Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS

Andrew J. Forbes, Steven M. Patrie, Gregory K. Taylor, Yong-Bin Kim, Lihua Jiang, and Neil L. Kelleher*

Department of Chemistry, University of Illinois at Urbana–Champaign, 600 South Mathews Avenue, Urbana, IL 61801

Edited by Stephen J. Benkovic, Pennsylvania State University, University Park, PA, and approved January 7, 2004 (received for review October 11, 2003)

For more complete characterization of DNA-predicted proteins (including their posttranslational modifications) a ''top-down'' approach using high-resolution tandem MS is forwarded here by its application to methanogens in both hypothesis-driven and discovery modes, with the latter dependent on new automation benchmarks for intact proteins. With proteins isolated from ribosomes and whole-cell lysates of *Methanococcus jannaschii* (≈1,800 genes) using a 2D protein fractionation method, 72 gene products were identified and characterized with 100% sequence coverage via automated fragmentation of intact protein ions in a custom quadrupole/Fourier transform hybrid mass spectrometer. Three incorrect start sites and two modifications were found, with one of each determined for MJ0556, a 20-kDa protein with an unknown methylation at ≈50% occupancy in stationary phase cells. The separation approach combined with the quadrupole/Fourier transform hybrid mass spectrometer allowed targeted and efficient comparison of histones from *M. jannaschii*, *Methanosarcina acetivorans* (largest Archaeal genome, 5.8 Mb), and yeast. This finding revealed a striking difference in the posttranslational regulation of DNA packaging in Eukarya vs. the Archaea. This study illustrates a significant evolutionary step for the MS tools available for characterization of WT proteins from complex proteomes without proteolysis.

The development of mass spectrometry (MS) to spearhead large-scale protein analysis continues its long maturation toward the global sample coverage achieved routinely with DNA microarrays (1). Of course, the field of proteomics involves a far more complicated measurement challenge, with posttranslational modification (PTM) of proteins one possible source of extra complexity even in Bacterial and Archaeal proteomes. Although identification of thousands of proteins (2, 3) with information about their relative abundance changes (4) is now possible, the task of detecting and localizing protein modifications is far more difficult (5, 6). Recent proteome-scale methods can use tryptic digestion of entire cell lysates into pools of peptides (7), producing mixtures of staggering complexity. Before such ''shotgun'' digestion methods (8), the classical approach of using 2D gels gave a different perspective of the proteome by visualizing intact proteins before their proteolytic digestion (9). Robotic systems now allow fast identification of proteins from 2D gels, but do not readily provide characterization of modifications (10).

Recent application of 2D gel technology to the proteome of a thermophilic (85°C) and barophilic methanogen, *Methanococcus jannaschii* (11), identified 170 proteins from 166 spots in multiple 2D gels. Few proteins >pI 8 (16 distinct proteins) or <15 kDa (22 distinct proteins) were identified. Furthermore, a few potential PTMs were postulated (from identifications of the same protein from multiple spots), but the peptide data from in-gel digestion did not provide direct evidence for the presence or absence of PTMs. In a separate study, Mukhopadhyay *et al.* (12) found differential expression of flagella proteins (by both electron microscopy and 2D gels/MS) based on the level of hydrogen partial pressure used during *M. jannaschii* cell cultivation.

M. jannaschii was first cultured in 1983 from material isolated from the base of a ''white smoker'' submarine hydrothermal vent (13) and was the first Archaeon for which the whole genome (1.8 Mb) was sequenced (14). About 34% of the predicted gene products have no functional predictions based on sequence homology and 87% are <50 kDa. In another methanogen, *Methanobacterium thermoautotrophicum*, two novel methylations were found in the active site of methyl coenzyme M reductase (15), which catalyzes the final reaction step in the formation of methane. Archaeal S-layer proteins in general (16) and *M. jannaschii* flagella proteins in particular (17) are thought to harbor PTMs.

One of the most active areas of technology development in proteomics is for the measurement of PTMs. Beyond protein arrays, MS methods fall into three general categories: targeted, bottom-up, and top-down. Measurement approaches involving targeted enrichment of tryptic peptides harboring specific PTMs (e.g., phosphopeptides) have been established (18–20). A more general method for measuring a greater diversity of PTMs is the ''sequence coverage'' approach, which has three incarnations. Bottom-up seeks to detect the maximal number of (overlapping) peptides (21), attempting to create a complete peptide map. Top-down involves direct fragmentation of protein ions in the gas phase (22). A hybrid approach first measures the intact protein by MS but uses subsequent proteolysis in solution and peptide mapping (23).

Intact proteins represent a major opportunity as the evolution of large-molecule MS continues. Advancing such MS technology using an ion trap instrument, McLuckey and coworkers (24) showed positive identification of six *Escherichia coli* proteins <10 kDa. McLafferty and associates (25, 26) have been expanding the capabilities of electron capture dissociation in its application to intact proteins >15 kDa. Lee *et al.* (27) recently demonstrated identification of many ribosomal proteins (RPs) <44 kDa from the large subunit of yeast. For proteins >50 kDa substantial limitations currently exist because of more protein charge states, more isotopic peaks, and signal suppression during electrospray ionization (ESI) (28). Our laboratory has lowered many barriers toward a wider realization of top-down advantages by establishing a method for 2D proteome fractionation (29), automating protein ion fragmentation (30), and developing an informatic framework (31) and tailored software (32). Here, we combine these improvements to demonstrate top-down proteomics. Seventy-two *M. jannaschii* proteins are processed with 100% sequence coverage, with detection of methylation, N-terminal processing, acetylation, and disulfide bonds. Further, the measurement platform allowed fast cross-species compari-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: PTM, posttranslational modification; Q-FTMS, quadrupole-Fourier transform MS; ALS, acid labile surfactant; RPLC, reversed-phase liquid chromatography; RP, ribosomal protein; ESI, electrospray ionization; MS/MS, tandem MS.

*To whom correspondence should be addressed. E-mail: kelleher@scs.uiuc.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0306575101

son of PTMs hypothesized to exist by examination of a sequenced genome. The study thus extends "hypothesis-driven" MS (33) from small molecules into the regime of WT proteins targeted from the complex proteome of *Methanosarcina acetivorans*, the Archaeon with the largest known genome (5.8 Mb).

Experimental Methods

Cells and Cell Lysis. *M. jannaschii* JAL-1 cells were grown by Biswarup Mukhopadhyay (Virginia Commonwealth University, Richmond), harvested, and stored at −80°C, as described (34). Two to three grams of wet cells were resuspended in 6 ml of 100 mM Tris·HCl (Fisher Scientific), pH 8.0, and lysed by using pulsed microtip sonication on ice. Fifty units of DNase (Sigma) and various protease inhibitors were added, and the mixture was incubated for 30 min at 37°C. During this time reversible or base-labile modifications may have been lost, making lysate processing under denaturing conditions and/or with inhibitors of modification-removing enzymes preferred. For *Saccharomyces cerevisiae* S288C grown to stationary phase in yeast extract/peptone/dextrose media, cell lysis was performed by French press in the presence of 2% acid-labile surfactant (ALS) before continuous elution gel electrophoresis; protease inhibitors were present. For *M. acetivorans* C2A, grown as described (35), histone acetylation was targeted and therefore sodium butyrate was used at 15 μM as a histone deacetylase inhibitor during cell growth (added 4 h before harvesting) (36). Lysates were cleared by centrifugation at 25,000 × *g* for 20 min at 4°C.

2D Fractionation. Whole-cell lysates of *M. jannaschii*, *M. acetivorans*, and yeast were fractionated by continuous elution gel electrophoresis (PAGE) with subsequent reversed-phase liquid chromatography (RPLC) as described (29). After preparation as described above, 3–4 ml of lysate (≈50–100 mg total protein) was added to an equal volume of PAGE sample buffer with an ALS. After incubation at 70°C for 5 min, samples were loaded on a preparative ALS-polyacrylamide gel at 9% T. The gel was run at constant power (12 W) for ≈8 h, and after elution of the dye front, fractions were collected for 3 min each at a flow rate of ≈1 ml/min.

The bulk buffer and surfactant were removed from fractions by acetone precipitation. Fractions 1–15 were processed directly, and adjacent fractions >15 were combined. Precipitated pellets were resuspended in water, acidified to pH 2 with trifluoroacetic acid, and incubated at room temperature for ≈1.5 h to hydrolyze any remaining ALS. These fractions (≈40–60 μg total protein) were then injected onto a Jupiter C4 (Phenomenex, Belmont, CA) or an ODS-I RPLC column (Eprogen, Darien, IL) and separated by using a linear gradient over 30 min. Resulting subfractions were frozen at −80°C before lyophilization.

Partial Purification of *M. jannaschii* RPs. A supernatant from the *M. jannaschii* lysate (described above) was kept and spun again at 100,000 × *g* for 1 h at 4°C. The resulting pellet was washed three times with 100 mM Tris·HCl, pH 8.0 buffer with increasing KCl concentration: 0, 0.5, and 1.0 M KCl. The resulting pellet constituted the crude ribosomal preparation, for which the complete composition and purity are unknown (Gary Olsen, personal communication). This pellet was resuspended by using 8 M urea, and then separated by RPLC using a PLRP-S column (Polymer Laboratories, Amherst, MA) using a standard gradient. Fractions were collected manually and lyophilized. Predictions of *M. jannaschii* RPs were obtained from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov).

ESI/Quadrupole-Fourier Transform MS (Q-FTMS). Dried samples were resuspended in electrospray solution (78% acetonitrile/20% water/2% acetic acid) and spun for 5 min at 16,000 × *g*. Electrospray conditions were as follows: 2,000–2,250 V on the needle, 120–250 V on the tube lens, and 330 nl/min flow rate. Total protein concentration in a given sample was ≈1–30 μM. Mass spectra were acquired by using external accumulation on an 8.5-T FTMS with a quadrupole ion filter between the source and cell (37) and gradient ion extraction from the accumulation region (38). MIDAS software with an accompanying data station (39) was used to control the voltages and timing of the instrument and several key external pieces of equipment, such as the IR laser, the status of the quadrupole filter, and the trigger for the NanoMate 100 nanospray robot (Advion BioSciences). Typically, 10 μl of each sample (≈30 pmol total protein) was loaded into a 96-well sample plate with adhesive aluminum seals used to prevent evaporation.

Automated Deconvolution, Isolation, and Fragmentation of Intact Protein Ions. For the analysis of both RPs and cytosolic proteins, automated (partial) isolation and fragmentation was accomplished in one of two ways: a data-dependent stored waveform inverse Fourier transform isolation followed by IR laser fragmentation as described (30) or a quadrupole-based selection of two to five proteins in an ≈40-m/z range followed by multiplexed collisionally activated dissociation.

Stored Waveform Inverse Fourier Transform (SWIFT)/IR Multiphoton Dissociation. The samples introduced via automated nanospray were analyzed on-the-fly by using in-house deconvolution software. The top two most abundant proteins were then isolated by using an automatically generated SWIFT where two to five charge states of a given protein were selected (30). Fragmentation was accomplished by using an IR CO_2 laser (λ = 10.6 μm) set at two static irradiation times for each sample. The total experiment time for a given sample was ≈15 min.

Quadrupole "Marching"/Axial-Collisionally Activated Dissociation. For a given sample of proteins from the ribosomal preparation, four or five quadrupole mass filter windows were applied sequentially. These windows were ≈40 m/z wide, spanned a total of 100–150 m/z, and enhanced the dynamic range for protein detection to ≈2,000. For each window, fragmentation was accomplished by lowering the accumulation Oct2 offset voltage from −10 V to anywhere between −25 and −40 V. The *b* and *y* ions that resulted were from the fragmentation of everything in the 40-m/z window.

Data Analysis and Protein Identification. Spectra of intact proteins were analyzed by using deconvolution to determine average molecular weight (M_r) values. Fragmentation spectra were analyzed by using a modified THRASH algorithm (40). The resulting protein and fragment peak lists were searched with a ±1,000-Da window against a MySQL database of predicted *M. jannaschii* protein forms by using the RETRIEVER module within PROSIGHT PTM (https://prosightptm.scs.uiuc.edu) (32). Probability scores were calculated by using a Poisson-based statistical model (31). All fragment mass searches were done with an error tolerance of 30 ppm or less. Data from multiplexed fragmentation of protein mixtures were handled in a similar manner, except that when a protein was identified, its matching fragment masses were subtracted from the peak list, enabling identification of subsequent proteins with increased confidence (i.e., lower *P* scores).

Mass values are reported with an italicized number after them indicating the number of heavy isotopes in the most abundant isotopic peak (22).

Results and Discussion

RPs from *M. jannaschii*. After RPLC fractionation of a sample enriched for ribosomes (Fig. 1*A*), 12 fractions were analyzed by ESI/Q-FTMS with three representative samples yielding 6 to 10 intact protein M_r values <28 kDa (Figs. 1 *B* and *C* and 2*A*).

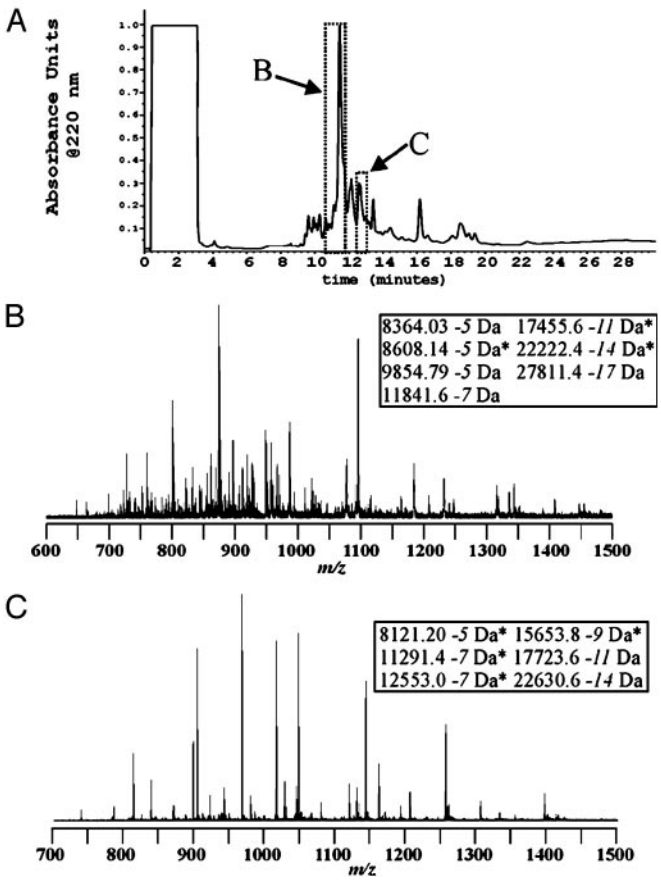

Fig. 1. (*A*) RPLC trace for RPs from *M. jannaschii*. (*B* and *C*) Mass spectra that correspond to samples from the labeled regions (B and C, respectively) in *A*. (*Insets*) Masses detected in these fractions. An asterisk denotes proteins identified unambiguously (see Tables 1 and 2). The masses listed are for the most abundant isotopic peak. The number in italics after the mass denotes the number of heavy isotopes in this most abundant species.

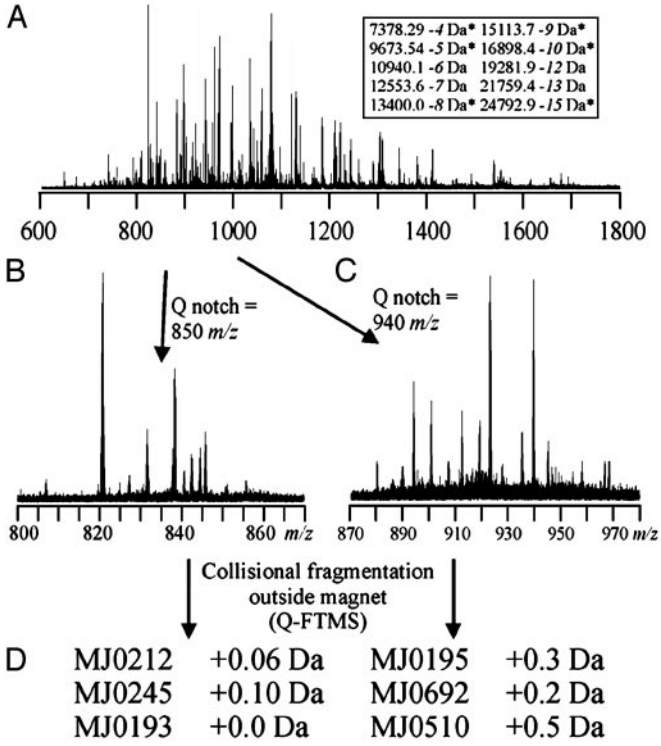

Fig. 2. (*A*) ESI/FT mass spectrum (100 scans) of a HPLC fraction from a separation of *M. jannaschii* RPs. (*B* and *C*) Quadrupole isolation (single scan) followed by axial-collisionally activated dissociation fragmentation for two quadrupole windows. (*D*) The six proteins positively identified from this experiment, also denoted by * in *A*.

Typical processing of such mixtures is illustrated in Fig. 2. For a "quad march" experiment (a portion of which is shown in Fig. 2 *B–D*), ions in ≈40-m/z sections of the spectrum were selectively accumulated (e.g., Fig. 2 *B* and *C*). Selected ions were dissociated by axial-collisionally activated dissociation as they entered an octopole after emerging from the notch-filtering quadrupole. The Fig. 2 mixture run in this manner revealed 10 distinct proteins, which ranged in size from 7378.29-*4* Da to 24,792.9-*15* Da (Fig. 2*A*). From four m/z windows, 34 b/y ions matched six RPs (Fig. 2*D*). Furthermore, each of the average M_r values matched the theoretical values to within 20 ppm. In general, the abundance of the fragment ions correlated well with the abundance of the intact protein ions when fragmentation was carried out in parallel.

The overall number of proteins identified in such experiments with ≥3 observed fragment ions (±25 ppm) was 26, including 8 proteins not predicted to be in the ribosome. None harbored covalent modifications apart from start methionine removal. In several of the ribosomal fractions, some of the M_r values detected were not identified through the multiplexed fragmentation. Manual tandem MS (MS/MS) of lower abundance species by quadrupole and/or stored waveform inverse Fourier transform isolation with IR dissociation was used to identify six more RPs and two non-RPs from these fractions. Of the 68 predicted proteins in the *M. jannaschii* ribosome, 24 were identified (Table 2, which is published as supporting information on the PNAS web site), with 18 of these predicted to be in the large subunit. Given the nonstandard behavior and undetermined purity of *M. jannaschii* ribosomes during isolation, the extent of sample coverage is unknown (percentage of detected vs. undetected RPs). Future comparison to bottom-up analyses of a similar preparation should illuminate the differences in performance. Such samples in this study proved beneficial for establishing the hardware and software for proteomewide analysis.

Proteomewide Analysis of WT Proteins. As a first dimension of fractionation, the ALS/PAGE step provided a basic ability to fractionate primarily by protein M_r values. Fractions with proteins in the 5- to 40-kDa range, determined by SDS/PAGE (Fig. 6*A*, which is published as supporting information on the PNAS web site), were carried forward. After surfactant removal, RPLC served to desalt and further fractionate the proteins. Subsequent analysis of these RPLC fractions by automated nanospray and ESI/FTMS yielded the Fig. 3*A* mass spectra, with 2 to 10 intact proteins observed per mixture. About 200 such mixtures were analyzed in this study, with IR fragmentation attempted on 2 proteins per sample. For fraction 4 in Fig. 3*A*, the automated data acquisition process is shown in Fig. 3 *B–E*. Isolation of the most abundant (Fig. 3*B*) and the next most abundant (Fig. 3*D*) proteins was followed by automated dissociation. Such MS/MS of the species isolated in Fig. 3*B* produced 55 distinct fragment ion masses, with 9 of these matching MJ0932 within ±30 ppm (Fig. 3*C*) for a P score of 2×10^{-5}. From MS/MS of the protein in Fig. 3*D*, 34 distinct fragment ion masses were observed with 12 of these matching b and y ions for MJ0595 within ±25 ppm (Fig. 3*E*, 6×10^{-13} P score). Fifty proteins were identified in this manner with P scores ranging from 10^{-2} to 10^{-16} (Tables 1 and 2).

Four proteins were found with significant mass shifts from

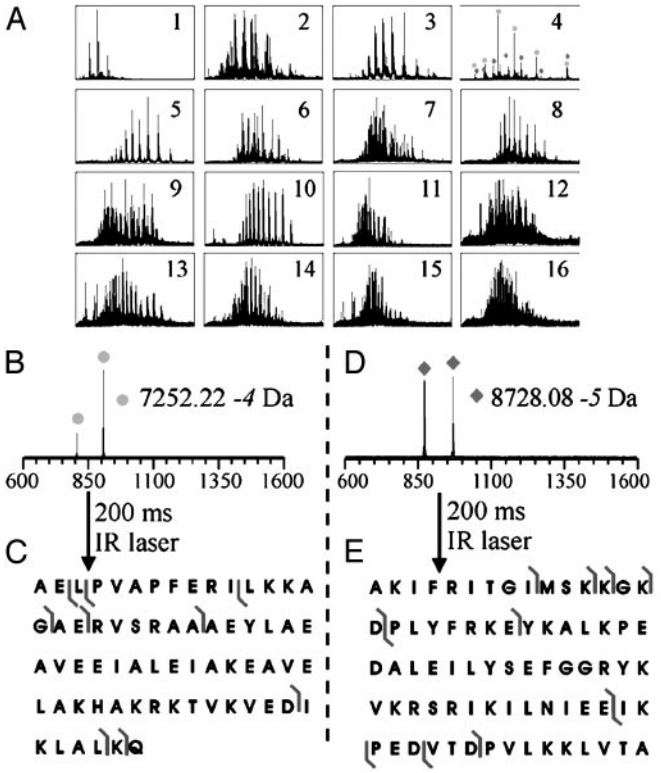

Fig. 3. (*A*) ESI/FT mass spectra of 16 fractions containing intact proteins from the 2D fractionation of a *M. jannaschii* whole-cell lysate. (*B* and *D*) Two automatically isolated species from fraction 4 in (*A Upper Right*). (*C* and *E*) Fragmentation maps of identified proteins from *B* and *D*, respectively, showing *b*- and *y*-type fragment ions.

those predicted—one internal oxidation, one probable acetylation (+42.1 Da), one incorrect start site, and one protein with both an incorrect start site and a +14-Da modification (described below). Some identified proteins contained mass shifts of either −2- or −4-Da from their predicted M_r values (Tables 1 and 2). One example is RP MJ0655, which was identified with a −4.1-Da mass shift. Initial studies of this protein and its fragmentation products revealed only three small fragment ion matches. When the search was performed in the "Δ*m*" mode of PROSIGHT PTM (which calculates and uses the mass difference between the observed and the theoretical M_r values of each candidate protein during the database query) (29), two more large *y* ions were assigned with the −4-Da shift. Manual inspection of the protein sequence revealed four total Cys residues, at residue numbers 36, 39, 73, and 76 (without the start Met). Five more fragments with −2-Da shifts were then found between residues 39 and 73, which indicated a disulfide linkage between Cys-36 and Cys-39 and another between Cys-73 and Cys-76. Five other disulfide linkages on four proteins were detected and localized in this manner.

Of the 72 proteins identified by MS/MS overall, 92% match their ORF-predicted M_r value within 50 ppm (with or without the start Met) and had no overlapping possibilities within 2 Da. With the difficulty involved in fragmenting and identifying every protein observed, 26 more putative identifications based solely on total molecular weight were assigned (Table 3, which is published as supporting information on the PNAS web site). About 70% of the 100 distinct proteins (many abundant proteins appeared multiple times in the 200 samples) for which MS/MS was executed yielded sufficient fragmentation for identification; the primary reason for unsuccessful identifications was insufficient intensity of fragment ions. Of those unambiguously identified and harboring PTMs, the degree of PTM localization ranged from 4 to 89 residues. The separation ability of the preparative gel for *M. jannaschii* proteins was compromised versus analogous work with the mesophilic yeast, *S. cerevisiae* (29). Further, the use of SDS instead of ALS shows improved size separation for *M. jannaschii* proteins (Fig. 6 *B* vs. *A*). This finding indicates that ALS may not fully denature many heat-resistant proteins (41), even when used in increased concentrations (0.2% ALS vs. 0.1% SDS). In all, only 18 of the 72 proteins identified in this study were also identified by bottom-up analysis of 2D gel spots (11), indicating that the ALS/PAGE/RPLC method gives complementary identifications while yielding far superior results for extensive characterization of primary structure.

Detection, Quantitation, and Localization of +14-Da Mass Discrepancy (Δ*m*). One ALS/PAGE/RPLC fraction gave the mixture spectrum of Fig. 4*A*. From the two protein forms isolated and fragmented together (20,219.7-*12* and 20,233.8-*12* Da, Fig. 4*B*), 55 fragment ions were detected, with 5 having a satellite peak of

Table 1. Selected proteins identified from *M. jannaschii*

Protein no.	Observed M_r*	Theoretical M_r†	Δ*m*	ppm error	*b*/*y* ions	*P* score‡	Function	Notes
MJ0168	7,248.28	7,248.15	0.13	18	6/3	1.54E-05	Archaeal histone A1	
MJ0932	7,248.22	7,248.11	0.11	15	6/3	1.54E-05	Archaeal histone A2	
MJ1258	7,108.76	7,109.02	−0.26	−37	1/2	3.09E-03	Archaeal histone A3	
MJ1338	36,841.9 (ave)	36,843.7 (ave)	−1.8	−49	3/7	1.11E-11	H (2)-dependent methylenetetrahydromethan-opterin dehydrogenase-related protein	Possible disulfide bond: Cys^{113}–Cys^{116}
MJ1540§	13,855.9	13,813.8	42.1		1/3	2.20E-03	Hypothetical protein	Internal acetylation incorrect start Met
MJ0556	20,207.7	21,677.6	−1,469.9		5/6	5.87E-05	Hypothetical protein	Putative methylation near N-term
MJ0655§	10,381.7	10,385.8	−4.1		2/8	6.50E-08	50S ribosomal protein L34e	Cys-36–Cys-39 and Cys-73–Cys-76 disulfide bonds
MJ0896	15,189.0	15,596.4	−407.4		1/8	2.41E-05	Putative flagella-related protein E	Incorrect start Met

*Masses are monoisotopic unless specified otherwise.
†The listed theoretical mass includes +71-Da cysteine alkylations for all proteins identified by using the ALS/PAGE separation.
‡*P* score calculated with the Poisson-based model (31).
§These proteins were identified with multiplexed data (crude ribosomal preparation).

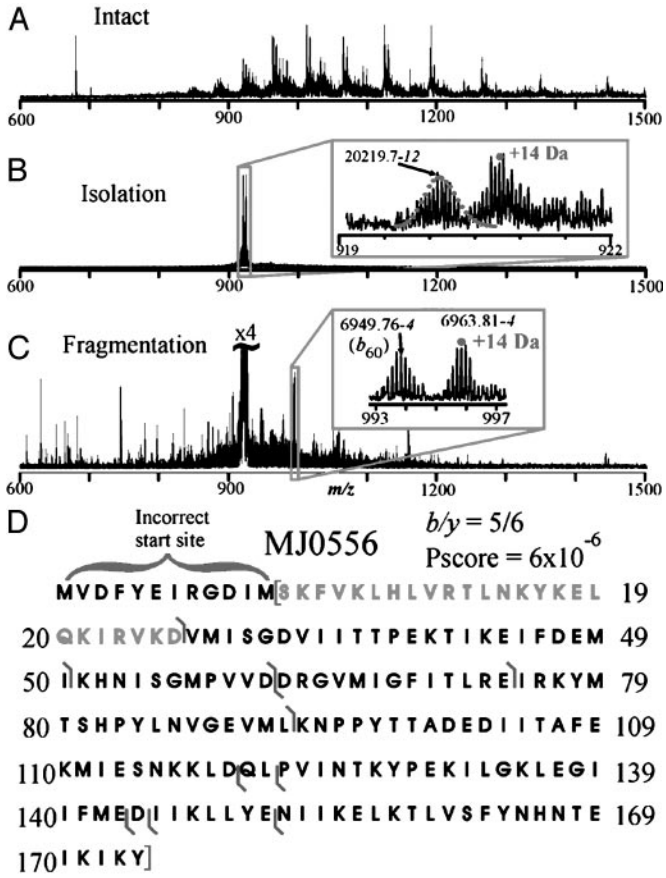

Fig. 4. (*A*) Broadband ESI/FTMS of an ALS/PAGE/RPLC fraction (20 scans). (*B*) Isolation of a 20219.7-*12*-Da species and another of slightly higher mass (five scans). (*Inset*) Expansion of the *m*/*z* scale for the intact ions isolated prior to MS/MS. (*C*) Fragmentation of the ions in *B* by IR photons (200 scans). (*Inset*) b_{60} fragment ion with a +14-Da satellite peak. (*D*) Graphical fragment map and database retrieval metrics for identification of MJ0556. The +14-Da mass shift was localized to the sequence in bold (Ser-2–Asp-27).

+ 14.04 ± 0.02 Da (Fig. 4*C Inset*). A database query returned MJ0556 with a *P* score of 7×10^{-3} (±2,000 Da), made poor because only *y* ions matched the predicted sequence. Closer inspection of putative *b*-type ions positioned the N terminus 13 aa from the predicted start Met. The data are most consistent with an incorrect prediction of the translation start site and posttranslational loss of the true start methionine (Fig. 4*D*). The smallest of these *b* ions (b_{25}) also had a +14-Da satellite peak of equal relative abundance. Further MS/MS experiments are required to precisely localize this Δm. Overall, the data indicate a methylated form (+14.02 Da, theoretical) of MJ0556 present at 50% relative abundance in stationary-phase cultures of *M. jannaschii*. No other proteins have shown +14-Da adducts, arguing strongly that this Δm is biologically relevant. Further, a homology search of this protein revealed a putative cystathionine β-synthase domain, which is involved in regulation by *S*-adenosyl methionine, a cofactor that donates methyl groups to a wide variety of substrates.

Efficient Comparison of Histone Modifications Across Species. Histone proteins from *M. jannaschii* gave abundant signals and high confidence *P* scores during this study. Evidence for expression of only three of the four predicted histones was accumulated, including the A1 and A2 pair with only a K → Q switch at their C termini. This switch creates a 0.036-Da Δm, in fact resolved for small *y*-type ions such as the y_7 (Fig. 5*A Inset*). From none of the

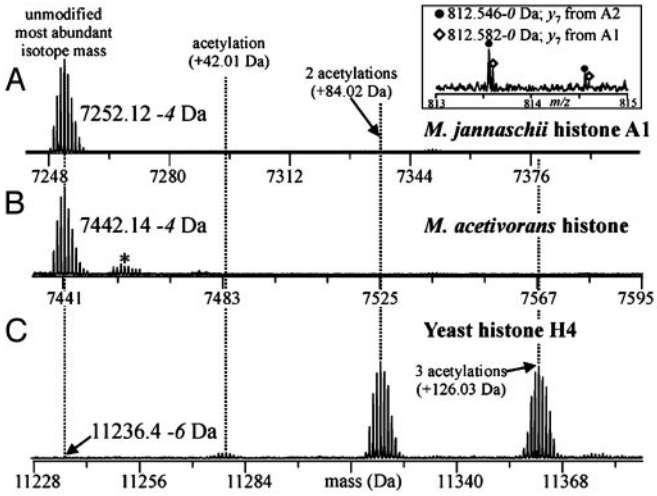

Fig. 5. (*A*) Comparison of core histone proteins from *M. jannaschii* (histone A1 from a 1.8-Mb genome, stationary phase cells. (*Inset*) Blow-up of the y_7 ions from fragmentation of a mixture of histone A1 and A2 (which differ by a K → Q switch at the C terminus). (*B*) Comparison of core histone proteins from *M. acetivorans* (histone from a 5.8-Mb genome, midlog phase cells). The isotopic cluster marked by * is an oxidation (+16 Da) of the base peak. (*C*) Comparison of core histone proteins from *S. cerevisiae* (histone H4 from a 12-Mb genome, stationary phase cells).

growth stages of this organism (34) were any of the histone proteins detected to harbor PTMs, consistent with a complete lack of histone-modifying enzymes in the *M. jannaschii* genome (13) and of histone tails (known to be PTM rich in Eukarya) (42). This finding is in stark contrast to eukaryotic histones such as those from histone H4 isolated from *S. cerevisiae*. Targeted analysis of histone H4 from ALS/PAGE samples created from yeast allowed directed detection of H4 protein forms with one to five acetylations (Fig. 5*C*). The data of Fig. 5*C* validate the proposal (30) that multiply modified forms largely coelute through the entire 2D ALS/PAGE/RPLC fractionation. A detailed study of the PTM dynamics present on yeast histones is beyond the scope of this study.

For *M. jannaschii*, with no histone-modifying enzymes, no "histone code" (43) involving modifications appears operative, making regulation by differential histone expression or changes in dimerization partner more likely (42). However, *M. acetivorans* has predicted histone-modifying enzymes, such as a histone acetyl transferase (Q8TSH0) and a histone deacetylase (HDAC, Q8TLY4) detectable by homology searching. Only one true histone is predicted from its genome, yet it clearly shows no N-terminal histone tail. The possibility of acetylations within the C-terminal core (44) and the strong HDAC homology (*E* value of 3×10^{-70}) warrant targeted study into the covalent state of the WT histone.

Application of the platform here in a hypothesis-driven mode (33) was initiated to search for the putatively acetylated histone from the *M. acetivorans* proteome (4528 predicted ORFs) (45). The following results were obtained from one experiment using 3 ml of cell lysate. Fractionation by ALS/PAGE of stationary-phase *M. acetivorans* cells grown at 37°C was of significantly better resolution versus that obtained with lysates from *M. jannaschii* (data not shown). From one ALS/PAGE fraction with ≈5- to 10-kDa proteins, a 7442.21-*4*-Da protein matched the M_r value of the unmodified target with its start Met removed (Fig. 5*B*). Isolation and fragmentation confirmed its identification with 11 matching fragment ions within 30 ppm (2×10^{-9} *P* score, data not shown). It is clearly seen from the spectra in Fig. 5*B* that no modifications other than a small amount of oxidation

(+16 Da) are present. Treatment with sodium butyrate, a nonspecific histone deacetylase inhibitor (36), for 4 h and harvesting at midlog phase followed by targeted detection by ALS/PAGE/RPLC and Q-FTMS provided no evidence for mono- or di-acetylation. This lack of histone acetylations indicates that either the vast excess of nucleosomes (>3,000:1) are not acetylated or that *M. acetivorans* (and all methanogenic Archaea) simply do not use any histone modifications to package their DNA or engage transcriptional machinery. Thus, the combination of the proteome fractionation with the quadrupole enhancement to FTMS enables facile detection of specific proteins for biochemical characterization from very complex mixtures of WT proteins, provided the copy number of the target is high enough.

Conclusion

The platform described here is capable of unattended interrogation of intact proteins by high-resolution MS/MS. Moderately complex samples of 2 to 10 proteins were analyzed, and 2 to 6 proteins per sample were reliably fragmented and identified. The combination of separations, Q-FTMS hardware, and new software will continue its evolution toward benchmarks set by peptide-based proteome analysis. Although PROSIGHT PTM was developed for true MS/MS experiments (isolation of a single protein), the data presented here show that it also functions well for proteins fragmented together. Future implementation of this platform on a larger scale promises to reveal the extent to which common and unique PTMs are present on proteins from thermophilic and methanogenic Archaea. The ability to specifically target a protein for PTM information based on hypotheses from genomic information will accelerate understanding of PTM regulatory roles in diverse organisms as technology for MS/MS at high mass reaches more laboratories.

We thank Biswarup Mukhopadhyay and William Metcalf for expert culturing of *M. jannaschii* and *M. acetivorans*, respectively, and Claudia Reich for the protocol for partial isolation of *M. jannaschii* ribosomes. This work was supported by the University of Illinois, National Institutes of Health Grant GM 067193, the Burroughs Wellcome Fund, the Searle Scholars Program, and the Packard Foundation.

1. Selinger, D. W., Cheung, K. J., Mei, R., Johansson, E. M., Richmond, C. S., Blattner, F. R., Lockhart, D. J. & Church, G. M. (2000) *Nat. Biotechnol.* **18,** 1262–1268.
2. Washburn, M. P., Wolters, D. & Yates, J. R. (2001) *Nat. Biotechnol.* **19,** 242–247.
3. Lipton, M. S., Pasa-Tolic, L., Anderson, G. A., Anderson, D. J., Auberry, D. L., Battista, K. R., Daly, M. J., Fredrickson, J., Hixson, K. K., Kostandarithes, H., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 11049–11054.
4. Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzeler, E. & Yates, J. R., III (2003) *Proc. Natl. Acad. Sci. USA* **100,** 3107–3112.
5. Mann, M. & Jensen, O. N. (2003) *Nat. Biotechnol.* **21,** 255–261.
6. Schweppe, R. E., Haydon, C. E., Lewis, T. S., Resing, K, A. & Ahn, N. G. (2003) *Acc. Chem. Res.* **36**, 453–461.
7. Davidsson, P., Brinkmalm, A., Karlsson, G., Persson, R., Lindbjer, M., Puchades, M., Folkesson, S., Paulson, L., Dahl, A., Rymo, L., *et al.* (2003) *Cell. Mol. Biol.* **49,** 681–688.
8. Wolters, D. A., Washburn, M. P. & Yates, J. R. (2001) *Anal. Chem.* **73,** 5683–5690.
9. Anderson, N. G. & Anderson, N. L. (1996) *Electrophoresis* **17,** 443–453.
10. Pandey, A. & Mann, M. (2000) *Nature* **405,** 837–846.
11. Giometti, C. S., Reich, C., Tollaksen, S., Babnigg, G., Lim, H., Zhu, W., Yates, J., III, & Olsen, G. (2002) *J. Chromatogr. B* **782,** 227–243.
12. Mukhopadhyay, B., Johnson, E. F. & Wolfe, R. S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 11522–11527.
13. Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. & Wolfe, R. S. (1983) *Arch. Microbiol.* **136,** 254–261.
14. Bult, C. J., White, O., Olsen, G. J., Zhou, L. X., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996) *Science* **273,** 1058–1073.
15. Selmer, T., Kahnt, J., Goubeaud, M., Shima, S., Grabarse, W., Ermler, U. & Thauer, R. K. (2000) *J. Biol. Chem.* **275,** 3755–3760.
16. Konrad, Z. & Eichler, J. (2002) *Biochem. J.* **366,** 959–964.
17. Giometti, C. S., Reich, C. I., Tollaksen, S. L., Babnigg, G., Lim, H., Yates, J. R., III, & Olsen, G. (2001) *J. Proteomics* **1,** 1033–1042.
18. Oda, Y., Nagasu, T. & Chait, B. T. (2001) *Nat. Biotechnol.* **19,** 379–382.
19. Conrads, T. P., Issaq, H. J. & Veenstra, T. D. (2002) *Biochem. Biophys. Res. Commun.* **290,** 885–890.
20. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F. & White, F. M. (2002) *Nat. Biotechnol.* **20,** 301–305.
21. MacCoss, M. J., McDonald, W. H., Saraf, A., Sadygov, R., Clark, J. M., Tasto, J. J., Gould, K. L., Wolters, D., Washburn, M., Weiss, A., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 7900–7905.
22. Zhu, K., Kim, J., Yoo, C., Miller, F. R. & Lubman, D. M. (2003) *Anal. Chem.* **75,** 6209–6217.
23. Kelleher N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K. & McLafferty F. W. (1999) *J. Am. Chem. Soc.* **121,** 806–812.
24. Reid, G. E., Shang, H., Hogan, J. M., Lee, G. U. & McLuckey, S. A. (2002) *J. Am. Chem. Soc.* **124,** 7353–7362.
25. Zubarev, R. A., Kruger, N. A., Fridriksson, E. K., Lewis, M. A., Horn D. M., Carpenter, B. K. & McLafferty, F. W. (1999) *J. Am. Chem. Soc.* **121,** 2857–2862.
26. Ge, Y., ElNaggar, M., Sze, S. K., Bin Oh, H., Begley, T. P., McLafferty, F. W., Boshoff, H. & Barry, T. P. (2003) *J. Am. Soc. Mass Spectrom.* **14,** 253–261.
27. Lee, S. W., Berger, S. J., Martinovic, S., Pasa-Tolic, L., Anderson, G. A., Shen, Y. F., Zhao, R. & Smith, R. D. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 5942–5947.
28. Pan, P. & McLuckey, S. A. (2003) *Anal. Chem.* **75,** 1491–1499.
29. Meng, F. Y., Cargile, B. J., Patrie, S. M., Johnson, J. R., McLoughlin, S. M. & Kelleher, N. L. (2002) *Anal. Chem.* **74,** 2923–2929.
30. Johnson, J. R., Meng, F. Y., Forbes, A. J., Cargile, B. J. & Kelleher, N. L. (2002) *Electrophoresis* **23,** 3217–3223.
31. Meng, F. Y., Cargile, B. J., Miller, L. M., Forbes, A. J., Johnson, J. R. & Kelleher, N. L. (2001) *Nat. Biotechnol.* **19,** 952–957.
32. Taylor, G. K., Kim, Y.-B., Forbes, A. J., Meng, F., McCarthy, R. & Kelleher, N. L. (2003) *Anal. Chem.* **75,** 4081–4086.
33. Kalkum, M., Lyon, G. J. & Chait, B. T. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2795–2800.
34. Mukhopadhyay, B., Johnson, E. F. & Wolfe, R. S. (1999) *Appl. Environ. Microbiol.* **65,** 5059–5065.
35. Metcalf, W. W., Zhang, J. K., Shi, X. & Wolfe, R. S. (1996) *J. Bacteriol.* **178,** 5797–5802.
36. Boffa, L. C., Vidal, G., Mann, R. S. & Allfrey, V. G. (1978) *J. Biol. Chem.* **253,** 3364–3366.
37. Schäfer, M., Schmuck, C., Heil, M., Cooper, H. J., Hendrickson, C. L., Chalmers, M. J. & Marshall, A. G. (2003) *J. Am. Soc. Mass Spectrom.*, **14,** 1282–1289.
38. Wilcox, B. E., Hendrickson, C. L. & Marshall, A. G. (2002) *J. Am. Soc. Mass Spectrom.* **13,** 1304–1312.
39. Senko, M. W., Canterbury, J. D., Guan, S. H. & Marshall, A. G. (1996) *Rapid Commun. Mass Spectrom.* **10,** 1839–1844.
40. Horn, D. M., Zubarev, R. A. & McLafferty, F. W. (2000) *J. Am. Soc. Mass Spectrom.* **11,** 320–332.
41. Cowan, D. A. (1997) *Comp. Biochem. Physiol. A Physiol.* **188,** 429–438.
42. Reeve, J. N. (2003) *Mol. Microbiol.* **48,** 587–598.
43. Jenuwein, T. & Allis, C. D. (2001) *Science* **293,** 1074–1080.
44. Zhang, L., Engeni, E. E., Parthun, M. R. & Freitas, M. A. (2003) *Chromasoma* **112,** 77–86.
45. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., *et al.* (2002) *Genome Res.* **12,** 532–542.