print

Automated Liquid Chromatography MALDI FT-ICR MS Platform for Proteomics: Automated High Performance Mass Spectrometry and Data Analysis

Ansgar Brock*, David M. Horn, Christopher M. Shaw, Eric C. Peters, Christer Ericson, Qui Phung, Scott B. Ficarro, Arthur R. Salomon, *Genomics Institute of the Novartis Research Foundation

The thorough analysis of even relatively simple native protein systems requires the development of new analytical tools due to the ever changing nature and the large dynamic range of protein expression in living organisms. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) can provide excellent dynamic range, mass accuracy, and throughput if the experimental design is adapted to suit this method of mass analysis. This paper shows how an optimized tool for protein analysis can be built around FT-ICR MS when combined with matrix-assisted laser desorption ionization (MALDI) and effective automation of sample preparation, data acquisition, and data reduction processes.

Introduction

From the considerations presented in a previous article of this series (Volume 5 Issue 3), it is clear that a proteomics approach based on MALDI [1] possesses significant advantages compared to the current predominant approach of on-line coupling of separations to the mass spectrometer through electrospray ionization (ESI) [2]. MALDI methods have recently been demonstrated on mass analyzers that are suitable for high-throughput protein identification using tandem mass spectrometry, including quadrupole ion trap, quadrupole time-of-flight, time-of-flight/time-of-flight (TOF/TOF), and Fourier transform ion cyclotron resonance. Although each of these instruments have their own operational advantages, the choice of mass analyzer to be employed in a proteomics platform must ultimately be based on which one possesses the best compromise of sensitivity, dynamic range, resolution, mass accuracy, and level of automation required for the successful analysis of complex protein mixtures.

The desire to study biological systems of ever greater complexity at the protein level requires the identification and characterization of individual proteins present in mixtures [3] ranging from affinity interaction partners [4] to native protein complexes [5, 6], and in the extreme case, the full protein content of a living organism (proteome). Whereas protein complexes may contain a large number of constituents, the relative abundances of the individual constituents are expected to be quite similar. This is quite different than the case of overall protein expression within an organism. The dynamic range of protein expression in yeast was reported to be around 10E4 [7], and concentration differences are estimated to reach 10E9 to 10E12 (8) in plasma. Therefore, high dynamic range must be a major selectioncriterion for any mass analyzer used in a general proteomics platform.

In addition to the challenges presented by the large dynamic range of protein expression, the mixtures to be analyzed can also be extremely complex, containing hundreds of thousands (or even millions) of individual peptides. Even though ultra high-resolution separations [9] and/or multi-dimensional separations [10] reduce the number of masses that need to be displayed simultaneously, it still often is necessary to resolve hundreds of isotopic clusters in a given spectrum. Thus, high-resolution mass analyses are an essential component of large-scale proteomics platforms.

The importance of mass measurement accuracy with respect to protein identification by both peptide mapping as well as tandem MS has already been demonstrated [11, 12]. Not only do high mass accuracy measurements provide greater confidence in protein identification assignments, but they also enable proteins to be identified with either less sequence coverage (in the case of peptide mapping) or fewer tandem MS fragments. The later is especially important as MALDI-generated ions are primarily singly-charged species, which generally yield fewer fragment ions than multiply-charged ions in tandem MS experiments. High mass measurement accuracy can also conceivably allow (at least in selected cases) protein identifications to be made on the basis of the mass of a single peptide [13]. This alone opens the door to even higher-throughputs in the analysis of mixtures due to the significant decrease in time spent on tandem MS experiments. In addition, a concomitant time saving in the cross correlation process of mass spectral data with in-silico digested databases [14, 15] would also be achieved.

In light of these considerations, it was decided that a FT-ICR mass spectrometer was the mass analyzer of choice due to its

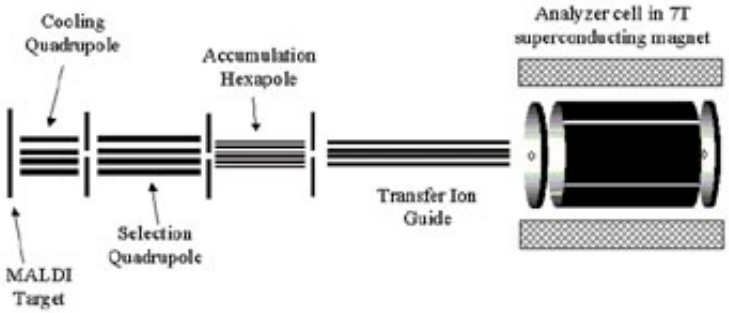
unparalleled mass accuracy (~1 ppm), high resolution (routinely >10E5), large dynamic range (~ 10E3, single scan), and good sensitivity (amol). Therefore, we developed a set of methodologies and instrumentation designed to leverage the full advantages of FT-ICR MS within an automated, robust analysis platform.

Experimental

Before MS analysis, several calibrant peptides are deposited with MALDI matrix in strips on each side of an industry standard 384 or 1536 microtiter-format MALDI target plate that carries the separated and deposited samples. This plate is mounted onto optically-encoded linear high precision x- and y-stages in a custom-built intermediate pressure MALDI source [16]. Sample or calibrant ions are generated by positioning the sample locations or calibrant strips within the focal point of a UV-laser beam (Figure 1). Ions are collisionally cooled by the surrounding nitrogen buffer gas (40 mTorr) and guided by a quadrupole (cooling quadrupole) to the entrance of a second quadrupole (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 an individual species previous to ion accumulation in the hexapole. Internal calibration, which is of utmost importance to ensure the high mass accuracy inherent in FT-ICR MS, is achieved by employing a novel gas phase mixing scheme. Specifically, after a sample is irradiated with the UV laser and the resulting ions are stored in the hexapole, the sample stage moves in a fraction of a second to the closer of the two calibrant strips on the side of the plates and irradiates it. The calibrant ions are mixed with the ions from the sample in the hexapole and the entire packet is transferred through the remaining ion optics into the mass analyzer, which is located inside the bore of a 7 T superconducting magnet. Ions are cooled by gas injection after being transferred to the analyzer cell and then detected. Software has been written to automate the acquisition of mass spectra for each spot on a sample plate without user intervention. The total time required for the acquisition of a typical mass spectrum is roughly 7 to 10 seconds, mostly due to the time required for pump down after gas injection (2 s) and the acquisition of a million points time domain signal (~3.5 s) [17]. Thus, internally calibrated mass spectra for all 384 samples fractions typically deposited during a reversed-phase mHPLC run are acquired in less than 1 hr.

Figure 1

Schematic of FT-ICR Ion Optics



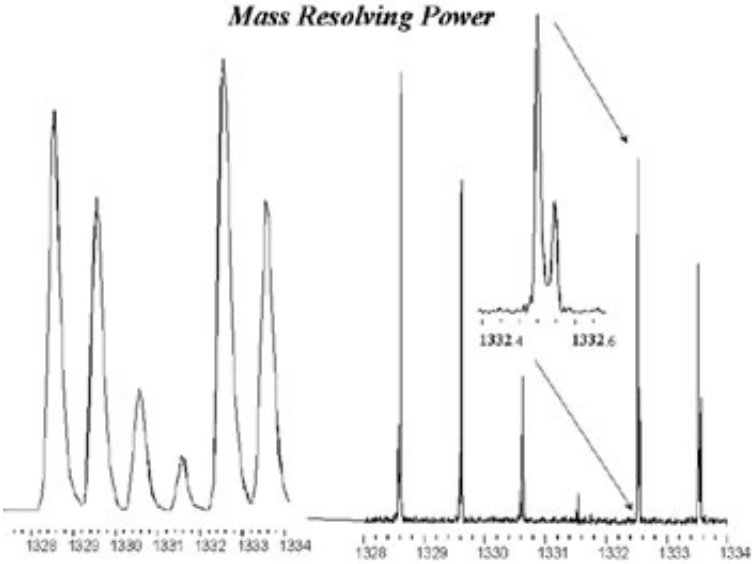
For tandem MS experiments, the precursor ions are first isolated by the selection quadrupole. Mass selection before ion accumulation is advantageous due to the “space charge” limit of the FT-ICR ion cell [17]. This “selected ion accumulation” allows the accumulation of only the species of interest in the cell, greatly improving the dynamic range for tandem MS experiments. After selective accumulation, ions are injected into the analyzer cell where they are fragmented either by sustained off-resonance irradiation collisionally activated dissociation (SORI-CAD) [18] or by infrared multi-photon dissociation (IRMPD) [19]. Software that performs automated tandem MS of selected peptides has been created such that it accepts a list of masses and their location on a given plate, automatically isolates these ions using the selection quadrupole, and subsequently fragments the ions. All data reduction has been automated using the THRASH algorithm [20]. Reduced mass spectral data from peptide mapping and tandem MS experiments are automatically submitted for database searching using the Mascot software package.

Results and Discussion

Figure 2 demonstrates the utility of high resolution measurements [21]. Two peptides differing in mass by 40 mDa were labeled separately with a 1:3 mixture of the N-Hydroxysuccinimide esters of nicotinic acid: d4-nicotinic acid for the lower mass peptide or 3:1 for the species of higher molecular weight [22]. The labeled peptides were combined and a mass spectrum of the resulting mixture was obtained on both a MALDI-TOF and our MALDI FT-ICR. The MALDI-TOF spectrum shows what appears to be only a

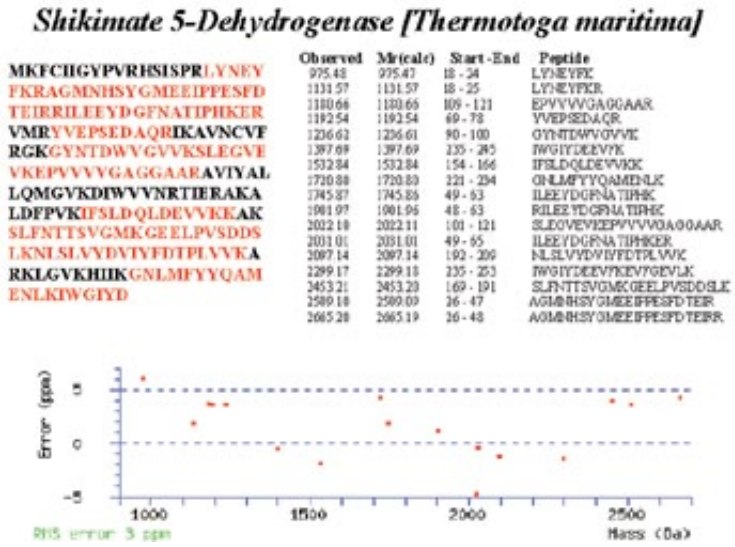
single peptide labeled in a 1:1 ratio, whereas the high resolution of the FT-ICR mass spectrum clearly shows the presence of the two differentially-labeled isotopic clusters. A resolution of at least 33,000 is required according to the full-width half maximum (FWHM) criterion in order to resolve the signals of the two peptides. Such high resolution measurements are only readily attainable using FT-ICR MS. In extremely complex mixtures that contain hundreds of thousands of peptides, the lack of such high resolution may lead to incomplete or even misleading information as demonstrated by the MALDI-TOF spectrum.

Figure 2



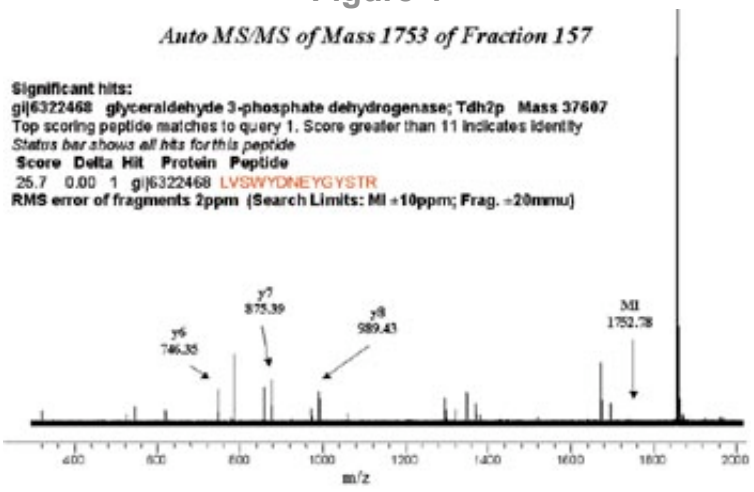
The high mass measurement accuracy afforded by FT-ICR is also highly advantageous for protein identification. Figure 3 shows the database search results for an internally calibrated peptide mass finger print of a tryptic digest. The root-mean-squared mass accuracy of 3 ppm for assigned peptides spanning a range of 1700 m/z resulted in the unambiguous identification of shikimate 5-dehydrogenase (thermotoga maritima) from the NCBI non-redundant database using the Mascot protein identification software, which returned a score of 290 (69% sequence coverage). Since a score of 45 for this search indicates 95% confidence in the protein identification and the returned Mascot score is proportional to the negative of the logarithm of the probability [15], there is a ~10E-27 chance that this identification is incorrect. Furthermore, the next most probable match is assigned a score of only 21, which is significantly below the confidence threshold. This spectrum was acquired as part of an automated MS run of tryptic digests of 96 protein samples. The entire process including data acquisition with internal calibration, data reduction, and protein identification was completed in less than two hours. Of these 96 samples, 91 were unambiguously identified in the NCBI non-redundant database, most with Mascot scores well above 100, while the remaining five samples could not be identified due to insufficient protein concentration. For comparison, the same samples were analyzed on our MALDI TOF instrument, which required several days of manual data aquistion and interpretation by one of the authors, and resulted in just 61 protein identifications with scores above the statistical threshold of 45. The average top score for TOF data was 63.5 versus 101.5 for FT-ICR, and the average score difference between first and second assignments was 38.8 for TOF data and 79.9 for FT-ICR data. These results clearly demonstrate the benefits of high mass accuracy and high throughput afforded by our instrument.

Figure 3



High mass accuracy is also extremely powerful for tandem MS experiments. Figure 4 shows the SORI-CAD spectrum of an unknown peptide originating from a tryptic digest of all the soluble cytosolic proteins in yeast. While only three fragment ion masses could be assigned in this experiment, this data was sufficient to unambiguously identify glyceraldehyde 3-phosphate dehydrogenase using the Mascot protein identification software due to the high mass measurement accuracy for both the parent and fragment ions (2 ppm error). The stringent search specificities employed (10 ppm for the parent ion, 0.020 Da for fragment ions) were enough to eliminate any possibility that this could be any other tryptic peptide in the whole yeast proteome. Thus, even with limited sequence information, the high mass accuracy of FT-ICR MS allows unambiguous assignment of peptides subjected to tandem MS.

Figure 4



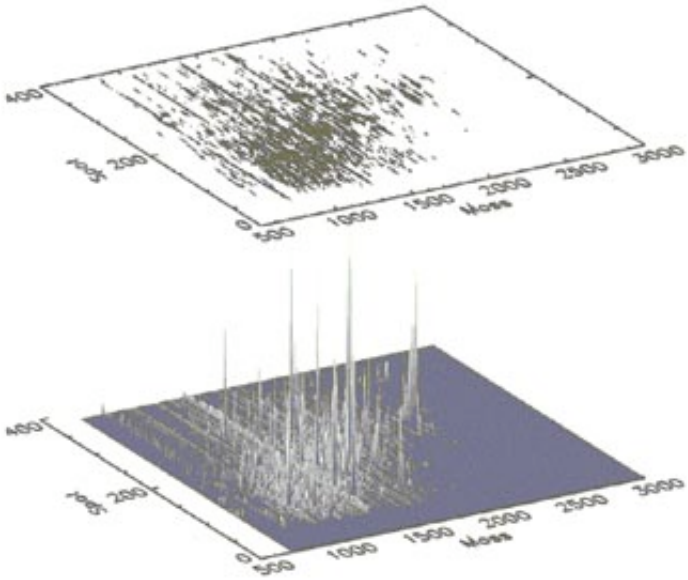
>

The high mass accuracy, resolution, and dynamic range of FT-ICR MS are critically important for the thorough analysis of protein expression patterns. In order to demonstrate the capacity of our instrument to handle such complex mixtures, a soluble yeast protein extract was prepared, reduced, alkylated, and digested with trypsin. 5 mg of this digest was separated on a 300 mm i.d. reversed-phase mHPLC column run at 3 mL/min, and 10 s fractions of the effluent were codeposited with matrix onto a MALDI plate. Figure 5 shows a representation of the reduced data in three-dimensional space spanned by mass, fraction number (called “spot” in the figure), and signal-to-noise ratio. Over 11,000 unique monoisotopic masses were found in this data set, with a considerable number of spectra exhibiting over 200 masses. The typical dynamic range observed in these single-scan spectra was 500 and in quite a few cases the dynamic range was over 1000. In additional experiments, an identical sample was first fractionated by strong cation–exchange (SCX) into 20 fractions before each fraction was subjected to mHPLC. In this 2D-LC MS experiment a total of 76,000 unique monoisotopic masses were detected as compared to 11,000 when only a single dimension of separation was employed. This powerfully demonstrates the increase in overall peak capacity and dynamic range of the analysis

platform due to adding a second dimension of separation.

Figure 5

LC MALDI FT-ICR MS Run of Yeast



>

Summary

The implementation of a proteomics platform based on MALDI FT-ICR MS has been demonstrated that enables peptide mapping, MS/MS, and complex mixture analyses in a fully automated and high-throughput fashion. A unique internal calibration scheme involving mixing of calibrant and sample ions in the gas phase was demonstrated, thus avoiding the problems of signal suppression or over spiking of calibrant into a sample typical of other MALDI MS internal calibration methods. Additionally, the successful identification of peptides from complex mixtures using automated tandem MS was demonstrated despite the poorer fragmentation characteristics of singly-charged ions. Future research will focus on improving both MS/MS efficiency as well as mass measurement accuracy of this platform, with improvements in the later further minimizing the need for MS/MS experiments and thus, also increasing the system throughput.

References

1. Karas, M.; Hillenkamp, F. Anal. Chem. 1988, 60, 2299-2301.

2. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Science 1989, 246, 64-71.

3. Peng, J.; Gygi, S. P. J. Mass Spectrom. 2001, 36, 1083-1091.

4. A. C. Gavin, e. a. Nature 2002, 415, 141-147.

5. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates III., J. R. Nat. Biotechnol. 1999, 17, 676-682.

6. Rout, M. P.; Aitchison, J. D.; Suprapto, A.; Hjertaas, K.; Zhao, Y.; Chait, B. T. J. Cell Biol. 2000, 148, 635-651.

7. Futcher, B.; Latter, G. I.; Monardo, P.; McLaughlin, C. S.; Garrels, J. I. Mol. Cell Biol. 1999, 19, 7357-7368.

8. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J.-C. Electrophoresis 2000, 21, 1104-1115.

9. Shen, Y.; Tolic, N.; Zhao, R.; Pasa-Tolic, L.; Li, L.; Berger, S. J.; Harkewicz, R.; Anderson, G. A.; Belov, M. E.; Smith, R. D. Anal. Chem. 2001, 73, 3011-3021.

10. Washburn, M. P.; Wolters, D.; Yates, J. R., III. Nat. Biotechnol. 2001, 19, 242-247.

11. Clauser, K. R.; Baker, P.; Burlingame, A. L. Anal. Chem. 1999, 71, 2871-2882.

12. Shevchenko, A.; Loboda, A.; Schevchenko, A.; Ens, W.; Standing, K. G. Anal. Chem. 2000, 72, 2132-2141.

13. Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Pa_a-Tolic, L.; Smith, R. D. Anal. Chem. 2000, 72, 3342-3354.

14. Eng, J. K.; McCormack, A. L.; Yates, J. R., III J. Am. Soc. Mass Spectrom. 1994, 5, 976-989.

15. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Electrophoresis 1999, 20, 3551-3567.

16. Krutchinsky, A. N.; Loboda, A. V.; Spicer, V. L.; Dworschak, R.; Ens, W.; Standing, K. G. Rapid Commun. Mass Spectrom. 1998, 12, 508-518.

17. \Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Mass Spectrom. Rev. 1998, 17, 1-35.

18. Gauthier, J. W.; Trautman, T. R.; Jacobsen, D. B. Anal. Chim. Acta 1991, 246, 211-225.

19. Little, D. P.; Speir, J. P.; Senko, M. W.; O'Cornnor, P. B.; McLafferty, F. W. Anal. Chem. 1994, 66, 2809-2815.

20. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. J. Am. Soc. Mass Spectrom. 2000, 11, 320-332.

21. Moseley, M. A. Trends Biotechnol. 2001, 19, S10-S16.

22. Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. Anal. Chem. 2000, 72, 4047-4057.

Dr. Brock *received his Ph.D. in Physical Chemistry from Baylor University in 1996. He served as a Postdoctoral associate at Stanford University during which Hadamard Transform TOFMS was developed. In 1999, he transferred to the newly founded Genomics Institute of the Novartis Research Foundation (GNF). His main responsibilities are in the areas of new technology development for proteomics and the support of a number of scientific projects at GNF and The Scripps Research Institute in La Jolla.*

Contact email: brock@gnf.org

*This article was published in the Winter 2002 issue of APR, Volume 5, Issue 4, on pages 94-99.