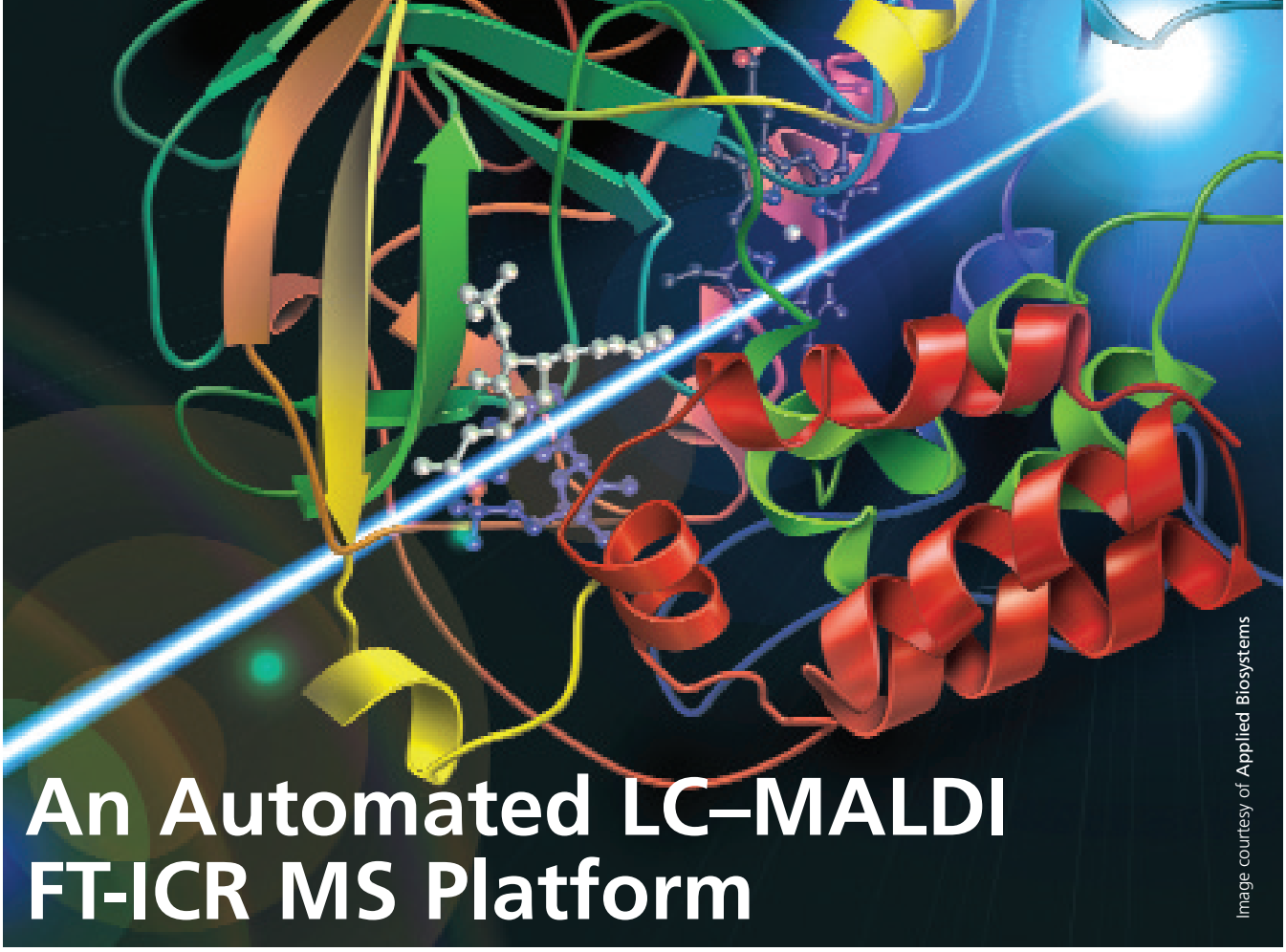


Image courtesy of Applied Biosystems

An Automated LC–MALDI FT-ICR MS Platform for High-Throughput Proteomics

Eric C. Peters, Ansgar Brock, David M. Horn, Qui T. Phung, Christer Ericson, Arthur R. Salomon, Scott B. Ficarro and Laurence M. Brill,
Genomics Institute of the Novartis Research Foundation, San Diego, California, USA.

The sheer complexity and ever-changing nature of the proteome necessitates the development of more powerful analysis systems. Each component of such a platform should be tailored and fully integrated with respect to the rest of the system in order to produce data of the highest quality in a high-throughput fashion. This article describes one such system built around MALDI FT-ICR MS in combination with effective automation of sample preparation, data acquisition and data reduction processes.

Introduction

Post-genomic biology is marked by the desire to study cells, cellular processes and disease at the protein level. However, such studies are considerably more difficult to perform than their genomic counterparts because RNA processing, proteolytic activation and many post-translational modifications (PTMs) can result in the production of numerous proteins from a single gene. Additionally, biological activity often results from the assembly of proteins into an active complex, the nature and composition of which can only be explored at the protein level. Finally, high-throughput analysis methods are required because, unlike the static nature of the genome, the expression and interactions of proteins are in a constant state of flux, varying over time, tissue type and in response to environmental changes.

Historically, two-dimensional gel electrophoresis (2D-GE) has been the dominant technique for simultaneously assessing the expression of multiple proteins.[1] The development and emergence of biological mass spectrometry (MS) in the early 1990s greatly increased the amount of information obtained using this technique,[2,3] enabling the identification of thousands of encoded proteins by peptide mapping or tandem MS experiments.[4] Although powerful, 2D-GE is laborious and possesses several other limitations.[5–7] For example, 2D-GE cannot fully address the extreme range of protein expression levels in complex living organisms as a result of sample loading restrictions imposed by the gels.[8] This limitation is of particular concern in that most regulatory proteins are often expressed at low copy numbers per cell.[8] Although this problem can be mitigated with extensive sample prefractionation,[9–11] system throughput is even further reduced.

These issues have prompted several groups to explore alternatives to 2D-GE based on multidimensional chromatography combined with MS.[12,13] The chromatographic separations serve to disperse the complexity of the initial sample, and can be performed at both the peptide[14] and protein levels.[15] Using these methodologies, the identification of proteins present in complexes, organelles and even whole organisms has been performed.[16,17] Additionally, novel methodologies for reducing the complexity of the analyte mixture based on the presence of particular amino acids[18,19] or specific PTMs[20–22] enable the analysis of low abundance regulatory proteins or PTMs.

These LC–MS methods exhibit tremendous potential for improved analyses of protein systems. However, the instrumentation and other methods employed are identical to those used for 2D-GE. Just as changes in the separation methodology enable more thorough analyses of highly complex protein samples, additional improvements in automation methods, component interfaces, MS instrumentation and bioinformatics strategies should allow similar improvements. Although each of these issues presents unique challenges, they must also be evaluated and optimized with respect to the other components of an analysis platform to produce data of the highest quality in a high-throughput manner. Here we describe one such integrated platform based on multidimensional LC–matrix-assisted laser desorption ionization (MALDI) Fourier-transform ion cyclotron resonance (FT-ICR) MS.

The MALDI FT-ICR MS Platform for High-Throughput Proteomics

The outline of our platform is shown in Figure 1. Although the outline delineates the experimental flow of a differential display experiment, the same platform can also be used for other studies, including peptide mapping, determination of the constituents of protein complexes, PTM identification and time-course studies. Samples are subjected to a series of chromatographic and affinity purification methods dictated by the experimental objectives in order to fractionate and/or enrich the sample. Although sample fractionation can be performed at both the intact protein or peptide level, all fractions are ultimately digested to peptides because of the increased ease of their quantitative chemical labelling. The peptide mixtures are then labelled in order to affect differential quantification, increase MALDI ionization efficiencies of peptides that lack arginine, and/or elucidate amino acid composition. The resulting peptide mixtures are subjected to reversed-phase micro high performance liquid chromatography (μHPLC) in parallel columns, and the separated peptides are automatically deposited in real time with matrix onto surface-patterned MALDI targets with minimal disruption to the integrity of the chromatographic separation. The resulting plates are then quickly scanned only in the MS mode using an FT-ICR MS equipped with a high-speed MALDI interface. The extremely high mass accuracies and resolutions afforded by the FT-ICR MS are used to identify peaks of interest for further investigation, thus restricting platform resources to the most interesting species. The deposited samples form a "permanent record", allowing species of interest to be investigated in greater detail without repeating the separation as required by real-time LC–MS coupling methods. Although deceptively simple, the successful implementation of this scheme requires the development of new proteomics instrumentation and methodologies to ensure effective systems integration. The following sections describe our progress towards implementation of this platform.

Although the implementation of highly efficient separation schemes is critical in multidimensional chromatography–MS proteomics analyses, equally important is the method chosen to interface the two components.

LC–MALDI MS: A Potentially Powerful Paradigm

Although the implementation of highly efficient separation schemes is critical in multidimensional chromatography–MS proteomics analyses, equally important is the method chosen to interface the two components. To date, LC–MS analyses of complex peptide mixtures have overwhelmingly employed electrospray ionization (ESI) methods because of the simplicity of their implementation. Although powerful, the operational parameters of ESI coupling methods also impose several limitations on high-throughput proteomics studies. Specifically, the separation system and mass spectrometer are coupled directly in real time, making parallel analysis systems difficult (or at least extremely costly) to implement, and often preventing the mass spectrometer from continually collecting useful data because of the equilibration and washing periods typical of HPLC separations. More importantly, current instrument-control and data-analysis software is not capable of real-time data-dependent processing during the course of a chromatographic separation except when employing simple selection criteria such as peak intensity. This often necessitates that upon the completion of a separation and subsequent analysis of the data, the same sample must be reanalysed to focus on those species that were not fully characterized during the initial run. Ultimately, these and other limitations result in dramatic reductions in overall platform throughput.

Thus, several groups have begun to explore MALDI-based analysis platforms that involve the creation of a "permanent record" of the multidimensional separation by depositing the effluents of the final separation columns directly onto MALDI target plates. Decoupling the separation from the mass spectrometer in this manner allows the separation to be performed free of artificial restrictions, while the mass spectrometer can operate at maximum throughput. The resulting plates can also be reanalysed as required without the need to repeat the separation step, thus decreasing sample requirements while simultaneously increasing the overall throughput of the system.

Clearly, the sample deposition method employed is critical to the success of creating high-fidelity "permanent records" of the separations. "Heart fractionation" methods[23] are relatively simple to implement, but potentially sacrifice a significant part of the chromatographic resolution. In contrast, sophisticated piezo-actuated microdispenser systems fabricated by the anisotropic wet etching of monocrystalline silicon have been described.[24,25] Similarly, specialized liquid junction-coupled subatmospheric pressure deposition chambers for the off-line

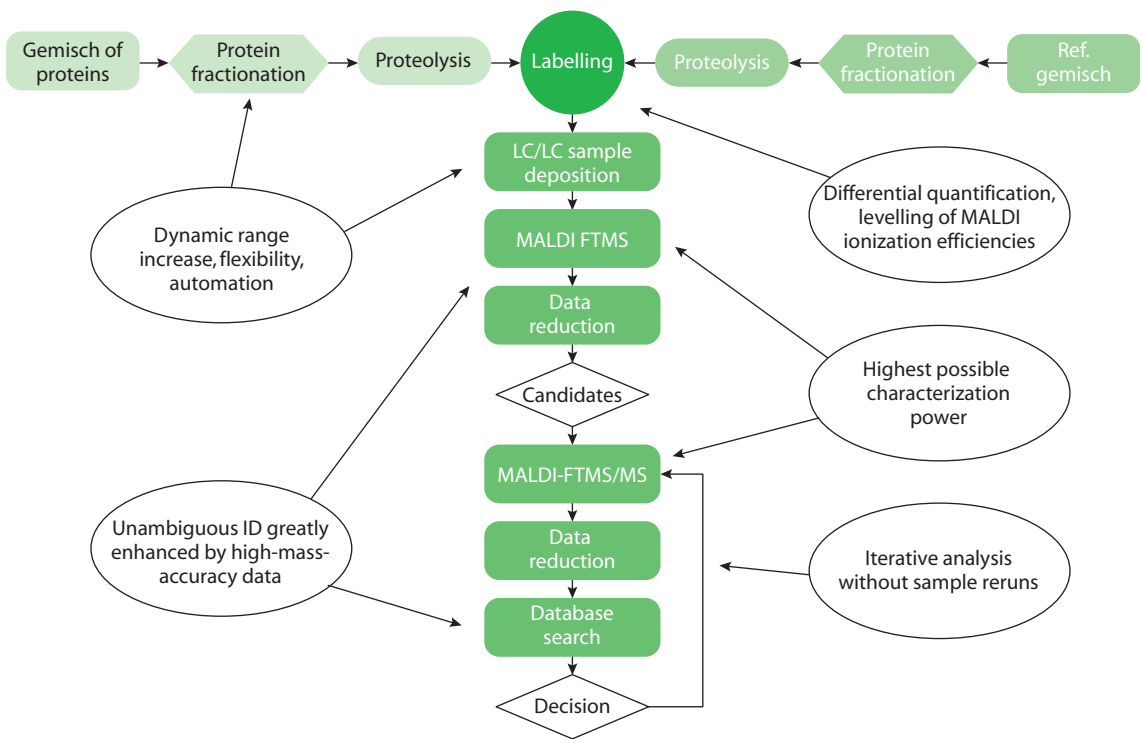
coupling of capillary electrophoresis with MALDI MS have also been reported.[26,27] In light of the potential power of a MALDI-based platform, our group has developed a simple, robust and fully automated system for the real-time deposition of the effluents of µHPLC columns and matrix onto MALDI target plates.

Figure 2 shows the concomitant deposition of a chromatographic run and matrix. The outlets of a series of µHPLC columns are arranged in parallel, and MALDI target plates on an *x,y*-translational stage are automatically moved underneath the columns. Importantly, the ends of the columns and the target plates never come into direct physical contact. Instead, the effluents of the columns are transferred to the plates through a charge induction mechanism by applying an intermittent negative potential to the plates. This results in a series of droplets of precisely controlled volume. The deposition system is also capable of electrospraying the column effluents as continual streams that should in theory more accurately maintain the fidelity of the separation. However, increases in the system's performance affected by localized deposition far outweigh the slight degradation of chromatographic resolution.

Our platform employs specially patterned target plates consisting of hydrophilic "spots" arrayed on an otherwise hydrophobic surface.[28] After deposition of a sample onto a spot, both the analyte and matrix localize into an area smaller than that occupied by the original droplet as the solvent evaporates, resulting in concentration of the analyte. The use of such target plates provides considerable advantages. For example, the sensitivities of ESI methods are known to be concentration dependent, often necessitating the use of nanochromatography to achieve maximum sensitivity.[12] Although effective, such nanoscale chromatography systems present practical problems and often require the manual loading of samples directly onto the separation column. By contrast, the spot target plates further concentrate the samples after the chromatographic process is complete, enabling the use of 300 µm internal diameter (i.d.) capillary columns and commercial autosamplers. In addition, the localization of analytes to precisely defined locations approximately 400 µm in diameter enables the MALDI stage to rapidly query only those regions that contain analyte. More importantly, increasing the size of the area irradiated by the MALDI laser to approximately 400 µm allows the entire sample to be queried simultaneously. This reduces the "sweet spot" problem often encountered when using the dried droplet method of sample preparation.[27] Together, these factors greatly increase the sample throughput of the overall platform.

The unique capabilities of this deposition system also address many of the issues unique to MALDI-based analyses. For example, the performance of MALDI MS can be greatly affected by competitive ionization effects, which are especially prevalent in complex mixtures.[29] The reversed-phase separation employed in conjunction with our automated deposition system minimizes these effects by providing a reproducible environment for the recrystallization of matrix and analytes with similar hydrophobicities. Additionally, the deposition system works equally well with aqueous solutions or a variety of organic solvents. This enables both the implementation of novel, automated on-plate recrystallization processes as well as the utilization of more hydrophobic matrices such as alpha-cyano-4-hydroxycinnamic acid (HCCA) that typically precipitate from acetonitrile/water mixtures before the desired localization process can be completed. This unique combination

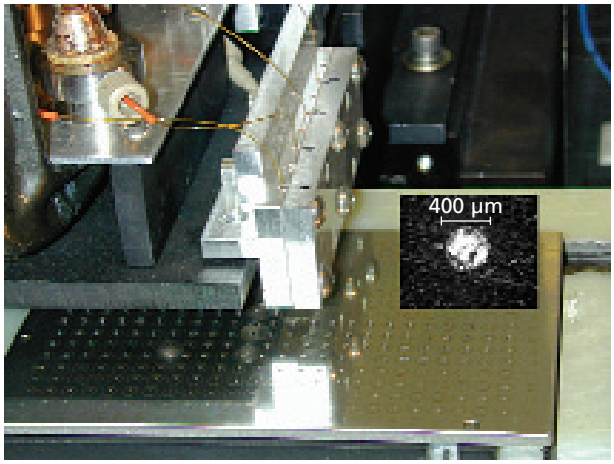
Figure 1: Flow chart of MALDI FT-ICR platform for the protein profiling of biological samples.



of automation and operational flexibility enables facile optimization of the many variables known to effect MALDI crystallization processes. Nevertheless, there are other sources of ion suppression in MALDI-based methods that cannot be rectified solely by optimized sample processing procedures.

Tryptic peptides containing a C-terminal arginine residue exhibit a four- to 18-fold increase in signal intensity compared with those peptides terminating in lysine.[30] Thought to result from the higher gas phase basicity of arginine residues, this bias can significantly limit the amount of information obtained using MALDI-based analysis methods. Several groups[31–33] have recently reported the use of *O*-methylisourea to convert lysine residues to more basic homoarginine residues, resulting in higher sequence coverages in peptide mapping experiments. Although effective in levelling ionization efficiencies, this label does not perform other functions typically enabled by other derivatization methodologies, such as differential quantification or directed tandem MS fragmentation. Therefore, our group developed a family of multifunctional labelling reagents that can affect several desirable functions simultaneously.[34] Figure 3 shows the reaction of 2-methoxy-4,5-dihydro-1*H*-imidazole where X = H with the ϵ-amino group of a lysine residue to form its 4,5-dihydro-1*H*-imidazol-2-yl derivative. Peptide mapping experiments of tryptic protein digests after reaction with this reagent showed total amino acid sequence coverages nearly double those of their unlabelled counterparts.[34] In addition, isotopic substitution of deuterium at the two methylene ring carbons simultaneously enables differential quantification by affecting a 4 Da mass difference per labelled lysine, while the label itself strongly biases tandem MS fragmentation patterns toward C-terminal fragment ions. This bias results in a highly simplified tandem mass spectrum consisting almost exclusively of *y*-ions. Other mass differences can also be affected by performing different functionalization reactions at these two ring positions. Although the term high-throughput has traditionally been equated simply with an increase in speed, implementation of our labelling methodology further increases the overall throughput of the system by also providing greater information content per measurement performed. For example, the levelling of peptide ionization efficiencies enables the detection of a greater number of peptides per sample. Similarly, the unique chemical selectivity of these labels for the ϵ-amino group of lysine residues over the α-amino group at a peptide's *N*-terminus combined with the 4 Da mass difference per differentially labelled lysine enable the determination of the exact number of lysines residues contained by a given peptide. When combined with accurate mass measurements, this additional compositional information further increases one's confidence in protein/peptide identifications resulting from database searches.

Figure 2: Automated deposition of a reversed-phase μHPLC chromatography run with concomitant addition of 3,5-dihydroxybenzoic acid as matrix.

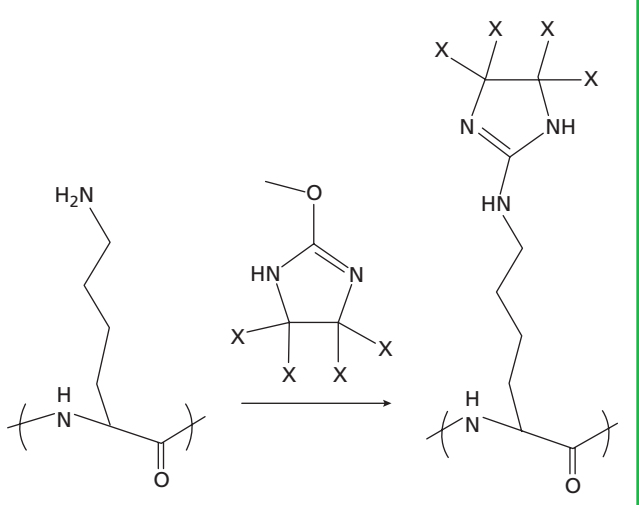


Automated High-Resolution/High-Mass-Accuracy MALDI MS

The sample processing methods described above have been designed to maximize the throughput of any MALDI-based analytical platform, and are therefore compatible with a wide range of commercially available instrumentation. However, the desire to analyse protein systems of ever greater complexity more thoroughly will also place increasing demands on the performance of the mass analyser employed. Thus, we decided to incorporate an FT-ICR mass spectrometer into our platform because of its unparalleled mass accuracy (~1 ppm), high resolution (routinely 100 000), large dynamic range (routinely 1000) and good sensitivity (low fmol to high amol).[35] More importantly, we also developed a set of methodologies and instrumentation modifications designed to leverage the full advantages of combining MALDI and FT-ICR MS within an automated, robust analysis platform.

A 384 or 1536-microtitre format target plate containing deposited analytes is mounted onto linearly encoded high precision *x*- and *y*-stages in a custom-built intermediate pressure MALDI source.[36] Following UV laser irradiation, the generated ions are collisionally cooled by the surrounding nitrogen buffer gas (pressure of 40 mTorr) and guided by a cooling quadrupole to the entrance of a selection quadrupole, through which they are passed into a hexapole ion guide for transient storage. The selection quadrupole can be operated in integral or mass-selective mode, allowing the isolation of a

Figure 3: Reaction of a lysine residue with 2-methoxy-4, 5-dihydro-1*H*-imidazole where X = H to form its 4,5- dihydro-1*H*-imidazol-2-yl derivative.



narrow mass range before ion accumulation. Internal calibration, which is required to ensure the high mass accuracy inherent in FT-ICR MS, is achieved by employing a novel gas-phase mixing scheme. Specifically, after sample irradiation and storage of the resulting ions in the hexapole, the stage quickly moves to a strip containing peptide calibrants embedded in a MALDI matrix located on the edge of the plate. Calibrant ions are then generated and mixed with the sample ions in the hexapole, and the entire packet is transferred into the mass analyser. Software has been written to both automate the acquisition of mass spectra without user intervention as well as deconvolute the resulting isotopic clusters.[37] The total time required for the acquisition of a typical mass spectrum is approximately 7 to 10 seconds, enabling internally calibrated mass spectra for 384 samples to be acquired in less than 1 hr. Similarly, automated tandem MS can be performed in the analyser cell by sustained off-radiance irradiation collisionally activated dissociation (SORI-CAD)[38] or by infrared multiphoton dissociation (IRMPD).[39]

Figure 4 demonstrates the utility of high-resolution measurements in a simulated differential display experiment.[40] Two peptides differing in mass by 40 mDa were labelled separately with a 1:3 mixture of the *N*-hydroxysuccinimide esters of nicotinic acid: d_4-nicotinic acid for the lower mass peptide or 3:1 for the larger mass species.[41] Equal amounts of each labelled peptide were combined and a mass spectrum of the resulting mixture was obtained on both a MALDI-TOF and our MALDI FT-ICR. The spectrum from the MALDI-TOF shows what appears to be a single peptide labelled in a 1:1 ratio, whereas the high resolution of the FT-ICR mass spectrum clearly shows the presence of the two differentially labelled isotopic clusters. Such high-resolution measurements are only feasible using FT-ICR MS. For extremely complex mixtures containing hundreds of thousands of peptides, lower resolution measurements may result in the loss or misinterpretation of data as demonstrated by the MALDI-TOF spectrum.

The importance of mass measurement accuracy for protein identification by both peptide mapping as well as tandem MS is documented.[42,43] High mass accuracy measurements provide greater confidence in protein identification assignments, and enable proteins to be identified with either less sequence coverage (peptide mapping) or fewer tandem MS fragments. The latter is especially important as ions generated by MALDI are primarily singly charged species, which generally yield fewer fragment ions than multiply charged ions in tandem MS experiments. Figure 5 shows the SORI-CAD spectrum of an unknown peptide from a tryptic digest of all soluble cytosolic proteins in yeast. While only three peptide fragments were detected, this data was sufficient to unambiguously distinguish this peptide from all other tryptic peptides in the yeast proteome. The highly accurate mass measurements obtained for both the parent as well as the fragment ions (2 ppm error) enabled the unequivocal identification of glyceraldehyde 3-phosphate dehydrogenase. To our knowledge, this level of mass accuracy has not been obtained in an automated fashion using any other MALDI MS combination.

To demonstrate the capacity of our integrated platform to effectively analyse especially complex mixtures, an extract of all soluble yeast proteins was prepared and digested with trypsin. This digest (5 µg) was loaded onto a 300 mm i.d. reversed-phase µHPLC column, and the effluent was automatically deposited every 10 s with matrix onto a MALDI plate. This "permanent record" of the separation was automatically analysed using our MALDI FT-ICR MS. Figure 6 shows a representation of the deconvoluted data with respect to mass, fraction number (called "spot" in the figure), and signal-to-noise ratio. Over 11 000 unique monoisotopic masses were found, with many spectra containing over 200 monoisotopic masses. Furthermore, numerous spectra exhibited dynamic ranges of over 1000. Because only a single chromatographic separation was performed, these results illustrate the excellent performance of our integrated platform. The same sample was

Figure 4: Demonstration of the importance of high-mass-resolution measurements. Mass spectra of the same sample analysed using (a) a commercial MALDI TOF or (b) our MALDI FT-ICR MS.

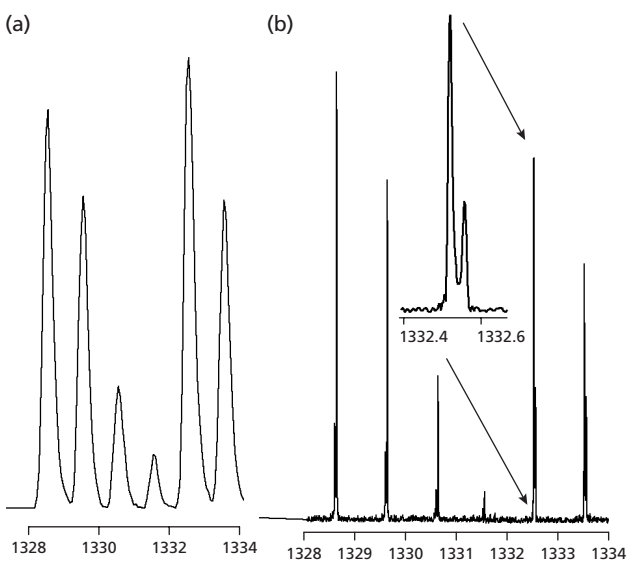


Figure 5: SORI-CAD spectrum of an unknown peptide with mass 1752.58, from a tryptic digest of all soluble cytosolic proteins in yeast.

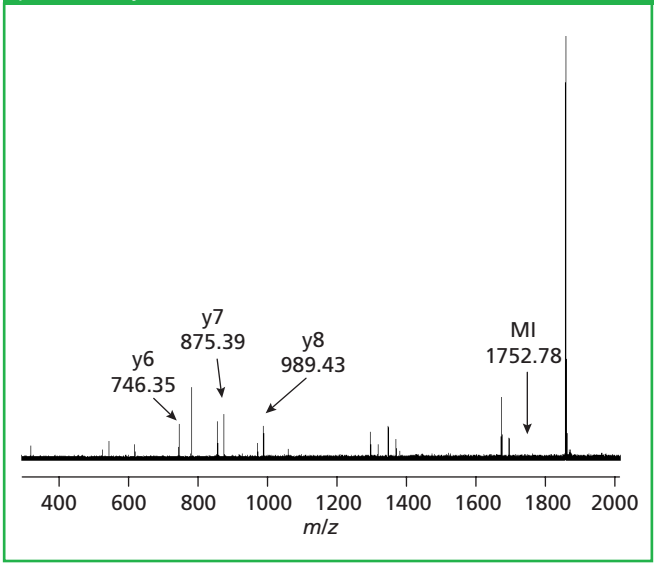
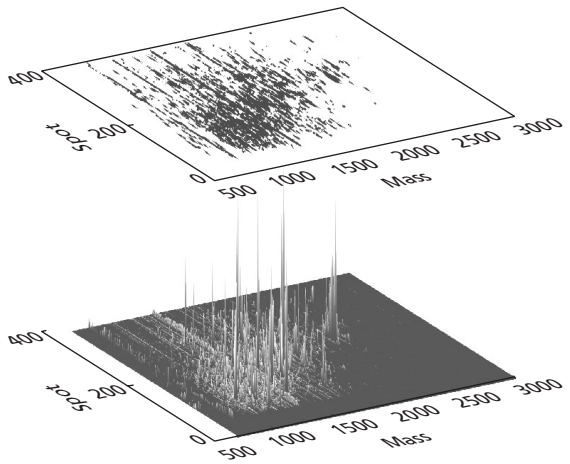


Figure 6: Three-dimensional plot of the reversed-phase μHPLC–FT-ICR MS analysis of a tryptic digest of the soluble proteins isolated from yeast.



then reanalysed using a multidimensional separation scheme employing strong cation exchange (SCX) fractionation before μHPLC. The sample was eluted in four salt steps from the SCX column, each of which was simultaneously separated on four μHPLC columns and deposited with matrix onto a 1536 format MALDI target plate. Analyses of these fractions produced 3D profiles that were entirely different, each of which contained a similar number of peptides as seen in the preceding experiment. This demonstrates the even greater peak capacity of the system when additional separation steps are employed.

Conclusion

The implementation of a high-throughput proteomics platform based on MALDI FT-ICR MS that enables protein identification, MS–MS, and complex mixture analyses has been demonstrated. In addition to the benefits of effective automation, higher throughputs are also effected by the high quality of the data produced by the FT-ICR mass analyser. The importance of high-mass-accuracy and high-resolution measurements should continue to grow as researchers seek to study protein systems of ever greater complexity.

References

1. G.L. Corthals et al., *Electrophoresis*, **21**, 1104–1115 (2000).
2. J.B. Fenn et al., *Science*, **246**, 64–71 (1989).
3. M. Karas and F. Hillenkamp, *Anal. Chem.*, **60**, 2299–2301 (1988).
4. S.D. Patterson and R. Aebersold, *Electrophoresis*, **16**, 1791–1814 (1995).
5. E. Bordini, M. Hamdan and P.G. Righetti, *Electrophoresis*, **21**, 2911–2918 (2000).
6. M. Galvani et al., *Rapid Commun. Mass Spectrom.*, **14**, 18–25 (2000).
7. T. Voss and P. Haberl, *Electrophoresis*, **21**, 3441–3459 (2000).
8. S.P. Gygi et al., *Proc. Natl Acad. Sci. USA*, **97**, 9390–9395 (2000).
9. R.S. Taylor et al., *Electrophoresis*, **21**, 3441–3459 (2000).
10. B. Herbert and P.G. Righetti, *Electrophoresis*, **21**, 3639–3648 (2000).
11. E. Jung et al., *Electrophoresis*, **21**, 3369–3377 (2000).
12. J.R. Yates III, *Trends Genet.*, **16**, 5–8 (2000).
13. R. Aebersold and D.R. Goodlett, *Chem. Rev.*, **101**, 269–295 (2001).
14. S.E. Martin et al., *Anal.Chem.*, **72**, 4266–4274 (2000).
15. D.B. Wall et al., *Anal. Chem.*, **72**, 1099–1111 (2000).
16. A.C. Gavin et al., *Nature*, **415**, 141–147 (2002).
17. A.J. Link et al., *Nat. Biotechnol.*, **17**, 676–682 (1999).
18. S.P. Gygi et al., *Nat. Biotechnol.*, **17**, 994–999 (1999).
19. S. Wang and F.E. Regnier, *J. Chromatogr. A*, **924**, 345–357 (2001).
20. P. Cao and J.T. Stults, *J. Chromatogr. A*, **853**, 225–235 (1999).
21. M.B. Goshe et al., *Anal. Chem.*, **73**, 2578–2586 (2001).
22. Y. Oda, T. Nagasu and B.T. Chait, *Nat. Biotechnol.*, **19**, 379–382 (2001).
23. T.J. Griffin et al., *Anal. Chem.*, **73**, 978–986 (2001).
24. S. Ekstrom et al., *Anal. Chem.*, **72**, 286–293 (2000).
25. S. Ekstrom et al., *Anal. Chem.*, **73**, 214–219 (2001).
26. J. Preisler, F. Foret and B.L. Karger, *Anal. Chem.*, **70**, 5278–5287 (1998).
27. J. Preisler et al., *Anal. Chem.*, **72**, 4785–4795 (2000).
28. M. Schuerenberg et al., *Anal. Chem.*, **72**, 3436–3442 (2000).
29. S.L. Cohen and B.T. Chait, *Anal. Chem.*, **68**, 31–37 (1996).
30. E. Krause, H. Wenschuh and P.R. Jungblut, *Anal. Chem.*, **71**, 4160–4165 (1999).
31. R.L. Beardsley, J.A. Karty and J.P. Reilly, *Rapid Commun. Mass Spectrom.*, **14**, 2147–2153 (2000).
32. F.L. Brancia, S.G. Oliver and S.J. Gaskell, *Rapid Commun. Mass Spectrom.*, **14**, 2070–2073 (2000).
33. T. Keough, M.P. Lacey and R.S. Youngquist, *Rapid Commun. Mass Spectrom.*, **14**, 2348–2356 (2000).
34. E.C. Peters et al., *Rapid Commun. Mass Spectrom.*, **15**, 2387–2392 (2001).
35. A.G. Marshall, C.L. Hendrickson and G.S. Jackson, *Mass Spectrometry Reviews*, **17**, 1–35 (1998).
36. A.N. Krutchinsky et al., *Rapid Commun. Mass Spectrom.*, **12**, 508–518 (1998).
37. D.M. Horn, R.A. Zubarev and F.W. McLafferty, *J. Am. Soc. Mass Spectrom.*, **11**, 320–332 (2000).
38. J.W. Gauthier, T.R. Trautman and D.B. Jacobsen, *Anal. Chem. Acta*, **246**, 211–225 (1991).
39. D.P. Little et al., *Anal. Chem.*, **66**, 2809–2815 (1994).
40. M.A. Moseley, *Trends Biotechnol.*, **19**, S10–S16 (2001).
41. M. Munchbach et al., *Anal. Chem.*, **72**, 4047–4057 (2000).
42. A. Shevchenko et al., *Anal. Chem.*, **72**, 2132–2141 (2000).
43. K.R. Clauser, P. Baker and A.L. Burlingame, *Anal. Chem.*, **71**, 2871–2882 (1999).

Dr Eric C. Peters received his BS (Chemistry) from Georgetown University in 1989, his MS (Polymer Chemistry) from Cornell University in 1996, and a PhD (Chemistry) from the University of California, Berkeley in 1999. At Berkeley, Dr Peters developed a number of novel monolithic polymer supports for use in chromatography, fluidics, catalysis and combinatorial chemistry applications. Since 1999, Dr Peters has been at the Genomics Institute of the Novartis Research Foundation in San Diego. He is currently a group leader of the Protein Profiling/ MS section, whose research has focused on the development of novel LC–MS-based methodologies for proteomics applications.

Dr Peters can be contacted at peters@gnf.org