

Ribosomal Proteins of *Deinococcus radiodurans*: Their Solvent Accessibility and Reactivity

William E. Running and James P. Reilly*

Department of Chemistry, Indiana University, Bloomington, Indiana, 47405

Received July 18, 2008

The structure of proteins in native ribosomes from *Deinococcus radiodurans* R1 was probed by S-methylthioacetimidate (SMTA) modification of amino groups. The extent of protein labeling was quantified using top down methods, and modified positions were identified using bottom up experiments. Each protein's reactivity was predicted by examination of the crystal structures of the *D. radiodurans* 50S subunit and the *T. thermophilus* HB8 30S subunit. The close phylogenetic relation between *D. radiodurans* and *T. thermophilus* allowed the evaluation of *D. radiodurans* small subunit protein reactivity by alignment of homologous sequences. As a result, we were able to observe and characterize the reactivity of all of *D. radiodurans* ribosomal proteins. The extent of protein amidination was well correlated with the solvent-exposed surface area of each protein and even better correlated with the number of visible lysine residues. Lysine residues that are in close contact with rRNA structural features or buried in protein tertiary structure are nonreactive with SMTA, while those that are surface exposed are modified. Crystallographic disorder and post-translational modifications lead to differences between the observed and predicted extents of reactivity. Comparison of unmodified and disassembled amidinated protein mixtures also shows great promise for the quality control of the proteomic sequences and has facilitated the identification of four sequencing errors in the ribosomal proteome of *D. radiodurans* R1.

Keywords: ribosomal proteins • post-translational modifications • protein-nucleic acid interactions • protein mass spectrometry • chemical modification of proteins

Introduction

The ubiquitous presence of large multiprotein complexes, such as the F_1F_0 ATPase or the type II fatty acid synthase, and larger macromolecular complexes, such as the ribosome, suggests that the machinery of cellular metabolism is a network of supra-organellar complexes.¹ To characterize the structure and function of macromolecular complexes, it is necessary to determine the identities and stoichiometries of their components and to define the sites of intra- and intercomplex association.^{2–4} A complete understanding of the role of a complexes components must include a description of the state of its proteins, including post-translational modifications that are not directly predictable from the sequenced genome.

Modern soft ionization techniques have enabled the application of mass spectrometry (MS) to the study of biomolecular complex structure.^{2,3} Top-down experiments allow the study of whole proteins in their naturally occurring forms, including post-translational modifications, while bottom-up proteomics generates an inventory of their components.^{5–7} Recent landmark success in the transfer of intact macromolecular complexes into the gas phase from physiologically relevant solution conditions have also increased the potential of macromolecular MS.³ Both crystallography and NMR studies

obtain structural data with resolution on the order of a bond length, at the expense of requiring several milligrams of protein and considerable time to obtain final results.^{8,9} MS-based techniques trade the Ångstrom-level resolution of NMR and crystallography for enhanced sensitivity, and can monitor protein conformation at the domain level with 0.1–10 picomoles of material in experiments that take a few hours.¹⁰ Intramolecular domain associations and intermolecular complex formation are studied by introducing mass shifts transiently by hydrogen–deuterium exchange,^{11–13} or permanently by covalent modification.^{2,14–24}

Mass spectrometric monitoring of amide hydrogen exchange provides information on protein structural flexibility and residue solvent accessibility.¹² This technique can also be used to detect conformational changes induced by binding interactions.¹³ However, subsequent analysis by fast proteolytic digestion and LC separation inevitably results in loss of the label by back-exchange. The broad range of exchange rates, the dependence of these rates on peptide structure and the difficulty in obtaining 100% label incorporation combine to make analysis of isotopic exchange data problematic.¹⁴

The reactivity of protein residues can be affected by secondary, tertiary and quaternary structure, and deductions about structure can be made from the patterns of reactivity displayed by a protein subjected to different modification reagents. Residue-selective reagents target specific amino acid side

* To whom correspondence should be addressed. E-mail: reilly@indiana.edu.

Ribosomal Proteins of *Deinococcus radiodurans*

chains based on their chemical reactivity, structural involvement, and solvent accessibility.^{14–21} Partial proteolytic reactivity of exposed “hinges and fringes” depends on solvent exposure and localized mobility.^{22–24} In both cases, the resulting mass differences can be predicted from the target protein’s sequence and the reagent or protease utilized, making these techniques an excellent match for mass spectrometric detection. Covalent modifications do not suffer from complications in interpretation due to back exchange because the modifications are permanent on the time scale of the analysis. However, care must be taken that the solution conditions do not affect the target molecule’s structure and that the modifications do not cause denaturation.^{18–21}

The goal of the research reported here was to validate covalent modification of lysine residues by S-methylthioacetamide (SMTA) as a probe of ribosome quaternary structure. Our group has investigated the application of thioamides as signal enhancers in peptide mass spectrometry and as probes of native protein structure.^{25–27} More recently we studied the quaternary structure of *Caulobacter crescentus* CB15N ribosomes, using the differential reactivity of lysine residues to SMTA to deduce which lysines were closely associated with rRNA and which were solvent exposed.²⁸ A good correlation between the extent of derivatization of each protein and its solvent accessible surface area was observed. The high sequence homology of ribosomal proteins²⁹ suggested that a better correlation might be found between the extent of modification and the number of lysine residues visible on the surfaces of ribosome crystal structures. *C. crescentus* protein sequences were aligned with those of *Deinococcus radiodurans* R1 and *Escherichia coli* K12. The positions of lysine residues in these sequence alignments and in the crystal structures of the *D. radiodurans* 50S³⁰ and *E. coli* 30S⁶⁴ subunit were used to predict the locations of these residues in *C. crescentus* ribosomal proteins. Very good agreement between the extent of amidination and the number of lysine groups visible in each protein’s crystal structure was obtained.²⁸ Here, we report the results of a study of the quaternary structure of the 50S subunit of *D. radiodurans* R1 ribosomes through amidination. A 3.1 Å resolution crystal structure of the 50S subunit of *D. radiodurans* ribosome was used for comparison.³⁰ To interpret labeling data on the 30S subunit, we employed a 2.8 Å resolution crystal structure of the 70S ribosomal particle from the closely related *T. thermophilus* HB8.³² Both bacterial species are closely related members of the same phylum, based on homologies in their 16S rRNA sequences.³³ We anticipated virtual identity between our observed extent of amidination and the number of visible modifiable groups found in the 50S subunit crystal structure. Any discrepancies between experimental and predicted extents of labeling are expected to arise from sequence errors, post-translational modifications, and deficiencies in the crystal structures.

Materials and Methods

Chemicals. Water was purified by a Barnstead and Noble Nanopure system. Ammonium chloride, 2-mercaptoethanol and methylamine as a 40% (v/v) solution in water were supplied by Aldrich. HPLC grade acetonitrile, methanol, trifluoroacetic acid, and formic acid were products of J. T. Baker. Bacto-tryptone and Yeast extract for bacterial growth media were provided by DIFCO. Reagent grade sodium chloride and (ethylenedinitrilo)tetraacetic acid (EDTA) were obtained from Mallinckrodt. Endoprotease Glu-C was provided by New En-

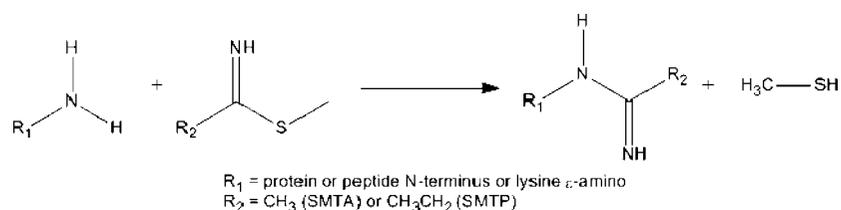
gland Biolabs. 2-Amino-2-(hydroxymethyl)-1,3-propanediol (TRIZMA, tris free base), Glucose, Sucrose, 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES, free acid), leucine enkephalin, horse heart myoglobin, Proteomics-grade alkylated porcine trypsin, Carboxypeptidase Y (from *Saccharomyces cerevisiae*) and Carboxypeptidase P (from *Penicillium janthinellum*) were purchased from Sigma.

Cell Growth and Ribosome Preparation. Ribosomes were prepared using Arnold and Reilly’s modification of Spedding’s procedure as reported for *Caulobacter crescentus* ribosomes, substituting HEPES for tris and omitting the second salt wash.^{34,35} Buffer A contained 20 mM HEPES, 100 mM NH₄Cl, 10.5 mM Mg(acetate)₂, 0.5 mM EDTA, and 5 mM 2-mercaptoethanol, adjusted to give a pH of 7.8 at 4 °C. Buffer B has the same composition as Buffer A except for 500 mM NH₄Cl. Buffer E has the same composition as Buffer A except for 60 mM NH₄Cl. *Deinococcus radiodurans* R1 cells were grown in Tryptone-Yeast extract-0.1% glucose (TY1G) medium overnight at 30 °C to midlog phase (OD₆₀₀ ≈ 0.8). Cells were centrifuged for 25 min at 6000 × g in a Beckman JA-10 rotor and washed once with one volume of Buffer A. A typical preparation yielded 8.8 g wet weight of cells. For some preparations, 3 mM phenylmethylsulfonylfluoride (PMSF, Sigma) was included to inhibit nonspecific proteolysis. After washing, cells were resuspended in a minimum volume of Buffer A containing miniComplete EDTA-free protease inhibitor cocktail tablets (1 tablet/10 mL solution, Boehringer-Mannheim). Cells were lysed by five passages through a French press at 16,000 psi (SLM-Aminco). Cell debris was cleared by spinning for 40 min at 30,000 × g in a Beckman JA-20 rotor. The cleared lysate was layered onto an equal volume of 1.1 M sucrose in Buffer B and spun in an ultracentrifuge for 16 h at 100,000 × g in a Beckman 60Ti rotor. After decanting the supernatant the ribosomes were resuspended in a minimum volume of Buffer E and aliquoted for storage at –80 °C. The ribosome suspension was not dialyzed since sucrose was not expected to interfere with any subsequent ion exchange or reversed phase chromatography steps. The final concentration of protein in this sample was estimated at 24.8 mg/mL by Bradford assay using BSA as a standard.

Ribosomal Protein Extraction. For direct analysis of the ribosomal proteins, rRNA was removed from whole ribosomes by mixing 1/3-volume of 1 M MgCl₂ and 2 volumes of glacial acetic acid with ribosomes stored in Buffer E. Samples were mixed and allowed to stand at room temperature for 10 min, then centrifuged for 10 min at 14,100 × g in an Eppendorf microfuge (Eppendorf North America, New York). The protein-containing acetic acid supernatant was removed by aspiration and analyzed directly or chemically modified as detailed below. The protein concentration of this acetic acid extract was estimated to be 8.7 mg/mL by Bradford assay with BSA as a standard. Typical analyses used 25–50 μL of this solution.

Amidination Reactions. The SMTA reagent was prepared as described by Beardsley and Reilly.²⁵ SMTA and S-methyl thiopropionimide (SMTP) have been used as signal enhancers and mass tags for peptide mass spectrometry and as derivatization reagents for whole protein mass spectrometry.^{25–27} These reagents react rapidly with free amino groups at room temperature, under solution conditions previously shown to preserve tertiary and quaternary structure,^{26–28} according to Scheme 1. Amidination with SMTA increases the average mass of the derivatized molecule by 41.05 Da per amidino group added.

Scheme 1



To prepare ribosomal proteins amidinated under native conditions, an aliquot of ribosome suspension in Buffer E was mixed with an equal volume of 43.4 mg/mL SMTA dissolved in 250 mM tris free base (pH 10.6).^{26–28} After incubating for 1 h at room temperature ribosomal proteins were extracted as described above. Addition of glacial acetic acid to the reaction mixture to precipitate rRNA also stops the amidination reaction by protonating all free amino groups in the sample. Because of the 2-fold dilution relative to unmodified acetic acid extracts, typical separations of this sample used 50–100 μ L of this solution.

Amino groups will not react with SMTA under strongly acidic conditions, so proteins from disassembled ribosomes were prepared by precipitation from the acetic acid supernatant with acetone. A 100 μ L aliquot of acetic acid extract was chilled on ice, and then mixed with 5 volumes of ice-cold acetone.³⁶ The mixture was allowed to stand on ice for 1 h, and then the precipitated proteins were separated from the supernatant by a brief spin (ca. 1 min) at 1000 \times g. The supernatant was removed by aspiration and the precipitate was resuspended in 50 μ L 6 M urea buffered with 25 mM ammonium bicarbonate. When the protein was fully redissolved, 50 μ L of 43.4 mg/mL SMTA in 250 mM tris was added, and the amidination reaction was allowed to proceed for 1 h at room temperature, then stopped by the addition of 10 μ L of glacial acetic acid. The final solution was not assayed for protein concentration. Typical separations used 35–70 μ L of the solution.

Two-Dimensional Chromatography. An automated, two-dimensional liquid chromatography system was used to fractionate acetic acid extracted, native amidinated, and disassembled amidinated ribosomal protein samples.^{28,35,37} Solvent handling was provided by Waters Alliance 2695 and 2795 chromatographs. The first dimension of this separation was a nonporous Tosoh-Haas SP-NPR SCX column (4.6 mm \times 35 mm, Tosoh Bioscience, Montgomeryville, PA) developed with a gradient shown in Supplemental Table 1 (Supporting Information). Effluent from this column was shunted directly to 20 Thermo BioBasic C4 Javelin guard columns (1.0 mm \times 20 mm) that serve as in-line traps for proteins. Switching between the traps was controlled by LabView (version 6.0, National Instruments, Austin, TX). A typical analysis involved serially loading three sets of twenty traps with acetic acid extracted, native amidinated and disassembled amidinated proteins. During the first 20 min of SCX elution, effluent was directed to Traps 1 and 2. Between 20 and 100 min, traps were switched every five minutes. Effluent from the last 10 min of the experiment was directed into Trap 20. After being loaded from the ion exchange dimension, traps were developed into a Waters Q-ToF micro mass spectrometer by running the 53 min long reversed phase gradient reproduced in Supplemental Table 2 (Supporting Information).

Liquid Chromatography–Tandem Mass Spectrometry (LC–MS/MS) of Peptide Samples. To corroborate identifications based on whole protein masses, proteins were eluted from

trapping columns by injecting 16 μ L of 80% isopropanol/20% water (v/v) with isocratic 95% mobile phase B for seven minutes. Samples were dried in a SpeedVac centrifugal concentrator (Thermo Jouan, Waltham, MA), neutralized and resuspended in 20 μ L of 10 mM ammonium bicarbonate solution and digested with 0.2 μ g of either proteomics grade alkylated porcine trypsin or endoprotease Glu-C for 16 h. Reactions were quenched by the addition of 20 μ L of 0.2% formic acid in water and analyzed by LC–MS/MS.

Solvent handling was provided by a Thermo Surveyor chromatography system (Thermo Scientific, Waltham, MA) with a flow splitter installed between the pump and the sample valve, so that flow from the pump could be split 1:20, providing a 5 μ L/min flow rate through the capillary columns used for these experiments. Peptide digests were analyzed on 254 μ m i.d. PEEK capillary columns (Upchurch Scientific) containing 5 μ m C18 silica beads (Phenomenex, Torrance, CA). The gradient used is shown in Supplemental Table 3 (Supporting Information).

Other Chromatography. The identifications of ribosomal proteins L11, L25, S2, and S5 were ambiguous due to post-translational modifications, endogenous proteolysis, or sequencing errors. To directly associate whole protein masses and tryptic peptides, and to obtain maximum sequence coverage of each protein, acetic acid extracts were chromatographed in a one-dimensional reversed phase experiment using a Phenomenex Jupiter C4 column (4.6 mm \times 25 cm). The gradient used is shown in Supplemental Table 4 (Supporting Information). Fractions containing single proteins or pairs of proteins were dried and digested with trypsin or endoprotease Glu-C as described above.

Fractions containing putative ribosomal protein S5 were also subjected to partial C-terminal sequence analysis using a mixture of carboxypeptidases Y and P as described previously.³⁵ After drying, the fractions were redissolved in 75 μ L of 100 mM pyridinium acetate buffer, mixed with 75 μ L of saturated (~12 M) urea and incubated at 37 $^{\circ}$ C for 15 min. Then, 4.2 μ g each of CPY and CPP were added. The reactions were incubated at 37 $^{\circ}$ C and quenched by withdrawing 45 μ L aliquots and mixing with 5 μ L of 90% formic acid at fixed times between 5 min and two hours. Samples were analyzed using 254 μ m i.d. PEEK capillary columns containing 5 μ m C4 silica beads (Phenomenex, Torrance, CA). The gradient shown in Supplemental Table 5 (Supporting Information) was used.

To make highly accurate mass measurements, aliquots of acetic acid extracted ribosomal proteins were chromatographed on fused silica nano-ESI tips (75 μ m \times 15 cm) prepared with a Sutter Instruments P-2000 Micropipette Puller (Sutter Instruments, Novato, CA) and packed with Phenomenex Jupiter C4 beads. The gradient shown in Supplemental Table 6 (Supporting Information) was delivered by a Dionex Ultimate-3000 chromatograph.

Whole Protein Mass Spectrometry. Whole protein 2DLC separations were detected using a Waters Q-ToF micro mass

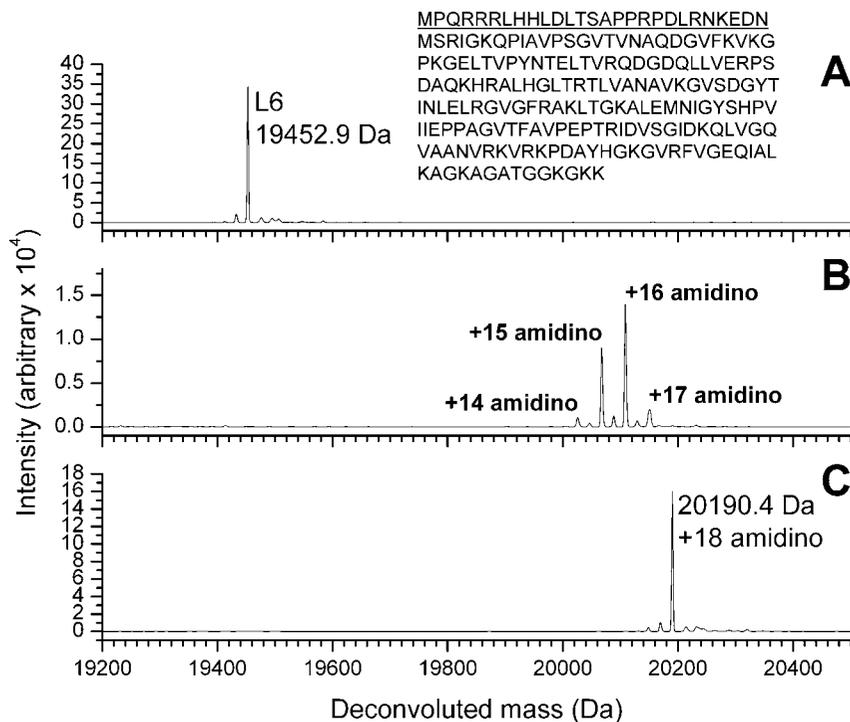


Figure 1. *D. radiodurans* ribosomal protein L6 mass spectra. (A) Unmodified protein, (B) native amidinated protein, and (C) disassembled, denatured amidinated protein. (Inset) L6 sequence from the proteome with a predicted mass of 22827.3 Da and 19 modifiable amino groups. Reassignment of methionine 28 as the start site and removal of the underlined portion of the sequence explains the experimentally observed mass of 19452.9 Da and leaves 17 lysines plus the amino terminus.

spectrometer. The flow rate was split from 50 $\mu\text{L}/\text{min}$ to 7 $\mu\text{L}/\text{min}$ into the ESI source. The instrument's spectral window was 600–1900 Th with 1 s scans and a 0.1 s interscan delay.

Total ion chromatograms (TICs) were analyzed by summing 100 scans at a time and deconvoluting to a target mass range of 4000 to 40000 Da using MaxEnt 1 (Waters/Micromass Milford, MA). Deconvoluted spectra were centroided, and tabulated masses and area intensities were used to calculate average masses for each protein. The mass spectrometer was calibrated before each set of experiments using a 50 $\text{pg}/\mu\text{L}$ solution of horse heart myoglobin in water/acetonitrile (50/50 v/v) containing 0.1% formic acid, resulting in deconvoluted masses that are accurate to ± 1.5 –2.0 Da.

Highly accurate masses were measured using a Thermo LTQ-FT Ultra hybrid ion trap/Fourier transform-ion cyclotron resonance mass spectrometer (Thermo Fisher, Bremen, Germany) equipped with a nano-ESI source. This instrument is calibrated weekly with the manufacturer's specified mixture [caffeine, methionyl-arginyl-phenylalanyl-alanine (MRFA, acetate salt) and Ultramark 1621] and tuned on the +3 charge state of bovine insulin chain B (Sigma, St. Louis, MO). Spectra were collected across a window of 300–2000 Th at a resolution of 100,000 and stored as centroided spectra. Raw spectra were extracted by summing across chromatographic peaks and deconvoluted using Xtract (Thermo Fisher, San Jose, CA). The reported isotopomer masses were obtained directly from the deconvoluted spectra.

Peptide MS/MS. Peptides in tryptic and Glu-C digests of unmodified proteins and native amidinated proteins were detected with a Thermo LCQ Deca XP Plus ion trap mass spectrometer.

Digest samples for the identification of ribosomal proteins L11, L25, S2, and S5 were also analyzed using an ABI 4700

MALDI-TOF/TOF instrument. Digestion was stopped by the addition of 5 μL of trifluoroacetic acid, and peptides were desalted and concentrated with small volume reverse phase columns containing 30–40 μm diameter octadecyl silica beads (Grace-Vydac, Deerfield, IL) immobilized in a polysulfone matrix. Peptides were eluted using a solution of 10 g/L α -cyano-4-hydroxycinnamic acid in 50/50 (v/v) water/acetonitrile with 0.5% trifluoroacetic acid and 1 μL of peptide-containing solution was deposited onto the probe.

Crystal Structures. The extent of labeling of proteins from the large ribosomal subunit was assessed using the published 3.1 Å resolution structure of the 50S subunit from *Deinococcus radiodurans* R1 (PDB file 1NKW).³⁰ The extent of labeling of proteins from the small ribosomal subunit was assessed using the 30S subunit structures from the published 2.8 Å crystal structures of the 70S ribosomal particle of *Thermus thermophilus* HB8 (PDB files 2J00 and 2J02). For *D. radiodurans* ribosomal proteins L1, L9, and L28, the 50S subunit files from *T. thermophilus* HB8 (PDB files 2J01 and 2J03) were also used as discussed below.³² Structures were visualized using Protein Explorer v. 2.79 beta.³⁸ To count visible, surface exposed lysines in a protein, solvent accessible surfaces were drawn around the protein and rRNA components of the structure using the default 1.4 Å probe sphere.³⁹ Then positions corresponding to lysine residues were displayed in a contrasting color visible against both the protein and rRNA surfaces.

Figures 7–10 were produced using PyMOL v. 0.99 (DeLano Scientific, www.pymol.org) and the recently released 2.9 Å resolution refinement of the original *D. radiodurans* crystal structure (PDB file 2ZJR)³¹ and one of the 2.8 Å 30S subunit structures from the *T. thermophilus* 70S ribosomal particle (PDB file 2J00).³²

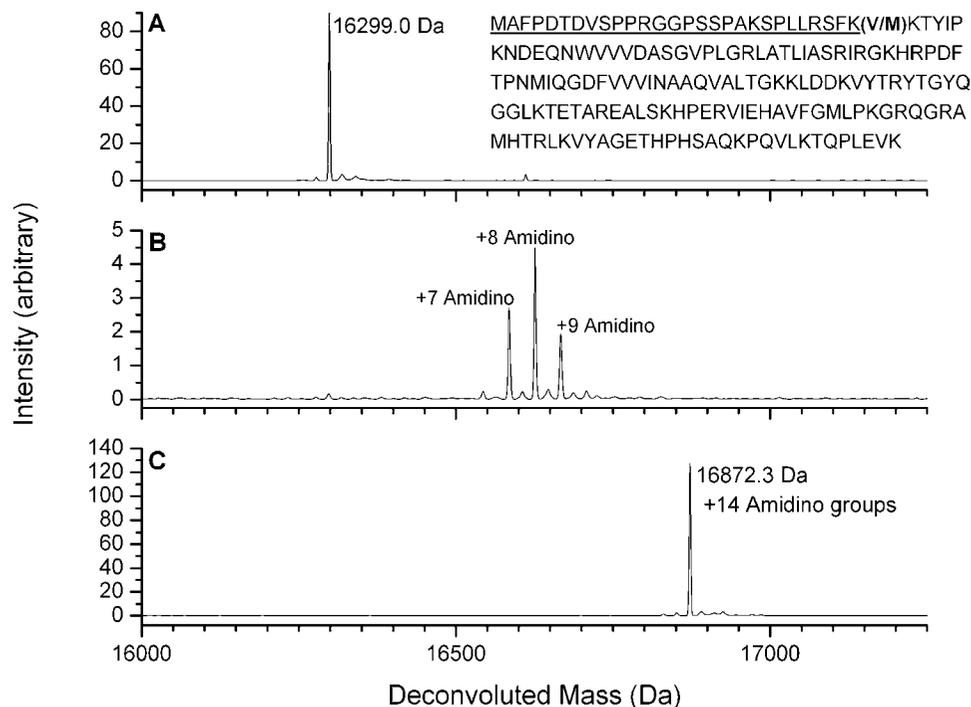


Figure 2. *D. radiodurans* L13 whole protein mass spectra. (A) Unmodified, (B) amidinated native, and (C) disassembled, denatured amidinated protein spectra. The inset shows the L13 sequence predicted by the proteome with a predicted mass of 19190.1 Da. The portion of the molecule that must be removed to match the experimental mass is underlined, and the proposed alternate start site that rationalizes this change is shown in parentheses.

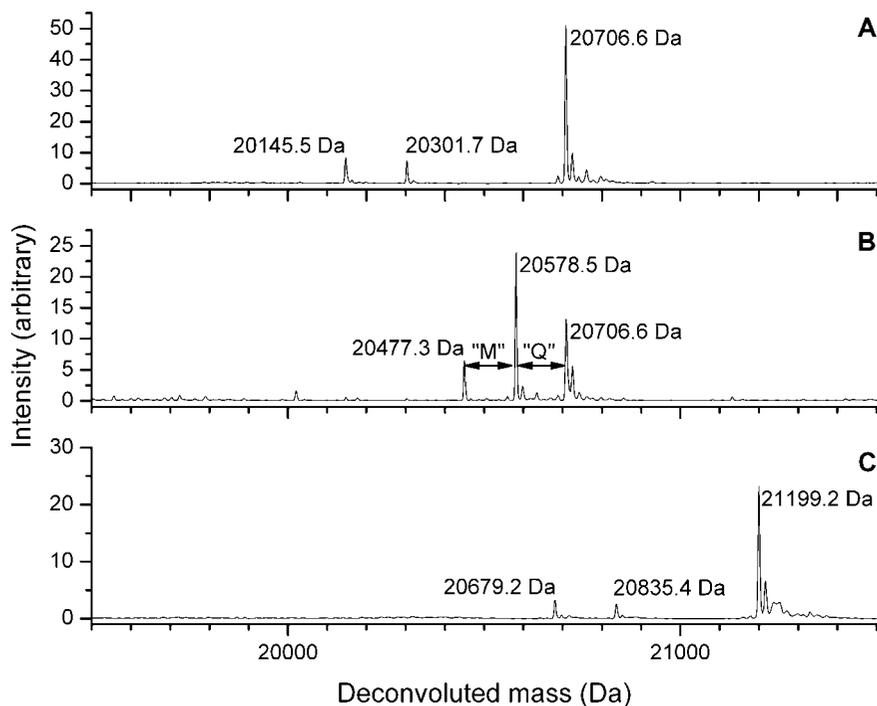


Figure 3. Mass spectra of ribosomal protein S5. (A) Deconvoluted spectrum showing the unmodified 20706.6 Da protein. (B) Deconvoluted spectrum from a CPY/CPP digest mixture (8.5 min reaction time). The mass differences between adjacent peaks are consistent with the C-terminal sequence of S5. (C) Deconvoluted spectrum from a disassembled, denatured amidination reaction. The mass 21199.2 Da corresponds to the addition of 12 amidino groups to 20707.6 Da. Subsidiary peaks (20301.7 Da/20835.4 Da and 20145.5 Da/20679.2 Da) are discussed in the main text.

Bioinformatics. The genomes and proteomes of *Deinococcus radiodurans* R1 and *Thermus thermophilus* HB8 were downloaded from The Institute for Genome Research's Comprehen-

sive Microbial Resource (<http://www.tigr.org>, now the J. Craig Venter Institute).⁴⁰ Other protein sequences were downloaded as necessary from the Swiss-Prot database, accessed at www.

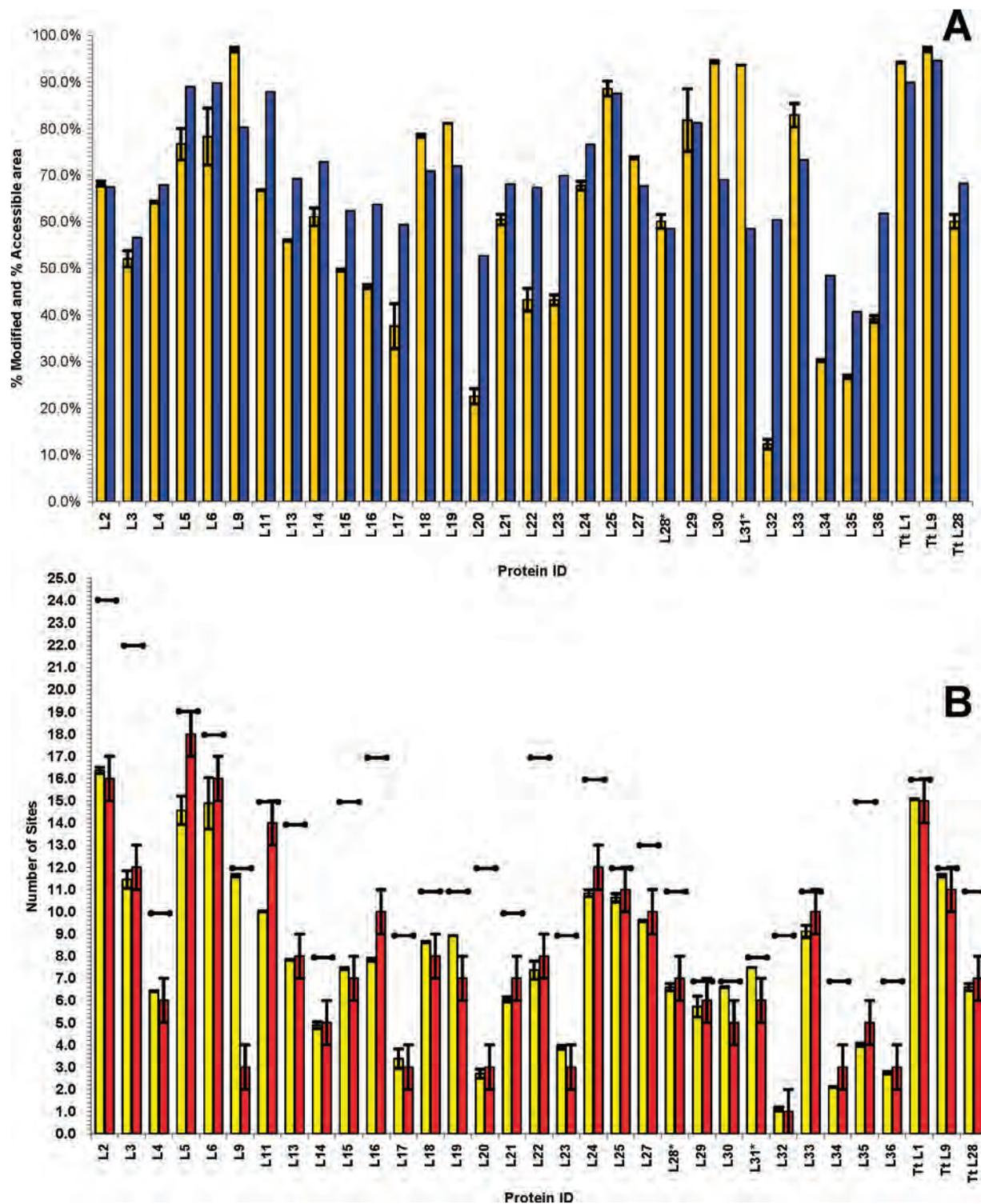


Figure 4. Comparison of labeling data for *D. radiodurans* large subunit ribosomal proteins using the *D. radiodurans* 50S crystal structure. (A) Weighted average percent of sites labeled (light bars) versus percentage of solvent accessible surface area (dark bars). (B) Weighted average extent of labeling (light bars) versus the count of lysine residues visible in the crystal structure (dark bars). Horizontal bars indicate the maximum extent of labeling. L28* and L31* are discussed in the text, as are the results for comparisons with *T. thermophilus* HB8 proteins labeled with a "Tt".

expasy.org. To determine the number of visible lysine residues in *D. radiodurans* R1 small subunit proteins using the crystal structures for *Thermus thermophilus* HB8 30S subunits, each *D. radiodurans* protein was aligned with its *T. thermophilus* homologue using ClustalW at the EMBL European Bioinfor-

matics Institute (<http://www.ebi.ac.uk/Tools/clustalw/>). Percentage identity or homology was calculated using the blast2p utility at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/blast/>). Solvent accessible surface areas (SASAs) for proteins in ribosomal subunits

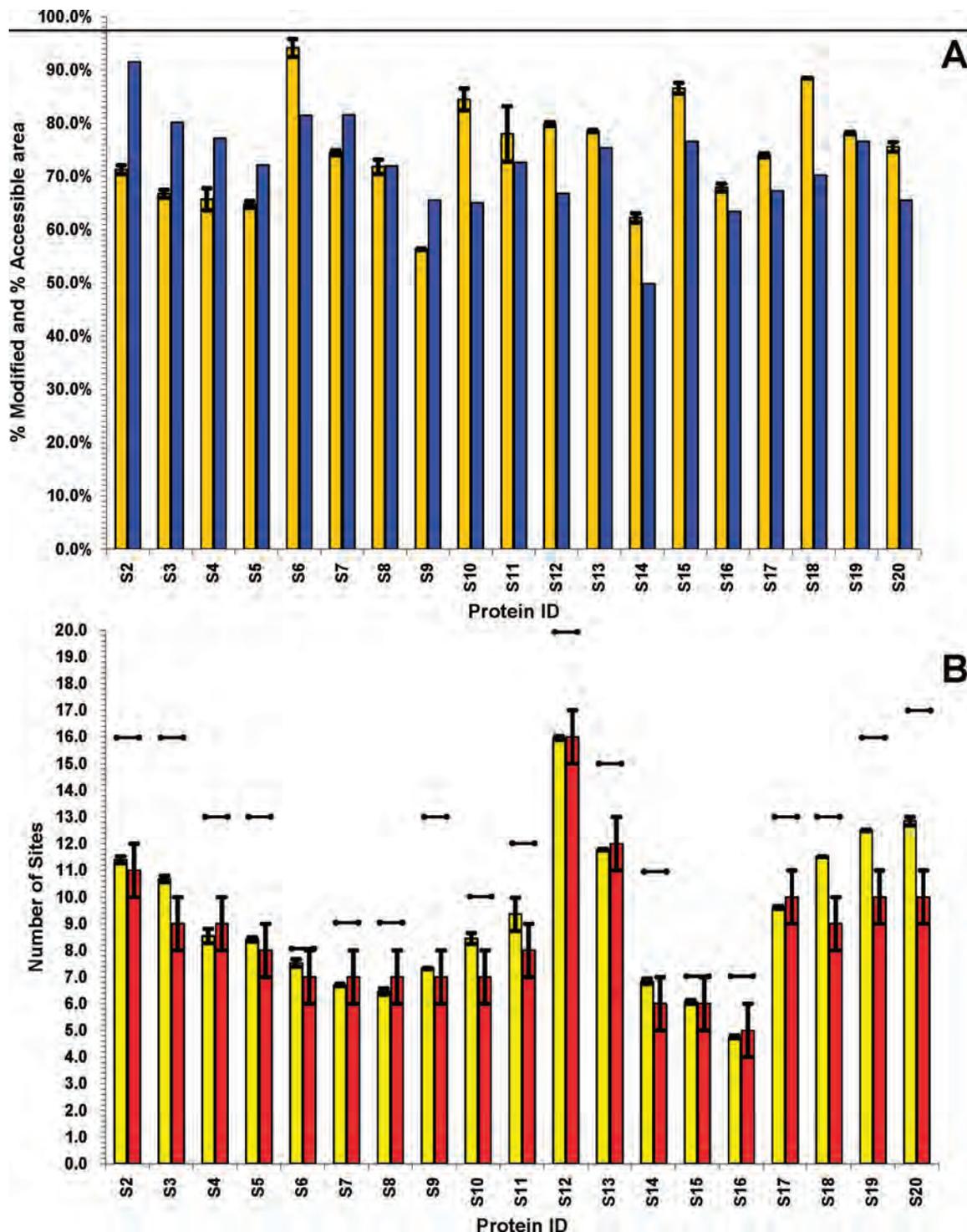


Figure 5. Comparison of amidination labeling data for *D. radiodurans* small subunit ribosomal proteins. Solvent accessible surface area and numbers of visible lysine residues are derived from sequence alignments between *D. radiodurans* and *T. thermophilus* proteins and the *T. thermophilus* 30S crystal structure. (A) Weighted average percent of sites labeled (light bars) versus percentage of solvent accessible surface area (dark bars). (B) Weighted average extent of labeling (light bars) versus the count of lysine residues visible in the crystal structure (dark bars).

were calculated using the program POPSCOMP (<http://zeus.cs.vu.nl/programs/popscompwww/>).⁴¹ The solvent accessible surface area of each protein in the intact ribosome was calculated as the difference between the SASA of the isolated molecule and one-half of the sum of the sequestered surface area calculated for each paired combination of ribosomal components.

Results

Protein Identifications. Several recent publications have used a bottom-up strategy to catalog the total proteome of *D. radiodurans* R1.^{42–44} These projects produced a global profile of protein expression as a function of growth state and environmental stress. In contrast, our research is focused on

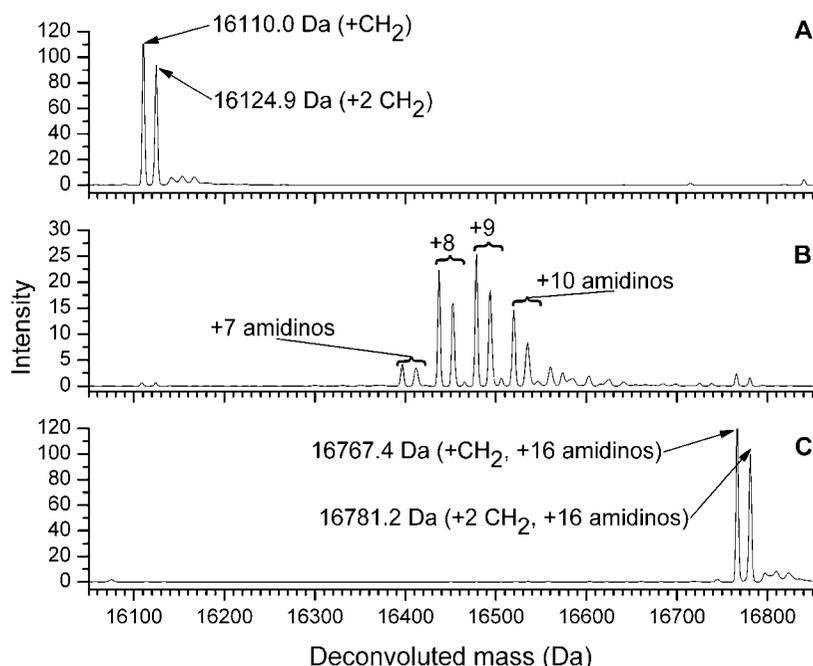


Figure 6. Whole protein spectra demonstrating post-translational modification of L16, and its effect on amidination. (A) Unmodified whole protein spectrum. Each mass corresponds to the predicted mass of L16, 16094.9 Da, plus one or two methyl groups. (B) Amidinated native whole protein spectrum. (C) Disassembled, denatured amidinated whole protein spectrum. This protein contains 17 modifiable groups, and the major form is singly methylated, presumably on the N-terminus. The position of the second methylation has not been determined.

the ribosomal proteome of this bacterium. The present results apply top-down and bottom-up methods to the proteome of a specific macromolecular complex, produce a description of its composition, post-translational modifications applied to its constituent proteins, and reveal several sequencing errors.

Protein identifications started with 2DLC separations of whole ribosomal proteins. Samples included unmodified proteins and those amidinated in the intact ribosome or following disassembly. After SCX fractionation, in-line trapping and C4 reverse phase chromatography, whole protein mass spectra were recorded. Tentative identifications in the unlabeled ribosomal protein sample were made by matching experimental masses to lists of isotopically averaged and monoisotopic masses derived from the translated proteome. Calculated masses in this list included those for potential post-translational modifications based on past observations in *E. coli*³⁴ and *C. crescentus*.³⁵ Examples include removal of N-terminal methionine by methionine aminopeptidase (MAP) when the second residue in a sequence is small (e.g., G, A, S, T, V, or P), methylation of amino groups, and acetylation of N-termini. The native amidinated and disassembled, denatured whole protein data were used to quantify the average extent of labeling in whole ribosomes and to confirm protein identifications by counting the number of reactive amino groups in each protein. Curiously, 14 of the 29 proteins (L1, l2, L4, L10, L12, L18, L20, L22, L23, L24, L33, S12, S18, and S20) predicted to be MAP substrates were found in two forms: the major form listed in Table 1, and one in which the N-terminal methionine was not removed by MAP. Relative quantities of the unmodified proteins range from 1–20% of the major, MAP-processed form. With only three exceptions, all of these proteins have alanine or proline as a second residue, and are expected to be good substrates for MAP.^{45,46} The biological significance of this violation of the canonical removal of the initiator methionine is not apparent, as the failure to process these proteins does

not appear to correlate with size, ribosome structure, or known regulatory functions of these proteins.

To confirm tentative identifications based on intact protein masses, separate 2DLC separations were performed on unmodified acetic acid extracts. Proteins adsorbed to each trap were eluted with high organic mobile phase, dried, digested with trypsin, and analyzed by LC-MS/MS. High sequence coverage for a protein in LC-MS/MS data from a trap containing a mass tentatively associated with that protein provided further supporting evidence for its identification, and Supplemental Scheme 1 (Supporting Information) presents a flowchart for this protein identification strategy. Sites of SMTA reactivity in the intact ribosomes were identified in similar experiments using C4 column fractions digested with either trypsin or Endoprotease Glu C.

The unlabeled and disassembled amidinated whole protein samples provide mutually reinforcing evidence for the identification of whole proteins. The total number of SMTA-reactive sites in a protein and the mass increase due to each modification can be calculated from the proteomic sequence. Pairing of an unmodified and a fully amidinated mass corroborates a protein identification because the two measurements contain both mass and composition (number of primary amines) data. This strategy is analogous to an approach used previously in peptide mass fingerprinting experiments, where unmodified tryptic digest mixtures were compared to digest mixtures that had been reacted with *O*-methylisourea. The mass shift introduced in a peptide by addition of an integral number of guanidino groups was used to count the number of lysine residues in the peptide, corroborating an identification based on the unmodified peptide mass.⁴⁷

Table 1 lists *D. radiodurans* ribosomal proteins, along with post-translational modifications necessary to rationalize the observed masses. The difference between the mass calculated from the proteomic sequence and the experimental value is

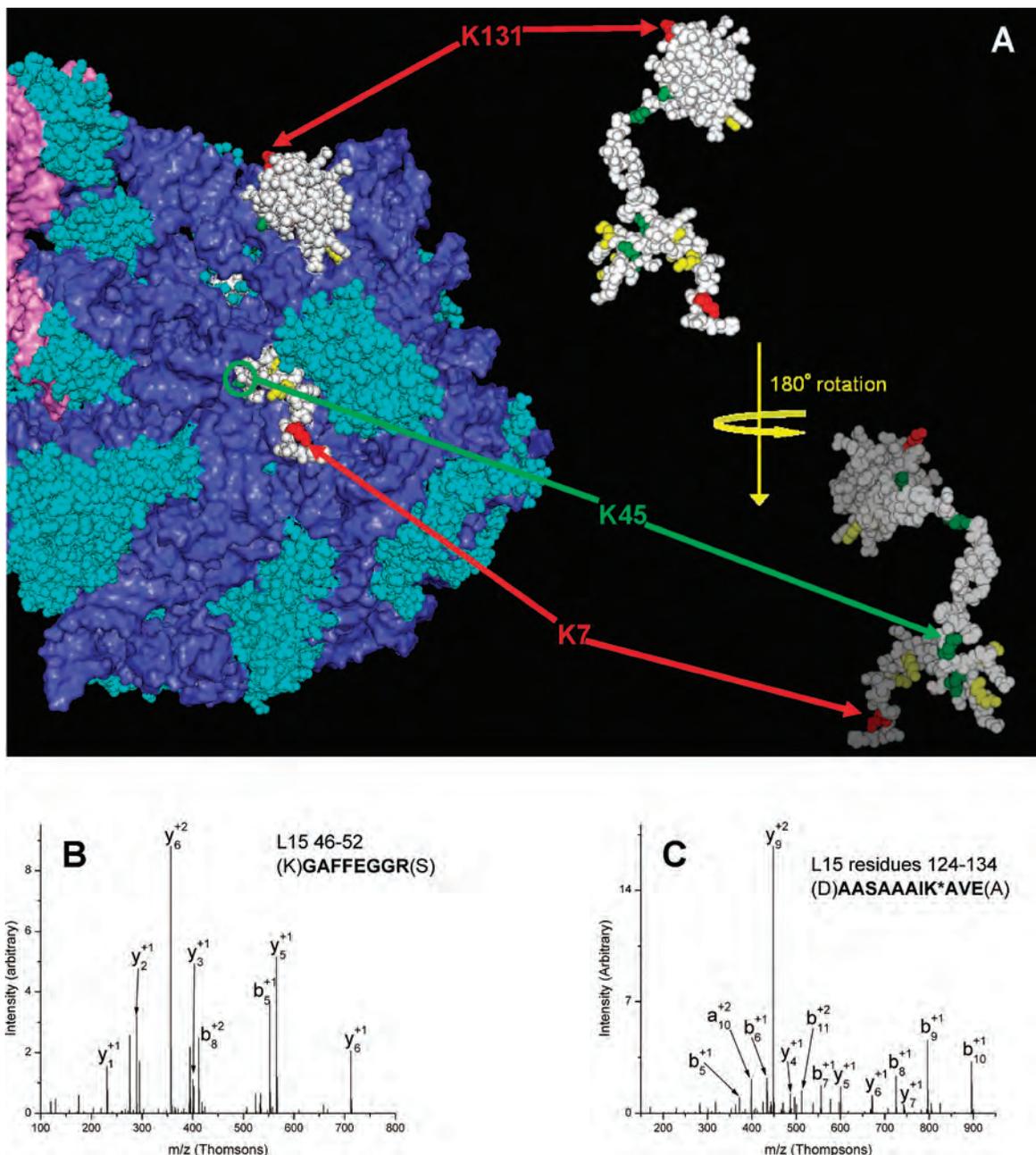


Figure 7. Large ribosomal subunit structures showing the position of protein L15 which is shown in white. Other proteins are depicted in light blue. The 23S rRNA is presented as a dark blue solvent accessible surface, while the 5S rRNA is shown in purple. (A) On the right, the isolated protein is shown. (B) Ion Trap CID MS/MS spectrum of a peptide demonstrating that K45 is not amidinated in native ribosomes. (C) MS/MS spectrum containing an amidinated K131 from native ribosomes. Peptides with amidinated residue K7 and unmodified residues K28, K62, K76, and K107 were also observed. Lysine residues depicted in red were observed with an added amidino group in LC-MS/MS analyses of enzymatic digests while those in green were found to be unmodified. Peptides containing lysines in yellow were not observed.

useful in diagnosing a protein's PTMs and is shown in the Δm column. For proteins not requiring some PTM to rationalize the observed mass, it can be seen that the experimental masses are well within the 1.5–2.0 Da error expected from our calibration with myoglobin. Identifications in Table 1 are based on the combination of consistently modified pairs of unlabeled and disassembled amidinated protein masses with high sequence coverage in peptide analyses.

An FT-ICR instrument was used to reinforce these identifications by making highly accurate whole protein mass measurements. Following the convention of several recent publications,

we have reported the most abundant isotopic mass in the deconvoluted spectrum of each protein in Table 1.^{48,49} The italicized number appended to each experimental mass indicates which isotopomer peak is reported. For example, the entry for protein L14, "14231.66-8", indicates that the most intense peak in this protein's spectrum corresponds to the eighth peak above the monoisotopic mass. The masses calculated for comparison with FT-ICR data in Table 1 are derived from the protein's molecular formula, including sequencing errors, deletions and PTMs, by adding a multiple of the average mass increment between peaks in a theoretical isotopomer

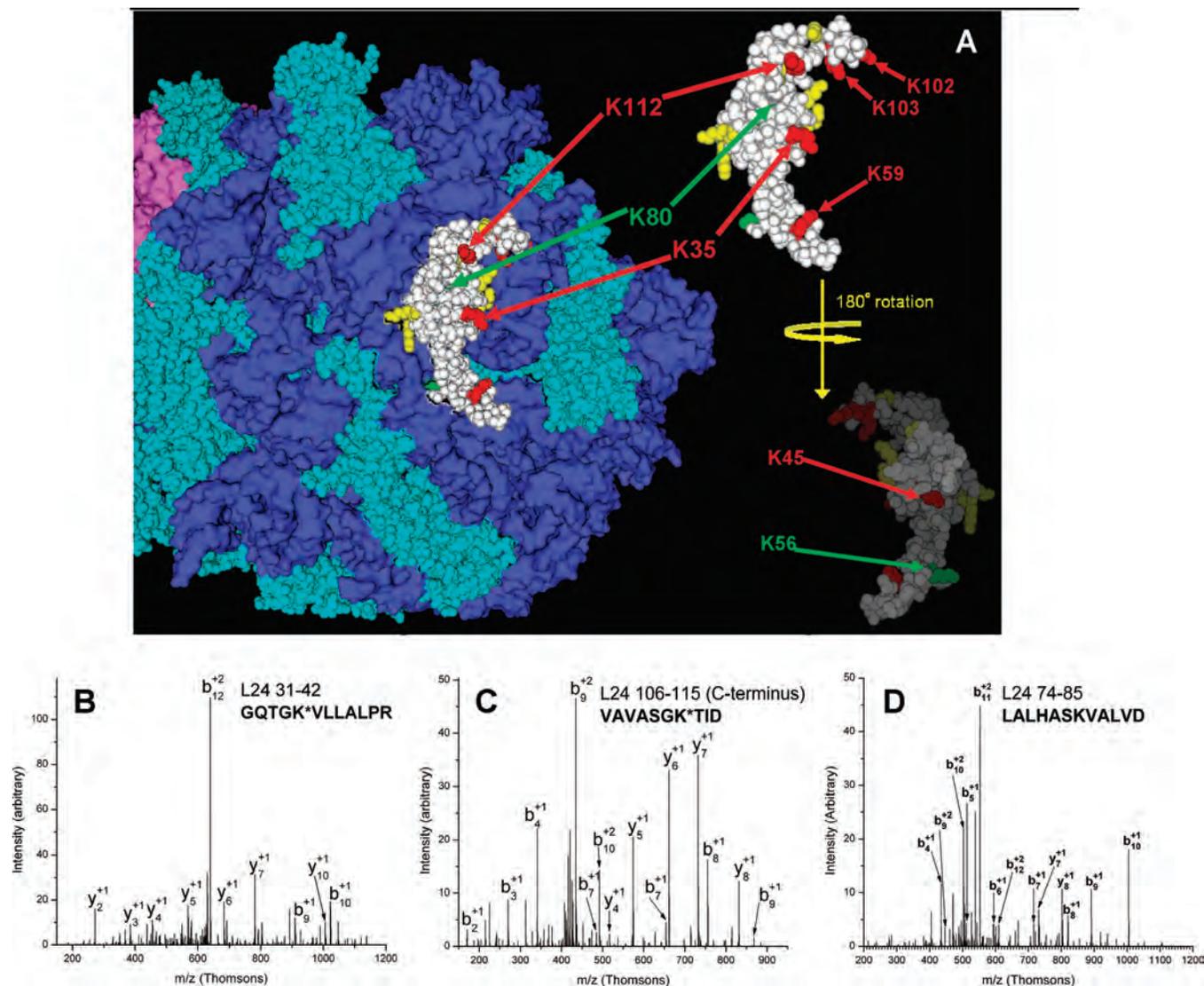


Figure 8. Large ribosomal subunit structures showing the position of protein L24. Color coding is as described in the caption of Figure 7. (A) Entire 50S subunit and the isolated protein. (B) and (C) lon trap CID MS/MS spectra demonstrating the amidination of lysine residues 35 and 112. (D) lon trap CID MS/MS spectrum of a peptide containing unmodified K80. Peptides containing amidinated K45, K59, K102, and K103 and unmodified K30 and K56 were also observed.

distribution (1.00235 Da) to the calculated monoisotopic mass.⁵⁰ Again using protein L14 as an example, the theoretical mass of the eighth isotopomer peak was calculated by adding 8×1.00235 Da to the monoisotopic mass calculated from the protein's molecular formula, 14223.66 Da. Calculated and experimental masses from FT-ICR experiments are compared using their relative mass difference in parts-per-million (ppm). These mass measurements show low relative errors (1–4 ppm, with a significant number of measurements showing errors between –1 and 1 ppm) and corroborate modifications that we have proposed for each protein, although not their locations or biological significance. The FT-ICR results are discussed further below.

This combination of techniques allowed the identification of several sequence errors in the *D. radiodurans* ribosomal proteome, as described next for proteins L6, L13, L21, and S5.

Ribosomal Protein L6. Although the L6 proteomic sequence predicts a 212-residue protein with a mass of 22827.3 Da, no protein of this mass was observed. Nevertheless, fractions containing an unknown with a mass of 19452.9 Da produced

high sequence coverage (67%) of peptides from L6. This result is understandable if the AUG codon at position 81 in the genomic sequence is the true start site. In that case, the protein's mass would be 19581.6 Da. Removal of the initiator methionine results in a predicted mass of 19450.4 Da, in good agreement with our observed mass. Figure 1A–C shows that this protein also contains 18 modifiable residues, consistent with this proposed sequence correction. The mass of the 10th isotopomer peak is within 2 ppm (0.04 Da) of the predicted mass. These results confirm the sequence correction in the Swiss-Prot database in database entry Q9RSL3.

Ribosomal Protein L13. The predicted mass of ribosomal protein L13, 19190.1 Da, was also not experimentally found, even after allowing for canonical post-translational modifications. This mass is 3 to 4 kDa larger than homologues from *Deinococcus geothermalis* and closely related strains of *T. thermophilus*. A multiple sequence alignment of L13 sequences is shown in Supplemental Figure 1 (Supporting Information). This alignment suggests that the actual initiator codon is the GTG at position 84 of the gene sequence, an alteration that

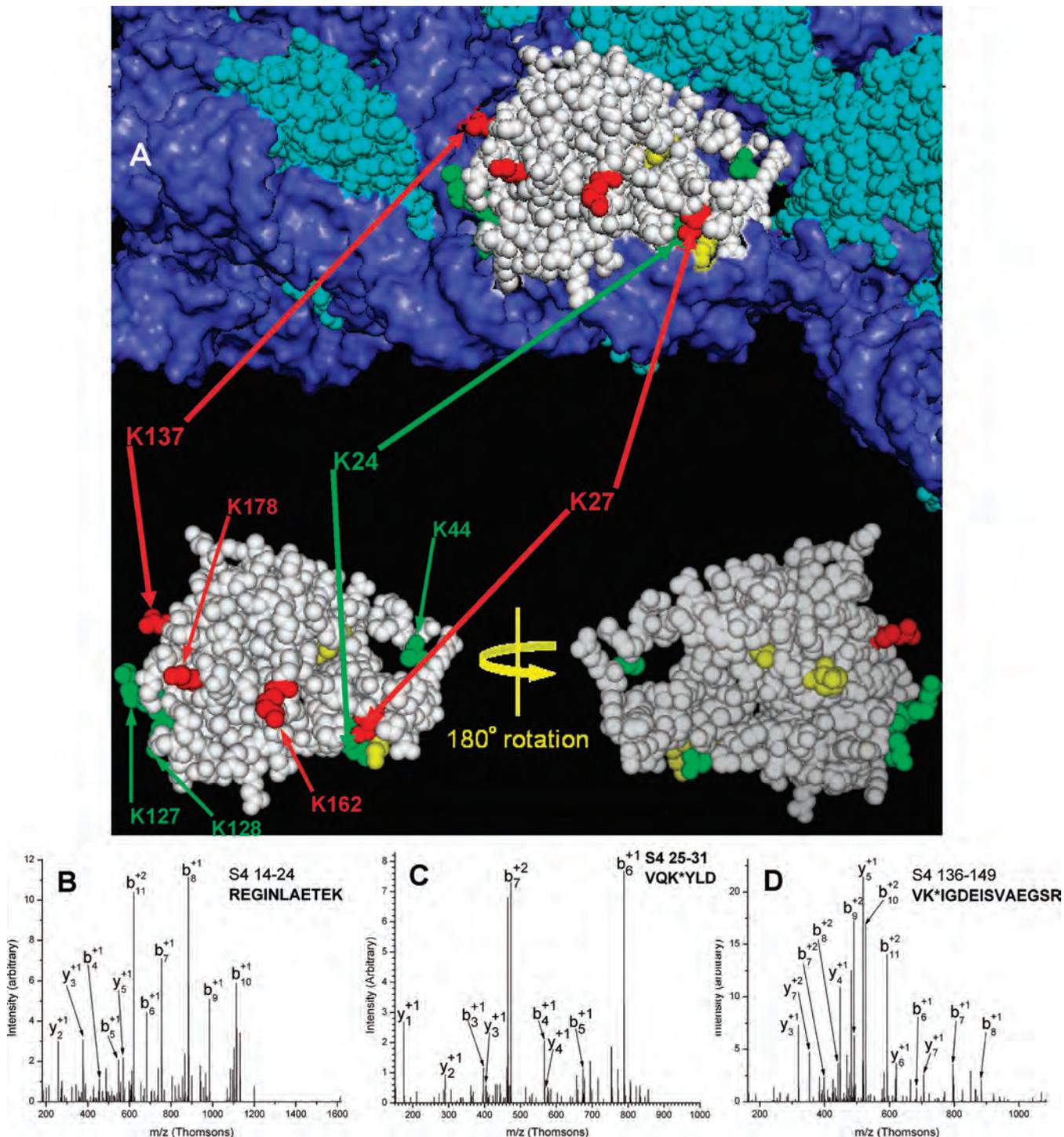


Figure 9. (A) Ribosomal protein S4 in position on the 30S subunit and isolated from it. Color coding of lysine residues is described in the caption of Figure 7. (B) Ion trap CID MS-MS spectrum demonstrating that lysine 24 is unlabeled in native amidated ribosomes. (C) and (D) MS/MS spectra demonstrating modification of lysine 27 and 137.

could be due either to a sequencing error or to the use of GTG as a nonstandard initiator codon.⁵¹ Correcting this error involves removing 28 amino acid residues from the protein sequence and substituting a methionine residue for the valine currently at position 29. This methionine for valine substitution would occur whether the error is a sequencing error or a misassigned start codon, and results in a predicted mass of 16298.6 Da. This predicted mass is within experimental error of an intense peak at 16299.0 Da. Tryptic digests of fractions containing this 16299.0 Da protein show 63% sequence cover-

age of L13 when analyzed with LC-MS/MS. This sequence correction removes two lysine residues and changes the total number of amidation sites from 16 to 14. Figure 2A-C shows the results obtained for this mass in unmodified, native amidation, and disassembled amidation experiments. The disassembled amidation experiment indicates that this protein has 14 labelable sites, matching exactly the number of sites available after the proposed sequence correction. The low, 1 ppm (0.02 Da) difference between the ninth isotopomer peak's calculated mass and the experimentally determined mass adds

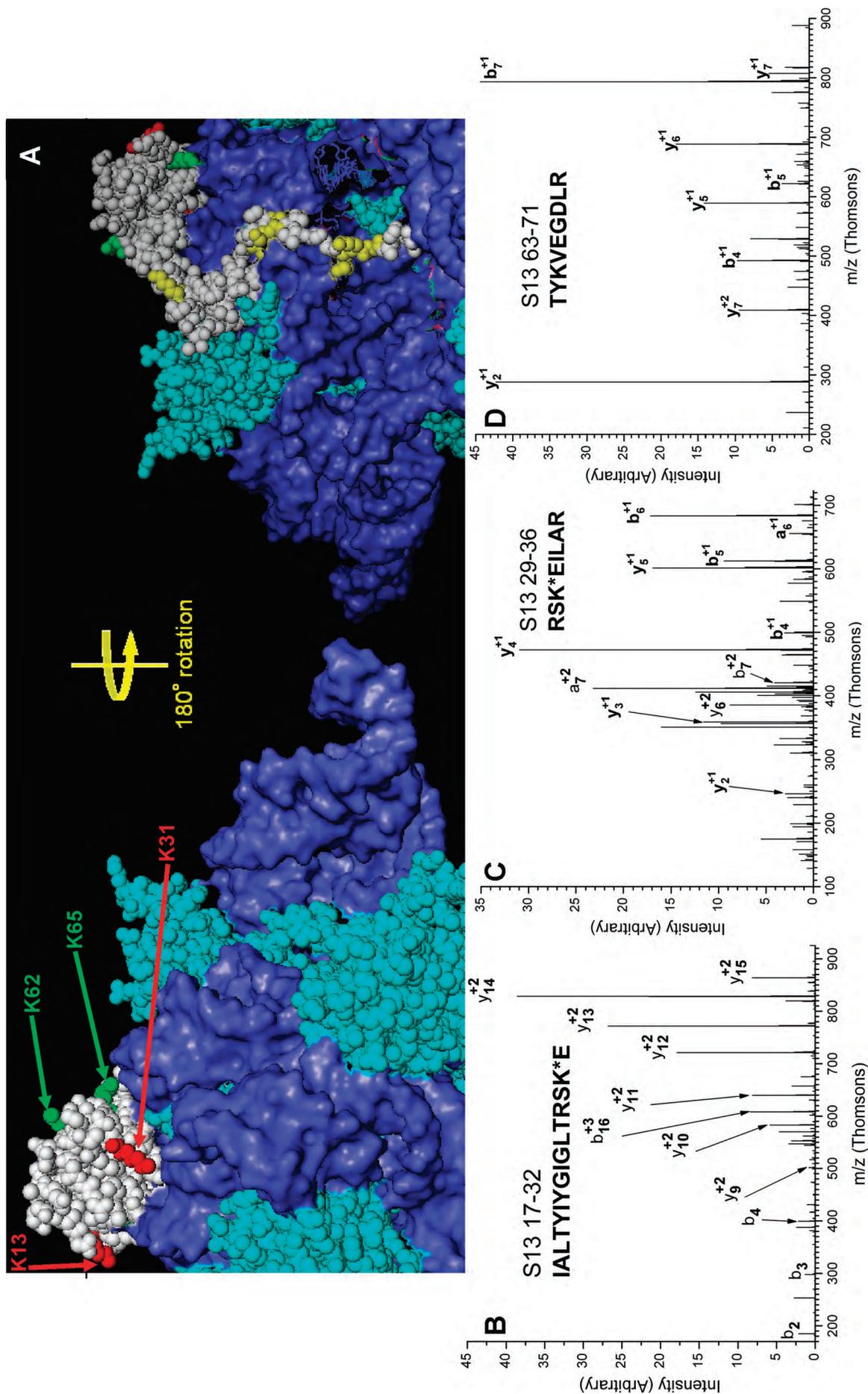


Figure 10. (A) Two views of ribosomal protein S13 in position on the 30S subunit, related by a 180° rotation around the vertical axis. Color coding is described in the caption of Figure 7. (B) and (C) Ion Trap CID MS/MS spectra of overlapping tryptic peptides of native ribosomes and tryptic peptides demonstrating the amidation of lysine residue 31 in native ribosomes. (D) Ion trap MS/MS spectrum of a peptide generated by tryptic cleavage at K62, containing unmodified K65.

Table 1. Observed Ribosomal Proteins *Deinococcus radiodurans* R1

protein	isotopically averaged masses			FT-ICR determined masses			modifications ^e	extent of labeling			%sequence coverage ^f
	calc.	obs.	Δm^a	calc. ^b	obs. ^c	ppm ^d		max ^f	native ^g	disassembled ^h	
L1	24359.9	24231.5	128.4	24227.86	24227.81-14	2	-Met	16	15.1 ± 0.0	16	68.3
L2	30044.6	29916.6	128.0	29911.29	29911.22-16	3	-Met	24	16.4 ± 0.1	24	66.2
L3	22437.1	22467.9	-30.8	22465.04	22465.00-14	2	+28	22	11.4 ± 0.4	21	68.3
L4	22278.1	22150.4	127.7	22146.62	22146.60-13	1	-Met	10	6.4 ± 0.0	10	78.1
L5	20349.9	20380.4	-30.5	20379.02	20378.98-14	1	+28	19	14.6 ± 0.5	18	65.6
L6	19581.6	19452.9	128.7	19448.64	19448.60-10	2	-Met, Seq. error ^j	18	14.9 ± 1.2	18	67.0
L7/L12	12627.5	12512.0	115.5	12509.64	12509.63-7	<1	-Met, +CH ₂ ^k	14	13.2 ± 0.1	13, 14	83.6
L9	16065.9	16068.0	-2.1	16064.57	16064.54-9	2		12	11.6 ± 0.1	12	70.6
L10	17754.0	17625.0	129.0	17622.31	17622.31-10	<1	-Met	12	9.8 ± 0.1	12	89.4
L11	15012.5	15014.7	-2.2	15011.04	15011.03-8	1	+10(CH ₂).	15	10.0 ± 0.0	11	63.2
L13	16296.8	16299.0	-2.2	16295.71	16295.69-9	1	Seq. error ^j	14	7.8 ± 0.2	14	63.2
L14	14232.5	14234.9	-2.4	14231.68	14231.66-8	2		8	4.9 ± 0.2	8	67.9
L15	16860.1	16862.1	-2.0	16859.06	16859.04-9	1		15	7.4 ± 0.0	15	67.3
L16	16094.9	16111.0	-16.1	16106.55	16106.51-8	3	+CH ₂ , +2(CH ₂)	17	7.8 ± 0.1	16	57.8
L17	12899.0	12900.5	-1.5	12898.13	12898.10-7	2		9	3.4 ± 0.4	9	64.7
L18	12138.9	12009.0	129.9	12007.58	12007.57-7	<1		11	8.6 ± 0.0	11	50.0
L19	18315.7	18317.7	-2.0	18313.76	18313.74-9	1		11	8.9	11	34.3
L20	13957.2	13828.1	129.1	13825.64	13825.62-8	1	-Met	12	2.7 ± 0.2	12	54.2
L21	11146.7	11148.0	-1.3	11145.94	11145.94-6	<1	Seq. error ^j	10	6.0 ± 0.1	10	37.9
L22	15158.8	15071.7	87.1	15068.45	15068.41-8	2	-Met, +Ac ^k	17	7.4 ± 0.4	16	55.2
L23	10522.2	10392.3	129.9	10390.60	10390.58-6	2	-Met	9	3.9 ± 0.1	9	76.8
L24	12358.3	12228.9	129.4	12226.80	12226.79-7	1	-Met	16	10.8 ± 0.2	16	71.3
L25	25390.4	25393.5	-3.1	25389.88	25389.85-15	1		12	10.6 ± 0.2	Not Seen ^l	52.2
L27	9590.0	9460.1	129.9	9458.11	9458.10-5	1	-Met	13	9.6 ± 0.0	13	52.8
L28	8962.5	8832.3	130.2	8831.01	8831.00-5	1	-Met	11	6.6 ± 0.2	11	33.7
L29	7760.0	7761.2	-1.2	7759.15	7759.14-4	<1		7	5.7 ± 0.5	7	64.2
L30	6067.2	6068.2	-1.0	6066.45	6066.45-3	<1		7	6.6 ± 0.0	7	63.6
L31	8581.8	8582.8	-1.0	8581.32	8581.32-5	1		8	7.5 ± 0.0	8	30.1
L32	6791.9	6657.2	134.7	6659.41	6659.41-3	1	-Met	9	1.1 ± 0.2	9	58.3
L33	6359.5	6229.2	130.3	6227.42	6227.42-3	<1	-Met	11	9.1 ± 0.3	11	31.7
L34	5608.5	5609.2	-0.7	5608.09	5608.08-3	<1		7	2.1 ± 0.0	7	31.9
L35	7426.0	7295.8	130.2	7294.09	7294.08-4	<1	-Met	15	4.0 ± 0.1	15	34.9
L36	4309.2	4308.1	1.1	4308.33	4308.33-2	<1		7	2.7 ± 0.1	7	Not seen
S2	29755.7	29670.0	85.7	29664.46	29664.43-16	1	-Met, +Ac	16	11.4 ± 0.1	Not Seen	57.6
S3	27307.4	27177.9	129.5	27176.93	27176.93-17	<1	-Met	16	10.7 ± 0.1	16	55.4
S4	23844.1	23728.1	116.0	23725.41	23725.34-13	3	-Met, +CH ₂ .	13	8.5 ± 0.3	13	73.7
S5	20792.7	20707.6	85.1	20702.90	20702.91-12	<1	+Ac, Seq. error ^j	13	8.4 ± 0.1	12	75.5
S6	11672.2	11673.5	-1.3	11671.06	11671.06-6	1		8	7.5 ± 0.1	8	89.2
S7	17941.7	17812.7	129.0	17809.44	17809.41-10	2	-Met	9	6.7 ± 0.0	9	55.8
S8	15048.4	15050.5	-2.1	15047.17	15047.15-8	2		9	6.5 ± 0.1	9	46.6
S9	14733.1	14603.8	129.3	14601.06	14601.03-8	2	-Met	13	7.3 ± 0.0	13	43.6
S10	12109.1	12110.4	-1.3	12108.64	12108.62-7	2		10	8.4 ± 0.2	10	59.8
S11	13790.7	13675.7	115.0	13674.19	13674.13-9	4	-Met, +CH ₂ .	12	9.4 ± 0.6	11,12	57.3
S12	14353.9	14270.5	83.4	14268.10	14268.07-8	2	-Met, +SCH ₃ ^m	20	16.0 ± 0.1	20	39.7
S13	14227.5	14098.3	129.2	14095.94	14095.91-8	2	-Met	15	11.8 ± 0.0	15	63.5
S14	10207.8	10077.6	130.2	10075.43	10075.41-5	1	-Met.	11	6.8 ± 0.1	11	12.2
S15	10150.6	10152.0	-1.4	10150.49	10150.46-6	3		7	6.1 ± 0.1	7	67.0
S16	9667.0	9668.4	-1.4	9666.19	9666.18-5	<1		7	4.8 ± 0.1	7	71.4
S17	10782.4	10783.5	-1.2	10781.84	10781.82-6	2		13	9.6 ± 0.0	13	64.6
S18	10534.3	10535.7	-1.4	10533.98	10533.96-6	2	-Met, +Ac	13	11.5 ± 0.0	12	46.7
S19	10822.6	10692.9	129.7	10690.78	10690.77-6	1	-Met	16	12.5 ± 0.0	16	55.8
S20	10002.7	9872.9	129.8	9870.62	9870.59-5	3	-Met	17	12.8 ± 0.2	17	27.2

^a Calculated minus Observed isotopically averaged mass. ^b Calculated isotopomer mass takes into account PTMs and sequence errors. ^c Most intense isotopomer peak. The hyphenated, italicized suffix indicates the isotopomer identified (see text). ^d Parts-per-million difference between calculated and experimental masses. “<1” indicates a value less than 1 or greater than -1. ^e Modifications: “-Met”: N-terminal methionine removed, “+n(CH₂)”: addition of n methyl groups, “+Ac”: addition of an acetyl group, “+28”: a dimethylation, retention of an N-terminal formyl group, or a lysine for arginine sequencing error. ^f Max: the maximum number of amidino groups that can be added to a protein. Equal to the number of lysines, plus the amino terminus. ^g Native: The intensity-weighted extent of labeling in native amidinated experiments. Errors are standard deviation of three determinations. L21 shows no standard deviation because detectable intensity was observed for this protein in only one of three experiments. ^h Disassembled: The number of amidino groups added to protein denatured in 6 M urea buffered with 25 mM ammonium bicarbonate. ⁱ Maximum percent sequence coverage seen in LC-MS/MS separations of tryptic digests. ^j “Seq. error”: Major sequencing error covered in the Discussion section. ^k L7/L12 that retains the N-terminal methionine is also methylated; L22 that retains the N-terminal methionine is not acetylated. ^l Proteolytically damaged version of L25 containing residues 1-209 was observed. This fragment contained 11 total sites (lacking K218), had a mass of 22556.3 Da unmodified, and appeared with a mass of 23005.6 Da (+11 amidino) in denatured experiments. ^m S12 homologues in *E. coli*, *R. palustris*, and *B. subtilis* are β-thiomethylated (+46.1 Da) at an aspartate homologous to D88 in *E. coli*.

further support to the conclusion that the start site of L13 has been misassigned.

Ribosomal Protein L21. An intense, recurring unknown with a mass of 11148.0 Da yielded peptides that gave 38% sequence

coverage of ribosomal protein L21. This mass matches that predicted for residues 70–129 of the L21 proteomic sequence, suggesting that the GAG codon of residue E69 in the current sequence is a misread UAG stop codon for a 69 residue protein immediately preceding the actual L21 sequence. This correction removes four lysine residues from the protein sequence. Supplemental Figure 2A–C (Supporting Information) shows that the extent of amidination of the 11148.0 mass is consistent with the corrected sequence: 10 amidino groups rather than 14. As with protein L6, these results corroborate a correction to the Swiss-Prot database in sequence Q9RY64. This protein's sixth isotopomer peak has a mass that is within less than 1 ppm (<0.01 Da) of the corresponding calculated mass, sufficiently accurate to support our conclusions. There is no sign of the 69 residue protein (MW 7332.6 Da) in our data. When a BLAST search (<http://blast.ncbi.nlm.nih.gov/>) was performed for this presumably untranslated sequence, the closest matches were to small segments of a transglutaminase isoform from *Rattus norvegicus* and *Mus musculus*, and to a portion of a transcriptional regulator from the β -proteobacterium *Thauera* sp. MZ1T.

Ribosomal Protein S5. We found no match for the predicted 20976.9 Da mass of ribosomal protein S5. An intense whole protein mass of 20706.6 Da, shown in Figure 3A, was observed in fractions giving high sequence coverage (70–74%) of ribosomal protein S5. The result of digestion of this protein with a mixture of CPY and CPP is shown in Figure 3B. The sequential loss of 128 and 131 Da from the 20706.6 Da mass implies a C-terminal sequence of “MQ”. The result of amidinating the protein under denaturing conditions is shown in Figure 3C. The 21199.2 Da mass corresponds to the addition of 12 amidino groups to the 20706.6 Da protein.

An alignment of ribosomal protein S5 sequences shown in Supplemental Figure 3 (Supporting Information), and in particular the sequence from *D. geothermalis*, whose S5 homologue shows 93% sequence identity to *D. radiodurans*' S5, presents an explanation for the 270.3 Da difference between S5's predicted and observed masses.⁵² The first nine residues in the predicted sequence of ribosomal protein S5 are MALTFNRRN. If the leucine codon at the third position, a TTG, was actually the start codon due either to a sequence error or its use as a nonstandard initiator codon, the predicted mass of the protein would be 20792.7 Da, or 20661.5 Da with removal of the N-terminal methionine. The addition of an acetyl group gives a predicted mass of 20704.0 Da, closely matching the observed mass of 20706.6 Da. The C-terminal sequence of S5 from residue 190 to the C-terminus is “ADTGGMQ”. Ribosomal protein S5 is predicted to contain 12 lysine residues and a free amino terminus, allowing the addition of 13 amidino groups and giving a mass increase of 533.6 Da for complete amidination under denaturing conditions. A mass increase corresponding to the addition of 12 amidino groups is consistent with the presence of a single unreactive amino group, possibly blocked by a post-translational modification, and N-terminal acetylation has been observed in *E. coli*'s S5 homologue.³⁴ The data in Figure 3 are all consistent with the identification of the 20706.6 Da mass as ribosomal protein S5 with a sequence error near the N-terminus and an acetylation. Removal of two amino acid residues and addition of an acetyl group yields a predicted mass that is within about 1 ppm of the measured mass of the protein, as shown by the twelfth isotopomer's mass of 20702.90 Da.

The identification of the 20706.6 Da mass as S5 and the proposed post-translational N-terminal acetylation are also

supported by two masses produced by an endogenous protease that was not inhibited by PMSF or the protease inhibitor cocktail. In Figure 3A, the 20706.6 Da mass is accompanied by low intensity masses that correspond to fragments of S5 containing residues R7-Q196 (20301.7 Da) and R8-Q196 (20145.5 Da). In Figure 3C, the masses 20835.4 and 20679.2 Da correspond to the addition of 13 amidino groups to the R7-Q196 and R8-Q196 fragments, respectively. Neither truncation removes any lysine residues from S5, and would expose a new, free amino terminus, unblocked by acetylation.

Differential Amidination. Labeling data for proteins from the large subunit are summarized in Figure 4. Intensity-weighted averages for the number of amidino groups added were calculated from the distribution of labels in native amidinated protein mass spectra and are reported in Table 1, with a horizontal black bar to indicate the maximum number of reactive groups for each protein in Figure 4B. Errors are the standard deviation from three determinations. The extents of labeling for the disassembled ribosomal protein sample are all integral numbers because with the exception of ribosomal proteins L7/L12 and S11 only fully amidinated protein was seen. Ribosomal protein L7/L12 is known to retain some structure and to self-associate even under strongly denaturing conditions, such as the 6 M urea solutions used by ourselves and others for SCX chromatography.^{53,54} The robust structure of this protein explains its partial reactivity, and a very stable structure is also the most likely explanation for incomplete modification of S11 in the disassembled, denatured reactions. Examples of the completeness of labeling of other proteins are shown in Figures 1C, 2C, 3C and Supplemental Figure 2C (Supporting Information). Visible residues were counted by inspecting the *D. radiodurans* R1 50S subunit crystal structure 1NKW as detailed above, and the error in this count was assumed to be ± 1 . Data for the last four proteins in Figure 4 were derived from the *T. thermophilus* HB8 crystal structure using the alignment procedure described above for small subunit proteins. Ribosomal protein L1 is not present in the *D. radiodurans* crystal structure³⁰ and the other three proteins, L9, L28, and L31 are discussed below.

Figure 4A shows a general correlation between solvent accessible surface area and the extent of labeling. However, differences between the percentage of modified lysine residues and the percentage of solvent accessible surface area for proteins L3–L6, L11–L17, L20–L25, and L34–L36 suggest a preference for burying lysine residues in tertiary or quaternary structure. A fractional solvent exposed surface area larger than the fraction of modified lysine residues suggests that the distribution of lysine residues on the surface of the protein is nonrandom, and is biased toward lysine interactions with the rRNA. When the labeling results are replotted to compare the intensity weighted average number of amidinations with the number of lysine residues not buried in protein tertiary structure or protein–rRNA quaternary structure, better agreement is obtained as shown in Figure 4B.

Comparisons of the extent of labeling with either SASA or with visible lysine counts for the proteins from the 30S subunit are shown in Figures 5A and B, respectively. Lysine counts were obtained by mapping the sequences of *D. radiodurans* small subunit proteins onto the *T. thermophilus* crystal structures using ClustalW to align the sequences of homologous proteins. The two sets of small subunit proteins show between 43% and 80% sequence identity to each other (average 67%; sequence homology ranges from 69 to 92%). The extent of modification

again correlates better with the count of visible modifiable groups than with the solvent accessible surface area.

Discussion

In a previous study, we found very good agreement between the experimentally determined extent of amidination of *Caulobacter crescentus* ribosomes and the solvent accessible surface area or the number of visible modifiable groups derived from crystal structures of *D. radiodurans* and *E. coli* ribosomal subunits.²⁸ In the present study, in which the extent of modification of *D. radiodurans* ribosomes was evaluated using the crystal structures of *D. radiodurans* own 50S subunit and the 30S subunit of the closely related *Thermus thermophilus* HB8, there is an even better correlation between experimental and predicted extents of labeling. The agreement is especially striking for the small subunit data.

As seen for a significant number of proteins in Figures 4A and 5A, the percentage of modified amino groups is lower than the percentage of solvent exposed surface area in the ribosomal particle. If the average number of amidinated amino groups is compared with the number of amino groups visible on the exposed surfaces of the proteins (Figures 4B and 5B), most of these discrepancies disappear. The improved correlation is due to the nature of each approach. Solvent exposed surface area is calculated from the continuous surface generated by computationally rolling a sphere with the average radius of a water molecule over the macromolecule. As such, no allowance is made for the chemical identity of the surface groups, only the volume that they occupy. Visually inspecting the crystal structure and counting exposed amino groups accounts for the well characterized chemical reactivity of these groups with SMTA, and is far more straightforward than tabulating residue or atom surface areas.

For example, in Figure 4A it can be seen that 60% of L14's reactive amino groups react with SMTA, while 72% of the protein's total surface area is still exposed in the ribosomal particle. This amounts to only 5 of the protein's 8 reactive groups, as shown in Table 1. The simplest explanation for the lack of modification of three amino groups in this protein is that they are unreactive with SMTA due to interactions with rRNA features or other proteins. This hypothesis is supported by the close agreement between the average extent of modification, 4.9 ± 0.2 amidinated groups per protein, and the number of surface visible lysine residues, 5. Similarly, for protein S4, the percentage of modified groups, 66%, and the remaining solvent accessible surface area, 77%, differ by 11%. However, the average number of amidinated groups in this protein, 8.5 ± 0.3 , is identical within error limits to the number of surface visible amino groups, 9.

The conclusion that residues protected from reaction with SMTA are buried in quaternary structural interactions with rRNA is supported by observations made on the original crystal structures of isolated ribosomal subunits and whole ribosomes. Ribosomal crystal structures show the sequestration of significant amounts of protein surface area. The portions of protein structure corresponding to this surface area contain lysine-, arginine-, and glycine-rich "tails" that wind into the bulk of the rRNAs, allowing extensive interactions between the cationic side chains of K and R residues and the phosphodiester oxyanions of the rRNA backbone.^{55–57} These results are also consistent with several recent computational studies of protein-RNA complexes that show R, K, S, and N residues to be most frequently involved in direct protein-RNA interactions.^{58–60} In

particular, lysine residues are more frequently found in the interfacial region of protein-RNA complexes and contribute a relatively larger amount of sequestered surface area to the interface between protein and RNA than any amino acid other than arginine. Lysine residues are also 40% more likely to interact with a backbone phosphate oxyanion than would be predicted from a random distribution of lysine residues across the surface of an RNA binding protein.⁵⁹ Casual inspection of the ribosome crystal structure indicates that this is indeed a common motif. The causes of any remaining discrepancies in the extent of labeling versus visible lysine residues data are discussed below.

Large Subunit Labeling Results. Significant differences between the experimental and predicted extents of labeling are evident in Figure 4B, the most notable involving L5, L9, L11, and L16. Ribosomal protein L9 is a special case and is considered below. The remaining three proteins are less modified by SMTA than predicted by the count of visible lysine residues. Table 1 shows that each of these proteins is apparently methylated at one or more positions. Methylation is a post-translational modification that usually occurs on lysine amino groups or the amino termini of proteins. Figure 6A shows that the observed mass of protein L16 is 15.1 Da heavier than the 16094.9 Da mass predicted for this protein by the proteomic sequence. This modification presumably involves methylation of the protein's amino terminus, as observed in *E. coli* and *C. crescentus*.^{35,61} Figure 6A also shows that about 50% of the total L16 is subjected to a second methylation, based on the reasonable assumption that the addition of one methyl group does not alter the protein's ESI ion yield. Protein L16 has 17 modifiable groups. However the mass shifts observed in the disassembled, fully amidinated spectrum, Figure 6C, correspond to the addition of only 16 amidino groups. One potential site of modification is apparently blocked in each form of L16; if the first methylation is the canonical one at the N-terminus, it is this modification that blocks amidination. The effect of this methylation is also visible in the native amidinated data of Figure 6B. If the methyl groups were not present, the Native Extent of Labeling for L16, currently 7.8 ± 0.1 would be shifted up by 1, and the comparison of results in Figure 4B would match to within the error limits for the measurements. The effect of post-translational methylations on the extent of amidination is more pronounced for L5 and L11. L5 is proposed to be dimethylated, and the difference between the observed extent of modification, 14.6 ± 0.5 , and the number of visible lysine groups, 18, is 2–3 modifiable positions when the assumed error in lysine counting is considered. Protein L11 is the most heavily modified protein in the ribosomal proteome of every bacterial species studied to date.^{34,35,49,62} L11 multiple methylations may represent a distribution of methyl groups or may be acetylations at some positions. *D. radiodurans*' L11 is methylated ten times and shows a difference of four modifiable sites (10.00 ± 0.0 versus 14 visible lysines). If the *D. radiodurans* protein is modified by trimethylations at the N-terminus and positions K2 and K39 (the canonical positions for modifications in *E. coli*'s L11), the methylation of a fourth, undetermined lysine explains the observed labeling discrepancy.

Figure 4 indicates a significant difference between the extent of labeling of L9 and the protein's number of visible lysines: nine more residues are amidinated predicted from the crystal structure. However, the X-ray results only include residues 1–52 of this 146-residue polypeptide, and residues 53–146 contain 8 lysines, within error of the number missing from our count

Ribosomal Proteins of *Deinococcus radiodurans*

of visible lysines. The TtL9 entry on right-hand side of Figure 4 shows the improved agreement obtained by mapping the *D. radiodurans* L9 sequence onto the *T. thermophilus* L9 crystal structure. This L9 homologue is 50% identical and 70% homologous to the *D. radiodurans* protein and its crystal structure contains all but the last two C-terminal residues of the protein. We conclude that the difference in extent of labeling and the number of visible amino groups shown in Figure 4B results from disorder in *D. radiodurans*' crystal structure.³⁰ No other large subunit proteins show as dramatic evidence of crystallographic disorder as L9, although the slightly higher average extent of labeling of proteins L15 and L19 seen in Figure 4B is most likely due to disordered sequences containing 2 and 1 lysine residues, respectively.

The *D. radiodurans* 50S subunit crystal structure contains no electron density for the protein L28 and a protein partially buried in the bulk of the 23S rRNA identified as L31. In the crystal structure of *T. thermophilus*' 50S subunit, both L28 and L31 are identified. The L28 protein in *T. thermophilus* occupies a position homologous to the protein identified as L31 in the *D. radiodurans* crystal structure. This conflicting identification of a polypeptide represents another type of crystallographic uncertainty, and we attempted to use our amidination data to distinguish between the two assignments. In Figure 4, the entries labeled L28* and L31* were derived by mapping the *D. radiodurans* L28 sequence onto the polypeptide currently identified as L31 (Chain Y in structure 1NKW) or using the polypeptide's current identification as L31. The entry on the right-hand side of the figure, labeled as TtL28, is the result of mapping the *D. radiodurans* L28 sequence onto the *T. thermophilus* protein (Chain 1 in structure 2J01). The correlation of labeling with either solvent exposed surface area or the count of visible amino groups is better when the polypeptide in question in the *D. radiodurans* structure is identified as L28, consistent with the *T. thermophilus* assignment. The ease of distinguishing between the alternate identifications of the polypeptide in the *D. radiodurans* 50S structure by using amidination patterns suggests the use of native amidination as a straightforward means of experimentally checking assignments in large, multimolecular crystal structures.

Small Subunit Labeling Results. Compared to data for the large subunit proteins, the distribution of the extent of modification compared to SASA is more uniform for small subunit proteins, as seen in Figure 5A. However, as seen in Figure 5B, correlation of labeling with the visible lysine count is even better. The apparent overlabeling of S11, S18, S19, and S20 is again due to disorder in the crystal structure of the small subunit. The N-termini of S11, S18, and S20 are disordered, leading to protein regions that contain five, four, and three lysines that do not appear in the crystal structure. Similarly, the disorder of the C-terminus of S19 prevents five lysines in this protein's sequence from being counted.

Comparison with *C. crescentus* Results. In our earlier publication on the amidination of *C. crescentus* ribosomal proteins, the initial comparison made was between the percentage of unmodified amino groups and the solvent inaccessible surface area of each protein (Figure 5 of ref 28). This depiction of the data is complementary to those used in Figures 4A and 5A of the present publication, that show comparisons of the percentage of modified amino groups with the solvent accessible surface areas of the ribosomal proteins from the large and small subunits respectively. Overall, the extents of SMTA modification for homologous ribosomal proteins in *C. crescentus*

and *D. radiodurans* are very similar. Proteins with more than 70% of their total amino groups modified in each organism include L2, L24, and L33. Proteins buried in quaternary interactions with rRNA that show less than 50% modification include L20, L22, L23, and L35. This similarity in labeling patterns suggests that structural features common to prokaryotic ribosomal proteins are being detected. This proposition is further supported by the dramatic improvement of the correlation between experimental results and structure based prediction when the number of modified groups is compared to the number of visible amino groups. Comparing Figures 5 and 6 of ref 29 with Figures 4 and 5 of this paper shows that the effect is especially notable for proteins L13, L17, L20, L22, S4, and S9 in each organism.

A few proteins do not show similar labeling patterns when *D. radiodurans* and *C. crescentus* amidination data are compared. Large subunit proteins L2 and L4 show clear differences in the labeling patterns between the two organisms. For example, the *C. crescentus* proteins are more extensively labeled than their *D. radiodurans* homologues whose labeling was quite consistent with the *D. radiodurans* crystal structure. This difference in amidination reactivity can be explained by comparing the sequences and amino acid compositions of the homologous proteins from each organism. *C. crescentus* proteins have more lysine residues—2 extra in an accessible C-terminal tail for L2, and a remarkable 11 extra scattered throughout the protein sequence for L4. Extensive modification of the lysines of *C. crescentus*' L2 and L4 proteins suggests substantial solvent accessibility for these extra residues. Similarly for proteins S10 and S17, the *C. crescentus* proteins have 3 more lysine residues than the *E. coli* proteins used to evaluate the labeling results. One implication of this observation is that protected lysines involved in rRNA interaction surfaces are more likely to be conserved between species. Solvent exposed residues are more likely to vary and allow the evolution of species-specific interaction surfaces with other intracellular complexes, such as the signal recognition particle or folding chaperones.

These labeling differences show that sequence homology and phylogenetic relatedness should be considered when using crystal structures to interpret chemical labeling data. The discrepancies discussed above arise when comparisons are made between proteins from only superficially related organisms like the α -proteobacterium *C. crescentus* and members of either the γ -proteobacterial phylum (*E. coli*) or the *Deinococcus/Thermus* phylum (*D. radiodurans*).

Structural Correlations. Bottom-up proteomics experiments were used to corroborate whole protein identifications and to determine the location of amidinated residues in the native amidinated samples. In Figures 7–10 data from four proteins (L15, L24, S4, and S13) are presented to demonstrate the direct correlation between reactivity and solvent exposure. In these figures, each specifically referred to protein is displayed as a white space-filling model *in situ*, and in one or two perspectives with the rRNA hidden. Other proteins not specifically referred to are depicted as light blue space-filling models. The largest rRNA in the subunit (23S or 16S) is presented as a blue solvent accessible surface. Smaller rRNAs (5S or tRNA where applicable) are shown in purple. In the protein structures, red lysine residues have been observed with an added amidino group in LC-MS/MS analyses of enzymatic digests. Green lysine residues are unmodified by the amidination reagent. Because amidination of a lysine residue blocks tryptic cleavage, green

lysines are either directly observed as unmodified in LC–MS/MS analyses of enzymatic digests or are deduced to be free due to the appearance of a peptide in an LC–MS/MS analysis of enzymatic digests. Yellow lysine residues are predicted by the proteomic sequence in regions of the protein that are not observed in any enzymatic digest experiments.

Large Ribosomal Subunit Proteins L15 and L24. Ribosomal protein L15 accelerates large subunit assembly and has significant RNA chaperone activity, preventing RNAs from being trapped in dead-end folded structures.^{63,64} Table 1 and Figure 4B indicate that on average half of this protein's reactive positions are amidinated. As seen in Figure 7A, ribosomal protein L15 has a globular, solvent exposed C-terminal domain and an extended N-terminal "tail" that winds through the bulk of the 23S rRNA. The extreme end of this tail (which terminates at residue H4 in the protein sequence) is also solvent exposed. One peptide from the N-terminal tail containing an amidinated residue K7 is observed. A few dozen residues up the chain from K7, lysines such as K45 are protected by the bulk of the rRNA, as indicated by the observation of a peptide containing residues 46–52 shown in Figure 6B. Trypsin would be unable to bind to and hydrolyze the peptide backbone if residue K45 were amidinated. The globular C-terminal domain contains three of the seven visible lysines that it is possible to see if the crystal structure is viewed over a range of orientations, and these residues' reactivity with SMTA is demonstrated by the observation of a peptide containing residues 124–134 from a Glu C digest in Figure 7C. This peptide's mass and MS/MS spectrum are consistent with amidination at K131.

Ribosomal protein L24 is more compact than L15 as shown in Figure 8. L24 is a primary binding protein, interacting with the 23S rRNA close to its 5' end, and is one of the first ribosomal proteins to bind to the 23S rRNA during 50S subunit assembly.⁶³ More than half of this protein's available sites are modified by amidination in native ribosomes. Amidinated lysines have been observed in peptides containing residues 31–42 (K35, Figure 8B), and 106–115 (K112, Figure 8C) of this 115 amino acid protein. Unmodified residues (e.g., K56 and K80) are buried beneath the bulk of the protein, presumably engaged in intramolecular salt bridges or in interactions with the rRNA backbone.

Small Ribosomal Subunit Proteins S4 and S13. Figures 9 and 10 show the position of small subunit proteins S4 and S13 in the crystal structure of *T. thermophilus* HB8's 30S subunit. These figures were created by aligning the sequences of *D. radiodurans* S4 and S13 with their *T. thermophilus* homologues, then color coding the crystal structures according to the *D. radiodurans* sequences.

Ribosomal protein S4 binds directly to the 16S rRNA during the assembly of the 30S ribosomal subunit.^{65,66} S4 also interacts with proteins S3 and S5 to form the mRNA entry pore, and is part of an ATP-independent RNA helicase activity associated with translation. Protein S4 has a compact globular structure and binds to the 30S particle at a junction formed by five rRNA helices.^{32,67} As shown in Figure 9, amidinated residues are spread across the exposed outer surface of the protein and include K27, K137, K162 and K178. Unmodified residues cluster around the protein-rRNA interface, and include K24, K44, K127 and K128. It is interesting to note that two residues in close proximity in the structure with similar solvent exposure, K24 and K27, have opposite labeling reactivity. This suggests that the interaction of K24 with adjacent rRNA plays a critical role in blocking amidination. Protein–protein interactions can also

block the reaction of lysine residues with SMTA, as shown in Supplemental Figure 5. Here a salt bridge interaction in the interfacial region between S4 aspartate 49 and lysine 20 of adjacent protein S5 protects the lysine from reaction, as shown by the MS/MS spectrum of a peptide containing unmodified S5 K20.

Ribosomal protein S13 forms one of the intersubunit bridges between the 30S and 50S subunits.^{32,68} The structure of protein S13 bound to the small subunit is shown in two perspectives, related by a 180° rotation, in Figure 10A. No large subunit features have been included. The protein has a globular N-terminal domain that hangs over the edge of the small subunit, and an extended C-terminal tail containing 18 basic residues (9 arginines and 9 lysines). Overlapping peptides from Endoprotease Glu-C (Figure 10B) and tryptic (Figure 10C) digests show a labeled residue K31 in the C-terminal domain on the outer face of the 30S subunit. Observation of a peptide containing residues 63 to 71 (Figure 10D) that was not amidinated at residue K65 demonstrates that both K62 and K65 are unmodified, accounting for two of S13's three unmodified residues, as shown in Table 1 (compare 11.8 amino groups modified on average with a total of 15 groups in the molecule). The structural features comprising Bridge B1 provide a likely explanation for the observed protection from SMTA labeling. Although K62 and K65 appear to be exposed to solution in Figure 10A, S13's close association with both large subunit protein L5 and the 23S rRNA suggests that intersubunit interactions are influencing the reactivity of these lysines.

Utility of FT-ICR Mass Measurements. Previous studies of ribosomal proteins have commented on the difficulty of distinguishing between the trimethylation and acetylation of amino groups due to the similarity of the mass shifts.⁴⁹ Acetylation causes a monoisotopic mass increase of 42.01 Da, while trimethylation induces a 42.05 Da mass increase. SE-QUEST searches of tryptic digest LC–MS/MS data, with variable modifications on lysine residues and protein N-termini did not retrieve any acceptable peptide matches. However, in each of the cases where we have proposed trimethylations or acetylations of *D. radiodurans* ribosomal proteins (L11, L22, S2, S5,, and S18), the accurate whole protein masses collected with the FT-ICR instrument are most closely matched by calculated masses that assume the PTMs listed in Table 1. For example, the mass calculated for the eighth isotopomer peak of protein L22 with an acetylation, 15068.45 Da, is within 2 ppm (0.03 Da) of the experimental value while the mass predicted for trimethylated L22 is 15068.48, which is 5 ppm (0.07 Da) over the experimental value. Similarly, various combinations of acetylation and methylation in protein L11 would result in masses between 1 and 6 ppm (–0.02 to –0.1 Da) lighter than the experimental value, due to the lower mass increase upon addition of an acetyl group compared to that from the addition of three methyl groups. Although the FT-ICR determinations are astonishingly accurate, they do not allow unambiguous identification of the cause of the 28 Da mass increase postulated for protein L3 and L5. Post-translational dimethylation, retention of the initiator methionine's N-terminal formyl group, or a single nucleotide sequencing error in which a lysine in the proteomic sequence should be an arginine⁶⁹ all give mass increases that are indistinguishable even by FT-ICR measurements of whole protein masses. Recent *de novo* sequencing experiments in this laboratory indicate that an alteration at position 78 of protein L5 is the cause of the +28 mass increase. When previously collected LC–MS/MS data are searched

against an edited version of the *D. radiodurans* R1 proteome with an entry for protein L5 containing an arginine at position 78, a match for a peptide containing residues 72–80 is obtained, as shown in Supplemental Figure 6 (Supporting Information). Although this result allows us to localize the alteration to an internal site and rule out N-terminal formylation, it is still not possible to distinguish between the PTM and sequencing error. Protein L3's mass increase is likely to have a similar explanation.⁷⁰

Conclusions

In the present paper, we have improved upon our earlier study of *C. crescentus* ribosomes by interpreting our SMTA labeling data from the ribosomal large subunit that was the subject of the first crystal structure of a bacterial 50S subunit, *D. radiodurans*. This organism's close phylogenetic relationship to *T. thermophilus* HB8 also allowed us to use *T. thermophilus*' 30S subunit structure to evaluate the labeling of *D. radiodurans*' small subunit proteins. As a result, we have been able to correlate reactivity and structure for all of the proteins in this organism's ribosomal proteome that are visible in the crystal structure: no proteins have been omitted due to an inability to obtain homologous sequence alignments between proteins, as was the case for several *C. crescentus* ribosomal proteins.²⁸ The acidic stalk proteins L10 and L7/L12 are largely solvent exposed and unprotected even in fully native ribosomes (Table 1). Small subunit protein S1 is not expected to appear in ribosomes prepared using our method. The excellent correlation between reactivity and surface accessible amino groups shown in Figures 4B and 5B provides solid validation for the use of SMTA as a probe of tertiary and quaternary structure, as first reported in our amidination experiments with *C. crescentus* ribosomes. As illustrated using data for ribosomal proteins L15, L24, S4, and S13, surface accessible lysine residues are modified while lysines buried in quaternary structure interactions are protected from reaction with SMTA and are likely to be involved in key structural and functional interactions.^{26,27,58–60} Further research on this topic will be pursued using the ribosomes of *Thermus thermophilus* HB8, for which higher resolution structures are available.³² In Selmer et al.'s pair of 70S ribosome data, the authors were able to model the disposition of lysine side chains into the structures, and this level of detail would make the prediction of each residue's interaction partners more certain. We would expect to be able to reliably identify protein–protein interactions like the salt bridge shown in Supplemental Figure 4 (Supporting Information) as well, and to be able to correlate changes in SMTA reactivity with changes in solution parameters such as pH, ionic strength and ionic composition. The direct examination of the effect of amidination on subunit association would also be useful. Whole 70S ribosomes and dissociated subunits can be separated using either ultracentrifugation in a sucrose density gradient or hydrophobic interaction chromatography.^{24,71}

The major sources of difference between the observed and predicted extents of reactivity are crystallographic disorder and post-translational modifications. Crystallographic disorder makes portions of proteins effectively invisible in a crystal structure, and causes an underestimation of the predicted extent of modification. Likewise, the canonical PTMs for ribosomal proteins block amino groups, preventing modification by SMTA. Post-translational modifications give an unambiguous signature of apparent under-labeling when experimental and predicted extents of amidination are compared, as discussed

above for proteins L5, L11, and L16. The ability to confirm the presence of post-translational modifications using amidination may be a useful tool for understanding why these commonly observed modifications occur.

Comparison of unmodified and disassembled amidinated protein mixtures also shows great promise for the quality control of the proteomic sequences derived from genome sequencing projects. We have identified four sequencing errors in the ribosomal proteome of *D. radiodurans* R1. These enable the interpretation of intense, unidentified masses in our whole protein and disassembled amidinated protein separations.

Acknowledgment. We thank Dr. Michael J. Daly for the gift of *Deinococcus radiodurans* R1 cells, Dr. Ignatius J. Kass of Waters/Micromass for AutoME 1.1, Matthew A. Lauber for insightful discussions on the depiction of ribosomal protein crystal structures, and Dr. Neil Kelleher for advice on the proper reporting of high resolution FT-ICR MS data. This research was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lily Endowment, Inc., and by National Science Foundation Grants CHE-0518234 and CHE-0832651.

Supporting Information Available: Supporting Information includes flowchart for our protein identification strategy (Supplemental Scheme 1), sequence alignments and spectra supporting the proposed sequence errors in proteins L13, L21, and S5 (Supplemental Figures 2–4), and a figure showing the salt bridge between S4 D49 and S5 K20 with a supporting MS-MS spectrum (Supplemental Figure 5). Also included are chromatographic gradients used for the separations described in the text (Supplemental Tables 1–6). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Ray, P. S.; Arif, A.; Fox, P. L. *Trends Biochem. Sci.* **2007**, *32*, 158–164.
- (2) Downard, K. M. *Proteomics* **2006**, *6*, 5374–5384.
- (3) Sharon, M.; Robinson, C. V. *Annu. Rev. Biochem.* **2007**, *76*, 167–193.
- (4) Krause, R.; von Mering, C.; Bork, P.; Dandekar, T. *BioEssays* **2004**, *26*, 1333–1343.
- (5) Resing, K. A.; Ahn, N. G. *FEBS Lett.* **2005**, *579*, 885–889.
- (6) 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.
- (7) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III *Nat. Biotech.* **1999**, *17*, 676–682.
- (8) Protein NMR Techniques. In *Methods in Molecular Biology*, Vol. 60; Reed, D. G., Ed.; Humana Press: Totowa, NJ, 1997; p 5.
- (9) Drenth, J. *Principles of Protein X-ray Crystallography*; Springer-Verlag: New York, 1994.
- (10) Akashi, S.; Sirouzu, M.; Terada, T.; Ito, Y.; Yokoyama, S.; Takio, K. *Anal. Biochem.* **1997**, *248*, 15–25.
- (11) Englander, S. W.; Mayne, L.; Sosnick, T. R. *Protein Sci.* **1997**, *6*, 1101–1109.
- (12) Ehring, H. *Anal. Biochem.* **1999**, *267*, 252–259.
- (13) Smith, D. L.; Deng, Y.; Zhang, Z. *J. Mass. Spectrom.* **1997**, *32*, 135–146.
- (14) Glocker, M. O.; Borchers, C.; Fiedler, W.; Suckau, D.; Przybylski, M. *Bioconjugate Chem.* **1994**, *5*, 583–590.
- (15) Glocker, M. O.; Nock, S.; Sprinzl, M.; Przybylski, M. *Chem.–Eur. J.* **1998**, *4*, 707–715.
- (16) Fiedler, W.; Borchers, C.; Macht, M.; Deininger, S.-O.; Przybylski, M. *Bioconjugate Chem.* **1998**, *9*, 236–241.
- (17) Mendoza, V. L.; Vachet, R. M. *Anal. Chem.* **2008**, *80*, 2895–2904.
- (18) Guan, J.-Q.; Chance, M. R. *Trends Biochem. Sci.* **2005**, *30*, 583–592.
- (19) Sharp, J. S.; Becker, J. M.; Hettich, R. L. *Anal. Chem.* **2004**, *76*, 672–683.

- (20) Sharp, J. S.; Guo, J.-T.; Uchiki, T.; Xu, Y.; Dealwis, C.; Hettich, R. L. *Anal. Biochem.* **2005**, *340*, 201–212.
- (21) Tong, X.; Wren, J. C.; Konerman, L. *Anal. Chem.* **2007**, *79*, 6376–6382.
- (22) Neurath, H. In *Protein folding*; Jaenicke, R., Ed.; Elsevier, New York, 1980; pp 501–523.
- (23) Hubbard, S. J. *BBA* **1998**, *1382*, 191–206.
- (24) Suh, M.-J.; Pourshahian, S.; Limbach, P. A. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1304–1317.
- (25) Beardsley, R. L.; Reilly, J. P. *J. Proteome Res.* **2003**, *2*, 15–21.
- (26) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2005**, *77*, 7274–7281.
- (27) Liu, X.; Broshears, W. C.; Reilly, J. P. *Anal. Biochem.* **2007**, *367*, 13–19.
- (28) Beardsley, R. L.; Running, W. E.; Reilly, J. P. *J. Proteome Res.* **2006**, *5*, 2935–2946.
- (29) Muller, E.-C.; Wittmann-Liebold, B. *Cell. Mol. Life Sci.* **1997**, *53*, 34–50.
- (30) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679–688.
- (31) Harms, J. M.; Wilson, D. N.; Schluenzen, F.; Connell, S. R.; Stachelhaus, T.; Zaborowska, Z.; Spahn, C. M. T.; Fucini, P. *Mol. Cell* **2008**, *30*, 26–38.
- (32) Selmer, M.; Dunham, C. M.; Murphy, F. V. IV.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935–1942.
- (33) Woese, C. R. *Microbiol. Rev.* **1987**, *51*, 221–271.
- (34) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105–112.
- (35) Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P. *J. Proteome Res.* **2007**, *6*, 337–347.
- (36) Barritault, D.; Expert-Bezancon, A.; Guerin, M F.; Hayes, D. *Eur. J. Biochem.* **1976**, *63*, 131–135.
- (37) Karty, J. A.; Running, W. E.; Reilly, J. P. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* **2007**, *847*, 103–113.
- (38) Martz, E. *Trends Biochem. Sci.* **2002**, *27*, 107–109, <http://proteinexplorer.org>.
- (39) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (40) White, O.; Eisen, J. A.; Heidelberg, J. F.; et al. *Science* **1999**, *286*, 1571–1577.
- (41) Kleinjung, J.; Fraternali, F. *Nucleic Acids Res.* **2005**, *33*, W342–W346.
- (42) Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Auberry, D. L.; Battista, J. R.; Daly, M. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11049–11054.
- (43) Zhang, C.; Wei, J.; Zheng, Z.; Ying, N.; Shaeng, D.; Hua, Y. *Proteomics* **2005**, *5*, 138–143.
- (44) Schmid, A. K.; Lipton, M. S.; Mottaz, H.; Monroe, M. E.; Smith, R. D.; Lidstrom, M. E. *J. Proteome Res.* **2005**, *4*, 709–718.
- (45) Sherman, F.; Stewart, J. W.; Tsunasawa, S. *BioEssays* **1985**, *3*, 27–31.
- (46) Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinel, T. *Mol. Cell. Proteomics* **2005**, *5*, 2336–2349.
- (47) Karty, J. A.; Ireland, M. M. E.; Brun, Y. V.; Reilly, J. P. *J. Proteome Res.* **2002**, *1*, 325–335.
- (48) Zabrouskov, V.; Senko, M. W.; Du, Y.; Leduc, R. D.; Kelleher, N. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2027–2038.
- (49) Strader, M. B.; VerBerkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B. *J. Proteome Res.* **2004**, *3*, 965–978.
- (50) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
- (51) Miller, J. H. *Cell* **1974**, *1*, 73–76.
- (52) Makarova, K. S.; Omelchenko, M. V.; Gaidamakova, E. K.; Matrosova, V. Y.; Vasilenko, A. *PLoS ONE* **2007**, *2*, e955, doi:10.1371/journal.pone.0000955.
- (53) Luer, C. A.; Wong, K.-P. *Biochemistry* **1980**, *19*, 176–183.
- (54) Diedrich, G.; Burkhardt, N.; Nierhaus, K. H. *Protein Expression Purif.* **1997**, *10*, 42–50.
- (55) Moore, P. B.; Steitz, T. A. *Annu. Rev. Biochem.* **2003**, *72*, 813–850.
- (56) Klein, D. J.; Moore, P. B.; Steitz, T. A. *J. Mol. Biol.* **2004**, *340*, 141–177.
- (57) Brodersen, D. E.; Clemens, W. M., Jr.; Carter, A. P.; Wimberly, B. T.; Ramakrishnan, V. *J. Mol. Biol.* **2002**, *316*, 725–768.
- (58) Jones, S.; Daly, D. T. A.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. *Nucleic Acids Res.* **2001**, *29*, 943–954.
- (59) Treger, M.; Westhof, E. *J. Mol. Recognit.* **2001**, *14*, 199–214.
- (60) Ellis, J. J.; Broom, M.; Jones, S. *Proteins* **2007**, *66*, 903–911.
- (61) Alix, J.-H. *Adv. Exper. Biol. Med.* **1988**, *231*, 371–385.
- (62) Cameron, D. M.; Gregory, S. T.; Thompson, J.; Suh, M.-J.; Limbach, P. A.; Dahlberg, A. E. *J. Bacteriol.* **2004**, *186*, 5819–5825.
- (63) Herold, M.; Nierhaus, K. H. *J. Biol. Chem.* **1987**, *262*, 8826–8833.
- (64) Semrad, K.; Green, R.; Schroeder, R. *RNA* **2004**, *10*, 1855–1860.
- (65) Held, W. A.; Ballou, B.; Mizushima, S.; Nomura, M. *J. Biol. Chem.* **1974**, *249*, 3103–3111.
- (66) Culver, G. *Biopolymers* **2003**, *68*, 234–249.
- (67) Wimberly, B. T.; Brodersen, D. E.; Clemmons, W. M., Jr.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2001**, *407*, 327–339.
- (68) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila, S.; Holton, J. M.; Doudna, C. A. *Science* **2005**, *310*, 827–834.
- (69) Lauber, M. A.; Running, W. E.; Reilly, J. P. manuscript in preparation.
- (70) Zhang, L.; Running, W. E.; Alley, W. S.; Reilly, J. P. De novo Sequencing of Peptide Mixtures using LC-MALDI-Photodissociation. Proceedings of the 56th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, June 1–5, 2008; Abstract 101 (A082979.4500.VER.3.pdf).
- (71) Clemons, W. M., Jr.; Brodersen, D. E.; McCutcheon, J. P.; May, J. L. C.; Carter, A. P.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *J. Mol. Biol.* **2001**, *310*, 827–843.

PR800544Y