research articles Journal of proteome research

B. subtilis Ribosomal Proteins: Structural Homology and Post-Translational Modifications

Matthew A. Lauber, William E. Running, and James P. Reilly*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received December 24, 2008

Ribosomal proteins of the model gram-positive bacterium *B. subtilis* 168 were extensively characterized in a proteomic study. Mass spectra of the 52 proteins expected to be constitutive components of the 70S ribosome were recorded. Peptide MS/MS analysis with an average sequence coverage of 85% supported the identification of these proteins and facilitated the unambiguous assignment of post-translational modifications, including the methylation of S7, L11, and L16 and the N-terminal acetylation of S9. In addition, the high degree of structural homology between *B. subtilis* and other eubacterial ribosomal proteins was demonstrated through chemical labeling with *S*-methylthioacetimidate. One striking difference from previous characterizations of bacterial ribosomal proteins is that dozens of protein masses were found to be in error and not easily accounted for by post-translational modifications. This, in turn, led us to discover an inordinate number of sequencing errors in the reference genome of *B. subtilis* 168. We have found that these errors have been corrected in a recently revised version of the genome.

Keywords: *B. subtilis* • ribosomal proteins • sequencing errors • structural homology • post-translational modifications • *S*-methylthioacetimidate • two-dimensional liquid chromatography • protein mass spectrometry

Introduction

The bacterial ribosome is a macromolecular complex composed of approximately one-third protein and two-thirds RNA by mass.[1] As a result, there has long been interest in the roles of both proteins and RNAs with respect to the function of the ribosome. But given that the catalytic properties of the ribosome have recently,[2] and perhaps misleadingly,[3] been attributed to rRNA, the salient roles of ribosomal proteins have been overshadowed.[4] Nevertheless, protein biosynthesis does depend on a number of factors other than catalytic RNA, including ribosomal proteins. In addition to stabilizing ribosome structure, ribosomal proteins function to integrate translation with other cellular pathways. This interactome includes trigger factor coupling the emergence of nascent polypeptide with chaperone activity by docking on ribosomal protein L23,[5−7] and translation factors, EF-G and Ef-Tu, interacting with the ribosome through the stalk protein complex of L7/L12.[4,8] Antibiotic resistance, a topic of utmost relevance, likewise demonstrates the intricate relationship between ribosome function and ribosomal proteins. Select mutations in S4, S5, and S12, for example, confer resistance against the error-inducing antibiotic, streptomycin, by increasing translation accuracy.[4] S12 mutations are, in particular, localized near two conserved loops that extend into the acceptor site of the ribosome.[9,10] This correlates with the significance of S12 and its involvement in the translational decoding process.[4] Interestingly, S12 is also subject to a novel post-translational modification: β-methylthiolation of an aspartic acid residue residing within one of these conserved loops.[11] Although possible implications, such as involvement in translational decoding, are numerous, the purpose of this modification is not well understood.[12] It is known that the requisite aspartic acid is universally conserved among bacteria, archea, and eukaryotes,[12] but the modification itself is not.[11,13−19] The importance of characterizing the post-translational states, structures and dynamics of ribosomal proteins is therefore clearly significant, particularly for the bacterium *B. subtilis*, which has long served as the model gram-positive organism for a host of closely related pathogenic species, such as *B. anthracis* and *S. aureus*.[20]

B. subtilis is also interesting in its own right. Recently, ribosomal proteins L31 and S14, as parts of the ribosome in *B. subtilis,* have been implicated in zinc homeostasis and stress response.[21,22] These two proteins have paralogs that, in contrast to their original counterparts, are regulated by the zinc repressor Zur and do not contain the CXXC motifs required to coordinate zinc. The role of these paralogs has just begun to be elucidated. Under zinc-deficient conditions, Zur derepresses the paralogous genes to allow expression of the paralogous proteins. Subsequently, the paralog of L31, YtiA, actively displaces its original zinc-binding counterpart, and the paralog of S14, YhzA, is preferentially incorporated into the ribosome during *de novo* assembly of the complex.[23−25] These mechanisms have been postulated to account for a modulated source of zinc accessible to the cell and a fail-safe mechanism for the continued synthesis of ribosomes during zinc-deficient conditions. Notably, ribosomal protein L33 is also expected to respond to alterations in zinc homeostasis, as L33 paralogs,

* To whom correspondence should be addressed. James P. Reilly, e-mail: reilly@indiana.edu.

10.1021/pr801114k CCC: $40.75

Published on Web 08/05/2009

similar to those of L31 and S14, have been identified in the *B. subtilis* genome.[21,22]

Furthermore, although *B. subtilis* has been extensively studied to the point of becoming a noteworthy paradigm, a crystal structure of its 70S ribosome has not yet been published. Information about its ribosomal proteins is therefore needed for demonstrating the degree of structural homology with other organisms, such as *T. thermophilus* and *E. coli*, whose 70S ribosome crystal structures have been derived.[26,27] An effective scheme for acquiring this type of information involves probing protein structure with chemical labeling.[28] The use of SMTA to amidinate primary amino groups in proteins has been thoroughly demonstrated, and the correlation between crystal structure solvent accessibility of targeted sites and their modification has been found to be excellent.[29,30] Moreover, Running and Reilly previously amidinated the primary amino groups of ribosomal proteins to effectively probe the structure of the *D. radiodurans* ribosome and also observed excellent agreement between crystal structure solvent accessibility of targeted sites and SMTA labeling.[19]

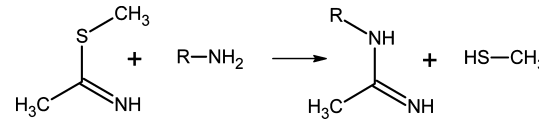
In this work, we present the results of a proteomic study of *B. subtilis* ribosomal proteins as facilitated by two-dimensional chromatography with online fractionation.[31] Our results catalog ribosomal proteins expressed under normal growth conditions, extend structural homology from sequence homology using amidination as a structural probe, and document post-translational modifications. Additionally, this work demonstrates complications in protein and peptide identifications that arise from the presence of sequencing errors in a reference genome. The reference genome of *B. subtilis 168* was one of the first to be sequenced and is known to contain an inordinate number of sequencing errors.[32] Here, phylogenetic and mass spectrometric techniques were employed to identify an alarming number of sequencing errors in the *B. subtilis* 168 reference genome. We have, in addition, found that these errors are corrected in a recently revised version of the genome.

Materials and Methods

Chemicals and Solvents. Acetonitrile, glacial acetic acid, trifluoroacetic acid, and urea were purchased from EM Science (Gibbstown, NJ). Water was purified by a Barnstead/ Thermolyne E-pure water filtration system (Barnstead/Thermolyne, Dubuqe, IA). Ammonium bicarbonate, anhydrous diethyl ether, and thioacetamide were obtained from Fisher (Fair Lawn, NJ). Proteomics grade trypsin (T-7575), magnesium acetate tetrahydrate, Trizma base, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and phenylmethylsulfonyl fluoride (PMSF) were purchased from Sigma (St. Louis, MO). Complete ethylene diamine tetraacetic acid (EDTA)-free Protease Inhibitor ("Mini") Tablets were supplied by Roche Applied Science (Indianapolis, IN). Ammonium chloride, formic acid, 2-mercaptoethanol, and 40% aqueous methylamine were obtained from Aldrich (Milwaukee, WI). Sodium chloride was purchased from Mallinckrodt Baker (Phillipsburg, NJ).

Bacterial Strains and Growth Conditions. *Bacillus subtilis* 168 (RL1, Laboratory strain 168 *trpC2,* Oxford University)[33] cells were grown for the preparation of 70S ribosomes. Precultures, derived from single colonies, were grown on Luria−Bertani medium to an OD_{600} of 4.0. A 130 μL volume of this preculture was used to inoculate 500 mL of Luria−Bertani medium (5 g NaCl, 5 g bactotryptone, and 2.5 g yeast extract in 500 mL of water) (BD Biosciences, San Jose, CA). The resulting cultures were grown for 10 h at 30 °C while being aerated at 180 rpm. Cells isolated were representative of late log phase growth.

Scheme 1. Amidination of Primary Amino Groups with *S*-Methylthioacetimidate (SMTA)



Ribosome Preparation. 70S ribosomes were isolated from cells as previously described by Arnold and Reilly.[18,34] Buffers A, B, and E were prepared with HEPES in place of Tris, ribosomes were salt-washed only once, and the final pellet resuspension was immediately aliquoted into 100 uL portions for storage at −80 °C.[35,36] A typical preparation yielded a ribosome aliquot with an A_{260}/A_{280} value of 0.92 and a total protein concentration of 2.05 mg/mL (Bradford Assay).[37] Upon acetic acid precipitation of RNA, the concentration of ribosomal proteins was determined by a Bradford assay to be 0.75 mg/mL.

Amidination of Ribosomal Proteins. Ribosomal proteins were amidinated both before and after disassembly of the ribosome. Primary amino groups were reacted with *S*-methylthioacetimidate (Scheme 1). The method for amidination used here was identical to that outlined in detail by Beardsley and Reilly, except that the amidination of denatured proteins was terminated by the addition of acetic acid not trifluoroacetic acid (TFA).[38]

Coupled 2D-LC−ESI−MS. Protein separations outlined in this work utilized a 60 column chromatography apparatus that has been extensively described.[31] For each sample, 75−100 μg of total protein was loaded on a SCX column (Toso-Hass SP-NPR, 4.6 × 35 mm, Tosoh Bioscience, Montgomeryville, PA), separated and directed onto an array of 20 C4 reversed-phase "trap" columns (Thermo Hypersil-Keystone Javelin, 1.0 × 20 mm, Bellefonte, PA). The contents of these traps were eluted onto and separated across a subsequent C4 reversed-phase analytical column (Thermo Hypersil-Keystone Pioneer, 1 × 100 mm, Bellefonte, PA). Proteins eluting from the analytical column were then directed to the ESI source of a mass spectrometer. In some separations, the "trap" contents were eluted with organic mobile phase and directly collected as fractions for tryptic digestion. Mobile phases and gradients used in these separations can be found in Supplemental Tables 1 and 2 (Supporting Information).

Whole Protein Mass Spectrometry. Whole protein mass spectra were acquired using a quadrupole time-of-flight mass spectrometer (Q-TOF, Waters, Manchester, UK) outfitted with an ESI source. Reversed-phase effluent from the coupled 2D-LC was split from a flow rate of 50 μL/min to 7 μL/min before being infused into the ESI source. In all experiments, the voltage applied to the ESI needle was +3.0 kV. Mass spectra (600−1800 m/z) were acquired over the time interval corresponding to 15 and 50 min of the second dimension LC gradient. Whole protein masses were obtained using Bioanalyte ProTrawler/Regatta, a software package for automating spectral extraction, simplification and deconvolution (Portland, ME). When analyzed manually, raw spectra corresponding to chromatographic peaks were summed together and deconvoluted using MassLynx and MaxEnt 1. The expected mass accuracy

and resolving power for these whole protein MS measurements was 2 Da and 5000, respectively.

Proteolytic Digestion of Trapping Column and Second Dimension Fractions. Proteins collected from eluting the contents of the C4 trapping columns were taken to dryness and resuspended in 25 μL of 10 mM ammonium bicarbonate buffer, containing either 0.2 μg of proteomics grade trypsin (Sigma, St. Louis, MO) or endoproteinase Glu-C (New England Biolabs, Ipswich, MA). Peptides were generated by enzymatic digestion for 13 h at 37 °C.

Capillary LC−ESI−MS/MS. Peptide analysis was performed via capillary LC−MS/MS using C18 reversed-phase trapping (100 μm × 10 cm, Magic 200 Å C18, Michrom Bioresources, Auburn, CA) and capillary LC columns (75 μm × 15 cm, Magic 100 Å C18, Michrom Bioresources, Auburn, CA), a Thermo LTQ-FT Ultra mass spectrometer and a Dionex chromatography system (Ultimate 3000, Dionex, Sunnyvale, CA). In each experiment, approximately 1 μg of a single trap protein digest was injected onto a trapping column to remove salts and contaminants by flushing for 20 min with 95% mobile phase A (0.1% formic acid in 97% water, 3% acetonitrile) and 5% mobile phase B (0.1% formic acid in acetonitrile) at a flow rate of 5 μL/min. Effluent from the trapping column was then directed onto the capillary LC column. The flow rate was reduced to 0.25 μL/min and an 80-min gradient between mobile phases A and B was implemented. Eluting peptides were electrosprayed into a Thermo LTQ-FT Ultra mass spectrometer operating in data-dependent mode to acquire a full FT-MS scan (400−2000 m/z) and subsequent CID MS/MS scans of the five most intense precursor ions in the LTQ. The resolution for the survey FTMS scan was set to 50000. Other settings included a 60 s exclusion duration time and collision energy of 35%. The expected mass accuracy for the peptide MS measurements was 1−4 ppm. MS and MS/MS spectra were subjected to data reduction using Mascot Distiller or ProtQuant Suite TurboRAW2Mgf (National Center for Glycomics and Glycoproteomics, Department of Chemistry, Indiana University, Bloomington, IN). The resultant Mascot Generic Format files were searched against the *B. subtilis* proteome using Mascot. Numerous variable modifications were included in these searches, such as methionine oxidation, N-terminal acetylation, N-terminal methylation, N-terminal formylation, lysine methylation, aspartic acid methylthiolation, and N-terminal/lysine acetamidination (for SMTA modified samples). Mascot error tolerant searches were also employed to facilitate detection of unexpected modifications and sequencing errors. MS/MS ion matches with scores less than the probability-based Mowse score for statistical significance (IS 25) were ignored. Mascot results from multiple experiments were combined using ProtQuant Suite ProtParser (National Center for Glycomics and Glycoproteomics, Department of Chemistry, Indiana University, Bloomington, IN).

Protein Sequence Data. The *B. subtilis* strain 168 proteome was obtained from The Institute for Genome Research (cmr.tigr.org, GenBank Accession.Version AL009126.1). Theoretical protein masses were calculated using an in-house computer program *Prodigies* or PAWS (freeware edition, ProteoMetrics, LLC). Additional sequence information was obtained from Swiss-Prot via ExPASy (www.expasy.org).

Phylogenetic Analysis. Phylogenetic analyses were completed via multiple sequence alignments with ClustalW v2.0.[39,40] Default settings were maintained for alignments.

Interpretation of Labeling and Crystal Structures of the Ribosome. *E. coli* and *T. thermophilus* served as models for investigating the structural homology of *B. subtilis* ribosomal proteins. ClustalW sequence alignments made of *B. subtilis* with *E. coli* and *T. thermophilus* were used to map modifiable sites to the sequences of ribosomal proteins found in published crystal structures (*E. coli* K12, PDB: 2AVY and 2AW4; *T. thermophilus* HB8, PDB: 2J00 and 2J01).[26,27] Aligned residues in the sequence of the crystal structure were interpreted with respect to solvent accessibility as a means to indicate the accessibility and count the number of modifiable sites (N-terminus and lysines) in a given *B. subtilis* ribosomal protein. PyMOL v. 0.99 (DeLano Scientific, www.pymol.org) was employed for the visualization and manipulation of crystal structures.[41] All interpretations of accessibility were derived from crystal structures displayed with Connolly surfaces.[42] Modifiable sites that aligned to absent crystal structure sequences were assumed to be accessible by argument of disorder and/or proteolytic cleavage. The inherent uncertainty associated with this manual interpretation was assumed to be plus or minus one modification. Additional uncertainty was introduced by gapped sequence alignments, in which case the accessibility of a modifiable site "aligned" with a gap was assumed to be unknown. Because of this, the positive uncertainty was increased by one, to represent the fact that such a modifiable site was not counted but could in fact be accessible.

Results and Discussion

Identification of Proteins. Assessing the protein components of the gram-positive *B. subtilis* ribosome requires different consideration from that previously given to the study of other bacterial ribosomal proteomes. For example, although S1 is a component of the *E. coli* ribosome, the protein is not believed to have a counterpart in *B. subtilis*, where the role of S1 is fulfilled through stronger interactions between the Shine-Dalgarno sequences of mRNA and the ribosome.[43,44] Likewise, in *B. subtilis*, ribosomal protein L25 is not a constitutive component of the ribosome under normal growth conditions. The homologue of *E. coli* L25 is instead general stress protein Ctc, which binds to the ribosome only during stress conditions, such as elevated temperature.[45−47] A total of 52 proteins were consequently expected to be present in this preparation of *B. subtilis* ribosomal proteins, and all were observed. In these experiments, ribosomal proteins from *B. subtilis* were separated by strong cation exchange (SCX) and fractions of the effluent were collected on 20 reversed-phase C4 "trapping" columns to provide material for subsequent LC−ESI−MS experiments. A typical result is outlined in Figure 1. In this example, three ribosomal proteins were collected on trap 17, as is evident in the total ion chromatogram of Figure 1A. The mass spectrum recorded during the elution of the third major peak before and after deconvolution appears in Figure 1B and C, respectively. It should be pointed out that some ribosomal proteins were isolated largely in a single trap while others were isolated across several. For this reason and because there may be significant differences in protein ionization efficiencies, the relative abundances of the ribosomal proteins cannot be reliably estimated from these measurements. In fact, deconvoluted peak intensities of different ribosomal proteins varied up to 30-fold, even though their abundances are expected to be similar.

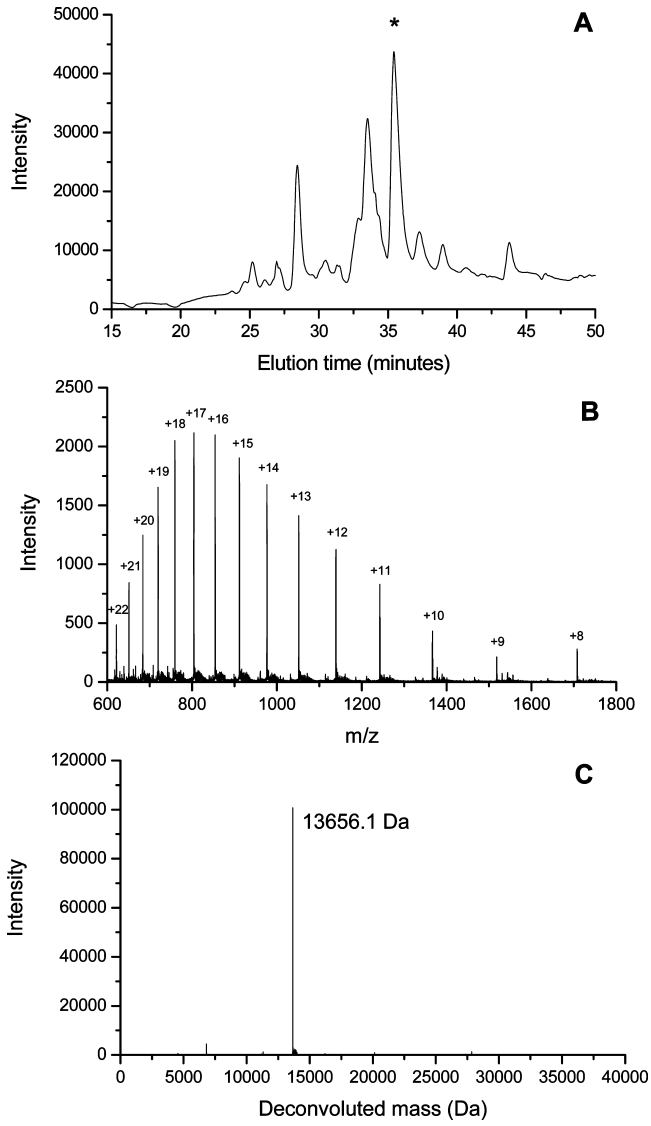


Figure 1. (A) Total ion chromatogram of trap 17. (B) Raw whole protein mass spectrum with charge states for the ribosomal protein represented by the third major chromatographic peak. (C) Deconvoluted whole protein mass spectrum.

Due to the possibilities of sequencing errors and post-translational modifications, the measurement of whole protein masses, such as 13656.1 Da seen in Figure 1C, does not directly translate to definitive protein assignments. It was therefore imperative to reproduce the separation of ribosomal proteins, proteolytically digest the contents of each trapping column, conduct peptide MS/MS analyses, and thereby complement whole protein masses with peptide analyses. In doing so, the identification of a whole protein mass was effectively constrained to just a few proteins, those detected by bottom-up analysis of a certain trap. This was most crucial for situations in which an observed whole protein mass did not match the theoretical mass of any ribosomal protein. In total, tryptic and Glu-C peptides from 51 of the 52 ribosomal proteins were detected and provided an impressive average sequence coverage of 85%.

The process for proposing protein identifications was involved, but again depended on peptide MS/MS analysis and whole protein MS of trapping column contents. Of the 52 noted ribosomal proteins, whole protein masses of only 13 matched their associated theoretical masses to within the expected mass accuracy of our Micro QTOF mass spectrometer (2 Da). However, simply accounting for expected N-terminal methionine cleavage by methionine aminopeptidase led to mass matches for 21 additional proteins.[48] Another 7 identifications were facilitated by reviewing entries for the ribosomal proteins of *B. subtilis,* as found on ExPASy of the Swiss Institute of Bioinformatics, and resolving conflicts between the published reference genome of *B. subtilis* 168[32] and other reports of sequencing information.[49−51] Masses for the remaining proteins were tentatively assigned assuming that they would not differ from theoretical masses by more than a few hundred Da, which could reasonably correspond to post-translational modifications or sequencing errors. This approach can fail if a start codon in the reference genome is misassigned, which can lead to a much greater difference between observed and theoretical masses. In practice, only one protein was assigned a mass that differed from genome prediction by more than several hundred Da. The genome sequence corresponding to this protein appears to have a misassigned start codon (see below).

Observed whole protein masses, trap locations, and peptide MS/MS sequence coverage values corresponding to protein identifications are shown in Table 1. Theoretical masses, as calculated from the proteome, and the mass differences between observed and theoretical masses are likewise provided.

Genome Sequencing Errors. Proteins with mass differences greater than our expected whole protein MS mass accuracy or those whose interpretation required noncanonical post-translational modifications were immediately suspected of sequencing errors. The occurrence of sequencing errors in the ribosomal proteome of *B. subtilis* 168 has been previously acknowledged. While investigating the sequence homology of *T. thermophilus* S12, Dahlberg and co-workers noticed that the S12 protein sequence derived from the published *B. subtilis* 168 reference genome conspicuously lacked the Asp residue that is methylthiolated in several organisms.[11,16,17,19,32] Sequencing of the gene unsurprisingly resulted in corrections to the reference genome (Genbank accession numbers DQ284750, DQ284751, DQ284752, and DQ284753).

Suspecting that some of the anomalous mass differences observed for our provisional protein identifications could be explained by sequencing errors, we conducted phylogenetic analyses via ClustalW multiple sequence alignments.[39] The approach here was to capitalize on the high sequence homology of ribosomal proteins and delineate conspicuous points of difference between the sequences of *B. subtilis* strain 168 and those of closely related species. For instance, the whole protein mass observed for L23 was 26.9 Da too heavy. Based on multiple sequence alignment (Figure 2), the sequence of *B. subtilis* L23 derived from the reference genome was found to contain two conspicuous residues (positions E24 and A39) that differed from the residues conserved at those positions in closely related species. Moreover, if these two residues were assumed to be incorrectly assigned and replaced by their normally conserved counterparts, the predicted and observed masses of L23 would agree. Two sequencing errors, E24K and A39 V, were therefore hypothesized.

Many of the mass differences observed for the remaining 11 proteins were readily explained by invoking 15 substitutions rationalized by genome sequencing errors. The theoretical masses of these proteins were accordingly corrected and are shown in Table 1. Notably, a significant number of the proposed sequencing errors have been confirmed by peptide

Table 1. Whole Protein and Digest Data for *B. subtilis* Ribosomal Proteins

protein	theo. mass	obs. mass	apparent mass diff.	corrected theo. mass[a]	corrected mass diff.	mass error	modifications	sequencing errors	whole protein location[c]	tryptic peptide MS/MS % sequence	glu-C peptide MS/MS % sequence	net MS/MS % sequence
50 S Ribosomal Proteins												
L1	24922.7	24857.7	−65.0	24988.8	−131.1	0.1	-Met	H17R[b], S53R[b], H56D[b]	Trap 8	100	69	100
L2	30331.9	30143.8	−188.1	30273.9	−130.1	1.1	-Met	S9T, [E256G[b], Q262K]	Trap 18	84	52	86
L3	22683.2	22552.4			−130.8	0.4	-Met		Trap 12	91	56	98
L4	22390.9	22260.4			−130.5	0.7	-Met		Trap 14	82	53	82
L5	20147.5	20147.6			0.1	0.1			Trap 9	93	83	100
L6	19509.3	19377.9			−131.4	−0.2	-Met		Trap 9	99	92	99
L9	16351.9	16351.7			−0.2	−0.2			Trap 9	26	28	41
L10	18028.7	17947.5	−81.2	18078.7	−131.2	0.0	-Met	S52F[b], P131S[b], K144Q[b]	Trap 6	99	72	99
L11	14917.4	14926.4	9.0	14931.4	−5.0	0.1	-Met, +9 CH_2	V35I[b]	Trap 7	95	56	95
L12	12750.6	12619.5			−131.1	0.1	-Met		Trap 5	100	95	100
L13	16291.9	16374.9	83.0	16375.0	−0.1	−0.1		C16W[b]	Trap 11	100	81	100
L14	13154.3	13153.5			−0.8	−0.8			Trap 7	46	22	48
L15	15382.6	15382.3			−0.3	−0.3			Trap 14	100	65	100
L16	16190.0	16218.2			28.2	0.2	+2 CH_2		Trap 18	78	62	87
L17	13750.8	13619.8			−131.0	0.2	-Met		Trap 10	95	45	99
L18	13017.8	12968.7	−49.1	12968.8	−0.1	−0.1		[Y37N[b]]	Trap 13	84	46	84
L19	13747.2	13386.9	−360.3	13386.8	0.1	0.1		[-MKT[b]]	Trap 15	75	59	84
L20	13638.1	13507.2			−130.9	0.3	-Met		Trap 18	75	25	75
L21	11275.0	11275.6			0.6	0.6			Trap 14	88	62	88
L22	12459.6	12459.4			−0.2	−0.2			Trap 10	82	69	97
L23	10928.6	10955.5	26.9	10955.7	−0.2	−0.2		E24K[b], A39V[b]	Trap 10	94	22	94
L24	11142.1	11141.6			−0.5	−0.5			Trap 11	96	100	100
L25	22055.5									21	15	36
L27	10371.8	9207.2	−1164.6	9207.4	−0.2	−0.2		-MLRLDLQFF	Trap 10	78	51	78
L28	6809.1	6677.5			−131.6	−0.4	-Met		Trap 15	58		58
L29	7713.0	7712.7			−0.3	−0.3			Trap 9	100	58	100
L30	6637.8	6506.3			−131.5	−0.3	-Met		Trap 7	100	92	100
L31	7443.5	7442.5			−1.0	−1.0			Trap 8	79	17	82
L32	6729.0	6596.0			−133.0	−1.8	-Met		Trap 15	68	41	71
L33.1	5987.0	5900.0	−87.0	5900.8	−0.8	−0.8		R13G[b], T22N	Trap 11	82		82
L34	5253.2	5253.0			−0.2	−0.2			Trap 19	43		43
L35	7557.0	7425.7			−131.3	−0.1	-Met		Trap 19	61		61
L36	4305.4	4304.9			−0.5	−0.5			Trap 13			
30S Ribosomal Proteins												
S2	27967.2	27836.6			−130.6	0.6	-Met		Trap 11	77	70	91
S3	24332.8	24191.6	−141.2	24322.8	−131.2	0.0	-Met	[P164S][b]	Trap 10	93	59	96
S4	22835.2	22704.3			−130.9	0.3	-Met		Trap 14	97	51	97
S5	17622.6	17622.5			−0.1	−0.1			Trap 9	93	65	93
S6	11124.5	11124.4			−0.1	−0.1			Trap 5	100	95	100
S7	17892.8	17765.9	−126.9	17882.8	−116.9	0.3	-Met, +1 CH_2	P39S[b]	Trap 11	99	35	99
S8	14843.3	14746.3	−97.0	14877.4	−131.1	0.1	-Met	[P37A[b], S63F[b]]	Trap 9	100	62	100
S9	14308.4	14177.5			−130.9	0.3	-Met, (also +C_2 H_2 O)		Trap 13	85	42	89
S10	11665.6	11534.6			−131.0	0.2	-Met		Trap 11	91	83	91
S11	13924.9	13795.1			−129.8	1.4	-Met		Trap 14	76	57	76
S12	15323.8	15084.4	−239.4	15215.6	−131.2	0.0	-Met	P128S[b], [N102D[b], R105G[b]]	Trap 15	91	36	93
S13	13801.0	13656.1	−144.9	13787.0	−130.9	0.3	-Met	[T85S[b]]	Trap 17	74	33	74
S14	7245.7	7113.2			−132.5	−1.3	-Met		Trap 16	34	56	62
S15	10573.1	10442.0			−131.1	0.1	-Met		Trap 15	89	17	89
S16	10134.8	10003.4			−131.4	−0.2	-Met		Trap 10	90	61	90
S17	10198.9	10067.5			−131.4	−0.2	-Met		Trap 13	79	63	93
S18	9201.8	8838.2	−363.6	8969.5	−131.3	−0.1	-Met	[-MT]	Trap 16	35	17	43
S19	10583.2	10451.9			−131.3	−0.1	-Met		Trap 11	95	37	95
S20	9500.0	9467.9	−32.1	9599.1	−131.2	0.0	-Met	[G18R[b]]	Trap 13	92	59	91
S21	6830.0	6698.6			−131.4	−0.2	-Met		Trap 16	65		65
Ribosome-Associated Protein												
YvyD	21979.6	21979.7			−0.1	−0.1			Trap 8	77	53	79

[a] Theoretical masses calculated from protein sequences with corrected sequencing errors. [b] Sequencing error confirmed by MS/MS of peptides. [c] Trap where the protein exhibited its highest intensity in whole protein spectra [ ] Sequencing error derived from a conflict between the published reference genome and other sequencing information.[17,32,49−51]

MS/MS. Such was the case of an aforementioned L23 sequencing error, in which a tryptic peptide of the protein was confirmed by MS/MS analysis to contain a Val rather than an Ala residue at position 39 (Figure 3). Detection of the b_5, b_6, y_{11}, and y_{12} ions unambiguously identified residue 39 as valine. Another notable case was that of a Lys to Gln sequencing error

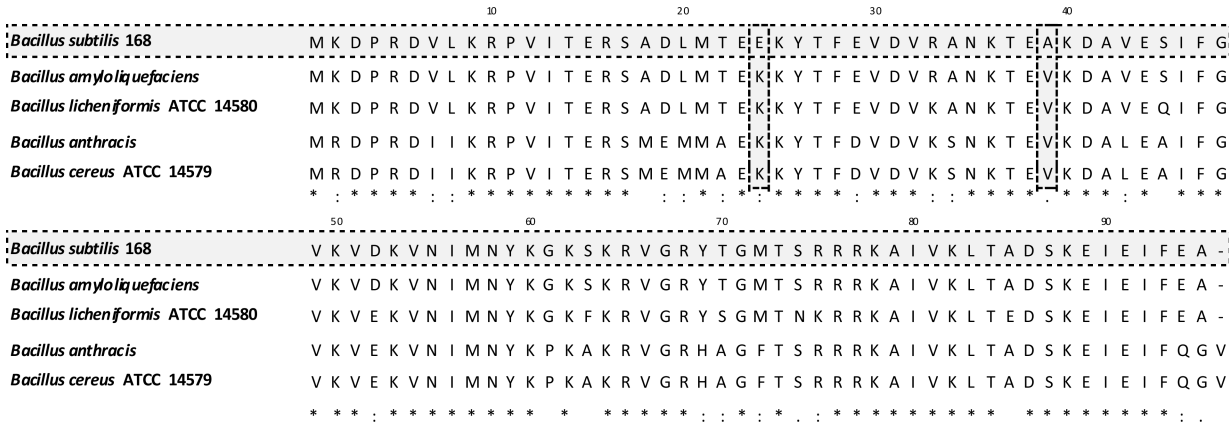


Figure 2. Multiple sequence alignment of ribosomal protein L23: *B. subtilis* and closely related species. Suspected sequencing errors at residues 24 and 39 are highlighted. An asterisk denotes identity, two dots denote strong similarity, one dot denotes weak similarity, and no marker denotes a significant difference among the aligned residues.

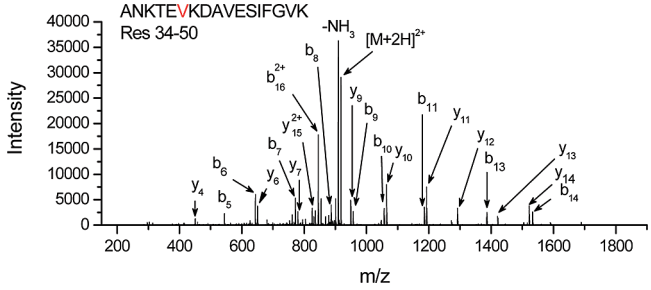


Figure 3. LTQ-FT MS/MS spectrum of an L23 peptide composed of residues 34−50 that confirms the A39 V sequencing error. Detection of the b_5, b_6, y_{11}, and y_{12} ions unambiguously confirms the identity of residue 39 as valine.

in L10 that was readily detected by high resolution FT-MS. The presence of a Gln in place of Lys typically is often not detected by mass spectrometry because of the small mass difference between the two residues (0.04 Da) and the resultant need for highly accurate mass determinations. However, the mass accuracy of the employed LTQ-FT (*ca.* 1−4 ppm or 0.001−0.004 Da for a typical peptide) easily enables these residues to be distinguished. Specifically, the masses determined for L10 peptides containing the noted sequencing error were 0.04 Da lighter than the theoretical masses of the sequences derived from the reference genome (Table 2). Altogether, 13 out of the 15 proposed sequencing errors have been confirmed by LTQ-FT based peptide mass measurements and MS/MS. The peptide MS/MS matches from which these errors were confirmed are shown in Supplemental Table 3 (Supporting Information). Those from which conflicts between the reference genome and other sequencing information were resolved are shown in Supplemental Table 4 (Supporting Information).

This method for proposing sequencing errors based on phylogenetic analysis was invariably proficient with one exception. Ribosomal protein S7 appeared by multiple sequence alignment to have two sequencing errors, S24T and P39S, that could explain an apparent mass error of 4.3 Da (Supplemental Figure 2, Supporting Information). Remarkably, based on peptide MS/MS, only one of these suspected sequencing errors (P39S) was confirmed (Figure 4A), while the other (S24T) was in fact discounted (Figure 4B). In place of the discounted sequencing error, a +14 Da mass differential was detected elsewhere in the protein sequence. Peptides containing lysine 149 were consistently observed to be 14 Da too heavy; and more importantly, fragmentation of one of these peptides resulted in a series of y-ions isolating the mass differential to lysine 149 and indicating an unexpected monomethylation at this residue (Figure 4C).

One other type of sequencing error was also identified by this combined phylogenetic/mass spectrometric analysis. The mass of ribosomal protein L27 was found to be 1164.6 Da too light. Meanwhile, multiple sequence alignment revealed the presence of a conserved start site further into the sequence of the protein than that derived from annotation of the reference sequence. Numerous ribosomal protein L27 sequences of closely related species originate at what had been assigned as residue 9 of *B. subtilis* L27. This suggested that the N-terminal residue of L27 should be residue 10, after considering methionine removal. Correcting the amino acid sequence in this way provided a theoretical mass consistent with the observed L27 mass to within 0.2 Da.

On the basis of the sequencing errors exclusively identified here and those previously implicated by alternative sequencing information,[17,49−51] the ribosomal protein genes in the *B. subtilis* 168 reference genome contain at least 25 nucleotide misassignments (the number of substitutions shown in Table 1). Using the ribosomal proteome as a predictor of genome quality then, we estimate the reference genome of *B. subtilis* 168 has an error frequency of 1 part per thousand. This is alarming considering that large-scale genome projects typically have error frequencies of less than one in every ten thousand bases, a value 10 times less than that of the *B. subtilis* reference genome.[52]

It is important to note that all 25 sequencing errors identified here have been corrected in an updated GenBank entry for the *B. subtilis* 168 reference genome (AL009126.3). This was concomitantly made available during the final revision of our manuscript. Although this revision of the *B. subtilis* 168 reference genome was done independently of the present study, it is clear that pragmatic use of proteomic data can lead to improved genomic data. Our work should serve as a prime reference for researchers in the burgeoning field of proteogenomics who are concerned with these issues.

Post-Translational Modifications. In addition to the previously mentioned monomethylation of S7, post-translational modifications were assigned to describe the predominant forms of two other proteins, L11 and L16. After accounting for N-terminal methionine removal and one suspected sequencing

Table 2. FT-MS of L10 Peptides Containing the K144Q Sequencing Error[a]

observed *m*/*z*	Mr(expt)	Mr(calc)	mass error	peptide	residues
863.5004	1724.9862	1725.0222	−0.0360	R.EGLLSMLLSVL**K**APVR.N	133−148
863.5004	1724.9862	1724.9859	0.0004	R.EGLLSMLLSVL**Q**APVR.N	
871.4966	1740.9786	1741.0172	−0.0386	R.EGLLSM^{ox}LLSVL**K**APVR.N	133−148
871.4966	1740.9786	1740.9808	−0.0022	R.EGLLSM^{ox}LLSVL**Q**APVR.N	
888.8538	2663.5395	2663.5771	−0.0376	E.GLLSM^{ox}LLSVL**K**APVRNLALAAKAVAE.Q	134−159
888.8538	2663.5395	2663.5407	−0.0012	E.GLLSM^{ox}LLSVL**Q**APVRNLALAAKAVAE.Q	

[a] The masses of sequences containing K or Q at residue 144 are indicated.

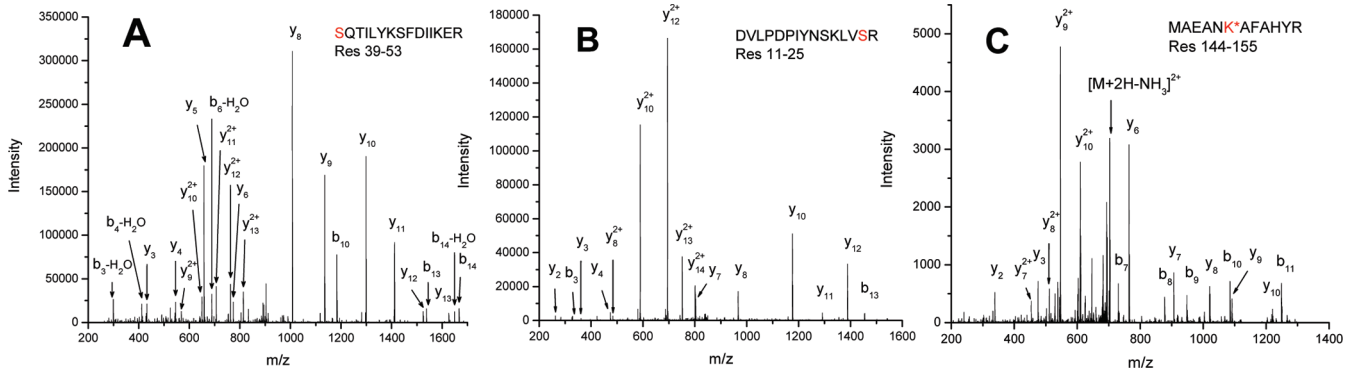


Figure 4. MS/MS spectra of tryptic peptides from ribosomal protein S7. (A) Sequencing error in the form of a serine residue at position number 39. (B) No sequencing error at residue number 24 despite the conspicuous sequence alignment shown in supplemental Figure 2 (Supporting Information). (C) Spectrum interpreted assuming monomethylated lysine at residue number 149.

Table 3. Observed Peptides that Confirm Proposed Post-Translational Modifications

	peptide	residues	*M*r (expt)	mass error	
		Trimethylation of K			
L11	LQIPAGK*ANPAPPVGPALGQAGVNIM GFCK*EFNAR	11−45	3616.95	0.01	+4 ion, *m*/*z* 905.24
		N-terminal monomethylation			
L16	*MLLPK	1−5	614.38	0.00	+1 ion
		N-terminal acetylation			
S9	*AQVQYYGTGR	2−11	1183.56	0.00	+2 ion, *m*/*z* 592.79
		Monomethylation of K			
S7	MAEANK*AFAHYR	144−155	1421.69	0.01	+2 ion, *m*/*z* 711.85
		Addition of 14.016 Da			
L16	VRMG*SGKGAPEGWVAVVKPGKVLFE	81−105	2611.43	0.00	+4 ion, *m*/*z* 653.86
	MG*SGKGAPEWVAVVKPGK	83−101	1868.00	0.01	+3 ion, *m*/*z* 623.68
	MG*SGKGAPEWVAVVKPGK	83−101	1868.00	0.00	+2 ion, *m*/*z* 935.01

error, the mass observed for L11 was still 126.2 Da too heavy. This mass shift was assumed to correspond to the addition of nine methyl groups, since trimethylation (+42 Da) on the primary amino groups of several L11 N-terminal domain residues is a canonical modification.[18,19,34,53−55] MS/MS analysis of tryptic peptides resulted in the confirmation of two +42 Da modifications on K17 and K40 (Table 3). And again, like the detection of a lysine to glutamine sequencing error, the identity of these +42 Da modifications could be attributed to trimethylations (+42.05 Da), and not another +42 Da modification such as acetylation (+42.01 Da), because of the mass accuracy of the FTMS instrument. The third site of +42 Da modification was unfortunately not conclusively identified. Potential sites could, however, be deduced based on observed tryptic cleavages, since trypsin is incapable of hydrolyzing peptide bonds adjacent to extensively methylated lysines.[56] Residues K7, K10, K71, K80, K81, and K104 could accordingly be eliminated from consideration. Considering that ribosomal protein L11 tends only to be methylated on its N-terminal domain, the remaining +42 Da modification likely occurs on the N-terminus (A2), K3, or K4. Residues aligning with *B. subtilis* L11 K3 are, however, unmodified in the well-characterized ribosomal proteomes of *E. coli* and *T. thermophilus*.[53,55,57,58] The most probable residues for the site of the third +42 Da modification therefore reduce to either the N-terminus (A2) or K4. Figure 5 shows the modified sites in *B. subtilis* L11 aligned with those of *E. coli* and *T. thermophilus* L11. Surprisingly, *B. subtilis* L11 exhibits the same extent of modification as *E. coli* L11 but does not appear to be modified at the same three residue positions.[57,58] Moreover, despite the N-terminal domain of *B. subtilis* L11 containing one more modifiable residue than *T. thermophilus* L11, *B. subtilis* L11 is not as extensively modified as *T. thermophilus* L11.[55]

Ribosomal protein L16, like L11, appeared to be post-translationally modified, since its whole protein mass was 28.2 Da too heavy. From peptide mass analysis, half of this mass differential could be localized to the N-terminus. The tryptic pentapeptide, MLLPK, from L16, in particular, was observed

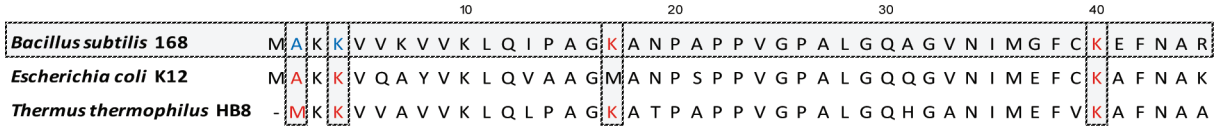


Figure 5. Multiple sequence alignment of the N-terminal domain of ribosomal protein L11: *B. subtilis*, *E. coli*, and *T. thermophilus*. Residues shown in red are confirmed sites of modification. Those in blue are potential sites of modification for the third trimethylation in *B. subtilis* L11. Uncolored residues are not modified.

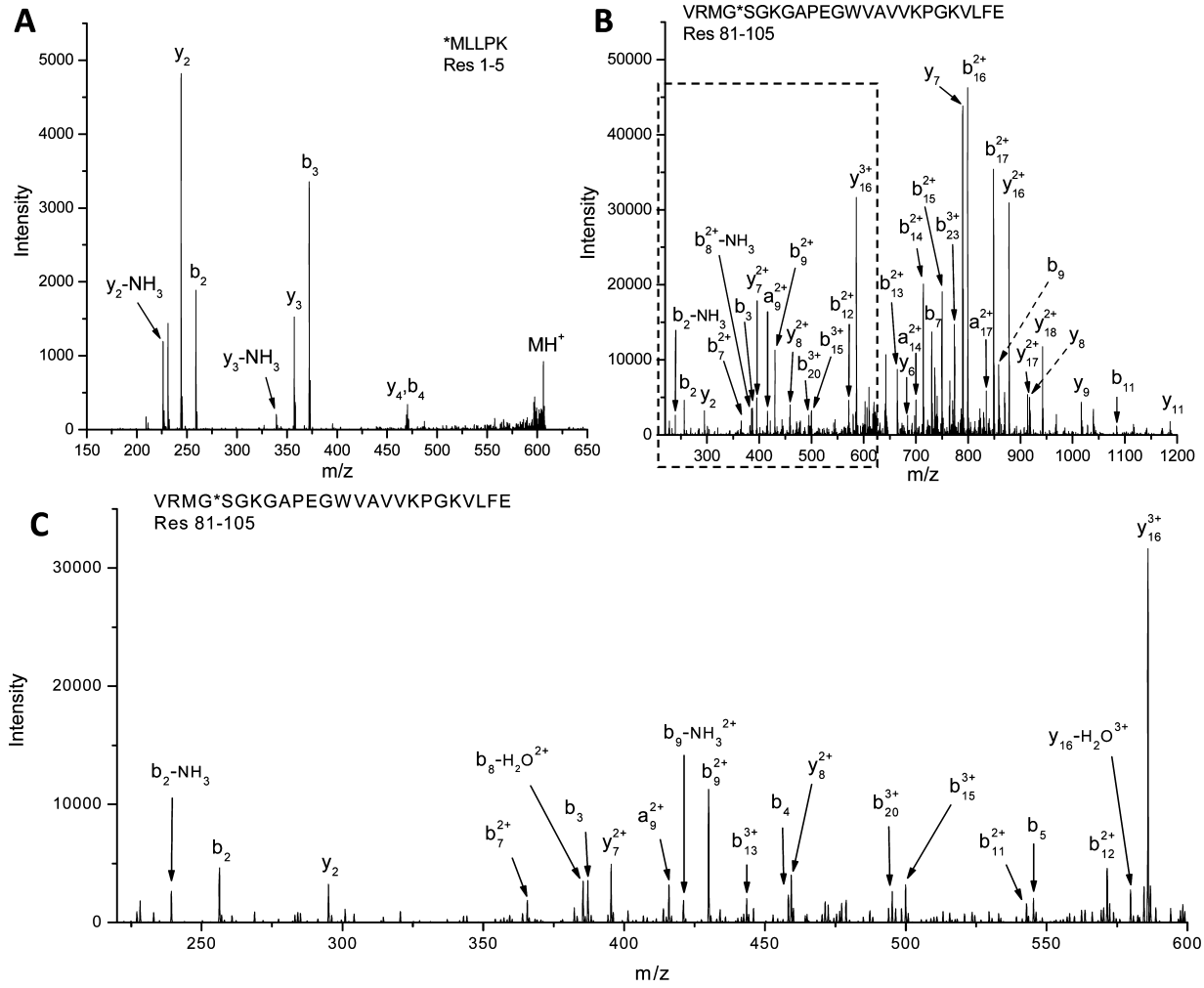


Figure 6. MS/MS spectra of apparently methylated peptides of ribosomal protein L16. (A) N-terminal peptide. (B) Internal peptide. (C) Expanded region from (B).

to be 14 Da too heavy. Figure 6A shows the MS/MS spectrum acquired for this precursor, in which an observed b_2 ion isolates the mass differential to the first two residues of the peptide and therefore protein. These observations are demonstrative of N-terminal methylation and are relatively unsurprising, since modification of ribosomal protein L16 in this way has been detected in several other bacteria.[18,34,59] Investigation of the remaining mass differential (+14 Da) was therefore all the more interesting. The site of a second methylation in a bacterial ribosomal protein L16 has not been reported, though the presence of such a modification has been speculated based on the observed whole protein masses of *E. coli* and *D. radiodurans* L16.[19,34] We have found, through our peptide measurements, the location of a second +14 Da mass differential in *B. subtilis* L16. L16 peptides containing residues 83−87 were found to be 14 Da too heavy. New theoretical masses were calculated for these peptides assuming the addition of a methyl group (+14.016 Da) to each sequence. These appear in Table 3 and are in excellent agreement with experimental masses measured for peptides containing residues 83−87. The MS/MS spectrum of a peptide from a Glu-C digest of purified L16 provided a strong indication for the location of this +14 Da mass differential. At first glance, the detected fragment ions seemed only to constrain the mass differential to somewhere between residues 84−87. However, less intense b-ions, notably b_3 and b_4, indicate that the site of the +14 Da mass differential is G84 (Figure 6B and C). Certainly, this would suggest the presence of a Gly to Ala sequencing error, but there are no suspected sequencing errors in the observed peptide, and the Gly at position 84 is actually conserved across all other prokaryotic ribosomal protein L16 sequences. Hence, the exact nature of the mass differential is not clear. If it is indeed localized on residue 84, possible explanations include a Gly to Ala mutation or an unprecedented post-translational modification, such as methylation of a backbone amide nitrogen. More

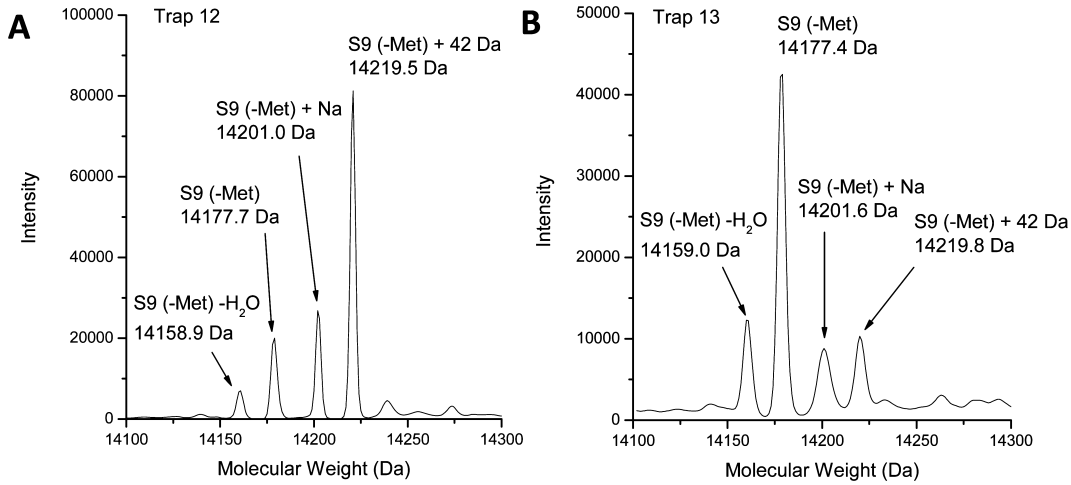


Figure 7. Mass spectra of ribosomal protein S9. (A) Acetylated form from trap 12. (B) Unacetylated form from trap 13.

extensive mass spectrometric characterization of this protein is needed to elucidate the source of this mass differential.

Although methylation was quite prevalent, another type of post-translational modification was also detected. Whole protein mass spectra of S9 exhibited an unmodified, save methionine removal, mass peak when derived from trap 13 and a near equally intense *ca.* +42.1 Da peak when derived from trap 12 (Figure 7). Elution of modified S9 from the SCX dimension thus preceded that of the unmodified protein by about 5 min (Figure 7A vs B), suggesting that the modified S9 had a less positive charge. Indeed, tryptic digestion of a trap 12 fraction yielded a peptide with a 42 Da modification on the N-terminal residue of S9 and an experimental mass in agreement with acetylation to within 0.0011 Da (Table 3). These observations are in accord with the previous detection of *B. subtilis* S9 exhibiting two post-translational states: unmodified and N-terminally acetylated.[49] It is interesting to conjecture about the role of S9 dimorphism, including that, as with L7, acetylation may enhance interaction with other components of the ribosome.[60] Alternative growth conditions may aid in elucidating the role of S9 N-terminal acetylation, particularly if stress conditions are shown to alter the relative abundance of this modification.

Finally, two other ribosomal proteins may likewise be post-translationally modified or just simply inefficiently processed. Whole protein mass spectra of L29 and S6 presented peaks at their theoretical molecular weights as well as one-third and one-tenth intense +28 Da peaks. Again, similar to modified S9, these +28 Da forms of L29 and S6 eluted from the SCX dimension one trap before their counterparts. The most convincing explanation for both these mass shifts is therefore the presence of N-formyl groups, in which the protein would have maintained a less positive charge during SCX. Furthermore, the relative intensities of these peaks suggest that the presence of these N-formyl groups was due to inefficient processing of protein N-termini. This explanation is consistent with detection of other inefficient processing in the *B. subtilis* ribosomal proteome, retention of N-terminal methionines on proteins expected to be processed by methionine aminopeptidase. For example, mass spectra of ribosomal proteins L3, L30, S2, S10, and S17 exhibited +131 Da peaks in addition to the 4 to 10 times more intense peaks representing the fully processed and predominant forms of the proteins.

In summary then, only ribosomal proteins L11, L16, S7 and S9 are subject to covalent post-translational modification under these experimental conditions. The absence of several canonical post-translational modifications is therefore striking. During this work, L12 and S12 were observed to be entirely unmodified, a particularly significant observation since these proteins are uniquely modified in a number of other bacteria.[11,15−17,19,53,54] For instance, in *E. coli*, L12 is subject to acetylation and methylation depending on growth conditions,[61,62] and S12 is subject to the previously mentioned methylthiolation. The absence of L12 modification in *B. subtilis* may demonstrate a dependence of modification on growth condition, as in *E. coli*, or the simple inability of *B. subtilis* to modify the protein in such a manner. The absence of S12 modification in *B. subtilis*, on the other hand, can be more thoroughly explained in light of recent findings published by Roberts and co-workers.[12] Their studies have shown that the methylthiotransferase responsible for modification of S12 in *E. coli*, RimO, represents a unique subfamily of likely methylthiotransferases and that this subfamily of enzymes contains members from *T.thermophilus* and *R. palustris*, both of which are known to modify S12,[15,55] but no member from *B. subtilis*. Thus, the absence of S12 methylthiolation in *B. subtilis* likely reflects the lack of a RimO-like methylthiotransferase capable of recognizing ribosomal protein S12 as a substrate for modification.

Structural Homology. Structural homology between *B. subtilis*, *E. coli*, and *T. thermophilus* ribosomal proteins was probed through chemical labeling with *S*-methylthioacetimidate. *B. subtilis* ribosomal proteins were amidinated before and after disassembly of the ribosome. When ribosomal proteins were amidinated after disassembly of the ribosome, all primary amino groups were labeled. However, amidination of the proteins before disassembly of the ribosome led to lesser extents of modification. This extent of modification for each protein was related to the solvent accessibility of modifiable sites and in turn the conformation of the protein as it is assembled on the ribosome. For example, 9 to 14 of the 16 modifiable sites of S4 were amidinated (Figure 8A). An intensity-weighted average for the experimental extent of modification (*ca.* 12) was subsequently compared with a predicted value derived from crystal structure data. Because the crystal structure of the *B. subtilis* ribosome has not yet been reported, we have aligned the modifiable residues of *B. subtilis* ribosomal proteins onto the sequences of *E. coli* and *T. thermophilus* ribosomal proteins and have interpreted the solvent accessibility of these aligned residues in the published crystal structures of these bacterial ribosomes. An example of these

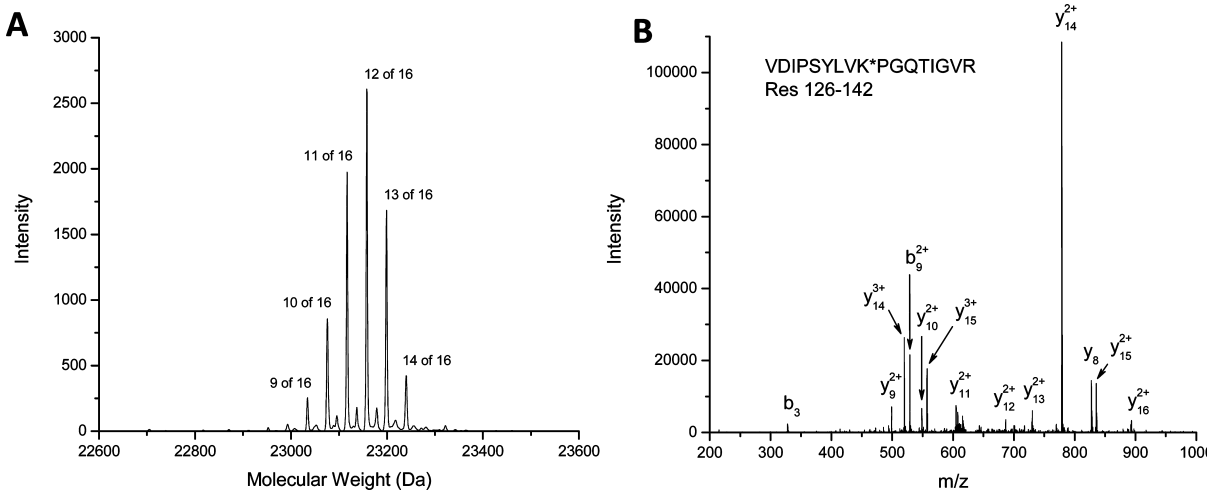


Figure 8. Mass spectrometric analysis of S4 amidinated before disassembly of the ribosome. (A) Deconvoluted whole protein mass spectrum. Numbers of amidino groups incorporated are indicated. (B) MS/MS spectrum of an amidinated tryptic peptide. Asterisked residue (K134) is shown to be modified.

interpretations is shown for ribosomal protein S4 in Figure 9A and B. From both interpretations, 13 out of 16 modifiable sites appeared to be solvent accessible (for details concerning this procedure see Material and Methods). This predicted extent of modification matched that of the observed value to within experimental error, thus ribosomal protein S4 of *B. subtilis* and both crystal structure bacteria were assumed to be structurally homologous.

Comparisons of the predicted and experimental extents of modification for all *B. subtilis* ribosomal proteins of the large and small subunit are shown in Figures 10 and 11, respectively. Theoretical and experimental extents of modification matched particularly well, with the exception of L11 and S3. L11 appeared to be significantly undermodified, while S3 appeared to be significantly overmodified with respect to theoretical interpretation. The experimental extent of modification of L11 is consistent with extensive post-translational methylation of otherwise modifiable sites given that we have previously found trimethylated amino groups to be unreactive with SMTA. The results for S3, however, suggest that either the protein has significantly more accessible Lys residues in its native structure than expected, or it tends to dissociate from the ribosome before/during the amidination reaction. Resultant excessive amidination of other ribosomal proteins, specifically that of S14, would only yield one additional amidinated Lys residue, a number within experimental error. We were therefore unable to determine whether S3 was truly dissociating from the ribosome or simply exhibiting significant flexibility in solution.

To provide further, higher resolution criteria for assessing structural homology, MS/MS analysis of peptides derived from native amidinated ribosomal proteins was completed. For instance, MS/MS of an amidinated S4 peptide facilitated the mapping of K134 as a site of modification (Figure 8B). In sum, the residue positions of 47 modified sites were mapped. The majority of these, 42 of 47, were consistent with solvent accessibility of both *T. thermophilus* and *E. coli* proteins (Supplemental Table 5, Supporting Information). Two were found to be solely consistent with interpretation of the *E. coli* structure, while two others were solely consistent with interpretation of the *T. thermophilus* structure. Amidination of L35 K59 suggests that the conformation of *B. subtilis* L35 more closely resembles that of E. *coli* L35. On the contrary, amidination of S4 K100 suggests that of *B. subtilis* S4 more closely resembles that of *T. thermophilus* S4. Additionally, individual mapped modifications of ribosomal protein L3, specifically K196 and K203, chimerically represent L3 of *T. thermophilus* and *E. coli*. Lastly, one modified site, K59 of L20, did not appear to be consistent with either prediction. This residue resides on the interface between L20 and the 23S rRNA, where it may be possible for some population of L20 to dissociate from the rRNA or for rRNA degradation to have rendered this residue sufficiently accessible for amidination.

Collectively, these data demonstrate the high structural homology of ribosomal proteins from the eubacteria, *B. subtilis*, *T. thermophilus* and *E. coli*. Without a doubt, requisite RNA-protein interactions and their stabilizing necessity dictate the tertiary and quaternary structures of ribosomal proteins to the point that only subtle variations in structure are evolutionarily tolerated.

Notable Proteins. Three putative zinc-binding ribosomal proteins from *B. subtilis*, RpsNA (S14), RpmE (L31), and RpmGA (L33.1), were identified as constituents of the ribosome under the growth conditions of this study. These proteins are referred to as the C^+ forms in the proteome, as they contain CXXC motifs suitable for coordinating zinc.[22] Their paralogs, not present in this study, are conversely all C^- forms, with the single exception of RpmGB (L33.2), because they do not contain the CXXC motifs needed to effectively bind zinc.[21,22] Previously, it has been demonstrated that under zinc-limited growth RpmE (L31) is actively displaced by the C^- paralog YtiA and that the C^- paralog of S14 (RpsNB) is incorporated into the ribosome upon *de novo* synthesis.[23−25] Implications of the role of these ribosomal proteins in zinc homeostasis and stress response are very intriguing. Helmann has postulated that they may provide a dynamic reservoir of zinc to the cell.[63] Under zinc-limited growth, Zur derepresses the paralogous genes and C^- ribosomal proteins are expressed. Reserves of zinc are then mobilized by the C^- ribosomal proteins replacing their C^+ counterparts in the ribosome. Our characterization of *B. subtilis* ribosomal proteins present under normal growth conditions may lead to a better understanding of this system.

Although weakly associated ribosomal protein L25 was neither detected by whole protein mass spectrometry nor expected to be present in these preparations of ribosomes,

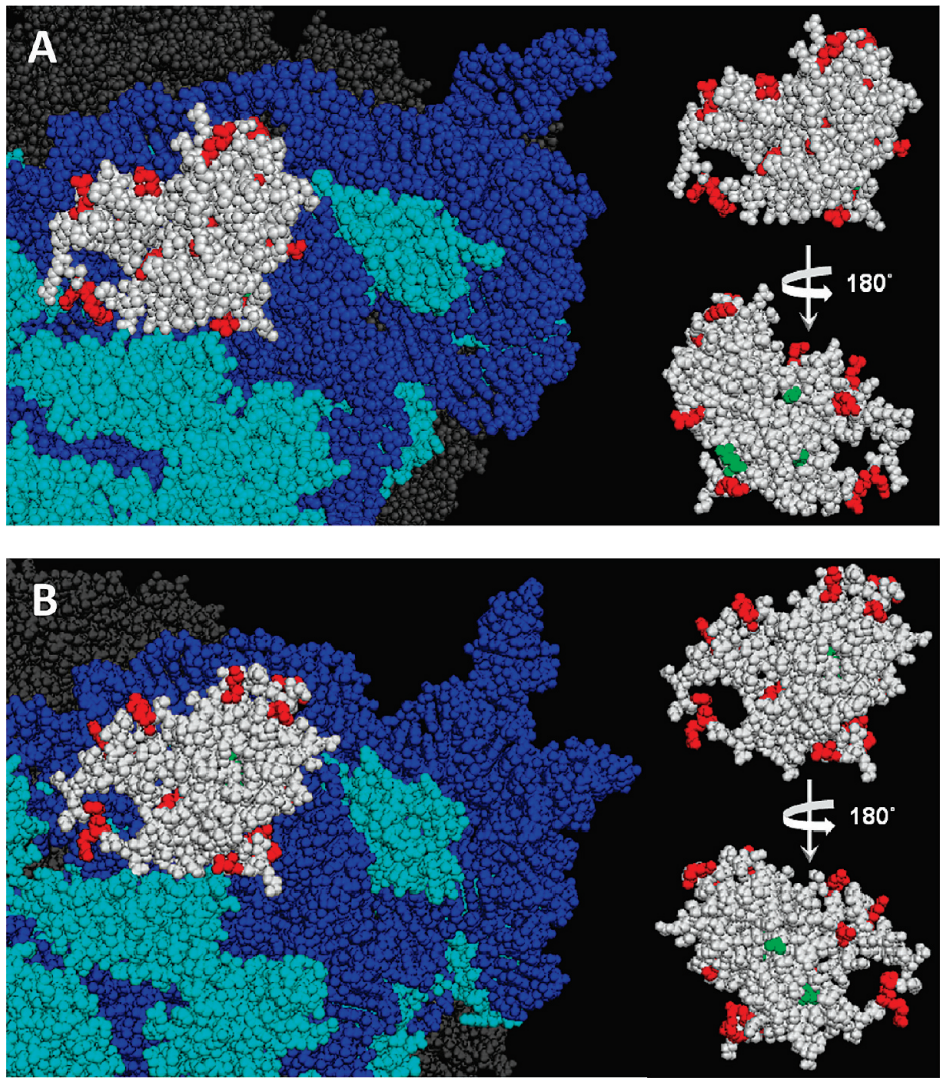


Figure 9. X-ray crystal structures of 70S ribosomes of (A) *T. thermophilus* HB8 (PDB: 2J00 and 2J01) and (B) *E. coli* K12 (PDB: 2AVY and 2AW4). Ribosomal protein S4 is shown in white, other small subunit proteins in cyan, 16S rRNA in blue, and the 50S subunit in dark gray. Residues that align with the modifiable sites of *B. subtilis* S4 are shown in either red or green, depending on their apparent solvent accessibility; those shown in red are accessible. Those shown in green are largely buried.

several peptides resulting from proteolyzed L25 were observed. The sequence coverage afforded by these peptides amounted to 36%. This is much lower than the average sequence coverage of 85% for ribosomal proteins and suggests that a trace amount of L25 (general stress protein Ctc) may associate with the ribosome under these experimental conditions. This would be consistent with previous observations that Ctc, despite not being considered a constitutive component of the ribosome, is present in *B. subtilis* even during exponential growth.[45−47,64]

In addition to the previously mentioned ribosomal proteins, 28 nonribosomal proteins were identified in the combined peptide MS/MS data set. Six of these were identified in peptide experiments of different ribosome preparations (Hag, SpoVG, YdbR, EF-Tu, YvyD, and HbsU). A few of these are likely high copy number proteins given their functional roles and their presence is not surprising. For example, Hag is a flagellin structural protein,[65] and SpoVG is a negative regulator of sporulation, specifically asymmetric septation.[66] Others, however, may be present not only because they are highly abundant in *B. subtilis* but because they may associate with the 70S ribosome. That is, YdbR is an RNA helicase that has been proposed to be involved in translation,[67] HBsu has been identified as an integral component of a *B. subtilis* SRP-like particle expected to associate with ribosomes,[68] EF-Tu is a translation factor, and YvyD is a homologue of two *E. coli* proteins, YfiA and YbhH, shown to bind to ribosomes. YvyD is possibly the most interesting of these, since, unlike the others, it was consistently observed in all ribosomal whole protein MS and bottom-up experiments and confidently identified based on an observed 79% sequence coverage and whole protein mass of 21979.6 Da. In crystallographic studies, *E. coli* YfiA was described as a cold shock protein that binds to the subunit interface of the 70S ribosome.[69] In other studies, *E. coli* YfiA and YbhH were shown to associate with ribosomes during stationary phase growth in order to regulate translation.[70−72] Both functional descriptions of the *E. coli* homologues suggest that YvyD may bind to 70S ribosomes as a means to inhibit translation during conditions unfavorable to protein biosynthesis.[73] There is a possibility that the 30 °C growth temperature or sample handling temperatures employed in this study were sufficiently low to induce a cold shock response in *B. subtilis* resulting in the association of YvyD with the 70S ribosome or

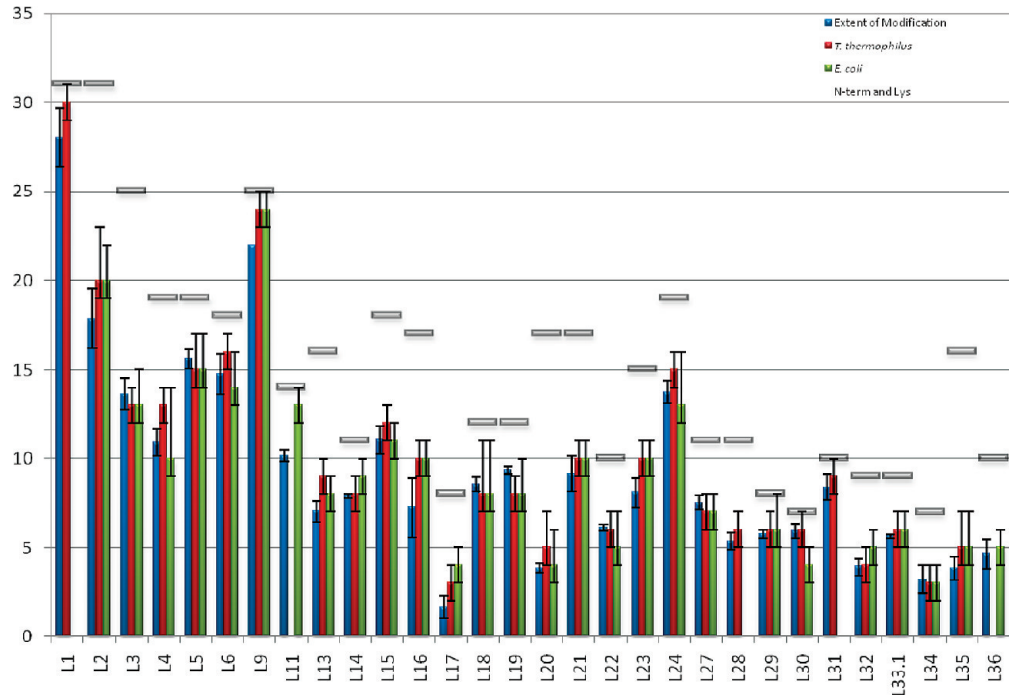


Figure 10. Comparison of extent of modification with the number of sequence-aligned solvent accessible sites of modification in the crystal structures of *E. coli* and *T. thermophilus* 50S subunits (*E. coli* K12, PDB: 2AW4; *T. thermophilus* HB8, PDB: 2J01). The counts of sequence aligned solvent accessible sites of modification for a given protein in the crystal structures of *E. coli* and *T. thermophilus* are shown in green and red, respectively. The intensity-weighted average extent of modification of a given *B. subtilis* ribosomal protein is shown in blue. Floating bars represent the count of modifiable sites (N-term + Lys) in the *B. subtilis* ribosomal proteins. Error bars are discussed in the text. Also, some proteins were not included in the ribosome crystal structures and are therefore not represented in this figure.

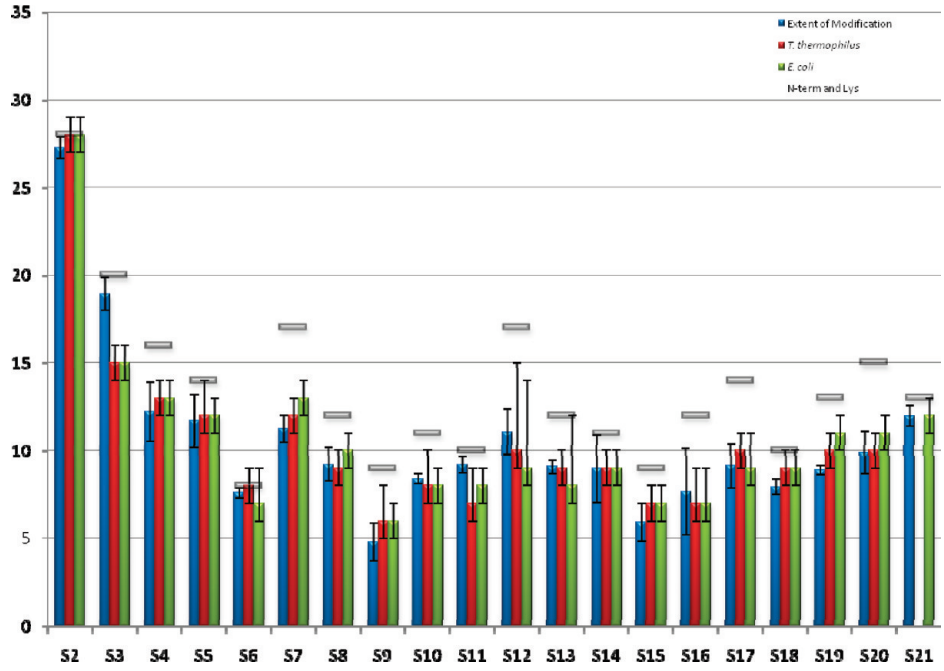


Figure 11. Comparison of extent of modification with the number of sequence-aligned solvent accessible sites of modification in the crystal structures of *E. coli* and *T. thermophilus* 30S subunits (*E. coli* K12, PDB: 2AVY; *T. thermophilus* HB8, PDB: 2J00). The counts of sequence aligned solvent accessible sites of modification for a given protein in the crystal structures of *E. coli* and *T. thermophilus* are shown in green and red, respectively. The intensity-weighted average extent of modification of a given *B. subtilis* 30S ribosomal protein is shown in blue. Floating bars represent the count of modifiable sites (N-term + Lys) in the *B. subtilis* ribosomal proteins. Note that the *T. thermophilus* ribosome does not contain a ribosomal protein S21.

that a detectable amount of YvyD, like Ctc, associates with 70S ribosomes even during exponential growth. Experiments involving different growth conditions may elucidate the roles of these proteins. Also, cross-linking studies should allow the

nonribosomal proteins present due to specific interaction with the ribosome to be discriminated from those present due to high levels of expression.

Conclusion

The characterization of *B. subtilis* ribosomal proteins outlined in this study has cataloged all expected gene products, extended the structural homology of ribosomal proteins, and documented post-translational modifications. We have also identified an alarming number of sequencing errors in the originally published reference genome of *B. subtilis* 168 but have found that these errors are entirely corrected in a recently revised version of the genome (GenBank AL009126.3). Our characterization demonstrates that some post-translational states of ribosomal proteins in *B. subtilis* under normal growth conditions are novel in comparison with other eubacteria. Ribosomal protein S9 is partially N-terminally acetylated, L11 is modified to the same extent as *E. coli* L11, but not at the same three residues, and S7 is noncanonically monomethylated on K149. Furthermore, some canonical post-translational modifications, namely β-methylthiolation of S12 and modification of L12, are entirely absent. At the moment, the significance of ribosomal protein modifications is, for the most part, poorly understood. Observations of their presence or absence in ribosomal proteomes of different bacterial phyla, as presented here, should lead to an improved understanding of their significance and function. Beyond these topics, the most interesting aspects of the *B. subtilis* ribosomal proteome have just begun to be elucidated. Demonstration of alternating ribosomal protein paralogs and association of L25 with the ribosome under stress conditions has been of utmost interest in recent literature.[23−25,45,46,63] The present results aid future studies, such as possible forthcoming investigations of L33 paralogs, by thoroughly characterizing the ribosomal proteome under normal growth conditions.

Acknowledgment. We thank Daniel Kearns for his gift of *B. subtilis* 168 and Jon Karty for his assistance with Bioanalyte ProTrawler/Regatta. This research was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc., and by National Science Foundation Grants CHE-0518234 and CHE-0832651.

Supporting Information Available: Tables of SCX dimension and RPLC gradients. Peptide MS/MS matches from which sequencing errors were confirmed. Peptide MS/MS matches from which conflicts between the published reference genome and other sequencing information have been resolved. An MS/MS spectrum of an L11 peptide containing the V35I sequencing error. A multiple sequence alignment of ribosomal protein S7. A table of mapped modified sites of ribosomal proteins amidinated before disassembly of the ribosome. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Voet, D.; Voet, J. G. In *Biochemistry*, 3rd ed.; J. Wiley & Sons: Hoboken, NJ, 2004; p 1310.
(2) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920–930.
(3) Polacek, N.; Gaynor, M.; Yassin, A.; Mankin, A. S. *Nature* **2001**, *411*, 499–501.
(4) Wilson, D.; Nierhaus, K. *Crit. Rev. Biochem. Mol. Biol.* **2005**, *40*, 243–267.
(5) Blaha, G.; Wilson, D. N.; Stoller, G.; Fischer, G.; Willumeit, R.; Nierhaus, K. H. *J. Mol. Biol.* **2003**, *326*, 887–897.
(6) Deuerling, E.; Schulze-Specking, A. *Nature* **1999**, *400*, 693–696.
(7) Kramer, G. T.; Rauch, T.; Rist, W.; Vorderwuelbecke, S.; Patzelt, H.; Schulze-Specking, A.; Ban, N.; Deuerling, E.; Bukau, B. *Nature* **2002**, *419*, 171–174.
(8) Wahl, M. C.; Moller, W. *Curr. Protein Pept. Sci.* **2002**, *3*, 93–106.
(9) Kurland, C. G.; Jørgensen, F.; Richter, A.; Ehrenberg, M.; Bilgin, N.; Rojas, A.-M. Through the accuracy window. In *The Ribosome—Structure, Function, and Evolution*; Dahlberg, A., Hill, W. E., Garrett, R. A., Moore, P. B., Schlessinger, D., Warner, J. R., Eds.; American Society Microbiology: Washington, D.C., 1990; pp 513−526.
(10) Kurland, C. G.; Hughes, D.; Ehrenberg, M. Escherichia coli and Salmonella. In *Cellular and Molecular Biology*, 2nd ed.; Neidhardt, F. C., Ed.; ASM Press: Washington, D.C., 1996; pp 974−1004.
(11) Kowalak, J. A.; Walsh, K. A. *Protein Sci.* **1996**, *5*, 1625–1632.
(12) Anton, B. P.; Saleh, L.; Benner, J. S.; Raleigh, E. A.; Kasif, S.; Roberts, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1826–1831.
(13) Louie, D. F.; Resing, K. A.; Lewis, T. S.; Ahn, N. G. *J. Biol. Chem.* **1996**, *271*, 28189–28198.
(14) Vladtmirov, S. N.; Ivanov, A. V.; Karpova, G. G.; Musolyamov, A. K.; Egorov, T. A.; Thiede, B.; Wittmann-Liebold, B.; Otto, A. *Eur. J. Biochem.* **1996**, *239*, 144–149.
(15) 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.
(16) Suh, M.-J.; Hamburg, D.-M.; Gregory, S. T.; Dahlberg, A. E.; Limbach, P. A. *Proteomics* **2005**, *5*, 4818–4831.
(17) Carr, J. F.; Hamburg, D. M.; Gregory, S. T.; Limbach, P. A.; Dahlberg, A. E. *J. Bacteriol.* **2006**, *188*, 2020–2023.
(18) Running, W. E.; Ravipaty, S.; Karty, J. A.; Reilly, J. P. *J. Proteome Res.* **2007**, *6*, 337–347.
(19) Running, W. E.; Reilly, J. P. *J. Proteome Res.* **2009**, *8* (3), 1228–1246.
(20) Sonenshein, A. L.; Hoch, J. A.; Losick, R. *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*; ASM Press: Washington, D.C., 2002.
(21) Panina, E. M.; Mironov, A. A.; Gelfand, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9912–9917.
(22) Makarova, K. S.; Ponomarev, V. A.; Koonin, E. V. *Genome Biol.* **2001**, *2*, research0033.110033.14.
(23) Natori, Y.; Nanamiya, H.; Akanuma, G.; Kosono, S.; Kudo, T.; Ochi, K.; Kawamura, F. *Mol. Microbiol.* **2007**, *63*, 294–307.
(24) Nanamiya, H.; Akanuma, G.; Natori, Y.; Murayama, R.; Kosono, S.; Kudo, T.; Kobayashi, K.; Ogasawara, N.; Park, S.; Ochi, K.; Kawamura, F. *Mol. Microbiol.* **2004**, *52*, 273–283.
(25) Akanuma, G.; Nanamiya, H.; Natori, Y.; Nomura, N.; Kawamura, F. *J. Bacteriol.* **2006**, *188*, 2715–2720.
(26) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935–1942.
(27) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. *Science* **2005**, *310*, 827–834.
(28) Suckau, D.; Mak, M.; Przybylski, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5630–4.
(29) Liu, X.; Broshears, W. C.; Reilly, J. P. *Anal. Biochem.* **2007**, *367*, 13–19.
(30) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2005**, *77*, 7274–7281.
(31) Karty, J. A.; Running, W. E.; Reilly, J. P. *J. Chromatogr. B.* **2007**, *847*, 103–113.
(32) Kunst, F.; et al. *Nature* **1997**, *390* (6657), 249–256.
(33) Kearns, D. B.; Losick, R. *Mol. Microbiol.* **2003**, *49* (3), 581–590.
(34) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105–112.
(35) Spedding, G. In *Ribosomes and Protein Synthesis: A Pratical Approach*; Rickwood, D., Hames, B. D., Eds.; Oxford Press: New York, 1990; pp 1−29.
(36) Good, N. E.; Winget, G. D.; Winter, W.; Connolly, T. N.; Izawa, S.; Singh, R. M. M. *Biochemistry* **1966**, *5* (2), 467–77.
(37) Read, S. M.; Northcote, D. H. *Anal. Biochem.* **1981**, *116* (1), 53–64.
(38) Beardsley, R. L.; Running, W. E.; Reilly, J. P. *J. Proteome Res.* **2006**, *5*, 2935–2946.
(39) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
(40) Combet, C.; Blanchet, C.; Geourjon, C.; Deléage, G. *Trends Biochem. Sci.* **2000**, *25*, 147–150.
(41) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, CA, 2002.
(42) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548–58.

(43) Farwell, M. A.; Roberts, M. W.; Rabinowitz, J. C. *Mol. Microbiol.* **1992**, *6* (22), 3375–83.
(44) Isono, K.; Isono, S. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 767–70.
(45) Schmalisch, M.; Langbein, I.; Stulke, J. *J. Mol. Microbiol. Biotechnol.* **2002**, *4*, 495–501.
(46) Korepanov, A. P.; Gongadze, G. M. *Biochemistry (Moscow)* **2004**, *69*, 607–611.
(47) Gryaznova, O. I.; Davydova, N. L.; Gongadze, G. M.; Jonsson, B. H.; Garber, M. B.; Liljas, A. *Biochimie* **1996**, *78*, 915–919.
(48) Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T. *Mol. Cell. Proteomics* **2006**, *5*, 2336–2349.
(49) Higo, K.; Otaka, E.; Osawa, S. *Mol. Gen. Genet.* **1982**, *185*, 239–44.
(50) Suh, J.-W.; Boylan, S. A.; Oh, S. H.; Price, C. W. *Gene* **1996**, (169), 17–23.
(51) Yasumoto, K.; Liu, H.; Jeong, S. M.; Ohashi, Y.; Kakinuma, S.; Tanaka, K.; Kawamura, F.; Yoshikawa, H.; Takahashi, H. *Microbiology* **1996**, (142), 3039–3046.
(52) Hill, F.; Gemund, C.; Benes, B.; Ansorge, W.; Gibson, T. J. *EMBO Rep.* **2000**, *1* (1), 29–31.
(53) Alix, J. *Adv. Exp. Med. Biol.* **1988**, *231*, 371–385.
(54) Polevoda, B.; Sherman, F. *Mol. Microbiol.* **2007**, *65* (3), 590–606.
(55) Cameron, D. M.; Gregory, S. T.; Thompson, J.; Suh, M.-J.; Limbach, P. A.; Dahlberg, A. E. *J. Bacteriol.* **2004**, *186* (17), 5819–5825.
(56) Poncz, L.; Dearborn, D. G. *J. Biol. Chem.* **1983**, *258* (3), 1844–50.
(57) Lederer, F.; Alix, J. H.; Hayes, D. *Biochem. Biophys. Res. Commun.* **1977**, *77* (2), 470–80.
(58) Dognin, M. J.; Wittmann-Liebold, B. *FEBS Lett.* **1977**, *84* (2), 342–6.
(59) Brosius, J.; Chen, R. *FEBS Lett.* **1976**, *68* (1), 105–9.
(60) Gordiyenko, Y.; Deroo, S.; Zhou, M.; Videler, H.; Robinson, C. V. *J. Mol. Biol.* **2008**, *380* (2), 404–414.
(61) Chang, F. N. *J. Bacteriol.* **1978**, *135* (3), 1165–6.
(62) Deusser, E.; Wittmann, H. G. *Nature* **1972**, *238* (5362), 269–70.
(63) Moore, C. M.; Helmann, J. D. *Curr. Opin. Microbiol.* **2005**, *8*, 188–195.
(64) Hilden, I.; Krath, B. N.; Hove-Jensen, B. *J. Bacteriol.* **1995**, *177* (24), 7280–4.
(65) Mirel, D. B.; Chamberlin, M. J. *J. Bacteriol.* **1989**, *171* (6), 3095–101.
(66) Matsuno, K.; Sonenshein, A. L. *J. Bacteriol.* **1999**, *181* (11), 3392–3401.
(67) Ando, Y.; Nakamura, K. *Biosci. Biotechnol. Biochem.* **2006**, *70* (7), 1606–1615.
(68) Nakamura, K.; Yahagi, S.-I.; Yamazaki, T.; Yamane, K. *J. Biol. Chem.* **1999**, *274* (19), 13569–13576.
(69) Vila-Sanjurjo, A.; Schuwirth, B.-S.; Hau, C. W.; Cate, J. H. D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1054–1059.
(70) Ueta, M.; Yoshida, H.; Wada, C.; Baba, T.; Mori, H.; Wada, A. *Genes Cells* **2005**, *10*, 1103–1112.
(71) Maki, Y.; Yoshida, H.; Wada, A. *Genes Cells* **2000**, *5* (12), 965–974.
(72) Ueta, M.; Ohniwa, R. L.; Yoshida, H.; Maki, Y.; Wada, C.; Wada, A. *J. Biochem.* **2008**, (143), 425–433.
(73) Wilson, D. N.; Nierhaus, K. H. *Crit. Rev. Biochem. Mol. Biol.* **2007**, *42* (3), 187–219.

PR801114K