

SANDIA REPORT

SAND2009-3855

Unlimited Release

Printed June 2009

Shotgun Protein Sequencing: LDRD Project 93501 Close-Out Report

Jean-Loup M. Faulon and Grant S. Heffelfinger

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.



Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: <http://www.osti.gov/bridge>

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



SAND 2009-3855
Unlimited Release
Printed June 2009

Shotgun Protein Sequencing: LDRD Project 93501 Close-Out Report

Jean-Loup M. Faulon and Grant Heffelfinger
Computational Biosciences
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1413

Abstract

While gene sequence information alone does not provide a complete picture of a protein final sequence and state of activity, the primary method employed in proteomics analysis is protein identification, mass spectrometry (MS), does not address this problem. Rather, all current MS-enabled protein identification protocols and computer codes are plagued by false-positive peptide sequence identifications.¹ To overcome the problem of protein identification, *de novo* protein sequencing using MS/MS data has been developed but is currently limited to short sequences. Protein sequencing can also be carried out using the traditional and time consuming Edman degradation technique. While the technique is more accurate than identification methods and enables *de novo* protein sequencing, it is slow, requires large amounts of proteins, and has a low throughput. It also requires proteins to be very pure, i.e., it cannot sequence proteins in a mixture.

For these reasons, a new *de novo* protein sequencing method is needed that can 1) sequence proteins that are the result of alternate splicing, 2) sequence proteins that are known to undergo PTMs, and 3) sequence genetically engineered proteins. This SAND reports the development of a protocol, algorithms, and computer code that enables *de novo* sequencing of proteins from MS/MS spectra.

This SAND report describes the development of a novel experimental and computational technique based on multiple enzymatic digestion of a protein or protein mixture that reconstructs protein sequences from sequences of overlapping peptides. This approach, analogous to shotgun sequencing of DNA, is to be used to sequence alternative spliced proteins, to identify post-translational modifications, and to sequence genetically engineered proteins. The interested reader is directed to two journal papers for more detailed analysis of the method and its usefulness.^{2, 3}

Acknowledgement

This work was funded by Sandia Laboratory Directed Research and Development. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. We also would like to thanks Professor Luke O'Neill from Trinity College, Ireland , for providing us plasmid constructs of the four IRAK2 spliced protein genes.

Contents

Nomenclature	6
Introduction.....	7
Background.....	7
Shotgun Protein Sequencing Methodology	9
Technical Approach.....	9
Proteins Selection.....	10
Enzymatic or Chemical Digestion	10
MS/MS data acquisition.....	11
MS/MS peptide identification.....	12
<i>De novo</i> protein assembly.....	13
Shotgun protein sequencing protocol development and benchmark	13
De novo sequencing of alternative spliced forms	14
Expression and purification of alternative spliced proteins.....	14
Multiple enzymes digestion of IRAK2b.....	14
IRAK2b assembly.....	14
Using tag assembly to distinguish IRAK2b from other IRAK isoform.....	15
Conclusion	16
References.....	17

Figures

1 Process schematic for de novo protein sequencing strategy.....	9
2 Tropomyosin Variants	10

Tables

1 Commercially available proteolytic agents utilized in this work.....	11
1 Highest probability assembled tag for IRAK2.....	15

Nomenclature

CAII	carbonic anhydrase II
DDA	data dependent acquisition
EST	expressed sequence tag
FN	false negatives
FP	false positives
MS	mass spectrometry
MS/MS	tandem mass spectrometry
TP	true positives

Introduction

A novel experimental and computational technique based on multiple enzymatic digestion of a protein or protein mixture that reconstructs protein sequences from sequences of overlapping peptides is described in this SAND report. This approach, analogous to shotgun sequencing of DNA, is to be used to sequence alternative spliced proteins, to identify post-translational modifications, and to sequence genetically engineered proteins.

Background

It is now well recognized that there are many reasons why gene sequence information alone does not provide a complete and accurate picture of a protein final sequence and state of activity. For example, the current estimates put number of human genes at ~30,000 while the number of proteins is undetermined but is reported to be in the vicinity of 90,000-100,000. There is thus a discrepancy between information provided by the genome and the proteome, and this fact has been observed experimentally for both eukaryotic and prokaryotic organisms. For instance, it is postulated that 40-60% of human genes have alternative splice forms, i.e., after transcription from DNA to RNA gene transcripts can be spliced in different ways prior to translation into proteins. Genome sequences also do not enable elucidation of post-translational modifications (PTMs): many eukaryotic and some prokaryotic proteins, undergo chemical modifications such as addition of carbohydrate and phosphate groups after expression. Consequently, in the cases of alternate splicing and PTMs, the information from a single gene may encode many different proteins. Another complicating factor is that current informatics algorithms do not always accurately identify start and stop codons. Hence, for many prokaryotic organisms, a significant percentage of identified genes do not represent actual proteins. Genetically modified proteins present a final case where the proteome will not correspond to genome.

The primary method in proteomics analysis is protein identification by mass spectrometry (MS). A typical identification technique consists of enzymatic cleavage of one or more proteins followed by separation and MS or tandem MS (MS/MS) measurements of the resulting peptides. Experimentally measured peptide masses or product ion masses are compared to the masses of theoretical peptides and ions resulting from *in silico* digestion and gas phase fragmentation of protein sequences. The *in silico* calculations are performed on protein sequences stored in genome databases, and for all the reasons mentioned above, in most cases the theoretically calculated masses poorly represent experimental MS or MS/MS results. In particular, all protein identification codes are plagued by false-positive peptide sequence identifications.¹ To overcome the problem of protein identification, *de novo* protein sequencing using MS/MS data has been developed but is limited to short sequences. Protein sequencing can also be carried out using the traditional and time consuming Edman degradation technique. While the technique is more accurate than identification methods and enable *de novo* protein sequencing, it is

slow, requires large amounts of proteins, and has a low throughput. It also requires proteins to be very pure, i.e., it cannot sequence proteins in a mixture.

For these reasons, a new *de novo* protein sequencing method is needed that can 1) sequence proteins that are the result of alternate splicing, 2) sequence proteins that are known to undergo PTMs, and 3) sequence genetically engineered proteins. This SAND reports the development of a protocol, algorithms, and computer code that enables *de novo* sequencing of proteins from MS/MS spectra. The method is illustrated with the sequencing of alternative spliced forms of IRAK2 a kinase involved in the innate immune system.

Shotgun Protein Sequencing Methodology

This SAND report provides an overview of our approach to shotgun protein sequencing. The interested reader is directed to two journal papers for more detailed analysis of the method and its usefulness.^{2,3}

Technical Approach

Our approach, illustrated below, is composed of the following steps:

- 1) The proteins of interest are selected and purified.
- 2) The proteins are specifically cleaved in serial and parallel reactions using enzymatic digestion.
- 3) For products of each reaction, peptide masses and ion masses are measured using MS/MS.
- 4) For each MS/MS spectrum, peptides are searched using several peptide identification codes.
- 5) A score is computed matching the *in silico* and experimental MS/MS spectra.
- 6) In case of poor score the MS/MS spectrum is subjected to *de novo* sequencing, which is used to uncover full peptide sequences and partial sequence tags, which in turn are used to search in databases.
- 7) The overlapping identified or *de novo* sequenced peptides are used by an assembly algorithm to *de novo* reconstruct protein sequences.

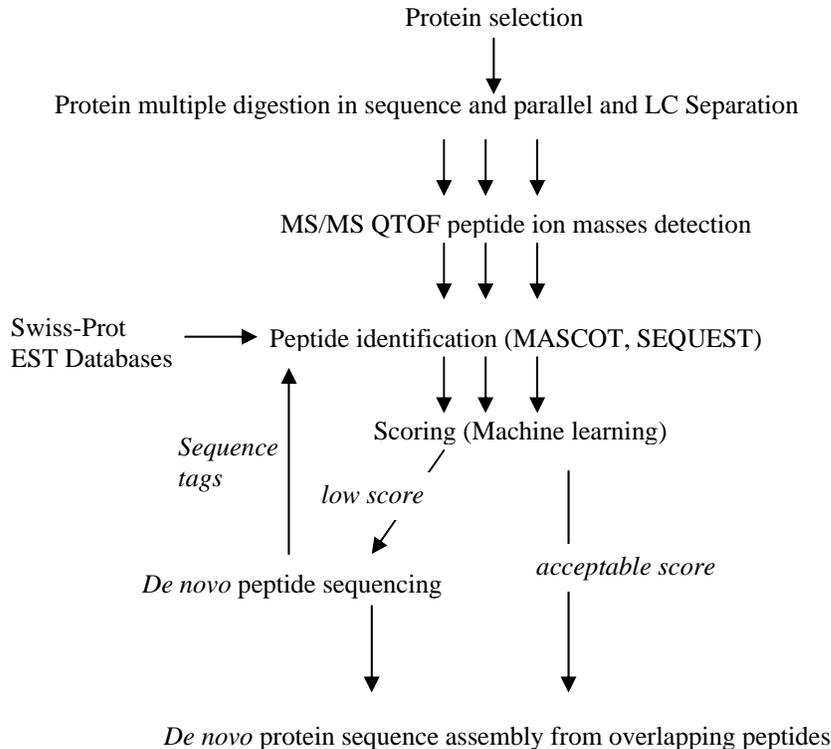


Figure 1. Process schematic for de novo protein sequencing strategy.

Proteins Selection

For each of the three applications outlined in Background, we selected appropriate proteins (and organisms) to study. To first validate our experimental and computational techniques, we used a set of commercially available proteins (among 18 proteins such as casein, carbonic anhydrase, cytoc, catalase, myosin, b-lactoglobulin etc.) that have been identified by other researchers using traditional MS/MS peptide identification methods.⁴ Next we targeted sequencing of proteins resulting from alternative splicing. Some examples include genes involved in sex determination in *Drosophila*, antibody response in humans and other

tissue or developmental stage of specific processes. We studied various tropomyosin variants, shown in Figure 2, produced by exon skipping for skeletal muscle, smooth muscle, fibroblast muscle, smooth muscle, fibroblast cells, liver and brain cells. We also studied the four alternative spliced forms of

IRAK2 (a,b,c and d), a kinase involved in transducing innate immune signals (cf. section 4). It appears that IRAK2a and IRAK2b enhance the signal, while IRAK2c and IRAK2d are inhibitory to the signal. For correcting genome annotation, candidates are proteins from *Desulfovibrio vulgaris* (an anaerobe involved in bioremediation of heavy metals and uranium). *D. vulgaris* has been sequenced and genome is available at TIGR website (www.tigr.org).

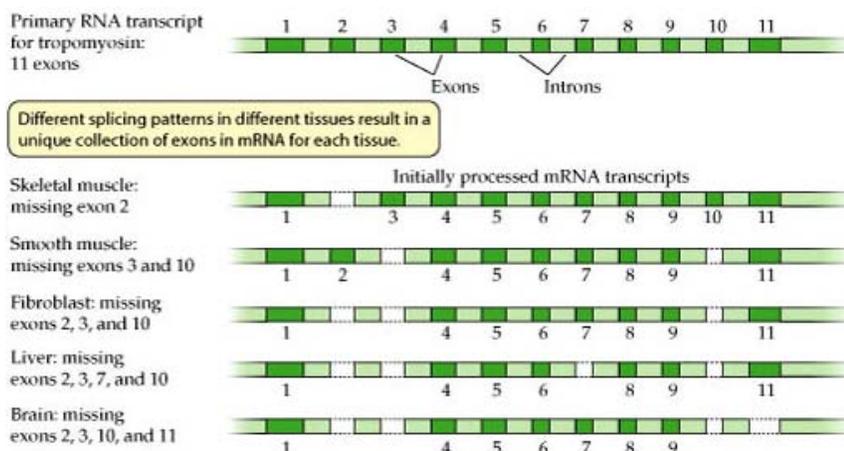


Figure 2. Tropomyosin Variants

Enzymatic or Chemical Digestion

Sets of mutually overlapping peptide fragments have been generated by chemically or enzymatically cleaving the purified protein in separate reactions. The table below lists some of the commercially available proteolytic agents that we utilized. It is of utmost importance that the proteolytic agents have different specificity in regards to the cleavage site(s). It is also important to choose agents that show high specificity, high yield (>90%), and high rate of reaction. Fortunately, a number of highly purified endoproteases and chemicals are commercially available that meet these requirements.

Proteolytic Reagent	Commercial Source	Amide Bonds Cleaved	Typical Reaction Conditions (buffer, pH, additives, temp, time)
Trypsin	Promega	K-X and R-X when X is P	50 mM ammonium bicarbonate, pH 7.8, 37 °C, 4-18 h
Endoprotease Lys-C	Roche	K-X	25 mM Tris-HCl, pH 7.7, 1 mM EDTA, 37 °C, 4-18 h
Endoprotease Lys-N	Seikagaku	X-K	100 mM phosphate buffer, pH 8, 25 °C, 18 h
Endoprotease Arg-C	Roche	R-X	10 mM Tris-HCl, pH 7.5, 2.5 mM DTT & 50 mM CaCl ₂ , 37 °C, 4-18 h
Endoprotease Asp-N	Roche	X-D	50 mM phosphate buffer, pH 7.8, 37 °C, 4-18 h
Endoprotease Glu-C (a.k.a. V8 Protease)	Roche	E-X, D-X	50 mM ammonium bicarbonate, pH 7.8, 37 °C, 4-18 h
Cyanogen bromide (CNBr)		M-X	0.1 M CNBr in 70% (v/v) formic acid, RT, 18h

Table 1. Commercially available proteolytic agents utilized in this work.

Our first attempt was to use trypsin-CNBr pair. Trypsin is highly specific and cleaves a polypeptide on the carboxy-side of arginine and lysine as long as the next amino acid on carboxy side is not proline. CNBr is very specific as well and cleaves after methionine. Methionine is a relatively low-abundant amino acid in a protein (average occurrence being 1.7%) and hence, CNBr cleavage for some proteins may lead to insufficient fragmentation.⁵ In those cases, we used other enzymes such as the endoproteinase Asp-N that is specific for cleavage after aspartic acid. A number of parameters require optimization to obtain fast and high-yield digestion including buffer and pH, temperature, time, and cleaving agent/protein ratio. For digestion in solution by enzymes, an enzyme/substrate ratio of 1:100 is used to minimize autolysis of the enzyme itself, but the low concentration of enzyme leads to longer incubation times (2 hours to overnight). An alternative is to use immobilized enzyme where a higher concentration of enzyme can be used without autolysis and significantly improves the speed of digestion.

MS/MS Data acquisition

The resulting peptide mixtures were analyzed by LC-MS/MS on a hybrid quadrupole-time-of-flight mass spectrometer, a Waters QTOF Ultima, fitted with a nanoelectrospray source. There are a number of distinct advantages for using a QTOF instrument for data acquisition including: 1) data dependant acquisition (DDA) capabilities – when MS/MS directly follows LC, data must be acquired "on the fly," 2) sufficient resolution (10,000) to differentiate precursor charge state (2+ versus 3+) and thus accurately determine the

peptide mass, 3) sufficient mass accuracy (less than 50 ppm) to greatly aid the sequence identification (cross-correlation analysis or *de novo* analysis), and 4) the ability to detect the presence of immonium ions in the MS/MS spectra indicative of peptide amino acid composition, greatly aiding *de novo* sequence analysis.

At the beginning of each run, the sample was loaded onto a trapping column and washed with 0.1% formic acid to remove any components (such as salts from the digest buffer) that would interfere with MS analysis. Acetonitrile was used to elute the peptides directly onto a C₁₈ column for reversed phase chromatography and subsequent MS and MS/MS analysis. Spectra were collected in DDA mode on the QTOF. In DDA mode the instrument toggles between MS and MS/MS experiments: A survey scan (MS) was acquired to identify the eluting peptide candidates for sequence analysis. The instrument was then used to select a number of these peptides according to parameters previously set by the user and to acquire MS/MS data on these peptides before switching back to acquire another survey scan.

MS/MS peptide identification and *de novo* peptide sequencing

We first used codes such as MASCOT and SEQUEST to identify peptides. Given our interest in *de novo* protein sequencing, however, these codes weren't used with specific genomes, rather we employed expressed sequence tag (EST) and other databases such as SWISS-PROT. ESTs are derived from fully processed mRNA, and thus provide a broad sample of mRNA diversity. We also used existing tools based on overlapping sequences to reconstruct the possible splice forms from ESTs.⁶ Both MASCOT and SEQUEST can be utilized on large databases and ESTs. The codes return a score (Xcorr and Mowse) based on correlation between predicted MS/MS spectra and experimental spectra, which are difficult to interpret. We then used a support vector machine method we developed learn to distinguish between correctly and incorrectly identified peptides. This task required performing an error analysis of MS/MS spectra based on multiple scans, and to establish relationships between peptide identification scores and rate of false positive identification, with the goal of establishing a new scoring metric to replace the current Xcorr and Mowse scores. For each MS/MS spectrum, if the identified peptide is found with high score it is kept for protein assembly, otherwise, the MS/MS spectra was fed to a *de novo* peptide sequencing algorithm. We employed modified *de novo* peptide sequencing techniques (e.g. PepNovo) and implemented our scoring metric into these sequence search algorithms. The resulting *de novo* algorithm was then able not only to sequence entire peptides but also to find sequence tags (the easier problem). These sequence tags were in turn used by identification methods (MASCOT, SEQUEST) with the overall goal of increasing the reliability of peptide identification/sequencing.

De novo protein assembly

The next step was to compile all peptide sequences in an overlapping graph. The construction of the overlapping graph was carried out as follows. For each peptide sequence retrieved from the MS/MS spectrum, a vertex was created. A color was assigned to the vertex depending on the origin of the peptide digest, and graph edges were added between pairs of vertices of different colors. Defining v_1 and v_2 as two such vertices, a directed edge was created from v_1 to v_2 if the suffix of v_1 overlapped with the prefix of v_2 , conversely a directed edge was created from v_2 to v_1 if the suffix of v_2 overlapped with the prefix of v_1 . Once all the edges are created, the protein sequence reconstruction problem is a classical physical mapping problem and reduced to the longest path problem in a directed acyclic graph. It is important to note that the problem of antisymmetric path has plagued *de novo* sequencing by making the sequence reconstruction intractable but does not apply here, an advantage of our approach.⁷ Further, when sequencing *de novo* protein from MS/MS spectra without using a site specific digest one typically creates a spectrum graph from the MS/MS ions. Each MS/MS ion has two corresponding vertices since every peak in the spectrum may be interpreted either as an N-terminal (*b* ion) or a C-terminal (*y* ion). In the present case, peptide sequences are oriented, or tagged, the last amino acid is the one at which the digestion occurred (Lys, Arg for trypsin digest, Met for CNBr digest, and Asp for Endo-Asp-N digest, etc), consequently, twins can be identified in the overlapping graph. This observation was a key element underlying our method's feasibility: it makes the reconstruction problem simple and computationally tractable.

Shotgun protein sequencing protocol development and benchmark

Carbonic anhydrase (CAII) was selected as our initial test case (i.e. a "standard" protein) because 1) CAII has been previously validated as a standard protein, 2) the molecular weight of CAII is 29 kDa, which does not represent either an extremely large or small sequence length, and 3) the sequence contains no cysteines, obviating the need for a reduction/alkylation step, which we deemed to be desirable for our first set of experiments. The details of the protocol development as well as the results can be found in the project's published papers.^{2, 3}

De novo sequencing of alternative splice forms

In murine macrophage cell line RAW 264.7, IRAK2 has four alternatively spliced forms. It appears that IRAK2a and IRAK2b enhance NF- κ B activity, while IRAK2c and IRAK2d are inhibitory in toll-like receptor (TLR) pathway. We digested IRAK2b with multiple enzymes, de novo sequenced the digestion products and reconstructed the sequence with the goal of being able to distinguish IRAK2b from the three other isoforms.

Expression and purification of alternatively spliced proteins

In June 2006, plasmid constructs of the four IRAK2 spliced protein genes were obtained from Professor Luke O'Neill at Trinity College, Ireland. The IRAK spliced genes were amplified using PCR technique from the plasmids, and inserted into five different expression vectors, pGEX2TK (GE Healthcare), pET-32 Ek/LIC, pET-41 Ek/LIC and pET-43.1 Ek/LIC (EMD Biosciences), pET101 (Invitrogen). Expression of the cloned genes was carried out *in vitro* in bacterial expression system *E.coli* BL21 (Invitrogen); it showed that all these vectors expressed the IRAK spliced proteins as insoluble inclusion bodies. Purification attempts revealed that the Nus-tagged IRAK proteins produced from the vector pET-41 Ek/LIC can be expressed and purified to the quantity and purity required for digestion using multiple enzymes.

Multiple enzymes digestion of IRAK2b

Using three enzymes Trypsin, Chymotrypsin and Proteinase K, and approximately 10 g Nus-IRAK2B, digestions were carried out at 37°C for 2 hours. Electrophoresis using SDS-PAGE confirmed that the protein was successfully digested. The digests were analyzed using MALDI; no reliable mass spectrometry data was generated. The digests were then further analyzed by Q-TOF, and MS/MS data generated was analyzed as described next.

IRAK2b sequence assembly

A total of 1306 MS/MS spectra for IRAK2b were analyzed by both MASCOT (a protein identification code) and PepNovo (a *de novo* peptide sequencing code). MASCOT was run against the entire SwissProt database (from which IRAK2b was removed) and returned 7598 sequences. As in our previous study with Carbonic Anhydrase II (CAII), MASCOT scores were ignored during the assembly process. PepNovo was run to generate sequence tags rather than peptide sequences because the probability of finding the right tags is higher than the probability of finding the right peptide sequence as long as tags are shorter in length than sequences. We carried out an *in silico* study demonstrating that most sequences can be uniquely assembled from their subsequences

of just five amino acids, consequently PepNovo was run to generate sequence tags of length five. PepNovo returned 6117 length-five tags. The MASCOT peptide sequences and PepNovo tags were fed to our longest path assembly algorithm and all longest subsequences were produced by piecing together peptide sequences and tags overlapping by at least four amino acids. The assembly algorithm was able to recover 186 amino acids (32.4%) among the 574 residues of IRAK2b. While this number is lower than in our previous study with CAII (82.7 % cf. FY06 LDRD report), this is still a notable achievement considering that in the present case only three enzymes were used, compared to the seven enzymes used with CAII. MASCOT returned far more information (142 residues) than PepNovo (44 residues). While peptides were search by MASCOT against SwissProt excluding IRAK2b, IRAK2a (the main isoform) was present in the database and consequently all the peptides returned by MASCOT are peptides shared by both IRAK2a and IRAK2b. MASCOT was thus not able to discriminate among the isoforms a and b, nor it was able to distinguished IRAK2b from the 4 known IRAK2 isoforms (a,b,c,d).

Using tag assembly to distinguished IRAK2b form other IRAK2 isoform

While only 44 residues were assembled when using PepNovo, could PepNovo alone have been used to identify IRAK2b? We raise this question not only because MASCOT cannot distinguish between IRAK2 isoforms, but also because we want to probe the utility of our assembly algorithm in a *de novo* situation, that is, in a situation where the protein to be sequenced is not found in a database. As our *in silico* study revealed, three enzymes are clearly not enough to assemble a sequence, however, a BLAST search may be able to identify homologues even for a low number of peptide and tags. To answer the above question all sequence tags of lengths 3-6 returned by PepNovo were collected. Tags that did not overlap with each other by 2-5 residues were removed. Assembled tags were ranked according to their probabilities (as computed by PepNovo), and for each assembled tag, BLAST was run against SwissProt. The results reported in Table 1 clearly identify IRAK2 as the top candidate sequence.

Table 2. Highest probability assembled tags fro IRAK2.

sequence	Probability	Best homologue	BLAST E-value
MANGSLHDRLWA	0.776	IRAK2	0.001
NGVAFAP	0.670	IRAK2	3.5
QASAAYLPE	0.430	IRAK2	7.5
KTDAPDSPQSYK	0.183	IRAK2	1.1
MENLVLYWAER	0.092	TRYCR Hypothetical protein	6.2
KSPEMLLELFR	0.077	SYNJB Glycyl-tRNA synthetase	0.79
RAGEVDDNDELDTSDTQHK	0.067	RPOC1 DNA-directed RNA polymerase	1.4

In order to discriminate among the four IRAK2 isoforms all tags of lengths 3-6 having a probability > 0.5 were searched in the four IRAK2 sequences. A total of 332 tags were found to belong to all four isoforms, 12 additional tags were found to belong to isoforms a,b,d, and 48 additional tags were found to belong to isoforms a and b. Thus using tag assembly we were able to identify IRAK2a and IRAK2b as the two highest ranking sequences. Additional digests should identify IRAK2b in the two remaining possibilities.

Conclusion

This work was published in two peer-reviewed journal articles “Use of a Designed Peptide Array to Infer Dissociation Trends for Non-Tryptic Peptides in Quadrupole Ion Trap and Quadrupole Time of Flight Mass Spectrometry”² published in *Analytical Chemistry* and “Genome Scale Enzyme-Metabolite and Drug-Target Interaction Predictions using the Signature Molecular Descriptor”³ published in *Bioinformatics*.

References

- ¹Vanable & Yates, *Anal. Chem.* 2004, 76, 2928
- ²S.P. Gaucher, J.A. Morrow, J-L. Faulon, Use of a Designed Peptide Array to Infer Dissociation Trends for Non-Tryptic Peptides in Quadrupole Ion Trap and Quadrupole Time of Flight Mass Spectrometry, *Analytical Chemistry*, vol 79, no. 20: 7822-30
- ³J-L. Faulon, M. Misra, S. Martin, K. Sale, and R. Sapro, Genome Scale Enzyme-Metabolite and Drug-Target Interaction Predictions using the Signature Molecular Descriptor, *Bioinformatics*, vol 24, no. 2, 2008: 225-33
- ⁴Keller et al, 2002, OMICS
- ⁵Klapper M.H., *Biochem. Biophys. Res. Commun.* 78, 1020, 1977
- ⁶Heber *et al.* *Bioinformatics* 2002, 18, S181
- ⁷Dancik, V. *et al.*, *J. Comp. Bio.* 1999, 327

Distribution

- 1 MS 1413 Grant Heffelfinger, 8630 (Electronic Copy)
- 1 MS 1413 Eric Ackerman, 8653 (Electronic Copy)
- 1 MS 9405 Glenn Kubiak, 8620 (Electronic Copy)
- 1 MS 0123 D. Chavez, LDRD Office, 1011 (Electronic Copy)
- 1 MS 0899 Technical Library, 9536 (Electronic Copy)