Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

Advantages of a new proteomic approach that uses accurate mass measurements, LC retention time, isoelectric point and dual enzymatic digestion

Konstantinos Petritis[1], Anoop M. Mayampurath[1], Daniel Lopez-Ferrer[1], Navdeep Jaitly[1], Jungbae Kim[1], David C. Schriemer[3], Darren F. Lewis[4], Matthew E. Monroe[1], Ljiljana Paša-Tolić[1], Nikola Tolić[1], Mikhail Belov[1], John Fjeldsted[2], Bryan Miller[2], Tom Vandegoor[2], Patrick Hoerth[2], Paul Goodley[2], and Richard D. Smith[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352; [2]Agilent Technologies, Waldbronn, Germany or Santa Clara, CA, USA; [3]University of Calgary, Calgary, Alberta, Canada [4]Upchurch Scientific Inc., Oak Harbor, WA

Overview

❖ A new shotgun proteomics approach has been developed based on the use of accurate mass measurements, liquid chromatography (LC) retention time/isoelectric point prediction and dual enzymatic digestion.

❖ The approach has been modeled and proof of concept is supported by preliminary results.

❖ In theory, the present approach presents several advantages over traditional shotgun proteomic approaches, such as higher proteome coverage, elimination of MS/MS gas phase fragmentation and its associated drawbacks (under-sampling, peptides that do not fragment well), better quantitation, etc.

❖ The model can be further extended for *de novo* sequencing and analysis of post-translational modifications.

Introduction

- High throughput proteomics based upon the analysis of peptides is the most widespread approach to proteomics, and generally involves a tryptic digestion followed by LC separation and MS analysis. While there are several variations of this approach, at some point they all utilize MS/MS. Limitations of MS/MS, such as under-sampling issues, low throughput, MS/MS spectral interpretation challenges, etc. lead to decreased dynamic range, low protein coverage, and loss of useful information (particularly with regard to protein modifications).
- Schriemer and co-workers [1] recently demonstrated the possibility of on-line protein digestion just seconds after separation by RPLC. One of the bottlenecks is the unavailability of highly stable and active immobilized enzymes that would allow their use for extended periods of time without enzymatic activity losses (either due to autolysis, proteolysis from other proteases contained in the sample or chemical instability).
- The present study uses a two-step digestion that simulates an "in-solution fragmentation" instead of the traditional gas fragmentation (i.e. MS/MS). Peptides from the first digestion are separated by RPLC, after which part of them are analyzed by MS directly and the other part is digested on-line by one or more proteases before analysis by MS. The peptide accurate mass information obtained from all digestions, in combination with other information such as the peptide LC elution time prediction and isoelectric point, narrow the possible peptide candidates, thereby allowing confident identification.

Methods

- For theoretical modeling of the approach, the *Homo sapiens* proteome was downloaded from European Bioinformatics Institute and contained about 61,000 proteins (H_sapiens_IPI_2006-08-22.fasta). *Homo sapiens* proteins were digested *in silico* by using the Protein Digestion Simulator, an in-house software written using VB.NET and available on-line at **http://ncrr.pnl.gov/software/.** The algorithm that calculates peptide uniqueness was written in C#, and the backend processing was written in C++ on a .NET framework.
- The proposed liquid chromatography-in-solution fragmentation-mass spectrometry (LC-ISF-MS) configuration is depicted in Figure 1. Briefly, the first digestion is performed at the protein level, on-line or off-line, and uses a chemical or enzymatic method to cleave the protein into relatively large peptides. The peptides are then separated, preferably using RPLC, and the liquid flow is split into two or more pathways. One pathway is analyzed as is, preferably by a TOF mass analyzer; the other liquid paths are directed for on-line fast digestion, using enzymatic or chemical methods that further cleave the peptides at different positions. After the on-line digestion, these peptides are also analyzed by MS. Optimally, each liquid stream is electrosprayed individually and analyzed one-by-one by quickly alternating from one liquid stream to the other (i.e., by using a two channel ion-funnel).
- Peptides from the first digestion (i.e., using LysC or AspN or GluC, etc.) provide the "parent ion" information, while the same peptides that have been further digested on-line (i.e., using trypsin and/or chymotrypsin, etc.) provide the "daughter ions". As both parent and daughter ions are analyzed with high mass accuracy, we end up with one unique peptide identification or a small number of possibilities. When combined with peptide LC predicted retention time or isoelectric point information (the latter requires an isoelectric point separation after the first digestion and before RPLC separation), this information should provide one highly confident peptide identification without the need for MS/MS fragmentation.
- In the present study, proof of concept was accomplished by injecting peptides from the first digestion twice, i.e., with and without on-line digestion.

Results

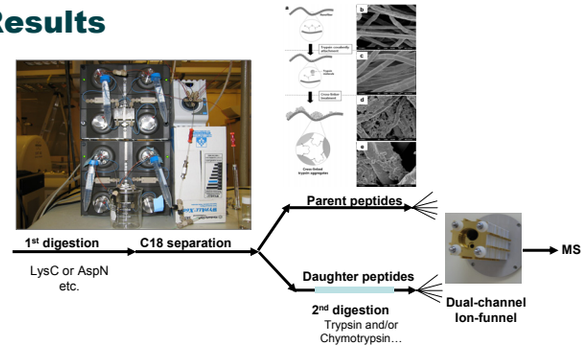
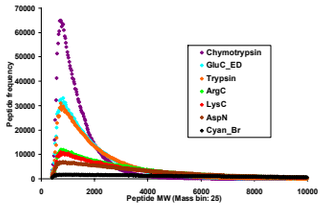


Figure 1. Proposed liquid chromatography-in-solution fragmentation-mass spectrometry (LC-ISF-MS) configuration. First digestion can be done on-line or off-line. After the C_{18} peptide separation, the mobile phase pH might need to be changed (depending on the enzyme used) so a second pump operates a mirror gradient in order for the solvent concentration to be kept constant and the pH before the digestion column to be modified. The mobile phase is re-acidified before introduction to the mass spectrometer when operating in positive ion-mode. The dual-channel ion-funnel interface is able to switch between parent and daughter peptides within MS. Furthermore, new enzyme immobilization solid supports and chemistries are investigated in order to produce immobilized enzymes with the appropriate characteristics (i.e. high activity/stability, proteolytic resistance to autolysis and other proteases contained in the sample, regeneration/reusability, more adequate for large peptide than protein digestion, etc). Figure (a-e) shows different steps of trypsin immobilization in nanofibers. Figure 1 (d-e) shows the creation of trypsin aggregates through trypsin crosslinking which, according to preliminary results, leads to higher activities, stability, and resistance to proteolysis [2].



Enzyme	Cleaves at:	# of peptides
Chymotrypsin	FWYL	2735830
GluC_ED	ED	2119184
Trypsin	KR	2056840
Pepsin	FLIWY	1992000
ArgC	R	1329093
LysC	K	1214811
AspN	D	1080247
Cyan_Br	M	438925

Figure 2. Theoretical and experimental MW peptide distributions for different proteases. Theoretical MW *Homo sapiens* peptide distributions for different proteases. The protease choice for the first and second digestion will be a compromise among peptide delectability, MW distribution of peptides, digestion orthogonality between first and second digestion, # of total peptides, # of peptides that can be confidently identified after in-solution fragmentation, MS ion-mode, etc.

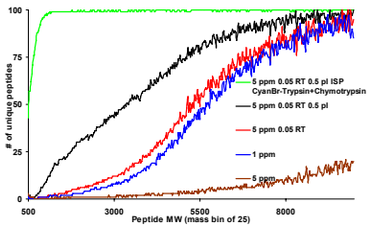


Figure 3. Percentage of unique peptides as a function to peptide MW and different theoretical conditions of mass accuracy (ppm), retention time (RT), isoelectric point (pI) and "in-solution fragmentation (ISF). In this particular example, ISF uses LysC for the first digestion and a combination of trypsin and chymotrypsin for the second digestion. The figure shows that in solution fragmentation in combination with 5 ppm mass accuracy, +/- 0.05% retention time prediction accuracy and +/- 0.5 pH unites isoelectric point accuracy provides enough specificity in order for the peptides with MW>1000 Da to be identified with high confidence. Indeed, at least >91% of the peptides with MW≥ 1000 Da are unique while >99% of the peptides with MW≥1500 are unique.

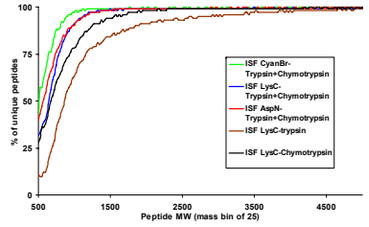


Figure 4. Percentage of unique peptides as a function to peptide MW and "in-solution fragmentation" conditions. All other conditions are 5 ppm of mass accuracy, ±0.05% of retention time prediction accuracy and ±0.5 pH unites isoelectric point accuracy. Different combinations of enzymes provide different degrees of specificity (peptide uniqueness). Higher specificities are obtained when a) in the first digestion we use enzymes that produce high MW peptides and b) in the second digestion we use enzymes that cleave at several additional (to the first digestion) amino acid residues.

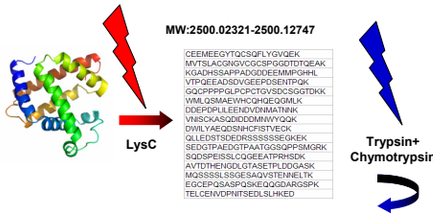


Figure 5. Theoretical example of "in-solution fragmentation". This example shows peptides with MW from 2500.02321-2500.12747 (50 ppm window) from the in-silico digest of *Homo sapiens* proteome. Although several of the LysC peptides are indistinguishable by only high mass accuracy, the second digestion provides several unique peptides for each peptide making their high confidence identification possible. Only amino acids/peptides in red have been found more than once.

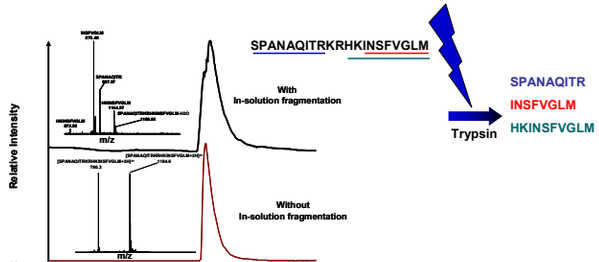


Figure 6. Experimental example of "in-solution fragmentation". Injection of 2 mM of Carassin peptide (SPANAQITRKRHKINSFVGLM) under LC-ISF-MS (daughter peptides) and LC-MS (parent peptide) conditions.

Conclusions

❖ A novel shotgun proteomics method is proposed and theoretically evaluated. The modeling and preliminary results imply that the present method can be a stand alone alternative to traditional MS/MS-based shotgun proteomics methods.

❖ The "in-solution fragmentation" approach is anticipated to provide high protein coverage and increased throughput. Furthermore it can be further extended and used in several other ways: In combination with accurate MS/MS information the "in-solution fragmentation" approach can provide much higher specificity and be used as a powerful *de novo* sequencing approach.

❖ The present method uses full MS spectra (rather than MS/MS) for peptide identification, making it possible to use/explore negative ion mode for their identification. An appropriate set of proteases might be necessary for optimized results (i.e., AspN as first digestion instead of LysC).

Acknowledgements

Portions of this work were supported by the NIH National Center for Research Resources (RR018522) and Battelle's internal research and development funds.

References

1. Detection and Identification of Sub-nanogram Levels of Protein in a NanoLC-Trypsin-MS System, NanoLC-Trypsin-MS System G. W. Slysz, D. F. Lewis, and D.C. Schriemer, *J. Proteome Res.* **2006**, *5*, 1959-1966.
2. Development of High Stability Immobilized Trypsin for Proteomic Applications Using Trypsin-Aggregate Coated Nanofibers B. C. Kim, D. Lopez-Ferrer, H.-K. Ahn, S. Nai, S. H. Kim, B. S. Kim, K. Petritis, D.G. Camp, R. D. Smith, M. B. Gu, J. Kim, J. in preparation.

Contact Information

Konstantinos Petritis, Ph.D.
Biological Sciences Division
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: konstantinos.petritis@pnl.gov
Phone: 509-376-4341