Supplemental Figures and Legends

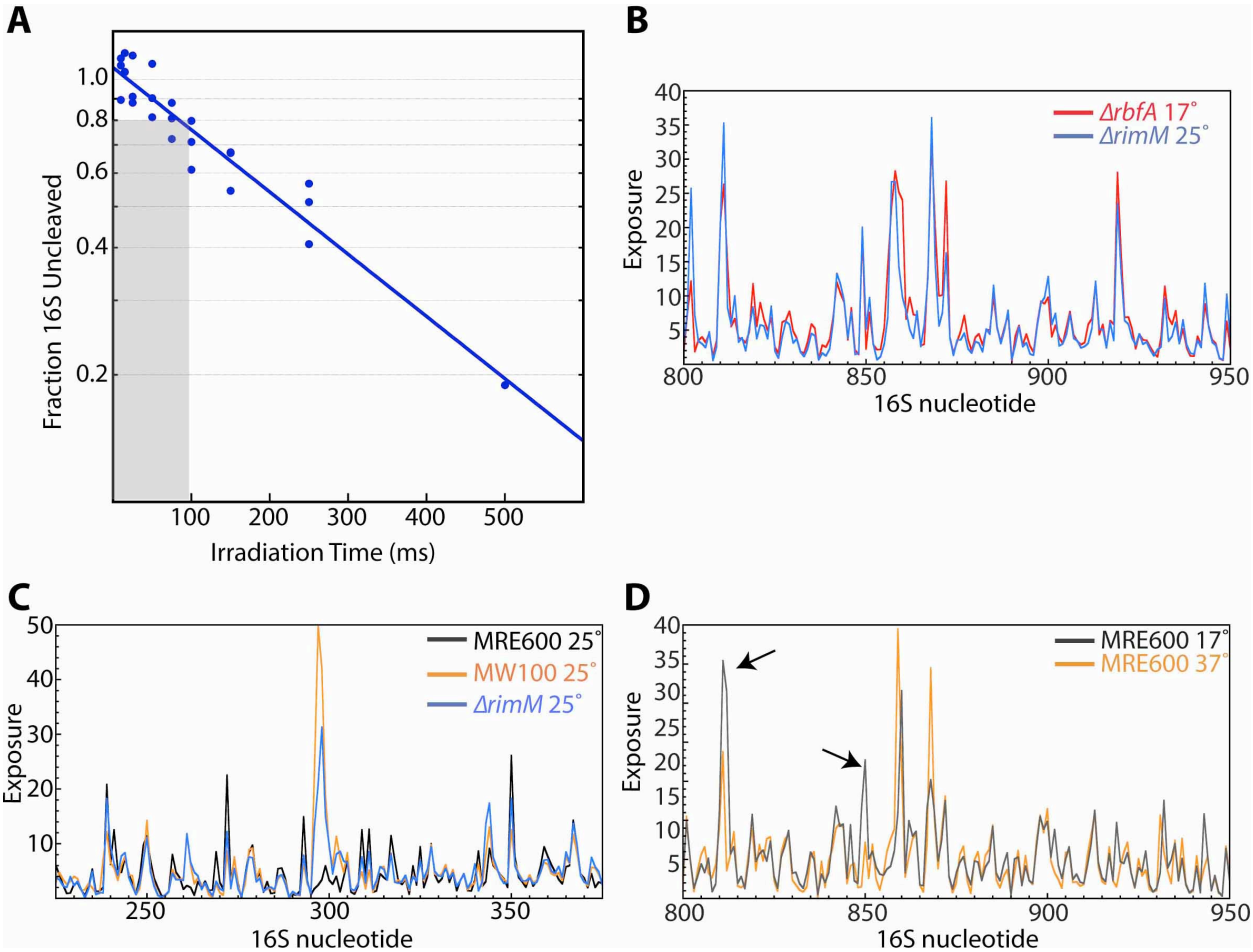

Figure S1, related to Figure 1. Integrity of *in vivo* footprinting data

(A) Frozen MRE600 cell samples were irradiated at NSLS X28C as described in Experimental Procedures. Total RNA was purified, and primer extension with primer 1486 (Table S1) was used to quantify the proportion of uncleaved 16S rRNA in the samples relative to an unirradiated control. After 100 ms, 80% of the 16S rRNA was full-length. At this dose, a Poisson distribution of cleavage events predicts only 2% of molecules are cleaved more than once (Brenowitz et al., 1986), minimizing skew in the fragment detection.

(B) Integrated cleavage intensities aligned to the 16S sequence. Cleavage intensities from the mutant strains are similar at most positions.

(C) Differences between the *ΔrimM* parental strain MW100 and MRE600. In a few regions, footprinting patterns from MW100 appear more like those from the assembly factor deletion strains than the patterns from MRE600. One example includes exposures of nucleotides 297-300, located at the tip of helix 12.

(D) Effects of growth temperature on MRE600. A few nucleotides experience changes in solvent accessibility due to temperature effects (arrows) and are excluded from annotation in the data sets for the mutant strains.

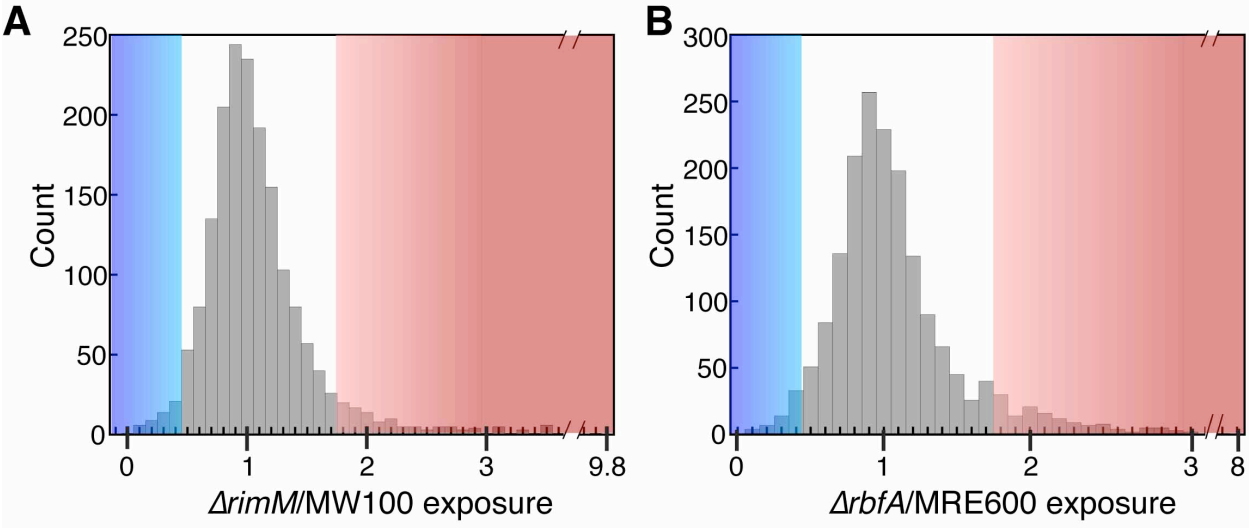

Figure S2, related to Figure 2. Histograms of 16S rRNA cleavage by hydroxyl radical

(A) For each nucleotide the ratio of the extent of cleavage (backbone exposure) in the *ΔrimM* MW37 strain over the parental MW100 strain was calculated. Ratios >1 indicate increased exposure in the mutant strain, while ratios < 1represent nucleotides with increased protection. Cut-off values for interpreting exposed or protected nucleotides were empirically set at 80% of the median. Colored areas denote the clusters used in Figure 2.

(B) As in (A), but for the *ΔrbfA* strain relative to MRE600.

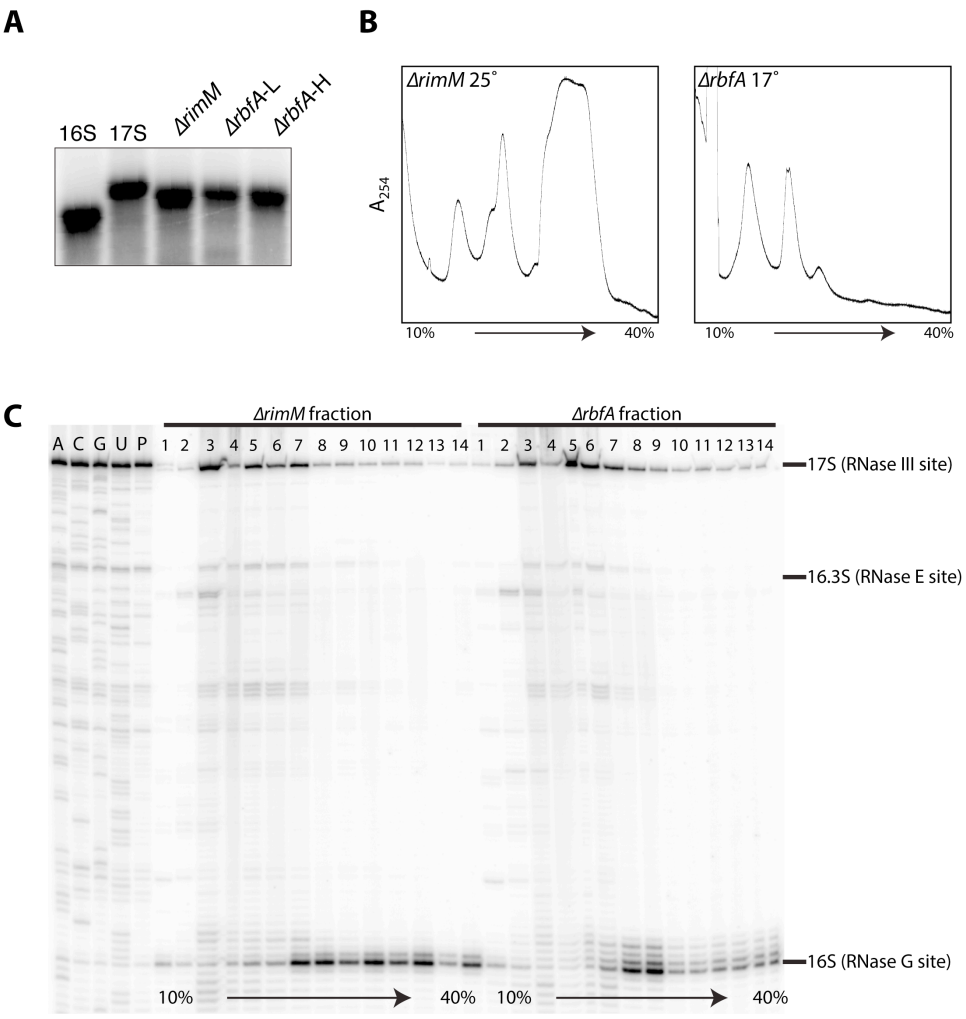

Figure S3, related to Figure 3. Analysis of 17S pre-rRNA in polysomes.

(A) Northern blot of 17S rRNA purified from isolated pre-30S complexes. L and H designate light and heavy Δ*rbfA* complexes. The intermediate mobility between in vitro transcribed 16S and 17S markers is due to cleavage near the RNase E site.

(B) A_{254} traces of sucrose gradient profiles, using chloramphenicol to preserve polysomes. These gradients were run in parallel to those fractionated in Figure 3G. Addition of chloramphenicol to *ΔrbfA* cells has little effect on the ribosome distribution, but in the *ΔrimM* strain chloramphenicol causes a large accumulation of an apparently aberrant +70S peak.

(C) Gel of primer extension experiment analyzed in Figure 3G. Primer 46 (Table S1) was used to extend the 5' end of the rRNA.

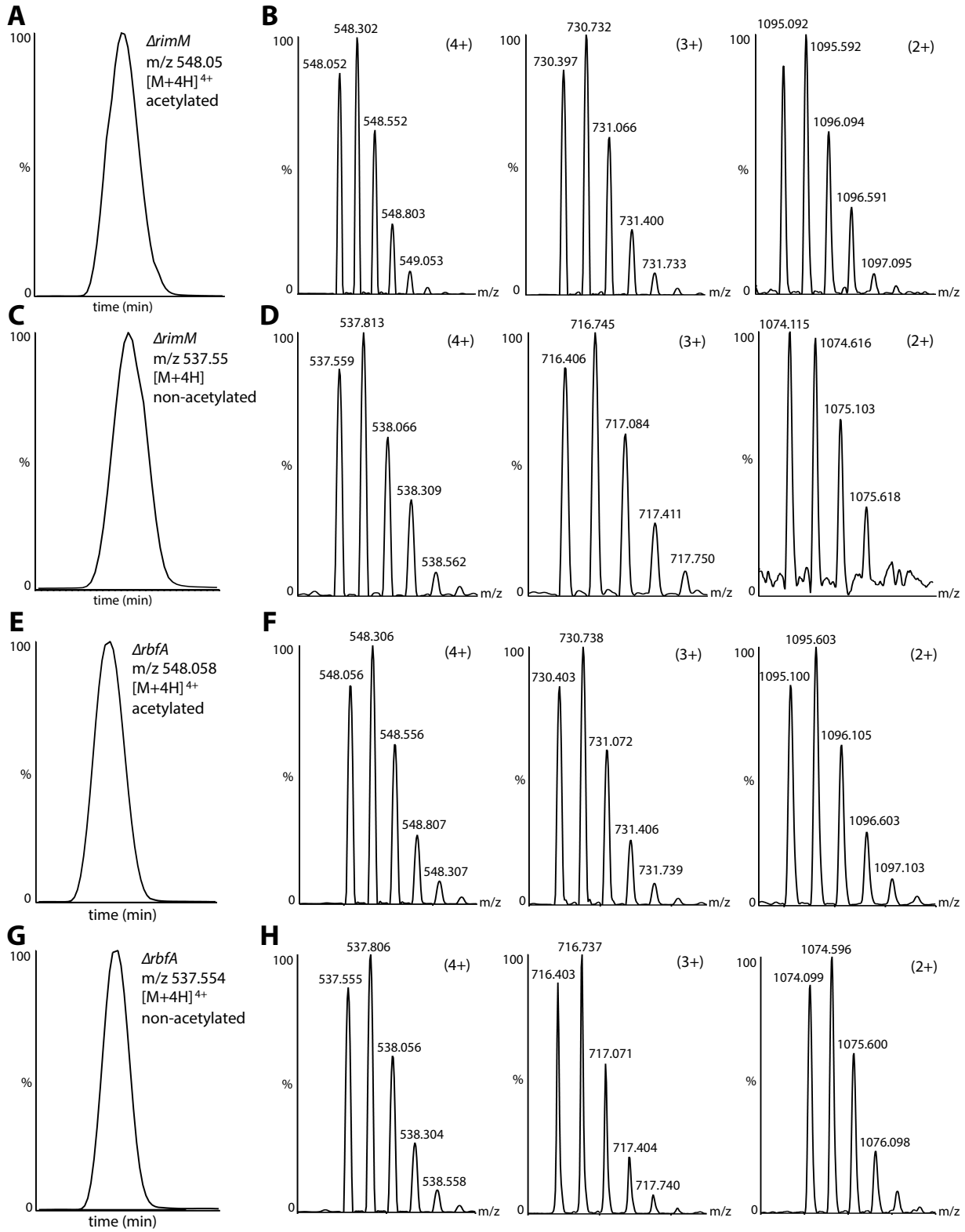
A
ΔrimM
m/z 548.05
[M+4H]4+
acetylated
time (min)
B
548.302
548.052
548.552
548.803
549.053
(4+)
730.732
730.397
731.066
731.400
731.733
(3+)
1095.092
1095.592
1096.094
1096.591
1097.095
(2+)
m/z
C
ΔrimM
m/z 537.55
[M+4H]
non-acetylated
time (min)
D
537.813
537.559
538.066
538.309
538.562
(4+)
716.745
716.406
717.084
717.411
717.750
(3+)
1074.115
1074.616
1075.103
1075.618
(2+)
m/z
E
ΔrbfA
m/z 548.058
[M+4H]4+
acetylated
time (min)
F
548.306
548.056
548.556
548.807
548.307
(4+)
730.738
730.403
731.072
731.406
731.739
(3+)
1095.603
1095.100
1096.105
1096.603
1097.103
(2+)
m/z
G
ΔrbfA
m/z 537.554
[M+4H]4+
non-acetylated
time (min)
H
537.806
537.555
538.056
538.304
538.558
(4+)
716.737
716.403
717.071
717.404
717.740
(3+)
1074.596
1074.099
1075.600
1076.098
(2+)
m/z

Figure S4, related to Figure 4. Acetylation of r-protein S5 in the *ΔrimM* and *ΔrbfA* strains

(A) Extracted ion chromatograms (XIC) of the acetylated peptide fragment (*m/z* 548) from *ΔrimM* TP30.

(B) Mass spectrum of the 4+, 3+, and 2+ charge states of the acetylated peptide from *ΔrimM* TP30.

(C) Extracted ion chromatograms (XIC) of the non-acetylated peptide fragment (*m/z* 537) from *ΔrimM* TP30.

(D) Mass spectrum of the 4+, 3+, and 2+ charge states of the non-acetylated peptide fragment from *ΔrimM* TP30.

(E) Extracted ion chromatograms (XIC) of the acetylated peptide fragment (*m/z* 548) from *ΔrbfA* TP30.

(F) Mass spectrum of the 4+, 3+, and 2+ charge states of the acetylated peptide from *ΔrbfA* TP30.

(G) Extracted ion chromatograms (XIC) of the non-acetylated peptide fragment (*m/z* 537) from *ΔrbfA* TP30.

(H) Mass spectrum of the 4+, 3+, and 2+ charge states of the non-acetylated peptide from *ΔrbfA* TP30.

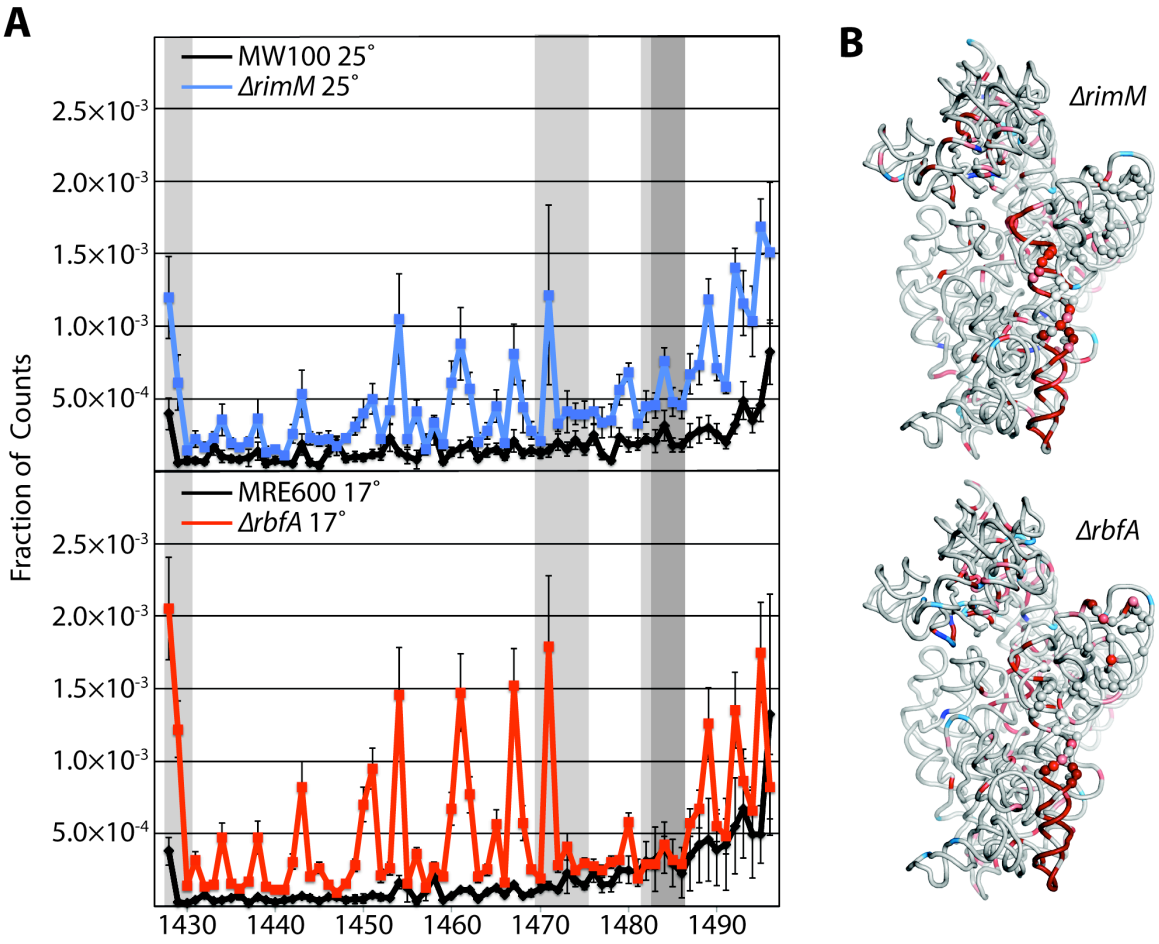

Figure S5, related to Figure 5. Helix 44 is undocked in pre-30S ribosomes

(A) Results of *in vivo* footprinting of helix 44. Integrate cleavage reactivity per residue show as mean ± S.D. of three irradiated samples. Gray bars depict areas protected by the 50S subunit (Merryman and Noller, 1999); strong protections are in dark gray, while medium protections are shown in light gray. Most of the change in exposure of helix 44 cannot be explained by an increase in free subunits.

(B) Three-dimensional mature 16S structure (PDB *2I2P*) colored as in Figure 2 to show differences in pre-30S and mature 30S ribosomes. Nucleotides with strong or medium protection by the 50S subunit (Merryman and Noller, 1999) are shown as spheres.

Table S1. Oligonucleotides used in this study.

Name	Sequence, 5'-3'
46	TCGACTTGCATGTGTTAGGC
161	GCGGTATTAGCTACCGT
323	AGTCTGGACCGTGTCTC
540	TTCCGATTAACGCTTGCACCC
812	AACCTCCAAGTCGACATCGTTTAC
1046	GACAGCCATGCAGCACC
1257	GCTCTCGCGAGGTCGCT
1486	GGTTACCTTGTTACGACTTCACCCC
1508	CCAACCGCAGGAACCCCTACGG
17S-5p-Affinity	Biotin-CTAGAGAGACTTGGTATTCATTTTTCGTCTTGCGACG
17S-Elution	CGTCGCAAGACGAAAAATGAATACCAAGTCTCTCTAG
C16S3 (2'-O-methyl RNA)	CCdCdGdAdAGGUUAAGCUACCU

Oligonucleotide primers used for primer extensions are named for the 16S nucleotide that anneals to the 3' end of the primer. Related to Figure 1.

Table S2. Quantification of S5 acetylation from various sources.

S5 Source	% non-acetylated	% acetylated
Recombinant S5	96	4
K12 Polysome	1	99
K12 70S	1	99
K12 30S	2	98
ΔrimM pre-30S	12	88
ΔrbfA pre-30S	34	66

Recombinant S5 was overexpressed in *E. coli* cells and purified before analysis. K12 proteins were isolated from sucrose gradient fractions as indicated. Pre-30 proteins were obtained from oligonucleotide affinity-purified complexes. Related to Figure 4.

Extended Experimental Procedures

Northern blots

Glyoxal-treated total RNA (~2 μg) was electrophoresed on a 1.4% 1:1 SeaKem LE:NuSieve GTG Agarose gel (Lonza) in 10 mM sodium phosphate pH 7.0. After transfer to a Nytran SPC membrane, the blot was hybridized overnight with 60 pmol ^{32}P-labeled primer 323 (Table S1) at 42˚C. The washed blot was exposed to a Phosphorimager screen, and band volumes were quantified with ImageQuant (Molecular Dynamics).

Primer extension and Data Analysis

Primer extension of 16S or 17S rRNA from irradiated cells was performed using 6 μL 250 ng/μL total *E. coli* RNA and 2 μL 10 mM Beckman D4-labeled primer (Table S1). Reverse transcription was carried out with SuperScript III (Invitrogen) as per the manufacturer's instructions, but using only 12.5 U per reaction. After ethanol precipitation, cDNA pellets were resuspended in 40 μL Sample Loading Solution (Beckman-Coulter) and 0.5 μL Beckman D1-labeled Size Standard 600 (Beckman-Coulter) was added. The cDNAs were separated on a CEQ 8000 (Beckman-Coulter) as described in (Mitra et al., 2008).

Raw CEQ traces were processed with MatLab scripts received as a kind gift from Alain Laederach. First, all traces from each primer were aligned using gui_profile_align2.m. Then, two sequencing ladders, an unexposed control, and an experimental trace were packaged for ShapeFinder using Assemble_CEQ. The assembled traces were processed with ShapeFinder (Vasa et al., 2008) to determine peak areas for each 16S nucleotide. Peak

assignments were manually adjusted where needed. Peak areas from ShapeFinder were analyzed without further normalization.

The results from three independent samples were averaged before calculating ratios between strains. Standard error varied by strain in the range of ±15-20%. Because the average peak areas from different samples are similar (Figure S1B-D), systematic error from reverse transcriptase pauses or sequence compression is reduced when two strains are compared. Ratios of peak areas (cleavage intensities) were binned, and nucleotides with ratios between 0.4 and 1.8 were not considered further, to minimize contributions of experimental noise and strain differences to our interpretation (grey, Figure S2). Based on histograms of cleavage ratios, ΔrbfA results were closer to footprinting data from MRE600 than from MW100, while the MW37 data were closer to the MW100 data, confirming that our chosen comparisons minimize rather than exaggerate differences between pre-30S and 30S ribosomes.

For technical reasons helix 44 was analyzed by sequencing PAGE, as previously described (Mayerle et al., 2011), using primer 1508. After calculating ratios between strains these data were treated identically to capillary electrophoresis data.

Affinity purification of pre-30S complexes

Total ribosomes were purified from deletion strains grown at non-permissive temperatures, as previously described (Nierhaus, and Dohme, 1979), but with the addition of 20 mM vanadyl complex to buffers to inhibit trimming of immature rRNAs. Ribosomes were split in two changes of Buffer E (50 mM Tris-HCl pH 7.5, 150 mM NH_4Cl, 1 mM $MgCl_2$, 6 mM β-mercaptoethanol) before purification of pre-30S complexes.

Affinity purification of pre-30S complexes was adapted from (Schnapp et al., 1998). Yeast tRNA (50 μg/mL) and 17S-5p-Affinity oligonucleotide (Table S1) (10 nmol/mL) were added to split 70S ribosomes and incubated at 30˚C for 15 minutes. This mixture was added to Ultralink-immobilized Neutravidin Plus beads (Pierce) that had been blocked twice with Buffer E containing 0.5 mg/ml UltraPure™ BSA (NEB), washed once with Buffer E, then incubated for 10 min at 30˚C. After rocking overnight at 4˚C, samples were centrifuged for 5 min at 1000 r.p.m. at 4˚C and the supernatant was removed. Beads were washed four times at 4˚C with Buffer E + 0.01% Nikkol, then two times at 25˚C with Buffer E. Pre-30S complexes were eluted by a 2.5-fold excess of 17S-Elution displacement oligonucleotide (Table S1) in buffer E containing 0.01% Nikkol for 30-min at room temperature, followed by centrifugation for 5 min at 1000 r.p.m., retaining the supernatant. This elution was repeated twice more, then once after increasing the Nikkol in the elution buffer to a concentration of 0.02%. Eluted complexes were concentrated using Ultracel-100K centrifugal filter units (Amicon) for final purification and verification on a 10-40% preparative sucrose gradient (buffer E).

Mass spectrometry of mature 30S and pre-30S complexes

^{15}N-labeling and data-dependent LC-MS/MS were used to determine the relative amounts of proteins associated with mature 30S and pre-30S complexes (Sykes et al., 2010; Gouw et al., 2011). Proteins from purified complexes were extracted by acetone as previously described (Spedding, 1990). Equal amounts (1:1) of unlabeled TP30S proteins from K12, *ΔrimM*, or *ΔrbfA* strains were mixed with ^{15}N-labeled TP30S proteins (*E. coli* MRE600). The proteins were reduced and alkylated using dithiothreitol (DTT) (Sigma) and iodoacetamide

(IAA) (Sigma), respectively. The protein mixture was digested with sequencing grade trypsin (Promega) (1:30 enzyme: protein ratio) overnight at 37°C.

Tryptic digests were desalted using C18 ZipTips (Millipore) and analyzed by data-dependent LC-MS/MS on a Synapt G2 mass spectrometer coupled to a nanoAcquity UPLC (Waters, Milford, MA). Reversed-phase chromatography on an BEH130 C18 column (1.7 μm, 100 μm x 100 mm, Waters, Milford, MA) was performed at a flow rate of 1.0 μL min^{-1} using 0.1% formic acid in water as mobile phase A and 0.1% formic acid in acetonitrile as mobile phase B. Two (2 μL) microliters of tryptic digest were injected on column (250 nanograms total protein concentration). Gradient elution was carried out starting with 5% B for the first min and 5% B min^{-1} to 10% B for another minute. A linear gradient from 10% B to 50% B for 58 min was then employed followed by a constant 50% B for 10 min. Finally, a 4.5% B min^{-1} to 5% B for 10 min was carried out and the column was re-equilibrated at 5% B for another 10 min. Mass spectra were recorded in positive ion mode with a source temperature of 100 °C, spray voltage of 3.5 kV. All LC-MS analyses were performed in high-resolution mode with at least 40,000 mass resolving power (FWHM). All samples were analyzed in triplicate.

Database Searching and Data Analysis

For peptide identification and quantitation, raw mass spectral data from a data-dependent acquisition mode were uploaded to Mascot Distiller version 2.3.2.0 (www.matrixscience.com) (Perkins, D. N. et al., 1999). Raw mass spectral data were processed into .mgf files and searched against the UniProtKB/Swiss-Pro protein database.

Searches were limited to *Escherichia coli* and Trypsin/P was selected as the protease. Search parameters used were monoisotopic masses with peptide mass tolerance of 0.6 Da and fragment mass tolerance of 0.3 Da. One (1) internal missed cleavage was allowed with cysteine residues being carbamidomethylated (+57.0215 Da) and methionine being oxidized (+15.9949 Da) as dynamic modifications. Quantification was set to "15N Metabolic [MD]". A composite decoy search was performed and specific cut-off scores were determined allowing <1% false discovery rate.

For peptide quantification, the threshold for the extracted ion chromatogram (XIC) was set to 0.1 Da- 0.2 Da. The minimum signal to noise (S/N) was 0.1 and the minimum and maximum *m/z* values were set from 50 to 100,000 respectively. The correlation threshold for quantification was set to 0.7- 0.9. A minimum of two light-heavy peptide pairs were used for quantification of the proteins, while manual inspection of the data was done for proteins with only one light-heavy peptide pair to assess data quality. Quantitative analysis was performed in triplicate and the L/H ratios were averaged and standard deviation of these multiple measurements were reported as error bars.

MALDI-TOFMS of intact proteins

All MALDI-TOF (Matrix-assisted laser desorption/ionization-time-of-flight) mass spectrometry experiments were performed on a MDS SCIEX 4800 MALDI TOF/TOF analyzer equipped with Nd:YAG 200 Hz laser (Applied Biosystems, Framingham, MA). Total 30S proteins (TP30) from wild type K12, *ΔrimM*, or *ΔrbfA* were acidified with trifluoroacetic acid (TFA) and mixed with equal volume of saturated sinapinic acid (Fluka, Milwaukee, WI) as matrix in 33% aqueous acetonitrile with 0.1% trifluoroacetic acid (Suh,

and Limbach, 2004). One (1) microliter of the mixture was spotted onto a MALDI target plate and analyzed in linear mode. Mass spectra were collected in positive ionization mode. The MALDI TOF mass spectrometer was calibrated using monoisotopic peaks of ubiquitin $[M+H]^+$ 8565.9, cytochrome C $[M+H]^+$ 12361.0, and myoglobin $[M+H]^+$ 16952.3. The spectra were then internally calibrated using well-resolved peaks of small subunit proteins of S15, S16, and S20.

LC-MS^E Analysis of Protein S5 Arg-C Digestion Products

Endoproteinase Arg-C (Sigma, St. Louis, MO) digestion of protein S5 was performed to determine the modification status of the protein as described elsewhere (Porras-Yakushi et al., 2007). Total 30S proteins from wild type K12, *ΔrimM* and *ΔrbfA* were digested with Arg-C and the digestion products were analyzed by alternate scanning mode, LC-MS^E (Silva et al., 2006). Peptide ions corresponding to the acetylated peptide (*m/z* 548, $[M+4H]^{4+}$) and non-acetylated peptide (*m/z* 537 $[M+4H]^{4+}$) were extracted from total ion chromatograms (TIC) and ion abundances were obtained by integrating their corresponding peak areas. Manual evaluation of collision-induced dissociation (CID) spectra of the precursor ions and database searching on PLGS (ProteinLynx Global Server, Version 2.5.2, Waters, Milford, MA) were performed to verify identity and modification status of protein S5.

Supplementary References

Brenowitz, M., Senear, D. F., Shea, M. A., Ackers, G. K. (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions. Methods Enzymol. *130*, 132-181.

Gouw, J.W., Tops, B.B., and Krijgsveld, J. (2011). Metabolic labeling of model organisms using heavy nitrogen (15N). Methods Mol. Biol. *753*, 29-42.

Mayerle, M., Bellur, D.L., and Woodson, S.A. (2011). Slow formation of stable complexes during coincubation of minimal rRNA and ribosomal protein S4. J. Mol. Biol. *412,* 453-465.

Mitra, S., Shcherbakova, I.V., Altman, R.B., Brenowitz, M., and Laederach, A. (2008). High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. Nucleic Acids Res. *36,* e63.

Nierhaus, K.H., and Dohme, F. (1979). Total reconstitution of 50 S subunits from Escherichia coli ribosomes. Methods Enzymol. *59*, 443-449.

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567.

Porras-Yakushi, T.R., Whitelegge, J.P., and Clarke, S. (2007). Yeast ribosomal/cytochrome c SET domain methyltransferase subfamily: identification of Rpl23ab methylation sites and recognition motifs. J. Biol. Chem. *282,* 12368-12376.

Schnapp, G., Rodi, H.P., Rettig, W.J., Schnapp, A., and Damm, K. (1998). One-step affinity purification protocol for human telomerase. Nucleic Acids Res. *26,* 3311-3313.

Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P., and Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol. Cell. Proteomics *5,* 144-156.

Spedding, G. (1990). Ribosomes and protein synthesis : a practical approach (Oxford ;New York: Oxford University Press).

Suh, M.J., and Limbach, P.A. (2004). Investigation of methods suitable for the matrix-assisted laser desorption/ionization mass spectrometric analysis of proteins from ribonucleoprotein complexes. Eur. J. Mass. Spectrom. (Chichester, Eng) *10,* 89-99.

Sykes, M. T., Shajani, Z., Sperling, E., Beck, A. H., & Williamson, J. R. (2010). Quantitative proteomic analysis of ribosome assembly and turnover in vivo. J. Mol. Biol., *403*, 331–345.

Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. (2008). ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA *14,* 1979-1990.