Characterization of the 70S Ribosome from Rhodopseudomonas p alustris Using an Integrated "Top-Down" and "Bottom-Up" Mass Spectrometric Approach

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · AUGUST 2004

Impact Factor: 4.25 · DOI: 10.1021/pr049940z · Source: PubMed

11 AUTHORS, INCLUDING:

Michael Brad Strader
U.S. Food and Drug Administration
26 PUBLICATIONS **490** CITATIONS
SEE PROFILE

David L Tabb
Vanderbilt University
97 PUBLICATIONS **6,987** CITATIONS
SEE PROFILE

Barry D Bruce
University of Tennessee
147 PUBLICATIONS **3,594** CITATIONS
SEE PROFILE

Robert L Hettich
Oak Ridge National Laboratory
252 PUBLICATIONS **6,907** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Characterization of the 70S Ribosome from *Rhodopseudomonas palustris* Using an Integrated "Top-Down" and "Bottom-Up" Mass Spectrometric Approach

Michael Brad Strader,[†] Nathan C. VerBerkmoes,[†,‡] David L. Tabb,[§] Heather M. Connelly,[†,‡] John W. Barton,[§] Barry D. Bruce,[||] Dale A. Pelletier,[§] Brian H. Davison,[§] Robert L. Hettich,[†,‡] Frank W. Larimer,[§] and Gregory B. Hurst*[,†]

Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6131, Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge Tennessee 37831, Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, Tennessee 37830-8026, and Biochemistry Cellular and Molecular Biology Department, University of Tennessee, Knoxville, Tennessee 37996

Received March 3, 2004

We present a comprehensive mass spectrometric approach that integrates intact protein molecular mass measurement ("top-down") and proteolytic fragment identification ("bottom-up") to characterize the 70S ribosome from *Rhodopseudomonas palustris*. Forty-two intact protein identifications were obtained by the top-down approach and 53 out of the 54 orthologs to *Escherichia coli* ribosomal proteins were identified from bottom-up analysis. This integrated approach simplified the assignment of post-translational modifications by increasing the confidence of identifications, distinguishing between isoforms, and identifying the amino acid positions at which particular post-translational modifications occurred. Our combined mass spectrometry data also allowed us to check and validate the gene annotations for three ribosomal proteins predicted to possess extended C-termini. In particular, we identified a highly repetitive C-terminal "alanine tail" on L25. This type of low complexity sequence, common to eukaryotic proteins, has previously not been reported in prokaryotic proteins. To our knowledge, this is the most comprehensive protein complex analysis to date that integrates two MS techniques.

Keywords: bottom-up • top-down • *Rhodopseudomonas palustris* • protein complexes • 70S ribosome • post-translational modifications • proteomics • mass spectrometry

Introduction

The development of strategies for identifying and characterizing multiprotein complexes on a genome-wide scale has drawn increasing attention. Mass spectrometry (MS) has been featured in many of these strategies because of the rapid growth of publicly available genome sequences, improved proteome bioinformatics, soft ionization methods, and the development of high-throughput MS based protein identification techniques.[1−8] Two important mass spectrometry methods used in identifying protein components of macromolecular complexes are the frequently applied "bottom-up" and the less widely used "top-down" strategies.

The bottom-up strategy involves enzymatic digestion of intact proteins to generate peptides that are analyzed by the mass spectrometer. Two recent reports involving the large-scale analysis of hundreds of protein complexes from yeast illustrate the potential of this technique.[9,10] One implementation of the bottom-up strategy for analyzing large protein complexes employs 1D or 2D gel electrophoresis followed by in-gel trypsin digestion and either peptide mass fingerprinting or tandem mass spectrometry (MS−MS) to identify protein components.[10−15] A higher throughput approach involves proteolytic digestion of the entire complex in solution, followed by one-dimensional (1D) or two-dimensional (2D) liquid chromatography coupled with electrospray ionization (ESI) MS−MS to generate peptide fragmentation spectra that can then be compared to translated genomic databases.[6,7,16] This "shotgun" approach is effective in comprehensively analyzing large macromolecular complexes by identifying large numbers of proteins in a single data acquisition.[17]

The bottom-up approach has also been applied to identifying post-translational modifications (PTMs).[18,19] Unfortunately, when using the bottom-up approach alone, a particular PTM

* To whom correspondence should be addressed. Phone: (865) 574−6142. Fax: (865) 576−8559. E-mail: hurstgb@ornl.gov.
[†] Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory.
[‡] Graduate School of Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratory.
[§] Life Sciences Division, Oak Ridge National Laboratory.
[||] Biochemistry Cellular and Molecular Biology Department, University of Tennessee.

10.1021/pr049940z CCC: $27.50

Published on Web 08/14/2004

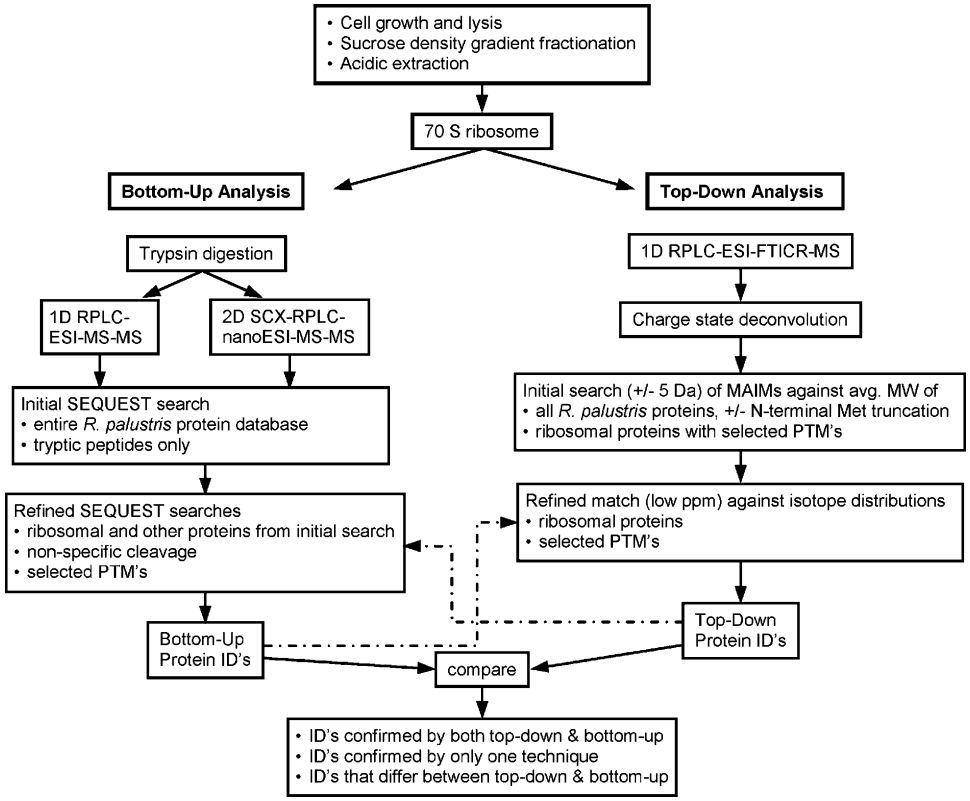

Figure 1. Strategy for top-down, bottom-up MS analysis of ribosomal proteins. Integration of results from the two approaches was achieved, as the dashed arrows show, by iteratively using the results from each approach to augment and expand the results of the other.

will be overlooked if the peptides bearing that modification escape detection. Furthermore, identifying modified peptides from a complex mixture does not provide information concerning different isoforms that may exist for a particular protein.

The top-down strategy involves MS or MS−MS of intact proteins.[20−23] Measuring the intact protein provides the advantage of potentially identifying translational start and stop sites, mRNA splice variants, and post-translational modifications (PTMs) of expressed gene products. To date, this method has predominantly been focused on characterizing intact protein molecular forms, including PTMs that result in isoforms of related proteins from either a macromolecular complex or crude protein mixture.[24,25] Although a lack of bioinformatics tools for data analysis and verification of PTMs has limited the general utility of this approach, the method is developing rapidly.[25,26]

Because of the unique capabilities each MS technique provides, a combination of bottom-up and top-down approaches would allow more complete characterization by potentially identifying the intact protein mass corresponding to the presence of a particular PTM, the modification position(s), and isoforms. Furthermore, a correlation between the two methods improves the certainty of the PTM identity and location.[26−30]

In previous studies of the *Shewanella oneidensis* proteome, we demonstrated the potential of integrating the bottom-up and top-down approaches.[31] Currently, we are extending this top-down, bottom-up (TDBU) approach to investigate the proteome and network of protein complexes from the organism *Rhodopseudomonas palustris*.[32] This characterization includes the identification of components of protein complexes as well as their PTMs. Figure 1 illustrates the strategy for the TDBU approach adopted in this study. Integration of results was achieved, as shown by the dotted-line arrows in Figure 1, by using protein identifications from analysis of top-down data to refine analysis of bottom-up data, and vice versa, in an iterative manner to increase the number of characterizations of ribosomal proteins obtained. For example, identification of a methylated protein by the top-down approach could provide motivation to examine more closely the bottom-up results for the presence of a methylated peptide from that protein.

R. palustris, a purple nonsulfur anoxygenic phototrophic bacterium, is found ubiquitously in soil and aqueous environments, where it can survive under a variety of growth conditions due to its high metabolic diversity. This microbe can grow in the presence or absence of oxygen, and can respond to environmental changes by converting among the four major life supporting modes of metabolism—photoheterotrophic (energy from light and carbon from organic compounds), photoautotrophic (energy from light and carbon from CO_2), chemoheterotrophic (carbon and energy from organic compounds), and chemoautotrophic (energy from inorganic compounds and carbon from CO_2).[33] *R. palustris* is a potential biofuel producer with the capability of producing hydrogen gas, and it can also degrade biomass such as the aromatic hydrocarbons in lignin monomers.[33−36] The genome of this microbe has been fully sequenced and annotated.[37] To better understand the network of complexes composing this diverse web of metabolic interactions, a detailed understanding of the individual complexes is needed. This paper describes a TDBU mass spectrometric analysis of the analog in *R. palustris* of one

of the most highly conserved and well studied protein complexes—the 70S ribosome.

The ribosome is the universal macromolecular machine involved in translating the genetic code into proteins. Bacterial ribosomes are composed of a small subunit (30S) containing about 20 proteins and a single rRNA (16S), and a large subunit (50S) consisting of over 30 proteins and two rRNAs (23S and 5S).[38,39] Crystallographic structures for different prokaryotic ribosomes indicate remarkable similarity in overall structure and composition, particularly in the rRNA folds where peptide bond synthesis occurs.[40−43] The ribosome from *Escherichia coli* is the most extensively characterized of the bacterial ribosomes. Ribosomes from bacterial species studied so far exhibit most if not all the homologues to ribosomal proteins found in *E. coli*.[38] Furthermore, PTMs of ribosomal proteins from other bacteria, and from eukaryotic-cell organelles thought to have evolved from bacteria by endosymbiosis, tend to be similar to PTMs of *E. coli* proteins, with variations in the corresponding modification positions.[38,44,45]

We present the TDBU characterization of the ribosome from *R. palustris*. Fifty-three orthologs to the 54 *E. coli* ribosomal proteins were identified from bottom-up MS analysis, and 42 intact protein identifications were obtained by the top-down approach. The integrated TDBU approach simplified PTM assignments by increasing the confidence of identifications, distinguishing between isoforms of ribosomal proteins and identifying the amino acid positions at which a particular PTM occurred. We also identified three ribosomal proteins with unusual extended sequences (including a low complexity highly repetitive "alanine tail" at the C-terminus of L25) by a combination of genomic database searches and manual evaluation of the MS data. The TDBU approach allowed a more comprehensive characterization of the ribosome than either technique alone. A study of the 40S subunit from rat fibroblast ribosomes employed both intact mass and peptide mass spectrometric measurements, although the former were obtained at lower resolution than described in the present paper.[27]

Materials and Methods

Materials. All salts, buffers, dithiothreitol (DTT), diethyl pyrocarbonate (DEPC), guanidine HCl, trifluoroacetic acid, glacial acetic acid, sucrose, and RNase-free DNase I, were obtained from Sigma Chemical Co. (St. Louis, MO). In addition to using DEPC treated water to make buffers, RNase Away (Molecular BioProducts, San Diego, CA) was also used to treat labware and benchtop surfaces to minimize RNase activity during the ribosome purification procedure. Sequencing-grade trypsin was purchased from Promega (Madison, WI). Formic acid was obtained from EM Science (Gibbstown, NJ). HPLC grade acetonitrile and water were used for all LC−MS analysis (Burdick & Jackson, Muskegon, MI). Ultrapure 18 MΩ water used for sample buffers was obtained from a Milli-Q system (Millipore, Bedford, MA). BCA assay reagent and standards were obtained from Pierce Chemical Co. (Rockford, IL). Fused silica was purchased from Polymicro Technologies (Phoenix, AZ).

Cell Growth and Preparation of 70S Ribosomes. The wild-type strain *Rhodopseudomonas palustris* CGA009 (a gift from Caroline Harwood, Department of Microbiology, University of Iowa), was grown either aerobically or anaerobically in a glass walled fermentation vessel (Biostat B, B. Braun Biotech, Allentown, PA).[46] Briefly, aerobic growth conditions, with air injected through the bottom of fermentation vessel, required media supplemented with 10 mM succinate (carbon source) without illumination (to eliminate photosynthesis). Anaerobic growth conditions required 10 mM succinate with the additional requirement of illumination and exclusion of air. All fermentations were run at 30° C at pH 6.8. Cells were harvested at mid-log growth phase ($O.D_{660}$ of ∼0.8), and washed twice in ice-cold French Press buffer (100 mM ammonium chloride, 50 mM magnesium acetate, 20 mM Tris-HCl (pH 7.5), 1.0 mM DTT, 0.5 mM EDTA). After resuspending cells in the same buffer, a French Pressure cell (Thermo Spectronic, Madison, WI) was used to disrupt cells by applying 16 000 psi 3× for 1 min. DNase I was added to the resultant suspension to degrade contaminant DNA for subsequent removal. Cellular debris was removed by centrifuging the lysate twice at 30 000 × *g* in a SS-34 Sorval rotor for 30 min at 4 °C. The collected supernatant was then quick-frozen with liquid nitrogen and stored at −80 °C.

To separate 70S ribosomes initially from the remaining cellular components, the supernatant was layered at a 1:1 ratio (wt/wt) over a high salt sucrose cushion (20 mM Tris-HCl, pH 7.5, 50 mM magnesium acetate, 100 mM ammonium chloride, 1 mM DTT, 0.5 mM EDTA, 1.1 M sucrose) and centrifuged at 100 000 × *g* in a Ti60 Beckman rotor for 16 h at 4 °C. The ribosomal pellet was resuspended in a small volume (1−3 mL) of French Press buffer, aliquoted and stored at −80 °C for further use.

70S ribosomes were further purified and fractionated using sucrose density fractionation.[47] Briefly, samples were layered on top of a 7−30% linear sucrose gradient (10 mM Tris-HCl, pH 7.5, 6 mM magnesium acetate, 50 mM ammonium chloride, 1mM DTT, 0.5 mM EDTA) and centrifuged at 85 000 × *g* for 4 h. After centrifugation, the gradients were fractionated and the absorbance at 260 nm was used to identify fractions containing ribosomes. Fractionated ribosomes were then pooled and recovered by centrifugation at 100 000 × *g* for 16 h.

Ribosomal protein extraction and the removal of contaminant rRNA was performed using the acid extraction method.[48] The resuspended ribosomes were combined with 0.1 volume of 1 M magnesium chloride, then with 2 volumes of glacial acetic acid, and mixed by inversion for 2 h at 4 °C. The insoluble fraction containing the contaminant rRNA was removed by centrifugation at 17 000 × *g* for 30 min at 4 °C. After overnight dialysis in a 3500 MWCO dialysis cassette (Slide-A-Lyzer, Pierce, Rockford, IL) against Ultrapure water, the protein samples were quantitated using the BCA assay according to the manufacturer's instructions.

Electrospray Fourier Transform Mass Spectrometry for "Top-Down" Proteomic Analysis. High-resolution mass spectra were acquired using an Ultimate HPLC (LC Packings/Dionex, Sunnyvale, CA) coupled to a 9.4 T HiRes electrospray Fourier transform ion cyclotron resonance mass spectrometer, ESI−FTICR MS (IonSpec, Lake Forest, CA). The HPLC flow rate was 4 μL/min with a 60 min linear gradient from 100% solvent A (95% H_2O/5% acetonitrile (ACN)/0.5% formic acid) to 100% solvent B (5% H_2O/95% ACN/0.5% formic acid). A C4 reverse phase column (model 214MS5.325, 300 μm i.d. × 15 cm, 300 Å pore size, 5 μm particles, Grace-Vydac, Hesperia, CA) was directly connected to the Analytica electrospray source with 100 μm i.d. fused silica capillary tubing. Ions were generated with a 3700 V potential between a grounded needle and heated transfer capillary, accumulated in an external hexapole for 2s, transferred into a high-vacuum region using a quadrupole lens system, and then detected in the mass analyzer.[49,50] A broad-band mass resolution of at least 50 000 (full width at half-

maximum) at m/z 1000 was possible because ion detection was achieved in an ultrahigh vacuum regime ($\sim 2 \times 10^{-10}$ Torr). Standard proteins (ubiquitin, myoglobin) and peptides (leucine enkephalin, gramicidin S) were used for mass calibration. The high-resolution mass measurement enabled isotopic resolution of multiply charged ions. The charge state of a multiply charged ion could, therefore, be determined by directly measuring its isotopic spacing.[51] Deconvoluted molecular mass spectra were generated with the IonSpec software. By calibrating on the calculated values of the most abundant isotopic peaks for the six different charge states (7^+ to 12^+) of the protein standard ubiquitin, the deconvoluted mass spectrum yielded a measured molecular mass that was within 0.025 Da (3 ppm) of the calculated value. The external calibration procedure enabled molecular measurement accuracy of ≤10 ppm for most proteins with molecular masses up to 40 kDa, although the mass errors were slightly larger for the more minor abundance proteins, for which the isotopic packets are somewhat distorted.

Because the mass resolution was at least 50 000 for the intact protein measurements, the molecular masses of these proteins could be measured with isotopic resolution. The measured most abundant isotopic mass (MAIM) of each molecular ion region was used as an approximation of the protein's isotopically averaged molecular mass in order to query a database of all possible *R. palustris* proteins (provided as Supporting Information; current version available at http://genome.ornl.gov/microbial/rpal/).[37] Although least-squares fitting algorithms such as THRASH have been developed to extract molecular mass information from isotopic clusters,[52] the MAIM value is easily obtainable from the measured mass spectra and may be superior to the value generated by THRASH for lower abundance peaks, since it is not affected by the somewhat distorted isotopic distributions of the minor species. This database query with the MAIM values was conducted with a reasonably large molecular mass tolerance window (± 5 Da) to accommodate the fact that the abundances (but not the mass values) of the ions in the *measured* isotopic packet may vary somewhat from their *calculated* values. This is especially noticeable in the larger proteins, where the abundances of the isotopomers around the average molecular mass are very similar. Even slight variations in the mass spectrometric measurements can result in peak abundance variations of a few percent, which can alter the most abundant isotope observed in these cases. This search usually revealed between 1 and 4 possible protein matches within the "crude" 5-Da window, with a close match to at least one protein in the database. Calculated masses for both intact proteins and proteins with N-terminal methionine truncation for all possible *R. palustris* proteins were searched in this initial screen. To refine a tentative protein match, the isotopically resolved molecular mass region of the suspected protein was calculated based on its sequence and compared to the measured data from the FTICR−MS experiment. Because these experiments were conducted under external calibration conditions, the mass of the most abundant isotopic peak for each matched protein from the database was required to be within 10 ppm (i.e., a few millidaltons) of the measured value for the more abundant signals, with somewhat lower mass accuracy (less than 30 ppm) permitted for more minor species. For the entire suite of 54 possible ribosomal proteins, an intact protein look-up table was extracted from the full *R. palustris* protein database; this intact protein table contained intact molecular masses, methionine-truncated molecular masses, and all possible combinations of methionine truncation with single acetylation and multiple methylations (up to 9). The experimental FTICR−MS data were used to query this look-up table for tentative PTM protein forms. All possible matches were compared against the results obtained from the bottom-up data, as described below.

LC−MS−MS for "Bottom-Up" Proteomic Analysis. All samples to be analyzed by the "bottom-up" approach were first digested with trypsin following the manufacturer's protocol and then desalted using C_{18} reverse-phase extraction (Sep-Pak, Waters, Milford, MA). Samples were then concentrated to ~0.1−1 μg/μL in a vacuum centrifuge (Savant Instruments, Holbrook, NY) and filtered with a 0.45 μm Ultrafree-MC filter (Millipore, Bedford, MA). Final peptide samples to be injected were in 100% H_2O with either 0.1% TFA (1D LC−MS−MS) or 0.1% formic acid (2D LC−MS−MS).

One-dimensional (1D) capillary LC−MS−MS experiments were performed with an Ultimate HPLC coupled to an LCQ-DECA or LCQ-DECA XP Plus quadrupole ion trap mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with an electrospray source. Injections of typically 10−20 μg peptide digest were made using a Famos (LC Packings) autosampler with a 50 μL loop directly onto the column. The flow rate was 4 μL/min with a 160 min linear gradient from 100% solvent A (95% H_2O/5% ACN/0.5% formic acid) to 100% solvent B (30% H_2O/70% ACN/0.5% formic acid). The C18 column (300 μm i.d. × 25 cm, 300 Å pore size, 5 μm particles; Vydac 218MS5.325 or Vydac 238EV5.325) was connected to the electrospray source with 100 μm i.d. fused silica tubing. Typical electrospray (ES) voltage was 4.5 kV and typical heated capillary temperature was 200−225 °C. The mass spectrometer was operated in the data dependent MS−MS mode with dynamic exclusion enabled and a repeat count of 1. In this mode, four parent ions from each mass spectrum were chosen automatically for MS−MS analysis based on an ion's (1) abundance in the mass spectrum, and (2) absence from an "exclusion list" of parent ions that had, more often than the "repeat count" setting, been subjected to MS−MS analysis in the previous 1 min time window. Data dependent LC−MS−MS was performed over a parent ion m/z range of 400−2000. In some experiments, to increase dynamic range, separate injections were made while scanning several narrower parent ion ranges (m/z 400−1000, m/z 980−1500, and m/z 1480−2000) in addition to the full m/z range of 400−2000.

Two-dimensional (2D) LC−MS−MS experiments were performed using a similar setup, with the following changes. Injections of 10 to 30 μg sample were made with the Famos autosampler onto a strong cation exchange column (LC Packings SCX, 500 μm i.d. × 15 mm), located on 10-port switching valve A of a Switchos system (LC Packings). The first dimension separation consisted of a series of step gradient elutions from the SCX column effected by 9−12 subsequent injections, using the Famos autosampler, of ammonium acetate salt at concentrations of 25 mM, 50 mM, 100 mM, 200 mM, 400 mM, 600 mM, 800 mM, 1 M, and one to four injections of 2M. Peptides eluting from the SCX column after each salt injection were captured on an LC Packings reverse phase precolumn (300 μm i.d. × 5 mm, 300 Å PepMap) on Switchos valve B. After washing salt from the precolumn, Valve B was switched to direct flow from the reversed-phase precolumn in backflush mode onto a nanoscale Vydac 218MS5.07515 C18 analytical column (75 μm i.d. × 15 cm, 300 Å pore size, 5 μm particles). This second dimension separation employed a 150 min gradient, going from solvent A (95% H_2O/5% ACN/0.1% formic acid) to solvent B (30% H_2O/70% ACN/0.1% formic acid) at 200 nL/min to elute

Table 1. Summary of "Bottom-Up" Analyses of Ribosomal Proteins[a]

						results of initial search[d]			
run no.	LC method	mass ranges[b]	sample amount (μg)	spectra produced[c]	measurement time (min)	proteins identified	high-scoring spectra[e]	identified peptides	average sequence coverage (%)
1	1D	1	10	4402	110	41	338	186	31
2	1D	4	72	10845	440	52	1071	604	51
3[a]	**1D**	**4**	**48**	**12906**	**440**	**51**	**1198**	**610**	**57**
4[a]	**2D**	**1**	**10**	**46033**	**2275**	**51**	**3737**	**821**	**60**
5	2D	1	30	54712	2275	50	5199	672	56

[a] Data from Runs 3 and 4 (shown in bold) subjected to more detailed SEQUEST analysis (see text for details). [b] Number of mass ranges for MS measurement (see Materials and Methods). [c] Total MS−MS spectra acquired. [d] Search considered all possible *R. palustris* proteins, peptides resulting from trypsin digestion, and no PTMs. [e] Spectra that met the default X_{corr} cutoffs (see Materials and Methods).

peptides into the mass spectrometer via a ThermoFinnigan nanospray source. The LCQ was run in the data dependent mode with dynamic exclusion enabled and a repeat count of 2.

Protein Identification from MS Data Analysis. The entire published *R. palustris* database[37] was used initially to analyze MS-MS spectra from bottom-up experiments using the SEQUEST algorithm (Thermo Finnigan).[53] Initial searches were configured to include only tryptic peptides. Data representing the best 1D and 2D runs (see Results and Discussion Section) from the preliminary results were then reanalyzed using SEQUEST by searching against *all* predicted peptides, without specifying tryptic cleavages. For SEQUEST post-translational modification searches, a subset of the *R. palustris* sequence database was used. This database contained all ribosomal proteins and other proteins for which at least one peptide was observed in either the best 1D run or the best 2D run. For PTMs we specified the following: The first search allowed mass shifts of 14 Da to detect methylation on lysine and arginine residues, and 16 Da to detect oxidations on methionine, cysteine, and tryptophan residues. The second search permitted mass shifts of 28 Da to detect dimethylations on lysines and arginines and 16 Da for methionine, cysteine and tryptophan residues. The third search permitted mass shifts of 42 Da to detect acetylations and trimethylations on lysine and arginine and 16 Da for methionine, cysteine, and tryptophan residues. The fourth search permitted 46 Da to detect β methylthiolation on aspartic acid and 16 Da for methionine, cysteine, and tryptophan residues. Two more searches aimed at identifying N-terminal modifications were performed to identify methylations (14 Da) and acetylation/trimethylation (42 Da) at the N-termini of peptides. Note that the PTMs specified for the bottom-up vs the top-down searches differ. Particular amino acid residues or termini can be specified for the bottom-up search, but not for the top-down search; furthermore, the tools for performing the searches differ in their nature and limitations.

The programs DTASelect and Contrast were used to assemble, filter, and compare the identifications from SEQUEST searches on various experimental data sets.[3] DTASelect groups, by protein, the peptides identified by SEQUEST as matching MS−MS spectra. This program filters peptide identifications based on SEQUEST parameters. DTASelect's default SEQUEST score cutoffs were used; spectra from singly charged peptides were required to exceed 1.8 in XCorr, whereas Xcorr values for doubly- and triply-charged peptides were required to exceed 2.5 and 3.5, respectively.[3] These conservative cutoffs which are recommended by the authors of the software have been found to work well in practice. The best matching sequence for each spectrum was required to have an X_{corr} at least 8% greater than the second best (DeltCN ≥ 0.08). Contrast combines DTASelect results for several different bottom-up experiments to summarize numbers of peptides identified and other parameters, grouped by protein.

Sequence Homology. Genomic database searches to retrieve orthologs to *R. palustris* ribosomal proteins were performed using the BLAST 2 software, developed by the National Center for Biotechnology Information, at the Swiss Institute of Bioinformatics (SIB) website (http://us.expasy.org/tools/blast/). The predicted amino acid sequences for ribosomal proteins were BLASTed against the UniProt knowledgebase. Multiple sequence alignments were performed at the SIB website using CLUSTAL W (version 1.74). Homology comparisons of predicted translated gene sequences were performed using the global gap alignment program EMBOSS (http://ccbioinfo.ornl.gov/emboss/).

Results and Discussion

The MS analysis presented in this paper identified a total of 53 ribosomal proteins. Our data indicated the presence of 21 proteins for the small subunit and 33 for the large subunit (S20 and L26 are identical). No ortholog of *E. coli* S22 was identified for *R. palustris* ribosomes. We also identified isoforms for L7/L12 from the large subunit. The traditional nomenclature for ribosomal proteins was adopted from studies of *E. coli*, where L1−L36 represent ribosomal proteins of the large subunit and S1−S22 denote proteins from the small subunit. In this paper, each of the *R. palustris* ribosomal proteins (RRP) is named after the corresponding ribosomal protein in *E. coli*. The L7/L12 isoforms were therefore named RRP-L7/L12A and RRP-L7/L12B (discussed later).

HPLC Separation Strategies for Bottom-Up Analysis. The mixture complexity of a tryptic digest from purified ribosomes is intermediate between that of a single protein digest and a digest from a whole proteome. Therefore, we compared several chromatographic strategies for the peptide separation in the bottom-up approach, including one-dimensional (1D) reverse phase liquid chromatography (RPLC), and two-dimensional (2D) separations employing both strong cation exchange (SCX) and RPLC. The criterion for this comparison was maximum sequence coverage of ribosomal proteins, which would be necessary for a comprehensive examination of post-translational modifications.

After optimizing both separation and MS protocols, and examining sequence coverage obtained from initial SEQUEST searches, we selected the best 1D and 2D data sets for further SEQUEST analysis tailored for identifying PTMs. Table 1 compares the results obtained using the various separation

strategies. A simple 1D RPLC separation required the least measurement time, but provided the smallest number of protein and peptide identifications. The same 1D RPLC separation, repeated four times, each time targeting a different range of precursor ions for fragmentation,[54,55] yielded a 3-fold larger number of peptide identifications. This four-mass-range 1D experiment was performed twice, identifying 604 different peptides in the first run and 610 in the second; the second run was selected for more extensive SEQUEST analysis due to slightly higher average sequence coverage per ribosomal protein and overall number of peptides identified, despite the use of less sample. Both the four-mass-range 1D and the 2D strategies resulted in similar numbers of identified ribosomal proteins and overall sequence coverage from the initial SEQUEST searches. Of the two 2D experiments, the run using 10 μg of sample produced a significantly larger number of peptide identifications than the 30 μg run, and so the former was chosen for further SEQUEST analysis. Although requiring the longest measurement time, the 2D method required less starting material than the multiple-mass-range 1D measurement, and produced the most raw spectra and confidently identified spectra, probably by decreasing the complexity of the peptide mixture introduced into the mass spectrometer at any particular time. Loading three times the sample amount on the 2D system resulted in a slight decrease in the mean sequence coverage of ribosomal proteins, while the sequence coverage on common contaminants was increased. This comparison suggests that 1D and 2D separations are complementary in regards to quality of results obtained, amount of sample required, and time requirements.

Protein Sequence Coverage and Protein Identifications. All but two ribosomal proteins were observed in the bottom-up analyses that were chosen for SEQUEST analysis tailored for PTM identification (see Table 2). The 1D, four mass-range analysis failed to observe RRP-L34, and the 2D analysis did not identify RRP-L34 or RRP-L36. These two proteins both have a high percentage of basic residues. RRP-L34 has five lysines and twelve arginines in its sequence of 44 residues, for an average of 2.6 residues between trypsin cut sites; RRP-L36 averages 2.7 residues between trypsin cut sites. Because these sequences are so rich in trypsin cleavage sites, many of the resulting peptides fall below the lower m/z limit for isolation and fragmentation. Interestingly, we identified the intact mass of RRP-L36 but not RRP-L34 from the FTICR analysis (discussed in more detail later).

Top-Down Characterization. Intact proteins from three separate aerobically grown ribosome samples were examined by LC−FTICR−MS, and the resulting data were pooled. From this top-down analysis, we identified 42 intact *R. palustris* ribosomal proteins. The four largest ribosomal proteins (RRP-S2 at 36 kDa, RRP-S1 at 62.8 kDa, RRP-L2 at 31.6 kDa, and RRP-S3 at 26.3 kDa) were not observed. Even though the FTICR−MS has sufficient mass range to observe these species, prior experience with intact proteins suggests that larger species such as these are difficult to elute from the C4 reversed-phase column under the experimental conditions employed for the top-down liquid chromatography. It is likely that the increased hydrophobicity of these larger proteins results in irreversible binding on the reversed-phase column, making these proteins difficult if not impossible to elute from the column.

Figure 2 presents an example of data from the top-down approach. Figure 2A shows a total ion chromatogram of the purified ribosome sample from the reversed-phase separation, and Figure 2B is the deconvoluted mass spectrum corresponding to the chromatographic peak at 1152 s. At least 10 different molecular species were observed in this spectrum, with molecular masses ranging from 7 to 11 kDa. For each observed species, the most abundant isotopic mass (MAIM) peak was used to query the entire *R. palustris* protein database for tentative protein identifications. This search was conducted by examining all intact and N-terminal methionine truncated proteins for possible matches. Note that this search did not consider all possible post-translational modifications of all the

Table 2. Sequence Coverage and Peptide Identifications for "Bottom-Up" **1D** and **2D** Analysis

	sequence coverage		no. peptide identifications[a]	
name	**1D (%)**	**2D (%)**	**1D**	**2D**
RRP-L1	66	80	25	43
RRP-L2	46	58	12	19
RRP-L3	79	92	23	45
RRP-L4	68	78	22	19
RRP-L5	54	45	10	14
RRP-L6	36	50	9	19
RRP-L7/L12	51	60	16	18
RRP-L9	90	56	13	17
RRP-L10	91	91	27	45
RRP-L11	61	69	10	20
RRP-L13	81	87	17	24
RRP-L14	80	84	11	16
RRP-L15	78	85	20	24
RRP-L16	65	77	14	26
RRP-L17	66	76	13	22
RRP-L18	73	73	12	27
RRP-L19	78	72	18	24
RRP-L20	51	57	10	11
RRP-L21	58	22	5	4
RRP-L22	63	57	12	18
RRP-L23	44	86	5	13
RRP-L24	89	89	11	14
RRP-L25	74	62	20	30
RRP-L27	63	87	8	15
RRP-L28	77	55	13	13
RRP-L29	45	32	5	5
RRP-L30	91	91	6	6
RRP-L31	80	100	6	9
RRP-L32	58	58	2	2
RRP-L33	56	66	4	9
RRP-L34	0	0	0	0
RRP-L35	27	49	3	5
RRP-L36	22	0	1	0
RRP-S1	37	39	12	16
RRP-S2	72	87	29	41
RRP-S3	57	68	16	26
RRP-S4	78	85	20	29
RRP-S5	72	70	22	25
RRP-S6	63	69	19	21
RRP-S7	72	82	22	28
RRP-S8	87	71	17	21
RRP-S9	84	74	18	27
RRP-S10	39	66	4	5
RRP-S11	86	33	10	13
RRP-S12	67	63	11	11
RRP-S13	33	72	6	21
RRP-S14	36	50	6	7
RRP-S15	87	99	14	17
RRP-S16	44	56	8	16
RRP-S17	83	79	12	10
RRP-S18	58	61	7	11
RRP-S19	98	84	12	19
RRP-S20	36	23	5	3
RRP-S21	23	51	3	9

[a] Number of different peptide identifications including +1, +2, and +3 charges states for identical peptides.

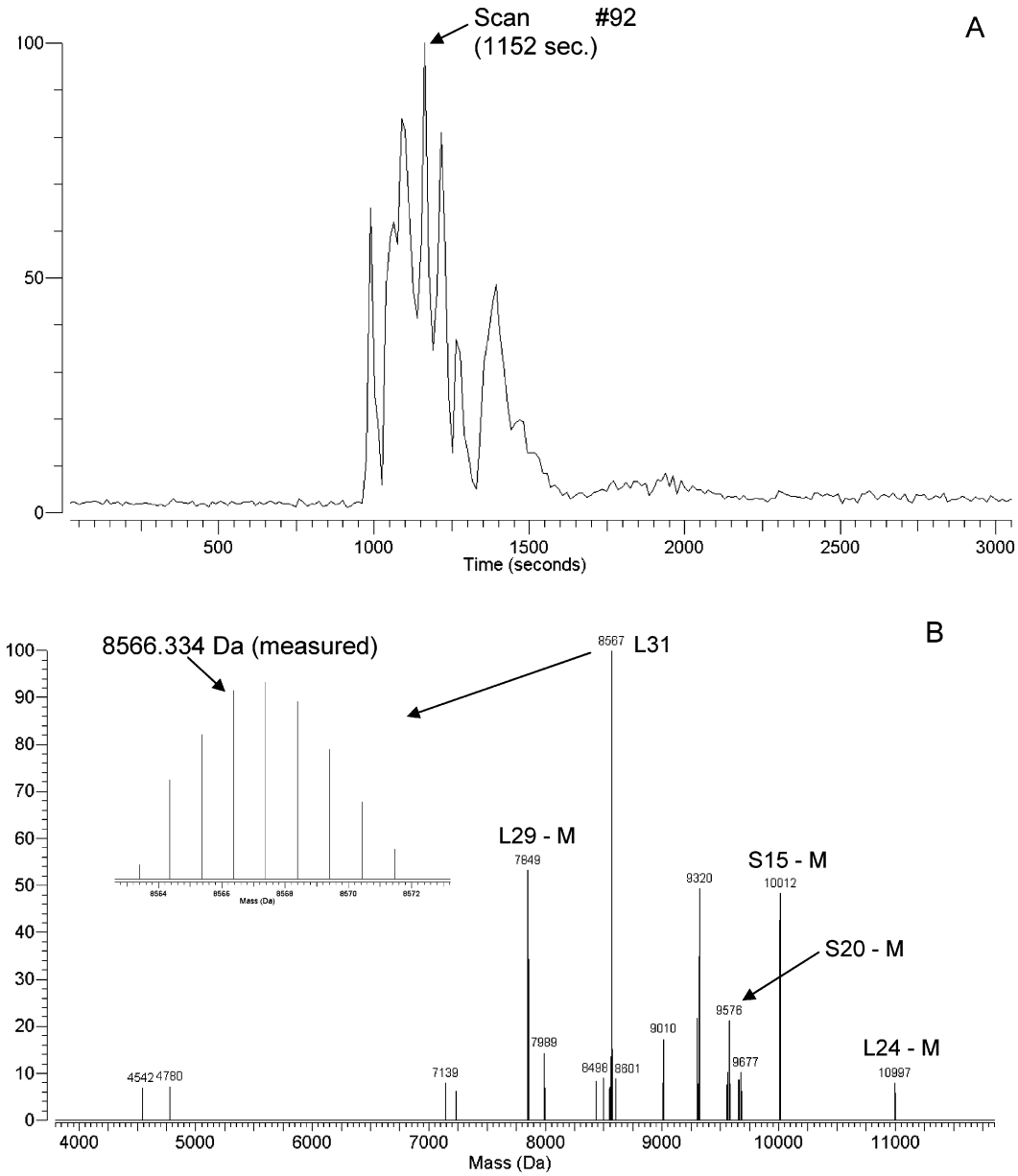

Figure 2. LC−FTICR measurement of intact masses. (A) Total ion chromatogram. (B) Deconvoluted mass spectrum corresponding to the chromatographic peak at 1152 s. Inset illustrates the isotopic resolution of the component at *nominal* mass 8567 Da.

possible proteins, as the number of such possibilities would preclude searching in a meaningful fashion. Although more definitive information could be obtained by conducting tandem MS on the intact proteins, this is difficult on the time scale of our chromatography. Our focus here was to compensate by correlating the top-down data with the bottom-up data for improved validation of tentative identifications. This approach, while probably less feasible for entire proteomes, is well suited to simpler systems such as the purified ribosome complex. An isotopically resolved pattern was then calculated from the elemental composition of each tentatively identified intact protein, and compared with the measured isotopic packet for final validation. The inset in Figure 2B illustrates the isotope pattern of the component at nominal mass 8567 Da. The measured isotopic packet of this species was consistent with the calculated isotopic packet of intact RRP-L31; the measured isotopically resolved peak at 8566.334 Da is within 2 parts per million of the calculated isotopically averaged value for this protein (8566.315 Da). If this measured protein mass at 8566.334 Da is used to query the entire *R. palustris* proteome, the next closest match is a methionine-truncated hypothetical protein (gene RPA 1934), which differs by 3 Da (360 ppm error) from the measured mass. In addition to the large mass error, RPA 1934 is a hypothetical protein that most likely would not be isolated in a ribosomal purification preparation. The next nearest ribosomal protein match to this measured value would be the methionine-truncated S18, which differs by 396 Da (45 000 ppm error) from the measured mass. This takes into account all possible ribosomal proteins, including intact, methionine truncated, or containing any variation of acetylation and/or methylation, to the extent specified in the Experimental section. Thus, within the constraints of our search, only RRP-L31 was found to be consistent with the measured mass of the 8567 Da species. Likewise, the component in Figure 2B at nominal mass 7849 Da had a MAIM of 7849.239 Da. This value is within 3 ppm of the calculated MAIM of 7849.213 Da

Table 3. Ribosomal Protein Identification by Top-Down ESI−FTICR−MS

protein	modification	calc. mass[a]	meas. mass[a]	mass error (ppm)
L1	loss of Met	23877.832	23877.449	16.0
L3	plus Methyl	25622.463	25622.159	11.9
L5	plus 2 Methyl	21064.992	21064.576	19.7
L6	loss of Met	19272.408	19272.674	−13.8
L7/L12	loss of Met + 3 Methyl	12754.070	12754.089	−1.5
L9	none	21178.022	21178.268	−11.6
L10	loss of Met	19067.739	19067.617	6.4
L11	loss of Met+Acet+ 9 Methyl	15507.107	15507.246	−9.0
L14	none	13488.498	13488.645	−10.9
L15	none	16836.243	16836.259	−1.0
L17	plus 3 Methyl	15716.353	15716.056	18.9
L18	loss of Met	12904.930	12905.157	−17.6
L19	none	14296.764	14296.899	−9.4
L21	loss of Met	13358.081	13358.533	−33.8
L22	loss of Met	13826.007	13825.644	26.2
L23	none	10907.949	10908.021	−6.6
L24	loss of Met	10998.226	10998.231	−0.5
L24	loss of Met + Methyl	11012.241	11012.146	8.6
L29	loss of Met	7849.213	7849.239	−3.3
L30	loss of Met	7092.967	7092.988	−3.0
L31	none	8566.315	8566.334	−2.2
L32	loss of Met	6860.730	6860.636	13.7
L33	loss of Met + Methyl	6248.504	6248.45	8.6
L35	loss of Met	7415.278	7415.278	0.0
L36	none	5063.971	5063.952	3.8
S4	loss of Met + Methyl	23441.536	23441.690	−6.6
S5	loss of Met	20522.086	20522.411	−15.8
S7	loss of Met	17556.270	17556.629	−20.4
S8	loss of Met	14477.631	14477.683	−3.6
S8	loss of Met+Acet+ 4 Methyl	14575.704	14575.619	5.8
S10	none	11667.363	11667.404	−3.5
S11	loss of Met + Methyl	13760.215	13760.314	−7.2
S12	none	13874.799	13875.167	−26.5
S13	loss of Met	14313.985	14313.596	27.2
S14	loss of Met	11331.399	11331.900	−44.2
S15	loss of Met	10010.563	10010.562	0.1
S16	loss of Met	12017.595	12017.575	1.7
S17	loss of Met	9553.253	9553.316	−6.6
S18	plus 6 Methyl	9178.219	9177.834	41.9
S19	loss of Met	10087.371	10087.379	−0.8
S20	loss of Met	9577.324	9577.387	−6.6
S21	none	10062.669	10062.722	−5.3

[a] MAIM (most abundant isotopic mass).

for the methionine truncated RRP-L29. The RRP-L31 species is only present as the intact gene product, whereas the RRP-L29 is only present in the methionine truncated form. By searching for intact gene products as well as methionine truncated forms, five ribosomal proteins could be identified in this mass spectrum. The remaining 4−5 species observed in this mass spectrum could not be identified. These may represent altered forms of ribosomal proteins that are not readily identifiable, such as other truncation products, or could be due to contaminant proteins isolated with the ribosome. Identification of truncation products of intact proteins, although difficult, is possible with the top-down MS approach,[31] but was not extensively pursued in this ribosome study.

Table 3 is the summary of intact protein identifications by the high-resolution FTICR−MS top-down technique. In total, 42 proteins were tentatively identified, with the majority (25) at better than 10 ppm mass accuracy, and only 3 differing by >30 ppm from the calculated value. Of these 42, 10 correspond directly to the predicted gene products, 21 are processed by only methionine truncation, and the remaining 11 appear to be modified by further acetylation and/or methylation. Two proteins, RRP-L24 and RRP-S8, were found to be present in two different forms. The most highly modified species identified was RRP-L11, which is methionine-truncated, and contains multiple methylations and/or acetylations. An analog to this highly decorated protein has been characterized in *E. coli*.[38,56] About 10 additional species were measured from the ribosome sample, but could not be identified. It is likely that these species correspond to the other ribosomal proteins, but are altered substantially (possibly by combinations of other PTMs, oxidation, and more extensive truncation) such that they are beyond the scope of our simple look-up table.

As described above, RRP-L34 and RRP-L36 were either identified poorly or not at all by bottom-up analysis, most likely because of their high basic content. Both proteins should, however, form positive ions quite readily in the ES source and therefore be detected by FTICR analysis. Although RRP-L36 could be matched to an isotopic packet from the FTICR analysis at 5063.952 Da, RRP-L34 was absent. Crystallographic structures of the 70S ribosome from *Thermus thermophilus* indicate that L34 is located at the base of the large subunit surface.[57] Biochemical isolation of *R. palustris* 70S ribosomes may have resulted in stripping of this protein from the subunit surface.

Post-translational Modifications of *R. palustris* Ribosomal Proteins. An important goal of this study was to search for PTMs of prokaryotic ribosomal proteins. Assignment of a particular PTM by only one proteomic technique is certainly possible, but PTM assignments can be strengthened by using the integrated TDBU approach, especially when results from the two approaches corroborate one another. The combined TDBU approach often allowed the identification of the modification positions and helped identify the presence of isoforms. For both analyses, we included in PTM searches the N-terminal modifications of methionine truncation, methylation, acetylation, and β-methylthiolation.[38,58] In addition, the search included β-methylthiolation of aspartic acids, single acetylations and mono-, di- and trimethylated lysines or arginines, all of which have been previously identified in ribosomal proteins from *E. coli* and eukaryotic-cell organellar ribosomes thought to have evolved from bacteria by endosymbiosis.[38,44,45,56,58] Phosphorylation, a common PTM in eukaryotic ribosomal proteins, has not been identified in prokaryotic ribosomal proteins, and was therefore not included in the subset of modification searches.[59]

N-Terminal Methionine Truncations. The most common PTM identified by the TDBU approach was truncation of the start methionine. We identified this modification in 32 *R. palustris* ribosomal proteins (Table 4). The top-down technique identified an N-terminal truncation if the measured intact mass for a protein matched that obtained by subtracting the mass contributed by a methionine residue (131.0405 Da) from the mass calculated from the DNA-derived amino acid sequence. Twenty-seven ribosomal proteins met this criterion. The bottom-up technique validated N-terminal truncations by identifying N-terminal peptides without a methionine; 16 truncated proteins could be identified in this way, while 10 were identified with the N-terminal methionine intact. For the proteins identified to have truncated N-termini, 10 were recognized by both techniques, 17 were identified by intact mass data alone and five were identified only by bottom-up data. Although a majority of these truncations were only identified by one

Table 4. Methionine Truncation Data for Bottom-Up and Top-Down Results

protein	observed top-down?	methionine truncated?	second residue[a]	N-terminal observations[b]
RRP-L1	Y	Y	A	T
RRP-L2	N	Y	A	1,2
RRP-L3	Y	N	R	1, 2, T
RRP-L4	N	N	E	1
RRP-L5	Y	*	A	1, T
RRP-L6	Y	Y	S	1, 2, T
RRP-L7/L12	Y	Y	A	1, 2, T
RRP-L9	Y	N	E	1, 2, T
RRP-L10	Y	Y	V	T
RRP-L11	Y	Y	A	T
RRP-L13	N	N	K	1, 2
RRP-L14	Y	N	I	1, 2, T
RRP-L15	Y	N	K	1, 2, T
RRP-L16	N	*	M	
RRP-L17	Y	N	K	T
RRP-L18	Y	Y	S	T
RRP-L19	Y	N	N	1, 2, T
RRP-L20	N	*	A	
RRP-L21	Y	Y	F	T
RRP-L22	Y	Y	S	T
RRP-L23	Y	N	K	1, T
RRP-L24	Y	Y	A	1, 2, T
RRP-L25	N	Y	T	1, 2
RRP-L27	N	*	A	
RRP-L28	N	Y	S	1
RRP-L29	Y	Y	A	T
RRP-L30	Y	Y	A	1, 2, T
RRP-L31	Y	N	K	1, 2, T
RRP-L32	Y	Y	A	T
RRP-L33	Y	Y	A	T
RRP-L34	N	*	K	
RRP-L35	Y	Y	P	T
RRP-L36	Y	N	K	T
RRP-S1	N	Y	A	1, 2
RRP-S2	N	*	S	
RRP-S3	N	Y	G	1, 2
RRP-S4	Y	Y	T	T
RRP-S5	Y	Y	A	T
RRP-S6	N	Y	P	1, 2
RRP-S7	Y	Y	S	T
RRP-S8	Y	Y	S	1, 2, T
RRP-S9	N	Y	S	1, 2, T
RRP-S10	Y	N	N	T
RRP-S11	Y	Y	A	T
RRP-S12	Y	*	P	1, 2, T
RRP-S13	Y	Y	T	T
RRP-S14	Y	Y	A	T
RRP-S15	Y	Y	S	1, 2, T
RRP-S16	Y	Y	S	T
RRP-S17	Y	Y	P	1, 2, T
RRP-S18	Y	Y	A	T
RRP-S19	Y	Y	V	1, 2, T
RRP-S20	Y	Y	A	1, T
RRP-S21	Y	N	Q	1, 2, T

[a] Amino acid occupying second position in polypeptide chain, if N-terminal met were present. [b] N-terminal observations supporting methionine truncation call: "T" = top-down, "1" = 1D LC-MS-MS, "2" = 2D LC-MS-MS. *Not identified.

technique, the combination of two MS approaches allowed a larger proportion of the N-termini to be surveyed than would have been possible with either one of these strategies, and provides increased confidence in the assignment for those proteins for which N-terminal methionine truncation was identified by both approaches.

Our results agreed well with previous reports in regards to the dependence of N-terminal methionine truncation on the charge and size of the amino acid side chain occupying the next position in from the N-terminal methionine. Residues bearing small uncharged side chains allow docking of methionine peptidases that cleave the N-terminal methionine. The data in Table 4 are consistent with this model. In nearly all cases where a methionine was truncated (column 3) the residue occupying the next position (column 4) was alanine, serine, proline, threonine, glycine, or valine. These six residues have previously been reported to occupy this second position in both eukaryotic and prokaryotic proteins that lose their N-terminal methionine.[59−61] For cases in which the N-terminal methionine was retained, the residue in the second position was lysine, isoleucine, glutamic acid, asparagine or glutamine; these residues possess bulky or charged residues that inhibit methionine aminopeptidases.[60,62]

Overall, the N-terminus truncation states of 45 proteins could be determined unambiguously. We were unable to determine whether seven ribosomal proteins contain an N-terminal methionine. For these species, it is possible that a longer truncation than simple methionine loss could have occurred. Among these, RRP-L16, RRP-L20, RRP-L27, RRP-L34, and RRP-S2 yielded neither bottom-up data for the N-terminal peptide, nor top-down molecular mass information. For RRP-L5 and RRP-S12, on the other hand, conflicting information resulted from the two techniques. For these two proteins, the bottom-up data indicated a methionine truncation while the top-down data did not. For each, the amino acid occupying the second position was small, suggesting the truncation product to be the most likely molecular form. Although this result is a bit puzzling, one possible explanation for the conflicting TDBU data is that these two proteins exist both with and without methionine truncations. Since the relative abundances of this pair are unknown, as are their relative chromatographic behavior and mass spectrometric responses as intact proteins or as N-terminal peptides, there is no expectation that they necessarily would be observed to the same extent in the top-down vs bottom-up experiments.

Table 5. Post-translational Modifications of *R. palustris* Ribosomal Proteins

protein	modification	residue(s)
RRP-L3	methylation	K155 or K158
RRP-L7/L12[a]	A: 2 methylations and 1 methylation	K69, K86
	B: trimethylation or acetylation	K86, K89
RRP-L11	Acetylation or trimethylation	K40
RRP-L30	methylation	N-terminus or K3[b]
RRP-L33	methylation	N-terminus or K3[b]
RRP−S12[c]	β-methylthiolation	D88

[a] Present as two isoforms, A and B. [b] Insufficient data to distinguish between methylation at the N-terminus or at K3. [c] Present in both modified and unmodified forms.

Methylation, Acetylation, and β-Methythiolation PTMs. In contrast to assigning N-terminal methionine truncations, identifying positions of acetylations, methylations, or β-methythiolation is more complex because these modifications often result in isoforms; furthermore, acetylation and methylation can occur either on residue side chains or N-termini. Table 5 summarizes the PTM assignments for ribosomal proteins determined from the TDBU approach. We assigned a particular PTM if at least one of the bottom-up data sets agreed with top-down data, or if bottom-up data from both the 1D and 2D separations were consistent.

RRP-L3. A MAIM peak in the top-down data at 25 622.159 Da was consistent with singly methylated RRP-L3. Bottom-up

ADLQKIVDDLSSLTVLEAAELAKLLEEKWGVSAAAAVAVAAAPGAGGAAAPAEEKTEFTVVLASA
GDK*KIEVIKEVRAITGLGLK*EAKDLVEGAPKPLKEGVNKEEAEKVKAQLEKAGAKVELK

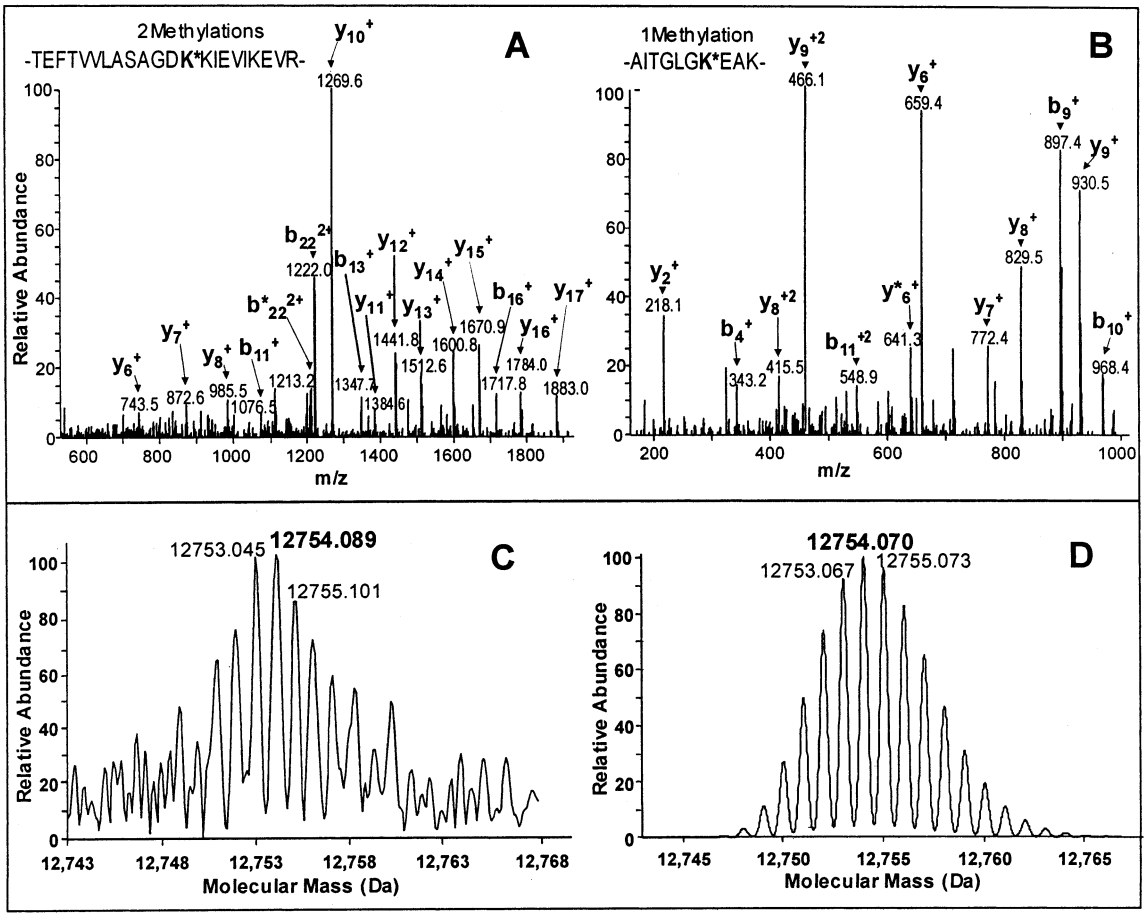

Figure 3. Comparison of top-down and bottom-up data for RRP-L7/L12. Fragmentation spectra from (A) peptide T57-R78 bearing 2 methylations at K69. (B) peptide A79-K89 bearing a single methylation at K86. (C) measured, and (D) calculated isotopic distributions for intact trimethylated RRP-L7.

analysis by 1D LC−MS−MS identified one peptide on which either K155 or K158 was methylated. In this spectrum (see Figure 1 in Supporting Information), y-series ions containing K155, K158, K160, and K161 (y_{17}, y_{18}) exhibited a 14 Da shift corresponding to methylation. Other y-series ions (y_8, y_{10}, y_{11}, y_{12}) showed no m/z shift relative to an unmethylated peptide; these unshifted y-series ions eliminated K160 and K161 as locations for the methylation. The SEQUEST search identified no unmethylated peptides covering this region of L3. While SEQUEST identified another tryptic peptide from L3 on which R170, R177 and R185 were all methylated, we assessed RRP-L3 to be singly methylated at either K155 or K158 because this scenario was supported by both the intact mass and the singly methylated peptide spectrum. L3 has previously been reported to be singly methylated in *E. coli* at Q150.[56] While the amino acid that is modified differs between RRP-L3 and the *E. coli* homologue, the data suggest that this single methylation is conserved between these two species.

RRP-L7/L12. A molecular mass from the top-down data at 12 754.089 Da indicated that this protein was modified by methionine truncation, plus either multiple methylation (three methyl groups) or acetylation. At low resolution, the latter two modifications are isobaric (i.e., 42 Da) and would not be resolved. However, this relatively small protein was observed in high abundance in the FTICR mass spectra and could be measured with high resolution and high mass accuracy, similar to the small L29 and L31 proteins illustrated in Figure 2. Thus, the measured MAIM of 12 754.089 Da suggested that this protein is trimethylated (calculated MAIM of 12 754.070 Da; 1.5 ppm less than measured value) rather than acetylated (calculated MAIM of 12 754.035; 4.2 ppm less than measured value), as shown in Figure 3, although more extensive measurements would be required to definitively make this assignment solely from intact mass data. 1D LC−MS−MS analysis suggested that K69 is dimethylated and K86 is singly methylated, with both multiple overlapping peptides and spectra for different charge states of the same peptide in evidence for K86. (Please note that our numbering for this protein begins with the truncated N-terminal methionine as residue 1.) Also identified from 1D data were acetylation or trimethylation at K86 and K89. The 2D LC−MS−MS analysis, on the other hand, identified K69 as dimethylated, and K6, K70, K86, and K100 as singly methylated (with K69 evidenced by two spectra representing multiple charge states of the same peptide, and K86 evidenced by multiple overlapping peptides), while other spectra indicated that K86 and K89 could be acetylated or trimethylated (on multiple overlapping peptides). These results are consistent with the existence of two isoforms of this protein; an increased abundance of one isoform over the other may explain why only one form is observed in the top-down data.

The isoform of RRP-L7/L12 for which an intact mass was observed is best explained by two methylations at K69 and a single methylation at K86. Figure 3 shows MS−MS spectra representing peptides with di-methylated K69 and mono-methylated K86 residues (Figure 3A,B) and the measured and calculated isotopic distributions determined for the intact mass (Figure 3C, D). Both MS−MS spectra indicate that y or b ions containing the modified residue are shifted by the appropriate mass. For example, in Figure 3A, y_{10}-y_{17} and b_{13}, b_{16} and b_{22}^{2+}, containing modified K69, are shifted by 28 Da (di-methylation), while y_6-y_9, and b_{11}, which do not contain K69, appear at the same m/z values as for an unmodified peptide. For these bottom-up data, the spectra from Figure 3A,B give information about the modified residues but not the number of isoforms that exist for RRP-L7/L12 with these modifications. Without the top-down data, it would not be possible to definitively assign these two modified peptides to a single isoform.

The second isoform of RRP-L7/L12 we assigned is acetylated or trimethylated at residues K86 and K89. Evidence for this isoform is found in MS−MS spectra corresponding to peptides with modified K86 and K89 in both 1D and 2D separations. An MS−MS spectrum (see Figure 2 in the Supporting Information) for this modified peptide shows b ions (b_{22}^{2+}, b_{11}, b_{12}, b_{14}, b_{16}) containing K86 and K89 shifted by 84 Da, indicating two acetylations or six methylations, whereas the y ions (y_5, y_6, y_9, y_{13}-y_{17}, y_{20}^{2+}) that do not contain K89 or K86 have no m/z shift. The L7/L12 protein in *E. coli* and other bacteria is known to exist in two isoforms: L7 is N-terminally truncated and acetylated, and L12 is N-terminally truncated and methylated at K81.[38,56]

RRP-L11. A definitive PTM analysis of this protein was not possible from our data. A molecular mass measured at 15 507.246 Da did not fit any unmodified ribosomal protein, but was consistent with the RRP-L11 protein containing methionine truncation and multiple acetylations and/or methylations. The *E. coli* homologue is methionine truncated, and trimethylated at each of three residues—K1, K3, and K39—for a total of nine methylations.[39,56] The measured intact mass suggests a slightly different level of modification for RRP-L11, with either one more acetylation or three more methylations than the *E. coli* L11 protein. The bottom-up data did not yield evidence for this level of modification. 1D and 2D LC−MS−MS did, however, identify peptides on which K40 was acetylated or trimethylated, with 1D results providing multiple overlapping peptides spanning this modification. In an MS−MS spectrum (See Figure 3 of the Supporting Information) representing a peptide with modified K40, a series of b-ions (b_{11}−b_{17} and b_{18}^{2+}−b_{20}^{2+}) containing K40 are shifted by 42 Da, corresponding to either acetylation or trimethylation. In contrast with the multiple modifications indicated by the top-down data, no PTMs other than this single acetylation/trimethylation could be identified in the bottom-up data.

RRP-L30. A molecular mass of 7092.988 Da corresponding to the methionine truncated L30 was observed by the top-down approach. In contrast, bottom-up data from both 1D and 2D analyses identified multiple peptides appearing to be methylated at K3, or at the N-terminal residue remaining after methionine truncation. Due to a lack of observed b or y ions containing either K3 or the N-terminus, but not both, it was not possible to distinguish between these two possible methylation sites. In one such spectrum assigned to peptide A2-R18, (Figure 4 in Supporting Information), b_6−b_{12}, b_{14}, and b_{16} were all shifted by 14 Da, indicating a single methylation, whereas y_4−y_{15} showed no m/z shift relative to an unmodified peptide.

RRP-L33. A methionine-truncated singly methylated form of RRP-L33 was identified at 6248.45 Da in top-down data. Methylation of RRP-L33 was also observed in multiple spectra of different charge states in both 1D and 2D bottom-up data. A representative spectrum (see Figure 5 in Supporting Information) indicates a scenario similar to that presented for RRP-L30, in which methylation at K3 is indistinguishable from methylation at the N terminus. All observed b series ion (b_2−b_8) are shifted by 14 Da, whereas y series ions show no m/z shift. A single methylation has been reported at the N-terminus (A1) of the *E. coli* L33.[56]

RRP-S12. A molecular mass of 13 875.167 Da corresponding to unmodified RRP-S12 was observed by the top-down approach. Bottom-up data from both 1D and 2D analyses, however, indicated the presence of both modified and unmodified RRP-S12. β-methylthiolation was observed in multiple spectra representing the same charge state. In an MS−MS spectrum assigned to peptide V86−R93 (see Figure 6 in the Supporting Information), y_6, y_7, and b_3 ions containing modified D88 are shifted by 46 Da corresponding to β-methylthiolation, whereas b_2, y_4, and y_5 ions not containing D88 are unshifted relative to an unmodified peptide. This novel PTM also occurs at D88 of the *E. coli* S12 ribosomal protein.[58]

Although some of our data suggest that other ribosomal proteins might possess PTMs, those reported in Table 5 include only cases for which supporting evidence from two or more different separation or MS approaches were found. For example, not included in the robust PTM assignments listed in Table 5 are several modified proteins identified from top-down data only (see Table 3). These include L5, L17, L24, S4, S8, S11, and S18. Similarly, although 2D LC−MS−MS provided bottom-up evidence for methylation at K6 and K100 of RRP-L7/L12, lack of evidence for these two modifications in the 1D LC−MS−MS data led to their exclusion from Table 5.

Unusual *R. palustris* Ribosomal Proteins. We searched for unique amino acid sequences in *R. palustris* ribosomal proteins by comparing their sequences with orthologs in other bacterial species. Orthologs to ribosomal proteins were retrieved by BLAST analysis (EXPASY) and then multiple sequence alignments were performed using CLUSTAL W. These studies revealed extended C-termini for RRP−S2, RRP-L9, and RRP-L25. *R. palustris* is a member of the bacterial group Proteobacteria and belongs to the α class. We found the greatest degree of similarity between ribosomal proteins from *R. palustris* and *B. japonicum*, two α proteobacteria that have previously been determined by rRNA analysis to be closely related.[63] With the exception of RRP-L25, RRP-L9, and RRP−S2, *R. palustris* ribosomal proteins were generally comparable in size with other bacterial orthologs. Table 6 summarizes information for RRP-L25, RRP-L9, and RRP−S2 and their sequence similarities with orthologs from organisms representing different proteobacterial classes.

RRP-L25. Figure 4 is an amino acid sequence alignment (CLUSTAL W) for the ribosomal protein L25 from candidates representing proteobacterial classes α, β, γ, and ϵ. Orthologs of L25 are generally some 200 amino acids in length, with distinct N- and C-terminal domains. The DNA derived sequence for RRP-L25 (233 amino acids), however, predicted an unusual extended C-terminus containing 26 alanines, with 8 other interspersed amino acids (lysines, glycines, and prolines). The only other bacterium found to possess an L25 protein with

```
R. palustris    L25 (α)          -MTSVLELKATARPKSGKGAARAERRAGRVPGVIYGDNQSPLPISVEEKE
B. japonicum L25 (α) Q89DJE      MATTVKELKATARPKSGKGAARAERRAGRVPGVIYGNNQPPVTISIEDRE
N. meningitidis L25 (β) Q9JZW3   -MT--YEIQASVREAQGTGASRRLRREGQIPGILYGEGQEPVAIAVDHKT
C. jejuni L25 (ε) Q9P118         ------MLEGIVRESIGRKAAKALKRDGYLIANIYGKGLENINAAFKVNE
E. coli L25 (γ) PO2426           ----MFTINAEVRKEQGKGASRRLRAANKFPAIIYGGKEAPLAIELDHDK

R. palustris    L25 (α)          LRLRILAGRFLTTVFDVVLDGKKHRVIPRDYHLDPVRDFPIHVDFLRLGA
B. japonicum L25 (α) Q89DJE      LRQRILAGRFLTTLVDIDLEGKKHRVIPRDYHLDPVKDFPIHVDFMRLGE
N. meningitidis L25 (β) Q9JZW3   VFYALEKESFHTALIKLSLNGETKDVIVRDFQMHPFRREVQHIDFQAVKA
C. jejuni L25 (ε) Q9P118         FIKEVR--KKTTLIFDVKVGSQTLSVVVVDYQKDPVTAELKHVDLKVAQK
E. coli L25 (γ) PO2426           VMNMQAKAEFYSEVLTIVVDGKEIKVKAQDVQRHPYKPKLQHIDFVRA--

R. palustris    L25 (α)          GATIRVSVPLHLKGLEVAPGVK-RGGTFNIVTHTVELEAPAENIPQFIEA
B. japonicum L25 (α) Q89DJE      GATIRISVPLHVVKSEASPGVK-RGGTVNIVAHAIELECGVESIPQYIEA
N. meningitidis L25 (β) Q9JZW3   DQLVRIRVPLHIVNAENSQAVKLQGGRVSLLNTSVEVVALPANIPAFLDL
C. jejuni L25 (ε) Q9P118         GVISKYMVPVKITGTAIGLKNK---GVLIQSKRRLKVKCAAENLPNFFEL
E. coli L25 (γ) PO2426           --------------------------------------------------

R. palustris    L25 (α)          DVSTLDIGVSLHLSDIALPTGVKSVSRDDVTLVTIVPPSGYNEDK-----
B. japonicum L25 (α) Q89DJE      DVGSLEIGHSLHLSDVKLPAGVKALTREDATLVTIVPPSGYAEEQKAAAA
N. meningitidis L25 (β) Q9JZW3   DCAEVVAGDILHLSDIKLPEGVESVSLKRNENLAVATVTGKKR--------
C. jejuni L25 (ε) Q9P118         DVSKLDVGDALLVRDIVVPAGVTMIDADRVAVVGVEKAR----------
E. coli L25 (γ) PO2426           --------------------------------------------------

R. palustris    L25 (α)          AAAGAAPAAAAAPAAAAKAPAAAAKAPAAAAPAA--KKK
B. japonicum L25 (α) Q89DJE      AAAGGAAPAAAAPAAGAAAPAAAAPAAAAKAPAGGDKKK
N. meningitidis L25 (β) Q9JZW3   ---------------------------------------
C. jejuni L25 (ε) Q9P118         ---------------------------------------
E. coli L25 (γ) PO2426           ---------------------------------------
```

Figure 4. Multiple sequence alignment using CLUSTAL W for the ribosomal protein L25 from candidates representing different proteobacterial classes (α,*β*,*γ* and *ϵ* classes).

Table 6. Homology Comparisons

R. palustris (α)		*B. japonicum* (α)		*E. coli* (*γ*)		*N. meningitidis* (*β*)		*C. jejuni* (*ϵ*)	
ribosomal protein	no. amino acids	no. amino acids	% similarity	no. amino acids	% similarity	no. amino acids	% similarity	no. amino acids	% similarity
L9	195	201	86	149	48	150	52	147	39
L25	233	238	83	94	22	190	49	178	50
S2	331	331	93	240	52	242	54	263	52

this C-terminal "poly-alanine tail" was *B. japonicum*. Interestingly, this type of low complexity, highly repetitive sequence, common to eukaryotes, has not previously been identified at the protein level in prokaryotes.[64] Within the limits of our search, no top-down evidence for L25 was found. Therefore, we manually evaluated our bottom-up data to identify peptides representing this C-terminal feature. Bottom-up data (Figure 7 in the Supporting Information) from both 1D and 2D LC separations included peptides for the entire C-terminal extension (with the exception of the last three lysine residues). To our knowledge, this represents the first experimental confirmation at the protein level in prokaryotes of a low complexity sequence composed primarily of a single amino acid.

RRP-L9. RRP-L9 also possesses a predicted C-terminal region that is 40 to 50 amino acids longer than orthologs of L9 in other bacteria, which are generally 140−150 amino acids in length, while RRP-L9 is predicted to be 201 amino acids in length. The N-terminal 140−150 amino acid segment of RRP-L9 shares significant similarity with L9 from other bacterial species. For example, the 147 amino acid *E. coli* L9 shares 66% sequence identity to the N-terminal 147 amino acids of RRP-L9. The C-terminal 40−50 amino acid extension of RRP-L9, however, was only identified in the seven α proteobacteria for which genome sequence is available. Among these, the L9 ortholog for *B. japonicum* shared 86% sequence identity (see Table 6). We were unable to identify MS−MS spectra corresponding to the C-terminal extension of RRP-L9 due to the acidity of this region; 17 of the 50 C-terminal residues are either aspartic acid or glutamic acid. Acidic peptides can often be observed as negative ions, but they do not readily acquire a net positive charge, and are therefore difficult to detect by electrospray-mass spectrometry in the positive ion mode employed for this work. We did, however, observe the intact mass for this protein in our top-down data. This intact mass assignment supported the presence of an extended C-terminus in RRP-L9, and, by extension, in L9 proteins in other α proteobacteria.

RRP-S2. Our MS data also allowed the experimental verification of an extended C-terminus in RRP-S2. S2 orthologs from other bacteria are generally 240−250 amino acids in length. RRP-S2, however, is 331 amino acids in length and shares 94% sequence similarity to the S2 ortholog in *B. japonicum*, which is also 331 amino acids in length. No other sequenced α

proteobacterium possessed an S2 ortholog with this C-terminal extension. Although we were unable to identify RRP-S2 from top-down data, we were able to map peptides to all but the last 16 amino acids of the C-terminus; see Figure 8 in the Supporting Information. This bottom-up analysis allowed the confirmation of a novel S2 ortholog found only in *R. palustris* and *B. japonicum*.

Conclusions

We have presented the mass spectrometric analysis of the component proteins of the 70S ribosome from *R. palustris*. The combination of bottom-up and top-down approaches enhanced several aspects of the analysis. The intact mass measurement includes the aggregate contribution of all modifications to the protein, allowing the discrimination of isoforms with different molecular masses, while the peptide-based analysis pinpoints the positions and masses of individual modifications. Additionally, the combination of genomic database searches and manual evaluation of MS data from both techniques helped validate ribosomal proteins that contained unusual extended C-termini.

Current limitations of the top-down method include nonuniform detection of proteins due to their widely varying physical and chemical characteristics, which affect both their chromatographic separation and introduction into the mass spectrometer via electrospray. Weaknesses of the bottom-up approach include missed PTMs due to undetected PTM-bearing peptides, difficulties in identifying isoforms of modified proteins, and lack of automated tools for identification of PTMs. When considering the integrated TDBU data, it is currently also a limitation that inconsistencies can arise when comparing data from the two approaches. However, as newer mass spectrometric and data analysis tools become available, these inconsistencies will become rarer. Despite these limitations, we have shown that the application of both approaches, and analysis of the data in an integrated fashion, provides more information than could be obtained from using either approach in isolation.

Acknowledgment. We thank Goran Mitulovic, Mark Van Gils and LC Packings, a Dionex Company, for providing and helping in optimizing the 2D HPLC system used in this study. We thank Grace Vydac Corp. for providing HPLC columns through a collaborative agreement. We thank Caroline S. Harwood, Department of Microbiology, University of Iowa for generously contributing the *R. palustris* strain used in this study. We also thank the John Yates laboratory at Scripps Research Institute for DTASelect and Contrast software. N. C. Verberkmoes and H. Connelly acknowledge support from the University of Tennessee-Oak Ridge National Laboratory Graduate School of Genome Science and Technology. We would like to thank Becky Maggard for administrative assistance in formatting the figures and final document. This research was funded by the U.S. DOE Office of Biological and Environmental Sciences Genomes to Life Program. Oak Ridge National Laboratory is operated and managed by the University of Tennessee-Battelle, L.L.C. for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

Supporting Information Available: Labeled tandem mass spectra supporting identification of post-translational modifications, and the annotated *R. palustris* protein database used for this work. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Yates, J. R., III.; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem*. **1995**, *67*, 1426−1436.
(2) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(3) Tabb, D. L.; McDonald, W. H.; Yates, J. R., III. *J. Proteome Res.* **2002**, *1*, 21−26.
(4) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem*. **1991**, *63*, 1193A−1203A.
(5) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(6) Wolters, D. A.; Washburn, M. P.; Yates, J. R., III. *Anal. Chem*. **2001** *73*, 5683−5690.
(7) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. *Anal. Chem*. **1997**, *69*, 767−776.
(8) Gygi, S. P.; Aebersold R. *Curr. Opin. Chem. Biol.* **2000**, *4*, 489−494.
(9) Gavin, A. C.; Bösche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Höfert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141−147.
(10) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L. Y.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sørensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann M.; Hogue, C. W. V.; Figeys, D.; Tyers, M. *Nature* **2002**, *415*, 180−183.
(11) Jensen, O. N.; Larsen, M. R.; Roepstorff, P. *Proteins, Suppl.* **1998**, *2*, 74−89.
(12) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5011−5015.
(13) Pappin, D. J. C.; Højrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327−332.
(14) Mann, M.; Højrup, P.; Roepstorff, P. *Biol. Mass Spectrom*. **1993**, *22*, 338−345.
(15) Wilm, M.; Mann, M. *Anal. Chem*. **1996**, *68*, 1−8.
(16) Hunt, D. F.; Yates, J. R., III.; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233−6237.
(17) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III. *Nat. Biotechnol.* **1999**, *17*, 676−682.
(18) Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nat. Biotechnol.* **2002**, *20*, 301−305.
(19) 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.; Clark, J. I.; Yates, J. R., III. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7900−7905.
(20) Mørtz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264−8267.
(21) Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H. J.; Begley, T. P.; McLafferty, F. W. *Protein Sci.* **1998**, *7*, 1796−1801.
(22) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806−812.
(23) Holmes, M. R.; Giddings, M. C. *Anal. Chem*. **2004**, *76*, 276−282
(24) Lee, S. W.; Berger, S. J.; Martinović, S.; Paša-Tolić, L.; Anderson, G. A.; Shen, Y. F.; Zhao, R.; Smith, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5942−5947.
(25) Meng, F. Y.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. *Nat. Biotechnol.* **2001**, *19*, 952−957.
(26) Taylor, G. K.; Kim, Y. B.; Forbes, A. J.; Meng, F. Y.; McCarthy, R.; Kelleher, N. L. *Anal. Chem*. **2003**, *75*, 4081−4086.
(27) Louie, D. F.; Resing, K. A.; Lewis, T. S.; Ahn, N. G. *J. Biol. Chem*. **1996**, *271*, 28 189−28 198.
(28) Zhu, K.; Kim, J.; Yoo, C.; Miller, F. R.; Lubman, D. M. *Anal. Chem*. **2003**, *75*, 6209−6217.
(29) Yan, F.; Subramanian B.; Nakeff A.; Barder T. J.; Parus S. J.; Lubman D. M. *Anal. Chem*. **2003**, *75*, 2299−2308.
(30) Hamler, R. L.; Zhu, K.; Buchanani, N. S.; Kreunin, P.; Kachman, M. T.; Miller, F. R.; Lubman, D. M. *Proteomics* **2004**, *4*, 562−577.

(31) VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Jr. *J. Proteome Res.* **2002**, *1*, 239−252.
(32) Buchanan, M. V.; Larimer, F. W.; Wiley, H. S.; Kennel, S. J.; Squier, T. J.; Ramsey, J. M.; Rodland, K. D.; Hurst, G. B.; Smith, R. D.; Xu, Y.; Dixon, D.; Doktycz, M. J.; Colson, S.; Gesteland, R.; Giometti, C.; Young, M.; Giddings, M. *OMICS* **2002**, *6*, 287−303.
(33) Blankenship, R. E.; Madigan, M. T.; Bauer, C. E. *Anoxygenic Photosynthetic Bacteria;* Kluwer Academic Publishers Group: Norwell, MA, 1995; Vol. 2, 4−11.
(34) Barbosa, M. J.; Rocha, J. M. S.; Tramper, J.; Wijffels, R. H. *J. Biotechnol.* **2001**, *85*, 25−33.
(35) Egland, P. G.; Pelletier, D. A.; Dispensa, M.; Gibson, J.; Harwood, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **1997** *94*, 6484−6489.
(36) Sasikala, C.; Ramana, C. V. *Adv. Microb. Physiol.* **1998**, *39*, 339−377.
(37) Larimer, F. W.; Chain, P.; Hauser, L.; Lamerdin, J.; Malfatti, S.; Do, L.; Land, M. L.; Pelletier, D. A.; Beatty, J. T.; Lang, A. S.; Tabita, F. R.; Gibson, J. L.; Hanson, T. E.; Bobst, C.; Torres, J. L. T. Y.; Peres, C.; Harrison, F. H.; Gibson, J.; Harwood: C. S. *Nat. Biotechnol.* **2004**, *22*, 55−61.
(38) Wittmann, H. G. *Annu. Rev. Biochem.* **1982**, *51*, 155−183.
(39) Ofengand, J. In *Ribosomes: Structure, Function, and Genetics*; Chambliss, G. et al., Eds.; University Park Press: Baltimore, MD, 1980; pp 497−530.
(40) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920−930.
(41) Mueller, F.; Sommer, I.; Baranov, P.; Matadeen, R.; Stoldt, M.; Wohnert, J.; Gorlach, M.; Van Heel, M.; Brimacombe, R. *J. Mol. Biol.* **2000**, *298*, 35−59.
(42) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Stietz, T. A. *Science* **2000**, *289*, 905−920.
(43) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679−688.
(44) Kamp, R. M.; Srinivasa, B. R.; von Knoblauch, K.; Subramanian, A. R. *Biochemistry* **1987**, *26*, 5866−5870.
(45) Yamaguchi, K.; Subramanian, A. R. *J. Biol. Chem.* **2000**, *275*, 28 466−28 482.
(46) Geissler, J. F.; Harwood, C. S.; Gibson, J. *J. Bacteriol.* **1988**, *170*, 1709−1714.
(47) Noll, M.; Hapke, B.; Schreier, M. H.; Noll, H. *J. Mol. Biol.* **1973**, *75*, 281−294.
(48) Hardy, S. J. S.; Kurland, C. G.; Voynow, P.; Mora, G. *Biochemistry* **1969**, *8*, 2897−2905.
(49) Sharp, J. S.; Becker, J. M.; Hettich, R. L. *Anal. Biochem.* **2003**, *313*, 216−225.
(50) Uchiki, T.; Hettich, R.; Gupta, V.; Dealwis, C. *Anal. Biochem.* **2002**, *301*, 35−48.
(51) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10 313−10 317.
(52) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am Soc. Mass Spectrom.* **2000**, *11*, 320−332.
(53) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(54) Spahr, C. S.; Davis, M. T.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Courchesne, P. L.; Chen, K.; Wahl, R. C.; Yu, W.; Luethy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 93−107.
(55) Davis, M. T.; Spahr, C. S.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Yu, W.; Luethy, R.; Patterson, S. D. *Proteomics* **2001**, *1*, 108−117.
(56) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105−112.
(57) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H. D.; Noller, H. F. *Science* **2001**, *292*, 883−896.
(58) Kowalak, J. A.; Walsh K. A. *Protein Science* **1996**, *5*, 1625−1632.
(59) Arnold, R. J.; Polevoda, B.; Reilly, J. P.; Sherman, F. *J. Biol. Chem.* **1999**, *274*, 37 035−37 040.
(60) Sherman, F.; Stewart, J. W.; Tsunasawa, S. *Bioessays* **1985**, *3*, 27−31.
(61) Huang, S.; Elliott, R. C.; Liu, P. S.; Koduri, R. K.; Weickmann, J. L.; Lee, J. H.; Blair, L. C.; Gosh-Dastidar, P.; Bradshaw, R. A.; Bryan, K. M.; Einarson, B.; Kendall, R. L.; Kolacz, K. H.; Saito, K. *Biochemistry* **1987**, *26*, 8242−8246.
(62) Tsunasawa, S.; Stewart, J. W.; Sherman, F. *J. Biol. Chem.* **1985**, *260*, 5382−5391.
(63) van Berkum, P.; Terefework, Z.; Paulin, L.; Suomalainen, S.; Lindstrom, K.; Eardly, B. D. *J. Bacteriol.* **2003**, *185*, 2988−2998.
(64) Huntley, M. A.; Golding, G. B. *Proteins: Structure, Function, and Genetics* **2002**, *48*, 134−140.

PR049940Z