Shotgun proteomics: Tools for the analysis of complex biological systems

Christine C Wu* & Michael J MacCoss

Address
Department of Cell Biology, SR11
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla
CA 92037
USA
Email: chriswu@scripps.edu

*To whom correspondence should be addressed

Current Opinion in Molecular Therapeutics (2002) **4**(3):242-250
 ISSN 1464-8431

Recent interest in proteomics has been fueled by the completion of multiple genome projects and ignited by the common need of biologists to rapidly and comprehensively evaluate complex samples of proteins on a global level. 'Shotgun proteomics' refers to the direct analysis of complex protein mixtures to rapidly generate a global profile of the protein complement within the mixture. This approach has been facilitated by the use of multidimensional protein identification technology (MudPIT), which incorporates multidimensional high-pressure liquid chromatography (LC/LC), tandem mass spectrometry (MS/MS) and database-searching algorithms. This review will focus on the most recent advances in methodologies for shotgun proteomics and address the limitations of the application of each to real biological samples.

Keywords Covalent modifications, mass spectrometry, MudPIT (multidimensional protein identification technology), quantitation, shotgun proteomics

Introduction

Cellular processes are mediated through protein-protein interactions, which are regulated spatially and temporally within the cell. While the genome of a cell is a static blueprint, the proteome of that same cell is dynamic and dependent on varying environmental conditions and developmental cues. Historically, biologists have taken a reductionist approach to studying the cellular functions of proteins, targeting specific proteins and looking for direct localizations and interactions to elucidate their biological roles. In this era of systems biology approaches, it is becoming increasingly clear that to truly understand a protein it is necessary to examine all its interactions concurrently because modulations in function are frequently the result of upstream, downstream and/or parallel interactions. But is this currently possible? Multiple strategies have been developed to systematically and comprehensively profile biological systems. 'Shotgun proteomics' refers to the direct and rapid analysis of the entire protein complement within a complex protein mixture. Implicit in this methodology is the ability to monitor the system both qualitatively and quantitatively. A comprehensive proteomic analysis should ideally include the following functions: (i) identify the entire protein complement; (ii) detect covalent modifications; and (iii) allow for quantitative comparisons between samples. This review will focus on the most recent advances in methodologies for shotgun proteomics and address the limitations of the application of each as applied to real biological systems.

Methods for global protein identification

Two-dimensional gel electrophoresis (2-DE) has traditionally been coupled with mass spectrometry to provide global proteomic analyses of complex protein samples. In this approach, proteins within a complex mixture are initially resolved by 2-DE (first by charge, or isoelectric point (p*I*) and then by relative molecular mass (M_r)) to reduce the complexity of each protein sample into resolved spots. Individual resolved spots are then sequentially analyzed by mass spectrometry (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) or electrospray ionization tandem mass spectrometry (ESI-MS/MS)). The laborious nature of this type of analysis hampers throughput and has only recently been minimized with advances in automation. However, major limitations to this approach still exist: (i) only the most abundant proteins are identified due to limitations in detection methods, as the dynamic range of most protein labels/stains is small compared to the dynamic range of protein levels in a given proteome; (ii) proteins with extreme p*I* (< 4 or > 9) and masses < 15 or > 200 kDa are not resolved, and complex samples frequently require multiple gels to resolve the entire p*I* and M_r range; (iii) membrane proteins are still under-represented due to poor solubility in the sample buffer and resolution in the gel; (iv) modified proteins are sometimes visualized as multiple spots and covalent modifications are difficult to assign; and (v) quantitation methods (usually by stain/densitometry or incorporation of radioactivity) are not practical, largely due to the limited detection range. However, despite these shortcomings, 2-DE remains the most common technique used for proteomic analyses and can often provide novel biological insights. Recent progress has been made in 2-DE technology and are reviewed in detail elsewhere [1•,2,3].

Multidimensional protein identification technology (MudPIT) provides a solution to overcoming many of the current limitations of 2-DE technology. This method combines the resolving power of high-pressure liquid chromatography (HPLC), the peptide analysis capabilities of tandem mass spectrometry (MS/MS) and the analytical power of database searching software [4•]. In this methodology (Figure 1), a complex protein sample (either soluble proteins or membrane proteins) is first digested with proteases to produce an even more complex peptide mixture. This peptide mixture is then loaded directly onto a fused-silica microcapillary column (100 µm inner diameter) packed with both reverse phase (RP) and strong cation exchange (SCX) stationary phases. Once the sample is loaded, no additional sample handling is required. The multiple dimensions of solid phase packing material facilitates peptide separations using two different physical characteristics (hydrophobicity and charge) [5••,6]. As peptides are eluted from the column, they are ionized, isolated by mass:charge ratio, and selectively fragmented by mass spectrometry. The peptide fragmentation spectrum is a 'fingerprint' that can be used to identify the protein from

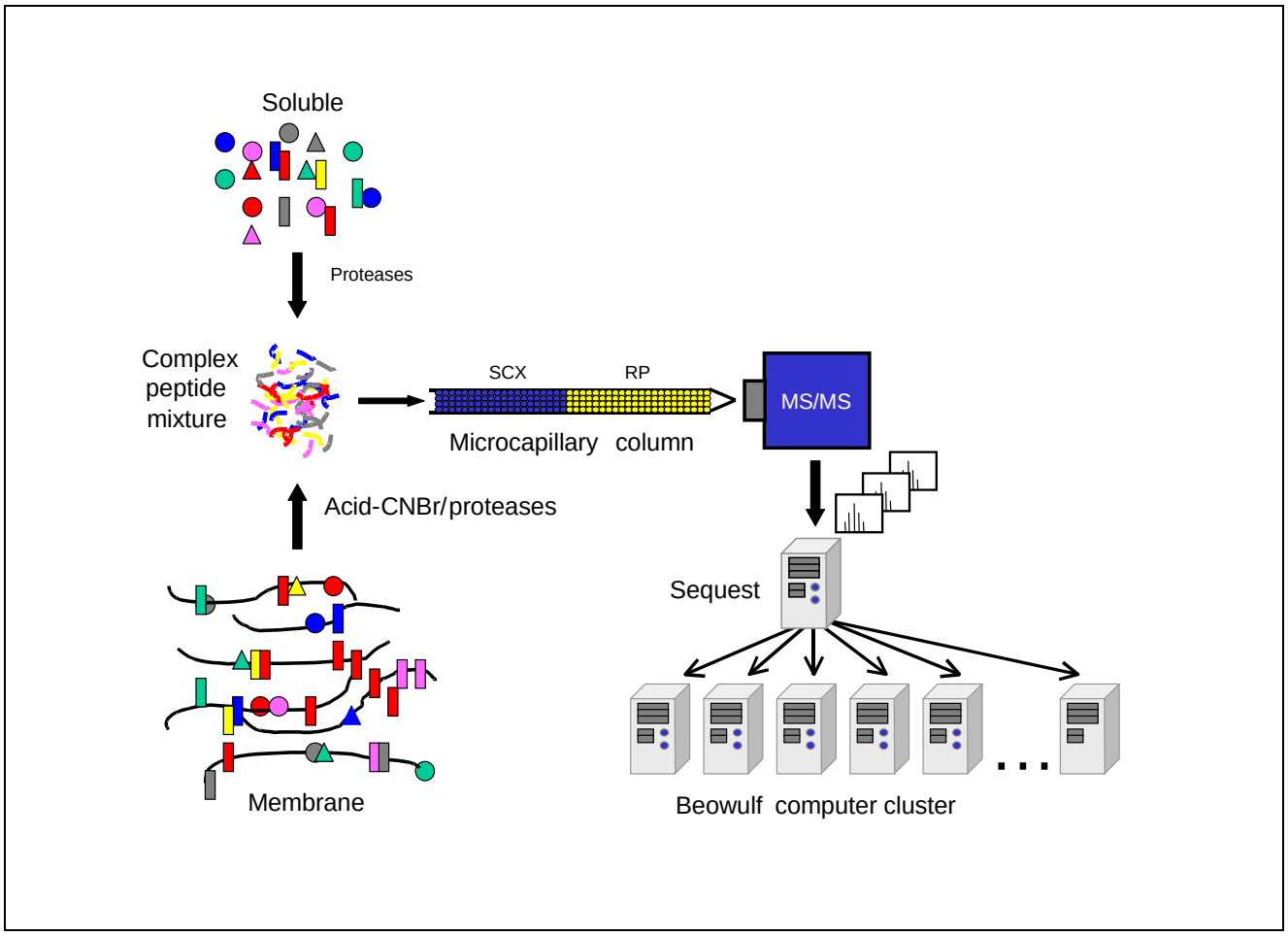
which they derive by searching against a sequence database with commercially available software (eg, SEQUEST). This method is largely unbiased, allowing both high- and low-abundance proteins, and proteins with extremes in p*I* and M_r, to be identified with equal sensitivities in the low femto-mole level [5••].

Modifications to the sample preparation have also allowed for the analysis of membrane proteins by MudPIT. It has been predicted that ~ 30% of all open reading frames in the human genome encode for membrane proteins [7], yet membrane proteins are largely under-represented in most proteomic analyses to date. Washburn and colleagues developed a method in which membranes were fractionated from *Saccharomyces cerevisiae* and solubilized in 90% formic acid. The proteins were then hydrolyzed using CNBr, and digested with endoproteinase-LysC and trypsin. The membrane acid solubilization step coupled with CNBr cleavage facilitated the analysis of peptides from hydrophobic proteins. This method, combined with MudPIT (Figure 1), resulted in the identification of 19% of the total membrane proteins predicted from the genome [5]. This study was impressive because it represented the first shotgun proteomic analysis that included a substantial proportion of membrane proteins. Recent progress by Wu and co-workers further improved on the methodology for proteomic analysis of membrane proteins by coupling protease-protection strategies with MudPIT technology to enable the global mapping of membrane protein topology [8].

Methods for global identification of covalent modifications

To provide comprehensive insight into protein function, covalent modifications of proteins need to be identified and included in proteomics analysis. There have been over 200 different protein modifications described in the literature to date [9], and because post-translational modifications cannot be determined by genomic sequence information alone, the direct identification of the type and position of the modification on the protein is vital to understanding its biological function [10].

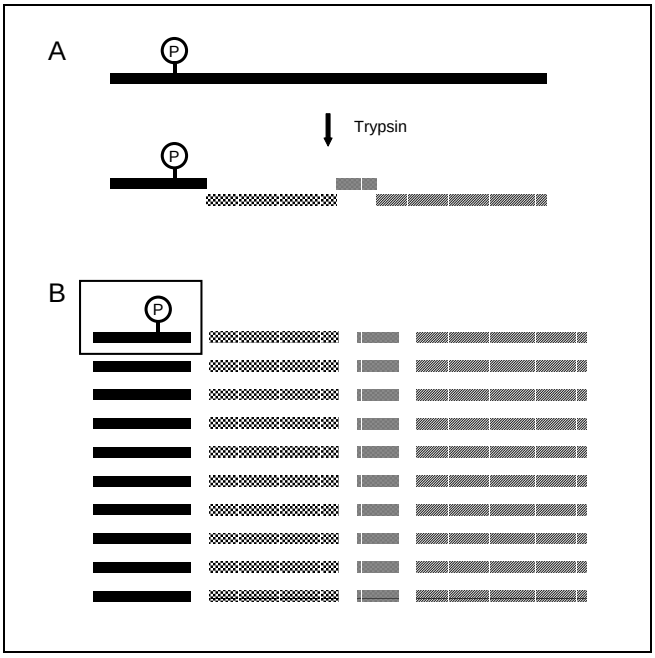
Figure 1. Multi-dimensional protein identification technology.



A complex soluble or membrane sample is digested with proteases to produce an even more complex peptide mixture, which is loaded directly onto a fused-silica microcapillary column (100 µm inner diameter) packed with RP material followed by SCX. The column is placed in-line with the mass spectrometer as part of a microelectrospray source and peptides are eluted sequentially into the mass spectrometer. Peptide fragmentation mass spectra are processed off-line using commercially available software (SEQUEST).

Methodologies based on mass spectrometry are ideal for this type of analysis because covalent modifications result in a change in peptide mass. However, the analysis of covalent modifications in protein mixtures has historically been difficult. In Figure 2A, tryptic digestion of Protein X results in four peptide fragments. The analysis of any of these four fragments by tandem mass spectrometry is often sufficient to identify the protein from which they derive. However, to identify and map the covalent modifications on Protein X requires the acquisition of a 'quality' tandem mass spectrum of the specific peptide containing the modification (in this instance the N-terminal peptide). The analysis is further complicated because covalent modifications are often present at low stoichiometry (the modified peptide is in low abundance relative to the unmodified peptide or to the peptides from the rest of the protein). As demonstrated in Figure 2B, if only 10% of Protein X is modified, then only one peptide out of 40 total peptides from the digestion of Protein X can potentially identify the type and the location of the modification.

Figure 2. Covalent modifications can be difficult to detect if present at low stoichiometry.



(**A**) Protein X is digested with trypsin to produce four fragments, one of which contains a covalent modification. (**B**) If 10% of Protein X is modified, then only one out 40 peptides will identify the type and location of the modification.

To date, efforts directed at the global analysis of protein modifications have been focused on phosphorylation. Reversible phosphorylation of proteins is central to the regulation of most aspects of cell function. At least 30% of all proteins encoded by the human genome are thought to be phosphorylated, and abnormal phosphorylation is often the cause or consequence of many human diseases [11]. Changes in protein phosphorylation are commonly studied by radiolabeling proteins with ^{32}P inorganic phosphate followed by proteomic analysis to compare relative spot intensities detected by 2-DE. However, this method requires that the protein still be identified by mass spectrometry and does not allow for the position of the phosphorylation site to be determined.

Recently, four phosphoproteome methods were reported using mass spectrometry [12,13,14,15••]. All four methods involve chemical modifications to enrich for phosphopeptides prior to analysis (Figure 3A). Smith and co-workers developed a method that uses phosphoprotein isotope-coded affinity tags (PhIAT) created by using a three-step affinity tag labeling approach similar to the approach reported earlier by Oda and co-workers (Figure 3A-1) [14,13]. The method employed by Oda *et al* only identified a single phosphorylation site in yeast [13], while Smith and co-workers failed to report any [14]. Zhou and colleagues used a 16-h six-step chemical derivatization strategy that resulted in the identification of only 14 unambiguous phosphorylated peptides (Figure 3A-3) [12]. Ficarro and colleagues reported a very impressive method that converted peptides to methyl esters and then enriched for phosphopeptides by Fe^{3+}-immobilized metal-affinity chromatography (IMAC) (Figure 3A-2) [15••]. The methyl esters reduced the non-specific binding of peptides to the IMAC column and allowed > 1000 different phosphoserine-, phosphothreonine- and even phosphotyrosine-containing peptides to be detected in yeast whole-cell lysate. Of the phosphopeptides identified by Ficarro *et al*, 216 were manually confirmed by comparisons to the MS/MS spectra of the respective synthetic phosphopeptides [15••]. Each of these methods is limited to the identification of phosphorylation and involves some form of chemical manipulation and enrichment of the protein sample (Figure 3). As with any *in vitro* chemical derivatization, each non-quantitative step decreases the total yield of the final product while increasing the potential for sample artifact and reaction side products. Some of these artifacts (eg, oxidized methionine, and glycation and carbamylation of amino groups) are similar or identical to ones observed in biological systems [16]. Therefore, an ideal method should involve minimal sample manipulation.

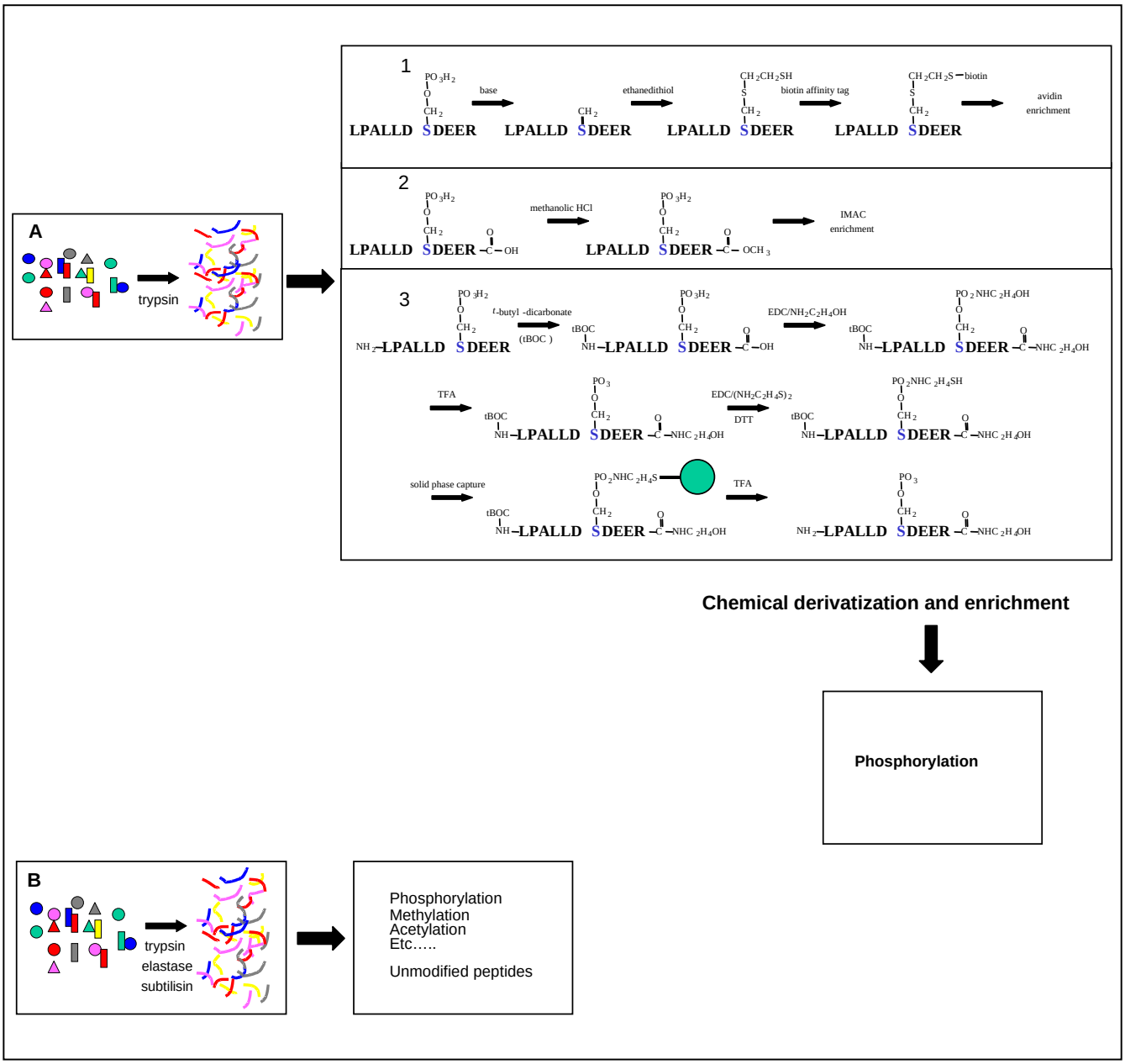
In a more recent study by MacCoss and colleagues, a totally different approach was taken to identify covalent modifications from complex protein mixtures (Figure 3B) [17••]. Unlike the methods described above, this approach does not use chemical derivatization or sample enrichment, and is not restricted to the detection of protein phosphorylation. Instead, this method is compatible with many different types of modifications, as demonstrated by the detection of protein phosphorylation, methylation and acetylation. The protocol uses non-specific proteases (subtilisin and elastase) in addition to a specific protease (trypsin) to generate multiple overlapping peptides covering each covalent modification. The complex mixture of peptides is then analyzed using MudPIT. As there is no enrichment, both modified and unmodified peptides are measured in a single analysis. Overlapping peptide coverage is generated throughout the proteins, so that data analysis is no longer dependent on a single spectrum from one peptide, but draws on multiple spectra from many different peptides. These overlapping peptides increase the probability that a particular modification locus is identified, reduces the ambiguity of single peptide identifications, and

provides multiple confirmations that the detected modification is real. This approach was validated against a protein complex (cdc2) from *Schizosaccharomyces pombe* and on the soluble fraction of a human cataract lens. Within a single family of lens proteins (crystallins), 74 protein modifications were reported and each redundantly identified with multiple overlapping peptides. More recently, this approach was optimized for the identification of modifications on membrane proteins [8].

Methods for global quantitation: Is there a true internal control?

A major goal in the study of biological regulation and function is to monitor and identify changes in global protein expression, and tools are being developed to enable quantitative comparisons of complex protein samples. Quantitative methods using mass spectrometry are based on measurements of the abundance of a protein relative to its corresponding internal standard. The major difficulty in proteomics is the production of an internal standard for every protein within the cell. Approaches for solving this problem fall into one of three categories: (i) chemical derivatization of one sample with an isotopically light reagent mixed with a second sample derivatized with an isotopically heavy reagent; (ii) one sample with no derivatization mixed with a sample containing a chemical derivative; and (iii) one sample with a natural abundant stable isotope composition mixed with a second sample metabolically labeled with a heavy stable isotope. Each method assumes that differences in sample preparation prior

Figure 3. Methods to detect phosphorylation and other covalent modifications.



(**A**) Summary of chemical derivatization strategies of functional groups to enrich for phosphopeptides. (**B**) Shotgun identification of protein modifications using multiple proteases to generate overlapping peptide coverage.

to mixing are minimal and occur equally between samples. In all approaches, the peptides from these mixtures are identified and the relative abundance between the two samples is measured using mass spectrometry.

Most studies have focused on differentially modifying proteins/peptides by *in vitro* chemical derivatization to create a complementary internal standard. Two samples are modified with either a tag containing only naturally abundant stable isotopes, or a tag containing selected atoms enriched in a heavy stable isotope. The samples are then combined and analyzed by mass spectrometry. Gygi and colleagues introduced a method in 1999 using isotope-coded affinity tags (ICATs) [18••], in which isotopic variants of a biotin-containing moiety to differentially label cysteine residues in peptides as a means for comparing two biological samples in a single analysis (Figure 4A-1). The relative abundance of the two samples is determined by monitoring the ratios of pairs of peptides offset by 8 Da (the number of deuteriums on the 'heavy' isotope tag). The ICAT method is unique in that the labeled cysteine-containing peptides are enriched using affinity chromatography (avidin-conjugated solid support), thereby decreasing the complexity of the mixture. Recently, this method was applied to the quantitation of 491 proteins contained in the microsomal fractions of naïve and *in vitro* differentiated human myeloid leukemia (HL-60) cells, making this study the largest quantitative proteomic analysis to date [19•]. Although impressive data have been reported using this methodology, drawbacks of ICAT include its expense and the difficulties associated with optimizing the recovery of biotinylated peptides.

Limitations in the ICAT method led to the subsequent development of cost-effective alternatives. Zhang and colleagues reported the use of the $^{2}H_{3}$ form of *N*-acetoxysuccinimide or the $^{2}H_{4}$ and ^{13}C form of succinic anhydride as alternative isotope tags (Figure 4A-3) [20,21]. Using this approach, differential labeling is directed at amine groups on peptides, thereby theoretically modifying every peptide in the mixture. Cagney and Emili developed a method similar to ICAT based on a different chemical modification involving differential guanidination of C-terminal lysine residues on tryptic peptides termed mass-coded abundance tagging (MCAT) (Figure 4A-4) [22]. This method is more global in that theoretically all C-terminal lysines produced by tryptic cleavages are tagged. However, because MCAT-derivatized peptides are directly compared to underivatized peptides, anything less than quantitative guanidination will significantly skew the results. Recently, the ICAT method was improved with a strategy devised by Zhou and colleagues [23], in which cysteine-containing peptides are labeled using a reagent that is immobilized onto a solid support (Figure 4A-2), thereby allowing tagging and enrichment to occur in one step. When compared directly to the ICAT method, this approach is reported to be faster, resulting in higher yield with lower background [23].

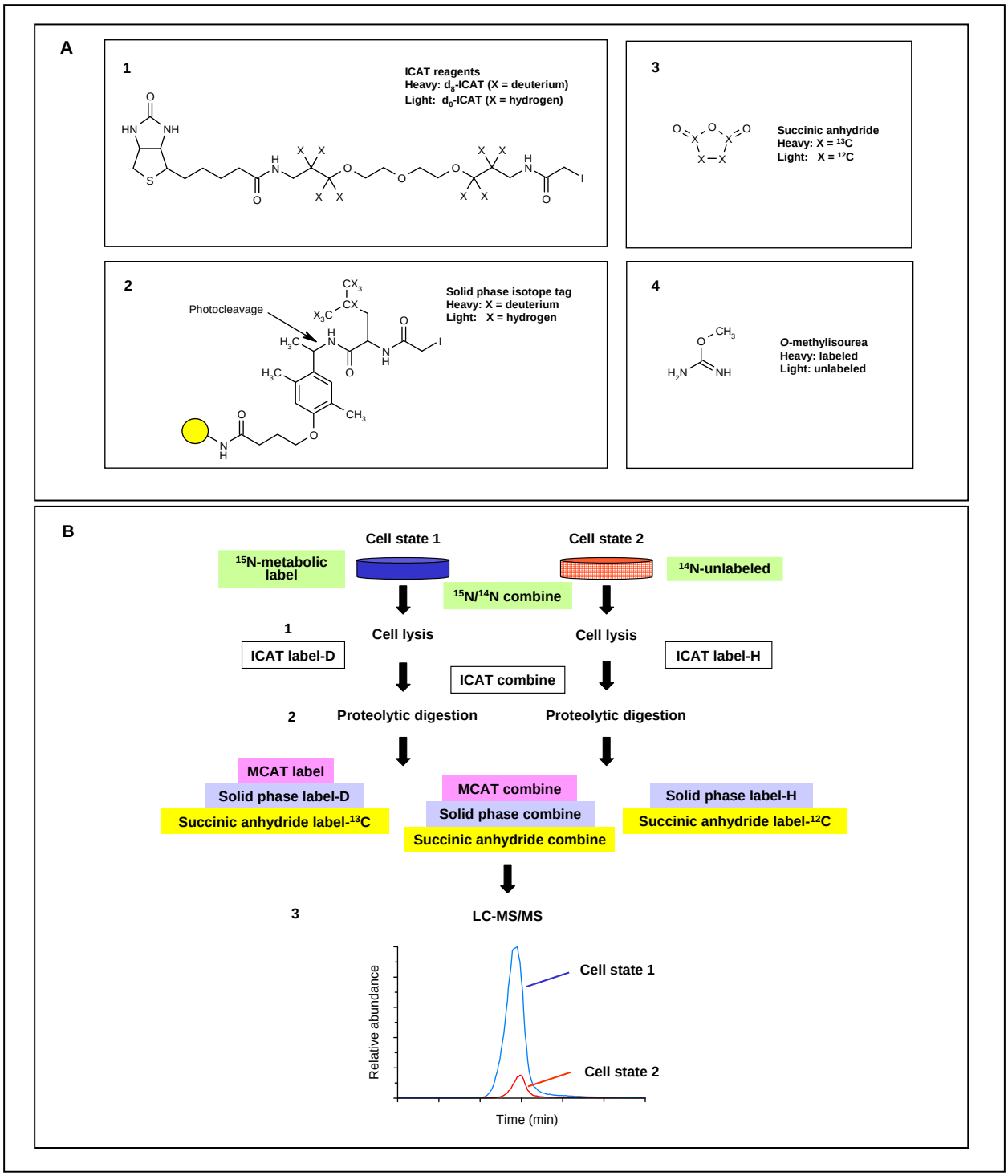
The various post-biosynthetic coding of peptides for quantitation detailed above is achieved by chemical derivatization *in vitro*. This approach is fundamentally flawed in that the internal control is synthesized after it has been extraction from the native cellular environment. Any chemical reaction will suffer from kinetic differences between the individual reactions within an extremely complex mixture. An internal standard should be incorporated at the earliest possible step to minimize differences occurring during sample preparation and subsequent analysis by mass spectrometry. Figure 4B outlines the major steps involved in quantitatively comparing the proteomes of two cell samples (cell state 1 and cell state 2). The cells are lysed, proteolytically digested and analyzed by LC-MS/MS. At some point during these steps, internal standards are incorporated to allow for the analysis of differences between the two samples. Sample mixing occurs after cell lysis using ICAT and after proteolytic digestion using MCAT, solid-phase isotope label and isotopic succinic anhydride. Any differences prior to mixing would not be accounted for using these methods.

Metabolic labeling is a well-established labeling technique and incorporates stable isotopes uniformly into metabolic products *in vivo*. The resulting labeled proteins provide an ideal internal standard. This strategy makes use of *in vivo* ^{15}N-labeling to measure the relative abundance of proteins isolated from cultured cells. The mass of a protein becomes offset by 1 Da for each nitrogen atom in the protein. This method has been severely under-appreciated due to the common misconception that labeling is only limited to organisms that can be grown in a defined media, and that isotope incorporation can vary between compartments in a cell or organism, making the mass difference between non-labeled and ^{15}N-labeled isoforms unpredictable and spectra difficult to interpret. Conrads and colleagues have successfully labeled bacteria, yeast and mammalian cells with ^{15}N-enriched culture media [24], as have Washburn and co-workers with yeast [25]. Given sufficient time, all proteins in any organism will reach isotopic equilibrium with the available amino acids in the cell and, therefore, can theoretically be labeled.

Methods for reducing the computational bottleneck

One of the major limitations of the application of shotgun proteomics to complex biological systems is trying to sort through the enormous amounts of data generated. Improvements in the resolving power of microcapillary separations and the sensitivity of mass spectrometers are currently such that a single MudPIT analysis can produce > 70,000 MS/MS spectra. These spectra are interpreted by comparing the experimental MS/MS spectrum against a theoretical spectrum produced from sequence information stored in databases. Furthermore, the number of spectra to be searched increases when electrospray ionization is used with a low-resolution mass spectrometer, because establishing the correct charge state of a multiply charged ion is difficult. Thus, a multiply charged peptide MS/MS spectrum has to be searched twice assuming both a +2 and +3 charge state. Because of the rapid expansion of protein databases, searching this many spectra can pose an enormous computational load. Finally, 'assembling' this peptide information back into protein identifications and trying to make sense of the real protein identifications versus the false positives can often be the rate-limiting step in proteomics.

Figure 4. Approaches for quantitative comparative proteomics.



(**A**) Summary of chemical derivatizing agents. (**B**) Sample preparation for the quantitative comparison of two cell samples (cell state 1 and cell state 2). The cells are (**1**) lysed, (**2**) proteolytically digested, and (**3**) analyzed by LC-MS/MS. Chemical derivatization for each method discussed creates paired internal standards that are incorporated into the sample during different steps of the sample preparation.

An obvious priority in improving shotgun proteomics approaches has to be to reduce the total number of spectra searched. Sadygov and co-workers reported a computer program called '2to3' that removes spectra of poor quality and determines the charge state of multiply charged peptides (Figure 5A) [26]. 2to3 was shown to substantially reduce the number of spectra searched without losing a single protein identification. In the same report, Sadygov *et al* also applied a fault-tolerant parallelization of the popular database search algorithm

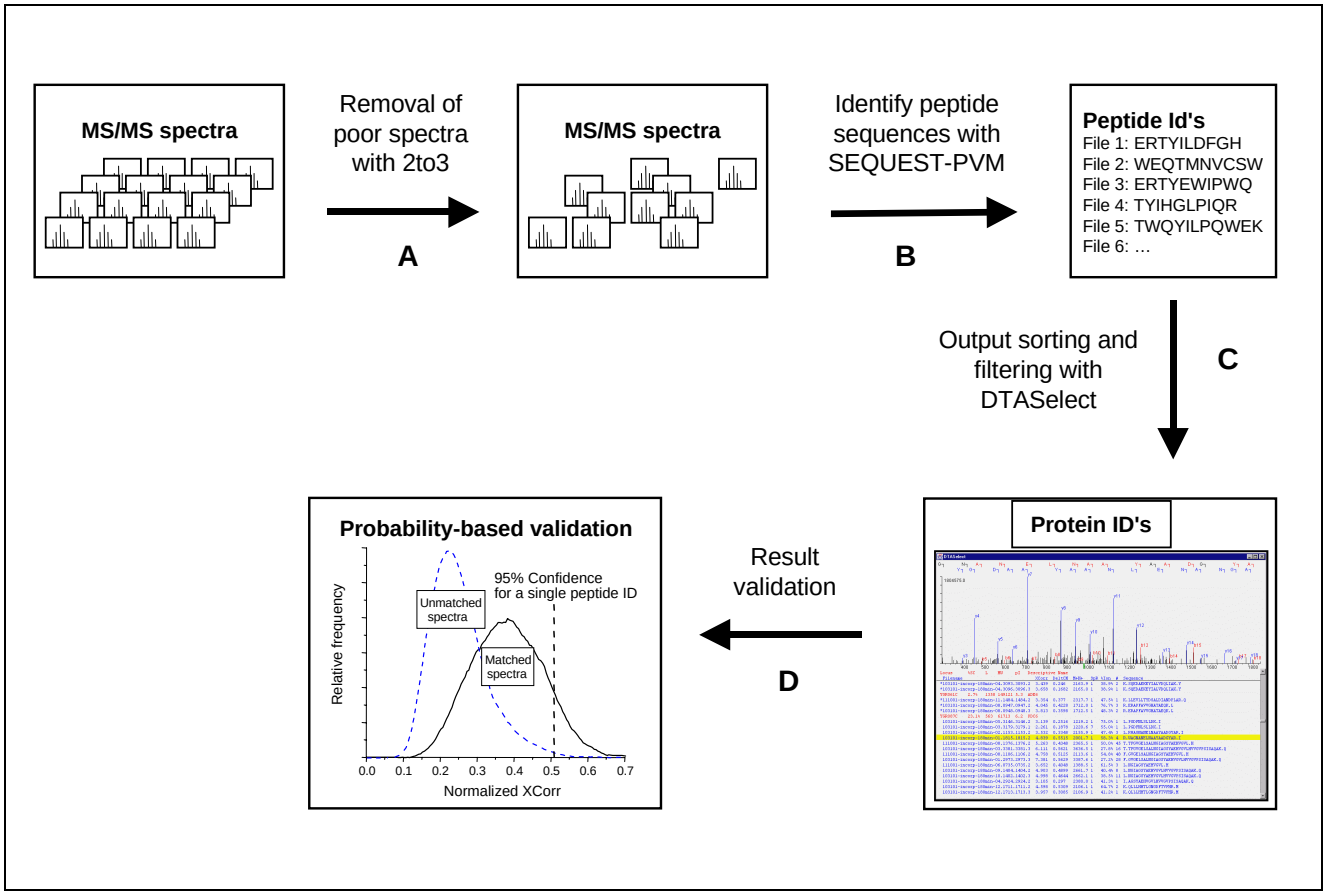
SEQUEST. This version, SEQUEST Parallel Virtual Machine (SEQUEST-PVM), distributes the MS/MS spectra to be searched across several computers in a cluster (Figure 5B). SEQUEST-PVM is extremely efficient and results in a significant drop in analysis time that is directly related to the number of computers in the cluster. In addition, this system is cost efficient, requiring little more than personal computers and a freely available operating system (LINUX).

For the filtering, sorting, assembly and comparison of SEQUEST data, Tabb and colleagues developed DTASelect and Contrast (Figure 5C) [27••]. DTASelect can filter SEQUEST results at both the peptide and protein level. A number of tools are integrated into DTASelect for the display and presentation of the results. Contrast allows the differences between DTASelect output files to be displayed for the comparison of different samples and/or the effect of different filtering criteria on the resultant dataset.

Despite the many software tools available, separating real identifications from false positives requires manual confirmation, a labor-intensive and inherently subjective process. SEQUEST further complicates this analysis because the resulting cross-correlation score (XCorr) is larger for larger peptides (ie, different peptides will have different scores even if their matches are of equal quality). MacCoss and colleagues address these issues by applying a normalization routine to all XCorr values generated [28••]. Normalization results in the comparison of all SEQUEST results on the same scale where 1.0 is a perfect match and 0.0 signifies no match. In addition, the authors use the normalized scoring with probability-based validation to eliminate any subjectivity in the analysis of SEQUEST results (Figure 5D). The effect of multiple peptides identifying the same protein on the confidence of the identification is also demonstrated.

Recent software algorithms greatly improve the interpretation of proteomic results. In the last year, software has been described for: (i) the removal of poor/redundant MS/MS spectra; (ii) parallelization of database searching software; (iii) filtering, sorting, assembly and comparison of separate analyses; and (iv) automated validation of protein identifications. These software developments still require the integration of proteomics quantitation software and the direct accessibility of all proteomics results through a public database.

Figure 5. Improvements in obtaining protein identifications from tandem mass spectra.



Software is available for (**A**) the removal of poor quality MS/MS spectra; (**B**) the parallelization of database searching software; (**C**) the filtering, sorting, assembly and comparison of different analyses; and (**D**) the probability-based validation of protein identifications.

Conclusion

Recent advances in methodologies enabling the direct analysis of complex protein mixtures have paved the way for a global 'shotgun proteomics' approach to the analysis of complex biological systems. Multiple laboratories continue to push the envelope of this technology, and limitations are being rapidly overcome. Major strides in technology development have made shotgun proteomics approaches possible for global protein identification, the detection of covalent modifications and comparative quantitative analyses. The ability to systematically profile dynamic proteomes is quickly becoming a reality and will represent one of the most significant advances in biology to date. However, successes in method development have led to an explosion in data production, and limitations in data analysis are quickly becoming the major bottlenecks in proteomics studies. Furthermore, a current dilemma lies in whether to create a large-scale center dedicated to high-throughput approaches to human proteomics (ie, information gathering), or instead to rely on the intense focus of individual investigators to thoroughly elucidate small pieces of the overall picture [29]. The high-throughput approach would be expensive and provide minimal insight into biological function, and yet it would create a uniform database from which specific biological questions could be asked. The scaled-down, focused approach would give rise to hundreds of uncoordinated databases, which may prove difficult to piece together without a coordinated large-scale effort. Regardless of whether proteomics is performed primarily in large institutes or individual laboratories, developments in methodology are likely to be continued at all levels for the foreseeable future. The implications of these technological and methodological advances, and of proteomics in general, on the future biomedical research will be enormous.

Acknowledgements

We gratefully acknowledge financial support from the National Institute of Health Grant F32DK59731 (MJM).

References

•• of outstanding interest
• of special interest

1. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.** *Proc Natl Acad Sci USA* (2000) **97**:9390-9395.

• *This paper thoroughly evaluates the limitations of 2-DE for global proteomic analyses.*

2. Santoni V, Molloy M, Rabilloud T: **Membrane proteins and proteomics: Un amour impossible?** *Electrophoresis* (2000) **21**:1054-1070.

3. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez J-C: **The dynamic range of protein expression: a challenge for proteomic research.** *Electrophoresis* (2000) **21**:1104-1115.

4. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* (1999) **17**:676-682.

• *This paper describes the proteomic analysis of the yeast ribosome using MudPIT.*

5. Washburn MP, Wolters DA, Yates JR III: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nat Biotechnol* (2001) **19**:242-247.

•• *In this study the authors used MudPIT to produce the largest proteome analysis to date. The methodology is largely unbiased and can even be applied to the analysis of transmembrane proteins.*

6. Wolters DA, Washburn MP, Yates JR III: **An automated multidimensional protein identification technology for shotgun proteomics.** *Anal Chem* (2001) **73**:5683-5690.

7. Stevens TJ, Arkin IT: **Do more complex organisms have a greater proportion of membrane proteins in their genomes?** *Proteins* (2000) **39**:417-420.

8. Wu CC, MacCoss MJ, Howell KE, Yates JR III: **Proteomic analysis of Golgi membrane proteins: identification, topology, and covalent modifications.** *Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics* (2002).

9. Krishna RG, Wold F: **Post-translational modification of proteins.** *Adv Enzymol Relat Areas Mol Biol* (1993) **67**:265-298.

10. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* (1999) **19**:1720-1730.

11. Cohen P: **The origins of protein phosphorylation.** *Nat Cell Biol* (2002) **4**:E127-E130.

12. Zhou H, Watts JD, Aebersold R: **A systematic approach to the analysis of protein phosphorylation.** *Nat Biotechnol* (2001) **19**:375-378.

13. Oda Y, Nagasu T, Chait BT: **Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome.** *Nat Biotechnol* (2001) **19**:379-382.

14. Goshe MB, Conrads TP, Panisko EA, Angell NH, Veenstra TD, Smith RD: **Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses.** *Anal Chem* (2001) **73**:2578-2586.

15. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM: **Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*.** *Nat Biotechnol* (2002) **20**:301-305.

•• *The authors report the optimization of immobilized metal-affinity chromatography for the analysis of protein phosphorylation in yeast. A total of 216 phosphorylation sites were identified in a single analysis, making it the largest phosphoproteome analysis to date.*

16. Volkin DB, Mach H, Middaugh CR: **Degradative covalent reactions important to protein stability.** *Mol Biotech* (1997) **8**:105-122.

17. MacCoss MJ, McDonald WH, Saraf A, Sadygov R, Clark JM, Tasto JJ, Gould KL, Wolters D, Washburn M, Weiss A, Clark JI, Yates JR III: **Shotgun identification of protein modifications from protein complexes and lens tissue.** *Proc Natl Acad Sci USA* (2002) 99:7900-7905.

•• *This method combines multi-enzyme digestion with the large dynamic range and sensitivity of MudPIT for the analysis of protein modification in mixtures. As there is no sample enrichment, both unmodified and modified peptides are measured in a single analysis.*

18. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* (1999) **17**:994-999.

•• *The original quantitative proteomics method using a stable isotope-labeled reagent.*

19. Han DK, Eng J, Zhou H, Aebersold R: **Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry.** *Nat Biotechnol* (2001) **19**:946-951.

• *The authors apply ICAT to the relative quantitation of 491 mammalian membrane proteins.*

20. Zhang R, Sioma CS, Wang S, Regnier FE: **Fractionation of isotopically labeled peptides in quantitative proteomics.** *Anal Chem* (2001) **73**:5142-5149.

21. Zhang R, Regnier FE: **Minimizing resolution of isotopically coded peptides in comparative proteomics.** *J Proteome Res* (2002) **1**:139-147.

22. Cagney G, Emili A: ***De novo* peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging.** *Nat Biotechnol* (2002) **20**:163-170.

23. Zhou H, Ranish JA, Watts JD, Aebersold R: **Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry.** *Nat Biotechnol* (2002) **20**:512-515.

24. Conrads TP, Alving K, Veenstra TD, Belov ME, Anderson GA, Anderson DJ, Lipton MS, Paša-Tolic L, Udseth HR, Chrisler WB, Thrall BD, Smith RD: **Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and ^{15}N-metabolic labeling.** *Anal Chem* (2001) **73**:2132-2139.

25. Washburn MP, Ulaszek R, Deciu C, Schieltz DM, Yates JR III: **Analysis of quantitative proteomic data generated via multidimensional protein identification technology.** *Anal Chem* (2002) **74**:1650-1657.

26. Sadygov RG, Eng J, Durr E, Saraf A, McDonald H, MacCoss MJ, Yates JR III: **Code developments to improve the efficiency of automated MS/MS spectra interpretation.** *J Proteome Res* (2002) in press.

27. Tabb DL, McDonald WH, Yates JR III: **DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics.** *J Proteome Res* (2002) **1**:21-26.

•• *This paper describes software developed for the sorting, filtering and assembly of SEQUEST results. Filters can be applied at both the peptide and protein level.*

28. MacCoss MJ, Wu CC, Yates JR III: **Probability based validation of protein identifications using a modified SEQUEST algorithm.** *Anal Chem* (2002) manuscript submitted.

•• *A modification to the SEQUEST cross-correlation algorithm that places the scoring of all peptide tandem mass spectra on the same scale. This normalized score permits the validation of protein identifications using a probability-based approach.*

29. Check E: **NIH ponders issues of scale in protein push.** *Nature* (2002) **417**:107.