An Automated Approach to Top-Down Biomarker Analysis

Robert Barkovich[1], Andrew Forbes[2], Richard LeDuc[2], Neil Kelleher[2]

[1]Thermo Electron Corporation, San Jose, CA USA; [2]Center for Top Down Proteomics, University of Illinois, Urbana, IL, USA

Overview

Purpose: A method for identifying and characterizing biologically generated peptides via a novel approach to top down protein identification is demonstrated.

Methods: Using the Biomarker search mode from ProSight® PC software, a *Saccharomyces cerevisiae* proteolytic product not previously determined via other search methods is characterized.

Results: A biologically derived peptide - not identified from other search modes - is characterized with high probability. Traditional top-down and bottom-up methods are unlikely to have obtained such characterization, including start and stop sites for the peptide.

Introduction

Mass spectrometric protein identification can be achieved by one of two general strategies. Using the *bottom-up* approach, proteins are digested and the resulting peptides are analyzed by MS and tandem MS with identification provided by a protein database search of the product ion spectra [1,2]. The *top-down* approach, on the other hand, uses the mass of the intact proteins, individually or in mixtures, and then fragments the intact proteins inside the mass spectrometer without prior enzymatic digestion[3]. The advantages of top-down proteomics are the ability to measure the actual intact protein molecular weight, preserving both the entire protein sequence and the integrity of post-translational modifications. Currently, top-down proteomics are limited to FTICR instruments because of requirements for high resolving power, mass accuracy and complementary fragmentation methods. Intact protein and fragment molecular weights can be searched against a corresponding database in a manner similar to that of the bottom up approach in order to provide protein identification[4-7]. At this point, the standard database search engine is ProSight PTM developed by the Kelleher research group at the University of Illinois[8,9]. ProSight PC is being developed for Thermo Electron by the Kelleher group to bring top-down proteomics to every user on a desktop PC and includes many features, such as the Biomarker search mode, that can only be found in this software product.

FIGURE 1. Intact and fragment mass lists for absolute mass search and Biomarker Search modes

Fragmentation Method: CID **Ion Type**: BY

Intact Mass List

ID	MZ Monoisotopic	MZ Average	Mass Monoisotopic	Mass Average	Intensity
1	.0000	.0000	10870.2000	.0000	1

Fragment Mass List

ID	MZ Monoisotopic	MZ Average	Mass Monoisotopic	Mass Average	Intensity
1	749.0398	749.4427	3740.1624	3742.1770	6.605056e-006
2	779.7210	779.9728	3114.8549	3115.8621	1.468191e-005
3	816.1216	816.5693	7336.0288	7340.0578	5.167678e-006
4	830.1229	830.5706	7462.0405	7466.0696	1.292058e-005
5	832.1895	832.6212	5818.2754	5821.2972	1.061433e-005
6	850.3527	850.7844	5945.4182	5948.4400	1.418269e-005
7	872.6019	873.1055	8715.9464	8720.9828	4.015638e-006
8	883.9175	884.4211	8829.1018	8834.1382	4.207119e-006
9	902.8913	903.4407	9920.7241	9926.7677	1.035522e-005
10	904.9968	905.5004	7231.9159	7235.9450	1.312595e-005
11	907.2327	907.7364	7249.8038	7253.8329	1.770501e-005
12	909.5232	910.0727	9993.6757	9999.7193	2.351011e-005
13	921.1599	921.7093	10121.6784	10127.7220	1.83778e-005
14	958.7323	959.2817	10534.9748	10541.0184	1.199389e-005
15	968.5532	969.0568	9675.4593	9680.4956	7.560144e-006
16	978.4669	978.9706	9774.5965	9779.6328	6.885026e-006
17	991.2633	991.8677	9902.5606	9908.6042	4.746259e-006
18	996.5699	997.1743	9955.6262	9961.6698	8.157484e-006
19	998.4695	999.0739	9974.6227	9980.6664	1.321867e-005
20	1011.3672	1011.9716	10103.5996	10109.6432	1.112067e-005
21	1013.0718	1013.6762	10120.6455	10126.6891	8.886895e-005
22	1013.3788	1013.9831	10123.7151	10129.7587	3.439902e-005
23	1109.4053	1110.0768	9975.5819	9981.6255	1.116529e-005

FIGURE 2. Search results from Absolute Mass Search Mode of ProSight PC. This is a standard top-down search algorithm as used in ProSight PTM. Four hits were found with the best expectation score being $4.25e^{-3}$.

Methods

Cell Growth and Protein Processing. A 7 mm (Mini Prep Cell) i.d. preparative gel was utilized as the first dimension of fractionation after cell lysis, with 1 mg of total protein loaded. Cells of *Saccharomyces cerevisiae* (strain S288C) were grown under aerobic conditions to stationary phase. The wet cell mass was resuspended in 10 mL of lysis buffer containing 0.05 M Tris-HCl, pH 7.2, 2% ALS, 2-5 mM EDTA, 5-10 mM TCEP, 2 µL DNase, and 2 protease inhibitor cocktail tablets. After lysis by French press (15 000 psi), the cellular debris was removed via centrifugation.

Iodoacetamide was used to alkylate cysteines by adding a 50 µL aliquot of stock solution (125 mM in water) to 300 µL of the reduced yeast cell extract. The resulting mixture was incubated in the dark for one hour at room temperature. The alkylated yeast cell extract (300 µL) was loaded on a Mini Prep Cell (7 mm i.d.; Bio-Rad, Hercules, CA) following the instructions described in the manual. The whole cell extract of yeast was fractionated according to protein size using continuous elution gel electrophoresis. Fractions from the Mini Prep Cell were first precipitated with cold acetone to remove bulk acid-labile surfactant (ALS), used in place of SDS.

An HP 1100 binary pump (Agilent) and a splitter (1:100 splitting ratio, LC Packings) were used to generate a flow rate of 5 µL/min. Standard solvents (H_2O, CH_3CN) and 0.5% formic acid were used. The column was washed at 10% CH_3CN for 15 min. before a linear gradient over 25 min. from 20% to 55% solvent B. The RPLC fractions containing ~0.5-1.0 µg of total protein were collected every 2 min. providing a 10 µL sample volume. The peptide shown here eluted relatively early in the LC run, around 30% B.

Nanospray Robotics and ESI-Q-FTMS. Effluent from capillary RPLC (5-10 µL) was collected in a 96 well sample plate for the NanoMate™ 100 system (Advion BioSciences, Ithaca, NY). Acquisition of intact protein spectra (10 or 25 scans) was obtained using ESI with a custom Q-FTMS with an 8.5 T actively shielded magnet. In general, ions generated from positive ESI were directed through a heated metal capillary, skimmer, and multiple ion guides into the ion cell (~10^{-9} Torr) of the FTMS.

The main strategy for protein ion isolation and fragmentation was initiated by on-the-fly deconvolution of intact protein spectra followed by automatic isolation with SWIFT and fragmentation via IRMPD in a collision cell (50% power, 250-350 ms irradiation time). The *b* and *y* type ions produced are the same as would be formed via CID in an Finnigan™ LTQ FT™ instrument.

Spectral Processing and Database Searching. Intact protein spectra were analyzed by an in-house deconvolution program. MS/MS spectral analysis was performed by a modified THRASH algorithm and the resulting peak lists together with the intact protein mass values were run in ProSight PC Beta 2.5.

FIGURE 3. Results from Biomarker Search mode of ProSight PC. The yeast ribosomal protein L27 is identified with expectation value of 6.79e-28, while no reasonable hits were found using Absolute Mass Search mode

Results

A yeast protein with an intact mass of 10870.2*(-0)* Da was isolated and fragmented as described in the Methods section. Figure 1 shows the observed intact and subsequent fragment ion masses. The MS/MS data was searched using the absolute mass search mode of ProSight PC against all known and predicted protein forms (via shotgun annotation[10]) in the yeast Swissprot database of the ProSight Warehouse™ within a ±2000 Da window of the observed intact mass. The resulting expectation scores show the probability of a retrieved protein matching the number of fragment ions listed by chance. A fragment ion mass tolerance of 25 ppm was used. The top result of the search was a hypothetical protein of mass 11916.8*(-0)* Da with an expectation score of 4.25e-3, an unlikely match (Figure 2).

There are many biological events that can proteolyze an intact protein to a biologically relevant or active form within the cell. Many of these events are uncharacterized or not well understood. As a result, many of these proteolytic products are not listed in standard protein databases. Standard database search tools based upon these protein databases will fail. ProSight PC comes with a Biomarker search mode specifically to find these biologically relevant protein fragments, including those that include known and predicted splice variants. ProSight PC uses the mass of the biomarker to search along the possible protein sequences to identify peptides (and their subsequent fragment ions) that may correspond to the biomarker.

Using the Biomarker search mode of ProSight PC with intact and fragment mass tolerances of 25 ppm, we were able to find a good match (a yeast ribosomal protein) with an expectation value of 6.79e-28 (see Figure 3). Ten of 23 fragments were matched using these tolerances. The Sequence Gazer (Figure 4) can be used to refine the results of the search and assign putative PTMs in iterative, hypothesis-driven tests of the MS/MS data.

FIGURE 4. The sequence of the top hit shown using Sequence Gazer. Ten of 23 fragments were matched using our tolerances.

Conclusions

Top-down proteomics has been shown to be an effective method of acquiring protein identifications, particularly for modified proteins. We have shown a new method of deducing the identification of biologically relevant proteolytic products via top-down methods, circumventing the need for observing the intact protein, as is the case in traditional top-down experiments. In this study, the biomarker search method was shown to characterize a yeast protein not identified via a standard top-down database search method. Bottom-up experiments may have identified this biologically derived peptide, but obtaining the start and stop sites would have been nearly impossible. This search mode will aid in obtaining complete characterization of biologically generated peptides, many of which are considered biomarkers, including those resulting from splice variants and other biologically relevant modifications not previously discovered.

References

1. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. *Proc Natl Acad Sci U S A* 1993; **90**: 5011.
2. Yates JR 3rd, Carmack E, Hays L, Link AJ, Eng JK. *Methods Mol Biol.* 1999; **112**: 553.
3. Reid GE, McLuckey SA. *J Mass Spectrom.* 2002; **37**: 663.
4. Senko MW, Beu SC, McLafferty FW. *Anal Chem.* 1994; **66**: 415.
5. Mortz E, O'Connor PB, Roepstorff P, Kelleher NL, Wood TD,McLafferty FW, Mann M. *Proc Natl Acad Sci U S A* 1996; **93**: 8264.
6. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. *Nature Biotech.* 2001; **19**: 952.
7. Meng F, Cargile BJ, Patrie SM, Johnson JR, McLoughlin SM, Kelleher NL.*Anal Chem.* 2002; **74**: 2923.
8. https://prosightptm.scs.uiuc.edu
9. Taylor GK, Kim YB, Forbes AJ, Meng F, McCarthy R, Kelleher NL. *Anal Chem.* 2003; **75**: 4081.
10. Pesavento JJ, Kim YB, Taylor GK, Kelleher NL. *J Am Chem Soc.* 2004; **126**: 3386

Patents pending on Biomarker search mode of ProSight PC. ProSight is a registered trademark and ProSight Warehouse and Sequence Gazer are trademarks of the University of Illinois. Nanomate is a trademark of Advion Biosciences. All other trademarks are the property of Thermo Electron Corporation and its subsidiaries.

Thermo
ELECTRON CORPORATION