Tandem mass spectrometry of integral membrane proteins for top-down proteomics

Julian Whitelegge

Transmembrane domains of integral membrane proteins present a formidable analytical challenge due to their general lack of polarity; recovery is enhanced by maintaining their association with polar-loop regions. Multi-dimensional separations of intact integral membrane proteins coupled with high-resolution electrospray-ionization mass spectrometers enable "top-down" proteomics, with coverage of bilayer domains.

Keywords: Bilayer; Proteome; Proteomics; Thylakoid; Top-down; 2D chromatography

Julian Whitelegge*
The Pasarow Mass Spectrometry Laboratory, Departments of Psychiatry & Biobehavioral Sciences, Chemistry & Biochemistry, and The Neuropsychiatric Institute, the Brain Research Institute, and the Molecular Biology Institute, University of California, Los Angeles, CA, USA

*Present address: Department of Chemistry, 405 Hilgard Avenue, Los Angeles, CA 90095, USA.
Tel.: +1 310 794 5156;
Fax: +1 310 206 2161;
E-mail: jpw@chem.ucla.edu

1. Introduction

Biological systems abound with lipid/protein bilayer membranes that play critical roles in life. Membrane proteins fall into two general classes: the integral or intrinsic proteins that form part of the membrane and must be extracted by its solubilization; and, the peripheral or extrinsic proteins that are more loosely associated with the membrane and can be displaced without its destruction. The integral proteins can be divided into two general structural classes: those exhibiting the transmembrane β-barrel porin-type fold; and, those with one or more transmembrane alpha helices. Transmembrane helix domains can be quite accurately predicted and analysis of assigned genomic open reading frames (ORFs) suggests that proteins with this motif constitute around 30% of the proteome [1]. When one considers the transmembrane porins, short transmembrane unassigned ORFs and peripheral membrane proteins, it is clear that membrane-associated proteins make up as much as 50% of the proteome. The critical role of membrane proteins in cellular processes is further emphasized by the estimate that 70% of drug targets fall in this category [2]. Consequently, it is essential that proteomics technologies address the bilayer proteome [3]. Unfortunately, integral membrane proteins tend to have awkward physico-chemical properties that result in their under-representation in general proteomics experiments. The origin of this discrepancy is probably the tendency of the hydrophobic transmembrane domain to aggregate/precipitate when removed from the bilayer, exacerbated by the sporadic presence of cysteine residues with reactive thiols. Consequently, sample-preparation protocols must be specially developed.

The bulk of the work in proteomics has fallen to mass spectrometry (MS) and the significance of the discovery of soft ionization techniques for large biological macromolecules was recognized with the award of the Chemistry Nobel Prize to Fenn and Tanaka in 2002. Since proteins can be matched to genomic sequence data by short, unique internal sequences, early proteomics protocols have broken up intact proteins into sets of peptides for MS. Data from one or more peptides is then matched to the genome to identify the parent gene. Integral membrane proteins are generally amenable to this type of analysis because loop regions yield soluble peptides convenient for MS [4]. Intact mass [5] or "top-down" [6] proteomics aims to include intact protein-mass measurements along with primary structure information to describe the entire protein, including transmembrane domains. The realization of this goal for integral membrane proteins will allow us to investigate post-translational modification within the bilayer domain.

0165-9936/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.trac.2005.04.010

2. The separation challenge in proteomics

Typical genomes contain thousands of ORFs such that the living organism contains several thousand different proteins of varying size, sequence and chemical properties. The first large-scale separation technology applied in proteomics was the two-dimensional gel (2D gel) that provided a useful visual representation of the abundant components of a complex mixture [7]. The first dimension of isoelectric focusing is limited to non-ionic or neutral zwitterionic detergents for solubilizing proteins under denaturing conditions with their native charge. As a consequence, some integral membrane proteins tend to precipitate and become poorly represented in 2D-gel studies. Improvements in choices of chaotrope and detergents have addressed the problem [8–10] and today 2D gels are still quite widely employed in membrane protein research [4]. Alternative approaches have side-stepped the first dimension, going straight to the sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS–PAGE) step used in the second dimension of the 2D-gel analysis; they use organic solvents to pre-fractionate the membrane system [11,12] or non-denaturing separations for separation of integral membrane-protein complexes [13].

More recently, ''shotgun'' protocols have appeared, improving throughput for proteome analysis [14]. The shotgun approach fragments the proteome into peptides using a specific proteolytic agent, such as the enzyme trypsin (that cleaves the peptide bond C-terminal to Arg and Lys amino-acid residues, provided the following residue is not Pro). The resulting mixture contains hundreds of thousands of different peptides with abundances varying over several orders of magnitude, presenting a phenomenal separation challenge. As a result, 2D chromatography involving strong-cation exchange followed by sequential reverse-phase runs has been employed prior to MS (multi-dimensional protein identification technology; MuDPIT) [15].

Modifications to shotgun protocols have been necessary to cover integral membrane proteins adequately [16,17] and good coverage is now achieved. Recovery of peptides from transmembrane domains is limited, illustrating the disadvantage of the shotgun approach – proteins become represented by the peptides most easily recovered and many studies rely on single-peptide recoveries for large proportions of the proteins identified. On the other hand, supplemental information on membrane protein topology and phosphorylation status can be collected [17,18].

The loss of information accompanying shotgun approaches balances enhancements in sensitivity and proteome coverage resulting from the improved ionization efficiency observed with tryptic and other peptides [19].

3. Intact mass measurement of integral membrane proteins

For intact mass and ''top-down'' proteomics, it is necessary to obtain intact proteins in a reasonably pure state in solvents compatible with electrospray ionization (ESI)-MS. In the case of integral membrane proteins, a suite of technologies has been developed for this purpose, allowing mass measurements of proteins with 1–15 transmembrane domains and masses to over 100 kDa [20,21]. Liquid chromatography (LC) is used to purify the proteins in aqueous/organic solvent mixtures and protein-mass measurements achieve 0.01% accuracy (±3 Da at 30 kDa) demonstrating that the integral membrane proteome is compatible with ESI-MS [5].

Using these techniques, we were able to record a Fourier-transform mass spectrum (FT-MS) of the seven-transmembrane bacteriorhodopsin apoprotein, achieving better than 10-ppm mass accuracy, demonstrating for the first time that ''top-down'' proteomics by FT-MS is applicable to the integral membrane proteome [22]. An intact protein-mass measurement is useful because it defines the native covalent state of a gene product and associated heterogeneity [5]. Based on the measured mass of the cytochrome *b* sub-unit of the cytochrome b_6f complex, we concluded that a covalently bound heme was present [23], and that was confirmed the following year by high-resolution structures from two groups [24,25].

There are limits to the ability of separations and mass spectrometers to accommodate complex protein mixtures. Analysis of the cytochrome b_6f complex from photosynthetic membranes demonstrated that around 10 different polypeptides could be resolved by size-exclusion chromatography ESI-MS (SEC-MS) [23], with the limit for reverse-phase chromatography ESI-MS (RPC-MS) being in the range 50–100 proteins [26]. Furthermore, some integral proteins elute with low efficiency from reverse-phase columns, requiring subsequent regeneration of the column. Fig. 1 illustrates how intact integral membrane proteins can be recovered from a reversed-phase column for downstream ESI-MS. A sample of bacteriorhodopsin (50 μg protein, 10 μL) was dissolved in formic acid (90%, 90 μL) and subjected to RPC-MS using organic buffers containing isopropanol as originally reported [27] and developed [21]. Approximately 60% of the protein eluted during the RPC-MS run, allowing collection of the mass spectrum (not shown). To elute the residual protein, the reverse-phase column was coupled to a size-exclusion column and the system equilibrated in chloroform/methanol/aqueous 1% formic acid (4/4/1; v/v), as described in Fig. 1(a). An injection of formic acid (90%) released the remaining protein to the mobile phase, thereby regenerating the column for further experiments without the possibility of

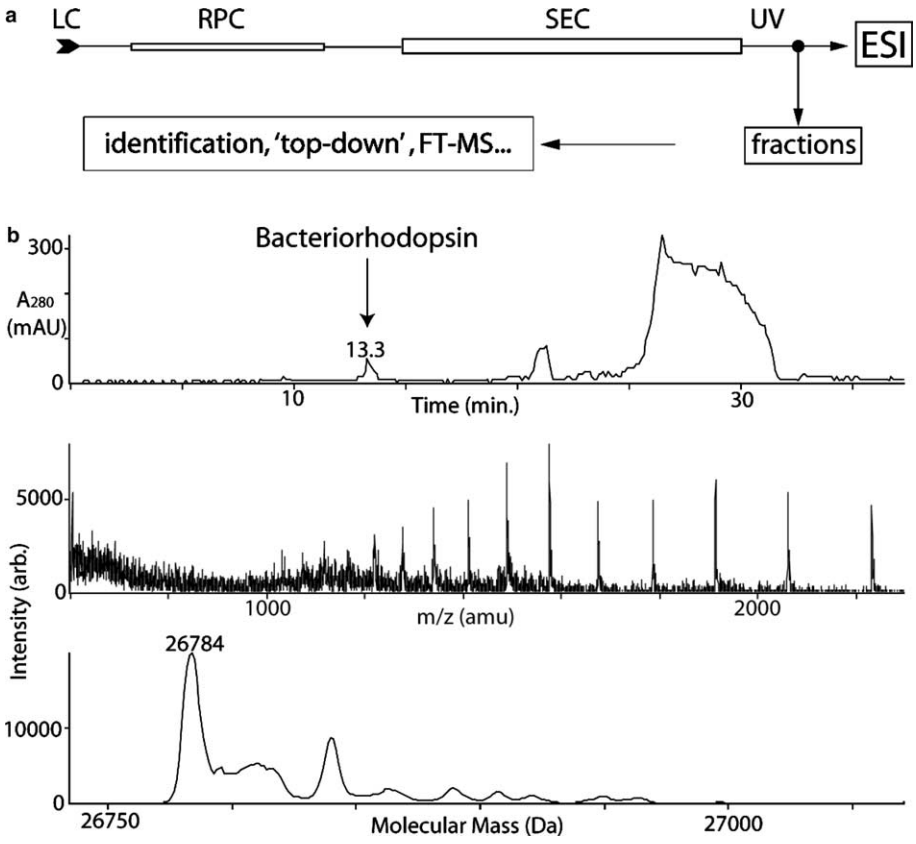

Figure 1. Recovery of integral membrane proteins from reverse-phase columns. (a) Schematic for regeneration of a polymeric reverse-phase column (RPC) with downstream size-exclusion chromatography (SEC) for recovery of membrane proteins for electrospray-ionization (ESI) MS and concomitant fraction collection (LC–MS+). (b) Residual bacteriorhodopsin was eluted from the RPC after equilibration of the system in chloroform/methanol/1% aqueous formic acid (4/4/1; v/v) and an injection of formic acid at time zero. The elution profile (upper panel) shows the protein was effectively separated from small molecules released from the RPC and from the formic acid, allowing ESI-MS (middle panel). The zero-charge reconstruction (lower panel) reveals a molecule with measured mass of 26,784 Da, within 0.01% of that calculated for the bacteriorhodopsin apoprotein residues 14–261 with N-terminal pyroglutamate (26783.55).

''ghost'' peaks. By including the size-exclusion column, the protein was separated from small molecule contamination, allowing the mass spectrum to be recorded (Fig. 1(b)). While this recovery/regeneration scheme is most useful, it does not overcome the limitations of spatial separation. We have therefore applied our protocols to purified integral membrane protein complexes [23] and membrane sub-domains [26,28] while application to complete membrane systems will require at least one more dimension of chromatography to provide reasonable coverage.

4. 2D chromatography

The goal of providing liquid samples for ESI-MS makes LC technologies preferable to gel-based technologies because recovery of intact proteins from gels is challenging and covalent side-reactions can modify the protein. 2D chromatography is therefore being evaluated for dissection of the membrane proteome.

Since we have demonstrated the ability to interface our LC systems to ESI-MS for successful analysis of integral membrane protein complexes, it is logical to reserve these for the second dimension and consider first dimensions that separate the different protein complexes of the membrane. Native gel experiments have demonstrated that at least some membrane systems can be solubilized with non-denaturing detergents, releasing intact membrane protein complexes for analysis. We have investigated SEC under non-denaturing conditions for separation of total membrane complexes with some success (Fig. 2(a)). For example, Photosystem 1 (PS1) could be enriched from total membranes of *Synechocystis* sp., PC 6803 allowing second dimension analysis of the two large sub-units, PsaA and PsaB, with 11 transmembrane helices and masses greater than 80 kDa. However, there was insufficient chromatographic resolution in the first dimension for analysis of the complexes of lesser abundance, without contamination by the dominant membrane complex. Respectable separations of this membrane system have been achieved using

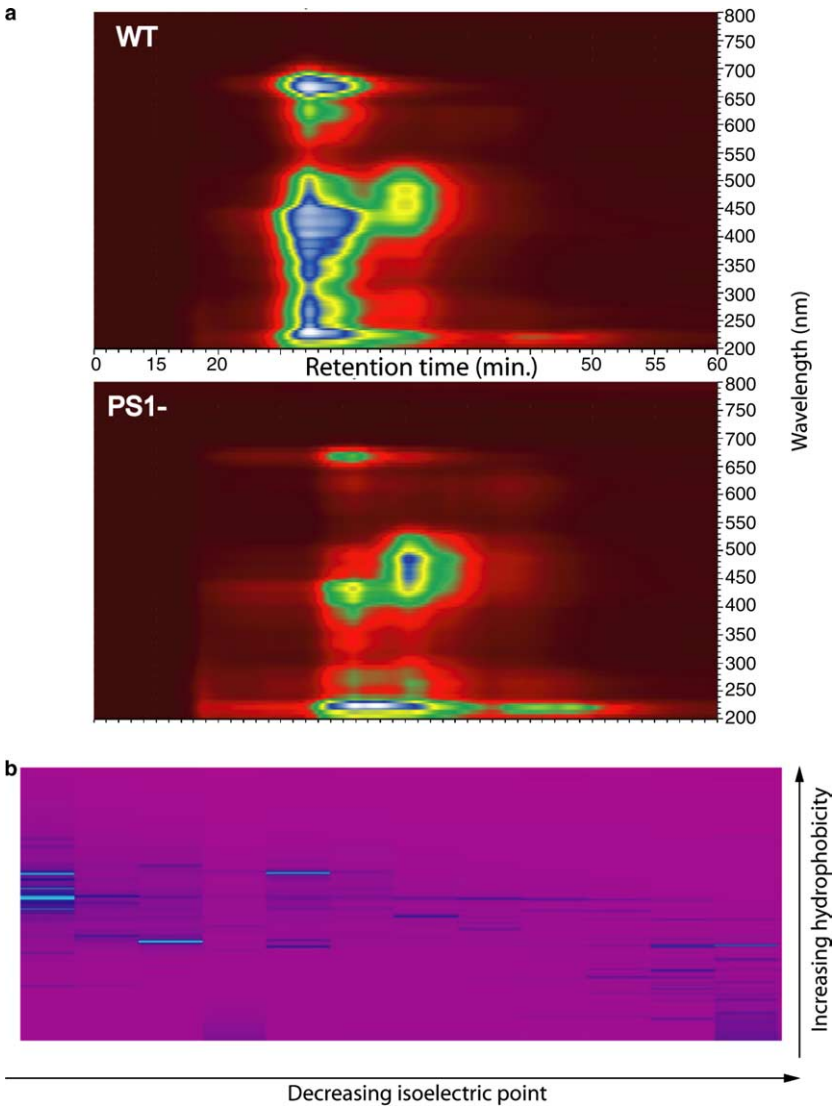

Figure 2. 2D chromatography of membrane proteins. (a) Non-denaturing first dimension separation of intact integral membrane protein complexes by size-exclusion chromatography (SEC). Total membranes from wild-type *Synechocystis* sp. PCC 6803 (WT) and a mutant that fails to accumulate Photosystem 1(PS1-) were solubilized in dodecylmaltoside prior to analysis. Eluent was passed through a diode-array detector (200–800 nm; Thermo) providing a compelling visual image of the changes in the pigmented photosynthetic complexes. (b) 2D chromatography of membrane-associated protein from wild-type *Synechocystis* sp. PCC 6803. Proteins were acetone precipitated prior to solubilization in a denaturing buffer for chromatofocusing on a commercial 2D-chromatography system (PF2D, Beckman Coulter). Fractions from the first dimension were separated by reverse-phase chromatography (RPC) with absorbance detection at 214 nm. The 2D image was rendered using Proteovue software (Beckman Coulter).

Blue-native gel technology [13], but this has not yet been interfaced with LC in the second dimension. An alternative 2D-chromatography system uses a denaturing first dimension step of chromatofocusing, which is performed under similar conditions to the isoelectric focusing separation performed with 2D gels (Fig. 2(b)). Fractions collected in the first dimension are separated by RPC with absorbance detection at 214 nm. While providing reproducible 2D chromatograms for quantitative proteomics, it is predictable that larger integral membrane proteins tend to precipitate before or during the first dimension and be difficult to recover from the reverse-phase columns used in the second dimension. Furthermore, the very useful proteomics information related to hypothetical protein–protein interactions in a complex is lost when a denaturing first dimension is used.

5. Top-down proteomics – technology

Once the integral membrane proteins are solubilized in the aqueous organic solvent mixtures used for RPC and

SEC, ESI is routine. The recorded mass spectrum yields protein-mass measurements of resolution and accuracy comparable to similar measurements on water-soluble proteins [5], so integral membrane proteins are amenable to Fourier-transform-ion cyclotron resonance MS (FT-ICR-MS) with the associated benefits to resolution, accuracy and so on [22]. Collision-activated dissociation (CAD) can be applied to intact integral membrane proteins and was used for tandem MS experiments to describe the primary structure of the four small sub-units of the cytochrome b_6f complex from both pro- and eukaryotes [23]. Interestingly, CAD did not yield any sequence information for the last few C-terminal residues of PetG from *Mastigocladus laminosus*, illustrating the need for alternative dissociation mechanisms. Electron capture dissociation (ECD) and the related electron-transfer dissociation (ETD) [29–31] have not yet been applied to integral membrane proteins, though this is predicted to be feasible. ECD, as well as ETD and related processes, rely upon reaction of an energetic electron with multiply charged protein/peptide ions via a non-ergodic mechanism that results in a more random backbone cleavage generating *c*- and *z*-type fragments distinct from the typical *b*-s and *y*-s from CAD. Integration of ECD with techniques to heat ions in the FT-ICR cell holds great promise, because dissociation is apparently enhanced when proteins are free of secondary and tertiary structural features [32].

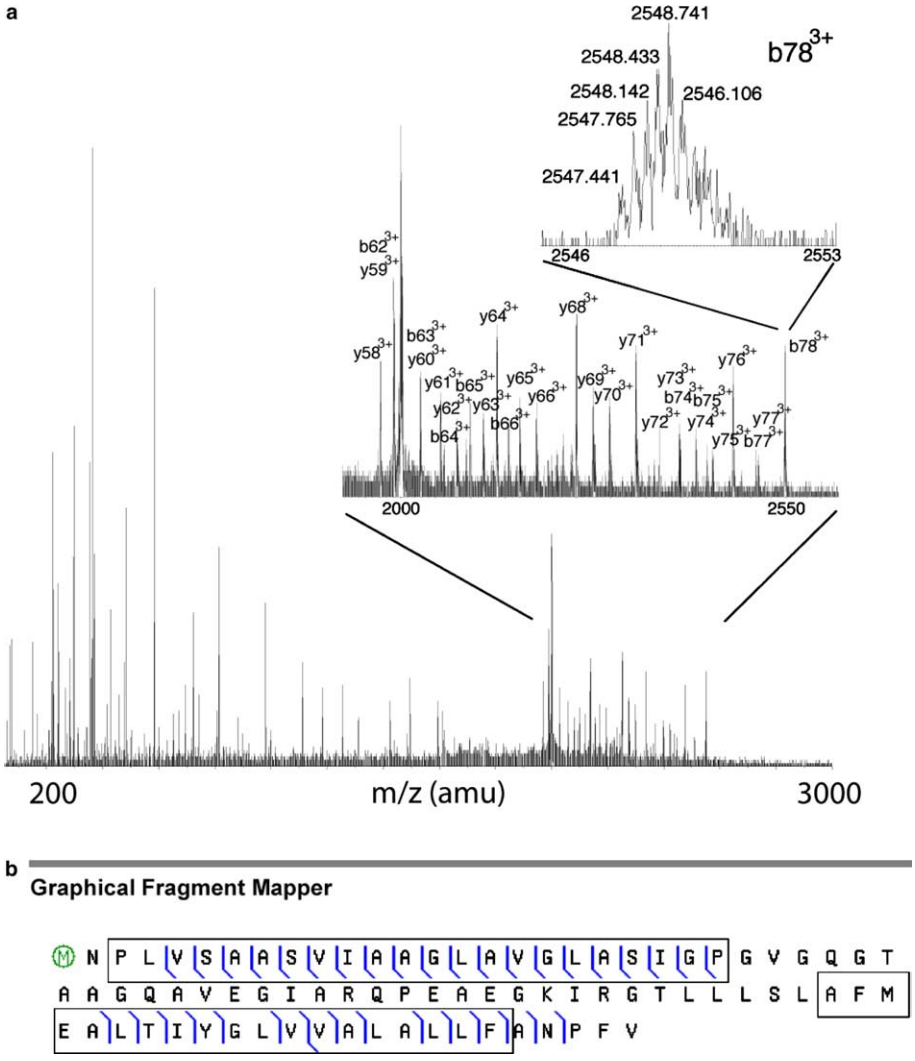

Figure 3. Top-down analysis of an integral membrane protein. (a) The ATP synthase sub-unit AtpH was isolated from *Arabidopsis thaliana* thylakoid proteolipids using size-exclusion chromatography (SEC) and tandem mass spectrometry (MS^2) on a quadrupole time-of-flight mass spectrometer (QSTAR XL, Applied Biosystems). MS^2 of the quadruply charged parent ion (2001 amu) yielded the tandem mass spectrum shown and specifically the set of triply charged product ions shown (expanded) corresponding to *b*62–*b*78 and *y*58–*y*77. (b) The masses of parent (8004 Da) and product molecules were submitted to Prosight PTM, identifying AtpH with high confidence and providing a graphical representation of the cleavage pattern. The transmembrane helices are boxed, demonstrating that collision-activated dissociation of the intact protein allowed sequencing through these domains. The N-terminal methionine residue is circled to indicate that it retains the formyl group of initiation.

Top-down analysis by MS was originally described for FT-MS only. However, any mass spectrometer with sufficient resolution to measure carbon-isotope spacing, and thus charge, on peptide/protein parent and daughter ions can be employed. A quadrupole-time-of-flight instrument, with sufficient resolution (~10,000 at m/z up to 1000) to measure charge state of ions with up to eight charges, was used to perform a top-down analysis on the integral membrane proteolipid that forms the major proton channel for ATP synthase (AtpH) (Fig. 3). The top-down mass spectrum demonstrates *b*- and *y*-series ions that read through both transmembrane helices. The information in the spectrum was sufficient to identify the protein using Prosight PTM, a web-based informatics tool for top-down proteomics [33]. The graphical output of Prosight is shown with the transmembrane domains highlighted to illustrate the sequence information derived from these domains (Fig. 3(b)). The intact mass and the tandem MS data all support retention of the formyl group on the initiating Met residue.

Top-down FT-MS experiments have traditionally been performed by direct sample infusion, to avoid changes in concentration that accompany LC. The development of new ion-trap FT-MS instrumentation that regulates the number of ions transferred to the FT-ICR cell is very exciting because it overcomes these earlier problems [34]. Using this technology, Wu and coworkers [35] achieved 95% sequence coverage of human growth hormone with a 200-fmol sample in a top-down LC–MS experiment. Top-down proteomics is coming of age and has been reviewed recently [36–38].

6. Ion sources for tandem MS of hydrophobic peptides

While ESI has covered our needs to date, alternatives are welcome in order to expand versatility. Photoionization has recently been applied to a collection of hydrophobic peptides, providing a potential alternative to ESI-MS. Interestingly, *c*- and *z*-type ions were present in the mass spectrum, suggesting some ECD/ETD-type chemistry occurring in the ionization source [39]. Extension of photoionization to proteins may be limited by the observation that singly charged ions dominate the mass spectrum.

7. Summary. An integrated solution

Integral membrane proteins can be analyzed by ESI-tandem MS, once they are purified in aqueous/organic solvent mixtures compatible with their solubility. The latest combination of linear ion-trap technology with Fourier-transform MS represents a particularly exciting development in this respect. Sample-preparation and separation technologies for reproducible, quantitative simplification of the bilayer proteome present the biggest challenge for an integrated solution.

Acknowledgements

Richard LeDuc, Yong-Bin Kim, Andrew Forbes and the rest of the software development team in Neil Kelleher's Laboratory are thanked.

References

[1] Y. Liu, D. Engelman, M. Gerstein, Genome Biol. 3 (2002) research0054.1.
[2] A.B. Weinglass, J.P. Whitelegge, H.R. Kaback, Curr. Opin. Drug Discov. Devel. 7 (2004) 589.
[3] J.P. Whitelegge, J. le Coutre, J.C. Lee, C.K. Engel, G.G. Privé, K.F. Faull, H.R. Kaback, Proc. Natl. Acad. Sci. USA 96 (1999) 10695.
[4] J.E. Nally, J.P. Whitelegge, R. Aguilera, M.M. Pereira, D.R. Blanco, M.A. Lovett, Proteomics 5 (2005) 144.
[5] J.P. Whitelegge, C.B. Gundersen, K.F. Faull, Protein Sci. 7 (1998) 1423.
[6] N.L. Kelleher, H.Y. Lin, G.A. Valaskovic, D.J. Aaserud, E.K. Fridriksson, F.W. McLafferty, J. Am. Chem. Soc. 121 (1999) 806.
[7] C. Hoogland, K. Mostaguir, J.C. Sanchez, D.F. Hochstrasser, R.D. Appel, Proteomics 4 (2004) 2352.
[8] V. Santoni, M. Molloy, T. Rabilloud, Electrophoresis 21 (2000) 1054.
[9] M.P. Molloy, B.R. Herbert, M.B. Slade, T. Rabilloud, A.S. Nouwens, K.L. Williams, A.A. Gooley, Eur. J. Biochem. 267 (2000) 2871.
[10] A. Görg, W. Weiss, M.J. Dunn, Proteomics 5 (2005) 826.
[11] M. Ferro, D. Salvi, S. Brugière, S. Miras, S. Kowalski, M. Louwagie, J. Garin, J. Joyard, N. Rolland, Mol. Cell. Proteomics 2 (2003) 325.
[12] G. Friso, L. Giacomelli, A.J. Ytterberg, J.-B. Peltier, A. Rudella, Q. Sun, K.J. van Wijk, Plant Cell 16 (2004) 478.
[13] M. Herranen, N. Battchikova, P. Zhang, A. Graf, S. Sirpiö, V. Paakkarinen, E.-M. Aro, Plant Physiol. 134 (2004) 470.
[14] M.P. Washburn, D. Wolters, J.R. Yates III, Nat. Biotechnol. 19 (2001) 242.
[15] J. Peng, J.E. Elias, C.C. Thoreen, L.J. Licklider, S.P. Gygi, J. Proteome Res. 2 (2003) 43.
[16] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nat. Biotechnol. 19 (2001) 946.
[17] C.C. Wu, M.J. MacCoss, K.E. Howell, J.R. Yates, Nat. Biotechnol. 21 (2003) 532.
[18] T.S. Nuhse, A. Stensballe, O.N. Jensen, S.C. Peck, Mol. Cell. Proteomics 2 (2003) 1234.
[19] J.P. Whitelegge, Proc. Natl. Acad. Sci. USA 99 (2002) 11564.
[20] J.P. Whitelegge, S.M. Gomez, K.F. Faull, Adv. Protein Chem. 65 (2003) 271.
[21] J.P. Whitelegge, Methods Mol. Biol. 251 (2004) 323.
[22] J.P. Whitelegge, Photosynthesis Res. 78 (2003) 265.
[23] J.P. Whitelegge, H. Zhang, R. Aguilera, R.M. Taylor, W.A. Cramer, Mol. Cell. Proteomics 1 (2002) 816.
[24] G. Kurisu, H. Zhang, J.L. Smith, W.A. Cramer, Science (Washington, DC) 302 (2003) 1009.

[25] D. Stroebel, Y. Choquet, J.-L. Popot, D. Picot, Nature (London) 426 (2003) 413.
[26] S.M. Gómez, J.N. Nishio, K.F. Faull, J.P. Whitelegge, Mol. Cell. Proteomics 1 (2002) 46.
[27] G.E. Tarr, J.W. Crabb, Anal. Biochem. 131 (1983) 99.
[28] S.M. Gómez, K.Y. Bil, R. Aguilera, J.N. Nishio, K.F. Faull, J.P. Whitelegge, Mol. Cell. Proteomics 2 (2003) 1068.
[29] R.A. Zubarev, Curr. Opin. Biotechnol. 15 (2004) 12.
[30] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Proc. Natl. Acad. Sci. USA 101 (2004) 9528.
[31] H.J. Cooper, K. Hakansson, A.G. Marshall, Mass Spectrom. Rev. 24 (2005) 201.
[32] K. Breuker, H.-B. Oh, C. Ling, B.K. Carpenter, F.W. McLafferty, Proc. Natl. Acad. Sci. USA 101 (2004) 14011.
[33] R.D. LeDuc, G.K. Taylor, Y.-B. Kim, T.E. Januszyk, L.H. Bynum, J.V. Sola, J.S. Garavelli, N.L. Kelleher, Nucleic Acids Res. 32 (2004) W340.
[34] J.S. Page, B. Bogdanov, A.N. Vilkov, D.C. Prior, M.A. Buschbach, K. Tang, R.D. Smith, J. Am. Soc. Mass Spectrom. 16 (2005) 244.
[35] S.L. Wu, I. Jardine, W.S. Hancock, B.L. Karger, Rapid Commun. Mass Spectrom. 18 (2004) 2201.
[36] N.L. Kelleher, Anal. Chem. 76 (2004) 197A.
[37] B. Bogdanov, R.D. Smith, Mass Spectrom. Rev. 24 (2005) 168.
[38] F. Meng, A.J. Forbes, L.M. Miller, N.L. Kelleher, Mass Spectrom. Rev. 24 (2005) 126.
[39] A. Delobel, F. Halgand, B. Laffranchise-Gosse, H. Snijders, O. Laprevote, Anal. Chem. 75 (2003) 5961.