OMICS A Journal of Integrative Biology
Volume 6, Number 1, 2002
Mary Ann Liebert, Inc.

Review

The Use of Accurate Mass Tags for High-Throughput Microbial Proteomics

RICHARD D. SMITH, GORDON A. ANDERSON, MARY S. LIPTON, CHRISTOPHE MASSELON, LJILJANA PAŠA-TOLIC´, YUFENG SHEN, and HAROLD R. UDSETH

ABSTRACT

We describe and review progress towards a global strategy that aims to extend the sensitivity, dynamic range, comprehensiveness, and throughput of proteomic measurements for microbial systems based upon the use of polypeptide accurate mass tags (AMTs) produced by global protein enzymatic digestions. The two-stage strategy exploits high accuracy mass measurements using Fourier transform ion cyclotron resonance mass spectrometry (FTICR) to validate polypeptide AMTs for a specific organism, from potential mass tags tentatively identified using tandem mass spectrometry (MS/MS), providing the basis for subsequent measurements without the need for routine MS/MS. A high-resolution capillary liquid chromatography separation combined with high sensitivity, and high-resolution accurate FTICR measurements is shown to be capable of characterizing polypeptide mixtures of more than 10^5 components, sufficient for broad protein identification using AMTs. Advantages of the approach include the high confidence of protein identification, its broad proteome coverage, and the capability for stable-isotope labeling methods for precise relative protein abundance measurements. The strategy has been initially evaluated using the microorganisms *Saccharomyces cerevisiae* and *Deinococcus radiodurans*. Additional developments, including the use of multiplexed-MS/MS capabilities and methods for dynamic range expansion of proteome measurements that promise to further extend the quality of proteomics measurements, are also described.

INTRODUCTION

TO UNDERSTAND HOW A BIOLOGICAL ORGANISM OPERATES and the functions of its component parts, it is enabling to be able to study how its components change and interact, in the most general sense, following a perturbation. The first step is presumably to understand the functions of its constituents at the cellular level, aiming then to integrate understandings at increasingly higher levels of organization, so as to ultimately develop a predictive capability for the living organism.

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington.

Microorganisms are the logical initial focus of many studies due to their relative chemical complexity and, realistically, are likely to be the only systems truly understandable and successfully modeled over the coming decade or so in a fashion that broadly (but perhaps not comprehensively) links the molecular and the cellular levels. Reaching this goal will require a global perspective in both modeling and experiment that allows the study of changes in cellular systems under different conditions, for example, to identify single genes or gene products most sensitive to perturbations or to establish a set of nodes most sensitive to multiple small perturbations needed to produce a specific systems-level response.

While methods to simultaneously assess the abundances of thousands of expressed genes at the mRNA level are now broadly applied (Adams, 1996; Velculescu et al., 1997) with more or less success, posttranscriptional processes play a major role in determining protein abundances and modification states, and protein abundances can show poor correlation with mRNA levels (Anderson, 1997; Gygi et al., 2000; Haynes et al., 1998). Thus, considerable attention is now focused on the proteome, the complement of proteins expressed by a particular cell, organism, or tissue at a given time or under a specific set of environmental conditions.

The currently existing proteome analysis capability is predominantly based upon protein separations using two-dimensional polyacrylamide gel electrophoresis (2D PAGE), which can resolve up to thousands of putative protein spots. Proteome coverage in 2D PAGE is problematic for proteins that have very high or low isoelectric points (approximately <3.5 and >9.5), and membrane proteins (due to solubility issues during sample processing), which typically account for more than half of all proteins. It has been shown that the number of spots is poorly correlated with the number of different proteins detected, which are predominantly of high abundance based on their codon bias (Gygi et al., 2000a). Furthermore, a single gene can give rise to multiple spots (Gygi et al., 2000) due to co- and posttranslational modifications, degradation intermediates, and alternative expression (e.g., alternative splicing of mRNAs, translational frame shifts). Conventionally, protein identification after 2D PAGE involves the separate extraction, digestion, and analysis of each spot using mass spectrometry (MS; Shevchenko et al., 1996a; Wilm et al., 1996; Yates et al., 1993), but remains lacking in proteome coverage, sensitivity, dynamic range, throughput, and the precision needed to discern small (but often biologically important) changes in protein abundances. The sensitivity of 2D PAGE is generally limited to femtomole levels (Shevchenko et al., 1996b; Wilm et al., 1996) by the need to visualize the protein spot on the gel and its subsequent processing and analysis. The largest study reported to date identified 502 proteins from *Haemophilus influenzae* (Langen et al., 2000). Similarly, the most comprehensive yeast proteome 2D PAGE/MS studies published to date (the broadest of which identified 279 proteins [Perrot 1999] and a combined total of only ~500 [Futcher et al., 1999; Garrels et al., 1997; Gygi et al., 1999b; Perrot, 1999; Shevchenko et al., 1996a]) provide a skewed codon bias distribution (Kitayama and Matsuyama, 1971), indicating that only more abundant proteins were detected. Many important regulatory proteins are expressed at such low levels (e.g., <1,000 copies per cell) that their detection is precluded unless 2D PAGE is preceded by extensive fractionation of large quantities of protein and/or the processing of a large number of gels. Finally, the precision of protein abundance determinations using 2D PAGE is based on comparison of protein spot intensities, limiting the capability for discerning subtle differences in protein abundances for large numbers of proteome-wide measurements (e.g., from time course studies).

Many of the possible alternative proteomics technologies presently being considered employ a liquid phase separation methodology combined with some form of MS, most typically applied after protein digestion using specific proteases (e.g., trypsin). Analysis of the polypeptides of size sufficiently large for protein identification, typically approximately >5–10-mer size based upon sequence uniqueness for a specific organism, is now effectively achieved by MS (Henzel et al., 1993; James et al., 1993; Mann et al., 1993; Pappin et al., 1993). A widely used approach involves MS selection of a polypeptide that is then dissociated to form fragments, the mass-to-charge (*m/z*) ratios of which are measured. This MS/MS analysis provides primary sequence-related information that allows the peptide (and most often its parent protein) to be identified (Yates et al., 1996). MS/MS analysis of only one polypeptide is often sufficient for protein identification (Ducret et al., 1998; Link et al., 1997; McCormack et al., 1997; Yates, 1998; Yates et al., 1996). Very recently, Washburn et al. (2001) demonstrated the use of this approach to identify 1,484 yeast proteins from three different protein fractions by a 2D capillary liquid chromatography (LC)–MS/MS strategy, in which peptides were separated using cation exchange LC in the first dimension into 15 fractions, which were subsequently separated by reverse-phase LC, for a total of >25 hrs for analysis of each fraction. This work demonstrated the potential for broad proteome coverage based upon the analysis of highly complex polypeptide mixtures. However, this approach leaves much to be desired in terms of speed, sensitivity, comprehensiveness, confidence of protein identifications, and the quantitative utility of the measurement method.

Here we describe and review progress towards a global proteomics strategy that aims to provide large improvements in sensitivity, dynamic range, comprehensiveness, and throughput based upon the use of polypeptide accurate mass tags (AMTs). The two-stage strategy exploits a single high-resolution capillary LC separation combined with Fourier transform ion cyclotron resonance mass spectrometry (FTICR) MS to validate polypeptide AMTs for a specific organism, tissue, or cell type (and often generated from potential mass tags from MS/MS analyses), and provides the basis for second-stage high-throughput studies using only AMTs obtained using FTICR. Key attractions of the approach include the feasibility of completely automated high-confidence protein identification, extensive proteome coverage, and the capability for exploiting stable-isotope labeling methods for high-precision abundance measurements.

MATERIALS AND METHODS

Sample processing

Yeast and *Deinococcus radiodurans* were cultured in appropriate media, and harvested by centrifugation at 10,000*g* at 4°C. *Deinococcus* R1 was cultured on a TGY media containing 0.5% tryptone, 0.3% yeast extract, and 0.1% glucose. For the extraction of proteins after harvest, Deinococcal cells were washed three times with PBS and then lysed by bead beating using three 1-min cycles, allowing a 5-min cool down on ice between cycles. Similarly, yeast cells were resuspended and washed three times with 100 mM ammonium bicarbonate and 5 mM EDTA (pH 8.4), and then lysed by bead beating using three 1-min cycles, allowing a 5-min cool down on ice between cycles.

Prior to LC/MS analysis, the protein samples were denatured and reduced by the addition of guanidine-HCl (6 M) and dithiothreitol (1 mM), respectively, boiled for 5 min and digested using sequencing grade modified trypsin (Promega, Madison, WI; trypsin/protein, 1:50, w/w) at 37°C for 16 h. MS studies (not shown) indicated that digestion was complete for most detected proteins based upon trypsin cleavage specificity (i.e., C-terminal of Lys and Arg residues), although some proteins displayed up to four missed cleavages, consistent with their greater stability. Such incomplete digestion products were explicitly included in the data analysis and did not significantly affect results.

High-efficiency capillary LC-FTICR MS of protein digests

We used both 5,000 and 10,000 psi reverse phase packed capillary (150 μm i.d. × 360 μm o.d. fused silica, Polymicro Technologies, Phoenix, AZ; 5 μm C18, 300-Å pores, Phenomenex, Torrance, CA) LC separations to obtain peak capacities of up to 1,000, significantly increasing the effective dynamic range and sensitivity of MS measurements (Shen et al., 2001b). The 11.4-tesla FTICR MS developed at our laboratory used an electrospray ionization (ESI) interface comprised of a heated metal capillary inlet, an electrodynamic ion funnel (Belov et al., 2000a; Kim et al., 2000), and three radiofrequency quadrupoles for collisional ion focusing and highly efficient ion accumulation and transport to the cylindrical ICR cell for analysis (Belov et al., 2000b, 2001c). FTICR simultaneously provides high sensitivity, resolution, dynamic range, and mass measurement accuracy (MMA) for large numbers of peptides (Jensen et al., 1999; Marshall et al., 1998; Shen et al., 2001b). Mass spectra were acquired with $\sim 10^5$ resolution. To obtain the desired 1 ppm MMA, a program that uses the multiple charge states (e.g., 2^+, 3^+) produced by ESI for many protonated polypeptides (Bruce et al., 2000) was applied, followed by the use of lock masses (i.e., confidently known species that serve as effective internal calibrants) for each spectrum derived from commonly occurring polypeptides that were identified with high confidence from a single capillary LC separation (for each organism) using FTICR MS/MS accurate mass measurements. In addition to the MS/MS measurements using FTICR, large numbers of polypeptides were tentatively identified from *D. radiodurans* using multiple capillary LC separations with a conventional ion trap MS (LCQ, ThermoFinnigan Corp.) using data-dependent MS/MS analyses and the SEQUEST identification program (Yates et al., 1995) searching against a genome sequence–derived data base. A multi-run MS/MS strategy (Spahr et al., 2000) was used to segment mass to charge (*m/z*) ranges and increase the number of peptide potential mass tags (PMTs), using the search/identification program SEQUEST and a minimum cross correlation score of 2. These analyses identified large numbers of polypeptide PMTs, of which ~70% were then validated as AMTs based upon the detection of a species having the predicted mass for the PMT to <1 ppm at the corresponding elution time in the FTICR analysis. This automated process increases the confidence for polypeptide identifications and allows the validated AMTs to be used in subsequent experiments without the need for MS/MS.

Quantitation based on stable-isotope labeling

For hydrogen peroxide stress studies, *D. radiodurans* was cultured on TGY media until mid-log phase and H_2O_2 was added to a final concentration of 60 uM, incubated for 2 h, and harvested. Cells cultured in ^{15}N-labeled media (Bioexpress, Cambridge Isotopes; Conrads et al., 2000a) were harvested at mid-log phase. ^{15}N-cultured and H_2O_2 (^{14}N)–stressed *D. radiodurans* were mixed, processed, and analyzed as described above. The two versions of each peptide AMT, differing in mass by the number of nitrogen atoms, allowed the pair to be identified in an automated fashion with high confidence. ^{14}N /^{15}N-metabolic labeling was also combined with a commercially available Cys-affinity tag, iodoacetyl-PEO-biotin, to derivatize and isolate Cys-polypeptides, similar to the ICAT approach described by Gygi et al. (1999a). Iodoacetyl-PEO-biotin was added to an estimated fivefold excess of the number of Cys residues and incubated while stirring for 90 min in the dark. The iodoacetyl-PEO-biotin–derivatized sample was desalted into 100 mM NH_4HCO_3, 5 mM EDTA, pH 8.4, and digested with trypsin (1:50 enzyme/protein ratio) overnight at 37°C. After boiling for 5 min, the samples were loaded onto an avidin column and washed with five bed volumes of 50 mM NH_4HCO_3 (pH 8.4). The bound Cys-polypeptides were eluted using 30% acetonitrile with 0.4% trifluoroacetic acid.

Data-dependent FTICR multiplexed-MS/MS

Data-dependent control was accomplished using the Odyssey data station coupled with an ancillary PC. The Odyssey data station was used to provide the experiment scripts (i.e., to control and generate all of the timing signals, potentials, and the transistor-transistor-logic [TTL] trigger signals). The combined capillary LC-FTICR experiment consisted of continuous repetition of two acquisitions: MS followed by MS/MS. During the MS acquisition, the PC was triggered to acquire raw data in parallel with the data station via a National Instruments 6070E Analog IO card. The ICR-2LS program converted the time domain raw data to an *m/z* spectrum during the experiment and rapidly generated an appropriate ion isolation waveform and ion excitation waveform, which were then downloaded to a National Instruments DAQ 5411 arbitrary waveform generator. Prior to the subsequent MS/MS acquisition, this arbitrary waveform generator was triggered by a TTL signal from the data station to isolate the selected precursor ions and (optionally) activate them during the CID step. After the acquisition, ICR-2LS processed the data based upon a series of user predefined processing steps. A log file, containing a list of selected parent ion masses and excitation parameters, was created for each LC-FTICR MS/MS analysis. The dynamic exclusion option (a user-defined signal intensity exclusion threshold) was used to prevent reacquisition of tandem mass spectra of ions for which an MS/MS spectrum has already been acquired during a given chromatographic time period (e.g., 1 min). This functionality allows for maximization of the number of peptides that can be identified in a single LC run by reducing the redundant fragmentation, thus further enhancing the throughput for multiplexed-MS/MS analysis. For multiplexed-MS/MS experiments, in addition to the generation of the waveform for isolation of multiple parent ions, multiple frequency irradiation waveforms were generated for dissociation as a superposition of the individual excitation waveforms.

FTICR MS/MS data analysis and database searching

Analysis of the large data sets arising from capillary RPLC FTICR experiments was performed in an automated fashion using ICR-2LS. Time domain signals were apodized (Hanning) and zero-filled twice before fast Fourier transform to produce mass spectra. Isotopic distributions and charge states in the mass spectra were deconvolved. The accurate molecular weight values of the parent ions and of the fragments were used for the database search. The *D. radiodurans*–predicted protein database was derived from genome sequence data (downloaded from ftp://ftp.tigr.org/pub/data/d_radiodurans/). The measured mass for each parent species was then searched against all masses for species on this list of possible tryptic digestion products, resulting in a set of candidates for each parent species. The subsequent search for the MS/MS data was performed using only a set of predicted fragment ion species originating from the possible candidate peptides. The list of possible fragment species for the candidates included all of the "b," "y," and "a" mode ion fragments as well as product ions corresponding to the loss of water or ammonia from these fragments. For all candidates, possible fragment masses were computed and compared to the list of masses from the LC multiplexed-MS/MS data.

RESULTS AND DISCUSSION

Our strategy for proteome analysis is based upon a combination of instrumental and methodological advances that provides broad coverage, high sensitivity, and the capability for greatly increased throughput compared with conventional technologies. After initial cell lysis, the recovered proteins are cleaved into polypeptide fragments (e.g., using trypsin) to produce tens to hundreds of potentially detectable polypeptides from each protein, and perhaps 10^5 to $>10^6$ in total (depending upon proteome complexity and the dynamic range of the measurements). This complex peptide mixture is then analyzed by combined high-resolution capillary LC-FTICR mass spectrometry. Variations on sample preparation that we have explored, among the many possible, to date include the use of Cysteine-labeling that incorporates a biotin affinity tag for isolation of the Cys-polypeptides using immobilized avidin column chromatography (Gygi et al., 2000; Conrads et al., 2001), and the incorporation of stable-isotope labeling, either by culturing in ^{15}N-labeled media or in conjunction with the Cys-peptide labeling (Conrads et al., 2001). The capillary LC-FTICR analysis can be preceded by additional sample fractionation to both simplify the analysis and potentially provide additional information on peptide composition (e.g., as in the use of ion exchange chromatography), thus allowing more complex proteomes to be studied in greater detail. However, our initial aim has been to optimize proteome coverage for the use of a single separation step so as to increase overall throughput. Additionally, every additional separation stage irrevocably leads to peptide losses and increases overall sample requirements.

Capillary LC-FTICR measurements of complex global protein digests

The extent of proteome coverage depends substantially on the achievable dynamic range of the MS measurements, which in turn depends significantly upon the resolution (or peak capacity) of the separation steps preceding MS analyses. The number of potentially detectable polypeptides from LC-FTICR measurements is the product of the capillary LC peak capacity (which can be as high as 1,000 in a one-dimensional separation [Shen et al., 2001b]) and the maximum number of peptides detectable per spectrum. In practice, the obtainable peak capacity (in terms of the number of measurable peaks) for a single FTICR mass spectrum is almost always limited by FTICR trap charge (ion) capacity and the dynamic range of each spectrum (to $\sim 10^3$) rather than by the MS resolution, but can potentially be as high as $\sim 10^5$ based upon the charge capacity of the FTICR trap ($\sim 10^7$ for our 11.4-tesla FTICR) and the minimum number of charges that give rise to a detectable signal ($\sim$30 with $S/N>3$) when the discrete isotopic peaks present for each species are considered. We have measured the masses of hundreds of peptides in a single mass spectrum during LC-FTICR analyses. The average number of polypeptides that can potentially be detected at any point in a separation can be further increased by improvements in experimental dynamic range, as described below.

Thus, the theoretical number of detectable species of the combined LC-FTICR approach is ~10^6 and can potentially exceed 10^7. It is worth noting, however, that the number of *distinguishable* species in each spectrum is a function of resolution, mass measurement accuracy (MMA) and the MS spectral space, and exceeds 10^6 due to the high resolution obtainable. Although difficult to quantify at the present time, it is obvious that the more the number of distinguishable species exceeds the number of detected species, the better will be the level of confidence in the identifications that results. The number of distinguishable species in the combined LC-FTICR analysis is presently $\sim 10^7$ and could potentially exceed 10^8 if precise elution time information could be effectively used.

To evaluate the complexity of polypeptide mixtures that can be addressed, we used trypsin to digest the soluble proteins from yeast grown to mid-log phase. A 10-microgram sample was separated by a gradient reverse phase LC separation in a 85-cm long capillary packed with C-18 bonded 3-μm particles that achieved a peak capacity of ~1,000. The online ESI-FTICR analysis consisted of ~1,200 high-resolution MS, or an average of approximately two spectra for the narrowest LC peaks. Figure 1 shows a capillary LC-FTICR total ion current (TIC) chromatogram reconstructed from the ESI-FTICR mass spectra where data collection began 30 min after sample injection, and the separation was completed in ~3 h. A high-efficiency separation with symmetric peaks was obtained throughout the elution using the conditions optimized for ESI-MS, and where many minor (low abundance) species were resolved from their neighboring major (high abundance) components. This high-efficiency separation was obtained using a simple connection of a replaceable emitter (electrospray tip) to the column outlet through a narrow bore (150 μm × 1 mm channel) union, which also allowed convenient replacement of ESI emitters. The excellent peak shapes and the sensitive MS detection (>100,000 detected putative polypeptides) support our belief that proper selection of the mobile phases is desirable for maintaining both separation and ESI efficiency.

Examples illustrating the qualities of the separation and the mass spectrometry are shown in Figure 2. Figure 2A shows a typical single spectrum with insets showing exploded views of several regions of the spectrum. Figure 2B shows a very narrow range (*m/z* 972.515–972.535) reconstructed ion chromatogram, where a number of both high- and low-abundance peaks eluted in this small *m/z* window during the separation. The power of the approach, however, is that this quality of information is obtained over a wide *m/z* range. Figure 2B also demonstrates the excellent peak shapes typically obtained. High-quality separations are of enormous importance for proteomics, since the detection methodology always places the ultimate limitations on the sample complexity and/or dynamic range that can be addressed. More abundant polypeptides were typically observed to elute over three to five spectra (one acquisition requires 5.7 sec), while minor components were observed to elute over only one to two spectra. Using an average peak width at the base of 25 sec, the chromatographic peak capacity (for a resolution of unity) corresponds to ~650 (1.5 × 180 × 60/25) under our ESI-FTICR analysis conditions. Using a shorter spectrum acquisition time of 2.5 seconds, a chromatographic peak capacity of ~1,000 has been achieved (Shen et al., 2001b). Figure 3 shows the typical data quality for a portion of the reconstructed LC chromatogram for a 3-*m/z* unit range and the mass spectra for the elution times of several peaks.

The combined effective resolving power supplied by the capillary LC-FTICR 2D separation is critical when utilizing accurately measured masses (e.g., a mass measurement accuracy of <1 ppm) to directly identify polypeptides. Extremely high resolving power of this type of 2D separation can be achieved using the combination of high-efficiency capillary LC with high-resolution FTICR. Due to the orthogonal relationship between capillary RPLC and FTICR, the combined peak capacity of 2D capillary LC-FTICR analyses was estimated at 6×10^7, providing the highest 2D separation capability of any technique yet reported (Shen et al., 2001a,b).

Figure 4 shows the detected polypeptide masses as discrete spots; however, neither the accuracy of the mass measurements nor peak intensities can be adequately conveyed by such a figure. More than 110,000 different species were detected in this analysis. If one assumes an average of 4 sec per MS/MS spectrum, at least 180 h would be required for characterization of the detected species, a factor of >30 difference, corresponding to the increased throughput provided by the AMT strategy.

It must be noted that although FTICR can supply a resolving power of $\sim 10^5$ species, this power by itself is not sufficient to resolve the extremely complex mixtures of cellular polypeptides. For example, >6,000 proteins are predicted to be potentially expressed by the yeast genome, which can yield >350,000

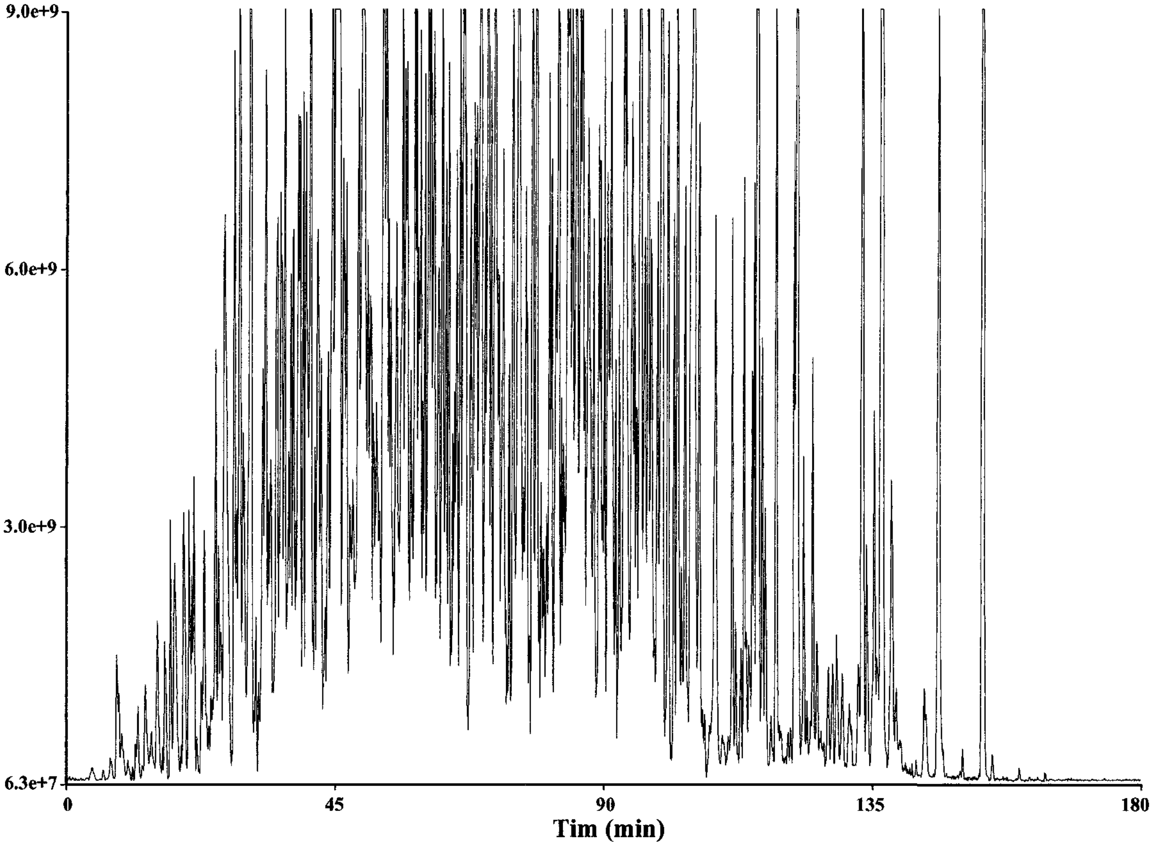

FIG. 1. Total ion current (TIC) chromatogram of capillary LC-FTICR of a global soluble yeast tryptic digest using the multiple-capillary LC system. The separation used a pressure of 10,000 psi and a mobile phase gradient from A (H_2O, 0.2% HOAc, 0.1% TFA, v/v) to 75% B (H_2O/ACN, 10:90, 0.2% HOAc, 0.1% TFA, v/v) over 180 min. The vertical axis is proportional to the ion signal.

different tryptic polypeptides, and ~195,000 having masses of 500–5,000. Even if only ~20% of the ~6,000 proteins are expressed under a specific set of conditions, an ideal tryptic digestion would conceivably yield ~40,000 different polypeptides, and a much larger number if modified and incompletely digested polypeptides are also considered. Many polypeptides will yield multiple charge states (e.g., 2^+, 3^+) following ESI, with each charge state comprising multiple isotopic peaks. Complexity of this type is illustrated in Figure 2B, which shows more than 20 peaks evident in a very narrow *m/z* range (0.02 Dalton) and having apparent LC retention factor differences of as little as 0.006, and that would be difficult to resolve using low-efficiency separations. Such complexity can result in the need for even greater FTICR resolving power and/or higher-efficiency separations prior to FTICR. While FTICR resolution can be increased, typically at the cost of an increase in the spectrum acquisition time or magnetic field strength, along with a more significant increase in the data storage requirements, the ion trap capacity issues mentioned above still impose the greatest limits on overall dynamic range. In practice, the need to address greater levels of complexity will strongly depend on the achievable dynamic range of proteome measurements, and will likely involve the use of additional separation stages.

The dynamic range of capillary LC-FTICR mass spectrometry analyses

As shown by the example in Figure 5, the dynamic range obtainable in a single FTICR mass spectrum exceeds 10^3. The most highly abundant polypeptide eluted over 13 spectra, while low-abundance polypep-

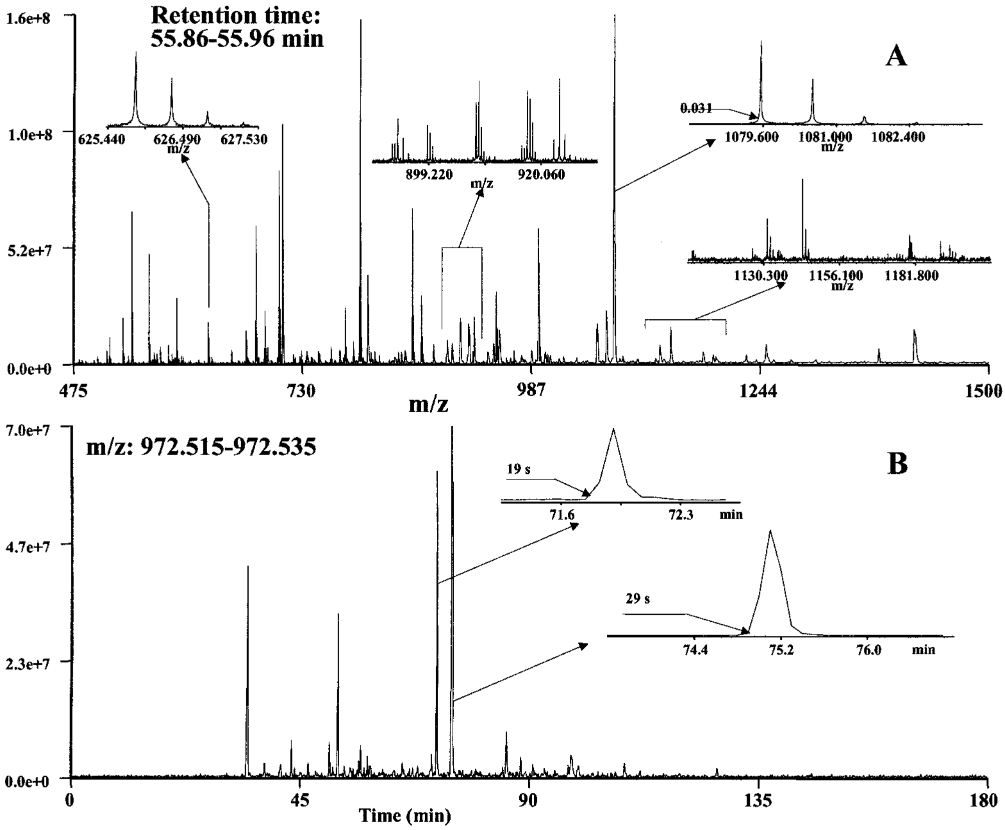

FIG. 2. Examples demonstrating the high resolution of the capillary LC and 11.5 tesla ESI-FTICR. Conditions given in Figure 1.

tides often elute over only a single spectrum. Therefore, the effective dynamic range for detection of polypeptides can approach ~10^4 if they have the same ionization or detection efficiency, just on this basis. Furthermore, if one's aim is protein identification, then a significant (perhaps 10-fold) increase in effective dynamic range will result due to the variable electrospray ionization or detection efficiency for different polypeptide sequences. This variation in overall detection efficiency is evident in the analysis of unseparated tryptic digest mixtures where polypeptide fragments, for many possible reasons, differ greatly in their signal intensities compared to their nominally expected equimolar abundances. More important, however, is the ion accumulation process used with FTICR trap. Ion introduction from ESI, transfer through the ion funnel, and into the external ion accumulation trap region are more efficient when ion production rates are lower since repulsive space charge effects are lower. Although we have not yet quantified this effect, it is clear that achievable sensitivity is greatly improved for low abundance peaks that are chromatographically separated from high-abundance peaks. This contribution likely accounts for at least an order of magnitude increase in dynamic range, and can potentially be much greater if ion bias effects that result from overfilling of the ion accumulation region can be mitigated (Belov et al., 2001b,d). Upon consideration of these factors, we estimate that the dynamic range currently achieved is approximately 10^4 to 10^5 (Shen et al., 2001a), and believe that an order of magnitude further gain is achievable if unbiased ion accumulation can be achieved. In this regard, we have previously noted that continuous external accumulation of ions between transfers to the FTICR trap would increase the ion accumulation times at least fivefold, potentially providing an equivalent gain in dynamic range if the quadrupole accumulation region was not over-filled, and could potentially increase the dynamic range of FTICR analysis to as much as 10^6. Of course, many issues (e.g., contaminants, background signals from low levels of ion dissociation) can prevent this from being realized.

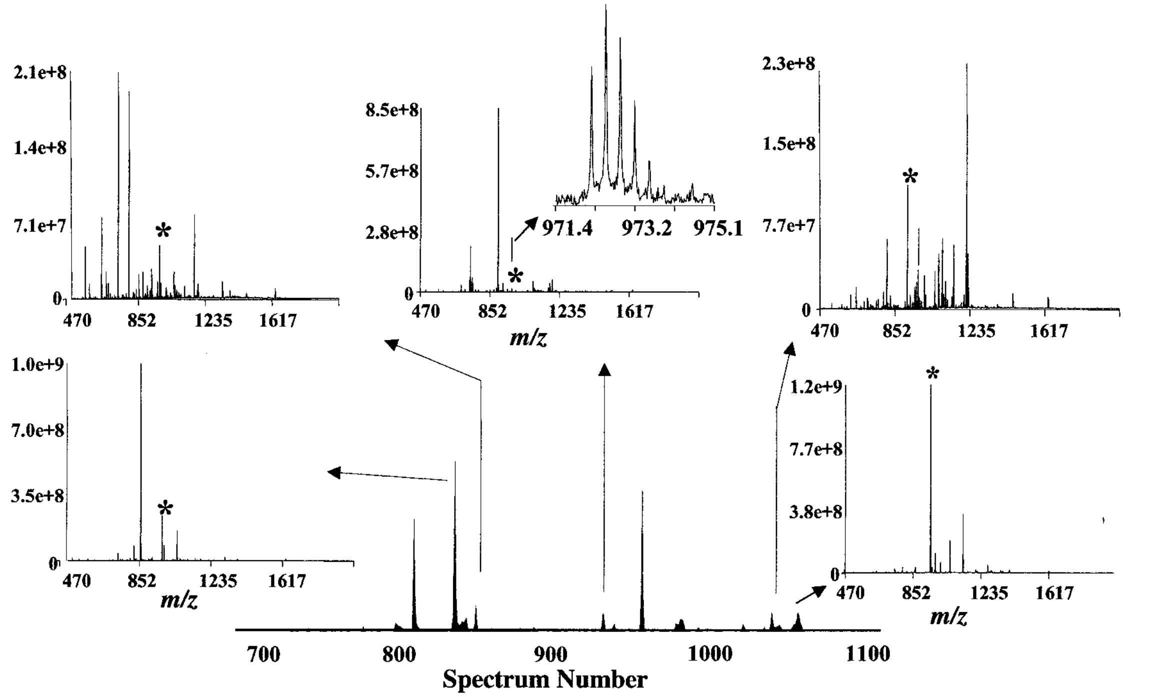

FIG. 3. Narrow range (*m/z* 970.0–973.9) reconstructed partial chromatogram (bottom center) showing several polypeptide peaks and representative mass spectra for the indicated elution times illustrating the dynamic range and typical data quality obtained for a capillary LC-FTICR analysis of tryptic polypeptides from a digestion of soluble yeast proteins. The LC peak from the narrow *m/z* range is indicated by "*" in each spectrum. Note that the relatively small peak at spectrum number ~940 (top, middle) provides high resolution MS results for the lower level component (inset).

We have directed considerable effort to achieving an extended dynamic range for proteome measurements. As can be seen from the above discussion, the dynamic range of a single FTICR mass spectrum is limited by the charge capacities of both the external ion accumulation region and of the FTICR analyzer trap. As noted earlier, the useful charge capacity of the external accumulation quadrupole trap is on the order of 10^7 charges if undesirable effects due to overfilling are to be avoided. These combined effects include bias due to charge stratification in the accumulation quadrupole and coalescence of closely spaced *m/z* ion packets in the FTICR trap. Improvements in the ESI source design and use of an electrodynamic ion funnel now allow currents of >1 nA of analytically useful ions to be transmitted to the ion accumulation region. This corresponds to ~10^{10} charges/sec, a factor of ~10^3 in excess of the ion population than can currently be analyzed by FTICR in a single spectrum even if only a 20% transfer efficiency is assumed with the present 11.4-tesla magnetic field instrument. Our efforts have thus been directed towards establishing a routinely useful active dynamic range enhancement capability, in which the information from a proceeding spectrum is used to remove the high abundance species in an RF-only quadrupole just prior to the ion accumulation quadrupole region. In this fashion, every other spectrum would dig deeper into the proteome and provide more information on lower abundance species. While the overall dynamic range potentially achievable with this approach should significantly exceed 10^6, its practical utility remains to be fully demonstrated.

DREAMS FTICR mass spectrometry for expanded dynamic range proteome measurements

The range of peptide (or protein) concentrations of interest in proteomic measurements can vary more than six orders of magnitude and include >10^5 components. When analyzed in conjunction with capillary LC separations, both the total ion production rate from ESI and the complexity of the mixture at any point can vary by more than two orders of magnitude. This temporal variation in ion production rate and spec-

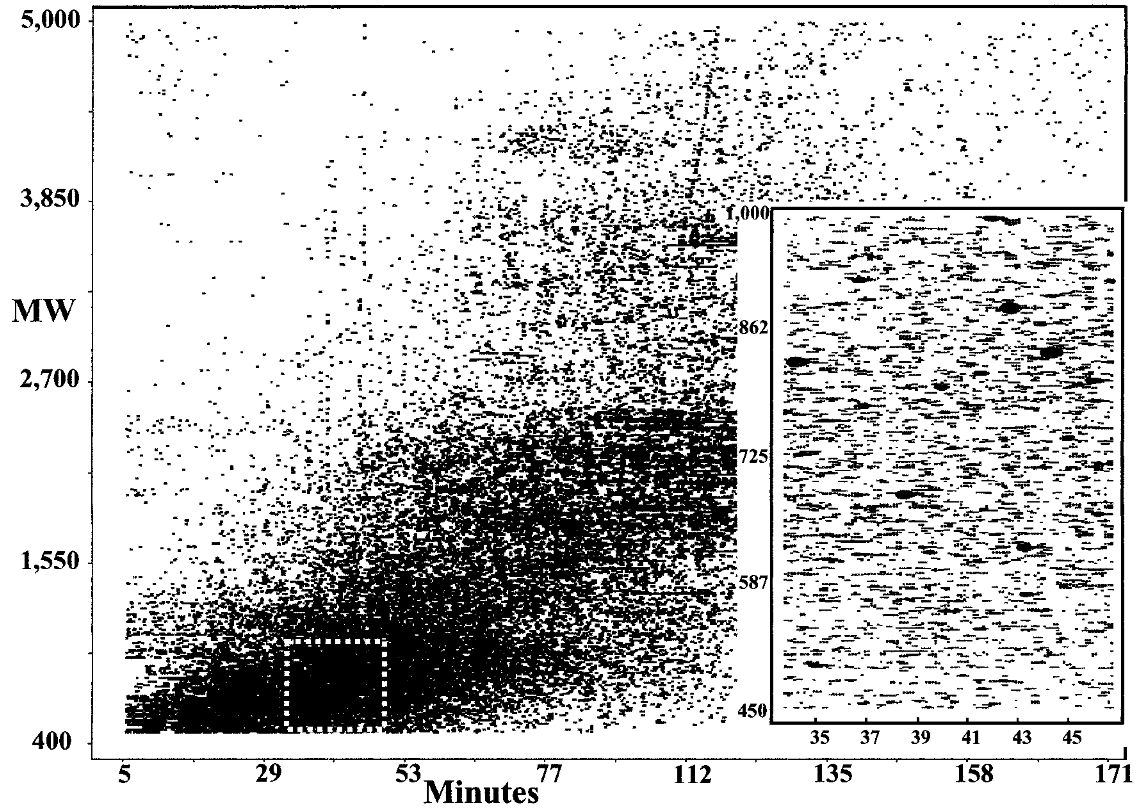

FIG. 4. Two-dimensional display of a capillary LC-FTICR analysis in which >110,000 putative polypeptides were detected form a tryptic digest of soluble yeast proteins. The inset shows detail for a limited mass and time segment from a region of high density indicated by the dashed box. Within the inset, spot size has been adjusted to show highly abundant species as larger spots.

tral complexity constitutes a major challenge for proteome analyses. For example, the elution of highly abundant peptides can restrict the detection of lower-level coeluting peptides, since the dynamic range presently achieved in a single spectrum is $\sim 10^3$. If the ion accumulation time is optimized for the most abundant peaks, the accumulation trap will not be filled to capacity during the elution of lower abundance components, and the overall experimental dynamic range will be significantly constrained. If, however, longer accumulation times are used, the conditions conventionally used result in an overfilling of the external accumulation trap in many cases, which will be manifested by biased accumulation or extensive activation and dissociation. Thus, we have attempted to develop methods that avoid or minimize the undesired artifacts associated with overfilling the external accumulation trap, and thus effectively expand the dynamic range of measurements (Belov et al., 2001b).

A more generally useful approach to dynamic range expansion involves the use of ion ejection from a linear quadrupole device external to the FTICR that is accomplished by resonant rf-only dipolar excitation (Belov et al., 2001a). The effective removal of the major species from a spectrum allows the lower abundance species to be accumulated for extended periods, resulting in an increase in the dynamic range. This dynamic range enhancement applied to mass spectrometry (DREAMS) approach thus provides the basis for a significant gain in the coverage of proteomic measurements.

The DREAMS methodology involves acquisition of sets of mass spectra during the non-selective accumulation, in which each spectrum is followed by software-controlled selective rf-only ejection of the most abundant species prior to external accumulation (for the next spectrum immediately following the non-selective normal spectrum). We initially evaluated the data-dependent selective external ion ejection with a

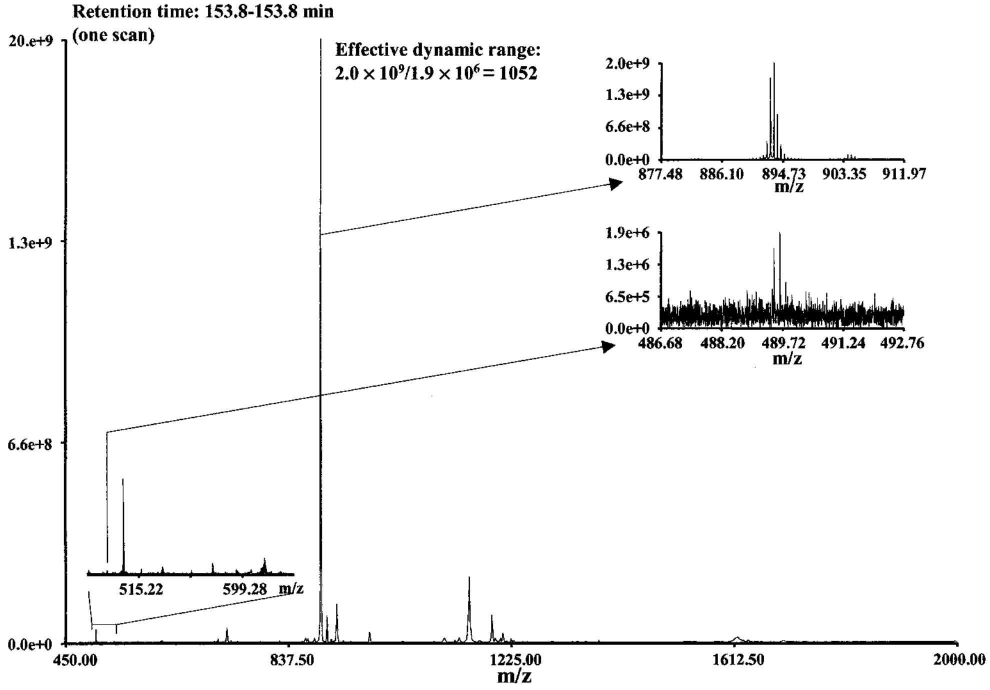

FIG. 5. Effective dynamic range of a single FTICR mass spectrum from a capillary LC-FTICR analysis of a global yeast soluble protein tryptic digest obtained under capillary LC data acquisition conditions.

mixture of peptides and then applied the DREAMS approach for the characterization of a global yeast proteome tryptic digest (Belov et al., 2001a). Figure 6 shows typical mass spectra acquired with these two alternating sequences. The most abundant species detected in the first acquisition were selectively ejected on the fly in the selection quadrupole prior to trapping the lower abundance species in the accumulation quadrupole for the subsequent spectrum acquisition (Fig. 6B). Removing the most abundant ion species in the selection quadrupole thus allowed accumulation of lower abundance species (not evident in the mass spectrum in Fig. 6A). Following the selective ejection of the most abundant species, a nonselective accumulation mass spectrum was again acquired (Fig. 6C), primarily showing the same species as in Figure 6A. This indicates that, in this case, the peptide with a monoisotopic mass of 1098.75 continued eluting from the LC column and was still the primary contributor to space charge effects in the accumulation quadrupole. Examination of the mass spectra acquired using the automated DREAMS data-dependent selective external ion accumulation showed that the experimental mass resolution during actual LC separation for rf-only ion ejection from the selection quadrupole was in the range of 30–50, depending on *m/z*.

In the initial demonstration of the DREAMS, the FTICR method generated two data sets comprising spectra detected after the nonselective and selective DREAMS accumulations. In order to evaluate our approach these data sets were processed and compared with a data set acquired in a separate capillary LC-FTICR analysis using only the standard nonselective external ion accumulation method. Throughout the LC runs, the intensities of the most abundant ion species were found to vary by approximately two orders of magnitude (consistent with variation in a chromatogram obtained using UV detection). The results obtained for the two data sets acquired using alternating sequences in one LC-FTICR run (i.e., the nonselective and DREAMS-selective external ion trapping) were compared with the number of putative peptides identified in a separate LC run using the nonselective external ion accumulation. It was found that the number of peptides detected with the alternating sequences (30,771 after subtraction of species detected in both) was greater by about 35% than that acquired using the nonselective ion accumulation (22,664). The same methodology was subsequently applied with data-dependent selective ion ejection of the two and three most abundant ion species. A 40% increase in the number of peptides was achieved when combining the nonselec-

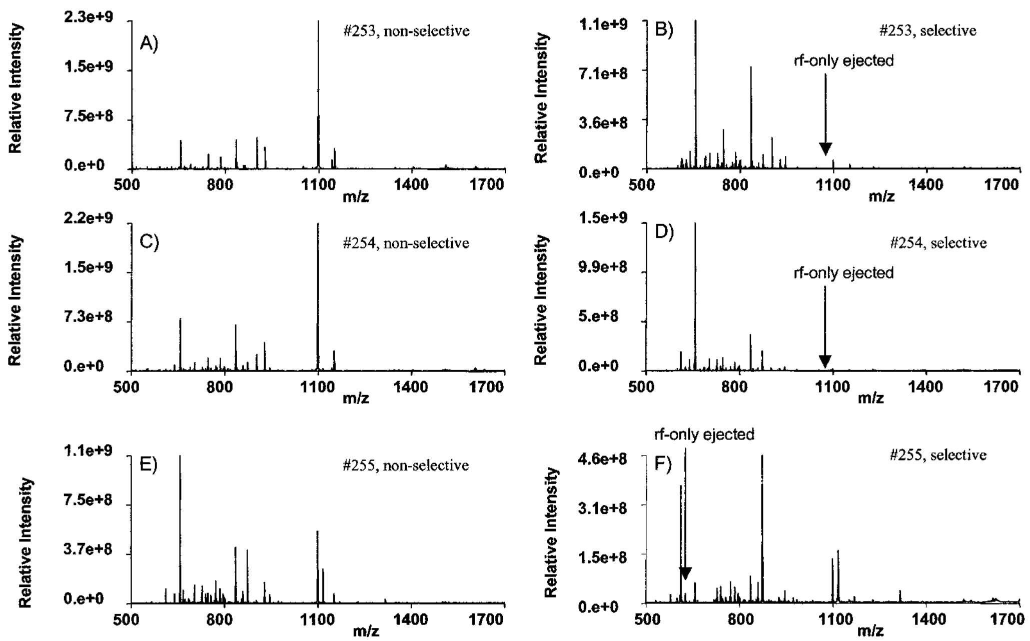

FIG. 6. Typical mass spectra obtained from a 1mg/mL soluble yeast proteome extract acquired using the data-dependent selective external ion accumulation for DREAMS FTICR mass spectrometry. (**A**) Nonselective ion accumulation, spectrum no. 253. (**B**) Selective ion accumulation, spectrum no. 253. (**C**) Nonselective ion accumulation, spectrum no. 254. (**D**) Selective ion accumulation, spectrum no. 254. (**E**) Nonselective ion accumulation, spectrum no. 255. (**F**) Selective ion accumulation, spectrum no. 255. The most abundant ion peak from the previous nonselective accumulation (e.g., *m/z* 1098.75 in A and C) was resonantly ejected on the fly through selection quadrupole using data-dependent rf-only dipolar excitation to yield the spectra immediately following each nonselective accumulation scan.

tive ion accumulation with data-dependent selective ion ejection of the three most abundant ion species (Belov et al., 2001a).

We have initially implemented the DREAMS approach with high-performance capillary reverse phase HPLC separations and high–magnetic field electrospray ionization FTICR mass spectrometry for obtaining more comprehensive quantitative measurements. The high-resolution FTICR mass spectrometric analysis allows the confident assignment of isotopically labeled peptide pairs (corresponding to the two distinctive versions of each peptide), and thus the basis for quantiative measurements when one of the two proteomes in the mixture are perturbed or altered in some fashion. In this work, we ejected up to the 10 most intense peaks in every other spectrum. In a subsequent analysis of a sample derived from a tryptic digest of proteins from mouse B16 cells cultured in both natural isotopic abundance and ^{15}N-labeled media, we showed that the DREAMS approach allowed assignment of approximately 80% more peptide pairs, providing quantitative information for approximately 18,000 peptide pairs in a single analysis. The observed 80% increase in the number of stable-isotope–labeled peptide pairs is attributed to the detection of lower-level peptides that would otherwise be missed in the current global proteomes analysis approach. The additional peptides detected from the DREAMS analysis represent high quality measurements from which useful quantitative information can be extracted.

Protein identification using AMTs

The power of MS for protein identification derives from the specificity of mass measurements for either the intact polypeptides or their fragments after dissociation from MS/MS measurements, and is implicitly based upon the relatively small number of possible polypeptide sequences for a specific organism compared to the total number of possible sequences (Table 1). MS/MS measurements provide typically partial sequence determination, and in relatively rare cases complete sequence information, and make a large frac-

TABLE 1. NUMBER OF POSSIBLE POLYPEPTIDES AND NUMBER PREDICTED FOR THREE ORGANISMS FROM DIGESTION WITH TRYPSIN

	Possible		*Predicted number of polypeptides*[c]		
Length	*Sequences*[a]	*Masses*[b]	*D. radiodurans*	*S. cerevisiæ*	*C. elegans*
10-mers	10^{13}	2×10^{7}	3,471	11,275	30,623
20-mers	10^{26}	7×10^{10}	1,292	3,463	9,475
30-mers	10^{39}	2×10^{13}	494	1,278	3,602
40-mers	10^{52}	1×10^{15}	195	405	1,295

[a] Assumes 20 possible distinguishable amino acid residues.

[b] The number of polypeptides of length *r* potentially distinguishable by mass based upon the number of possible combinations of *n* different amino acids:

$$\frac{(n + r - l)!}{r!\,(n - l)!}$$

The actual number of possible masses is somewhat smaller due to some mass degeneracy. The number of distinguishable polypeptides in actual measurements depends upon the MS resolution.

[c] Predicted from the identified open reading frames and applying the cleavage specificity of trypsin.

tion of the ~10^{26} possible 20-mer polypeptide sequences distinguishable and potentially identifiable. Thus, highly confident polypeptide identifications using MS/MS methods can often be achieved from only limited sequence data due to the enormously smaller numbers of polypeptides predicted for an organism; for example, only 3,463 different 20-mer polypeptides are predicted from an ideal tryptic digestion of all yeast proteins. The distinctiveness of polypeptide sequences increases with size, but in practice the utility of increased size for identification is mitigated by the increased likelihood that a peptide will be unpredictably modified. Indeed, we have found that exact MW measurements *alone* have very limited utility for the identification of intact proteins (Jensen et al., 1999).

Though much smaller than the number of possible sequences, the number of potentially distinguishable polypeptide *masses*, given sufficient resolution and accuracy, also dwarfs the number of predicted peptides from any organism. For example, the number of potentially distinguishable 30-mer polypeptide masses at high MMA, estimated from the number of possible combinations, is $>10^{13}$, compared to the 494, 1,278, and 3,602 predicted for tryptic digestion of all predicted proteins for *D. radiodurans*, *E. coli*, *S. cerevisiae* (yeast), and *C. elegans*, respectively. As shown in Table 2, an ideal tryptic digestion of all yeast proteins would produce 194,239 polypeptides having masses of 500–4,000 Da, the range typically studied by MS. Of these, 34% are unique at ±0.5 ppm MMA and can potentially serve as AMTs. (A larger fraction are useful as AMTs if constrained by additional information resulting from any prior sample fractionation steps or the use of LC elution times.) These distinctive polypeptide AMTs would cover 98% and 96.6% of all predicted *S. cerevisiae* and *C. elegans* proteins, respectively.

TABLE 2. PREDICTED NUMBER OF POLYPEPTIDES[a] FOR IDEAL GLOBAL TRYPTIC DIGESTIONS

Organism	*Peptides*[a]	*Unique*[b]	*ORF coverage*[c]	*Cys-peptides*[a]	*Unique*[b]	*ORF coverage*[c]
D. radiodurans	60,068	51.4%	99.4%	4,906	87.2%	66%
E. coli	84,162	48.6%	99.1%	11,487	83.6%	80%
S. cerevisiæ	194,239	33.9%	98%	27,483	72.7%	84%
C. elegans	527,863	20.9%	96.6%	108,848	52.5%	92%

[a] Peptides or Cys-peptides in mass range of 500–4,000 Da, assuming ideal trypsin cleavage specificity.

[b] Percent unique to ± 0.5 ppm (by mass not using elution time).

[c] Percent of ORFs (or predicted proteins) covered by unique peptides.

Thus, given sufficient MMA, a polypeptide mass measurement can often be confidently attributed to a single protein within the constraints provided by a single genome sequence and its predicted proteome (i.e., serve as an AMT). The limited MMA achievable with conventional MS technologies generally requires extensive separations (e.g., using 2D PAGE) or the use of MS/MS methods for protein identification. The AMT strategy obviates the routine need for MS/MS, and thus reduces sample requirements. Since the masses of many peptides will generally be obtained in each mass spectrum, requiring equivalent or less time than one MS/MS measurement, the increase in throughput is at the least equal to the average number of polypeptides in each spectrum. In practice, the increase in either throughput or proteome coverage is even greater since the lower abundance peptides are often not analyzed by conventional MS/MS approaches, or require the need for additional time for extended ion accumulation or spectrum averaging to yield spectra of sufficient quality. Thus, the AMT approach provides increased sensitivity, coverage and throughput, and facilitates quantitative studies involving many perturbations or time points.

The practical utility of high MMA in defining AMTs for global proteomic measurements is obviously dependent upon the complexity of the system being studied. As indicated above, if the universe of all possible peptides were to be considered, essentially nothing could be identified by mass spectrometry (as conventionally practiced) with high confidence. When the use of AMTs is confined to a specific (e.g., sequenced microbial) system, in which a much more constrained set of possible peptide masses exist, the situation becomes much more tractable. If the set of proteins sequences could be predicted with confidence, were not modified during or after translation, and were digested by proteolytic enzymes such as trypsin in an ideal fashion to cleave only certain sites and do so with 100% efficiency, the use of AMTs would be relatively straightforward. In the real world, however, the situation is intermediate between these extremes. First, if one has a fully sequenced microbe, the set of open reading frames (ORFs), and the resulting set of possible polypeptide masses from, for example, a tryptic digestion, can be predicted with *some reasonable facility*. This defines what can be considered the best-case scenario. At the opposite extreme, one can consider the range of possible masses that can result from all possible polypeptide cleavage sites, all possible modifications of these peptides, sequence variants (e.g., all possible single amino acid residue substitutions for each predicted peptide), contaminants. This situation similarly becomes intractable unless constrained in some fashion.

Figure 7 compares the uniqueness for polypeptides within the predicted proteome of *D. radiodurans* as a function of peptide molecular weight and for three different levels of mass measurement accuracy (1, 10, and 100 ppm). The calculations used the 3,187 proteins predicted from the DNA sequence by White et al. (1999) and assumes an ideal tryptic digestion involving protein cleavages only after lysine and arginine amino acid residues (Fig. 7A), and the case where all possible peptide fragments are considered (i.e., where cleavage at every amino acid residue is possible; Fig. 7B). The number of peptides having molecular masses of 500–4,000 is 60,068 for case A and 14,671,278 for case B. As shown in Figure 7A, a MMA of 1 ppm provides unique mass tags for a substantial fraction of the peptides generated for the ideal case. Measurements at 10 ppm MMA retain some utility, but a MMA of $>$100 ppm (a level more typical of conventional mass spectrometers) is of very limited value except for the largest peptides. In case B, where any peptide fragment is allowed, the utility of high MMA is largely negated (even for 1 ppm) due to the large set of possibilities. For a more realistic situation, however, where possible peptide fragments will also be formed from chymotryptic-type enzymatic activity, missed cleavage sites, and various posttranslation modifications, the MMA requirements are intermediate between cases A and B.

Thus, the two scenarios illustrated in Figure 7 span a range that extends from the extremes where the AMT approach would appear as either straightforward (A) to where it would be largely ineffective (B) for levels of performance that are currently achievable using mass spectrometry. The unknown extent of complexity beyond that suggested by Figure 7A needs to be addressed in some fashion. We do so by two additional aspects of our approach: the use of elution time information to increase the specificity of the measurements and the initial use of MS/MS measurements to initially validate the use of accurate mass and elution time data for subsequent studies.

The validation of AMTs from PMTs generated by MS/MS

The above analysis neglects the use of elution time information for the identification of peptides, and includes consideration of many possible peptides that will not ever be observed. The development of a ca-

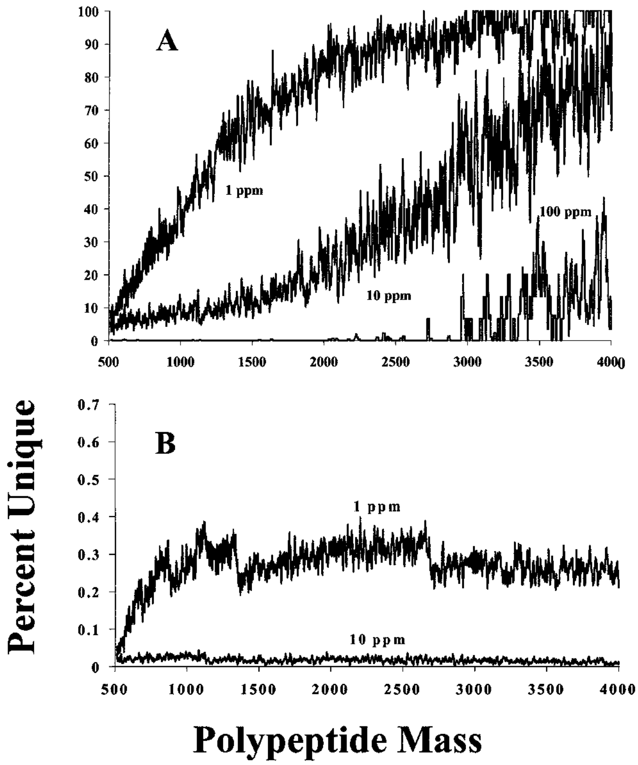

FIG. 7. The number of unique peptides as a function of molecular weight for three different levels of mass measurement accuracy for the microorganism *D. radiodurans*. (**A**) The case of a whole proteome ideal tryptic digestion. (**B**) Where peptides can be formed by cleavage at every amino acid (note the different scales). The real-world level of complexity needs to consider the many issues discussed in the text and thus the use of AMTs in our approach is augmented by the use of elution time data and the initial validation using tandem mass spectrometry.

pability for the prediction of peptide separation times would be an extremely valuable adjunct to the mass analysis, but a useful capability for this does not yet exist. Thus, our AMT strategy uses tandem MS/MS for the initial screening for AMTs as well as for the (initially limited at this point) identification of modified or otherwise unexpected peptides (that potentially arise due to sequence errors, frame shifts, ORFs missed by gene calling software). FTICR MS/MS measurements are also used to establish the initial set of highly confident polypeptide lock masses for FTICR spectrum calibration and for identification of peptides where higher resolution, mass accuracy or sensitivity are needed (e.g., the lowest level peptides). The generation of most AMTs by our approach presently uses a two-stage process (Fig. 8). The proteome sample is digested with trypsin and analyzed by high-efficiency capillary LC-tandem MS using either FTICR (using multiplexed MS/MS, as described below) or a conventional (LCQ ion trap or Q-TOF) mass spectrometer operating in a data-dependent mode. In this mode, a single MS scan is followed by three consecutive MS/MS analyses where three (or more) parent ions from a spectrum are sequentially (or simultaneously in the case of FTICR MS/MS) selected for analysis based upon predefined criteria designed to minimize repeated analysis of the same species. Due to the high MMA of the FTICR MS/MS measurements, identified peptides having distinctive masses are immediately assigned as AMTs, while the ion trap MS/MS measurements yields potential mass tags (PMTs) that are subsequently validated as AMTs if the predicted peptide's accurate mass is observed using FTICR in a corresponding sample and at an equivalent elution time (Fig. 8C).

Ion trap MS/MS generated PMTs are presently identified using scores produced by the SEQUEST search program based upon the similarity of the spectrum with a set of peaks predicted on the basis of the known most common peptide fragmentation processes. Due to the nature of the analysis, the results will invariably span the range from low scores where identifications are highly doubtful, to high scores where iden-

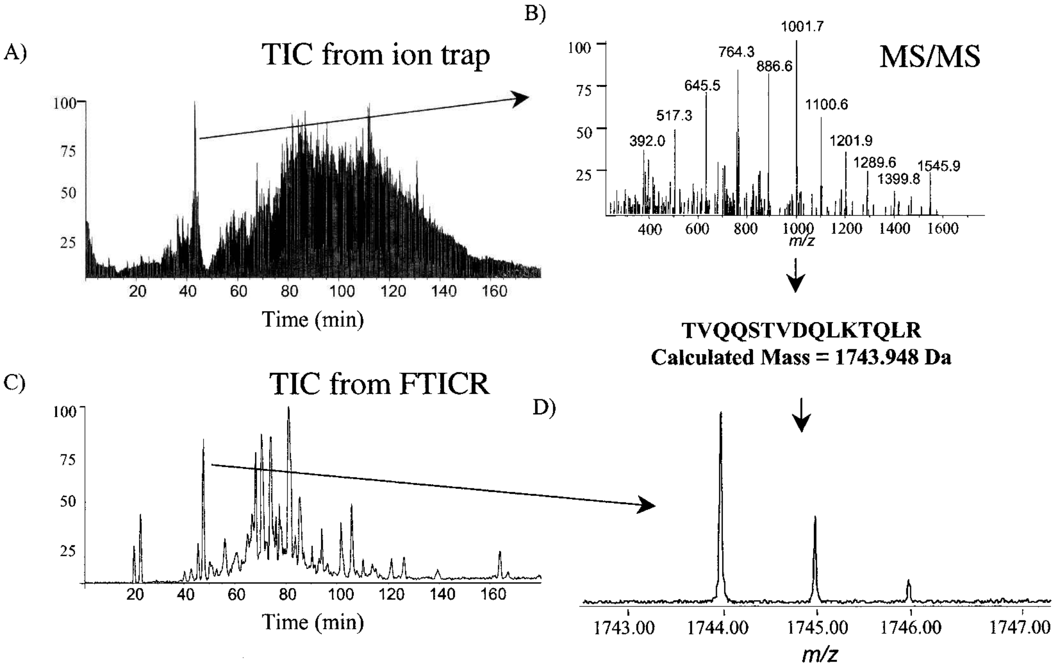

FIG. 8. Experimental steps involved in establishing an accurate mass tag (AMT). (**A**) In the first stage, a proteome sample is analyzed by capillary LC-MS using a conventional ion trap mass spectrometer. (**B**) Peptides are automatically selected for collisional induced dissociation (CID) and identified by the resulting sequence information as a potential mass tag (PMT). (**C**) In the second stage, the same proteome sample is analyzed under the same LC-MS conditions using a high-field FTICR mass spectrometer. (**D**) An AMT is established when a peptide eluting at the same time and corresponding to the calculated mass (within 1 ppm) of the PMT identified in the first stage is observed. This peptide then functions as a biomarker to identify this particular protein in all subsequent experiments analyzing a proteome sample from a specific organism.

tifications are quite reliable, with no clear line of demarcation. If one uses only the highest scores for identification, fewer proteins will be identified; uncritical use of lower scores will result in many false identifications. Conventionally, many MS/MS spectra and the program search results need to be manually examined for the less confident identifications to evaluate both spectrum quality and the ranking of peptide scores so as to establish acceptable confidence for identifications. This process generally results in discarding a substantial fraction of the peptides identified with lower scores, and serves to increase the confidence to an extent that is difficult to quantify. In our approach, the use of highly accurate mass measurements provides an additional and high-quality test for tentative peptide identifications that can be applied in the data analysis using software developed at our laboratory. Thus, an additional advantage of the automated validation of AMTs from PMTs is the increased confidence that results.

As mentioned above, promoting PMTs to validated AMTs also exploits the distinctive elution times for peptides. Although run-to-run variations in the gradient capillary LC separations limited the use of elution times to 10–20% in our initial work, better calibration procedures or flow control will likely reduce this to less than a few percent and significantly increase the utility of elution times, and the ability to exploit AMTs having otherwise indistinguishable masses. The use of the same lock masses that correspond to peptide AMTs would serve as LC elution time calibrants, allowing correction for any differences in mobile phase gradient and flow rate that may occur between separations.

In our initial studies with *D. radiodurans*, a peptide is validated as an AMT if its observed mass, as measured by FTICR, agrees with the theoretical calculated mass of the PMT within 1 ppm MMA. Additionally, the LC elution times in the experiment wherein the PMT was first identified has to agree with that observed in the FTICR experiment (Fig. 8D). Also, in the initial studies with *D. radiodurans*, we required that the peptide must be unique within the annotated genomic database (although distinctive elution time data should allow the designation of AMTs having otherwise indistinguishable masses). Only when these three criteria were met was a peptide designated as an AMT, and subsequently used to confidently identify a specific protein in subsequent proteome studies from the same species. Without the need to reestablish the identity of a peptide using time-consuming MS/MS analyses, multiple high-throughput studies to measure changes in relative protein abundances between two (or more) different proteomes can be completed in rapid fashion based solely on the highly accurate mass measurements provided by FTICR. Once a protein has been identified using AMTs, its subsequent identification (and quantitation) in other studies is based on FTICR measurements (and its elution time), which provide much greater sensitivity and throughput than conventional MS instrumentation.

An example of the generation of an AMT for the protein elongation factor Tu (EF-Tu) is shown in Figure 9. A tryptic digest of *D. radiodurans* proteins is analyzed by capillary LC-MS/MS using a conventional ion-trap mass spectrometer. In this example, the peptide selected for dissociation was identified as GITINTAHVEYQTETR from the protein EF-Tu and is considered a PMT. The theoretical mass of this peptide is then calculated based on its amino acid sequence and its LC elution time recorded. Next, the same or similarly derived digested proteome sample is analyzed by LC-FTICR using the identical LC separation conditions. In the LC-FTICR analysis, the test involves whether a peptide with an observed mass that closely agrees with the calculated theoretical mass of this peptide (e.g., within 1 ppm) and has a corresponding elution time is actually detected. If a peptide of predicted accurate mass and elution time is detected, it is now considered to confidently function as an AMT for EF-Tu. While we have focused our initial efforts towards obtaining highly confident protein identifications of unmodified tryptic fragments, the AMT validation approach can be used to identify any class of peptide or modified peptide, providing that the modified peptide can be characterized by MS/MS.

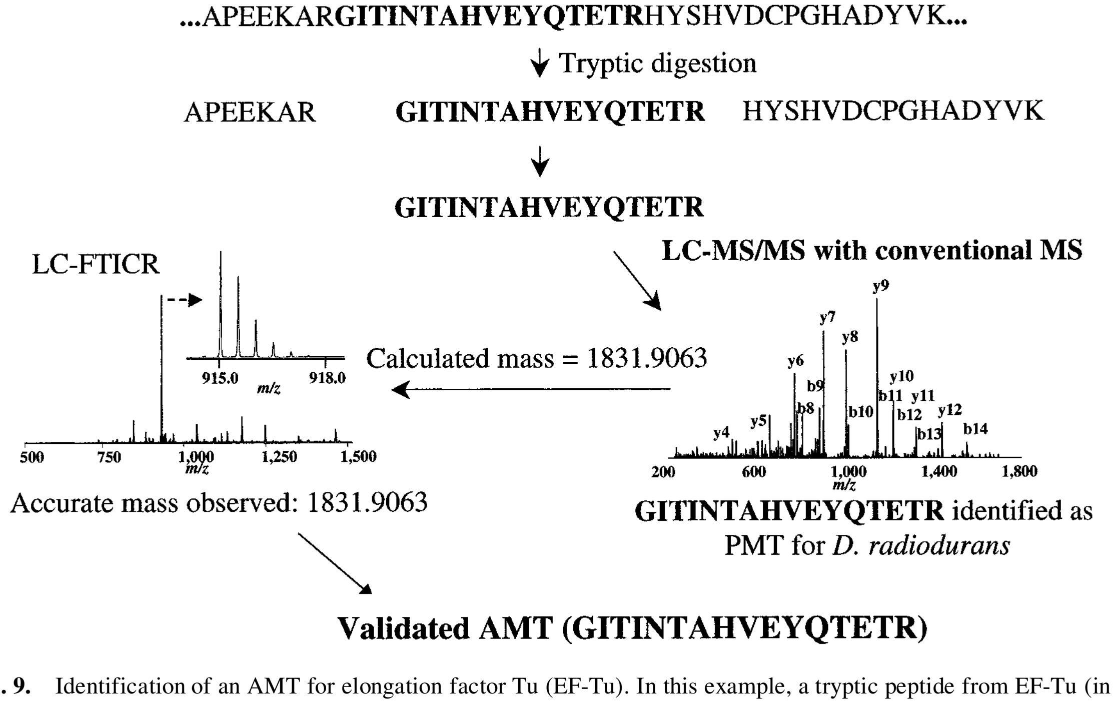

FIG. 9. Identification of an AMT for elongation factor Tu (EF-Tu). In this example, a tryptic peptide from EF-Tu (in bold) was identified by tandem MS using an LCQ ion trap mass spectrometer. The accurate mass of this PMT was calculated based on its sequence (i.e., 1831.9063 Da) and its elution time recorded. In the LC-FTICR analysis of the same sample, a doubly charged peptide was observed at this same elution time, having a mass within 1 ppm (i.e., 1831.9063 Da) of the calculated mass of this peptide. This peptide is then considered an AMT and can be used to identify EF-Tu within the *D. radiodurans* proteome.

Increased confidence in protein identifications using AMTs

The generation of AMTs for a specific proteome has two key advantages over conventional protein identification practices. Firstly, correlating PMT calculated masses to observed accurate masses using high mass accuracy FTICR provides a much higher level of confidence in those peptides identified. Using the search/identification program SEQUEST and a minimum cross correlation score of 2, a large number of polypeptide PMTs have been identified for *D. radiodurans*, however, only ~70% were then validated as AMTs. An analysis of the conversion of PMTs to AMTs as a function of SEQUEST quality score is shown in Figure 10 for all peptides identified from ion trap MS/MS measurements. This analysis shows that more than 95% of the PMTs identified with a SEQUEST cross-correlation (X_{corr}) value of >4.0 are converted into AMTs. A rapid decrease in the conversion of PMTs to AMTs, however, is observed for PMTs identified with an X_{corr} value of <3.0. This result illustrates the conventional approach's effectiveness for peptide identification for high X_{corr} values, but also reveals a rapid decrease in peptide identification confidence (which we believe is attributed to incorrect peptide identifications) at X_{corr} values of <3, which account for the majority of peptides identified. The results shown in Figure 10 support the need for validating peptide identifications obtained using programs such as SEQUEST (we have observed similar results with the program MASCOT [Perkins et al., 1999]) and demonstrate that a significant increase in confidence in peptide identifications is obtained through applying the AMT validation criteria.

The key downstream advantage in generating AMTs is their applicability for high throughput studies designed to compare changes in the relative abundances of proteins between two separate proteome samples (i.e., control versus treated) based solely on the accurate mass measurements provided by FTICR. Once an AMT has been established, it can be used to confidently identify a specific protein in subsequent proteome studies. Without the need to reestablish the identity of a peptide using MS/MS analyses, multiple high throughput studies focused on measuring changes in relative protein abundances between two (or more) different proteomes are facilitated. In such comparative studies, stable isotope labeling methods can be used to provide a means to measure protein relative abundances, a process that also benefits from the resolution and sensitivity of the FTICR measurements.

Increasing proteome coverage using AMTs

Several strategies have been applied in our initial work with *D. radiodurans* to increase the number of AMTs so as to subsequently routinely allow lower-level proteins to be analyzed by this approach. First, samples were analyzed several times using the same capillary LC-MS/MS strategy, but with different *m/z* ranges and with the exclusion of parent ions that were previously selected for MS/MS, resulting in the selection of different peptides and generation of many additional PMTs. Beyond variations in instrumental approaches, proteome samples extracted from cells harvested at different growth phases (i.e., mid-log, stationary phase) or cultured under a variety of different conditions (i.e., nutrients, perturbations) were also analyzed. By varying growth conditions and harvesting stages, the potential pool of PMTs increases significantly since the absolute number of proteins collectively present in the different samples is significantly greater than the number expressed by the organism under a single growth condition. Finally, since any additional sample fractionation will increase the overall dynamic range achievable, we also analyzed peptide fractions first separated off-line by ion exchange chromatography, which again resulted in the generation of large numbers of additional PMTs for peptides that would otherwise have too low abundance for conventional MS/MS analyses. It should be noted that any number of alternative sample fractionation and analysis strategies can be preformed to increase the number of PMTs and AMTs generated, and that the extra efforts at this stage are more than off-set by the resulting ability to make subsequent comprehensive proteome measurements with much greater sensitivity and speed. We continued PMT generation efforts for *D. radiodurans* for over 200 different ion trap MS/MS runs, and until the rate of generation of novel PMTs decreased significantly. Since the analysis procedure can be totally automated, this corresponds to a one-time effort requiring approximately 3 weeks using a single ion trap instrument, and additional experience should significantly reduce the number of runs required for PMT generation.

Thus, while a significant number of samples are analyzed to generate the MS/MS spectra used for generating the set of AMTs for a specific organism, this initial investment of effort obviates the need for rou-

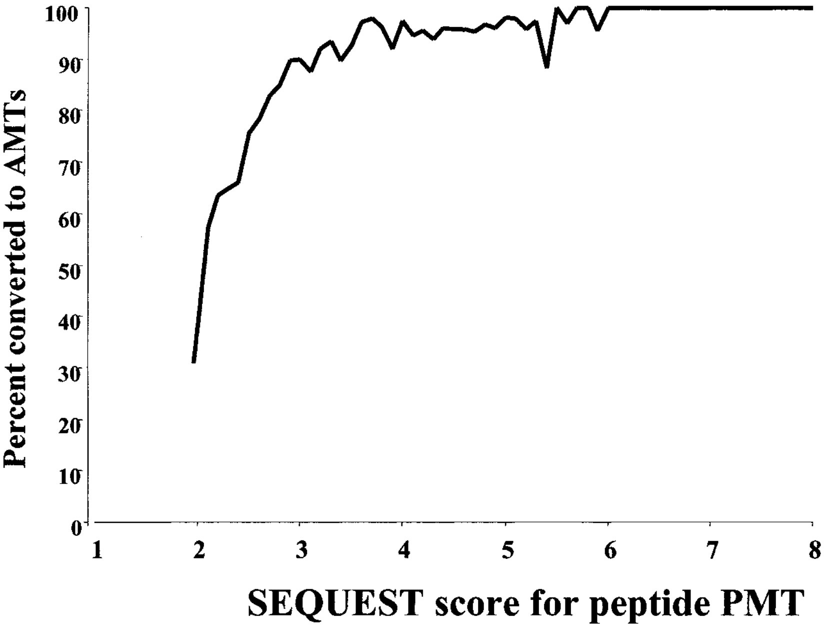

FIG. 10. An analysis of the conversion of *D. radiodurans* PMTs identified from ion trap MS/MS measurements to AMTs as a function of SEQUEST cross correlation score. A polypeptide is converted to an AMT if it is detected using FTICR with a MMA agreement within 1 ppm and a normalized elution time within 20%, resulting in elimination of approximately one-third of all PMTs, and particularly those with lower cross correlation scores. The AMT validation step is automated and results in a significant improvement in the confidence for protein identifications.

tine use of MS/MS in future analyses. The dividends for such an investment are realized in proteome studies designed to quantify changes in the relative abundance of proteins as a function of time or environment. In addition, a significant fractionation of samples prior to LC-MS/MS analysis is necessary to generate a large number of useful spectra using conventional instruments, but these lower-level peptides can then be routinely detected using the FTICR without the need for sample fractionation. Without the need to reidentify the expressed proteins through time-consuming MS/MS analyses and extensive sample fractionation, studies designed to quantify the relative abundances of proteins between two distinct proteome samples can be completed in a high-throughput manner by the exclusive use of the high mass accuracy measurements afforded by FTICR and with low attomole level sensitivity.

Multiplexed-MS/MS for high-throughput polypeptide identification

Due to the complexity of the proteomic samples, multiple peptides generally coelute and often hundreds can be observed in a single FTICR spectrum, even for the highest resolution LC separations. On the other hand, conventional MS/MS analysis is sequential (i.e., can only address one peptide at a time), and the data acquisition rate typically fails to allow selection and fragmentation of all detected peptides in the time available for analysis (i.e., the elution time of a peak). Consequently, the dynamic range is effectively reduced since typically the low abundance ions are not selected for MS/MS analysis even when dynamic exclusion methods are used to prevent repetitive selection of the same peak. To help alleviate the problem, various peak-parking schemes have been developed in which the chromatographic or electrophoretic peak elution time is extended to allow additional MS/MS experiments to be conducted (Davis, 1998; Davis et al., 1995; Goodlett et al., 1993). For example, Martin et al. (2000) recently described a variable-flow HPLC apparatus for on-line tandem mass spectrometric analysis of tryptic peptides. While such an approach alleviates the problem to some extent, comprehensive MS/MS analysis remains impractical for complex proteome digest samples. The intermittent reduction of LC flow rates not only significantly increases the overall separation time but may also decrease sensitivity for MS detection (particularly for the use of very small i.d.

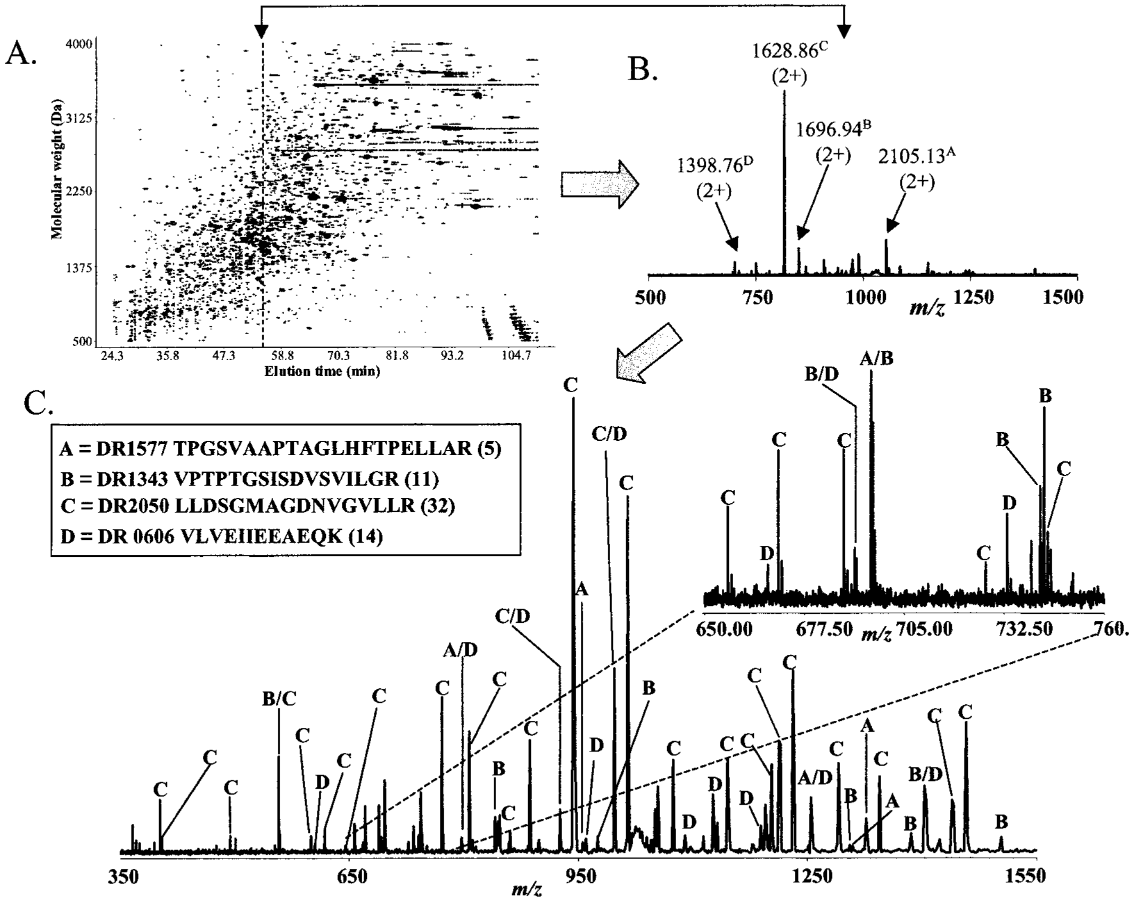

FIG. 11. (**A**) Two-dimensional display reconstructed from (7 tesla) FTICR spectra obtained during a capillary LC-FTICR analysis of a global tryptic digest from a *D. radiodurans* cell lysate. Over 13,000 peptide "spots" (i.e., isotopic distributions) were detected in the FTICR set of mass spectra during the separation. (**B**) A mass spectrum (indicated by dotted line in two-dimensional plot shown in A, with four most abundant ions selected for the subsequent MS/MS acquisition labeled. (**C**) Multiplexed-MS/MS spectrum of the four peptide species selected in B with fragment ions attributed to each individual peptide after searching against the *D. radiodurans* protein database. The four proteins uniquely identified from this spectrum are listed in the inset box with their open reading frame reference number (e.g., DR1577). The numbers listed in parenthesis after each peptide indicated the number of fragment ions detected for each tryptic peptide. An expanded view of *m/z* 650–760 with sequence-specific fragment ions labeled is shown as an inset.

capillaries where the electrospray ionization efficiency is already close to its maximum [Smith et al., 1993]) and the resolution obtained in the separation, and is less useful for high-efficiency LC separations that are based upon the use of high pressures (Shen et al., 2001a).

One approach for addressing this issue involves the effective collection of LC effluent fractions for subsequent analysis using MALDI MS, an approach that allows as much time as needed for MS/MS analyses, or the use of approaches that presume to identify the interesting peptides. However, many issues related to the throughput, dynamic range, sensitivity, sample preparation, storage stability, and matrix/bias effects remain to be addressed to reasonably assess the viability of such an approach.

We have developed an approach that utilizes the multiplexing capability derived from the high resolution and MMA of FTICR for the simultaneous MS/MS analysis of multiple peptides (Masselon et al., 2000) during on-line capillary LC separations. A unique attribute of FTICR is the ability to select and simultaneously dissociate multiple precursor peptides. While repeating MS/MS experiments with different subsets of parent ions allows the assignment of all fragments to the corresponding parent species, these methods require a large amount of sample, and are too slow for the use with on-line separations. Our multiplexed-MS/MS strategy simultaneously obtains sequence information for multiple peptides, providing both enhanced sensitivity and a gain in throughput.

As an initial demonstration of the use of the multiplexed-MS/MS approach for protein identification from complex proteome samples, Figure 11A shows a 2D display for the capillary LC-FTICR analysis of tryptic peptides from *D. radiodurans* whole cell lysate, where the peptide spots are based on their molecular mass and LC elution time. Over 13,000 peptide spots were observed in this single analysis. Figure 11B shows an example of a mass spectrum obtained in a single MS acquisition corresponding to the dotted line in the 2D display, while Figure 11C shows the corresponding (i.e., immediately following) multiplexed MS/MS spectrum with the dissociation products from the four most abundant parent ions selected from spectrum shown in Figure 11B. Table 3 lists fragment ion assignments from this MS/MS spectrum based upon a search of the *D. radiodurans* protein database. A significant number of sequence-specific fragments were attributed to each selected peptide. The extensive fragmentation allowed the identification of the four selected tryptic peptides, with each identifying a protein *D. radiodurans* protein. For example, the peptide with M_r = 1,628.85 Da was identified as LLDSGMAGDNVGVLLR from elongation factor TU (DR2050 and DR0309 which happen to be duplicated open reading frames) and the peptide with M_r = 1,696.94 Da was identified as a tryptic fragment from glyceraldehyde 3-phosphate dehydrogenase (DR1343). We have found that peptides having significant differences in abundances (visualized by variation of the spot size in 2D display and the ion intensity in MS spectrum) could be readily fragmented to yield useful sequence information.

Obviously, the initial stages of proteomic research with any cell or tissue type by this approach will require a larger effort to establish and validate AMTs. These efforts, together with the need to unambiguously identify modified polypeptides or unexpected (e.g., due to frame shifts) polypeptides, can be significantly facilitated by the multiplexed-MS/MS approach. The ability to selectively eject the most abundant species prior to external ion accumulation and the transfer of ions to the ICR cell should allow us to obtain broader proteome coverage by measuring low abundance species that are undetectable using other methodologies. The coupling of DREAMS active dynamic range enhancement methodology with the multiplexed-MS/MS provides a basis for protein identification to be performed with much greater sensitivity and speed.

Identification of Deinococcus radiodurans *proteins*

The use of AMTs approach has been initially demonstrated for the small, radiation-resistant, prokaryotic organism *D. radiodurans*. *D. radiodurans* is a gram-positive, nonmotile, red-pigmented bacterium whose most distinguishing feature is its effectiveness in coping with insults that cause DNA damage (i.e., desiccation, UV, or ionizing radiation; Makarova et al., 2001). *D. radiodurans* has a ~3.1-Mbase genome that initial annotation efforts *predict* to code for 3,187 possible proteins (White et al., 1999). An ideal tryptic digest of all proteins would yield 60,068 polypeptides of 500–4,000 Da, of which ~51% would be unique at 1 ppm MMA, affording coverage of >99% of all predicted proteins. In work to be described elsewhere (Lipton et al., manuscript in preparation), proteins from the organism cultured under a number of different growth conditions typically resulted in the detection of 20,000 to >50,000 peptides by capillary LC-FTICR analysis. (In contrast, while conventional ion trap MS instrumentation generate on the order of 10,000 MS/MS spectra in a single run, a very large fraction of these spectra provide generally no useful peptide identifications.) Using capillary LC with tandem MS measurements (including ion trap tandem MS measurements that generated >9,000 PMTs), a total of 6,997 polypeptides were validated as *bona fide* AMTs. These AMTs provide confident identification of 1,910 predicted proteins (with an average of >3 AMTs per protein), covering ~61% of the predicted proteome which span every category of predicted protein function from the annotated genome. This level and comprehensiveness of proteome coverage exceeds that achieved for any other organism to date, and allowed (depending upon culture conditions) 15 to 25% of the predicted proteome to be identified from AMTs detected in single FTICR runs. In a typical single FTICR analysis we detect ~1,500 AMTs, corresponding to approximately 600 to 700 unique ORFs, representing ~20% of the *D. radiodurans*–predicted proteome. Our initial analysis of the detected ORFs indicates that

TABLE 3. FRAGMENTS OF THE FOUR TRYPTIC PEPTIDES FROM *D. RADIODURANS* IDENTIFIED IN A SINGLE LC FTICR MULTIPLEXED-MS/MS SPECTRUM (SEE FIGURE 11)

m/z	*Measured* $[M + H]^+$	*Calculated* $[M + H]^+$	*Error (ppm)*	*Assignment*	
				Peptide	*Fragment*
401.288	401.288	401.288	−1.0	C	y3
405.200	1212.573	1212.557	13.1	C	b13-NH_3
500.356	500.356	500.356	−0.4	C	y4
554.357	554.357	554.355	3.1	D	b5
557.380	557.380	557.3775	4.5	B/C	y5
559.289	599.289	559.286	5.0	C	b6-H_2O
604.301	604.301	604.294	11.6	D	y5
617.305	617.305	617.297	12.9	C	b6
644.844	1288.678	1288.668	7.7	C	y13
656.452	656.452	656.446	9.1	C	y6
667.445	667.445	667.439	9.0	D	b6
670.329	670.329	670.323	8.3	C	b7-H_2O
688.337	688.337	688.334	4.4	C	b7
691.893	1381.770	1381.757	9.2	D	M-H_2O
701.392	1400.766	1400.775	−6.5	B	y14
727.352	727.352	727.345	9.6	C	b8-H_2O
733.342	733.342	733.337	6.8	D	y6
742.414	1483.820	1483.812	5.4	B	y15-H_2O
743.487	743.487	743.478	12.1	B	y7
745.365	745.365	745.355	13.4	C	b8
770.506	770.506	770.489	22.1	C	y7
796.499	796.499	796.482	21.3	D	b7
797.934	1593.85	1593.875	−15.6	A	y15
806.445	1611.881	1611.852	18.2	C	M-H_2O
840.476	1678.935	1678.925	6.0	B	M-H_2O
860.388	860.388	860.382	6.5	C	b9
885.535	885.535	885.516	21.5	C	y8
925.522	925.522	925.511	11.9	C	y9-NH_3
925.522	925.522	925.525	−3.2	D	b8
942.555	942.555	942.537	19.1	C	y9
946.552	946.552	946.536	16.9	A	y8
955.475	955.475	955.474	1.0	B	b10
959.512	959.512	959.505	7.3	D	y8
996.556	996.556	996.548	8.0	C	y10-NH_3
996.556	996.556	996.562	−6.0	D	b9
1013.59	1013.59	1013.574	15.8	C	y10
1073.514	1073.514	1073.494	18.6	C	b11
1088.557	1088.557	1088.547	9.2	D	y9
1113.503	1113.503	1113.489	13.0	C	b12-NH_3
1125.614	1125.614	1125.604	8.9	D	b10
1144.646	1144.646	1144.615	27.1	C	y11
1187.639	1187.639	1187.616	19.4	D	y10
1201.651	1201.651	1201.636	12.5	C	y12
1212.573	1212.573	1212.557	13.2	C	b13-NH_3
1229.613	1229.613	1229.584	23.6	C	b13
1253.707	1253.707	1253.701	4.8	A	y11
1253.707	`1253.707	1253.663	35.1	D	b11
1288.68	1288.68	1288.668	9.3	C	y13
1303.747	1303.747	1303.722	19.2	B	y13
1307.695	1307.695	1307.675	15.3	A	b14

TABLE 3. (CONT'D) FRAGMENTS OF THE FOUR TRYPTIC PEPTIDES FROM *D. RADIODURANS* IDENTIFIED IN A SINGLE LC FTICR MULTIPLEXED-MS/MS SPECTRUM (SEE FIGURE 11)

				Assignment	
m/z	*Measured* $[M + H]^+$	*Calculated* $[M + H]^+$	*Error (ppm)*	*Peptide*	*Fragment*
1324.728	1324.728	1324.738	−7.5	A	y12
1342.683	1342.683	1342.668	11.2	C	b14
1380.768	1380.768	1380.750	13.1	B	M-H_2O
1383.768	1383.768	1383.748	14.3	B	y14-NH_3
1400.766	1400.766	1400.775	−6.4	B	y14
1437.738	1437.738	1437.741	−2.1	C	b15-H_2O
1455.760	1455.760	1455.752	5.5	C	b15
1501.824	1501.824	1501.822	1.33	B	y15

Peptide assignment corresponds to the open reading frame (ORF) reference number listed in the inset box in Figure 12C.

even greater proteome coverage would be obtained by separately processing and analyzing the insoluble (membrane) protein fraction.

Figure 12 illustrates the high density of data obtained in a single LC-FTICR analysis of a *D. radiodurans* proteome sample where >22,000 putative peptides were observed. The inset within Figure 12 shows

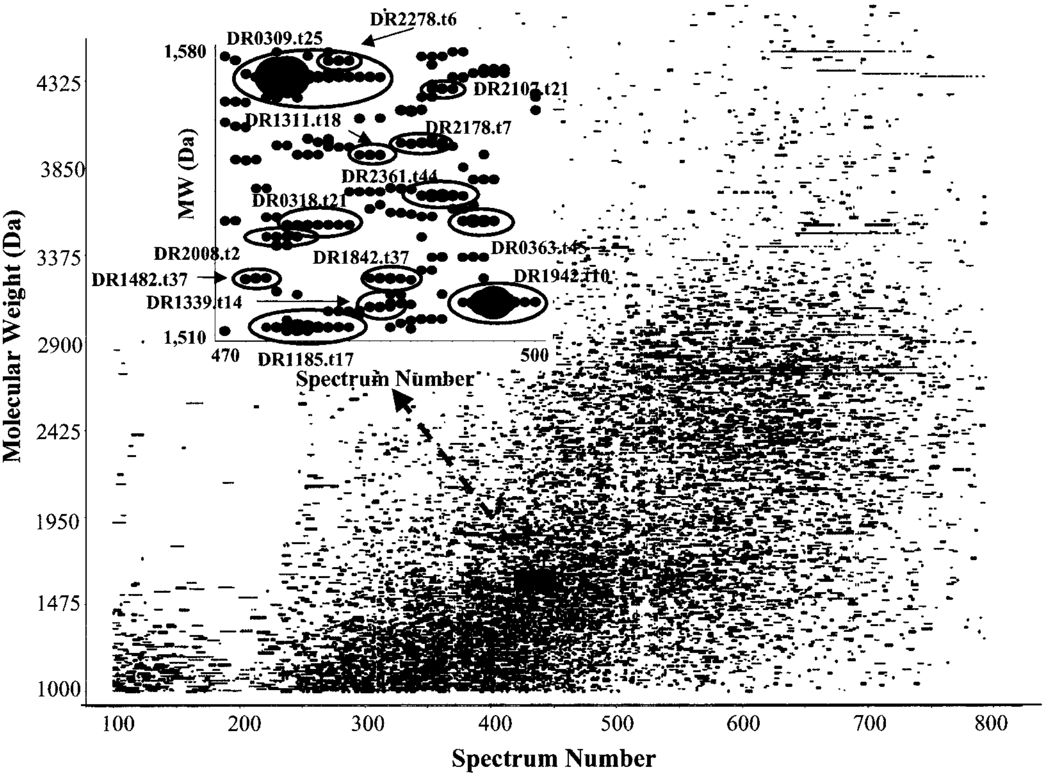

FIG. 12. Two-dimensional display of identified AMTs from *Deinococcus radiodurans*. In this display, all of the peptides observed in an LC-FTICR analysis of the *D. radiodurans* proteome are displayed based on their molecular weight and elution order (i.e., FTICR spectrum number). The circled spots labeled in the inset show spots that were identified as AMTs. The spots are labeled based on their annotation within the organism's genome sequence (i.e., DR0309; elongation factor Tu) and the tryptic peptide of the protein that was identified (i.e., t25; the 25th tryptic peptide counting from the amino terminus, based on complete digestion). A comprehensive list of the AMTs identified, along with their calculated accurate mass and protein of origin is given in Table 4.

detail for a typical region where peptides identified as AMTs are annotated with the ORFs that encode the parent proteins. Table 4 lists proteins identified in this small segment of the 2D display. While a significant fraction of the species detected were not validated as AMTs due to possible ambiguities, many also do not correspond to the masses of possible peptides predicted from the sequences genome. The large MS/MS and accurate mass data sets from such studies thus contain extensive information that can be subsequently mined for information unpredictable from genomic sequence data, such as posttranslational modifications, sequence variations, and frame shifts, upon the development and application of improved bioinformatic software tools. The large fraction of detected species not validated as AMTs includes those not unique in mass (~50%), peptides unexpectedly modified, peaks due to contamination, or peptides that are otherwise unexpected or missed by gene-calling algorithms. Indeed, an initial limited analysis of this data has revealed protein modifications, as well as a number of instances of apparent frame shifts during protein expression or potential DNA sequencing errors that we will report elsewhere.

The proteins identified from *D. radiodurans* using our approach include many predicted to be present in low abundance based on their predicted codon adaptation index (CAI; Sharp and Li, 1987). CAI has been shown to be a crude but useful predictor of protein abundance; proteins with high CAI values tend to be highly expressed, and those with low CAI values tend to be expressed at very low levels. Figure 13 shows the distribution of CAI values for the identified *D. radiodurans* proteins compared to the distribution for all proteins predicted from the genome of this organism. In general, CAI values for the proteins identified by AMTs group in a Gaussian-like distribution similar to that for all of the predicted proteins, with more than 90% of predicted proteins having CAI values >0.8 being detected. In addition, we detected 18 out of the 54 (i.e., 33%) *D. radiodurans*–predicted proteins that have a CAI value of <0.2. (In comparison for yeast, 2D PAGE detects very few proteins in this range [Emmert-Buck et al., 2000], whereas Washburn et al. [2001] reported 16% using multidimensional LC online with ion trap tandem MS.) These results further support that the AMT approach provides a good representation of the proteins expressed.

In this initial work, we applied methods based solely upon a global tryptic digestion that would be expected to be much less effective for membrane proteins. Remarkably, we find that a significant fraction of the most hydrophobic proteins are still detected. Figure 14 shows that the percentage of proteins detected as a function of the portion of each protein predicted to reside in a membrane decreases by ~50% for the most hydrophobic proteins. Further refinements of sample solubilization and digestion methods, such as those utilized by Washburn et al. (2001), should further increase membrane protein coverage.

TABLE 4. ACCURATE MASS TAGS AND THEIR CORRESPONDING PROTEINS IDENTIFIED WITHIN THE INSET OF FIGURE 12

AMT	*Peptide sequence*	*Calculated AMT M_r (Da)*[a]	*Protein of origin*
DR0309.t25	VQDEVEIVGLTDTR	1572.7994	Elongation factor TU
DR0318.t21	VMFEVAGVTEEQAK	1536.7493	Ribosomal protein L16
DR0363.t45	TMALPDSFPGYDPK	1537.7122	Putative peptide ABC transporter
DR1185.t17	APGFADYTTTITVR	1511.7619	S-layer-like array-related protein
DR1311.t18	TGDIGHAIQSLAESR	1553.7797	Methionine aminopeptidase
DR1339.t14	TATADDAEELAAAIR	1516.7368	Triosephosphate isomerase
DR1482.t37	ETYEIMNAELVGR	1523.7289	2-Isopropylmalate synthase
DR1942.t10	FGVTIPDEAAETIR	1517.7725	Acyl carrier protein
DR2008.t2	VAIVGATGAVGHELLK	1533.8878	Aspartate-semialdehyde dehydrogenase
DR2107.t21	GENLGGLIITHYQR	1569.8263	ABC transporter
DR2178.t7	HGEVPAEAHAALVQK	1555.8106	Adenylosuccinate lyase
DR2278.t11	GSDGQVQGFDIDIAR	1576.7481	Amino acid ABC transporter
DR2361.t44	AAQQLGSITMVIGQK	1543.8391	Putative acyl-CoA dehydrogenase

[a]Monoisotopic molecular weight.

Quantitative high throughput proteome-wide measurements

Useful proteome measurements often require comparing protein abundances between two cellular populations resulting from, for example, some insult or perturbation. The predominant method for measuring changes in protein expression levels using current proteomic technology is to compare the intensities of the corresponding 2D PAGE spots. Measurements of peptide abundances using MS signal intensities can vary significantly for reasons that include variations in ionization efficiencies and losses during sample preparation and separations, and while useful for large differences in abundances, are unsuited to study more subtle variations. The generation and use of AMTs enables high-throughput and high-precision expression studies based upon stable isotope labeling by directly comparing two proteomes in the same analysis (e.g., utilizing a reference proteome to which perturbed systems are compared). A stable-isotope–labeled reference proteome, for example, provides an effective internal standard for each protein, and hence their tryptic peptides, allowing changes in protein abundances based upon the relative abundances of AMTs to be assessed, potentially to precisions better than 10% (Gygi et al., 1999a, 2000b; Oda et al., 1999; Pasa-Tolic et al., 1999). While such measurements require both versions of the protein or peptide to be present, it should be feasible to combine this information with absolute peak intensity data to provide less precise abundances for cases where only one peptide is detected, and to also establish approximate absolute abundances (albeit, with less precision).

A 2D display is shown in Figure 15 where the relative peptide abundances for *D. radiodurans*–untreated cells are compared with those treated with H_2O_2. The colored spots represent the measured relative abundance levels of peptides (green represents a decrease in abundance, black unchanged and red an increase). In this preliminary study, *D. radiodurans* cells were cultured in both normal and ^{15}N-enriched growth media to constitute a reference proteome; the inset in Figure 15 shows the mass spectral region for two peptide pairs corresponding to AMTs for catalase and S-layer protein. As previously observed using conventional methods (Carbonneau et al., 1989; Wang, 1995), significant differences in a large fraction of peptide abundances following H_2O_2 exposure are observed. Although methodological development is required to develop and define its limitations, this general approach should provide a global view of the response of a proteome to a perturbation, and the ability to conduct many such experiments will significantly contribute to an increased understanding of the functions of the proteins and their interactions. Recent work has also highlighted the promise of applying such methods to selected subproteomes, for example, by isotopic labeling and affinity isolation of phosphopeptides (Goshe et al., 2001; Oda et al., 2001; Zhou et al., 2001).

It should be noted that high-resolution capillary LC-FTICR experiments on complicated samples produced huge data sets (>1.5 GB) for each analysis. Even using the most current desktop PC computers, nearly two days for data processing are required. The development of higher-speed computers, the implementation of multiprocessor, PC clusters, and distributed computing platforms will be a necessary aspect of high-throughput proteome analyses.

DREAMS for the future

The possibilities for application of the new quantitative proteome measurement technology and methods described in this work are effectively unlimited. While major challenges remain to realize the full potential of the technology, particularly for obtaining useful quantitative measurements for low-abundance proteins and for extending the approach to modified proteins, it is clear that many useful applications to microbial systems are already tractable. The combination of sensitivity, dynamic range and throughput should enable new types of studies to be contemplated. In particular studies that would otherwise demand excessive quantities of protein or present too much complexity should now be tractable. As an example, it may be possible to begin the study of modal microbial communities as new approaches provide DNA sequence information for their constituents.

We believe, for example, that the DREAMS FTICR technology is an important component of an approach that provides the basis for a significant gain in the coverage of proteomic measurements. It is clear that such technology development efforts can have a significant impact upon the practice of proteomics, particularly for applications where measurements of the highest quality and broadest scope are beneficial.

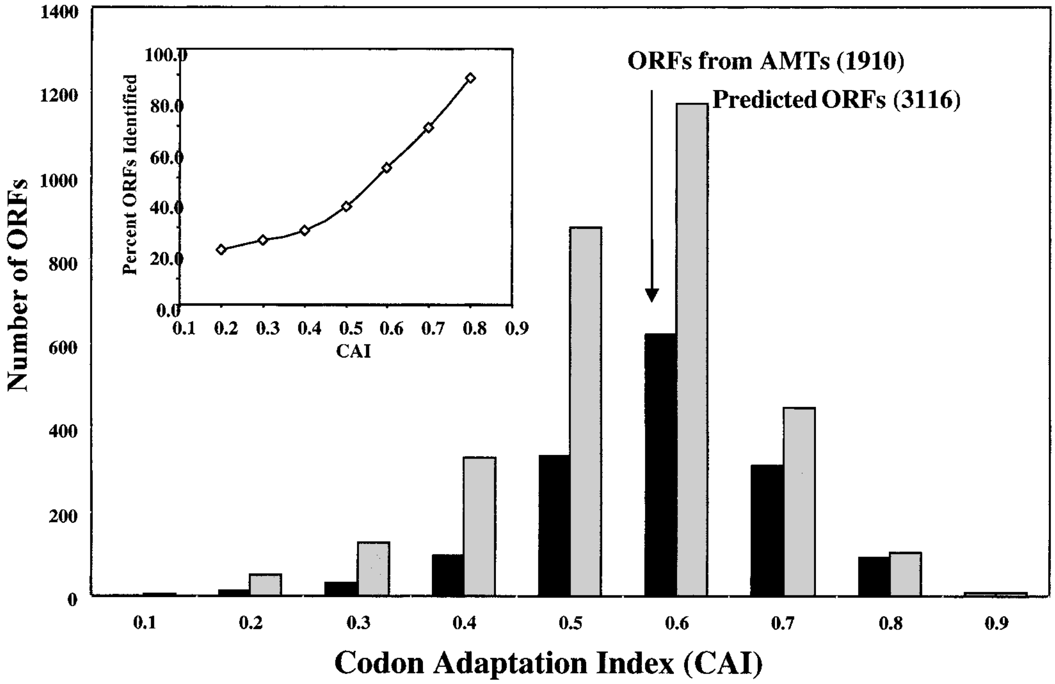

FIG. 13. Distribution of CAI values for detected *D. radiodurans* proteins compared to the distribution for all predicted proteins (Sharp and Li, 1987). This comparison indicates that, while there is some bias for high CAI proteins, many proteins with low CAI values, and thus predicted to be expressed at only low levels (and not generally detected by 2D PAGE), are observed.

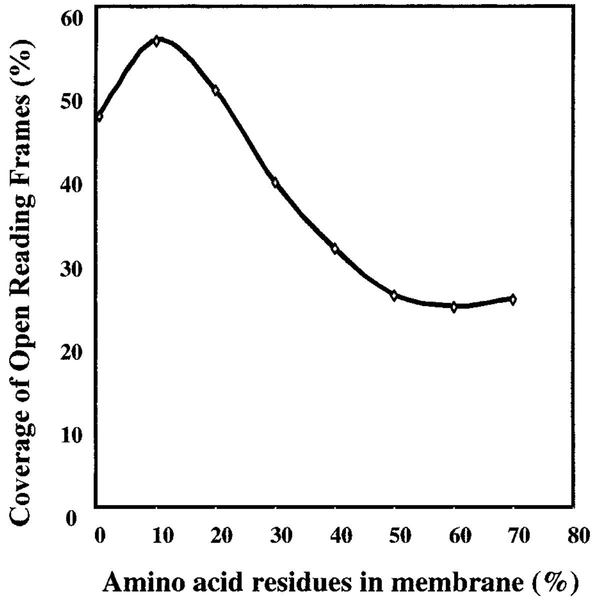

FIG. 14. The percentage of detected *D. radiodurans* proteins as a function of the portion of each protein predicted to reside in a membrane is shown, indicating there is a bias for detection against the most hydrophobic proteins, which is attributed to the sample processing and digestion conditions used in this initial work.

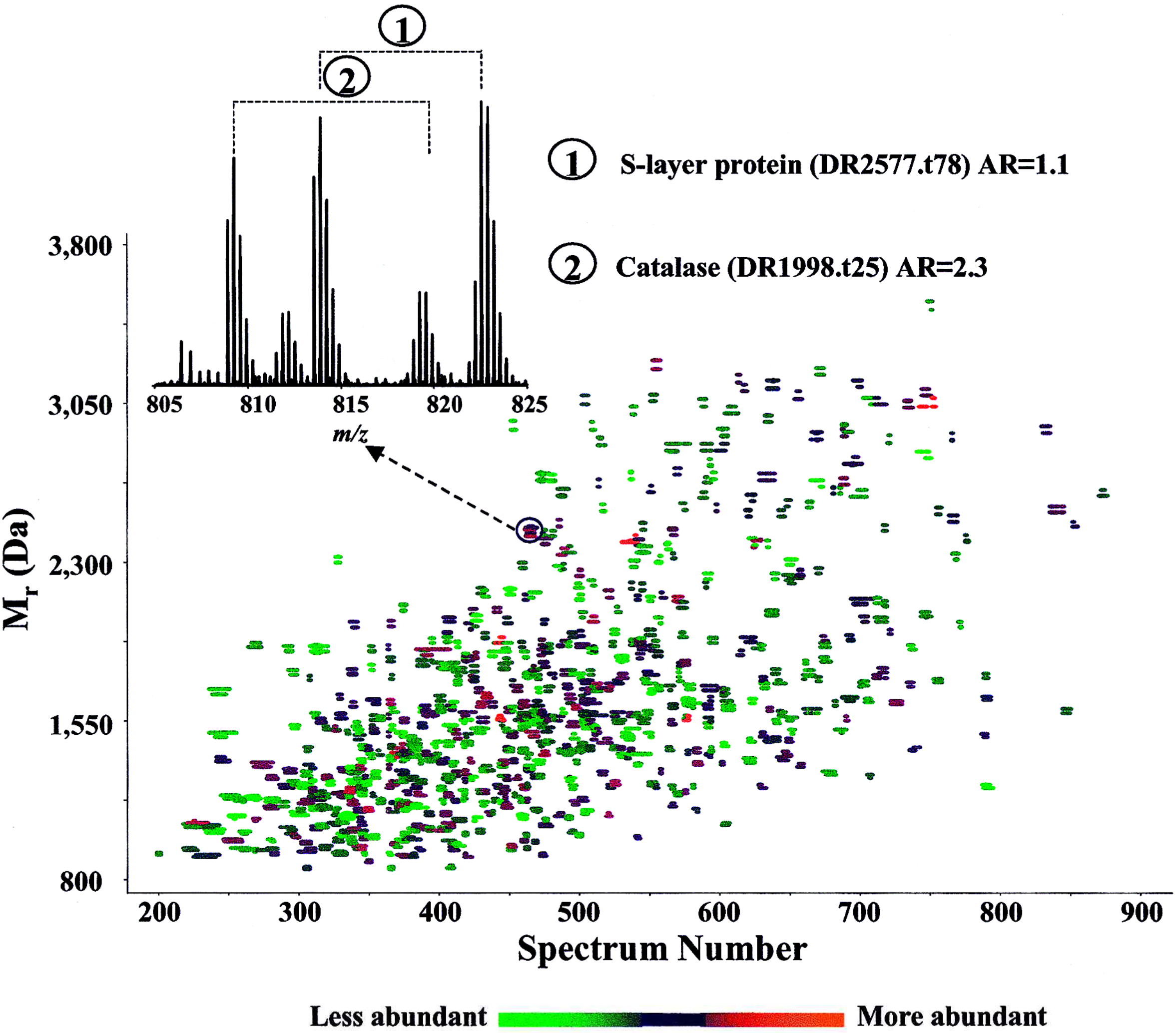

FIG. 15. Two-dimensional display comparing relative protein abundances of control and H_2O_2-treated *D. radiodurans*. The colored spots provide a representation of the relative expression level of the peptides (green represents a decrease in abundance, black unchanged, and red an increase). The inset shows the peak pair results for two selected AMTs corresponding to S-layer protein and catalase observed in the control and H_2O_2-treated cells, along with their calculated abundance ratios (AR).

Over the next couple of years the initial application of this new technology to the study of microbial systems should clarify its role.

ACKNOWLEDGMENTS

We acknowledge the contributions of Kim Hixson, Ron Moore, Rui Zhao, Richard Harkewicz, David Anderson, Nikola Tolić, Lingjun Li, Ken Auberry, Deanna Auberry, Margie Romine, and Brian Thrall (all of PNNL) to the work reviewed here. We thank the U.S. Department of Energy Office of Biological and Environmental Research for long time support of the *D. radiodurans* research and the FTICR technology development, as well as the National Institutes of Health, through NCI (CA81654), NINDS (NS39617), and NCRR (RR12365), for support of portions of this work. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830.

REFERENCES

ADAMS, M.D. (1996). Serial analysis of gene expression: ESTs get smaller. Bioessays **18,** 261–262.

ANDERSON, L., and SEILHAMMER, J. (1997). A comparison of selected mRNA and protein abundances in human liver. Electrophoresis **18,** 533–537.

BELOV, M.E., GORSHKOV, M.V., UDSETH, H.R., et al. (2000a). Zeptomole-sensitivity electrospray ionization—Fourier transform ion cyclotron resonance. Anal Chem **72,** 2271–2279.

BELOV, M.E., GORSHKOV, M.V., UDSETH, H.R., et al. (2000b). Initial implementation of an electrodynamic ion funnel with FTICR mass spectrometry. J Am Soc Mass Spectrom **11,** 19–23.

BELOV, M.E., ANDERSON, G.A., ANGELL, N.H., et al. (2001a). Dynamic range expansion applied to mass spectrometry based on data-dependent selective ion ejection in capillary liquid chromatography Fourier transform ion cyclotron resonance for enhanced proteome characterization. Anal Chem **73,** 5052–5060.

BELOV, M.E., GORSHKOV, M.V., ALVING, K., et al. (2001b). Optimal pressure conditions for unbiased external ion accumulation in a 2D rf-quadrupole for FTICR mass spectrometry. Rapid Commun Mass Spectrom **15,** 1988–1996.

BELOV, M.E., NIKOLAEV, E.N., ANDERSON, G.A., et al. (2001c). Electrospray ionization—Fourier transform ion cyclotron mass spectrometry using ion pre-selection and external accumulation for ultra-high sensitivity. J Am Soc Mass Spectrom **12,** 38–48.

BELOV, M.E., NIKOLAEV, E.N., HARKEWICZ, R., et al. (2001d). Ion discrimination during ion accumulation in a quadrupole interface external to a Fourier transform ion cyclotron resonance mass spectrometer. Int J Mass Spectrom **208,** 205–225.

BRUCE, J.E., ANDERSON, G.A., BRANDS, M.D. et al. (2000). Obtaining more accurate FTICR mass measurements without internal standards using multiply charged ions. J Am Soc Mass Spectrom **11,** 416–421.

CARBONNEAU, M., MELIN, A., PERROMAT, A., et al. (1989). The action of free radicals on *Deinococcus radiodurans* carotenoids. Arch Biochem Biophys **275,** 244–251.

CONRADS, T.P., ALVING, K., VEENSTRA, T.D., et al. (2001). Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and ^{15}N-metabolic labeling. Anal Chem **73,** 2132–2139.

CONRADS, T.P., ANDERSON, G.A., VEENSTRA, T.D., et al. (2000). Utility of accurate mass tags for proteome-wide protein identification. Anal Chem **72,** 3349–3354.

DAVIS, M.T., and LEE, T.D. (1998). Rapid protein identification using a microscale electrospray LC/MS system on an ion trap mass spectrometer. J Am Soc Mass Spectrom **9,** 194–201.

DAVIS, M.T., STAHL, D.C., HEFTA, S.A., et al. (1995). A microscale electrospray interface for on-line, capillary liquid chromatography tandem mass spectrometry of complex peptide mixtures. Anal Chem **67,** 4549–4556.

DUCRET, A., VANOOSTVEEN, I., ENG, J.K., et al. (1998). High-throughput protein characterization by automated reverse-phase chromatography electrospray tandem mass spectrometry. Protein Sci **7,** 706–719.

EMMERT-BUCK, M.R., STRAUSBERG, R.L., KRIZMAN, D.B., et al. (2000). Molecular profiling of clinical tissue specimens. Am J Pathol **156,** 1109–1115.

FUTCHER, B., LATTER, G.I., MONARDO, P., et al. (1999). A sampling of the yeast proteome. Mol Cell Biol **19,** 7357–7368.

GARRELS, J.I., McLAUGHLIN, C.S., WARNER, J.R., et al. (1997). Proteome studies of saccharomyces cerevisiae—identification and characterization of abundant proteins. Electrophoresis **18,** 1347–1360.

GOODLETT, D.R., WAHL, J.H., UDSETH, H.R., et al. (1993). Reduced elution speed detection for capillary electrophoresis mass-spectrometry. J Microcolumn Separations **5,** 57–62.

GOSHE, M., CONRADS, T., PANISKO, E., et al. (2001). Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. Anal Chem **73,** 2578–2586.

GYGI, S.P., RIST, B., GERBER, S.A., et al. (1999a). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol **17,** 994–999.

GYGI, S.P., ROCHON, Y., FRANZA, B.R., et al. (1999b). Correlation between protein and mRNA abundance in yeast. Mol Cell Biol **19,** 1720–1730.

GYGI, S.P., CORTHALS, G.L., ZHANG, Y., et al. (2000a). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci USA **97,** 9390–9395.

GYGI, S.P., RIST, B., and AEBERSOLD, R. (2000b). Measuring gene expression by quantitative proteome analysis. Biotechnology **11,** 396–401.

HAYNES, P.A., GYGI, S.P., FIGEYS, D., et al. (1998). Proteome analysis: biological assay or data archive? Electrophoresis **19,** 1862–1871.

HENZEL, W.J., BILLECI, T.M., STULTS, J.T., et al. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci USA **90,** 5011–5015.

JAMES, P., QUADRONI, M., CARAFOLI, E., et al. (1993). Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun **195,** 58–64.

JENSEN, P.K., PAŠA TOLIC´, L., ANDERSON, G.A., et al. (1999). Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem **71,** 2076–2084.

KIM, T., TOLMACHEV, V., HARKEWICZ, R., et al. (2000). Design and implementation of a new electrodynamic ion funnel. Anal Chem **72,** 2247–2255.

KITAYAMA, S., and MATSUYAMA, A. (1971). Mechanism for radiation lethality in *M. radiodurans*. Int J Radiat Biol Rel Stud Phys Chem Med **19,** 13–19.

LANGEN, H., TAKÁCS, B. EVERS, S., et al. (2000). Two-dimensional map of the proteome of *Haemophilus influezae*. Electrophoresis **21,** 411–429.

LINK, A.J., HAYS, L.G., CARMACK, E.B., et al. (1997). Identifying the major proteome components of *Haemophilus influenzae* type–strain nctc 8143. Electrophoresis **18,** 1314–1334.

MAKAROVA, K.S., ARAVIND, L., WOLF, Y.I., et al. (2001). Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomes. Microbiology **65,** 44–79.

MANN, M., HOJRUP, P., and ROEPSTORFF, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol Mass Spectrom **22,** 338–345.

MARSHALL, A.G., HENDRICKSON, C.L., and JACKSON, G.S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: a primer. Mass Spectrom Rev **17,** 1–35.

MARTIN, S.E., SHABANOWITZ, J., HUNT, D.F., et al. (2000). Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem **72,** 4266–4274.

MASSELON, C., ANDERSON, G.A., HARKEWICZ, R., et al. (2000). Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. Anal Chem **72,** 1918–1924.

McCORMACK, A.L., SCHIELTZ, D.M., GOODE, B., et al. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. Anal Chem **69,** 767–776.

ODA, Y., HUANG, K., CROSS, F.R., et al. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. Proc Natl Acad Sci USA **96,** 6591–6596.

ODA, Y., NAGASU, T., and CHAIT, B. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat Biotechnol **19,** 379–382.

PAPPIN, D.J., HOJRUP, P., and BLEASBY, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol **3,** 327–332.

PAŠA TOLIC´, L., JENSEN, P.K., ANDERSON, G.A., et al. (1999). High-throughput proteome-wide precision measurements of protein expression using mass spectrometry. J Amer Chem Soc **121,** 7949–7950.

PERKINS, D., PAPPIN, D., CREASY, D., et al. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20,** 3551–3567.

PERROT, M. (1999). Two-dimensional gel protein database of saccharomyces cerevisiae. Electrophoresis **20,** 2280–2298.

SHARP, P., and LI, W. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res **15,** 1281–1295.

SHEN, Y., TOLIĆ, N., ZHAO, R., et al. (2001a). High-throughput proteomics using high efficiency multiple-capillary liquid chromatography with on-line high performance ESI-FTICR mass spectrometry. Anal Chem **73,** 3011–3021.

SHEN, Y., ZHAO, R., BELOV, M.E., et al. (2001b). Packed capillary reversed-phase liquid chromatography with high-performance electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for proteomics. Anal Chem **73,** 1766–1775.

SHEVCHENKO, A., JENSEN, O.N., PODTELEJNIKOV, A.V., et al. (1996a). Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci USA **93,** 14440–14445.

SHEVCHENKO, A., WILM, M., VORM, O., et al. (1996b). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal Chem **68,** 850–858.

SMITH, R.D., WAHL, J.H., GOODLETT, D.R., et al. (1993). Capillary electrophoresis/mass spectrometry. Anal Chem **65,** A574–A584.

SPAHR, C.S., SUSIN, S.A., BURES, E.J., et al. (2000). Simplification of complex peptide mixtures for proteomic analysis: reversible biotinylation of cysteinyl peptides. Electrophoresis **21,** 1635–1650.

VELCULESCU, V.E., ZHANG, L., ZHOU, W., et al. (1997). Characterization of the yeast transcriptome. Cell **88,** 243–251.

WANG, P., and SCHELLHORN, H. (1995). Induction of resistance to hydrogen peroxide and radiation in *Deinococcus radiodurans*. Can J Microbiol **41,** 170–176.

WASHBURN, M.P., WALTERS, D., and YATES, J.R.I. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol **19,** 242–247.

WHITE, O., EISEN, J.A., HEIDELBERG, J.F., et al. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* r1. Science **286,** 1571–1577.

WILM, M., SHEVCHENKO, A., HOUTHAEVE, T., et al. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature **379,** 466–469.

YATES, J.R. (1998). Mass spectrometry and the age of the proteome. J Mass Spectrom **33,** 1–19.

YATES, J.R.I., SPEICHER, S., GRIFFIN, P.R., et al. (1993). Peptide mass maps: a highly informative approach to protein identification. Anal Biochem **214,** 397–408.

YATES, J.R., ENG, J.K., MCCORMACK, A.L., et al. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem **67,** 1426–1436.

YATES, J.R., McCORMACK, A.L., and ENG, J. (1996). Mining genomes with MS. Anal Chem **68,** A534–A540.

ZHOU, H., WATTS, J., and AEBERSOLD, R. (2001). A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol **19,** 375–378.

Address reprint requests to:
Dr. Richard D. Smith
Environmental and Molecular Sciences Laboratory
MSIN: K8-98
Pacific Northwest National Laboratory
P.O. Box 999
Richland, WA 99352