

# Assessment of the nucleotide modifications in the high-resolution cryo-electron microscopy structure of the *Escherichia coli* 50S subunit

Vanja Stojković<sup>1</sup>, Alexander G. Myasnikov<sup>2</sup>, Iris D. Young<sup>3</sup>, Adam Frost<sup>2,4,\*</sup>, James S. Fraser<sup>3,4,\*</sup> and Danica Galonić Fujimori<sup>1,4,5,\*</sup>

<sup>1</sup>Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA 94158, USA, <sup>2</sup>Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA 94158, USA, <sup>3</sup>Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA, <sup>4</sup>Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA 94158, USA and <sup>5</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16th St, MC2280 San Francisco, CA 94158, USA

Received July 05, 2019; Revised January 09, 2020; Editorial Decision January 10, 2020; Accepted January 14, 2020

## ABSTRACT

Post-transcriptional ribosomal RNA (rRNA) modifications are present in all organisms, but their exact functional roles and positions are yet to be fully characterized. Modified nucleotides have been implicated in the stabilization of RNA structure and regulation of ribosome biogenesis and protein synthesis. In some instances, rRNA modifications can confer antibiotic resistance. High-resolution ribosome structures are thus necessary for precise determination of modified nucleotides' positions, a task that has previously been accomplished by X-ray crystallography. Here, we present a cryo-electron microscopy (cryo-EM) structure of the *Escherichia coli* 50S subunit at an average resolution of 2.2 Å as an additional approach for mapping modification sites. Our structure confirms known modifications present in 23S rRNA and additionally allows for localization of Mg<sup>2+</sup> ions and their coordinated water molecules. Using our cryo-EM structure as a testbed, we developed a program for assessment of cryo-EM map quality. This program can be easily used on any RNA-containing cryo-EM structure, and an associated Coot plugin allows for visualization of validated modifications, making it highly accessible.

## INTRODUCTION

Ribosomes are complex cellular machines responsible for protein synthesis. Accurate translation by the ribosome re-

quires specific post-transcriptional and post-translational modifications of ribosomal RNA (rRNA) and ribosomal proteins, respectively. The roles of individual rRNA modifications are still poorly understood, and while some modulate the function of the ribosome, others are important in RNA folding and stability (1–4). The majority of rRNA modifications are located in functional regions of the ribosome: the decoding center, the peptidyl transferase center (PTC), the exit tunnel or the interface between ribosomal subunits. In *Escherichia coli*, not a single rRNA modification is essential for the survival of the cell. However, the presence of individual modifications confers advantages under certain growth or stress conditions (3). For instance, modifications of rRNA have emerged as one of the most clinically relevant mechanisms of resistance to ribosome-targeting antibiotics (5–8). Differential expression and post-translational modifications of ribosomal proteins, as well as differences in rRNA modifications, can lead to heterogeneity in ribosome composition which can result in 'specialized ribosomes'. These ribosomes can have a substantial impact on the relative abundances of proteins being produced and consequently influence an organism's adaptability to a variety of environmental factors (9–11). For example, in bacteria, ribosome heterogeneity has been identified as a mechanism for stress adaptation (12).

Visualizing rRNA modifications is important for understanding their impact on the translating ribosome, and can aid in the development of novel antibiotics. Recently, rRNA and ribosomal protein modifications have been visualized in several high-resolution X-ray and cryo-electron microscopy (cryo-EM) structures of bacterial ribosomes (2.3 Å resolution) (13), including *E. coli* ribosome (2.4 Å) (14), parasitic

\*To whom correspondence should be addressed. Tel: +1 415 514 0147; Email: danica.fujimori@ucsf.edu  
Correspondence may also be addressed to James Fraser. Email: jfraser@fraserlab.com  
Correspondence may also be addressed to Adam Frost. Email: adam.frost@ucsf.edu

(2.8 Å – cryo-EM) (15) and human ribosomes (2.8 Å – cryo-EM) (16,17). In certain instances, assignments of nucleotide modifications have been subsequently revised in the light of improved data (18), highlighting the challenges in assigning nucleotide modifications in near-atomic resolution structures.

Here, we report a high-resolution cryo-EM structure of the *E. coli* 50S subunit. Motivated by the high resolution of this map and previous claims in the literature about the ability to identify post-transcriptional modifications directly from the cryo-EM maps (17), we developed a program, qPTxM (quantifying Post-Transcriptional Modifications), to validate the presence of assigned rRNA modifications. These metrics provide a sensitive readout of map quality by probing the local density around all possible modifications of each nucleotide. Modifications that are well-supported by the map can be easily examined in Coot using a custom plugin (19). The performance of this program in confirming biochemically-validated modifications provides orthogonal evidence of the superior quality of this map relative to previous EM maps. Additionally, we corroborated the assigned nucleotide conformations using the X3DNA software system (20). Overall, these analyses confirm the quality of this new high-resolution cryo-EM map of the *E. coli* 50S ribosome and suggest that even higher quality maps are required before post-transcriptional modifications can be confidently identified *de novo* from cryo-EM data alone.

## MATERIALS AND METHODS

### *E. coli* 50S ribosomal subunit purification

50S ribosomal subunit was purified from *E. coli* MRE600 strain using modified version of previously published protocol (21). In short, cells were grown to an OD<sub>600</sub> of 0.5 in LB media at 37°C and 220 rpm. Cells were pelleted by centrifugation, washed with wash buffer (20 mM Tris pH 7.5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>) and stored at –80°C. The pellet corresponding to 750 ml of growth was resuspended in Buffer A (20 mM Tris pH 7.5, 300 mM NH<sub>4</sub>Cl, 10 mM MgCl<sub>2</sub>, 0.5 mM EDTA, 6 mM β-mercaptoethanol and 10 U/ml of SuperASE-In (Ambion)). PMSF was added to a final concentration of 0.1 mM. The cell resuspension was lysed using microfluidizer at 10 000 psi with three passages. The lysate was centrifuged at 16 100 rpm (Beckman Ti45 rotor) for 30 min, at 4°C, to remove cell debris. Without disturbing the pellet, supernatant was removed and centrifuged for additional 30 min. The resulting supernatant was loaded onto a 32% sucrose cushion prepared in Buffer A and centrifuged at 28 000 rpm (Beckman SW41Ti rotor) for 16 h, at 4°C, to obtain the crude 50S pellet. This pellet was slowly resuspended at 4°C in Buffer A to homogeneity. Any presence of non-resuspended particles was removed by a short centrifugation at 10 000 rpm, for 10 min at 4°C.

Crude 50S ribosomal subunits were further purified on 10–40% sucrose gradients. Sucrose gradients were prepared in Buffer A using BioComp Gradient Master. Approximately 650 pmol of crude 50S subunits were loaded on each gradient and centrifuged at 28 000 rpm (Beckman SW41Ti rotor) for 16 h, at 4°C. Fractions were collected from top to bottom using BioComp Piston Gradient Fractionator.

The sample absorbance was recorded using UV reader and the peak corresponding to 50S was pooled for precipitation with PEG 20 000. A final concentration of 10.5% PEG 20 000 was slowly added to the pooled fractions, incubated on ice for 10 min and centrifuged at 10 000 rpm, for 10 min at 4°C. The pure 50S pellet was dissolved in Buffer A and filtered using 0.22 μm low-binding Durapore PVDF filters (Millipore).

### Cryo-EM analysis

Purified 50S ribosomal subunits were diluted from 2 to 0.5 mg/ml in Buffer A, applied to 300-mesh carbon coated (2nm thickness) holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools) and flash-frozen as described in (22). Data were collected on the in-house Titan Krios X-FEG instrument (Thermo Fisher Scientific) operating at an acceleration voltage of 300 kV and a nominal underfocus of  $\Delta z = 0.2$  to 1.5 μm at a magnification of 29 000 (nominal pixel size of 0.822 Å). We recorded 1889 movies using a K2 direct electron detector camera in super-resolution mode with dose fractionation (80 individual frames were collected, starting from the first one). Total exposure time was 8 s, with the total dose of 80 e<sup>–</sup> (or 1 e<sup>–</sup>/Å<sup>2</sup>/frame). Images in the stack were aligned using the whole-image motion correction and patch motion correction (5 × 5 patches) methods in MotionCor2 (23). Before image processing, all micrographs were checked for quality and 1610 best were selected for the next step of image processing. The contrast transfer function of each image was determined using GCTF (24) as a standalone program. For particle selection we have used Relion3.0 autopicking procedure which picked 193 000 particles (25). For the first steps of image processing we used data binned by a factor of 8 (C8 images). During the first round of 2D classification we removed only images with ice or other contaminants (162 746 particles left). Subsequently, the initial structure was generated using the *ab initio* procedure in CryoSPARC v2.0. Following this step, we performed Relion 3D classification with bin by four data (C4) in order to exclude bad particles (21 197 non-ribosomal particle images). The resulting 141 549 particle images of ribosomes were used for subsequent classification and refinement procedures. For the initial refinement we used spherical mask, which was followed by further refinement using mask around stable part of 50S (excluding L1 stalk, L7/L12 region). A further improved cryo-EM map was obtained by using CTF-refinement procedure from Relion 3.0. The Post-processing procedure implemented in Relion 3.0 (25) was applied to the final maps with appropriate masking, B-factor sharpening (automatic B-factor estimation was –55.85) and resolution estimation to avoid over-fitting (final resolution after post-processing with 50S mask applied was 2.68 Å). Subsequently the stack of CTF-refined particles was processed in a new version of CryoSPARC v2.0 (26). After homogeneous refinement resolution was improved to 2.5 Å. The same stack of particles was additionally refined in cisTEM (27). After Auto-Refine (with automasking within cisTEM) we attained 2.4 Å resolution. Then we performed local refinement using 50S mask (the same one used for refinement in Relion) and also ap-

**Table 1.** Cryo-EM data collection, refinement and validation statistics

<b>Data collection and processing</b>	
Electron microscope	Krios
Magnification	29 000
Number of micrographs	2688
Number of particles in the map	300 000
Number of particles after classification	144 000
Pixel size (Å)	0.822
Defocus range (µm)	−0.2 to −1.5
Voltage (kV)	300
Electron dose (e-/Å <sup>2</sup> )	80
<b>Map refinement</b>	
Model resolution (Å)	2.2
FSC threshold	0.143
Model resolution range (Å)	2.2–20
Map sharpening <i>B</i> -factor (Å <sup>2</sup> )	−70
<b>Refinement and model statistics</b>	
Clashscore, all atoms	4.98
<u>Protein geometry</u>	
MolProbity score	1.87
Rotamer outliers (%)	2.41
Cβ deviations >0.25 Å (%)	0.5
Ramachandran (%)	
- Favored	95.26
- Allowed	4.68
- Outliers	0.07
Deviations from ideal geometry	
- Bonds (%)	0.01
- Angles (%)	0.18
<u>Nucleic acid geometry</u>	
Probably wrong sugar puckers (%)	0.67
Bad backbone conformations (%)	14.15
Bad bonds (%)	0.05
Bad angles (%)	0.02

plied per particle CTF refinement as implemented in cis-TEM software. After such refinement the resolution was improved to 2.2 Å (Figure 1). This map after Sharpen3D (27) was used for model building and map interpretation.

### Atomic model building and refinement

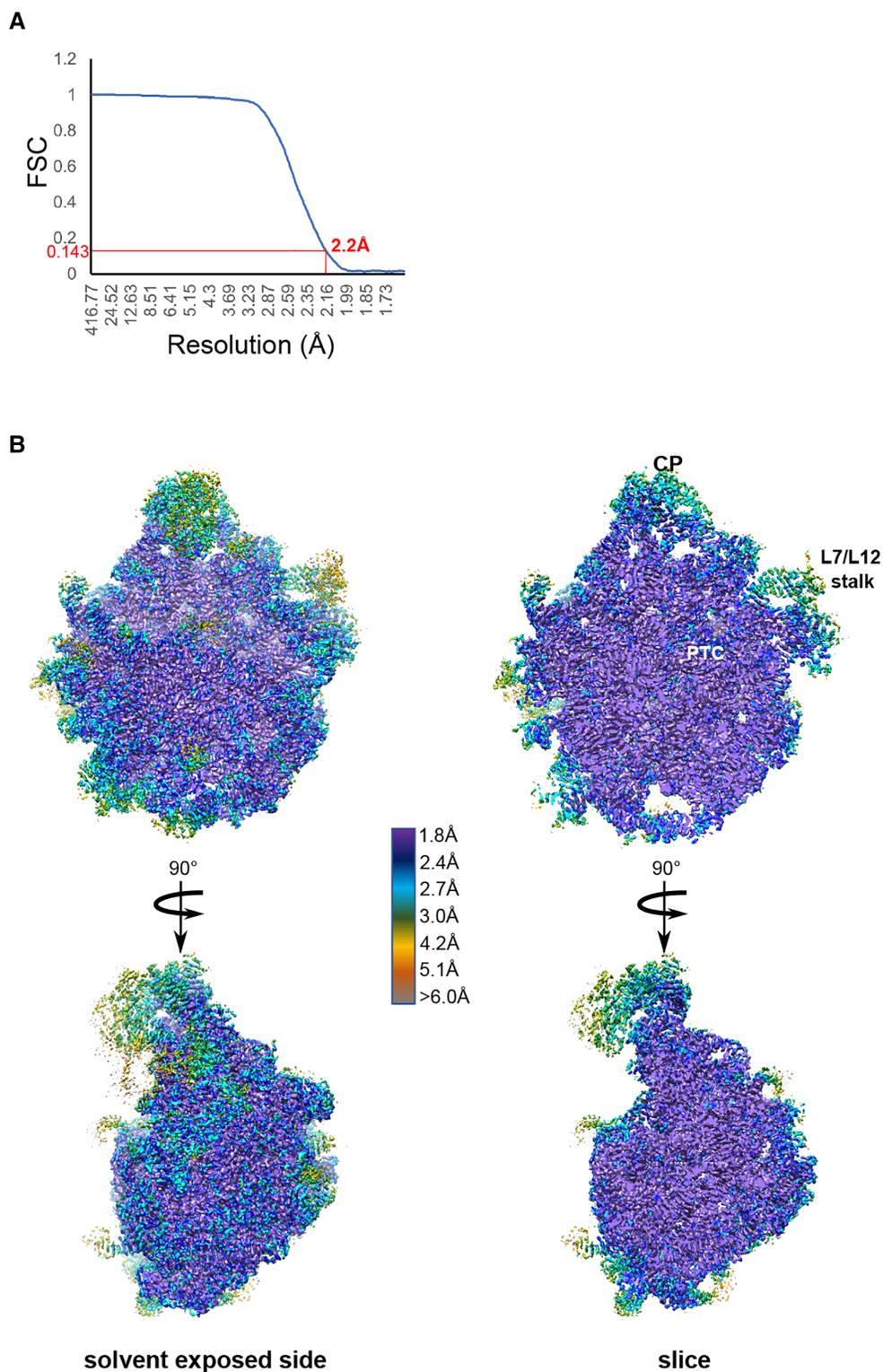
The final model of the 50S subunit was generated by multiple rounds of model building in Coot (19) and subsequent refinement in PHENIX (28). Modified nucleotides were drawn and the restraints for the atomic model fitting and refinements were generated using JLigand (29). The atomic model of the 50S subunit from the *E. coli* ribosome structure (PDB 4YBB) (14) was used as a starting point and refined against the experimental cryo-EM map by iterative manual model building and restrained parameter-refinement protocol (real-space refinement, positional refinement, and simulated annealing). Final atomic model comprised ~154 836 atoms (excluding hydrogens) across the 3016 nucleotides and 3356 amino acids of 29 ribosomal proteins. Proteins L3, L10 and L31 were not modelled in. In addition, 185 Mg<sup>2+</sup>, 3897 water molecules, one Zn<sup>2+</sup> and one Na<sup>+</sup> were included in the final model. Prior to running MolProbity (30) analysis, nucleotides 878–898, 1053–1107, 2099–2188 of 23S rRNA, and ribosomal proteins L9 and L11 were removed, due to their high degree of disorder. Overall, protein residues and nucleotides show well-refined geometrical parameters (Table 1). Figures were prepared using UCSF Chimera 1.13 (31).

### qPTxM program for quantitative identification of post-transcriptional modifications: development and analysis

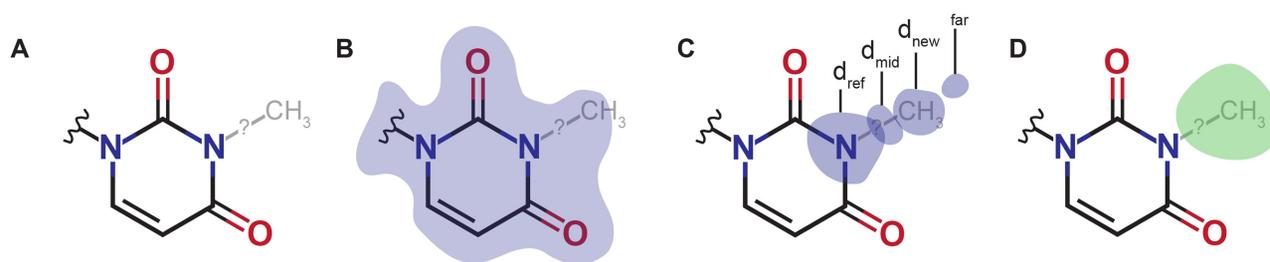
qPTxM was developed for validation and detection of post-transcriptional modifications based on evidence from an X-ray or EM map. It requires a map, an atomic model refined into it, and the estimated average resolution of the map. First, qPTxM reads the model and constructs a copy containing only the major conformer of each residue or nucleotide, removes all hydrogens, and removes any post-transcriptional modifications already modeled. The resulting model is used to generate a calculated map. Using the calculated and experimental maps, a difference map is generated, where positive difference density features will indicate sites of possible unmodeled atoms. qPTxM then examines each nucleotide and corresponding density in the experimental and difference maps for possible modifications—this excludes pseudouridines, which differ from uridines only in the attachment of the base to the ribose (Figure 2, Supplementary Figure S1). First, the correlation coefficient between the calculated and experimental maps is calculated using 8-point interpolation at the position of each atom in the unmodified nucleotide. If the correlation coefficient is worse than 0.6, the quality of the density is considered insufficient to support the presence of modifications, and no modifications will be suggested at this nucleotide. The default threshold of 0.6 is a lower bound that may be adjusted upward for higher confidence and lower sensitivity where high resolution data permit more stringent criteria.

Second, densities at four groups of positions in the experimental map are calculated and compared. For each modification we define one or more reference atoms (ref), one or more new atoms (new), positions midway between the reference and the new atoms (mid), and positions half a covalent bond beyond the new atoms along the direction of the bond (far). We obtain the experimental map densities at each position by 8-point interpolation and calculate the average density within each of these categories. Ratios of these densities  $d_{ref}$ ,  $d_{mid}$ ,  $d_{new}$  and  $d_{far}$  are used as indications of how closely the shape of the density around the modification matches our expectation, and we specifically select three comparisons that help eliminate a large number of cases where features in the map at the proposed modification position can be explained by water molecules, ions, and protrusions from other parts of the protein or nucleic acid model. We have found that requiring  $d_{far} \leq 1.5d_{new}$ ,  $d_{far} \leq d_{mid}$  and  $d_{ref} \leq 3d_{new}$  filters out most of these cases. These multipliers may also be adjusted with optional arguments to qPTxM.

Third, the collection of possible modifications is sorted by  $d_{ref}$  and  $d_{new,diff}$ , where the latter quantity is the density measured by 8-point interpolation at the new atom positions in the difference map, and the modifications in the lower half of  $d_{ref}$  and the lower 90% of  $d_{new,diff}$  densities are excluded from consideration. This step improves selectivity for nucleotides that have strong densities and strong evidence for additional unmodeled atoms. At all stages, proposed modifications that are marked as rejected are not removed from memory, so that all filters operate on the full set of proposed modifications. The full set is written to



**Figure 1.** Cryo-EM data collection and processing. (A) Fourier shell correlation (FSC) curve indicating the overall resolution of the *E. coli* 50S subunit using the gold-standard criterion (FSC = 0.143). (B) 3D cryo-EM map of the 50S *E. coli* subunit colored according to local resolution. Local resolution map was calculated with ResMap. Left, solvent exposed side view; right, cut-through view. Features in the 50S subunit include the central protuberance (CP), peptidyl-transferase center (PTC), L7–L12 region (L7/L12 stalk).



**Figure 2.** Methodology for assignment of post-transcriptional modifications (PTxMs). (A) For each possible modification, the new atoms are modeled with ideal geometry on a copy of the model. (B) Correlation of the model with the map is used to remove modifications for which the model is already a poor fit in the density. (C) Ratios of densities at several locations on the modified residue are used to assess whether the density profile along the vector of each new covalent bond matches expectations. By default,  $d_{\text{far}}$  may be no more than twice  $d_{\text{new}}$ ,  $d_{\text{far}}$  may not be more than  $d_{\text{mid}}$ , and  $d_{\text{ref}}$  may be no more than three times  $d_{\text{new}}$ . (D) A difference map is calculated between the experimental map and a map calculated from the (unmodified) model, and only the modifications with the strongest experimental and difference map densities at the new atom positions (by default, the 50% strongest by each measure) are considered. An illustration of these parameters is shown for N3-methyluridine.

all\_tested\_ptms.out to facilitate rapid filter-only runs of qP-TxM, while ptms.out contains all accepted modifications.

Finally, modifications are scored by taking the ratio  $d_{\text{new}}/d_{\text{ref}}$  and inflating it by a factor between 1.2 for atoms close to the aromatic ring (in modifications such as 7-methylguanosine ( $m^7G$ )) and 3 for atoms distant from the aromatic ring (in modifications such as 2-methylguanosine ( $m^2G$ )). Thus, the score is normalized depending on how far the modified position is from the aromatic ring and whether the reference atoms are part of the aromatic ring or protrude from it. This scaling factor helps account for the steeper dropoff in density at positions further from other heavy atoms. Scores are compared against a default threshold of 0.5, which can again be adjusted with an optional argument.

Modifications passing all tests are built into a copy of the model for visualization, with the possibility of multiple modified residues at a single position, which is written to a file ptms.pdb. qPTxM also generates a plain text file ptms.out, a script goto\_ptms.py and pdfs of histograms of the correlation coefficients, scores and densities. It is possible to rerun just the filtering steps of qP-TxM to regenerate plots and output (.out) files for rapid fine-tuning of the optional parameters using the argument `adjust_filters_only=True`, since these analyses can be run on the ptms.out file alone without repeating each of the calculations. This can be useful for adjusting sensitivity until a reasonable number of modifications pass all tests. To generate the custom script and the copy of the model with modifications in place, qPTxM should be run again without this argument once the user has settled on final parameters. The command `'coot goto_ptms.py pruned.pdb ptms.pdb'` will load two copies of the model in the Coot model viewing program, where pruned.pdb contains no modifications and ptms.pdb contains those passing all tests (Supplementary Figure S2). The script goto\_ptms.py adds functionality to jump directly to each modification for visual inspection.

We used synthetic data to tune the behavior of qPTxM under ideal (noise-free) conditions. We are including the tools we used to generate the synthetic data in qPTxM for the benefit of other users and developers, accessed by supplying the command line argument `synthetic_data=True`. Beginning with a refined map and model, we strip off any modifications already present, any alternate conformers,

and all hydrogens, just as for a model to be tested. This model is then modified at random at 10% of all recognized nucleotides (i.e. all standard, unmodified nucleotides as well as any modified nucleotides the program recognizes how to strip modifications off of). Then, instead of using the experimental map supplied, the modified model is used to generate calculated structure factors (substituting for a noise-free experimental map) and the experimental map supplied is used only to inform the dimensions of the calculated map. Subsequently, qPTxM proceeds through all other steps as described above, and finishes by reporting on its accuracy. The list of (true) modified sites in the synthetic data is also written to a file synthetic\_ptms.out and can be examined independently.

We primarily intended detection of post-transcriptional modifications to be useful as a validation step when the modifications are already known by some other method. The performance of the program should improve with the quality of the map, which should track with its resemblance to a noise-free synthetic map. To this end, we trained a random forest classifier on features from synthetic data generated at a range of resolutions (2, 2.2, 2.4, 2.6, 2.8, 3.2, 3.8 and 5 Å) and a range of B-factors (10–90 in increments of 10) from a collection of 22 publicly available ribosome models after stripping any recognized modifications and then modifying 10% of nucleotides at random. For datasets 6EK0 and 6QZP, the deposited map and model did not match, so the map was adjusted to 0.85 Å voxel size from 0.9 Å and fit with Chimera to a map calculated from the model. Features were the same map densities, ratios and correlation coefficients used to filter proposed modification sites, the scores determined by qPTxM, and the resolution and B-factors used to generate the synthetic maps. The trained classifier has an out-of-bag score estimate of 99.5% and was able to identify detectable (non-pseudouridine) modifications in the present dataset with a mean accuracy score of 99.8%. The classifier can be used to predict a set of modifications by running a separate script, rf\_util.py, with python3 (a requirement for loading the classifier object). The output from this script can be fed back to qPTxM with the `selected_ptms` argument to produce the matching model for visualization if desired. We used the classifier to predict modifications for the same 22

ribosomes in addition to the current dataset and an additional group of synthetic datasets generated at 2.2 Å resolution and a B-factor of 10. Even with >99% accuracy, few true positives could be detected in most experimental maps. Both sensitivity and specificity were markedly better for the noise-free synthetic datasets compared with the experimental datasets (Supplementary Figure S3, Supplementary Tables S1 and S2), validating our hypothesis that unambiguously identifiable modifications in EM maps may be used as evidence of high map quality. The classifier and script to use it are available as part of the qPTxM github repository (<https://github.com/fraser-lab/qptm>). The data used to train the classifier and scripts reproducing these steps are also available there. Curious developers are highly encouraged to reproduce and improve upon our classifier and any other aspect of the implementation.

## RESULTS

Using a Titan Krios X-FEG instrument equipped with a K2 direct electron detector camera operated in super-resolution mode and performing reconstruction refinement in Relion3, CryoSPARC and cisTEM software, we obtained a structure of the *E. coli* 50S subunit at an average resolution of 2.2 Å (Table 1 and Figure 1). In order to avoid any model bias, the initial *ab initio* structure was generated in CryoSPARC and then used as a reference in Relion3. To achieve higher resolution, later steps of 3D refinement included only the well-ordered regions of the 50S subunit, and excluded the L1 stalk and L7/L12 region. An improved cryo-EM map was subsequently obtained using the CTF-refinement function of Relion3 and further improved using per-particle CTF refinement in cisTEM (see Methods). The resulting local resolution map is uniformly sampled, with a large portion of the map showing resolution of ~2 Å (Figure 1B). Highly dynamic peripheral regions of the 50S subunit, encompassing the L1 arm, the stalk proteins L10, L11 and L7/L12 region, ribosomal proteins L9 and L31, and sections of the GT-Pase center, however, are disordered. While proteins L1, L7, L10 and L31 are not built in our model, the amino-terminal domains of the proteins L9 and L11 are structurally well-defined and thus these two proteins are built into the final model. Conformations and complete models of proteins L9 and L11 were modeled based on the high-resolution X-ray crystal structure (14). Overall, our structure is the highest resolution cryo-EM ribosome structure to date, allowing us to unambiguously visualize many rRNA modifications, Mg<sup>2+</sup> ions and coordinated waters.

### Ribosomal modifications

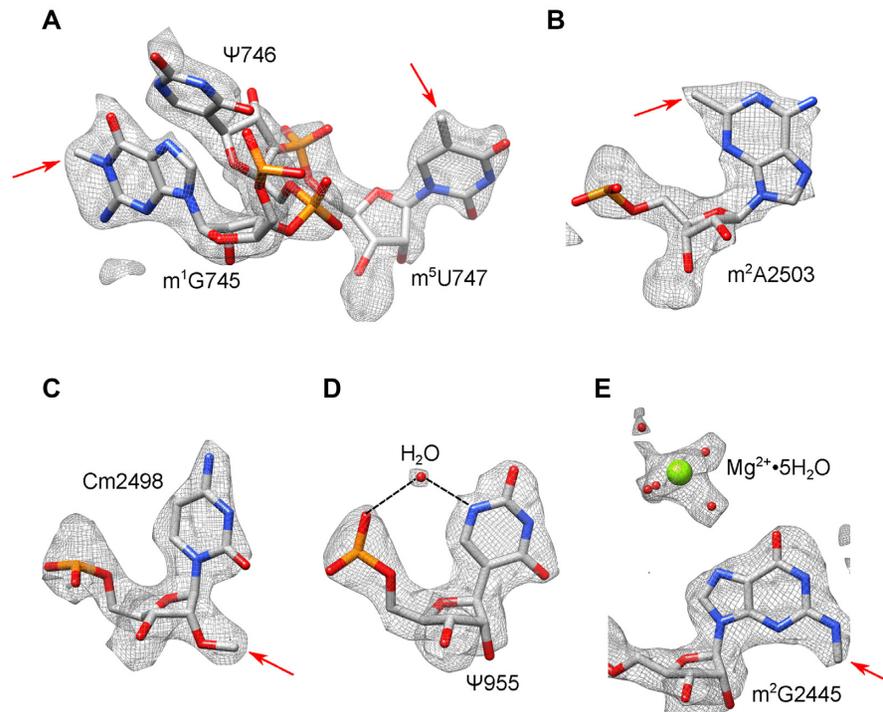
Most ribosomal modifications in the 50S subunit are clustered in the PTC region and the nascent peptide exit tunnel (NPET). Excluding pseudouridines (Ψ), we modeled 23 out of 25 known post-transcriptional modifications in the 23S rRNA, including the non-planar base of dihydrouridine (D) at position 2449 (Figure 3, Supplementary Figure S4 and Supplementary Table S3). The only modifications not observed are substoichiometric thionation of cytidine 2501 (s<sup>2</sup>C) and disordered methylated pseudouridine 1915 (m<sup>3</sup>Ψ1915). Nucleotide 1915 is located at the interface between the large and the small ribosomal subunits

and it has been proposed to strengthen the intersubunit contact (13). Seven out of ten known pseudouridines were well-ordered and thus are modelled in our structure (Supplementary Figure S4). Unlike uridines, pseudouridines contain an extra hydrogen-bond donor in the form of an N1 imino group, and water-mediated contact between the pseudouridine N1 imino group and the rRNA phosphate backbone is taken as an indication of pseudouridine presence. This water-mediated contact was observed for several pseudouridines (Supplementary Figure S4 and Supplementary Table S4).

### Validating map quality by detection of ribosomal modifications: qPTxM

The observation that post-transcriptional modifications were discernible in this map presented us with an opportunity to use these features to benchmark map quality. Others have noted (e.g. (17)) that the sites of modifications can potentially be discovered from the map directly. We reasoned that the success rate of recovering the correct modification positions could serve as a proxy for overall map quality. To test this hypothesis, we developed a new program, qPTxM (for quantifying post-transcriptional modifications). The premise of qPTxM is similar to EMRinger (32), which quantifies agreement between the protein backbone and EM density maps. In both programs, the part of the structure that can be most confidently built (the nucleobase or the protein backbone) is used to assess the strength of the EM density support at positions that are predicted, with ideal geometry, to be occupied by features that require high resolution data to model accurately (the nucleotide modification or the gamma atom of the protein side chain). These approaches leverage prior knowledge of ideal geometry to objectively quantify the strength of density at positions where a model might be extended, in this case identifying possible sites of post-transcriptional modification. qPTxM reports on true positives, false positives, false negatives, and true negatives. It can alternatively be used for modification discovery in very high resolution maps: qPTxM produces a ranked list of candidate sites with support for interactive viewing in Coot (19) (Supplementary Figure S2) (see Materials and Methods), but is subject to a high false-positive rate.

To test for modifications on an EM model and map, we examine the evidence for each possible modification on each nucleotide and apply a series of filters to progressively narrow the pool of candidates. We optimized these filters for noise-free, synthetic maps we generated for this purpose, achieving sensitivity >50% and specificity >99% in all synthetic test datasets (see Materials and Methods and Supplementary Figure S3). First, where the local map quality at a given nucleotide is insufficient for interpretation, all possible modifications at this nucleotide are rejected (3542 out of 11 712 possible modifications in the ribosome model and experimental EM map considered here). Second, we test whether there is density supporting a new atom at the appropriate position, and whether a covalent bond is more likely than a nearby water molecule or ion (7459 out of 8260 remaining modifications are rejected in this step). Third, we select for the most well-resolved reference atoms and the



**Figure 3.** Cryo-EM density map allows modeling of rRNA modifications of the *E. coli* 50S ribosomal subunit. (A–D) Cryo-EM density map of selected modified nucleotides present in 23S rRNA. The modifications shown are: (A) 1-methylguanosine ( $m^1G$ ) 745, pseudouridine ( $\Psi$ ) 746, 5-methyluridine ( $m^5U$ ) 747, (B) 2-methyladenosine ( $m^2A$ ) 2503, (C) 2'-O-methylcytidine (Cm) 2498, (D) pseudouridine ( $\Psi$ ) 955 and (E) 2-methylguanosine ( $m^2G$ ) 2445. (D, E) Coordinated waters and  $Mg^{2+}$  are also shown.

modification positions with the strongest difference density, rejecting another 600 of the remaining 801 modifications. Finally, we produce a score for each modification, and all 201 remaining modifications pass a final filter on score.

qPTxM was able to identify 8 of 13 modifications theoretically visible in the density. This represents a higher sensitivity than we observed for the other six ribosome cryoEM datasets that we tested, and even with 193 false positives, a comparable specificity to the other structures (Supplementary Figure S3). Although our 8 true positives include some of the highest scoring candidates (Supplementary Table S5), many of the other top candidates are residues that have never been detected as modified in *E. coli* 23S rRNA by sensitive mass spectrometry methods (33). This result suggests that although the map quality is high enough to confirm known modified residues and to guide their refinement, it is not yet sufficient to unambiguously determine the modification state of bases. Thus, a major caveat in automatic map interpretation by qPTxM is a high false positive rate, an attribute that is likely shared by a blinded, manual interpretation of the density map.

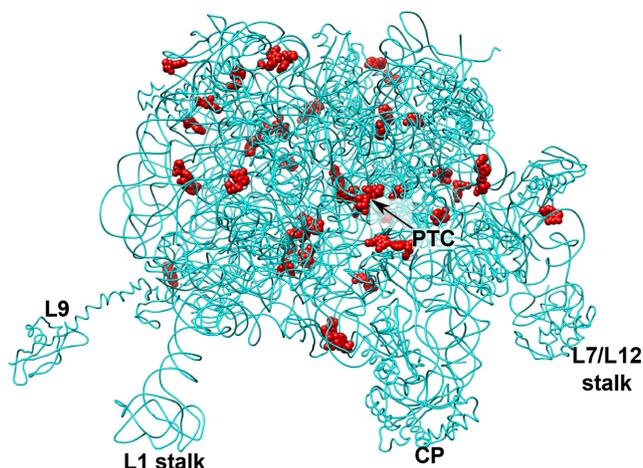
To attempt to address the high false positive rate, we trained a random forest classifier on a collection of synthetic data generated from publicly available structures and tested its performance on a separate group of synthetic maps (Supplementary Figure S3, Supplementary Table S2). Predictions by the classifier exhibited higher sensitivity (higher true positive rate) but lower specificity (higher false positive rate) than modifications identified by qPTxM with default parameters. Results were comparable by both methods (and

more widely variable generally) for the experimental maps, with the present dataset outperforming most others in both cases (Supplementary Figure S3).

Because qPTxM searches for modifications using only the information from a map and unmodified atomic model, a second application of the program for modification discovery is also possible. When examining a high-resolution structure without prior knowledge about the modifications on any of the nucleotides, qPTxM can alternatively reduce the labor required to search for possible modifications by supplying the user with a limited number of candidate sites to be investigated by biochemical methods. The researcher's independent examination of the candidate sites in Coot may still identify a number of cases where features in the density can be otherwise explained. In addition to reducing the number of sites to examine by a factor of  $\sim 50$ , qPTxM applies criteria uniformly to all nucleotides, eliminating any confirmation bias favoring sites that appear modified when the map is contoured to a particular level (Supplementary Figure S5).

### Nucleotide conformations in the ribosome

RNA nucleobases can adopt two major conformations, *syn* and *anti*, that differ in the position of the base with respect to the sugar ring, and are interchangeable through rotation around the glycosidic bond. While the *anti* conformation is energetically more favorable than the *syn* conformation, the *syn* state affords a more compact form of the nucleotide aiding in the stability of RNA (34,35). Detailed analysis



**Figure 4.** *Syn* nucleotides with good density present in cryo-EM structure are shown as red spheres. *Syn* conformation is defined by glycosidic torsion angle ( $\chi$ ) of  $0^\circ \pm 110^\circ$ . Features in the 50S subunit include the central protuberance (CP), L1 arm (L1 stalk), protein L9 (L9), L7–L12 region (L7/L12 stalk) and the peptidyl transferase center (PTC).

of several RNAs and RNA-containing complexes including ribosomes indicated that most *syn* nucleobases participate in tertiary stacking and tertiary base-pairing (34). Consequently, this conformation often occurs in the functionally important regions of RNA, where *syn* pyrimidines are rarer than *syn* purines, as pyrimidine derivatives have a larger energetic penalty to adopting this conformation. The importance of *syn* conformations to RNA function is still not completely understood, and high-resolution structures are instrumental in identifying conserved *syn* nucleotides and aiding in assessment of their functional role.

Glycosidic torsion angle ( $\chi$ ) for each nucleotide present in our structure was measured using X3DNA suite, an integrated software system for analysis of nucleic-acid containing structures (20). The *syn* conformation was defined with a  $\chi$ -angle in a range of  $0^\circ \pm 110^\circ$ , following recommendations from recent analysis of high-resolution RNA datasets (36). In the 23S rRNA, we observed 43 *syn* purines and five *syn* pyrimidine with good densities (Figure 4, Supplementary Table S6 and Supplementary Table S7). Assigned *syn* nucleotides were compared to *syn* bases reported in other high resolution *E. coli* ribosome structures, such as the one presented in (14) (Supplementary Table S6 and Supplementary Table S7). We observed high similarity between the two structures, suggesting that differences in experimental conditions between two techniques had minimal effect on the conformations of the nucleotides, and that designated *syn* nucleotides most likely have an important structural or functional role (34). In our structure, *syn* nucleotides are observed throughout the 23S rRNA (Figure 4), however when examining each domain, the largest number of *syn* nucleotides are found in the domain V of 23S rRNA, whose central loop lines the PTC and NPET (Figure 4, Supplementary Table S6 and Supplementary Table S7). One example is highly conserved nucleotide G2576, which occupies *syn* conformation extending the stacking between G2576 and G2505. This stacking supports the single stranded rRNA segment of the central loop of domain

V, which lines the NPET and is a binding site for several antibiotics (Supplementary Figure S6A). Additionally, in the PTC region, nucleotide A2503 methylated at the C2 position ( $m^2A$ ) also adopts the full *syn* conformation. It has been proposed that the presence of this methylated nucleotide in the *syn* conformation extends the stacking between A2059 and A2503, stabilizing the fold of the two single-stranded rRNA sections (13). *Syn* nucleotides also support tertiary structure through tertiary base-pairing. For example, in the helix H20 of domain I, A330 assumes the *syn* conformation to form *trans* Watson–Crick/sugar edge A-G base pair (37) with G307, a nucleotide part of the helix H19 (Supplementary Figure S6B). This base pairing stabilizes the interaction between two helices, an interaction which is additionally facilitated by the ribosomal protein L24. Protein L24 is an assembly initiator protein and organizer of rRNA folding (38).

### Solvation of the PTC and NPET

Most clinically relevant antibiotics that target the 50S subunit bind in the PTC or in the NPET region. These antibiotics inhibit protein synthesis by either perturbing the binding of tRNA at the A- or P-sites or by preventing the passage of the nascent peptide chain through the ribosomal tunnel (7). In the previous X-ray crystal structures, ordered water molecules were observed in the known antibiotic-binding sites, including PTC and NPET (14). High-resolution of the new cryo-EM structure allowed for the modelling of water molecules which were assigned using a utility supplied within Coot (19). However, while water molecules coordinating magnesium ions were evident and well-resolved throughout the structure (Supplementary Figure S7), only a few ordered water molecules were observed in the PTC and NPET. We hypothesize that water molecules and ions are diffused throughout the NPET, and could become ordered in the presence of the nascent peptide or an antibiotic.

### DISCUSSION

In the current study, we have harnessed several technological advances to obtain the highest resolution cryo-EM ribosome structure to date at an average resolution of 2.2 Å. This resolution allowed for the modelling of known RNA modifications. Additionally, we present a new program for assessing map quality, named qPTxM, which demonstrates the superior quality of the present map.

It is now possible to resolve small features such as post-transcriptional modifications in sufficiently high-resolution cryo-EM maps. For structures where modifications have been determined by biochemical methods, such as the *E. coli* ribosome, we can use our ability to rediscover these features in the map as a proxy for map quality. This is the design principle guiding development of qPTxM. Because program execution involves rediscovering these modifications, it can also be used for *de novo* discovery of modifications in a map of sufficiently high quality. We caution against using cryoEM as the sole discovery method for assignment of post-transcriptional modifications, as recent work has shown the potential for false identification of

modifications (17,18). Consistent with that emerging view, our analyses indicate that it is not yet possible to identify true modifications without also producing many false positives, although we anticipate that additional methodological advancements in cryo-EM will soon make this viable. We therefore describe this alternative use of qPTxM in advance of a map suitable for this use of the software.

The present structure can be used for comparison to other ribosome structures. While ribosomal RNA is extensively modified across all three kingdoms of life, the nature and the location of the rRNA modifications differ between kingdoms. For example, C1962 is a 5-methylcytidine ( $m^5C$ ) in *E. coli* and *Thermus thermophilus* 23S rRNA, and unmethylated in archaea and eukaryotes (39). Presence of a methyl group on C1962 extends the stacking between G1935–C1962 within the H70–H71 helical junction. As helix H71 participates in the formation of bridge B3 with 16S rRNA, it is possible that C1942 aids in strengthening of the intersubunit contact and subsequently in maintaining the 70S stability (40). Differences in the extent of modification at this nucleotide could indicate dissimilarities in the optimization of intersubunit contact and 70S stability at distinct temperatures and in different environments.

Another interesting discovery from comparison of our structure to other ribosome structures is the extent of solvation of functional regions, such as the NPET. Water molecules are known to play both functional and structural roles, especially in charged molecules such as nucleic acids. While in our structure the water molecules appear to be disordered throughout the NPET, in the high-resolution X-ray structure several well-ordered waters are identified (14). We hypothesize that these differences in the hydration pattern between the two structures are real (not a limitation of the resolution) and likely due to differences in experimental methods. One possibility is that rigid packing in the crystal lattice and dehydration procedures prior to cryocooling likely promotes ordering of the water in the X-ray structure. Alternatively, quicker vitrification during preparation of the cryo-EM sample might have aided in capturing a more biologically relevant state (41). Ultimately, understanding the hydration pattern in the ribosome structure will be important for future computational studies, calculations of the electrostatic contribution of water molecules (42) and other ions (43), and might aid in the development of new antibiotics that target the bacterial ribosome.

Lastly, our structure also allowed us to determine the conformation of each nucleotide in the structure. We have identified 48 *syn* nucleotides, predominantly in the functionally important regions of the ribosome, where they likely have a functional or structural role. Our findings are highly consistent with previous *syn* designations in X-ray crystal structure (14) and previous analysis that revealed the overabundance of *syn* nucleotides in the antibiotic-binding sites (44). *Syn* conformation has been proposed to be associated with the higher flexibility of rRNA in these sites, resulting in pockets that are capable of the induced fit-binding mode. Thus, the identification of the *syn* nucleotides in the high-resolution structures, and especially *syn* nucleotides conserved across kingdoms could facilitate the identification of new functional sites suitable for the design of ribosome targeting drugs.

## DATA AVAILABILITY

Atomic coordinates have been deposited in the Protein Data Bank under accession number 6PJ6, the density map has been deposited in the EMDB under accession number 20353.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the Fujimori lab members for helpful discussions and comments on the manuscript, Pavel Afonine for help with cctbx map tools and Saulo de Oliveira for help with machine learning. We thank the UCSF Center for Advanced CryoEM, which is supported by the National Institutes of Health [S100D020054 and 1S100D021741] and the Howard Hughes Medical Institute (HHMI).

## FUNDING

National Institute of Health [NIAID R01AI137270 to D.G.F.]; HHMI Faculty Scholar grant [to A.F.]; A.F. is a Chan Zuckerberg Biohub Investigator and is supported by the National Institutes of Health [1R01GM127673-01 and 5P50GM082545-12]; J.S.F. is supported by the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation, a Sangvhi-Agarwal Innovation Award and the National Institutes of Health [R01GM123159]. Funding for open access charge: National Institute of Health [NIAID R01AI137270].

*Conflict of interest statement.* None declared.

## REFERENCES

- Green, R. and Noller, H.F. (1996) In vitro complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for Escherichia coli 50S ribosomal subunit assembly and function. *RNA*, **2**, 1011–1021.
- Li, S. and Mason, C.E. (2014) The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.*, **15**, 127–150.
- Sergeeva, O.V., Bogdanov, A.A. and Sergiev, P.V. (2015) What do we know about ribosomal RNA methylation in Escherichia coli? *Biochimie*, **117**, 110–118.
- Nachtergaele, S. and He, C. (2017) The emerging biology of RNA post-transcriptional modifications. *RNA Biol.*, **14**, 156–163.
- Grosjean, H. (2009) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. CRC Press.
- McCusker, K.P. and Fujimori, D.G. (2012) The chemistry of peptidyltransferase center-targeted antibiotics: enzymatic resistance and approaches to countering resistance. *ACS Chem. Biol.*, **7**, 64–72.
- Wilson, D.N. (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.*, **12**, 35–48.
- Lin, J., Zhou, D., Steitz, T.A., Polikanov, Y.S. and Gagnon, M.G. (2018) Ribosome-targeting antibiotics: modes of action, mechanisms of resistance, and implications for drug design. *Annu. Rev. Biochem.*, **87**, 451–478.
- Genuth, N.R. and Barna, M. (2018) The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol. Cell*, **71**, 364–374.
- Byrgazov, K., Vesper, O. and Moll, I. (2013) Ribosome heterogeneity: another level of complexity in bacterial translation regulation. *Curr. Opin. Microbiol.*, **16**, 133–139.
- Ferretti, M.B., Barre, J.L. and Karbstein, K. (2018) Translational reprogramming provides a blueprint for cellular adaptation. *Cell Chem. Biol.*, **25**, 1372–1379.

12. Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H. and Moll, I. (2011) Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell*, **147**, 147–157.
13. Polikanov, Y.S., Melnikov, S.V., Söll, D. and Steitz, T.A. (2015) Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.*, **22**, 342–344.
14. Noeske, J., Wasserman, M.R., Terry, D.S., Altman, R.B., Blanchard, S.C. and Cate, J.H.D. (2015) High-resolution structure of the *Escherichia coli* ribosome. *Nat. Struct. Mol. Biol.*, **22**, 336–341.
15. Shalev-Benami, M., Zhang, Y., Matzov, D., Halfon, Y., Zackay, A., Rozenberg, H., Zimmerman, E., Bashan, A., Jaffe, C.L., Yonath, A. *et al.* (2016) 2.8-Å Cryo-EM structure of the large ribosomal subunit from the eukaryotic parasite *Leishmania*. *Cell Rep.*, **16**, 288–294.
16. Myasnikov, A.G., Kundhavi Natchiar, S., Nebout, M., Hazemann, I., Imbert, V., Khatter, H., Peyron, J.-F. and Klaholz, B.P. (2016) Structure-function insights reveal the human ribosome as a cancer target for antibiotics. *Nat. Commun.*, **7**, 12856.
17. Natchiar, S.K., Myasnikov, A.G., Kratzat, H., Hazemann, I. and Klaholz, B.P. (2017) Visualization of chemical modifications in the human 80S ribosome structure. *Nature*, **551**, 472–477.
18. Taoka, M., Nobe, Y., Yamaki, Y., Sato, K., Ishikawa, H., Izumikawa, K., Yamauchi, Y., Hirota, K., Nakayama, H., Takahashi, N. *et al.* (2018) Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.*, **46**, 9289–9298.
19. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
20. Lu, X.-J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
21. Mehta, P., Woo, P., Venkataraman, K. and Karzai, A.W. (2012) Ribosome purification approaches for studying interactions of regulatory proteins and RNAs with the ribosome. *Methods Mol. Biol.*, **905**, 273–289.
22. Khatter, H., Myasnikov, A.G., Natchiar, S.K. and Klaholz, B.P. (2015) Structure of the human 80S ribosome. *Nature*, **520**, 640–645.
23. Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y. and Agard, D.A. (2017) MotionCorr2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.
24. Zhang, K. (2016) Gctf: real-time CTF determination and correction. *J. Struct. Biol.*, **193**, 1–12.
25. Scheres, S.H.W. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, **180**, 519–530.
26. Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, **14**, 290–296.
27. Grant, T., Rohou, A. and Grigorieff, N. (2018) cisTEM, user-friendly software for single-particle image processing. *eLife*, **7**, e35383.
28. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
29. Lebedev, A.A., Young, P., Isupov, M.N., Moroz, O.V., Vagin, A.A. and Murshudov, G.N. (2012) JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 431–440.
30. Chen, V.B., Arendall, W.B. 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
31. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
32. Barad, B.A., Echols, N., Wang, R.Y.-R., Cheng, Y., DiMaio, F., Adams, P.D. and Fraser, J.S. (2015) EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods*, **12**, 943–946.
33. Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crécy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A. *et al.* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
34. Sokolowski, J.E., Godfrey, S.A., Dombrowski, S.E. and Bevilacqua, P.C. (2011) Prevalence of syn nucleobases in the active sites of functional RNAs. *RNA*, **17**, 1775–1787.
35. Murthy, V.L., Srinivasan, R., Draper, D.E. and Rose, G.D. (1999) A complete conformational map for RNA. *J. Mol. Biol.*, **291**, 313–327.
36. Jain, S., Richardson, D.C. and Richardson, J.S. (2015) Computational methods for RNA structure validation and improvement. *Methods Enzymol.*, **558**, 181–212.
37. Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
38. Stelzl, U., Spahn, C.M. and Nierhaus, K.H. (2000) Selecting rRNA binding sites for the ribosomal proteins L4 and L6 from randomly fragmented rRNA: application of a method called SERF. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 4597–4602.
39. Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. (2008) The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.*, **36**, D178–D183.
40. Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S.M., Van Roey, P., Agrawal, R.K., Harvey, S.C., Sali, A. *et al.* (2003) Study of the structural dynamics of the *E. coli* 70S ribosome using real-space refinement. *Cell*, **113**, 789–801.
41. Halle, B. (2004) Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 4793–4798.
42. Kulik, M., Goral, A.M., Jasiński, M., Dominiak, P.M. and Trylska, J. (2015) Electrostatic interactions in aminoglycoside-RNA complexes. *Biophys. J.*, **108**, 655–665.
43. Rozov, A., Khusainov, I., El Omari, K., Duman, R., Mykhaylyk, V., Yusupov, M., Westhof, E., Wagner, A. and Yusupova, G. (2019) Importance of potassium ions for ribosome structure and function revealed by long-wavelength X-ray diffraction. *Nat. Commun.*, **10**, 2519.
44. David-Eden, H., Mankin, A.S. and Mandel-Gutfreund, Y. (2010) Structural signatures of antibiotic binding sites on the ribosome. *Nucleic Acids Res.*, **38**, 5982–5994.

## SUPPLEMENTARY INFORMATION

### High-resolution cryo-electron microscopy of the *Escherichia coli* 50S subunit and validation of nucleotide modifications

Vanja Stojković<sup>1</sup>, Alexander G. Myasnikov<sup>2</sup>, Iris D. Young<sup>3</sup>, Adam Frost<sup>2,4\*</sup>, James S. Fraser<sup>3,4\*</sup>, Danica Galonić Fujimori<sup>1,4,5\*</sup>

#### SUPPLEMENTARY FIGURES AND TABLES LEGENDS

Supplementary Figure S1. Diagram of the series of tunable filters applied to candidate modifications by qPTxM. Individual filters select the modification sites where map density or some related measurement most closely matches expectations. Examples of the filters and their associated measurements are shown at right for a candidate N3-methylation on a uridine. Two filters, difference densities fraction and reference densities fraction, are determined as the thresholds selecting particular proportions of the possible modifications, while other filters are encoded as absolute thresholds. All thresholds or proportions, including the score cutoff, can be adjusted by passing optional command line arguments to qPTxM. The accepted modifications are those that pass all the filters, as represented by the funnel.

Supplementary Figure S2. A custom plugin script `goto_ptms.py` is written out with each run of `qptm.py`. Supplying this script to Coot on launch along with the original and modified models allows the user to step through all suggested sites of post-transcriptional modifications.

Supplementary Figure S3. Plot of qPTxM results for the test ribosome maps in receiver operating characteristic (ROC) space. True positive rate is calculated as the ratio of the number of true positives to the sum of the true positives and false negatives. False positive rate is calculated as the ratio of the number of false positives to the sum of the false positives and true negatives. An inset at top right calls attention to the sub-5% false positive rate across all datasets. Four groups of datasets are plotted, corresponding to the two groups of datasets (either the maps and models available in the PDB, or noise-free synthetic maps generated from these models with a random 10% of sites modified) and the two ways of assessing candidate modifications (either by running qPTxM with all default parameters except resolution or by predicting modifications with a random forest classifier). The random forest classifier was trained on a collection of synthetic maps from the same models with a different random 10% of sites modified in each instance, and was not trained on any of the data plotted or on any map generated from the model presented here. True positive rates could not be calculated for experimental maps with no true positives, so experimental points are shown only for structures with known modifications. We added any modifications known to be present in *E. coli*, *S. cerevisiae* or *H. sapiens* to ribosome structures in these organisms where not already modelled.

Supplementary Figure S4. Cryo-EM density map of modified nucleotides in the 50S subunit.

Supplementary Figure S5. Visual similarity of modified and unmodified positions in the present map (A) and a 2.9 Å resolution human ribosome map, PDB 6EK0 (B). The danger of manual visual analysis of maps and propensity for confirmation bias is demonstrated by comparing the site of a true positive to three sites of false positives with similar density for each map. Maps are contoured at 5 sigma in all cases.

Supplementary Figure S6. Example of *syn* nucleotides participating in tertiary base stacking (A) and tertiary base pairing (B). (A) Nucleotide G2576 in 23S rRNA adopts a *syn* conformation that extends the stacking between G2576 and G2505. G2505 lines the peptide exit tunnel and interacts directly with the nascent peptide. Model of VemP nascent peptide chain (PDB 5NWX) (1) is superimposed on the *E. coli* 50S structure. (B) Nucleotide A330 assumes the *syn* conformation to form trans Watson-Crick/sugar edge A-G base pair with G307. G307 and A330 are part of the two loops in helices H19 and H20, respectively. These loops are additionally stabilized through interaction with ribosomal protein L24 (cyan color). Hydrogen bonds are indicated by black dashed lines.

Supplementary Figure S7. An example of water coordinated to magnesium ions. Phosphate of a nucleotide U567 in 23S rRNA also coordinates shown magnesium ion.

Supplementary Table S1. Comparison of the qPTxM default behavior and random forest predictions on the 22 publicly available ribosomes tested and the present dataset. Where modifications were known to be present from other structures from the same organism, but not modeled in a particular model, these modifications were added to the model and contribute to true positives and false negatives. The reported resolution in the Protein Data Bank was passed as the only non-default parameter to qPTxM.

Supplementary Table S2. Comparison of the qPTxM default behavior and random forest predictions on synthetic datasets generated from the 22 publicly available ribosomes tested and the present dataset. Where modifications were known to be present from other structures from the same organism, but not modeled in a particular model, these modifications were added to the model and contribute to true positives and false negatives. Reported resolutions of the original models (presumably affecting model geometries) is listed, but all synthetic maps were generated at 2.2 Å resolution and with all residues reset to a B-factor of 10.

Supplementary Table S3. Modified nucleotides in the *E. coli* 50S subunit.

Supplementary Table S4. Solvation of pseudouridines.

Supplementary Table S5. List of nucleotides known to be modified in *E. coli* 50S subunit with their qPTM scoring percentile. Scores for the assigned modifications were determined using a calculated cryo-EM map at a 2.2 Å resolution. Scores were calculated on nucleotides whose correlation coefficients between experimental and calculated cryo-EM maps were at least 0.6. Modifications were limited to the 60% of nucleotides that had the strongest experimental map density at reference atom positions and the 10% of sites with the strongest difference map density at the positions of modifications. Modifications were also filtered by three thresholds on the ratios of densities in the experimental map,  $d_{\text{far}} \leq 0.6 \cdot d_{\text{new}}$ ,  $d_{\text{far}} \leq 0.4 \cdot d_{\text{mid}}$  and  $d_{\text{ref}} \leq 3.5 \cdot d_{\text{new}}$  (see Methods). Scores are scaled ratios of difference and experimental map densities at the proposed and reference atom positions, respectively, and percentiles were calculated among the 176 sites passing all tests and scoring at least 0.5. NP indicates that the modification did not meet these conditions and was not assigned a score. For the nucleotide 1915, methylation was tested starting from a modelled uridine, as pseudouridine cannot be identified from the map alone.

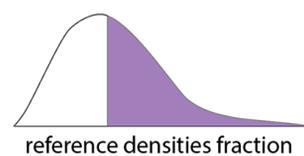
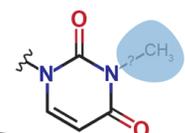
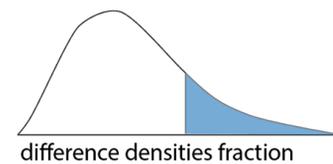
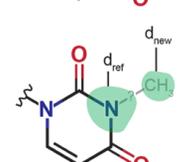
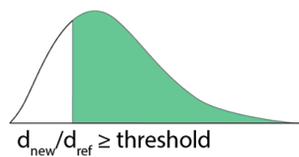
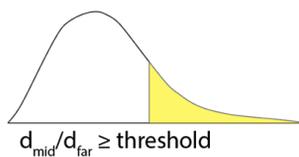
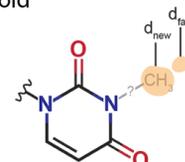
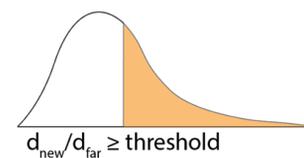
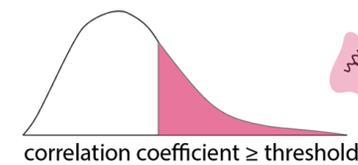
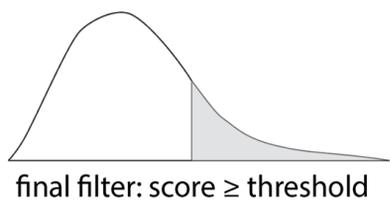
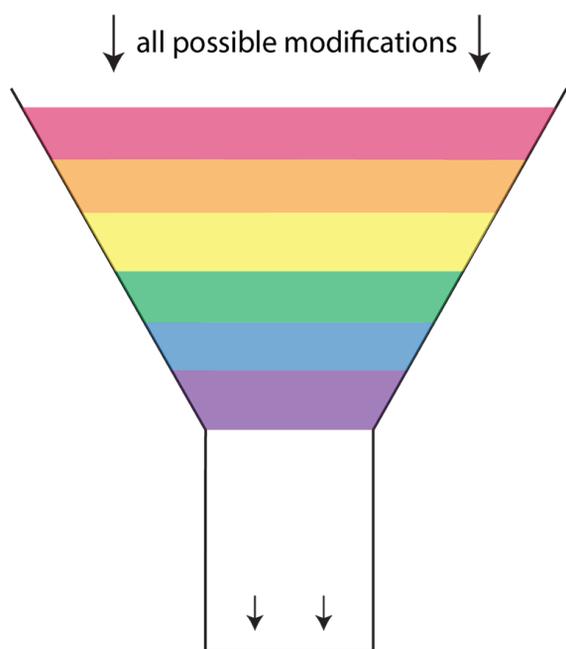
Supplementary Table S6. Comparison of *syn* purines with good density between cryo-EM structure and high-resolution X-ray crystal structure (2). *Syn* conformation is defined by glycosidic torsion angle ( $\chi$ ) of  $0^\circ \pm 110^\circ$ . *Syn* purines present only in high-resolution X-ray crystal structure are shown in bold.

Supplementary Table S7. Comparison of *syn* pyrimidines with good density between cryo-EM structure and high-resolution X-ray crystal structure (2). *Syn* conformation is defined by glycosidic torsion angle ( $\chi$ ) of  $0^\circ \pm 110^\circ$ . *Syn* pyrimidines present only in high-resolution X-ray crystal structure are shown in bold.

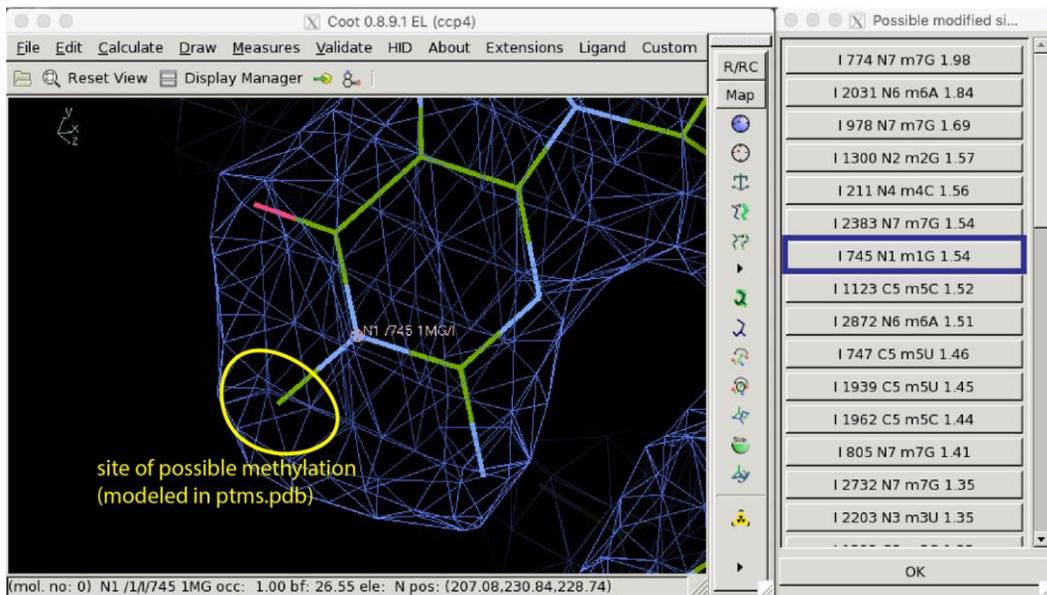
## REFERENCES

1. Su, T., Cheng, J., Sohmen, D., Hedman, R., Berninghausen, O., von Heijne, G., Wilson, D.N., Beckmann, R. (2017) The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *Elife*. **6**, e25642.
2. Noeske, J., Wasserman, M.R., Terry, D.S., Altman, R.B., Blanchard, S.C. and Cate, J.H.D. (2015) High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.*, **22**, 336–341.

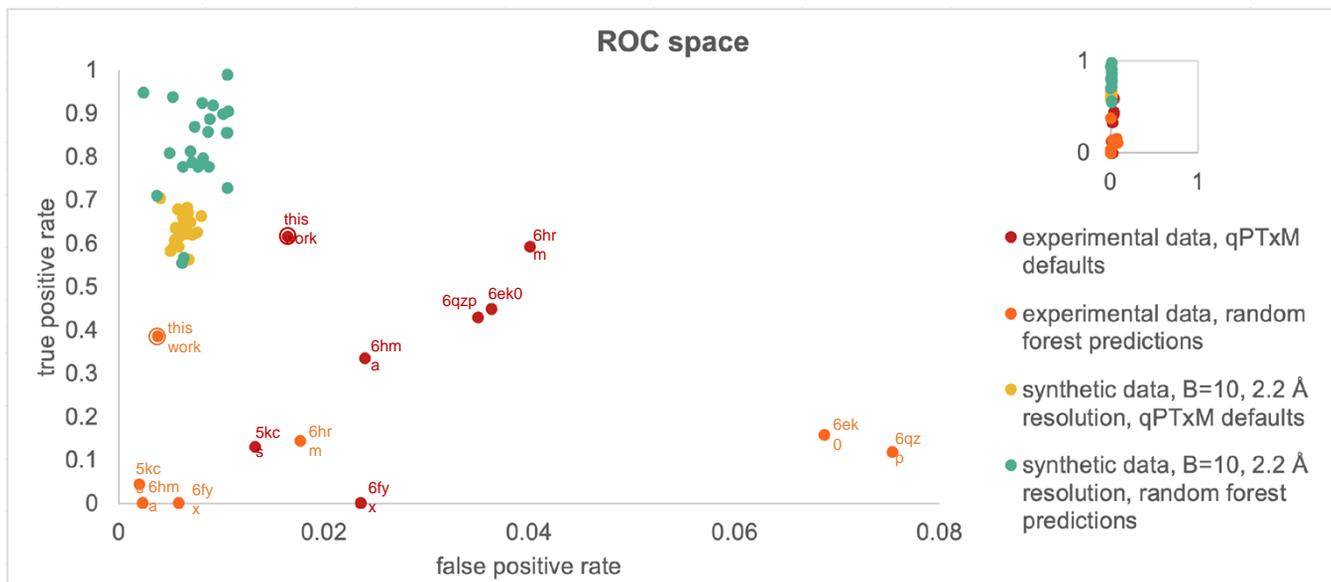
Supplementary Figure S1.



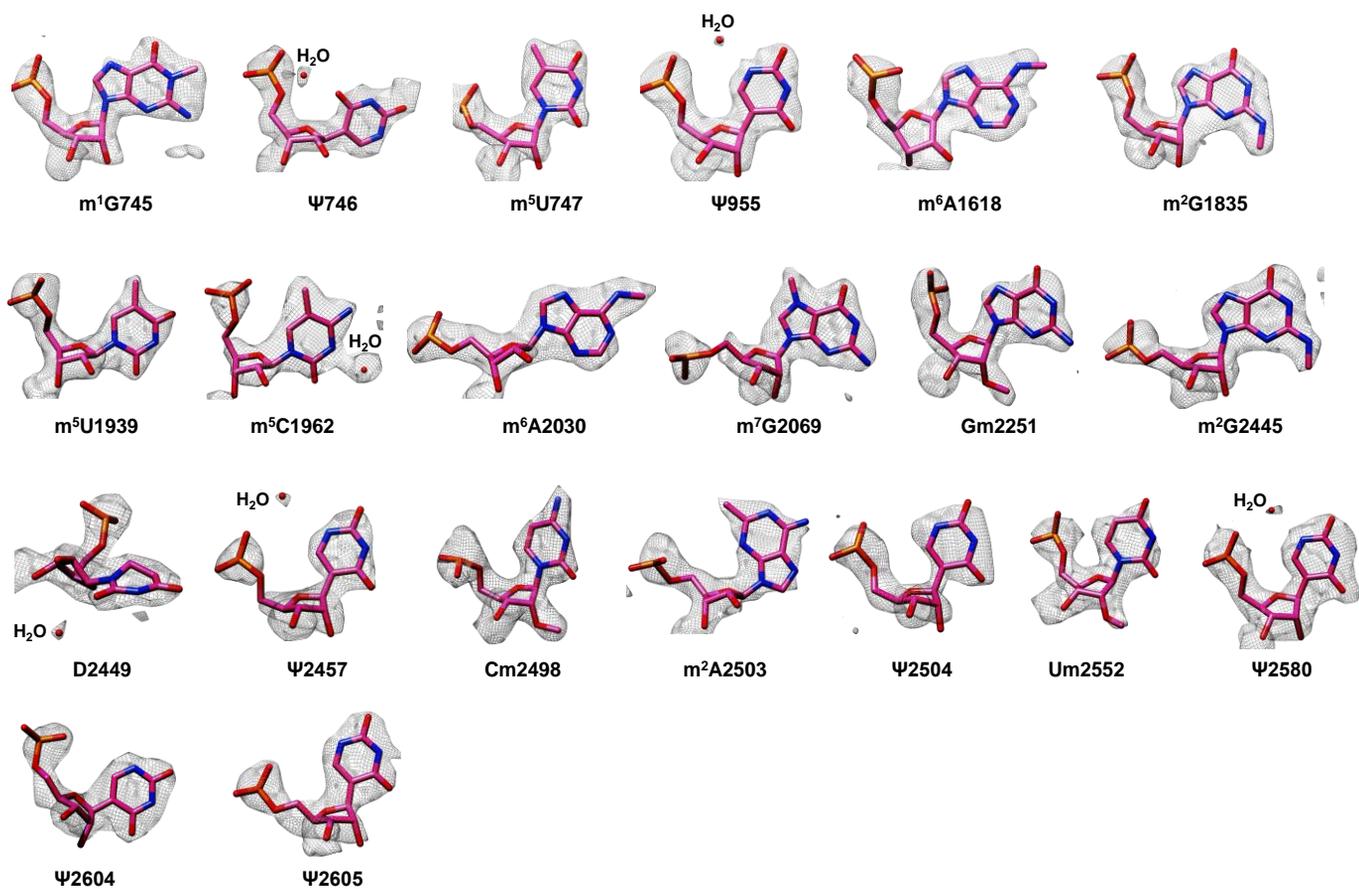
# Supplementary Figure S2.



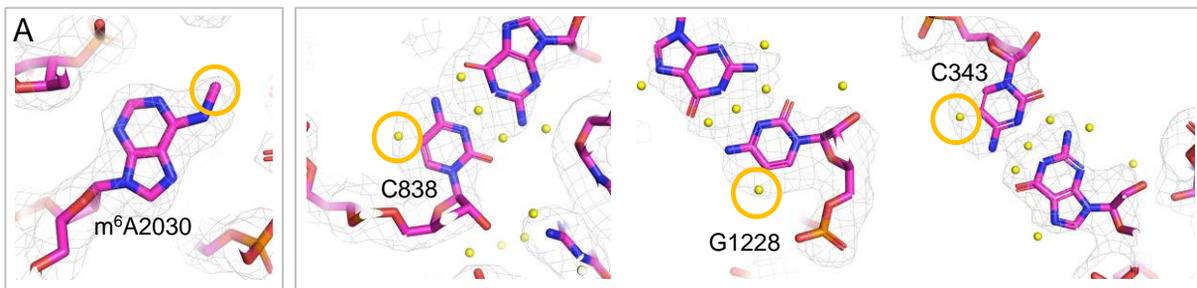
Supplementary Figure S3.



Supplementary Figure S4.

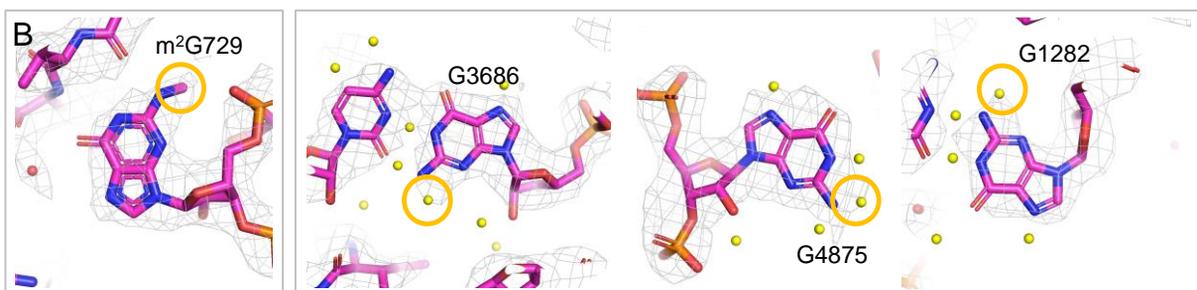


Supplementary Figure S5.



true positive

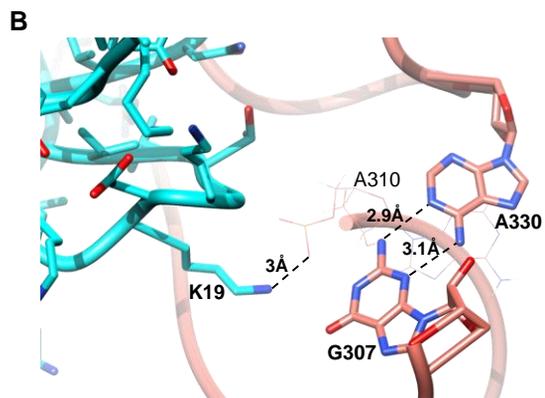
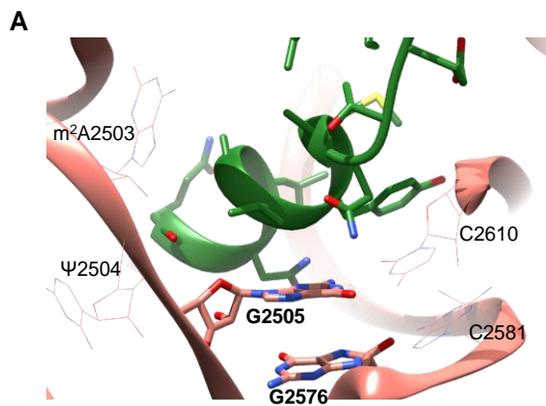
false positives



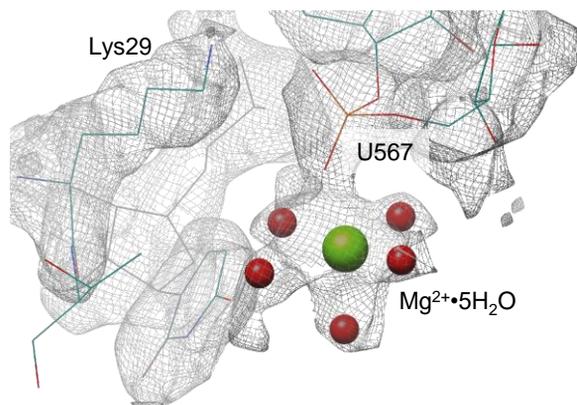
true positive

false positives

Supplementary Figure S6.



Supplementary Figure S7.



Supplementary Table S1.

model	subunit	organism	dmin (Å)	experimental data qPTxM default parameters				experimental data random forest predictions			
				true +	false +	false -	true -	true +	false +	false -	true -
5kcs	70S	<i>E. coli</i>	3.9	3	239	20	17733	1	37	22	17935
5zeb	70S	<i>M. smegmatis</i>	3.4	0	422	0	18322	0	750	0	17994
6dzi	70S	<i>M. smegmatis</i>	3.46	0	505	0	18025	0	1380	0	17150
6dzk	30S	<i>M. smegmatis</i>	3.6	0	143	0	5760	0	398	0	5505
6dzp	50S	<i>M. smegmatis</i>	3.42	0	331	0	12296	0	1416	0	11211
6ek0	80S	<i>H. sapiens</i>	2.9	34	803	42	21280	12	1519	64	20564
6fyx	48S	<i>S. cerevisiae</i>	3.5	0	166	10	6865	0	41	10	6990
6gaw	55S	<i>S. scrofa</i>	3.2	0	244	0	9655	0	112	0	9787
6gaz	28S	<i>S. scrofa</i>	3.1	0	103	0	3786	0	444	0	3445
6gb2	39S	<i>S. scrofa</i>	3.2	0	134	0	6136	0	462	0	5808
6gq1	80S	<i>S. cerevisiae</i>	4.4	0	458	0	19983	0	908	0	19533
6gxm	70S	<i>E. coli</i>	3.8	0	320	0	17718	0	469	0	17569
6gz3	80S	<i>S. cerevisiae</i>	3.6	0	455	0	21787	0	69	0	22173
6gzq	70S	<i>T. thermophila</i>	3.28	0	409	0	17488	0	1305	0	16592
6h4n	100S	<i>E. coli</i>	3	0	184	0	17839	0	567	0	17456
6ha1	70S	<i>B. subtilis</i>	3.1	0	512	0	17475	0	107	0	17880
6hiv	mt-LSSU	<i>T. brucei</i>	7.8	0	52	0	4331	0	98	0	4285
6hiw	mt-SSU	<i>T. brucei</i>	3.37	0	52	0	2196	0	188	0	2060
6hix	mt-LSU	<i>T. brucei</i>	3.39	0	66	0	2069	0	297	0	1838
6hma	50S	<i>S. aureus</i>	2.65	1	274	2	11129	0	27	3	11376
6hrm	70S	<i>E. coli</i>	2.96	13	711	9	17007	3	314	18	17404
6qzp	80S	<i>H. sapiens</i>	2.9	33	778	44	21426	9	1674	68	20530
this work	50S	<i>E. coli</i>	2.2	8	193	5	11506	5	45	8	11654

Supplementary Table S2.

model	subunit	organism	dmin (Å)	synthetic data qPTxM default parameters				synthetic data random forest predictions			
				true +	false +	false -	true -	true +	false +	false -	true -
5kcs	70S	<i>E. coli</i>	3.9	283	96	183	17433	376	87	90	17442
5zeb	70S	<i>M. smegmatis</i>	3.4	317	102	208	18117	408	141	117	18078
6dzi	70S	<i>M. smegmatis</i>	3.46	323	146	165	17896	276	114	212	17928
6dzk	30S	<i>M. smegmatis</i>	3.6	92	38	54	5719	125	50	21	5707
6dzp	50S	<i>M. smegmatis</i>	3.42	227	81	139	12180	284	108	82	12153
6ek0	80S	<i>H. sapiens</i>	2.9	337	125	215	21482	306	133	246	21474
6fyx	48S	<i>S. cerevisiae</i>	3.5	110	40	76	6815	151	48	35	6807
6gaw	55S	<i>S. scrofa</i>	3.2	171	65	84	9579	229	98	26	9546
6gaz	28S	<i>S. scrofa</i>	3.1	65	22	31	3771	90	20	6	3773
6gb2	39S	<i>S. scrofa</i>	3.2	105	41	49	6075	112	65	42	6051
6gq1	80S	<i>S. cerevisiae</i>	4.4	353	140	192	19756	423	124	122	19772
6gxm	70S	<i>E. coli</i>	3.8	265	120	206	17447	403	186	68	17381
6gz3	80S	<i>S. cerevisiae</i>	3.6	358	155	220	21509	410	81	168	21583
6gzq	70S	<i>T. thermophila</i>	3.28	297	133	178	17289	429	186	46	17236
6h4n	100S	<i>E. coli</i>	3	260	94	179	17490	345	126	94	17458
6ha1	70S	<i>B. subtilis</i>	3.1	286	105	177	17419	396	184	67	17340
6hiv	mt-LSSU	<i>T. brucei</i>	7.8	80	27	44	4232	110	38	14	4221
6hiw	mt-SSU	<i>T. brucei</i>	3.37	38	9	16	2185	43	18	11	2176
6hix	mt-LSU	<i>T. brucei</i>	3.39	37	13	19	2066	53	5	3	2074
6hma	50S	<i>S. aureus</i>	2.65	185	60	128	11033	289	90	24	11003
6hrm	70S	<i>E. coli</i>	2.96	281	113	146	17200	422	184	5	17129
6qzp	80S	<i>H. sapiens</i>	2.9	345	120	199	21617	499	201	45	21536
this work	50S	<i>E. coli</i>	2.2	160	58	115	11379	239	85	36	11352

Supplementary Table S3.

Position	Known modification	Cryo EM density
745	m <sup>1</sup> G	+
746	Ψ	+
747	m <sup>5</sup> U	+
955	Ψ	+
1618	m <sup>6</sup> A	+
1835	m <sup>2</sup> G	+
1911	Ψ	-
1915	m <sup>8</sup> Ψ	- (a)
1917	Ψ	-
1939	m <sup>5</sup> U	+
1962	m <sup>5</sup> C	+
2030	m <sup>6</sup> A	+
2069	m <sup>7</sup> G	+
2251	Gm	+
2445	m <sup>2</sup> G	+
2449	D	+
2457	Ψ	+
2498	Cm	+
2501	s <sup>2</sup> C	- (b)
2503	m <sup>2</sup> A	+
2504	Ψ	+
2552	Um	+
2580	Ψ	+
2604	Ψ	+
2605	Ψ	+

(a) disordered region

(b) partial modification

(c) poor density for modification

Supplementary Table S4.

<b>Pseudouridine</b>	<b><i>Syn/anti</i> conformation</b>	<b>H<sub>2</sub>O-binding to N1</b>	<b>H<sub>2</sub>O mediated contact to</b>
746	<i>syn</i>	no	N/A
955	<i>anti</i>	yes	OP1 955 OP2 954
1911	N/A	N/A	N/A
1915	N/A	N/A	N/A
1917	N/A	N/A	N/A
2457	<i>anti</i>	yes	OP2 2456 OP2 2457
2504	<i>anti</i>	no	-
2580	<i>anti</i>	yes	OP2 2580
2604	<i>anti</i>	no	-
2605	<i>anti</i>	no	-

N/A = not applicable; no electron density present

Supplementary Table S5.

<b>Nucleotide</b>	<b>Modification</b>	<b>Scoring percentile</b>
745	m <sup>1</sup> G	0.86
747	m <sup>5</sup> U	NP
1618	m <sup>6</sup> A	0.57
1835	m <sup>2</sup> G	1.00
1915	m <sup>3</sup> Ψ**	NP
1939	m <sup>5</sup> U	0.43
1962	m <sup>5</sup> C	NP
2030	m <sup>6</sup> A	NP
2069	m <sup>7</sup> G	0.14
2251	Gm	NP
2445	m <sup>2</sup> G	0.71
2498	Cm	0.29
2503	m <sup>2</sup> A	0.00
2552	Um	NP

Supplementary Table S6.

23S rRNA					
Domain I	Domain II	Domain III	Domain IV	Domain V	Domain VI
A 49	G 620	G 1288	A 1669	G 2238	G 2645
A 71	G 729	A 1301	G 1695	G 2250	G 2751
A 74	A 783	G 1311	G 1929	A 2267	A 2765
A 101	A 845	G 1332	A 1936	G 2286	G 2825
G 177	A 933	G 1452		A 2287	A 2879
A 196	A 984	A 1490		A 2430	
A 330	A 1021			A 2503	
<b>A 404</b>	A 1142			A 2518	
A 528	G 1210			G 2576	
A 532				G 2581	

Supplementary Table S7.

<b>23S rRNA</b>	
Nucleotide	Sugar pucker
C 323	C2'-endo
$\Psi$ 746	C1'-exo
U 1313	C4'-exo
<b>U 1758</b>	C2'-endo
U 1779	C4'-exo
<b>U 2689</b>	C2'-endo
U 2884	C2'-endo