Shotgun proteomics and biomarker discovery

W. Hayes McDonald* and John R. Yates III
Department of Cell Biology, SR11, The Scripps Research Institute, La Jolla, CA 92037, USA

Abstract. Coupling large-scale sequencing projects with the amino acid sequence information that can be gleaned from tandem mass spectrometry (MS/MS) has made it much easier to analyze complex mixtures of proteins. The limits of this "shotgun" approach, in which the protein mixture is proteolytically digested before separation, can be further expanded by separating the resulting mixture of peptides prior to MS/MS analysis. Both single dimensional high pressure liquid chromatography (LC) and multidimensional LC (LC/LC) can be directly interfaced with the mass spectrometer to allow for automated collection of tremendous quantities of data. While there is no single technique that addresses all proteomic challenges, the shotgun approaches, especially LC/LC-MS/MS-based techniques such as MudPIT (multidimensional protein identification technology), show advantages over gel-based techniques in speed, sensitivity, scope of analysis, and dynamic range. Advances in the ability to quantitate differences between samples and to detect for an array of post-translational modifications allow for the discovery of classes of protein biomarkers that were previously unassailable.

Keywords: Proteomics, proteome, biomarker, MudPIT, mass spectrometry, multidimensional chromatography

1. Introduction

Returning to the central dogma of gene function found in all cell biology texts, most genes are transcribed into RNA that in turn is translated into the proteins which perform the actual function of those genes within the cell. In eukaryotic cells there are myriad regulatory mechanisms at each step in this pathway that in sum control the activity of the protein. Assuming that it is this activity, or lack thereof, which is associated with a disease state, it follows that directly measuring the proteins would provide a rich source for important biomarkers. In other words, even if there were mutations at the DNA level or changes in the amount of mRNA expression, these would only be important if they affected the protein and its activity. In addition, there are hundreds of examples in which a protein's function is controlled by some post-translational mechanism such as covalent addition of a phosphate group or the targeted degredation of the protein.

Even though techniques to study specific aspects of protein structure and function have been developed and refined over the last several decades, most of these have been limited to the study of individual proteins or at most small groups of proteins. Genome sequencing has opened up entirely new realms of possibility. Most obviously, it provides the basic infrastructure to detect genetic differences (mutations) which can help to understand observed phenotypic differences (Genomics and SNPs). This sequence infrastructure also facilitates the measurement of the global transcription profile within a cell using DNA/RNA microarray technologies. This same genomic sequence infrastructure coupled with advances in analytical techniques has enabled the emerging field of proteomics.

While precise definitions may vary, in most general terms, proteomics encompasses a set of techniques that allows one to more rapidly or more comprehensively study proteins. Because protein identification is usually the rate-limiting step in any proteomic strategy, many of the advances in proteomics have focused on improving the speed and sensitivity with which proteins can be identified. Two mass spectrometry based techniques have proven most useful for this task, peptide mass mapping and peptide "sequencing" via tandem mass spectrometry. In order to understand the scope and possibilities of proteomics, it is important to first describe these techniques and how they are cou-

*Corresponding author: Tel.: +1 858 784 8876; Fax: +1 858 784 8883; E-mail: hayesmcd@scripps.edu.

pled with separative strategies allowing the most robust experiments to be performed for a given sample-type.

2. Proteomics

Proteins possess an incredible diversity of chemical properties, with broad ranges of catalytic activity, molecular weight, and solubility. To simplify the analytical challenges posed by these differences, both peptide mass mapping and peptide tandem mass spectrometry are performed on sub-sections (peptides) of the protein. These peptides are more readily analyzed because of their more uniform size and chemistry. Normally they are generated from the protein being analyzed through the use of proteolytic enzymes (proteases).

For mass mapping, a protease of known specificity, such as trypsin, is used to digest the protein. Not only does this produce peptides of a size more readily analyzed in a mass spectrometer (MS), but based on the amino acid specificity of the enzyme, will produce a mass fingerprint specific enough to allow identification of the protein. For an unknown protein, the masses of the peptides are searched against a "virtual" digest of a protein database to find a protein that would yield a similar peptide pattern if it were digested with the same specific protease. This technique can be very rapid, easily automated, is not excessively computationally intensive, even for large databases. For these reasons, mass mapping has become a cornerstone technology in many proteomic strategies.

Peptide sequencing by tandem mass spectrometry (MS/MS), also starts with proteolytically derived peptides. However, as we will discuss more extensively later, a protease of known specificity need not be used. This technique, takes advantage of the tandem mass spectrometer's ability to select a specific peptide ion and further analyze it. Addition of energy to the peptide causes it to break and produce series of ions that are fragments of the "parent" peptide. Because fragmentation occurs most often between the amide bonds along the peptide backbone, the differences in masses between these "daughter" ions allows the inference of amino acid sequence.

Besides protein identification, the immense complexity of the proteome poses a challenge of separation. One strategy is to first separate proteins and then use a distinct analytical step to identify the protein in question. This classic combination is probably best illustrated using two-dimensional gel electrophoresis to
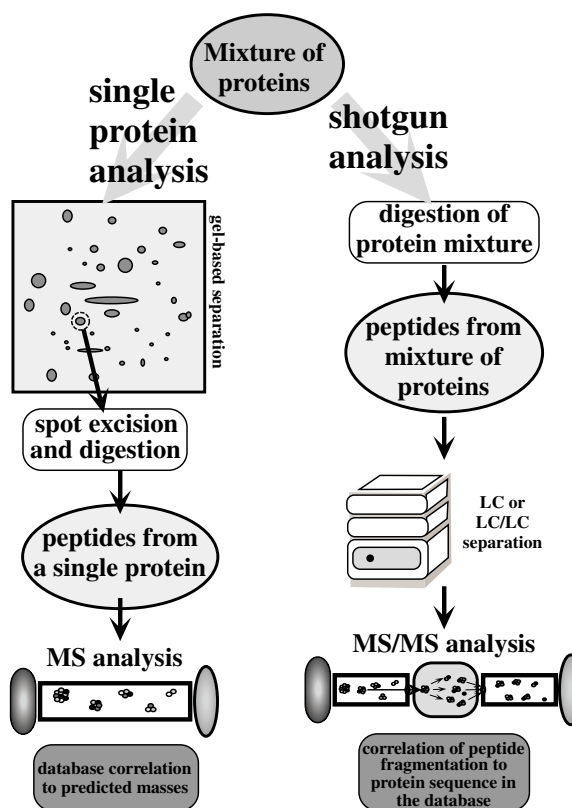

Fig. 1. "Single protein" versus "shotgun" analysis of proteins. There are two commonly applied strategies for analyzing a mixture of proteins. For years, the predominant techniques involved first separating the mixture to single proteins and then identifying them one protein at a time. Illustrated here is the common combination of resolving protein by two dimensional gel electrophoresis (2D-GE) and protein identification by peptide mass mapping. Another technique involves digesting the entire mixture of proteins and then using the ability of the tandem mass spectrometer (MS/MS) to derive amino acid sequence data from the peptides. Very complex mixtures of proteins can be assayed in this manner if the peptides are separated by single dimensional (LC) or multidimensional (LC/LC) high performance liquid chromatography.

separate the proteins and then mass mapping to identify them (Fig. 1). The emergence of tandem mass spectrometry has made viable an alternative strategy in which proteins are not separated prior to digestion. By analogy to the "shotgun" DNA sequencing techniques, shotgun proteomics allows for identification of the protein components from a mixture using the tandem mass spectrometry based identifications of individual peptides. In the shotgun proteomics experiment, the problem of complexity is addressed by separating a peptide mixture prior to the data collection using MS/MS (Fig. 1). High pressure liquid chromatography tandem mass spectrometry (LC-MS/MS) provides a very effective methodology for the basic shotgun proteomics

experiment. For more complete analysis or analysis of very complex protein mixtures, two-dimensional chromatography (LC/LC-MS/MS) strategies such as MudPIT [14,26,27] can be employed.

Despite sometimes strident discussions over which general method is "the answer" for proteomics, the current state of the art can be viewed as a continuum. At one end, proteins are completely separated prior to identification and at the other, no protein separation is performed prior to digestion into peptides (Fig. 2). At the "single protein" end of the spectrum falls the two-dimensional gel electrophoresis experiment, resolving proteins for identification by mass mapping. At the far right "shotgun proteomic" end, would be found LC/LC-MS/MS experiments such as the MudPIT analysis of whole cell lysates [26,27] which would prove impossible without the additional amino acid sequence data provided by tandem spectra.

Even with the intense interest in being able to analyze an entire "proteome", some of the most important experiments fall closer to the center of the spectrum. An excellent example of this is in the analysis of purified multiprotein complexes. These can be analyzed a variety of different ways including: first resolving the proteins by single dimensional SDS polyacrylamide gel electrophoresis (SDS-PAGE) and then analyzing them by mass mapping, resolving the proteins by SDS-PAGE and then analyzing them by LC-MS/MS, or by directly analyzing the protein complexes by LC-MS/MS or LC/LC-MS/MS (see Fig. 2 for example references for each). Thus, it is important to note that the MS/MS techniques that are essential for shotgun proteomics can be applied almost across the entire spectrum. In general these would be used when dictated by the complexity of the protein mixture, if the range of protein amounts within the mixture exceeds the resolving/visualization capacity of gel-based techniques, or if more specific structural information such as defining sites of post-translational modifications are needed. For the remainder of this article, we will focus on aspects of shotgun proteomic-based techniques and their potential applications both for answering basic science questions and for biomarker discovery.

3. Biomarker discovery

The presence of a particular protein in a given disease state not only provides a potential biomarker for that disease, but also could provide some insight into the basic etiology of that disease. Much of the current biomarker related proteomic work has used standard 2D-GE based techniques with some more recent work utilizing protein chip technologies [4,24]. However, a publication from the proteomics group at Bristol-Myers Squibb shows the potential for using shotgun proteomics based techniques for biomarker discovery [17]. Their use of several different techniques on the same types of samples allows a comparison of the results yielded by these strategies.

They analyzed healthy and diseased urine samples from an individual who had been diagnosed with an inflamed pilonidal abcess. Differences in the urine proteomes between the samples were probed for potential biomarkers using 2D-GE and two shotgun-proteomic based techniques, LC-MS/MS and LC/LC-MS/MS. 2D-GE allowed the identification of 5 differentially expressed proteins. With the LC-MS/MS based approach, they identified 28 proteins in healthy samples and 23 in the diseased sample, 16 of which were common between the two experiments. Even amongst these commonly expressed proteins they were able to infer a rough estimate of relative abundance using a normalization of the average number of peptides identified for a protein across duplicate experiments. The big advantages of this experiment over their 2D-GE one was the speed of the experiment, 36 hours versus approximately 5 days, and that about one tenth the amount of protein was needed for the analysis. The added separative capacity of the LC/LC-MS/MS based approach yielded both more protein identifications, 51 and 67 respectively, and more proteins that were unique to the two samples, 19 and 39. Again even for the 28 proteins in common between the two experiments, differences in protein amounts could be inferred by numbers of peptides identified. While requiring more work and time than the LC-MS/MS based approach, it was still significantly less than was required for 2D-GE.

The ability to perform these experiments on very limited sample amounts was probably the biggest advantage of the shotgun-based techniques. This difference could be even more important when screening for biomarkers from patient samples that are far more limiting than urine or blood. One of the major challenges of using shotgun proteomics to compare two samples is that one does not get precise quantitation of the relative protein amounts. While this group used techniques which were effective for judging rough differences, several shotgun based techniques are available which allow much more precise comparisons between two samples.
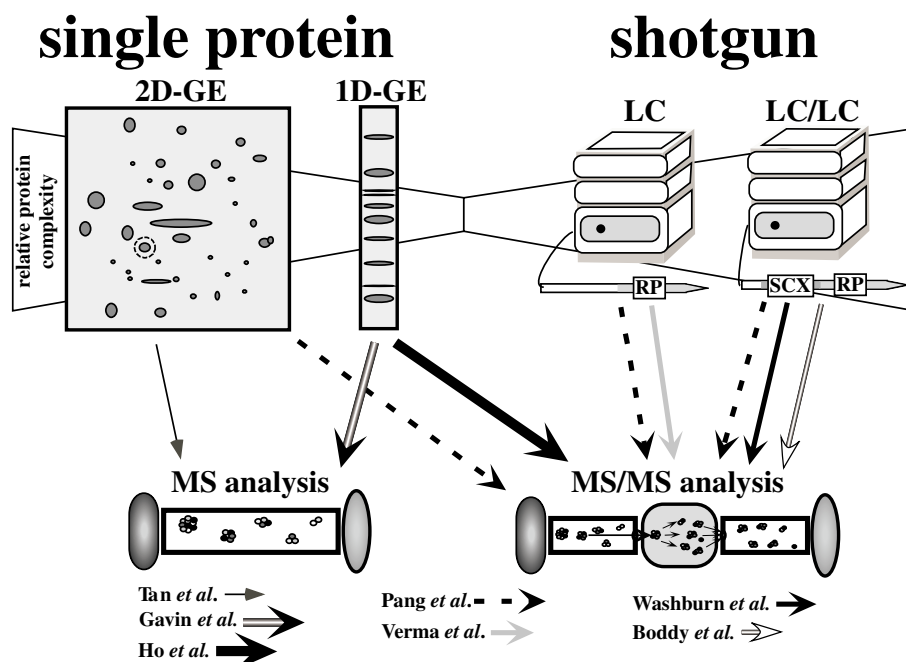
Fig. 2. The "single protein" to "shotgun" continuum. Many common experiments in proteomics do not use purely single protein or shotgun proteomic strategies. The most complex mixtures (far left and far right of the figure) require the maximum separation. From these types of mixtures one must use two dimensional gel electrophoresis (2D-GE) in order to have a chance of resolving the proteins to single-protein spots for subsequent identification. To analyze similarly complex mixtures using a shotgun approach, one must utilize a multidimensional separation strategy (LC/LC) in order to separate and to collect tandem mass spectral (MS/MS) data from the resulting peptides. Less complex mixtures such as multiprotein complexes afford greater choices of techniques. They can be adequately separated by single dimensional gel electrophoresis for subsequent identification by MS or MS/MS analysis. Depending on the number of proteins present within the complex being analyzed they can also be digested and directly analyzed by LC or LC/LC coupled MS/MS analysis. The included references are meant to be illustrative of how different combinations of these technologies can be used. Tan et al., 2002 used the classic combination of 2D-GE and peptide mass mapping [22]. For protein complex analysis Gavin et al., 2002 first resolved proteins by 1D-GE and then predominantly identified them by mass mapping [6]. Ho et al., 2002 while still resolving their purified protein complexes by 1D-GE chose to identify the proteins contained in the bands by LC-MS/MS [12]. Shotgun analysis of protein complexes can be achieved either through LC-MS/MS [23] or LC/LC-MS/MS [1]. Pang et al., 2002 showed the widest variety of techniques to their analysis of urine proteins using 2D-GE coupled with MS/MS identifications, LC-MS/MS analysis, and LC/LC-MS/MS analysis [17].

4. Quantitation

The challenge to quantitation that proteins/peptides pose to mass spectrometry is that different peptides will ionize with different efficiencies depending on the precise chemical properties of the polypeptide. Because the effects of these differences are currently impossible to predict, one cannot infer how much of a given protein is present just based on the measurement of the intensity of an ion from that protein. The general analytical way around the differential ionization problem has been to include a known quantity of stable isotope labeled control which is otherwise identical to the experimental compound to be measured. Because ionization efficiencies will be the same, it is possible to compare very precisely the amount of heavy (control) to light (experimental) of a given compound (Fig. 3).

Because one cannot include a control peptide for every possible protein which could be present within a proteomic sample, slightly different strategies must be employed. These involve labeling the proteins in a way in which the peptides from one condition will be stable isotope labeled and in the other conditions they will not. Thus the relative amount of protein can be compared by measuring the ratio of heavy to light versions of a particular peptide. While there are a variety of methods which are being developed for this, they all fall into one of two categories: isotopically labeling all of the proteins within a cell by growing the cells up in media (usually 15N) that will incorporate that isotope into all cellular proteins or chemically derevitizing proteins from the samples with isotopic "tags".

One method, the isotope coded affinity tag (ICAT) [9–11], takes advantage of the standard reduction and alky-
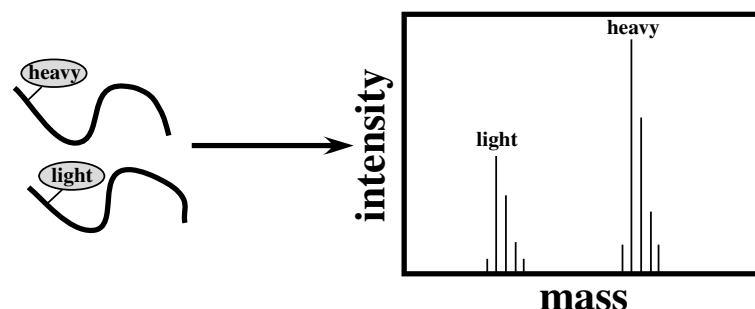
Fig. 3. Quantitation by mass spectrometry. Relative quantitation of peptides and proteins can be achieved using the mass spectrometer's ability to resolve isotopically distinct versions of otherwise identical compounds. The relative amount of each compound can be measured by comparing the signal from "light" and "heavy" versions. For comparisons of relative amounts of proteins between samples, the proteins in each sample must be made isotopically distinct (see text). After digestion the relative amounts of each protein can be judged by comparing peptides from the differentially "labeled" proteins.

lation step that precedes most proteomic digestions by placing an isotopically differentiated tag on cysteine residues within the protein mixture. In addition to isotopically labeling the proteins within a sample, the presence of biotin within the tag allows those labeled peptides to be further purified using avidin affinity capture. Other reagents have been reported [18,25] and it remains an area of active development (reviewed in [19]).

Cagney et al have recently reported another tagging system in which lysine residues are modified with O-methylisourea to form homoarginine [2]. This modification changes the fragmentation patterns of peptides during MS/MS fragmentation and the authors propose that this will be of added benefit for de novo, i.e. without relying on database information, determination of the sequence of these peptides. For relative quantitation purposes, one compares the modified (homoarginine) to the same peptide in the unmodified (lysine) state. More rigorous validation will have to be done to make certain that the chemical differences do not have confounding effects on the chromatography and/or ionization efficiency of tagged peptides.

5. Post-translational modifications

Clearly not all regulation of protein activity takes place by simply controlling the overall quantity of that protein within the cell. Teleologically speaking, cells have taken advantage of the over 200 described post-translation modifications [13] to regulate the activities of their complement of proteins in ways that are more subtle and less energetically expensive than simply producing or destroying a given protein. These post-translational modifications vary from the extremely well understood and broadly utilized modifi-

cations such as protein phosphorylation to those whose physiolgical role are almost complete mysteries. Because of this incredible diversity of structure, function and utilization within the cell, post-translation modifications of proteins provide an incredibly rich source of potential disease biomarkers.

Even with agreement that globally surveying all protein post-translational modifications within a cell would be a good place to look for biomarkers, it is too great of an analytical problem for the current state of technology. However, strides are being made. The ability of 2D-GE to resolve different isoforms and modification states proteins has been well documented and allows one to visualize changes those proteins which are abundant enough to be observed using the technique [21]. However, the standard bottleneck of having to identify the protein of interest is even more challenging if one has to determine not only the reason for the change in protein mobility but also the type and site of the modification within the protein. While these determinations would not necessarily have to be made if 2D-GE was planned to remain as the diagnostic test, they would be essential for migrating to more sensitive, robust, and better established techniques such as ELISA for the clinical diagnostic setting.

Returning our focus back to shotgun-based techniques there are several that could provide the analytical specificity and dynamic range necessary to look for post-translational modifications within the proteome. The first are similar in principal to the ICAT reagent except that the chemistry is targeted to phosphate-modified serines and threonines [7,8,16,28]. While these techniques show some promise, the extensive front-end chemistry and the fact that they reported relatively few modifications from *in vivo* sources of protein, a single [16] and 12 [28] modified protein(s) re-

spectively, suggest that optimizations must be made before they can be applied to biological samples in which starting material is limited.

A more recent publication utilized an optimized version a technique that has been used for some time, imobilized metal affinity chromatography (IMAC) [3]. They increased the specificity of binding of phosphopeptides to the IMAC column by first making methyl esters out of all peptides within the mixture. Using this technique coupled with LC-MS/MS they were able to detect phosphopeptides down to sensitivity of 5 fmol. More importantly, they were able to apply this to a yeast whole-cell lyasate and identified 216 phosphopeptides from 171 different proteins. Many of these phosphopeptides were from proteins expressed at very low levels, demonstrating the general applicability of the protocol.

Our group has been attempting to explore how far the principle of shotgun proteomics using MudPIT can be taken for determining sites of post-translation modification. This more recent work expands on earlier work [5] in which a multienzyme digestion was used to increase the percentage of amino acid sequence for which MS/MS spectra could be collected. In that case it was used in combination with a modification to the SEQUEST search algorithm to search for single nucleotide polymorphisms (SNPs) in hemoglobin. Further optimizations in the digestion protocol, including an additional non-specific protease, and in the search algorithm has made this technique applicable for looking for protein modification from diverse biological sources [15].

As a proof of principal, modifications were mapped in a mixture of proteins that co-purified with the *Schizosaccharomyces pombe* cell cycle regulatory protein, cdc2. Both known and previously unreported modifications were found in cdc2 and in associated proteins [15]. Interestingly, not only were sites of phosphorylation mapped, some in cdc13 which had proven reticent to mapping by standard *in vivo* labeling based techniques, but also sites of methylation were found in the same experiment. Even more sites and many more different types of modifications were observed when the experiment was performed on human lens tissue.

Human lens was a particularly good test-bed for the technology for several reasons. One is that proteins do not turn over within the lens, making any techniques to survey mRNA or protein levels relatively uninformative. Another is that numerous modifications have already been described for lens proteins and many of these have been suggested to modulate activity. Fi-

nally, the total protein composition of the lens is relatively simple, with most of the protein by mass being comprised of a relatively small family of proteins. This technique returned numerous known and previously unreported sites of phosphorylation, oxidation, methylation, and acetylation. 53 proteins returned greater than 40% sequence coverage and some sites of modification were mapped for most of these.

However, the practical dynamic range of this technique remains to be determined. That is, for what range of protein amounts could one reasonably expect to characterize most of the sites of modification. Our group is currently undertaking a survey of lens tissue through development in both healthy and diseased lenses [20]. This global post-translational modification data set should provide some insight into the changes that take place and how these might relate to disease progression within the lens. Ultimately they may lead to the discovery of biomarkers, but for them to be useful in any sort of clinical setting, one would need use different techniques in order to detect them at much lower levels and requiring much less tissue.

6. Conclusion

Proteomic based biomarker discovery, while incredibly promising because of the potential wealth of protein biomakers, is still rather a new field. Both the technologies discussed in this review and many others not covered here are providing the necessary tools to tap a previously inaccessible classes of biomarkers. Protein biomarkers could not only be based on differences in the levels of specific proteins but also on differences in their post-translational modification state. 2D-GE and shotgun techniques such as MudPIT will play a role in the discovery process as well as other techniques that are even earlier in development. As these sources start to produce greater and greater amounts of data, database management and integration will become key to the process. The next few years should be an exciting time as more groups begin to apply these proteomic techniques on a systematic basis to both isolate biomarkers and to better understand disease.

Acknowledgements

References

[1] M.N. Boddy, P.H. Gaillard, W.H. McDonald, P. Shanahan, J.R. Yates III and P. Russell, Mus81-Eme1 are essential components of a Holliday junction resolvase, *Cell* **107** (2001), 537–548.

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

[3] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt and F.M. White, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae, *Nat Biotechnol* **20** (2002), 301–305.

[4] E.T. Fung, V. Thulasiraman, S.R. Weinberger and E.A. Dalmasso, Protein biochips for differential profiling, *Curr Opin Biotechnol* **12** (2001), 65–69.

[5] C.L. Gatlin, J.K. Eng, S.T. Cross, J.C. Detter and J.R. Yates III, Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry, *Anal Chem* **72** (2000), 757–763.

[6] A.C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga, Functional organization of the yeast proteome by systematic analysis of protein complexes, *Nature* **415** (2002), 141–147.

[7] M.B. Goshe, T.P. Conrads, E.A. Panisko, N.H. Angell, T.D. Veenstra and R.D. Smith, Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses, *Anal Chem* **73** (2001), 2578–2586.

[8] M.B. Goshe, T.D. Veenstra, E.A. Panisko, T.P. Conrads, N.H. Angell and R.D. Smith, Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins, *Anal Chem* **74** (2002), 607–616.

[9] T.J. Griffin, D.K. Han, S.P. Gygi, B. Rist, H. Lee, R. Aebersold and K.C. Parker, Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein identification by mass spectrometry, *J Am Soc Mass Spectrom* **12** (2001), 1238–1246.

[10] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb and R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat Biotechnol* **17** (1999), 994–999.

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

[12] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sorensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W. Hogue, D. Figeys and M. Tyers, Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry, *Nature* **415** (2002), 180–183.

[13] R.G. Krishna and F. Wold, Post-translational modification of proteins, *Adv Enzymol Relat Areas Mol Biol* **67** (1993), 265–298.

[14] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik and J.R. Yates III, Direct analysis of protein complexes using mass spectrometry, *Nat Biotechnol* **17** (1999), 676–682.

[15] M.J. MacCoss, W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark and J.R. Yates III, Shotgun identification of protein modifications from protein complexes and lens tissue, *PNAS* **12** (2002), 7900–7905.

[16] Y. Oda, T. Nagasu and B.T. Chait, Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome, *Nat Biotechnol* **19** (2001), 379–382.

[17] J.X. Pang, N. Ginanni, A.R. Dongre, S.A. Hefta and G.J. Opiteck, Biomarker discovery in urine by proteomics, *Journal of Proteome Research* **1** (2002) 161–169.

[18] E.C. Peters, D.M. Horn, D.C. Tully and A. Brock, A novel multifunctional labeling reagent for enhanced protein characterization with mass spectrometry, *Rapid Commun Mass Spectrom* **15** (2001), 2387–2392.

[19] F.E. Regnier, L. Riggs, R. Zhang, L. Xiong, P. Liu, A. Chakraborty, E. Seeley, C. Sioma and R.A. Thompson, Comparative proteomics based on stable isotope labeling and affinity selection, *J Mass Spectrom* **37** (2002), 133–145.

[20] A. Saraf et al., in preparation.

[21] A. Sickmann, K. Marcus, H. Schafer, E. Butt-Dorje, S. Lehr, A. Herkner, S. Suer, I. Bahr and H.E. Meyer, Identification of post-translationally modified proteins in proteome studies, *Electrophoresis* **22** (2001), 1669–1676.

[22] C. Tan, J. Li, J. Wang, Q. Xiang, X. Zhang, L. Dong, S. Shen, S. Liang and G. Li, Proteomic analysis of differential protein expression in human nasopharyngeal carcinoma cells induced by NAG7 transfection, *Proteomics* **2** (2002), 306–312.

[23] R. Verma, S. Chen, R. Feldman, D. Schieltz, J. Yates, J. Dohmen and R.J. Deshaies, Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes, *Mol Biol Cell* **11** (2000), 3425–3439.

[24] A. Vlahou, P.F. Schellhammer, S. Mendrinos, K. Patel, F.I. Kondylis, L. Gong, S. Nasim and G.L. Wright Jr, Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine, *Am J Pathol* **158** (2001), 1491–1502.

[25] S. Wang and F.E. Regnier, Proteomics based on selecting and quantifying cysteine containing peptides by covalent chromatography, *J Chromatogr* **A924** (2001), 345–357.

[26] M.P. Washburn, D. Wolters and J.R. Yates III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology, *Nat Biotechnol* **19** (2001), 242–247.

[27] D.A. Wolters, M.P. Washburn and J.R. Yates III, An automated multidimensional protein identification technology for shotgun proteomics, *Anal Chem* **73** (2001), 5683–5690.

[28] H. Zhou, J.D. Watts and R. Aebersold, A systematic approach to the analysis of protein phosphorylation, *Nat Biotechnol* **19** (2001), 375–378.