., Xiao, Z., Conrads, T.P., Veenstra, T. 2005. Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. Clinical Chemistry 51(8): 1525-1528.

Kjellström, S. and Jensen, O.N. 2003. In situ liquid-liquid extraction as a sample preparation method for matrix-assisted laser desorption ionization MS analysis of polypeptide mixtures. Analytical Chemistry 75: 2367-2369.

Koomen, J.M., Li, D., Xiao, L., Liu, T.C., Coombes, K.R., Abbruzzese, J. and Kobayashi, R. 2005. Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. Journal of Proteome Research 4: 972-981.

Leman, E.S., Schoen, R.E., Magheli, A., Sokoll, L.J., Chan, D.W. and Getzenberg, R.H. 2008. Evaluation of colon-cancer specific antigen 2 as a potential serum marker for colorectal cancer. Clinical Cancer Research 14: 1349-1354.

Li, X., Xu, S., Pan, C., Zhou, H., Jiang, X., Zhang, Y., Ye, M. and Zou, H. 2007. Enrichment of peptides from plasma for peptidome analysis using multiwalled carbon nanotubes. Journal of Separation Science 30: 930-943.

Lim, Y., Bendelja, K., Opal, S.M., Siryaporn, E., Hixson, D.C. and Palardy, J.E. 2003. Correlation between mortality and the levels of inter-alpha inhibitors in the plasma of patients with severe sepsis. Journal of Infectious Diseases 188: 919-926.

Loboda, A.V., Krutchinsky, A.N., Bromirski, M., Ens, W. and Standing, K.G. 2000. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: Design and performance. Rapid Communications in Mass Spectrometry 14: 1047-1057.

Neal, B.L., Asthagiri, D. and Lenhoff, A.M. 1998. Molecular origins of osmotic second virial coefficients of proteins. Biophysical Journal 75: 2469-2477.

Nishimura, H., Kakizaki, I., Muta, T., Sasaki, N., Pu, P.X., Yamashita, T. and Nagasawa, S. 1995. cDNA and deduced amino acid sequence of human PK-120, a plasma kallikrein-sensitive glycoprotein. FEBS Letters 357: 207-211.

Malik, G., Ward, M.D., Gupta, S.K., Trosset, M.W., Grizzle, W.E., Adam, B., Diaz, J.I. and Semmes, O.J. 2005. Serum levels of an isoform of Apolipoprotein A-II as a potential marker for prostate cancer. Clinical Cancer Research 11: 1073-1085.

Marshall, J., Kupchak, P., Zhu, W., Yantha, J., Vrees, T., Furesz, S., Jacks, K., Smith, C., Kireeva, I., Zhang, R., Takahashi, M., Stanton, E. and Jackowski, G. 2003. Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. Journal of Proteome Research 2: 361-372.

Marshall, J., Jankowski, A., Furesz, S., Kireeva, I., Barker, L., Dombrovsky, M., Zhu, W., Jacks, K., Ingratta, L., Bruin, J., Kristensen, E., Zhang, R., Stanton, E., Takahashi, M. and Jackowski, G. 2004. Human serum proteins preseparated by electrophoresis or

chromatography followed by tandem mass spectrometry. Journal of Proteome Research 3: 364-382.

Mann, M., Ong, S., Grønborg, M., Steen, H., Jensen, O.N. and Pandey, A. 2002. Analysis of protein phosphorylation using mass spectrometry: Deciphering the phosphoproteome. Trends in Biotechnology 20(6): 261-268.

Masuda, Y. and Sugiyama, T. 2001. Human fibrinopeptide A mediates allergic reaction in mice in the acute phase. Peptides 22: 1511-1513.

McLafferty, F.W. and Bockhoff, F.M. 1978. Separation/identification system for complex mixtures using mass separation and mass spectral characterization. Analytical Chemistry 50(1): 69-76.

Merrell, K., Southwick, K., Graves, S.W., Esplin, M.S., Lewis, N.E. and Thulin, C.D. 2004. Analysis of low-abundance, low molecular weight serum proteins using mass spectrometry. Journal of Biomolecular Techniques 15(4): 238-248.

Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R.S., Hoyes, J. and Bateman, R.H. 1996. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. Rapid Communications in Mass Spectrometry 10 (8): 889-896.

Oleschuk, R.D., McComb, M.E., Chow, A., Ens, W., Standing, K.G., Perreault, H., Marois, Y. and King, M. 2000. Characterization of plasma proteins adsorbed onto biomaterials by MALDI-TOFMS. Biomaterials 21: 1701-1710.

Olsen, J.V., Ong, S. and Mann, M. 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular and Cellular Proteomics 3(6): 608-614.

Omenn, G.S., States, D.J., Adamski, M., Blackwell, T.W., Menon, R., Hermjakob, H., Apweiler, R., Haab, B.B., Simpson, R.J., Eddes, J.S. et al. 2005. Overview of the HUPO plasma proteome project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 5: 3226-3245.

Ong, S., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Molecular and Cellular Proteomics 1(5): 376-386.

Ornstein, D.K., Rayford, W., Fusaro, V.A., Conrads, T.P., Ross, S.J., Hitt, B.A., Wiggins, W.W., Veenstra, T.D., Liotta, L.A. and Petricoin, E.F. III. 2004. Serum proteomic profiling can discriminate prostate cancer from benign prostates in men with total prostate specific antigen levels between 2.5 and 15.0 ng/mL. The Journal of Urology 172: 1302-1305.

Palmblad, M., Wetterhall, M., Markides, K., Håkansson, P. and Berqquist, J. 2000. Analysis of enzymatically digested proteins and protein mixtures using a 9.4 Tesla Fourier transform ion cyclotron mass spectrometer. Rapid Communications in Mass Spectrometry 14(12): 1029-34.

Pearson, W.R. and Lipman, D.L. 1988. Improved tools for biological sequence comparison. Proceedings of the National Academy of Science USA 85: 2444-2448.

Petricoin EF III, Ardekani AM, Hitt BA, et al. 2002. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359: 572–575.

Plumer, A., Duan, H., Subramaniam, S., Lucas, F.L., Miesfeldt, S., Ng, A. and Liaw, L. 2008. Development of fragment-specific osteopontin antibodies and ELISA for quantification in human metastatic breast cancer. BMC Cancer 8: 38.

Pruitt, K.D., Tatusova, T. and Maglott, D.R. 2007. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35: D61-D65.

Qian, W., Liu, T., Monroe, M.E., Strittmatter, E.F., Jacobs, J.M. Kangas, L.J., Petritis, K., Camp II, D.G. and Smith, R.D. 2005. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: The human proteome. Journal of Proteome Research 4: 53-62.

Ramos-Fernández, A., Paradela, A., Navajas, R., Albar, J.P. 2008. Generalized method for probability-based peptide and protein identification from tandem mass spectrometry data and sequence database searching. Molecular and Cellular Proteomics Epub ahead of print May 31, 2008 Manuscript M800122-MCP200.

Reid, K.B. 1983. Proteins involved in the activation and control of the two pathways of human complement. Biochemical Society Transactions 11(1): 1-12.

Richter, R., Schulz-Knappe, P., Schrader, M., Standker, L., Jurgens, M., Tammen, H. and Forssmann, W. 1999. Composition of the peptide fraction in human blood plasma: Database of circulating human peptides. Journal of Chromatography B 726: 25-35.

Ruckenstein, E. and Shulgin, I.L. 2006. Effects of salts and organic additives on the solubility of proteins in aqueous solutions. Advances in Colloid and Interface Science 123-126: 97-103.

Saguchi, K., Tobe, T., Hashimoto, K., Sano, Y., Nakano, Y., Miura, N. and Tomita, M. 1995. Cloning and characterization of cDNA for inter-alpha-trypsin inhibitor family heavy chain-related protein (IHRP), a novel human plasma glycoprotein. Journal of Biochemistry 117: 14-18.

Schagger, H. and von Jagow, G. 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Analytical Biochemistry 166: 368-79

Seydewitz, H.H. and Witt, I. 1985. Increased phosphorylation of human fibrinopeptide A under acute phase conditions. Thrombosis Research 40(1): 29-39.

Shen, Y., Jacobs, J.M., Camp, D.G. II, Fang, R., Moore, R.J., Smith, R.D., Xiao, W., Davis, R.W. and Tompkins, R.G. 2004. Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. Analytical Chemistry 76: 1134-1144.

Shen, Y., Kim, J., Strittmatter, E.F., Jacobs, J.M., Camp, D.G. II, Fang, R., Tolie, N., Moore, R.J. and Smith, R.D. 2005. Characterization of the human blood plasma proteome. Proteomics 5: 4034-4045.

Shevchenko, A., Chernushevich, I., Ens, W., Standing, K.G., Thompson, B., Wilm, M. and Mann, M. 1997. Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. Rapid Communications in Mass Spectrometry 11: 1015-1024.

Shilov, I.V., Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H., Keating, S.P., Hunter, C.L., Nuwaysir, L.M. and Shaeffer, D.A. 2007. The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Molecular and Cellular Proteomics 6.9: 1638-1655.

Song, J., Patel, M., Rosenzweig, C.N., Chan-Li, Y., Sokoll, L.J., Fung, E.T., Choi-Miura, N., Goggins, M., Chan, D.W. and Zhang, Z. 2006. Quantification of fragments of human serum Inter-alpha-trypsin inhibitor heavy chain 4 by a surface-enhanced laser desorption/ionization based immunoassay. Clinical Chemistry 52(6): 1045-1053.

States, D.J., Omenn, G.S., Blackwell, T.W., Fermin, D., Eng, J., Speicher, D.W. and Hanash, S.H. 2006. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. Nature Biotechnology 24 (3): 333-338.

Steen, H. and Mann, M. 2004. The ABC's (and XYZ's) of peptide sequencing. Nature Reviews Molecular Cell Biology 5: 699-711.

Strader, M.B., Tabb, D.L., Hervey, W.J., Pan, C. and Hurst, G.B. 2006. Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. Analytical Chemistry 78: 125-134.

Tammen, H., Schulte, I., Hess, R., Menzel, C., Kellmann, M., Mohring, T. and Schulz-Knappe, P. 2005. Peptidomic analysis of human blood specimens: Comparison between plasma specimens and serum by differential peptide display. Proteomics 5: 3414-3422.

Taylor, A. 1993. Aminopeptidases: Structure and function. The FASEB Journal 7(2): 290-298.

Tian, R., Ye, M., Hu, L., Li, X. and Zou, H. 2007. Selective extraction of peptides in acidic human plasma by porous silica nanoparticles for peptidome analysis with 2-D LC-MS/MS. Journal of Separation Science 30: 2204-2209.

Tirumalai, R.S., Chan, K.C., Prieto, D.A., Issaq, H.J., Conrads, T.P. and Veenstra, T.D. 2003. Characterization of the low molecular weight human serum proteome. Molecular and Cellular Proteomics 2: 1096-1103.

Tiss, A., Smith, C., Camuzeaux, S., Kabir, M., Gayther, S., Menon, U., Waterfield, M., Timms, J., Jacobs, I. and Cramer, R. 2007. Serum peptide profiling using MALDI mass spectrometry: Avoiding the pitfalls of coated magnetic beads using well-established ZipTip technology. Practical Proteomics 1: 77-89.

Tucholska, M., Scozzaro, S., Williams, D., Ackloo, S., Lock, C., Siu., K.W.M., Evans, K.R. and Marshall, J.G. 2007. Endogenous peptides from biophysical and biochemical fractionation of serum analyzed by matrix-assisted laser desorption/ionization and electrospray ionization hybrid quadrupole time-of-flight. Analytical Biochemistry 370: 228-245.

Tsur, D. Tanner, S., Zandi, E., Bafna, V. and Pevzner, P.A. 2005. Identification of post-translational modifications by blind search of mass spectra. Nature Biotechnology 23(12): 1562-1567.

Verhaert, P., Uttenweiler-Joseph, S., de Vries, M., Loboda, A., Ens, W. and Standing K.G. 2001. Matrix-assisted laser desorption/ionization quadrupole time-of-flight mass spectrometry: An elegant tool for peptidomics. Proteomics 1: 118-131.

Villanueva, J., Shaffer, D.R., Philip, J., Chaparro, C.A., Erdjument-Bromage, H., Olshen, A.B., Fleisher, M., Lilja, H., Brogi, E., Boyd, J., Sanchez-Carbayo, M., Holland, E.C., Cordon-Cardo, C., Scher, H.I. and Tempst, P. 2006. Differential exoprotease activities confer tumor-specific serum peptidome patterns. Journal of Clinical Investigation 116(1): 271-284.

Werle, M. and Bernkop-Schnürch, A. 2006. Strategies to improve plasma half life time of peptide and protein drugs. Amino Acids 30: 351-367.

Williams, D., Zhu, P., Bowden, P., Stacey, C., McDonell, M., Kowalski, P., Kowalski, J.M., Evans, K., Diamandis, E.P., Siu, K.W.M. and Marshall, J. 2006. Comparison of methods to examine the endogenous peptides of fetal calf serum. Clinical Proteomics 2: 67-89.

Wu, W.W., Wang, G., Baek, S.J. and Shen, R. 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel or LC-MALDI TOF/TOF. Journal of Proteome Research 5: 651-658.

Yi, J., Kim, C. and Gelfand, C.A. 2007. Inhibition of intrinsic proteolytic activities moderates preanalytical variability and instability of human plasma. Journal of Proteome Research 6: 1768-1781.

Zhang, N. and Li, L. 2004. Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. Rapid Communications in Mass Spectrometry 18: 889-896.

Zheng, X., Baker, H. and Hancock, W.S. 2006. Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer. Journal of Chromatography A 1120: 173-184.

Zimmerman, L.J., Wernke, G.R., Caprioli, R.M. and Liebler, D.C. 2005. Identification of protein fragments as pattern features in MALDI-MS analyses of serum. Journal of Proteome Research 4: 1672-1680.