

MudPIT: multidimensional protein identification technology

Claire M. Delahunty and John R. Yates III

Department of Physiological Chemistry, The Scripps Research Institute, La Jolla, CA, USA

Large-scale biology emerged out of the efforts to sequence genomes of important organisms. Based on resources created by whole genome sequencing, large-scale analyses of messenger RNA (mRNA) and protein expression are now possible. With the availability of large amounts of genomic sequence information, a convenient method for the identification and analysis of proteins based on proteolytic digestion into peptides emerged. Processes to fragment peptides using collision-activated dissociation (CAD) in tandem mass spectrometers and computer algorithms to match the tandem mass spectra of peptides to sequences in databases enable rapid identification of amino acid sequences, and hence proteins, present in mixtures. The inherent complexity of the peptide mixtures has necessitated improvements in methodology for mass spectrometry (MS) analysis of peptides.

Introduction

Multidimensional protein identification technology (MudPIT) (1) developed as a method to analyze the highly complex samples necessary for large-scale proteome analysis by electrospray ionization, tandem mass spectrometry (MS/MS), and database searching. As it is most frequently used, MudPIT couples a two-dimensional liquid chromatography (2D-LC) separation of peptides on a microcapillary column with detection in a tandem mass spectrometer (Figure 1). In the MudPIT experiment, a protein or mixture of proteins is first reduced (to break cysteine disulfide bonds), alkylated (to prevent reformation of disulfide bonds), and digested into a complex mixture of peptides. The digested sample is pressure-loaded directly onto a 100 μm i.d. fused silica microcapillary column that has a tip that has been pulled to 5- μm diameter with a capillary puller. The microcapillary column is packed with approximately 10 cm of C_{18} resin followed by 3 cm strong cation exchange (SCX) resin. If samples contain a high concentration of salt, they can either be desalted offline before loading onto the biphasic column, or they can be directly loaded onto a triphasic column, which has a second C_{18} phase directly upstream from the SCX phase. After loading, the column is mounted on a platform attached to the mass spectrometer and is aligned with the inlet. The column is interfaced with a quaternary high-performance liquid chromatography (HPLC) pump that supplies four different buffers through a three-way connecting T, which allows the flow from the HPLC to be split and a 100–300 nL/min flow rate to be maintained. A voltage supply is directly interfaced with the microcapillary column, so that peptides are ionized as they elute from the column. A typical separation strategy uses three buffers: (i) Buffer A [5% acetonitrile (ACN), 0.1% formic acid], (ii) Buffer B (80% ACN, 0.1% formic acid), and (iii) Buffer C (500 mM

ammonium acetate, 5% ACN, 0.1% formic acid). The microcapillary column is first washed with Buffer A, and then a short segment of Buffer C is applied, followed by a long gradient of increasing Buffer B. The cycle is repeated with an increase in the concentration of Buffer C in each step. Each application of Buffer C serves to displace a fraction of the peptides from the SCX resin to the hydrophobic C_{18} resin. Displaced peptides are then separated on the C_{18} resin using a reversed-phase (RP) gradient. Using this stepwise cycled separation scheme, peptides of similar isoelectric point (pI) are sequentially displaced from the SCX resin to the C_{18} , where they are separated on the basis of their size and hydrophobicity. The strength of the MudPIT separation lies in the orthogonality of the chromatographic phases: a sub-

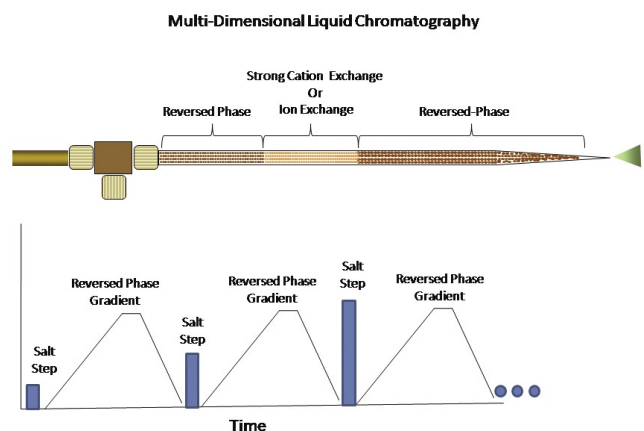


Figure 1. An integrated multidimensional liquid chromatography (LC) column can be used to create a multidimension separation. By using a series of salt steps, peptides can be eluted from the strong cation exchange (SCX) resin onto the reversed-phase (RP) resin. Peptides can then be eluted off the RP material using an increasing gradient of organic solvent. By increasing the concentration of salt at each step, a different population of peptides can be eluted.

set of peptides are selectively displaced (based on charge) from the SCX phase by controlling the salt concentration of Buffer C, and these peptides are separated on C_{18} (based on hydrophobicity) using an RP gradient. The number of salt steps and the length of the RP gradient can be modified to accommodate samples of different complexity. Once peptides are separated and eluted from the microcapillary column, they are ionized and enter the mass spectrometer, where they are isolated based on their mass-to-charge ratio (m/z). Peptide ions are selectively fragmented via collision-activated dissociation (CAD) in the tandem mass spectrometer using computer control of the experimental process. Tandem mass spectra are generated and are searched against a protein database using SEQUEST (2) or other database searching algorithms (3), allowing the peptide sequence and their protein of origin to be determined. These algorithms match the fragment ions of the peptide to those predicted from peptides of equal size in the database and determine the closeness of fit by mathematical means.

Applications of MudPIT

MudPIT has been used in a wide range of proteomics experiments, including large-scale catalogues of proteins in cells and organisms, profiling of organelle and membrane proteins, identification of protein complexes, determination of posttranslational modifications, and quantitative analysis of protein expression (4–9). In a recent study, MudPIT was used to identify spermatogenesis-specific proteins necessary for male fertility in *Caenorhabditis elegans* (10). MudPIT analysis was performed on chromatin-associated factors from two cell types: (i) those undergoing spermatogenesis

and (ii) those undergoing oogenesis. After subtraction of the 812 proteins associated with oogenesis from the 1099 identified in spermatogenic chromatin, 132 unique sperm-specific proteins were identified. RNA interference (RNAi) was used to knock down gene function, and 50 spermatogenesis-specific factors involved in DNA compaction, chromosome segregation, and fertility were identified. A combination of protein enrichment and a subtractive comparison between two different cell types, together with MudPIT, focused the analysis on important functional proteins.

RNAi has become an important tool for functional analysis of proteins, and as a result, researchers are trying to understand the mechanism of the process. A key component of the RNAi is Dicer (DCR-1). To identify DCR-1 interacting proteins in *C. elegans*, MudPIT was used to analyze DCR-1 complexes (11). The Dicer RNase III-related endoribonucleases are known to process long double-stranded RNAs (dsRNAs) into small duplex species that are involved in a number of small-RNA-mediated gene-silencing mechanisms. The interacting protein identifications were used to create deletion mutants in *C. elegans* and allowed the elucidation of protein function. This study uncovered new interacting proteins participating in novel silencing pathways and provided a first description of the biological role of Dicer.

Measuring the impact of gene deletions on protein expression can help determine the targets of specific pathways. To measure alterations in protein expression, quantitative analysis is necessary. In a recent example of MudPIT as a quantitative tool, Dong et al. measured protein abundance changes in response to perturbation of the *daf-2* gene in *C. elegans* (12). *daf-2* is known to encode an insulin-like growth factor receptor that regulates expression of a number of genes involved in metabolism and aging in *C. elegans*. Worms with *daf-2* mutations have been shown to have healthy life spans twice the length of normal. Using MudPIT, targets downstream of *daf-2* were directly measured at the protein level, allowing identification of proteins not regulated at the transcription level. Proteins in the *daf-2* mutated worms were labeled with the stable isotope ^{15}N , which allowed them to be distinguished and compared with ^{14}N proteins from the wild-type worm. MudPIT analysis of a 1:1 mixture of *daf-2* mutated and wild-type proteins identified 1685 worm proteins, 86 of which differ in abundance between *daf-2* mutated and wild-type. RNAi was used to disable the genes that encode seven of the identified proteins, and the effect on longevity was measured. In each case, life span was affected, indicating the role of the gene in influencing longevity.

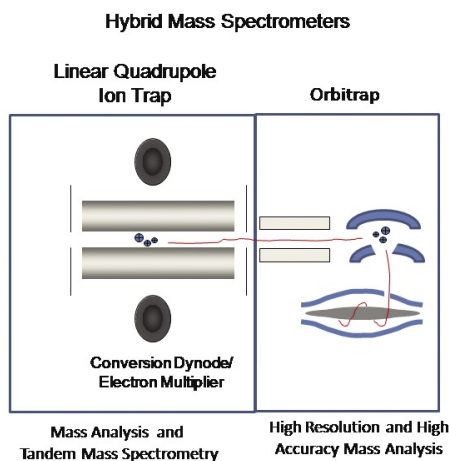


Figure 2. A typical hybrid mass spectrometer combines two different types of mass analyzers to achieve more powerful performance. A linear ion trap-Orbitrap is depicted, which combines a low resolution tandem mass spectrometer with a mass analyzer that can achieve high resolutions and high mass accuracy measurements. This configuration allows mass-to-charge ratio (m/z) values to be measured at either low or high resolution or at low or high mass accuracy. In addition, tandem mass spectra can be recorded with high resolution and high mass accuracy as well.

Advances in MudPIT Technology

The success of MudPIT for proteomics is a result of the two-dimensional resolution of peptides and the ability of database searching programs to identify proteins based

on a search with one or more peptides. By using peptides for identification, unbiased identification of proteins can be made; even proteins of relatively low abundance, extreme hydrophobicity or pI, and large molecular weight can be identified. MudPIT has been proven robust for the resolution of complex mixtures of peptides, but given the enormous complexity of the proteome, improvements in the resolution of peptides are desirable to identify proteins present at widely varying expression levels and to provide increased protein coverage. A new strategy, ultra-high-pressure MudPIT (UHP-MudPIT) (13), improves the separation of peptides on a MudPIT column by increasing the length of the C₁₈ phase and reducing the particle size of the resin used. While the advantage of increased C₁₈ phase length and reduced particle size to improve resolution has long been recognized, the application to online MudPIT has been limited by the lack of commercial HPLC instruments able to handle the back pressure necessary to accommodate adequate flow through these columns. The columns typically require pressures in excess of 10 kpsi, while conventional HPLC instruments are generally limited to pressures below 6 kpsi. However, recent offerings from LC manufacturers have made HPLCs available that are capable of operating pressures of up to 15 kpsi. UHP-MudPIT uses a 60-cm analytical column packed with 3 μm RP packing material coupled to a standard biphasic SCX/C₁₈ trapping column. Using increased pressure to maintain a flow rate of approximately 125 nL/min, a 30% increase in protein identifications was achieved over conventional MudPIT analysis of a soluble fraction of yeast lysate.

As illustrated in the UHP-MudPIT method, efforts to improve resolution in MudPIT have generally focused on the C₁₈ separation. We have recently reported a method to increase peptide identification through modification of the SCX phase. Orthogonal separation through the coupling of SCX to C₁₈ in an online MudPIT is possible because the method uses elution conditions advantageous to both stages of the separation. SCX separation is typically run under low pH conditions to enhance interactions between the protonated basic amino acid residues of the analyte and the SCX resin. Under these conditions, resolution of peptides on SCX is compatible with RP separation, which also requires low pH to decrease undesirable interactions between the sample and the C₁₈ resin. In addition, the elution conditions used in the MudPIT method are compatible with direct elution into the mass spectrometer and result in enhanced sensitivity in electrospray ionization. By contrast, strong anion exchange (SAX) resin cannot be coupled to RP in an online fashion, since elution from the SAX resin requires neutral-to-basic pH, which is incompatible with both the C₁₈ resin and elution into the mass spectrometer. Anion and cation mixed-bed ion exchange (ACE) (14) is a strategy that uses a mixture of anion and cation exchange resins instead of SCX in the MudPIT

experiment. By mixing anion and cation exchange resins in a single dimension, it is possible to take advantage of the anion-exchange mode while still using a low pH buffer system. ACE shows increased orthogonality to C₁₈ over SCX through increased retention of acidic proteins and reduced retention of neutral-to-basic peptides, resulting in the separation of acidic and basic peptides in a single chromatographic step. When applied to the tryptic digest of a whole cell lysate, ACE resulted in a 100% increase in the number of peptides identified over SCX alone. ACE is particularly effective for the separation of phosphopeptides, which are characteristically acidic. The use of ACE to analyze an enriched sample of phosphopeptides from a HeLa nuclear extract showed an increase in phosphopeptide identifications by 94% over SCX. ACE methodology resulted in an increase in the total number of phosphopeptides identified, but particularly in the number of acidic phosphopeptides (pI 3.00–4.99).

The growing analytical challenges of proteomics have driven advances in MS technologies. The most common arrangement for a MudPIT experiment couples electrospray ionization of peptides with CAD activation in a quadrupole or linear ion trap mass spectrometer. Ion trap mass spectrometers allow ions of specific m/z to be trapped and fragmented. The fragment ions can then be scanned to generate MS/MS spectra, or they can be isolated and further fragmented to generate higher order MS/MS spectra. Ion traps have been extensively used for proteomics because they are versatile, easy-to-use, and provide high sensitivity at relatively low cost. Recently, advances in MS technologies have provided a dramatic improvement in instrument performance through the creation of high-power hybrid instruments. These hybrid mass spectrometers couple the versatility of the ion trap with higher mass accuracy detection through novel geometric arrangements of existing mass analyzers (Figure 2). Hybrid ion trap-Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry (FTMS) (15) allows the fragmentation of peptides or proteins in the ion trap and detection with the highest level of mass accuracy and resolution by FT-ICR. The hybrid linear quadrupole ion trap (LTQ)-Orbitrap (16) combines a linear ion trap with an Orbitrap mass analyzer, a novel mass analyzer that provides the high mass accuracy of FTMS with the lower cost and ease of operation of an ion trap instrument. The result is an instrument capable of high mass accuracy, high resolution, large ion capacity, and large dynamic range. One example of the power of these instruments is the analysis of modifications to histones. Histones have an array of different types of modifications that help control transcription of genes. It is important to identify the state of modification of a histone as a function of the other modifications present, and these new types of instruments allow longer stretches of amino acid

sequence to be analyzed (17). These instruments and future developments in MS will provide exciting new options for proteomic experiments and enable exciting and powerful biological experiments.

Technologies like MudPIT for the analysis of proteins have accelerated biological discovery. Advances come from the development of new and improved separation methods and strategies, mass spectrometers, and computer software. This creates a highly dynamic technological environment in the field of proteomics, permitting new applications and driving new discoveries. As new developments are reduced to practice and commercialized, many of these methods and technologies will provide nonspecialists with the ability to conduct their own experiments.

References

1. Link, A.J., J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, and J.R. Yates III. 1999. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17:676-682.
2. Eng, J., A. McCormack, and J.R. Yates III. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a database. *J. Am. Mass Spectrom.* 5:976-989.
3. Sadygov, R.G., D. Cociorva, and J.R. Yates III. 2004. Large-scale database searching using tandem mass spectra: looking the answer up in the back of the book. *Nat. Methods* 1:195-202.
4. Cagney, G., S. Park, C. Chung, B. Tong, C. O'Dushlaine, D.C. Shields, and A. Emili. 2005. Human tissue profiling with multidimensional protein identification technology. *J. Proteome Res.* 4:1757-1767.
5. Yates III, J.R., A. Gilchrist, K.E. Howell, and J.J. Bergeron. 2005. Proteomics of organelles and large cellular structures. *Nat. Reviews* 6:702-714.
6. Speers, A.E. and C.C. Wu. 2007. Proteomics of integral membrane proteins-theory and application. *Chem. Rev.* 107:3687-3714.
7. Cantin, G.T., J.D. Venable, D. Cociorva, and J.R. Yates III. 2006. Quantitative phosphoproteomic analysis of the tumor necrosis factor pathway. *J. Proteome Res.* 5:127-134.
8. Florens, L., M.P. Washburn, J.D. Raine, R.M. Anthony, M. Grainger, J.D. Haynes, J.K. Moch, N. Muster, et al. 2002. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 419:520-526.
9. Fujii, K., T. Nakano, T. Kawamura, F. Usui, Y. Bando, R. Wang, and T. Nishimura. 2004. Multidimensional protein profiling technology and its application to human plasma proteome. *J. Proteome Res.* 3:712-718.
10. Chu, D.S., H. Liu, P. Nix, T.F. Wu, E.J. Ralston, J.R. Yates III, and B.J. Meyer. 2006. Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. *Nature* 443:101-105.
11. Duchaine, T.F., J.A. Wohlschlagel, S. Kennedy, Y. Bei, D. Conte, Jr., K. Pang, D.R. Brownell, S. Harding, et al. 2006. Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* 124:343-354.
12. Dong, M.-Q., J.D. Venable, N. Au, T. Xu, S.K. Park, D. Cociorva, J.R. Johnson, A. Dillin, and J.R. Yates III. 2007. Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* 317:660-663.
13. Motoyama, A., J.D. Venable, C.I. Ruse, and J.R. Yates III. 2006. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomics samples. *Anal. Chem.* 78:5109-5118.
14. Motoyama, A., T. Xu, C.I. Ruse, J.A. Wohlschlagel, and J.R. Yates III. 2007. Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal. Chem.* 79:3623-3634.
15. Syka, J.E., J.A. Marto, D.L. Bai, S. Horning, M.W. Senko, J.C. Schwartz, B. Ueberheide, B. Garcia, et al. 2004. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* 3:621-626.
16. Yates III, J.R., D. Cociorva, L. Liao, and V. Zabrouskov. 2006. Performance of a linear ion-trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* 78:493-500.
17. Mikesch, L.M., B. Ueberheide, A. Chi, J.J. Coon, J.E. Syka, J. Shabanowitz, and D.F. Hunt. 2006. The utility of ETD in proteomic analysis. *Biochim. Biophys. Acta* 1764:1811-1822.