Title
Strategies for High-Resolution Investigation of Bacterial Ribosome Function by Cryo-Electron Microscopy

Author
Watson, Zoe Lee

Publication Date
2020

Peer reviewed|Thesis/dissertation

Strategies for High-Resolution Investigation of Bacterial Ribosome Function by Cryo-Electron Microscopy

By

Zoe Lee Watson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

In

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jamie H. D. Cate, Chair
Professor Naomi S. Ginsberg
Professor Eva Nogales

Fall 2020

Abstract

Strategies for High-Resolution Investigation of Bacterial Ribosome Function by Cryo-Electron Microscopy

By

Zoe Lee Watson

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Jamie H. D. Cate, Chair

This dissertation comprises projects focused broadly on cryo-electron microscopy (cryo-EM) methods and their application to high-resolution structural biology of the ribosome. The ribosome is the cell's machinery for creating proteins in a series of steps referred to as translation. During translation, the large and small subunits of the ribosome participate in a cyclical process of association and disassociation, complexing with messenger RNAs, transfer RNAs, and many other translation factors along the way. From the intricacy of this process and of the ribosome itself comes a high degree of efficiency and fidelity in the synthesis of polypeptides. Structure is at the core of our current (and evolving) understanding of how the ribosome accomplishes this feat. Structures of the ribosome specifically are also rich with information implicating the fundamentals of RNA structure and folding, ribosomal assembly, medicine, and evolutionary relationships.

Much of what has historically made an ideal specimen for single-particle cryo-EM is embodied by the ribosome: it is large, measuring approximately 25 nm in diameter and weighing approximately 2.6 MDa in bacteria, and it displays conformational flexibility that is difficult to capture and resolve by other methods. The recent widespread use of direct electron detectors has pushed the limits of what is achievable by single-particle cryo-EM, whose development over the past several decades has been punctuated by important advances involving the ribosome. Herein are described multiple stories from both sides of the coin – cryo-EM and ribosomal biology – including the most well-resolved ribosome structure to date, whose global resolution broke the 2 Å barrier for the first time for a large ribosomal subunit. Concluding thoughts consider future directions for cryo-EM methods development, informed by a specialized focus on the ribosome during a period of rapid growth in the field.

Acknowledgments

I am not that effusive, but here goes: thank you to Jamie Cate for providing a supportive environment where I could contribute and explore in the way that suited me. To Bob Glaeser and BG Han for being my very generous cryo-EM mentors, and all of the coffee gang at Donner for some of my favorite Ph.D. memories. To many other labmates, colleagues, collaborators, and committees – especially Fred Ward for years of indispensable collaboration, expertise, support, and friendship; Wenfei Li for patience, expertise, humor, and friendship; Alanna Schepartz for opportunities and support while I have been involved in C-GEM; and all members of my qualification exam and dissertation committees. To all of the Kanai Lab at UNC – Yosuke Kanai, Shahar Keinan, Kyle Reeves, Lesheng Li, Yi Yao, and Dillon Yost – for the undergraduate research experience that in part led me here, and ongoing friendships. To my housemates and friends for most of these years, Xinyi Zhou and Yifan Li. To Erin Sullivan for support while I finish this thing.

Most of all, to my parents and Lisa for everything.

Table of Contents

Chapter 1: Introduction

Cryo-EM: Background

Cryo-electron microscopy (cryo-EM) has recently taken center-stage in the world of structural biology. Formerly sometimes known as "blobology," cryo-EM has reached a point where the resolution we can achieve rivals or even exceeds that of X-ray crystallography. Cryo-EM is a subcategory of transmission electron microscopy (TEM), which has been used to visualize the structures of biological macromolecules since the mid-20th century. Work dating back to the late 1950s originated the use of heavy salts to coat the specimen and produce contrast in the images[1] (so-called "negative stain" methods). Techniques relying on contrast from the molecules themselves developed in the 1970s[2,3], allowing us to see interior structural features, not just the solvent-accessible shell surrounding the molecule. This gain in information content comes at the cost of a poor signal-to-noise ratio (SNR), a result of weak contrast and the use of a low electron dose to accommodate the radiation sensitivity of biological samples. Flash-freezing the sample in a thin film of vitreous ice became the standard preparation in the field for its preservation of native structure, improved contrast compared to other embedding media, and the increased protection from radiation damage at cryogenic temperatures. Thus, the basic identity of the field as it exists today was established.

Early efforts towards high resolution were mostly restricted to proteins that would easily assemble as thin crystals or ordered filaments, taking advantage of the combined signal from many asymmetric units as in X-ray crystallography, but with the benefits of real-space imaging to determine phases in the power spectrum[4]. The strength of the interaction between the electron beam and matter requires thinness of the sample, while the low SNR requires the averaging of information from many individual copies of the specimen, or "particles". The thin (ideally 2D) crystal condition was often met by specimens that were recalcitrant to structure determination by other means; by the same token, this has also been one of the reasons why electron crystallography (EC) did not take off outside these specialized cases (although recently, there has been somewhat of a renaissance of crystallographic EM with electron diffraction of microcrystals[5]). Still, EC was responsible for important landmarks in structural biology, including the first demonstration of high-resolution diffraction from a biological specimen in 1976[3] and the first atomic model of a protein based on an EM map in 1990[6]. "Single-particle" cryo-EM offers another approach, which is not a true single-particle technique but refers to the collection of many images (micrographs) of fields of individual particles and averaging them computationally to render the 3D structure from the ensemble of poses[7].

Rapid strides in cryo-EM have recently followed the advent of direct electron detectors (DEDs), which were preceded first by film and then by CCD cameras. Earlier detection methods had major limitations, including that it was only possible to record a single, continuous exposure of each area, with no way to mitigate blurring from electron beam-induced motion of the sample[8]. DEDs have revolutionized the field with their ability to capture dose-fractionated movies with high speed and improved detective quantum efficiency. From movie-based detection alone, immediate improvements in data quality have followed from the necessity of combining the signal from individual frames. In particular, motion tracking[8–10] and dose weighting[11] provide a way to capture enough signal with less blurring of particles and less contribution from frames where too much

damage to the chemical structure has accrued. Higher-quality data has in turn spurred the development of more sophisticated image processing tools to correct for lens aberrations and sample motion on a finer scale, as well as higher-order optical aberrations and the Ewald sphere, to obtain reconstructions that fully extract the information contained within the images[12–15]. With the current most advanced hardware and a very well-behaved sample, it is now possible to achieve true atomic resolution with single particle cryo-EM[16,17]. However, this regime is still elusive for the majority of cases.

The sample remains to be the most limiting factor in routinely achieving well-resolved structures by cryo-EM. Susceptibility to damage from the electron beam and poor SNR from the requisite low dose, as well as conformational heterogeneity, are all inherent challenges imposed by the nature of frozen-hydrated biological specimens. Newer software and hardware allow us to better circumvent some of these, but innovation on the side of sample preparation and/or better ways to ensure particle quality are still needed for reproducibly high-quality results. A hurdle is that we have limited ability to predict the quality of particles until a dataset has been fully processed and the answer is revealed through the EM density of the final reconstruction. At the outset of data collection, many flaws (too-thick ice, broken holes, aggregates, poor particle density, etc.) are quite obvious, while well-preserved atomic and near-atomic resolution information is not as easy to guarantee. With little guidance in this regard, the field will benefit from new approaches that help to avoid excess effort and costly microscope time used on mediocre data.

Cryo-EM and the Ribosome

The ribosome is a ribozyme that catalyzes the formation of peptide bonds to create proteins, the most abundant and varied macromolecules in biology. It performs this task with remarkable precision and speed, doing so by reading a transcription of the cell's genetic instructions via messenger RNA (mRNA), with the help of transfer RNAs (tRNAs) to decode the message. It is not only essential for life, but foundational for our knowledge of RNA structure, and full of promise for engineering new functionality to usher in new biochemical tools and polymer syntheses. The ribosome has historically been an important target for single-particle analysis by cryo-EM, parallel to significant discoveries in the ribosome field that have been driven by X-ray crystal structures. Indeed, only recently has cryo-EM surpassed the resolution afforded by X-ray crystallography, which until now has enabled our most detailed atomic models and mechanistic insights of translation[18,19]. However, even at lower resolution, cryo-EM and the ribosome have mutually served each other for several decades.

The first 70S ribosome structure determined by single-particle cryo-EM was published in 1976, showing the two subunits individually and their association with each other[20]. Because the ribosome is large and asymmetrical, it is an easy-to-see particle with distinctive features even at quite low resolution. Composed of two-thirds RNA and one-third protein, the comparatively high density of RNA lends it somewhat increased contrast compared to protein-only particles. While the ribosome does tend to display some degree of preferred orientation on typical support films, and is less resilient in open-hole conditions, these caveats are not greatly limiting compared to the ribosome's usefulness. It has been the topic of numerous publications developing foundational techniques in cryo-EM, including the random conical tilt method to assign Euler angles[21], angular refinement to improve orientation assignment[22], and local motion correction for individual particles[23], to name a few. A recent review from Brown and Shao nicely summarizes many of the biological insights that have accompanied this journey[24].

Perhaps most importantly for contemporary ribosomal structural biology, structural dynamics and conformational heterogeneity play a significant role in the ribosome's functioning. This makes it particularly well-suited to cryo-EM as our ability to classify and visualize 3D variability within the ensemble of particles has expanded[25–28] and we move beyond static crystal structures. Translation involves a series of many distinct steps during initiation, elongation, and termination that require complex formation with numerous other proteins and RNAs, as well as inter- and intra-subunit conformational changes[29]. Time-resolved studies on the translating ribosome have been able to elucidate a staggering number of distinct states in the range of ~3-4 Å resolution, with laborious classification strategies to tease out the closely-related structures[30]. Progress towards better resolution with fewer particles will facilitate a more detailed understanding of the chemistry and key intermolecular interactions involved in ribosome function and assembly.

In terms of reciprocity between cryo-EM and the ribosome, other specimens have proven to be more useful for pushing resolution boundaries in recent years, especially as a test for new hardware and image processing tools. In particular, apoferritin broke the 2 Å barrier two years before the large subunit of the ribosome[31,32]. However, the ribosome differs from apoferritin in key ways – asymmetry and flexibility – which leave it poised to serve as a benchmark for the many other targets for which these are concerns. The longstanding relationship between cryo-EM and the ribosome is expected to continue to be a mutually fruitful one, and the chapters contained in this dissertation recount stories to those ends.

Summary of Chapters

Chapters 2 and 3 present two published works focused on cryo-EM methods development, done in collaboration with the group of Robert Glaeser at Lawrence Berkeley National Laboratory (efforts led by first author Bong-Gyoon "BG" Han, for both publications). In Chapter 2, we describe a new method of preparing cryo-EM grids with monolayer streptavidin (SA) crystals as a support layer for biotinylated particles, using the ribosome as a test case for performance[33]. This work builds upon previous investigations from the same group (as well as others) on the utility of SA as a "structure-friendly" support layer for biological samples[34–36]. As second author, my primary contributions here were testing and screening of crystallization conditions to reliably achieve single-crystal SA within the field of view of a typical micrograph. This publication serves as proof of principle for the method and shows that the streptavidin affinity layer and random biotinylation can successfully lead to a ribosome reconstruction without apparent artifacts. Furthermore, we show a distribution of orientations that correlates with possible biotinylation sites rather than typically observed orientations.

Chapter 3 is a continued analysis of the same dataset as Chapter 2, showing that SA crystals provide an internal standard for particle quality in addition to the original reasons it was investigated[37]. I am again second author on this publication, and this project has been influential on my thinking for ways to ensure collection of high-quality data in the future, which are expanded on in Chapter 5 and the concluding remarks (Chapter 7).

Chapter 4 is a publication presenting the use of high resolution cryo-EM to determine the structure of a mutant ribosome, P7A7, highlighting the detrimental effects of the mutations on the ribosome's structure and function[38]. I am co-first author on this publication, with my colleague Fred Ward who carried out the biochemical and mass spectrometry work and analysis. We sought the structure of this ribosome to understand its previously-shown ability to incorporate β-amino acids into a protein chain *in vivo*[39]. However, moving to *in vitro* conditions in order to obtain the structure made it apparent that these ribosomes struggled to assemble properly outside the cellular

milieu, a consequence that had gone under-appreciated in *in vivo* screens for new functionality. The structure we obtained bears similarity to previously published late-stage intermediate structures in ribosomal assembly[40], supporting this conclusion.

Chapter 5 describes an effort to implement a "focal pair" strategy of data collection for cryo-EM. The aim here was to make it possible to work with very low-dose, close-to-focus movies that may do better to preserve the highest-resolution information in a micrograph. The combination of these two imaging conditions would typically make the contrast too low to work with the images, however, a second movie for each acquisition target can be collected at a higher defocus value, allowing for ease of handing the data during early processing stages. Analysis of datasets collected in this fashion compared to a control dataset and another conventional dataset suggests that at best, any possible benefit from this method is overshadowed by other factors contributing to data quality for a given microscope session; at worst, focal pairs are detrimental to data quality.

Chapter 6 presents a 2 Å structure of the wild type *E. coli* ribosome, the highest resolution achieved for a ribosome at the time of its publication[32]. The quality of the EM map in this work has led to many new insights, including the discovery of two post-translational modifications seen in the backbones of ribosomal proteins for the first time, a more detailed picture of the antibiotic paromomycin binding site, and has enabled the modeling of numerous ions and thousands of water molecules needed for proper RNA folding. In fact, the map has been central to the development of automated water molecule building into cryo-EM maps in a commonly used model refinement suite[41]. The map and resulting atomic model are expected to find additional use in the future for ribosome biology, benchmarking new model-building and validation methods for cryo-EM, and RNA structure prediction methods.

Chapter 7 is a brief conclusion addressing future directions for ribosomal structural biology and cryo-EM in view of the preceding chapters and the Appendix.

Appendix 1 contains excerpts from a co-authored publication on the discovery of a new ribosomally synthesized and post-translationally modified peptide (RiPP), phazolicin, which binds in the ribosome's exit tunnel[42]. The work was led by collaborator and first author Dmitrii Travin. I am the second author, contributing the cryo-EM structure and modeling work to determine the exact location of azole modifications and the mode of binding of phazolicin.

REFERENCES

1. Brenner, S. & Horne, R. W. A negative staining method for high resolution electron microscopy of viruses. *BBA - Biochim. Biophys. Acta* (1959). doi:10.1016/0006-3002(59)90237-9
2. Unwin, P. N. T. & Henderson, R. Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* (1975). doi:10.1016/0022-2836(75)90212-0
3. Taylor, K. A. & Glaeser, R. M. Electron microscopy of frozen hydrated biological specimens. *J. Ultrasructure Res.* (1976). doi:10.1016/S0022-5320(76)80099-8
4. Amos, L. A., Henderson, R. & Unwin, P. N. T. Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Progress in Biophysics and Molecular Biology* (1982). doi:10.1016/0079-6107(83)90017-2
5. Nguyen, C. & Gonen, T. Beyond protein structure determination with MicroED. *Current Opinion in Structural Biology* (2020). doi:10.1016/j.sbi.2020.05.018
6. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution

electron cryo-microscopy. *J. Mol. Biol.* (1990). doi:10.1016/S0022-2836(05)80271-2
7. Frank, J. Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* (1975). doi:10.1016/S0304-3991(75)80020-9
8. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* (2013). doi:10.1038/nmeth.2472
9. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.02.003
10. Campbell, M. G. *et al.* Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* (2012). doi:10.1016/j.str.2012.08.026
11. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* (2015). doi:10.7554/eLife.06980
12. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* (2019). doi:10.1107/S205225251801463X
13. Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* (2020). doi:10.1107/S2052252520000081
14. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.7 Å inside cells. *bioRxiv* 2020.06.05.136341 (2020). doi:10.1101/2020.06.05.136341
15. Russo, C. J. & Henderson, R. Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy* (2018). doi:10.1016/j.ultramic.2017.11.001
16. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* (2020).
17. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
18. Noeske, J. *et al.* High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. &Amp; Mol. Biol.* **22**, 336 (2015).
19. Polikanov, Y. S., Steitz, T. A. & Innis, C. A. A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat. Struct. Mol. Biol.* (2014). doi:10.1038/nsmb.2871
20. Lake, J. A. Ribosome structure determined by electron microscopy of Escherichia coli small subunits, large subunits and monomeric ribosomes. *J. Mol. Biol.* (1976). doi:10.1016/0022-2836(76)90200-X
21. Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J. Microsc.* (1987). doi:10.1111/j.1365-2818.1987.tb01333.x
22. Penczek, P. A., Grassucci, R. A. & Frank, J. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* (1994). doi:10.1016/0304-3991(94)90038-8
23. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. W. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* (2013). doi:10.7554/eLife.00461
24. Brown, A. & Shao, S. Ribosomes and cryo-EM: a duet. *Current Opinion in Structural Biology* (2018). doi:10.1016/j.sbi.2018.07.001
25. Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. in

Methods in Enzymology (2016). doi:10.1016/bs.mie.2016.04.012
26. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. W. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* (2018). doi:10.7554/eLife.36861
27. Punjani, A. & Fleet, D. J. 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *bioRxiv* 2020.04.08.032466 (2020). doi:10.1101/2020.04.08.032466
28. Zhong, E., Bepler, T., Berger, B. & Davis, J. CryoDRGN: Deep Generative Models for Reconstructing Heterogeneous 3D Structures from Cryo-electron Micrographs. *Microsc. Microanal.* (2020). doi:10.1017/s1431927620019492
29. Javed, A. & Orlova, E. V. Unravelling Ribosome Function Through Structural Studies. in *Subcellular Biochemistry* (2019). doi:10.1007/978-3-030-28151-9_3
30. Loveland, A. B., Demo, G. & Korostelev, A. A. Cryo-EM of elongating ribosome with EF-Tu•GTP elucidates tRNA proofreading. *Nature* (2020). doi:10.1038/s41586-020-2447-x
31. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* (2018). doi:10.7554/eLife.42166
32. Watson, Z. L. *et al.* Structure of the bacterial ribosome at 2 Å resolution. *Elife* (2020). doi:10.7554/ELIFE.60482
33. Han, B.-G. *et al.* Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. *J. Struct. Biol.* **195**, 238–244 (2016).
34. Crucifix, C., Uhring, M. & Schultz, P. Immobilization of biotinylated DNA on 2-D streptavidin crystals. *J. Struct. Biol.* (2004). doi:10.1016/j.jsb.2004.02.001
35. Wang, L., Ounjai, P. & Sigworth, F. J. Streptavidin crystals as nanostructured supports and image-calibration references for cryo-EM data collection. *J. Struct. Biol.* (2008). doi:10.1016/j.jsb.2008.07.008
36. Han, B.-G. *et al.* Electron microscopy of biotinylated protein complexes bound to streptavidin monolayer crystals. *J. Struct. Biol.* **180**, 249–253 (2012).
37. Han, B.-G., Watson, Z., Cate, J. H. D. & Glaeser, R. M. Monolayer-crystal streptavidin support films provide an internal standard of cryo-EM image quality. *J. Struct. Biol.* (2017). doi:10.1016/j.jsb.2017.02.009
38. Ward, F. R., Watson, Z. L., Ad, O., Schepartz, A. & Cate, J. H. D. Defects in the Assembly of Ribosomes Selected for β-Amino Acid Incorporation. *Biochemistry* (2019). doi:10.1021/acs.biochem.9b00746
39. Melo Czekster, C., Robertson, W. E., Walker, A. S., Söll, D. & Schepartz, A. In Vivo Biosynthesis of a β-Amino Acid-Containing Protein. *J. Am. Chem. Soc.* **138**, 5194–5197 (2016).
40. Davis, J. H. *et al.* Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell* **167**, 1610-1622.e15 (2016).
41. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* (2019). doi:10.1107/S2059798319011471
42. Travin, D. Y. *et al.* Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition. *Nat. Commun.* (2019). doi:10.1038/s41467-019-12589-5

Chapter 2: Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules

Reproduced with permission from: Han BG, Watson Z, Kang H, Pulk A, Downing KH, Cate J, Glaeser RM. J Struct Biol. 2016;195(2):238-244. doi:10.1016/j.jsb.2016.06.009

ABSTRACT

We describe a rapid and convenient method of growing streptavidin (SA) monolayer crystals directly on holey-carbon EM grids. As expected, these SA monolayer crystals retain their biotin-binding function and crystalline order through a cycle of embedding in trehalose and, later, its removal. This fact allows one to prepare, and store for later use, EM grids on which SA monolayer crystals serve as an affinity substrate for preparing specimens of biological macromolecules. In addition, we report that coating the lipid-tail side of trehalose-embedded monolayer crystals with evaporated carbon appears to improve the consistency with which well-ordered, single crystals are observed to span over entire, 2 μm holes of the support films. Randomly biotinylated 70S ribosomes are used as a test specimen to show that these support films can be used to obtain a high-resolution cryo-EM structure.

INTRODUCTION

There are a number of reasons to consider using streptavidin monolayer crystals as "affinity" support films for cryo-electron microscopy (cryo-EM). Macromolecules of interest can be easily tagged with biotin or a streptavidin-binding peptide and then bound to streptavidin (SA) with high affinity and specificity. Furthermore, tagging followed by affinity binding is expected to pose less risk to the native structure of the macromolecule than does (1) adsorption of particles to the surface of carbon film, even when rendered hydrophilic by exposure to a glow discharge, or (2) repeated collision with the air-water interface that occurs when freely diffusing macromolecules are confined to a thin aqueous film[1].

Monolayer crystals of SA have been considered previously by several authors for use as an affinity-support film. One early study viewed SA as being a "general adaptor" for linking any kind of biotinylated molecule to a lipid monolayer[2]. Chemically biotinylated ferritin was used in that work to show that a high density of randomly distributed particles could be bound to 2-D crystals of SA. In an extension of the adaptor-molecule idea, Crucifix et al. first randomly decorated SA monolayer crystals with biotinylated dsDNA molecules, and then used the immobilized DNA as bait to bind yeast RNA Pol I particles[3]. Wang et al. showed that biotinylated proteoliposomes could be bound at high density[4,5], and they introduced the further innovation of eliminating the periodic background due to SA by masking out the Bragg reflections in the computed Fourier transforms of images. Han et al. then went on to demonstrate the generality with which chemical biotinylation of soluble-protein complexes could be used[6].

In spite of these promising demonstrations, SA monolayer crystals have not been adopted as support films for routine data collection. Our recent attempts to do so made it clear that two major problems remained. (1) The standard protocol for growing monolayer crystals involves an

additional incubation step of 2 hours or more[7], which both slows and complicates the process of preparing cryo-EM specimens. (2) In addition, while the results can be excellent, the growth of large, well-ordered crystals on micro-wells, together with their transfer onto EM grids, is quite inconsistent.

We now describe a simplified, on-grid crystallization protocol that yields large SA crystals in times as short as 10 minutes. In addition, we demonstrate that trehalose-embedding makes it possible to prepare these grids in advance, with their useful shelf life expected to be months or longer. We also find that a thin layer of evaporated carbon can be deposited on the back side (lipid-tail side) of the trehalose-embedded SA crystals in order to add mechanical stability. In a practical test of these grids, *E. coli* 70S ribosomes were used to obtain a 3-D reconstruction at a global resolution estimated to be ~4.0 Å, which improved to ~3.9 Å when focused refinement was used for the large subunit.

MATERIALS AND METHODS

Lipids

The biotinylated lipid used here is 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-(cap biotinyl), supplied as a 10 mg/mL solution in chloroform/methanol/water (Avanti Polar Lipids). This was diluted to 1.0 mg/mL with a solution of chloroform/methanol/water and aliquoted into small volumes intended for a single usage. The aliquots were sealed under nirogen gas and stored at -80 °C. No deterioration as a function of time was observed in the ability of such aliquots to produce high-quality streptavidin monolayer crystals. Nevertheless, as a precaution, we prepare new aliquots after a period of about 6 months.

Streptavidin

Streptavidin (SA) was purchased from New England Biolabs (catalog number N7021S). This sample is provided at a concentration of ~1 mg/mL, dissolved in 10 mM sodium phosphate pH 7.2 with 0.15 M NaCl. This was aliquoted in quantities intended for single use, frozen in liquid nitrogen, and stored at -80 °C. Similar to what we do for the lipid, as a precaution, we prepare new aliquots of streptavidin after a period of about 6 months.

Protocol for growing monolayer crystals directly on holey-carbon EM grids

A lipid monolayer, cast on an air-water interface, is first picked up by touching the lipid from above with a hydrophobic, holey-carbon EM grid. This results in Langmuir-Schaefer transfer of patches of the monolayer that span the holes of the carbon film, as was discovered by Kubalek et al.[8]. We presume that an additional, unwanted lipid monolayer is also transferred to the air-water surface of the small volume of buffer that adheres to the (now hydrophilic) face of the EM grid. Thus, to remove as much lipid as possible from the surface of the adhering drop, we touch the grid to three successive, 50 µL drops of subphase buffer sitting on parafilm. Next, SA is added to the small droplet of buffer that adheres to the face of the EM grid. The grid is then incubated long enough to allow binding and subsequent crystallization of SA. We refer to this technique as the "on-grid" crystallization method. **Figure S1** shows two photographs that further illustrate the steps just described.

We cast the above-mentioned lipid monolayer on a small trough holding ~5 mL of subphase buffer consisting of 50 mM HEPES (pH 7.5), 0.15 M KCl, and 10 % trehalose (Swanson Health Products). Including trehalose in the subphase buffer serves only to eliminate the need for

a buffer-exchange "wash" before air-drying (see the following section). This simplification became possible after we observed that adding 10% trehalose has no effect on the quality of monolayer crystals of streptavidin. Before applying lipid, however, we first "dust" the air-water interface lightly with unscented talcum powder, and then we apply a small droplet (~10 μL) of castor oil at the center. As the castor oil spreads, the advancing front of oil sweeps away contaminants at the air-water interface and compresses them to the perimeter of the trough. A Hamilton syringe is then used to deliver ~0.5 μL of previously aliquoted lipid to the center of the trough. At this point the thin film of castor oil also serves as a "piston" to control the surface pressure of the lipid monolayer[9]. We believe that the castor-oil piston is a desirable but not necessary element of our protocol. Using a trough, rather than smaller, individual wells, facilitates use of the castor-oil piston, and allows one to prepare multiple grids from a single lipid monolayer.

We wash Quantifoil grids by first dipping them into chloroform and then into 95% ethanol. We also apply an additional ~5 nm of evaporated carbon to the top side of Quantifoil grids and then allow these to "age" for at least three days to make the freshly evaporated carbon more hydrophobic. Just before use, we again wash these grids by dipping them into 95% ethanol. We have less experience with C-flat grids, but we have successfully used them as received, i.e. without depositing additional carbon. As is mentioned in the Supplemental Material, in the section "*Issues still to be addressed* ", we prefer to use gold or molybdenum grids to copper grids.

In our current crystallization protocol, we dilute an aliquot of SA to a concentration of 0.2 mg/mL with subphase buffer, and we use 4 μL of diluted SA for each grid. After applying SA to the EM grid, the crystals are grown within a humidity chamber. Care is taken to minimize evaporation by placing crushed ice on the tweezers, with the intent to cool the grid slightly below the ambient dew point. Whereas an incubation time of 1-2 minutes appears to be too short to ensure full crystal formation, we observed crystallization to be completed within 10 minutes.

Following an incubation time that, for convenience, is often about half an hour when doing many grids at a time, most of the unbound SA is washed away by placing the grid on top of a 200 μL drop of wash buffer. The composition of the wash buffer is 10% trehalose in 10 mM HEPES (pH 7.5), as before, but with the KCl concentration now reduced to 50 mM. After waiting several seconds, the floating grid is caught with a tweezers and lifted vertically until it separates from the wash drop. As noted further in the Supplemental Material, we believe that this step in the protocol may be the one in which the monolayer crystals of SA are at greatest risk of becoming fragmented or even lost completely. Even when crystals have been severely damaged (at this stage, as we believe), most, if not all, holes in the holey carbon film still remain covered by a monolayer of SA bound to biotinylated lipid.

Streptavidin crystals are then embedded in trehalose and backed with evaporated carbon

After washing the grids to remove unbound SA, excess trehalose solution is gently "wicked off" by touching the edge of the grid to a piece of filter paper. The grid then is left on a filter paper with the wet side facing the air, and any remaining liquid on the grid is allowed to dry.

After the trehalose solution has dried, a thin layer (~5 nm or less) of evaporated carbon is deposited on the back side (lipid-tail side) of the EM grid. Grids are placed ~20 cm below a carbon-arc source, and evaporation is done at a vacuum of 10^{-5} torr or lower. We use carbon rods with a tip width of 1 mm (Ted Pella catalog number 62-107), which require less power to evaporate, in order to minimize heating of the trehalose-embedded SA "target". To further minimize the risk of radiant heating, the carbon rod is heated very rapidly, resulting in breakage of the thin tip and a

concomitant flash of evaporated carbon. A cartoon showing the structure of the resulting carbon-backed, trehalose-embedded streptavidin support films is presented in **Supplemental Figure S2**.

We store the carbon-backed, trehalose embedded SA crystals at room temperature in a sealed container. We prefer to store grids over silica gel that is pink (but not white) to maintain a relatively constant value of humidity.

Just before use, the grid is rehydrated by touching to two successive, 50 µL drops of a solution of 10 mM HEPES (pH 7.5) with 150 mM KCl without trehalose and then left on a 100 µL drop of the same solution for 10 minutes. This same process is repeated a second time, with the intent being to rinse away all remaining trehalose. After that, the grid is further washed with whatever buffer is optimal for the macromolecular sample under investigation.

Preparation of grids and electron microscopy of 70S ribosomal particles

Ribosomes were purified from E. coli strain MRE600 using sucrose gradient centrifugation, as previously described[10]. Ribosome complexes were formed by incubating 1.5 µM deacylated $tRNA^{Phe}$ and 3 µM mRNA of sequence 5'-GGCAAGGAGGUAAAAUUCUACAAA-3' (Thermo Scientific) with 0.5 µM ribosomes at 37 °C for 15 minutes in the buffer A: 20 mM HEPES, pH 7.5/ 70 mM KCl/ 6 mM MgOAc/ 1 mM TCEP. The antibiotic spectinomycin (Sigma-Aldrich) at a concentration of 20 µM was added to the pre-formed complex and incubated an additional 10 minutes at 37 °C.

Ribosomes were biotinylated by adding 5-fold excess of biotin-labeling reagent (Solulink Catalog No. B-1007-110) and incubated 20 minutes at room temperature. Excess unreacted biotin was removed by using 1 mL S300 Sephacryl (GE Healthcare) gel-filtration spin column. The column was equilibrated with Buffer A, and usually 50 µL of ribosome complex was loaded on the column and spun on a table centrifuge at 3200 rpm for 1 minute.

Grids were first washed 3 times with 50 µL drops of cold ribosome buffer, and then a 4 µL aliquot of ribosomes was applied at a concentration in the range of 20-40 nM. After incubating for 20 minutes, chilled and in a humidity box to minimize evaporation, as described above for growth of monolayer crystals, unbound ribosomes were washed away by touching to three successive, 50 µL drops of buffer.

To prepare negatively stained specimens, we added 4 µL of 2% uranyl acetate to the lens of buffer adhering to the rinsed grid, which then was mixed by repeated, gentle pipetting while on the grid. This was followed by two cycles in which 3 µL was removed from the grid and 4 µL of 2% uranyl acetate was again added. After the second cycle, the excess uranyl acetate solution that remained was removed by blotting with filter paper. Images were recorded with a Gatan CCD camera on a JEOL 1200 electron microscope.

To prepare cryo-EM samples, the grid is transferred to the Vitrobot tweezers after washing away unbound ribosomes with ribosome buffer. The tweezers and grid were then loaded into the Vitrobot Mark IV chamber, which was previously equilibrated at a temperature of 15 °C and a relative humidity setting of 100%. In order to standardize the volume of liquid on the grid before blotting, excess liquid was first wicked off by touching the bottom edge of the grid with filter paper, which was brought in through a side port of the Vitrobot chamber. Following this, 1.2 µL of sample buffer was then added to the wet face of the grid. The blotting time used was 5 s, with a force setting of 8, a blotting time of 3 seconds, and zero wait/drain time.

For routine evaluation of cryo-grids, images were recorded with a Gatan CCD camera on a Philips CM 200. High-resolution images were obtained with a Gatan K2 camera on an FEI low-base Titan, using a Gatan cryo-holder. In this latter case, images were recorded as dose-

fractionated movies consisting of twenty 300 ms frames, each with an exposure of 1.2 electrons/Å^2 at the specimen. The pixel size in these images was 1.3 Å, referred to the specimen. The movie frames were aligned and summed with the motion-correction software developed by Li et al.[11].

Data processing

The "2dx" software package[12] was used to unbend the SA lattice in a few of the images. This was done to sharpen the Bragg spots in the Fourier transforms, thereby increasing the resolution at which spots could be detected with a good signal-to-noise ratio.

In all other cases, Fourier filtration of Bragg peaks was used to remove the image of the SA crystal, without unbending, before particles were boxed for further analysis. To do this, a script, available upon request, was written in MATLAB to identify all pixels where the magnitude of the Fourier transform of an image was higher than a user-defined value. The Fourier-transform magnitudes in such pixels, plus those in several adjacent pixels, were replaced by the average value of the surrounding background, and the phases were replaced by random values. The threshold value was adjusted manually, while looking at a display of the Fourier transform of an image, until all visible diffraction spots were removed.

Candidate ribosome particles were automatically boxed with a software tool provided in RELION[13]. Images, with candidate particles outlined, were edited manually, using the BOXER tool provided in EMAN[14] to remove initial candidates that were obviously aggregates or other undesired material. Three-dimensional classification of particles and subsequent refinement again used tools provided in RELION.

RESULTS

Characterization of the carbon-backed SA monolayer crystals

Obtaining monolayer crystals by the on-grid technique is quite reliable and reproducible. On the basis of our current work using LEGINON for automated data collection[15], we estimate the success rate to be well over 90 percent for getting images in which single SA crystals cover the full field of view. The estimated ice thickness was the only criterion used when selecting areas to be added to the data-collection queue, and no attempt was made to decide, in advance, whether the holes contained SA crystals. Nevertheless, until the monolayers have been backed with evaporated carbon, our experience indicates that their crystallinity may still be at considerable risk. More information on factors that we believe can cause preparation of these grids to fail is presented in the Supporting Material.

The fact that these SA crystals remain well ordered after carbon coating, removal of trehalose, and application of a sample has been established by recording high-resolution images. One such image is shown in **Figure 1A**, and the FFT of this raw, unprocessed image is shown in **Figure 1B**. In addition to showing clear Thon rings from the evaporated carbon, the FFT also shows many Bragg peaks from the streptavidin monolayer crystal. The Fourier transform of the unbent image is shown in **Figure 1C**, and an enlarged version is shown in **Figure S3**. The Bragg spots in the FFT of the unbent image are represented by squares of various sizes, the largest corresponding to an IQ[16] of 1 (estimated phase error of ~4 degrees) and the smallest corresponding to an IQ of 7 (estimated phase error of ~45 degrees). As in this example, Bragg spots with an IQ of 3 (estimated phase error of ~14 degrees) commonly extend to a resolution of 4 Å.

Finally, these grids are easy to prepare and use. The incubation time required to grow streptavidin-monolayer crystals by the on-grid technique is much shorter than the ~2 hours or more

that has been recommended for crystallization by the micro-well technique[7]. While we often use the SA support films within a week after first making them, we have confirmed that the SA lattice shows no signs of deterioration when stored for a month. We thus expect that the room-temperature shelf life will prove to be much longer than that.

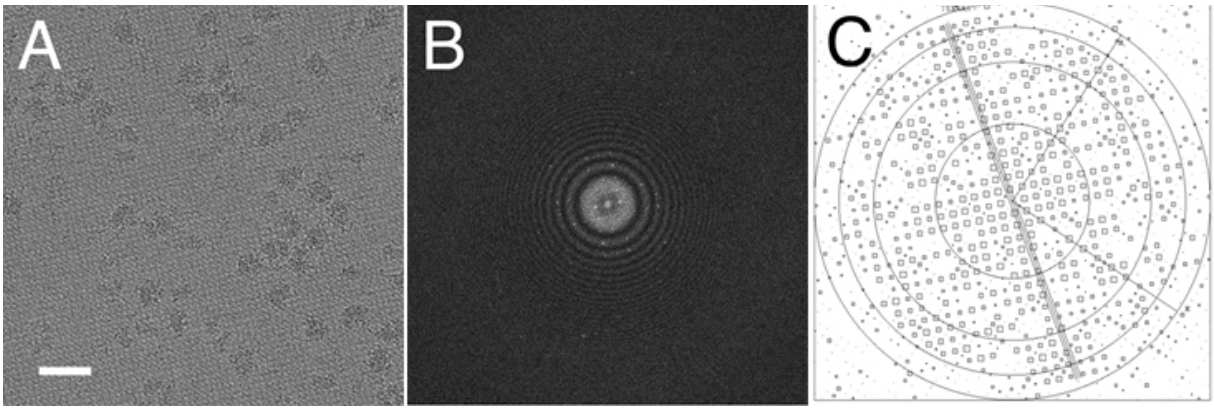


Figure 1. Demonstration that the SA monolayer crystals remain well ordered after trehalose embedding, backing with evaporated carbon, and subsequent removal of trehalose just before use. Panel (**A**) shows a cryo-EM image recorded with a K2 camera on a FEI Titan electron microscope, operated at 300 keV. In this case, images were recorded as movies consisting of 20 frames, using an electron exposure of 1.2 electrons/Å^2 per frame. Movie frames were aligned with the software developed by Li et al.[11]. Panel (**B**) shows the raw Fourier transform of this image. Panel (**C**) shows the Fourier transform of the "unbent" image, in which the Bragg spots are displayed as a so-called IQ plot[16], and a larger version of this panel is shown as Figure S3. The 2dx software package[12] was used to unbend the lattice. Indexing of the lattice is indicated by the H and K coordinate system, and the orientation of the specimen-tilt axis, as well as the tilt angle, is also shown in the figure. The signal-to-noise ratio at the reciprocal-lattice points is represented by square boxes of different sizes. Circles are shown at resolutions of 9 Å, 5 Å, 4 Å, and 3.5 Å, respectively. Numerous diffraction spots with IQ values of 4 or less (indicating an expected phase error of 22 degrees or less) are seen out to a resolution of better than 4 Å.

Use of randomly biotinylated 70S ribosome particles as a test specimen

Figure 2B shows that a high density of ribosome particles is obtained for the conditions of sample concentration and subsequent grid-washing described in the Materials and Methods section. In a corresponding control experiment, using ribosomes that had not been biotinylated, few ribosomes remain on the support film, as is shown in **Figure 2A**. We thus conclude that the biotin-binding functionality of streptavidin is retained throughout the process of trehalose embedding, carbon evaporation, and storage.

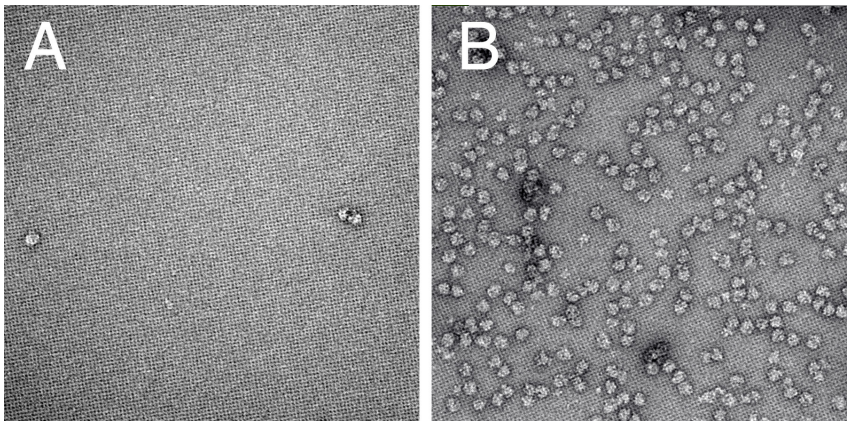


Figure 2. Images demonstrating the effectiveness of binding randomly biotinylated 70S ribosomes to the SA monolayer-crystal support film. (**A**) This panel shows that only a sparse density of ribosomes was obtained in the control experiment, for which the ribosomes were not biotinylated. (**B**) By comparison, a uniform, high density of ribosomes was obtained when the ribosomes were biotinylated.

A set of 1497 movies were recorded in order to test the use of SA support films for single-particle data collection. After removing the SA background from these images, candidate ribosome particles were boxed automatically. This raw set was edited manually to produce a working data set of 101,213 particles. 3-D classification was then used to obtain structurally more homogeneous subsets of the working set.

Refinement of the largest such subset, consisting of 52,433 particles, produced the density map shown in **Figure 3A**. The overall resolution of this map is estimated to be 4.0 Å, based on the point at which the "gold standard" FSC curve, shown in **Figure S4**, falls to a value of 0.143. Focused refinement, based on the 50S subunit, improved this only slightly, to just under 3.9 Å. An example of the quality of the map in the interior of the large subunit is shown in **Figure 3B**.

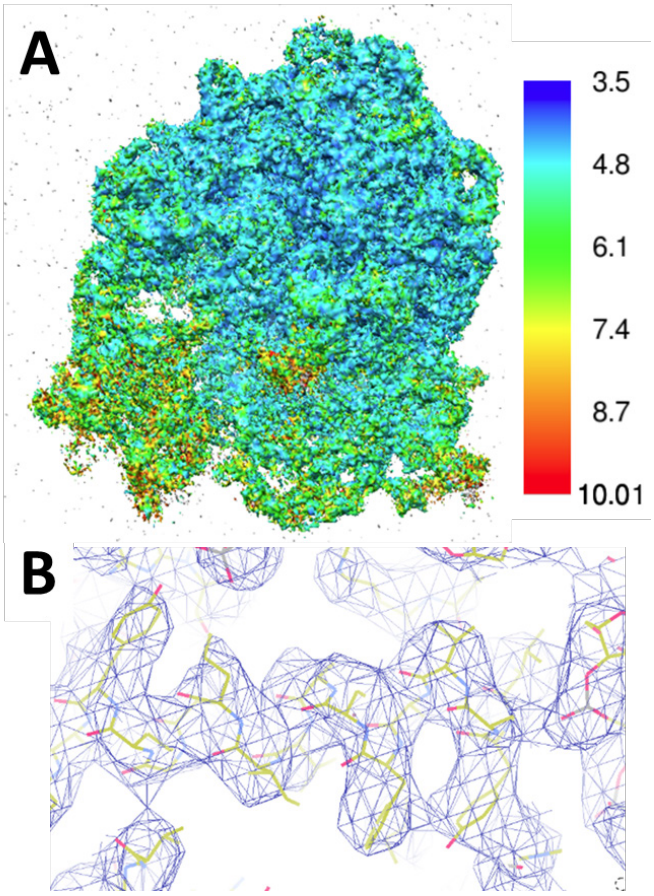


Figure 3. Density map of E. coli 70S ribosome particles obtained when using SA monolayer crystals as a support film. (**A**) Surface representation of the map, color-coded by the local resolution, with the color scale shown to the right. (**B**) Example of the map in the best region, with an atomic model of the structure fitted to the map.

A plot of the Euler angles assigned during refinement is shown in **Figure 4**. These assignments are not distributed uniformly over the surface of the unit sphere, nor do we believe that they ought to be. As is shown in **Figure S5** of the Supplemental Material, for example, a significant fraction of the surface of 70S particles is made up of RNA. Since no biotinylation of RNA can occur, we do not expect particles to bind in the corresponding orientations. In practice, the areas of heaviest assignment of Euler angles correlated roughly, but not always precisely with orientations where there are lysine residues that can be biotinylated. In addition, large portions of the Euler angle plot in **Figure 4** have only one or two particles that were assigned to each point. As described by results shown in **Figure S6** in the Supplemental Material, we are currently investigating whether this low-frequency, uniform assignment of orientations may reflect errors that are occasionally made in the process of projection matching, rather than non-specific binding.

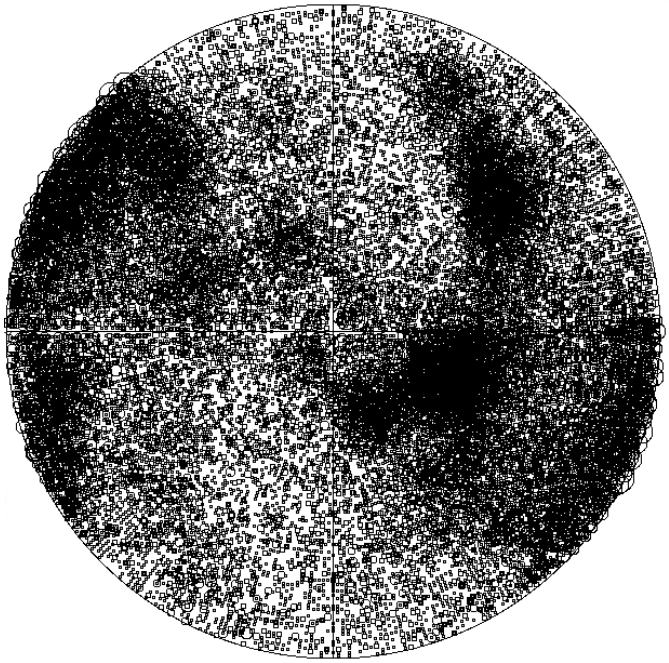


Figure 4. Plot of the angular orientations assigned to 52,433 ribosome particles during refinement of the 3-D density map. Assignments are confined mainly to broad, continuous patches. The remaining particles are distributed quite uniformly, but usually with only one or two particles assigned per point on the unit sphere. As is described in Results, and again in the Discussion, biotinylation (and thus binding) is not expected for those parts of the surface that are made up of RNA. We are currently investigating whether this low-frequency, uniform assignment of orientations might not be accurate, but rather may reflect major errors that are occasionally made in the process of projection matching.

DISCUSSION

Streptavidin affinity-grids can be used to obtain high-resolution structures

The resolution of the 3-D density map that we obtained for *E. coli* 70S ribosomes, using SA monolayer crystals as a support film, is in line with that of maps obtained previously by others, using continuous carbon as a support film. A map at a somewhat lower resolution of ~5.1 Å was obtained by Bai et al.[17] for *T. thermophilus* 70S ribosomes, but from a much smaller data set consisting of 15,202 particles. On the other hand, significantly higher-resolution maps have been obtained, this time for *E. coli* 70S ribosomes, when using much larger data sets. A 3.0 Å map was obtained from 164,353 particles by Brown et al.[18], and an even higher-resolution structure was obtained by from 417,201 particles[19].

It is natural to be concerned that the periodic motif of the SA crystal might interfere with alignment and assignment of Euler angles for boxed particles. In principle, this should not be a problem because this background is easily removed by Fourier filtration when it is no longer of value to have it there[5,6]. Our success in obtaining a high-resolution density map of the 70S ribosome particle confirms that removal of the periodic SA motif by Fourier filtering was as effective as it was expected to be.

Another concern is that nearly irreversible affinity binding could, in principle, trap the structure of a large complex in one or more off-pathway states. This could happen if some biotin residues are normally inaccessible for binding, for example those that are coupled to lysine residues within deep indentations of the surface. Transient fluctuations of the structure might nevertheless bring such biotin residues sufficiently far out that they would have access to a

streptavidin binding site. We can again say that our success in obtaining a high-resolution density map rules out the hypothesis that the trapping of non-native structural variants contributes significantly to the structural heterogeneity of the data set.

Random biotinylation can result in particles adopting many orientations

One of the major concerns about using affinity grids is that particles might be bound in a preferred orientation with respect to the support film. Chemical biotinylation of one or a few lysine residues per particle, used here, addresses this concern by placing affinity ligands as uniformly as possible over the surface of the particle.

Although the distribution of particle orientations in our case is clearly not uniform, there is sufficient coverage to adequately sample the 3-D Fourier transform of the ribosome particle. Fortunately, not all particle orientations need to be presented for the resulting 3-D structure factors to still be sampled completely. In fact, a complete sampling of Fourier space already occurs for rotation of an object about a single axis. In that case, the location of all projection directions is restricted to a great circle on the unit sphere. The Euler-angle distribution shown in **Figure 4** is actually far more complete than that.

As was indicated in Results, the distribution of surface locations where there are potential biotinylation sites agrees only partially, but not completely, with the pattern of densely populated Euler angles. The reasons for there being discrepancies can be complex, and the explanations, once determined, may not be generalizable. Variations in reactivity with the biotinylation reagent (which could be determined by mass spectroscopy of peptide fragments) could, for example, be one factor. Flexibility of protein subunits that are biotinylated is very likely to be another factor. In the latter case, flexible tethering of particles to the SA lattice might allow them to search locally for nonspecific, but preferred binding interactions that would, otherwise, be too weak to produce significant binding.

In addition, large portions of the Euler angle plot in **Figure 4** have only one or two particles that were assigned to each point. As described in the Supplemental Material, we are currently investigating whether an almost uniform background distribution, in which there are only one or two particles at a given angle, may reflect errors that are occasionally made in the process of projection matching, rather than binding that was not mediated by biotinylation – i.e. non-specific binding.

Significant benefits are expected from using streptavidin affinity-grids

Affinity support films in general provide a number of features that are expected to be helpful for preparing cryo-EM grids. These include the fact that particles are bound at a single plane (single Z-height, for untilted specimens), which is important for CTF correction at high resolution. In addition, tethered particles are prevented from interacting with the air-water interface, provided that the aqueous sample is not too thin.

Affinity binding is, in general, expected to leave structures in a close-to-native state. Streptavidin affinity grids have the special feature that binding of the biotin ligand is essentially as strong as a covalent bond.

In addition, the background image of the SA lattice might serve as a fiducial to aid alignment of movie frames, and even sub-regions of frames. The background image of the SA lattice might also serve to track changes in tilt orientation that occur during the course of recording dose-fractionated images.

Value is added by the long shelf-life feature of these affinity grids

By separating the step when these affinity grids are made from the time when they are actually used, the process of making cryo-specimens becomes little more complicated than it is for using just holey-carbon grids. It is true, of course, that the process of making these affinity grids is still somewhat time consuming. It thus is very helpful that it is possible to make enough of them in advance that the supply does not run out in the middle of a session of freezing specimens.

The step of preserving SA monolayer crystals over open holes required some optimization of the trehalose concentration, in combination with the blotting technique (i.e. wicking from the edge of the grid), used to leave a residual volume of solution to dry. The requirement for optimization is not too surprising in light of the fact that millimeter-scale droplets of aqueous trehalose can leave complex patterns of convection (colloidal-particle deposition), thickness, and stress (cracking) after drying by evaporation[20]. While we have found the protocol described here to be very reliable and simple to use, it is likely that other variations may work equally well. Indeed, spin-coating may be especially well-suited as a way to spread a thin film of trehalose[21].

Trehalose is known to have superior properties for room-temperature preservation of freeze-dried phospholipid vesicles and proteins[22]. It has also been used extensively to preserve the high-resolution structure of thin protein crystals adsorbed to a carbon support film, a few examples of which are described in the literature[14,23,24]. We note that other hydrophilic solutes might also work as well as does trehalose – see, for example, section 6.4 of Glaeser et al.[25].

ACKNOWLEDGMENTS

This work has been supported in part by NIH grants R01 GM083039, P01 GM051487, and R01 GM065050.

SUPPLEMENTAL MATERIAL

I. FURTHER DETAILS ABOUT HOW THE LONG SHELF-LIFE STEPTAVIDIN AFFINITY GRIDS ARE MADE

Please refer to the MATERIALS AND METHODS section of the main paper for the context in which **Figures S1** and **S2**, below, were mentioned.

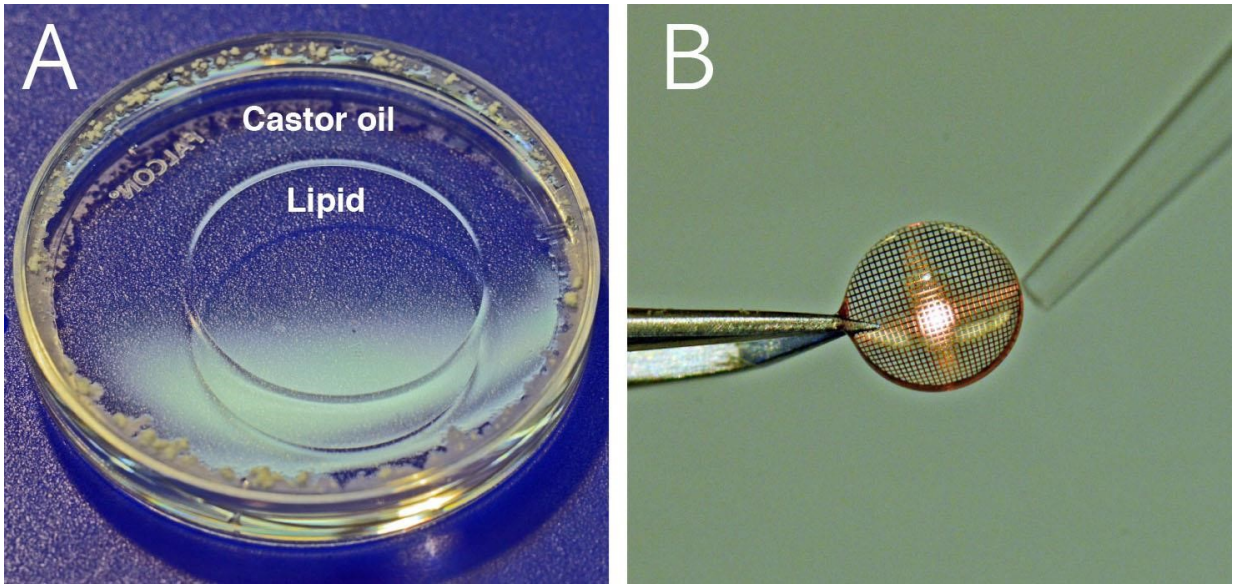


Figure S1. Illustration of two major steps of the "on-grid" crystallization technique. As indicated in (**A**), after applying a light dusting of talcum powder (optional, to mark the initial air-water interface), approximately 10 µL of castor oil is spread over the surface of buffer. The castor oil initially creates a clean oil-water interface, sweeping and compressing any surface contaminants to the perimeter of the trough, along with the talcum powder. We routinely use the lid of a 35 mm disposable petri dish as the "trough". A lipid monolayer is then spread by applying ~0.5 µL of previously aliquoted lipid at the center of the trough, using a Hamilton syringe to deliver the volatile organic solution at the center of the thin layer of oil. As is shown here, the contact line between the resulting lipid monolayer and the thin film of castor oil is visible in reflected light, especially when viewed against a dark background. A hydrophobic, holey-carbon grid is touched to the air-lipid interface and then removed, resulting in Langmuir-Schaefer transfer of the lipid monolayer spanning the open holes in the carbon film. Note that we use a so-called "anticapillary" tweezers, and note the orientation of the carbon-coated surface of the grid with respect to the asymmetric points of the tweezers. As shown in (**B**), a thin lens of subphase buffer now adheres to the formerly hydrophobic face of the grid, into which 4 µL of streptavidin solution is injected. The grid, still held in the EM tweezers, is then incubated for at least 10 minutes in a high-humidity chamber to allow time for streptavidin tetramers to bind to the biotinylated lipid and form monolayer crystals.

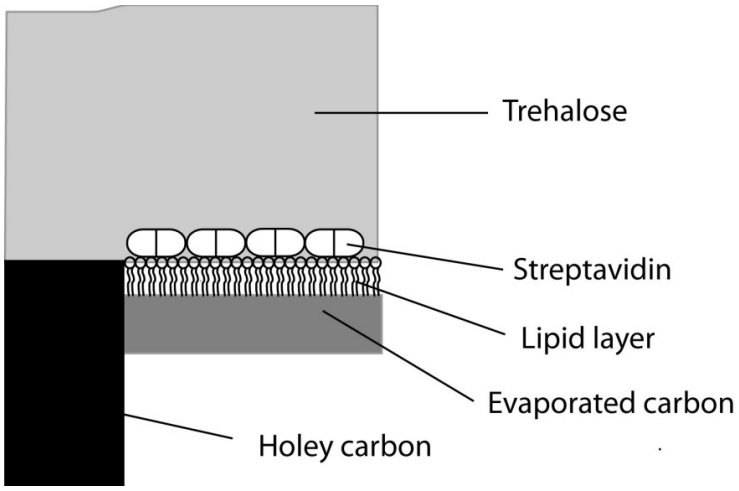


Figure S2. Schematic cartoon illustrating the structure of the long shelf-life SA support films. This drawing shows only an area that is close to, and includes one edge of, a hole in the carbon film. The EM grid bars, not shown here, would be below the holey-carbon film. While the location of the lipid monolayer relative to the surface of the holey-carbon film is not known in molecular detail, one possible arrangement is for the hydrophobic lipid tails to be in contact with the edge of the hydrophobic carbon, possibly similar to the way shown in this cartoon. Immediately after the SA crystals have been grown, they are embedded in a thin film of trehalose. The thickness of the trehalose embedment is not known experimentally, but it is likely to be as much as 100 nm or more. A layer of carbon, estimated to be 5 nm or less in thickness, is then evaporated onto the lipid tails. Just before use, the trehalose is washed from the SA crystals, which remain bound to the lipid monolayer, which in turn is stabilized by the layer of evaporated carbon.

II. ENLARGED VERSION OF THE IQ PLOT SHOWN IN FIGURE 1

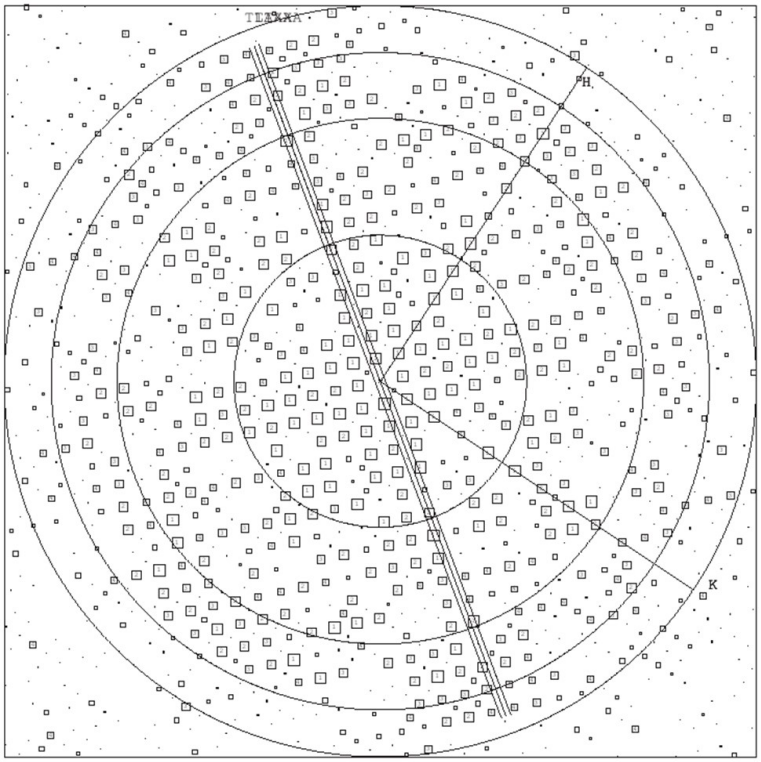


Figure S3. A copy of **Figure 1C**, enlarged here to make it easier to recognize the differences in IQ values for different Bragg spots.

III. FOURIER SHELL CORRELATION (FSC) CURVES FOR THE REFINED STRUCTURE OF THE 70S RIBOSOME OBTAINED USING THE STREPTAVIDIN AFFINITY GRIDS

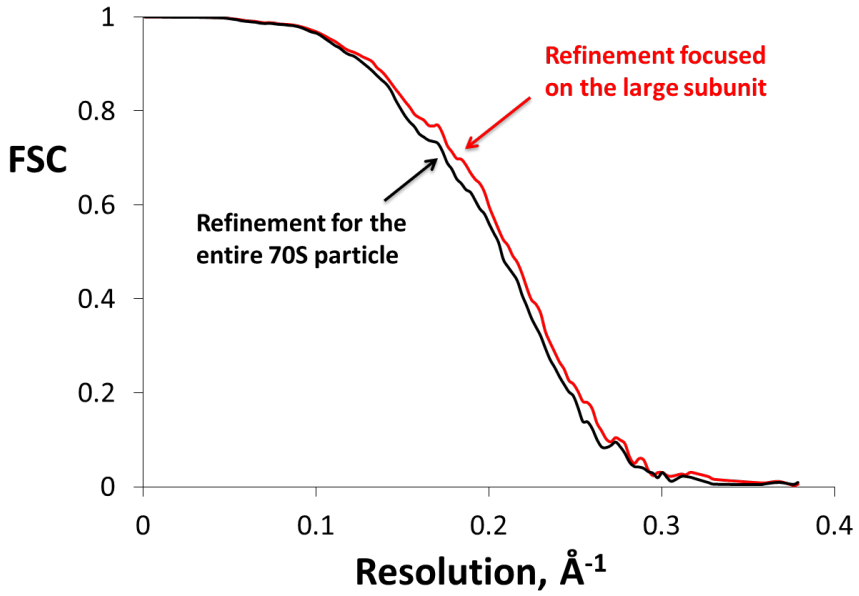


Figure S4. Fourier shell correlation (FSC) curves for the refined density map of the 70S ribosome particle, produced with the RELION software package. The value of the resolution at the point where the FSC falls to 0.143 is 4.0 Å when the entire particle is used for refinement, and this improves to 3.9 Å when the refinement is focused on the large subunit.

IV. DISTRIBUTION OF POTENTIAL SITES FOR BIOTINYLATON ON THE SURFACE OF THE 70S RIBOSOME

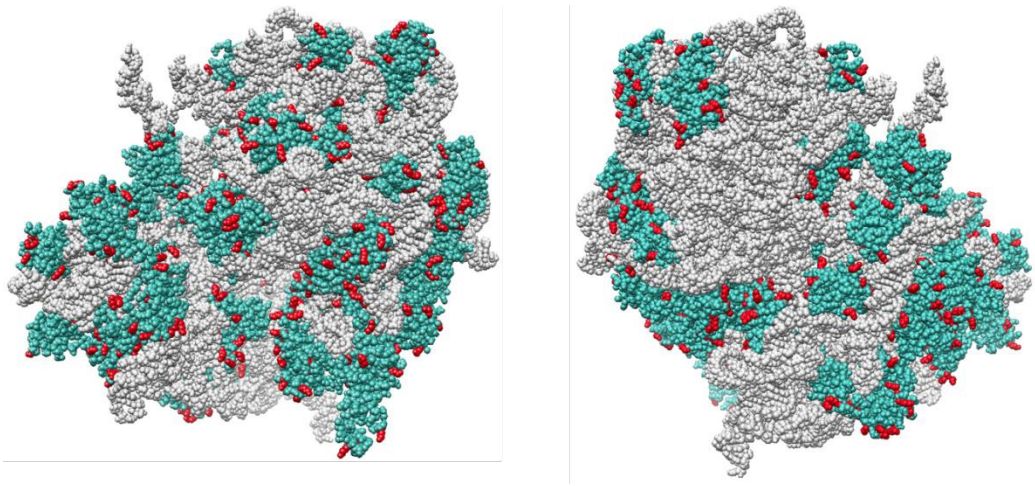


Figure S5. Two views of the surfaces, front and back respectively, of a model of the *E. coli* 70S ribosome particle. Red balls represent lysine residues, cyan represents the rest of the protein, and white (gray) represents RNA. The atomic model on which this surface image is based was obtained by separately docking atomic models of the 30S and 50S subunits into our experimental density map. The 30S model used was Protein Data Bank (PDB) file 4ADV, and the 50S model used was Protein Data Bank (PDB) file 3J7Z.

V. FURTHER ANALYSIS OF THE DISTRIBUTION OF EULER ANGLES DURING REFINEMENT

The distribution of Euler angles shown in **Figure 4** is not uniform. Instead, the orientations are clustered in diffuse islands. Nevertheless, **Figure S6** shows that there is a nearly uniform distribution of orientations to which only one particle was assigned. However, even the orientations with a low occupancy (one or two particles) show a preference to be in the highly occupied orientations.

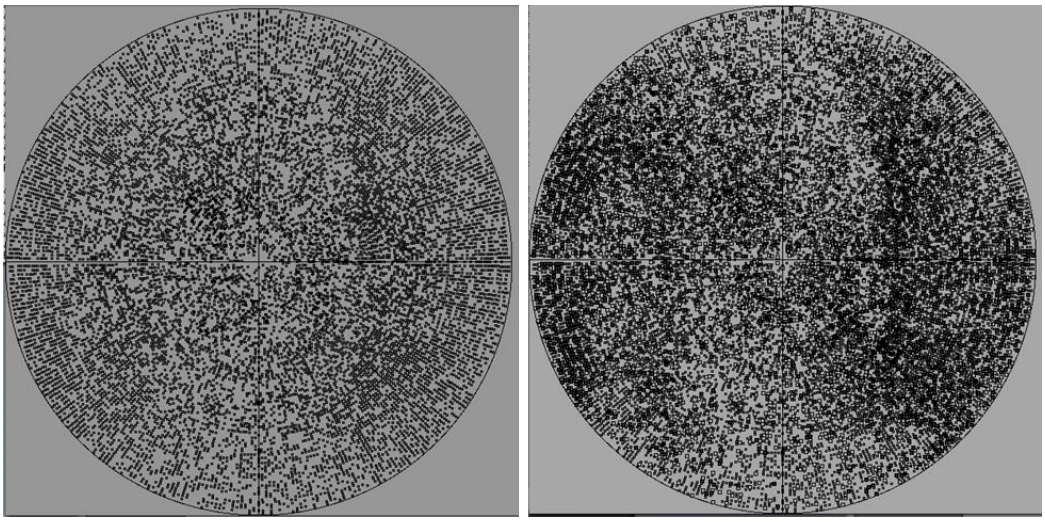


Figure S6. Representation of the Euler-angle distribution. Black dots show (left) the locations where only one particle has been assigned that orientation and (right) the locations where either one particle or two particles had been assigned that orientation.

VI. CARBON-BACKED STREPTAVIDIN MONOLAYER CRYSTALS TOLERATE DETERGENT

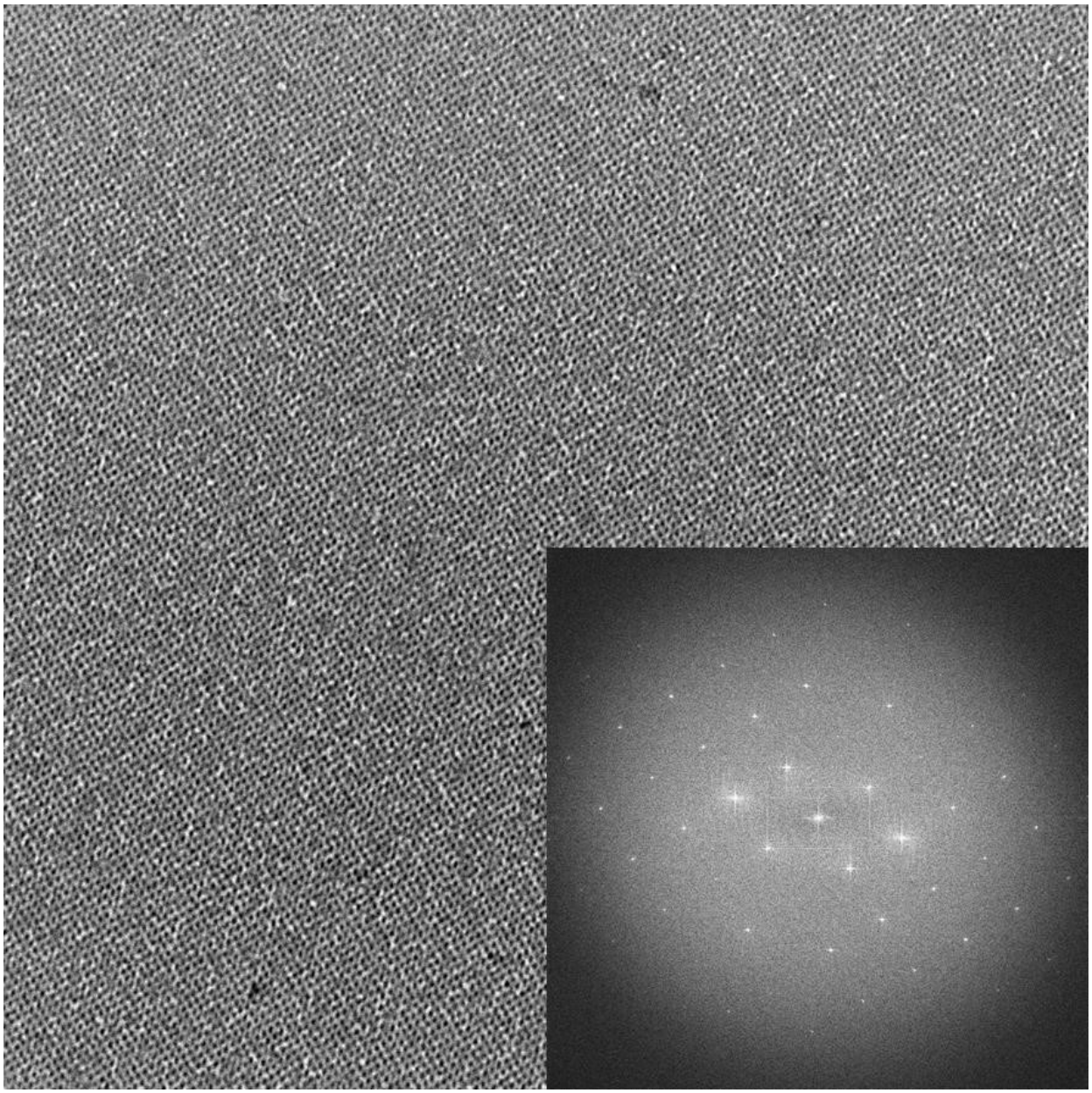


Figure S7. Image of a negatively stained, carbon-backed SA monolayer crystal, showing that the crystal lattice remains intact after the grid is washed in 1% SDS. This experiment was done only to see whether SDS treatment would remove all particulate material from the crystals. It has not yet been followed up by a study to determine whether these carbon-backed crystals might be used with specimens containing detergents.

VII. FURTHER DISCUSSION AND DETAILS ABOUT THE ON-GRID CRYSTALLIZATION OF SA MONOLAYER CRYSTALS

Reasons why on-grid crystallization may fail

When using the micro-well technique to grow SA monolayer crystals, and in the early stages of developing the current on-grid crystallization technique, we frequently had inconsistent results in getting individual monolayer crystals that spanned most of the holes on a holey-carbon grid. While some of the time we did obtain the desired result, at other times the crystals were small and fragmented, or there were no crystals at all. Examples of some of the failed outcomes are shown in **Figure S8**.

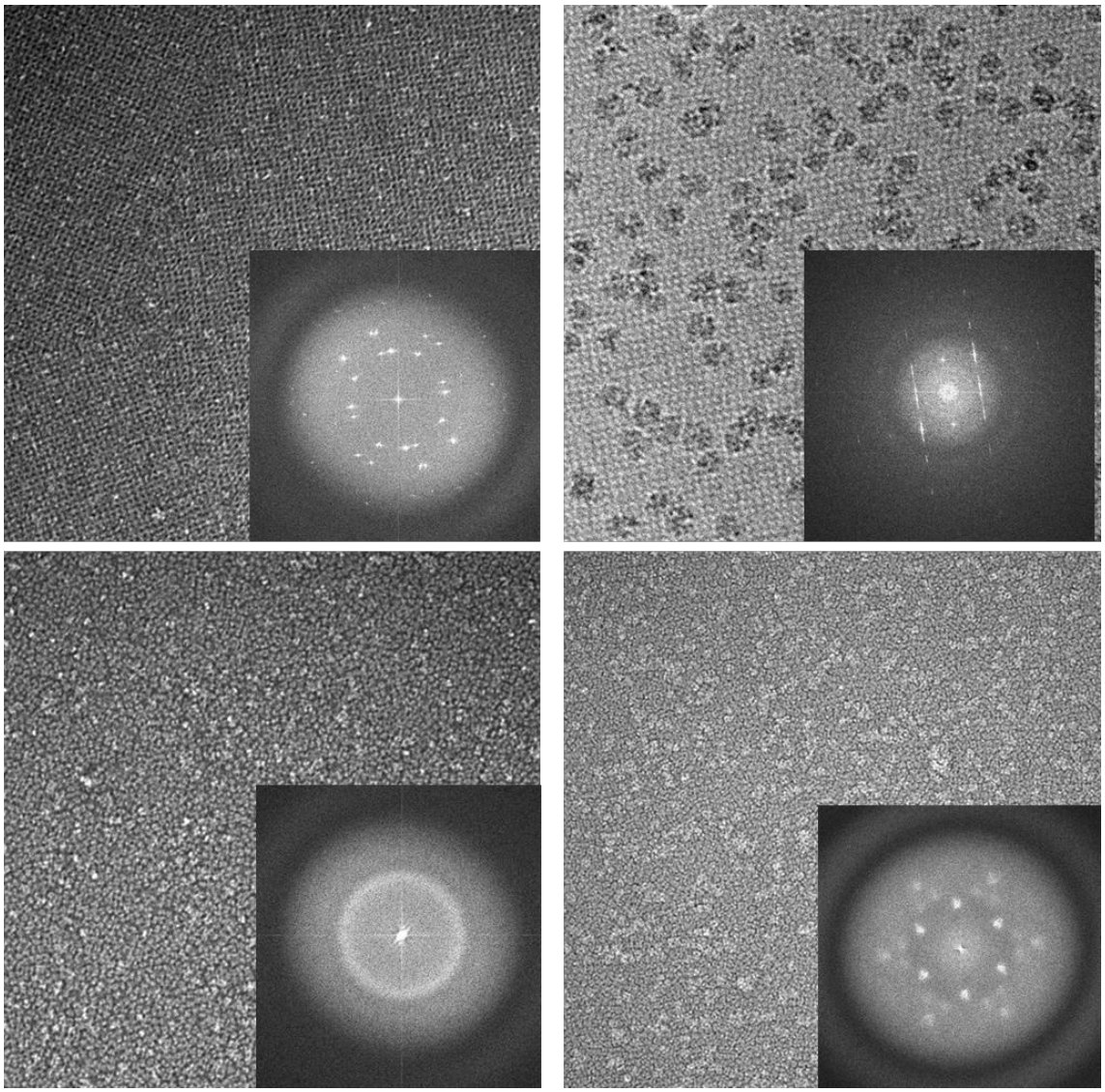


Figure S8. Gallery of examples of cases when the attempt failed to produce single crystals of SA, spanning an entire hole in the holey-carbon support film. The example shown at the top right, while a single crystal (with bound ribosome particles, in this case), is shown because it is an example in which the crystal contains unwanted stacking faults, whose presence is indicated by streaking of selected Bragg spots.

On the basis of this experience, we believe that the integrity of the SA monolayer crystals may be especially at risk at any stage before a carbon backing is deposited onto the lipid tails. The reason for our concern is that pressure fluctuations must occur every time a grid is either touched to, or is separated from, a drop of buffer, due to the high surface tension. Our hypothesis is that these transient changes in pressure can cause the thin, membrane-like monolayers to bulge above or below the plane of the holes in the holey carbon film. Although the lipid monolayer is fluid and extensible, the SA crystals are not. Thus, if pressure-induced bulging causes the area of the lipid monolayer to increase by only a few per cent, the SA crystals are likely to crack and not anneal properly as the bulging relaxes.

We now take care to minimize the mechanical stresses that we believe to be associated both with fusion between, and rupture of, the air-water interfaces as the grid is moved from one drop to the next. We believe that, prior to backing the SA crystals with evaporated carbon, this step in the protocol may be the one in which the monolayer crystals of SA are at greatest risk of becoming fragmented or even lost completely.

Another step that we feel may be important enough to mention is to thoroughly wash the grid after binding a specimen to the streptavidin. It is possible to do this quite stringently because the biotinylated particles, once bound, are not going to come off again. Our concern here is that some types of protein can rapidly form a monolayer "skin" of denatured protein at the air-water interface, on the surface of the drop, while the specimen is incubating on the grid. Holoferritin, for example, is one such example[26], and it has been reported that lysozyme also denatures very rapidly at the air-water interface[27].

Although perhaps rarely, such a skin may be retained nearly everywhere across the grid, when the underlying buffer is blotted away. This is, in any case, our interpretation of what has happened, whenever we encounter a grid where a thin layer, or skin, can be seen to be stretched over some of the open squares, "bunching up" near the grid bars. Examples are shown in **Figure S9**. In some grid squares a thin layer seemed to be peeled back from the grid bars, and folded out into the adjacent open area of the square.

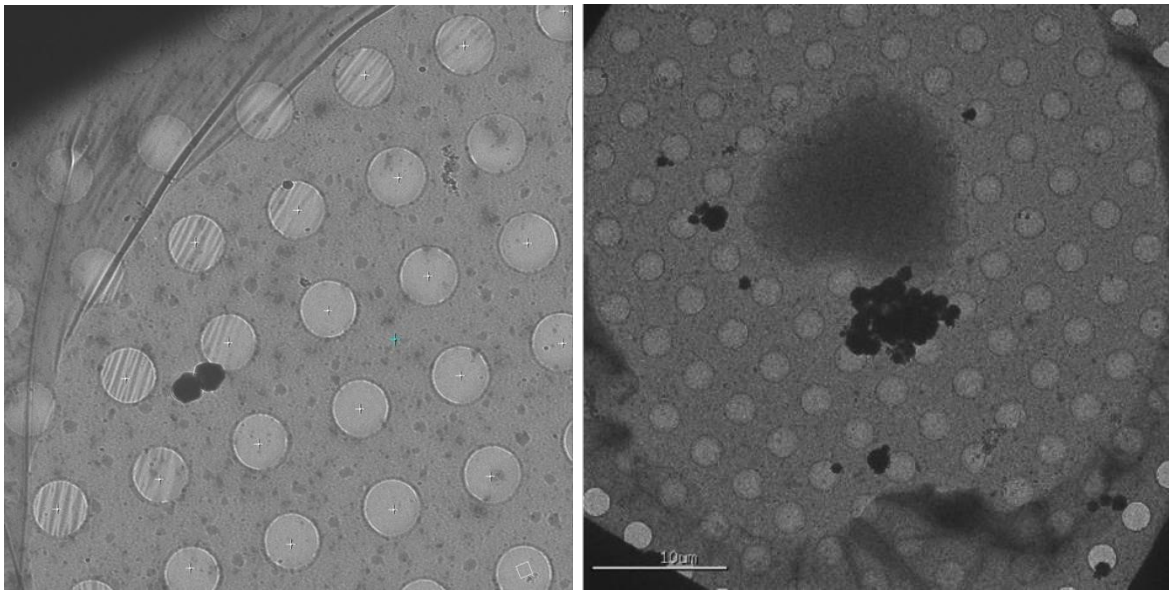


Figure S9. Two examples of low-magnification images that show that cryo-grids sometimes have a denatured-particle "skin" that is collapsed onto the holey-carbon grid. The holey-carbon grid itself has carbon-backed streptavidin monolayer crystals spanning the area of most of the holes.

Advantages and disadvantages of adding a carbon backing

The original reason for evaporating carbon onto the back side of SA affinity grids was, as mentioned above, to prevent bulging of the thin, membrane-like support film during the washing and blotting steps. Since an evaporated-carbon layer cannot be stretched to any practical extent, it was reasoned that it should provide a strong restraint against an increase in area that would be caused by such bulging.

An unintended benefit of the carbon-backing is that the SA crystals remain well ordered when washed with 1% sodium dodecyl sulfate detergent (SDS), as is shown in **Figure S7**. The fact that carbon-backed SA crystals are stable when washed with a detergent solution, as harsh as SDS, indicates that these support films may be used for any samples containing detergent, including membrane proteins. This came as a surprise to us, because our experience with SA monolayer crystals – without the carbon backing – is that they do not tolerate exposure to even much milder detergents. The stability to SDS, after carbon backing of monolayer crystals, was discovered in the course of preliminary experiments in which we tested a variety of reagents in attempts to clean off the contaminating particles that sometimes are present, mentioned below.

In closely related work, evaporating carbon onto the tails of Ni-NTA lipid monolayers, spanning the holes of holey-carbon films, was introduced previously by Kelly et al.[28] to stabilize pre-made affinity grids during long-term storage. In addition, Jap et al. had earlier evaporated carbon onto monolayer films of denatured ferritin, again spanning the holes of holey-carbon support films, in this case to stabilize them against rupture when irradiated by electrons[29].

Although the thin-carbon backing layer contributes to the noise in the image, it also adds features with significant, compensating value. Much stronger Thon rings, helpful for accurate determination of the CTF, are provided when amorphous carbon is present in the image. Furthermore, the electrical conductivity over the open hole should be improved relative to that of amorphous ice.

The major limitation on the resolution of the data recovered from single SA images seems to be the ice thickness, with the best crystal images being obtained from areas where the ice is estimated to be less than ~40 nm. The thickness of ice is also the major limiting factor on the resolution to which the Thon rings can be discerned.

Issues still to be addressed

We assume, but we have not tested whether it is true, that some of the specific elements of our on-grid crystallization protocol can be varied, or even dispensed with, and still one would get equally good results. We did not test, for example, whether comparable results were obtained when using SA purchased from other vendors. It should also be noted that we have limited experience on the extent to which 2-D crystal growth is sensitive to pH, ionic strength, and ionic composition. It seems unlikely, however, that these parameters would be any more important for on-grid crystallization than what had been found for crystallization on a Langmuir trough[30–32].

Although a great deal of our work initially used dioleoyl lipid, we now prefer to use dipalmitoyl lipid in order to avoid complications that might be caused by oxidation of the fatty acid tail. Fortunately, this saturated-lipid remains fluid as a monolayer when compressed by the castor-oil piston. It is nevertheless apparent to the eye that the dipalmitoyl lipid monolayer is more viscous than is the dioleoyl lipid monolayer, as it only slowly adopts a fully circular profile like that shown in **Figure S1A**.

We prefer to use EM grids that have a hole size of 2 μm in diameter. We use gold or molybdenum grids rather than copper grids because the latter can react with azide (as well as amine

groups) when the protein solution is incubated on the grid. Copper ions are known to catalyze the oxidation of lipids, and we suspect this to have been a major factor in producing previously variable SA crystal quality. We emphasize that none of the reagents that we now recommend contain azide, out of caution for the possibility that azide may react even with gold (or molybdenum).

A problem that still remains is that particulate contamination (non-crystalline material) sometimes, but not always, is found on the SA monolayer crystals. This is illustrated by the example shown in **Figure S10**. Although the source of this contamination is unknown at present, it may be that it comes from the air-water interfaces on the surfaces of drops of buffer that are used in our protocol. Even when this particulate contamination is present, it can be avoided when boxing images of the macromolecules of interest, provided that size of the particle of interest is large enough. For the moment, however, further work is needed to understand how to consistently avoid the appearance of such contamination if one intends to use the SA support films for smaller macromolecules.

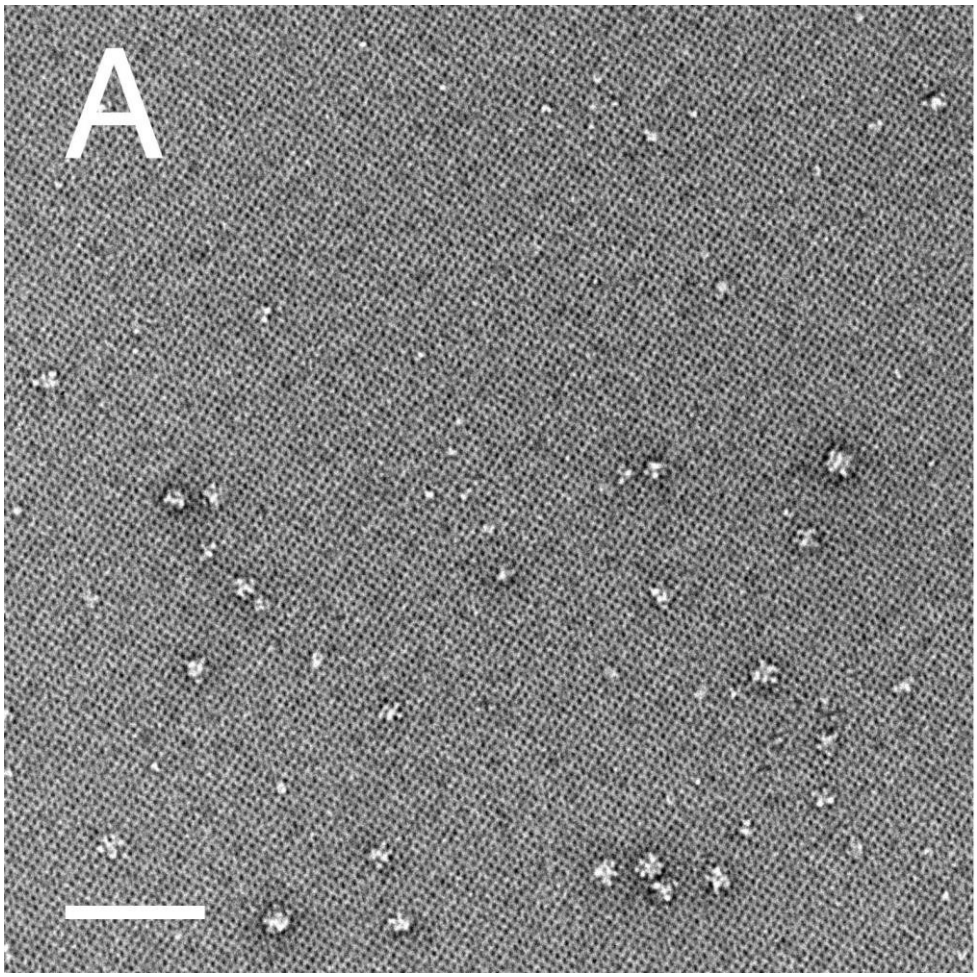


Figure S10. Image of a negatively stained specimen, showing one of the worst examples of particulate contamination on the SA monolayer crystal. The scale bar corresponds to 100 nm.

REFERENCES

1. Taylor, K. A. & Glaeser, R. M. Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. *J. Struct. Biol.* (2008). doi:10.1016/j.jsb.2008.06.004
2. Darst, S. A. *et al.* Two-dimensional crystals of streptavidin on biotinylated lipid layers and their interactions with biotinylated macromolecules. *Biophys. J.* (1991). doi:10.1016/S0006-3495(91)82232-9
3. Crucifix, C., Uhring, M. & Schultz, P. Immobilization of biotinylated DNA on 2-D streptavidin crystals. *J. Struct. Biol.* (2004). doi:10.1016/j.jsb.2004.02.001
4. Wang, L. & Sigworth, F. J. Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy. *Nature* (2009). doi:10.1038/nature08291
5. Wang, L., Ounjai, P. & Sigworth, F. J. Streptavidin crystals as nanostructured supports and image-calibration references for cryo-EM data collection. *J. Struct. Biol.* (2008). doi:10.1016/j.jsb.2008.07.008
6. Han, B. G. *et al.* Electron microscopy of biotinylated protein complexes bound to streptavidin monolayer crystals. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.04.025
7. Wang, L. & Sigworth, F. J. Liposomes on a streptavidin crystal. A system to study membrane proteins by cryo-EM. in *Methods in Enzymology* (2010). doi:10.1016/S0076-6879(10)81007-9
8. Kubalek, E. W., Kornberg, R. D. & Darst, S. A. Improved transfer of two-dimensional crystals from the air/water interface to specimen support grids for high-resolution analysis by electron microscopy. *Ultramicroscopy* (1991). doi:10.1016/0304-3991(91)90082-H
9. Langmuir, I. The constitution and fundamental properties of solids and liquids. II. Liquids. *J. Am. Chem. Soc.* (1917). doi:10.1021/ja02254a006
10. Blaha, G. *et al.* Preparation of functional ribosomal complexes and effect of buffer conditions on tRNA positions observed by cryoelectron microscopy. *Methods Enzymol.* (2000). doi:10.1016/s0076-6879(00)17021-1
11. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* (2013). doi:10.1038/nmeth.2472
12. Gipson, B., Zeng, X., Zhang, Z. Y. & Stahlberg, H. 2dx-User-friendly image processing for 2D crystals. *J. Struct. Biol.* (2007). doi:10.1016/j.jsb.2006.07.020
13. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.09.006
14. Tang, G. *et al.* EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* (2007). doi:10.1016/j.jsb.2006.05.009
15. Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* (2005). doi:10.1016/j.jsb.2005.03.010
16. Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. Structure of purple membrane from halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* (1986). doi:10.1016/0304-3991(86)90203-2
17. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. W. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* (2013). doi:10.7554/eLife.00461
18. Brown, A., Fernández, I. S., Gordiyenko, Y. & Ramakrishnan, V. Ribosome-dependent activation of stringent control. *Nature* (2016). doi:10.1038/nature17675

19. Fischer, N. *et al.* Structure of the E. coli ribosome-EF-Tu complex at. *Nature* (2015). doi:10.1038/nature14275
20. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* (2015). doi:10.7554/eLife.06980
21. Abazari, A., Chakraborty, N., Hand, S., Aksan, A. & Toner, M. A raman microspectroscopy study of water and trehalose in spin-dried cells. *Biophys. J.* (2014). doi:10.1016/j.bpj.2014.09.032
22. Zhou, Z. H. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Current Opinion in Structural Biology* (2008). doi:10.1016/j.sbi.2008.03.004
23. Hebert, H. *et al.* The 3.0 Å projection structure of microsomal glutathione transferase as determined by electron crystallography of p21212 two-dimensional crystals. *J. Mol. Biol.* (1997). doi:10.1006/jmbi.1997.1216
24. Leong, P. A., Yu, X., Zhou, Z. H. & Jensen, G. J. Correcting for the Ewald Sphere in High-Resolution Single-Particle Reconstructions. in *Methods in Enzymology* (2010). doi:10.1016/S0076-6879(10)82015-4
25. Glaeser, R. M., Downing, K., DeRosier, D., Chiu, W. & Frank, J. *Electron crystallography of biological macromolecules*. (Oxford University Press, 2007).
26. Yoshimura, H., Scheybani, T., Baumeister, W. & Nagayama, K. Two-Dimensional Protein Array Growth in Thin Layers of Protein Solution on Aqueous Subphases. *Langmuir* (1994). doi:10.1021/la00021a062
27. Yano, Y. F., Arakawa, E., Voegeli, W. & Matsushita, T. Real-time investigation of protein unfolding at an air-water interface at the 1 s time scale. *J. Synchrotron Radiat.* (2013). doi:10.1107/S0909049513023741
28. Kelly, D. F., Abeyrathne, P. D., Dukovski, D. & Walz, T. The Affinity Grid: A Pre-fabricated EM Grid for Monolayer Purification. *J. Mol. Biol.* (2008). doi:10.1016/j.jmb.2008.07.023
29. Jap, B. K. *et al.* 2D crystallization: from art to science. *Ultramicroscopy* (1992). doi:10.1016/0304-3991(92)90007-7
30. Blankenburg, R., Meller, P., Ringsdorf, H. & Salesse, C. Interaction between Biotin Lipids and Streptavidin in Monolayers: Formation of Oriented Two-Dimensional Protein Domains Induced by Surface Recognition. *Biochemistry* (1989). doi:10.1021/bi00446a037
31. Ratanabanangkoon, P. & Gast, A. P. Effect of ionic strength on two-dimensional streptavidin crystallization. *Langmuir* (2003). doi:10.1021/la026176r
32. Yatcilla, M. T., Robertson, C. R. & Gast, A. P. Influence of pH on two-dimensional streptavidin crystals. *Langmuir* (1998). doi:10.1021/la970237h

Chapter 3: Monolayer-crystal streptavidin support films provide an internal standard of cryo-EM image quality

Reproduced with permission from: Han BG, Watson Z, Cate JHD, Glaeser RM. J Struct Biol. 2017;200(3):307-313. doi:10.1016/j.jsb.2017.02.009

ABSTRACT

Analysis of images of biotinylated *Escherichia coli* 70S ribosome particles, bound to streptavidin affinity grids, demonstrates that the image-quality of particles can be predicted by the image-quality of the monolayer crystalline support film. The quality of the Thon rings is also a good predictor of the image-quality of particles, but only when images of the streptavidin crystals extend to relatively high resolution. When the estimated resolution of streptavidin was 5 Å or worse, for example, the ribosomal density map obtained from 22,697 particles went to only 9.5 Å, while the resolution of the map reached 4.0 Å for the same number of particles, when the estimated resolution of streptavidin crystal was 4 Å or better. It thus is easy to tell which images in a data set ought to be retained for further work, based on the highest resolution seen for Bragg peaks in the computed Fourier transforms of the streptavidin component. The refined density map obtained from 57,826 particles obtained in this way extended to 3.6 Å, a marked improvement over the value of 3.9 Å obtained previously from a subset of 52,433 particles obtained from the same initial data set of 101,213 particles after 3-D classification. These results are consistent with the hypothesis that interaction with the air-water interface can damage particles when the sample becomes too thin. Streptavidin monolayer crystals appear to provide a good indication of when that is the case.

INTRODUCTION

Affinity support films have a number of features that are attractive for preparing specimens for single-particle electron cryo-microscopy (cryo-EM for short). It is thought that affinity binding should be generally more structure-friendly than adsorption of biological macromolecules to continuous carbon films, and certainly less hazardous than interaction with the air-water interface within open holes. In addition, binding to an affinity substrate offers the possibility of achieving a high density of particles per unit area, even when the particle concentration is relatively low, such as ~40 nM[1].

A number of different types of affinity grid have been introduced in recent years. One of these is based on Ni-NTA lipid monolayers that are picked up on holey-carbon films[2–4], following a technique introduced originally by Kubalek et al., 1994[5]. Another is based on chemical functionalization that can be performed on oxidized, continuous carbon films[6]. A more recent idea has been to simply bind antibodies non-specifically to a carbon support films, and then use these as affinity grids[7].

Using streptavidin (SA) monolayer crystals to make affinity grids represents yet another approach[1,8,9]. There are numerous ways to take advantage of the high binding affinity of biotin, and also of streptavidin-binding peptides, to immobilize particles of interest onto the SA support

film. Examples include the use of biotinylated "adaptor molecules"[8]; random chemical biotinylation of lysine residues on the surface of proteins[10]; and – for membrane proteins – incorporation into proteoliposomes that include a biotinylated lipid[11]. A further feature of the SA affinity grid is that the image of the SA crystal can be subtracted by Fourier filtering, when desired[9].

We now report that a further benefit of using monolayer crystals of SA is that they provide a convenient way to assess single-particle cryo-EM image quality. The image-quality of the SA lattice itself is easily evaluated, because the signal is confined to reciprocal lattice points ("Bragg spots"). By comparison, it takes much more effort to evaluate the image quality of single particles because of the lengthy process involved in obtaining a high-resolution density map. It thus is quite useful that the image of the SA lattice provides a good indication of the quality of images of single particles bound to the lattice.

To illustrate the use of SA crystals as an indicator of single-particle image quality, we obtained 3-D reconstructions of *Escherichia coli* 70S ribosomes using images in which the resolution of the SA lattice was either (1) no better than 5 Å or (2) 4 Å or better. Importantly, the quality of the Thon rings was the same in both cases. When the images of SA crystals extended to only 5 Å or worse, the resolution of the density map obtained from 22,697 particles extended to only 9.5 Å. On the other hand, when the resolution of the SA lattice was better than 4 Å, the density map from the same number of particles extended to a resolution of 4.0 Å. The resolution of the map improved even further, to 3.6 Å, when 57,826 particles from such images were used.

MATERIALS AND METHODS

The data set analyzed in this study is the same one that was collected as part of our initial description of how to make long shelf-life SA affinity grids[1]. To summarize, biotinylated *E. coli* 70S ribosomes, which had been incubated with 20 μM spectinomycin, were bound to carbon-backed SA monolayer crystals that spanned the open holes of Quantifoil grids. Cryo-EM images were collected at 300 keV with a Gatan K2 camera, using a low-base FEI Titan electron microscope, equipped with a Gatan 626 cold holder. LEGINON[12] was used to automate the data collection. The movie frames were re-aligned and summed with the most recent version of software developed at UCSF, referred to as MotionCor2[13]. An example of such an image is presented in Figure 1 of Han et al.[1]. The ribosome particles used in the current work were the same ones that were boxed previously.

A key step added in the present work consisted of using an automated method to quantitatively characterize the quality of images of SA crystals. The first step in this process was to unbend the images of SA lattices, using scripts provided in the 2dx package[14]. Another script, written in MATLAB (The MathWorks, Inc.), was then used to define the resolution zone within which the average IQ value, as reported in 2dx after unbending, increased to a value of 7.

In addition, the values reported by CTFFIND4[15], for the cross correlation between the experimental Thon rings and the corresponding theoretical models of the CTF, were used as a second measure of the quality of the image. In this instance, the cross correlation was limited to the region between 1/(8 Å) and 1/(4 Å), and Bragg spots from the SA lattice were first removed, using the script for Fourier filtering described in Han et al.[1].

Single-particle data-processing of images of ribosome particles was done with tools in the RELION package[16], version 1.4. Final refinements were always carried out with a mask placed

around the large ribosomal subunit. Further details of the processing are mentioned in the Results section, below.

RESULTS

Although all holes in the carbon film that were queued up for automated data collection appeared to be equally good when viewed in SEARCH mode, the corresponding high-magnification images of streptavidin were not all of equal quality. After unbending[17] the streptavidin lattice, the highest resolution at which Bragg spots with IQ values of 3 or 4 could be seen extended beyond 4 Å in many, but not all, of the images. For other images, however, Bragg spots with IQ values of 3 or 4 did not go beyond 10 Å. Examples of these two cases are shown in **Figure 1**. The initial interpretation might be that some of the streptavidin crystals simply were "not good to begin with". However, a deeper analysis of the data, presented below, indicates that this interpretation is not correct.

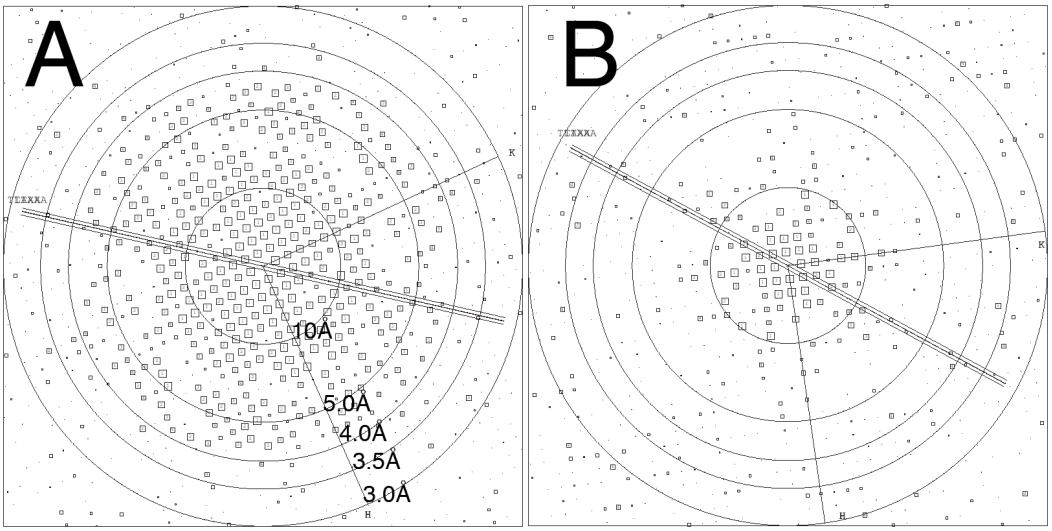


Figure 1. Representative examples of IQ plots produced after unbending images of streptavidin monolayer crystals. Circles of different radius are superimposed at different resolution values, as labeled in panel A. (**A**) The IQ plot for an image selected from within the middle of the distribution of "good images" that is shown in Figure 2B. Diffraction spots with IQ values of 4 or less, which are shown as larger boxes, extend to a resolution of 4 Å, with a few being present at even higher resolution. (**B**) The IQ plot for an image selected from within the middle of the distribution of "poor images" that is shown in Figure 2B**.** Diffraction spots with IQ values of 4 or less extend only to a resolution somewhere between 5 Å and 10 Å.

In order to estimate objectively the image quality of each SA crystal, we first plotted the reciprocal of the values of IQ, averaged within successive resolution shells, as a function of resolution. The IQ value of individual Bragg spots in unbent images was first introduced as a figure of merit to evaluate the quality of the data by Henderson et al.[17]. When IQ = 7, the estimated phase error is 45° - see Table 10.1 in Glaeser et al.[18], and the contributions of signal and noise within such a Bragg spot are equal.

Figure 2A shows examples of such plots, using the two images whose Fourier transforms are shown in **Figure 1**. Also shown in **Figure 2A** is the fit of a Gaussian function to the data that is of the form

$$y = ae^{-\frac{x^2}{b}} + 0.13,$$

where y is the reciprocal of the average value of IQ in a given resolution zone, and x is the resolution corresponding to the midpoint of each such zone. The fixed constant, 0.13, is an estimate of the reciprocal of the average IQ values for the structure factors that are provided by 2dx, when there is little or no signal in the Bragg spots.

The resolution at which $\frac{1}{<IQ>} = \frac{1}{7} = 0.143$ for the fitted curve is then used as an indication of the quality of the image of the streptavidin crystal. Unexpectedly, the distribution of the quality (i.e. "resolution") of the ~1500 images in the data set was found to be bimodal, as is shown in **Figure 2B**. There is a relatively narrow peak in the distribution centered at an estimated resolution of ~3.5 Å, and a second, relatively broad peak centered at a resolution of ~6.5 Å.

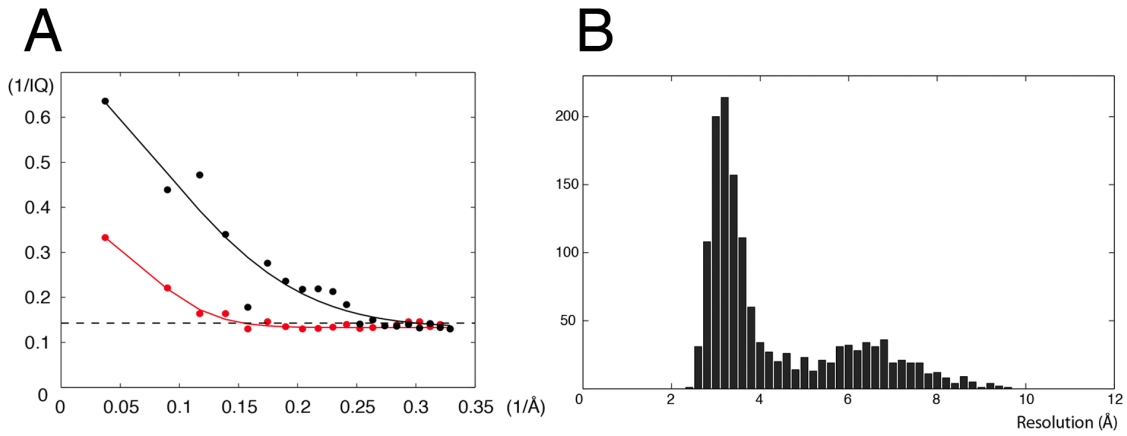


Figure 2. Explanation of how the IQ plot is used to evaluate the quality (i.e. resolution) of an image. (**A**) The values of IQ after unbending, as reported by 2dx, are first averaged within successive resolution shells. The reciprocal of this average value is then plotted as a function of spatial frequency. Finally, a Gaussian function, defined in Equation 1, is fitted to the data. The resolution of the image of the SA crystal is then estimated to be the value at which the fitted curve falls to 1/7. (**B**) Histogram showing the number of images in resolution bins versus the resolution. The bimodal distribution seen here was not anticipated.

We first wanted to understand why the distribution of images with respect to their quality was bimodal. The simplest possibility would be that "some grids are good and some are not", due to unknown factors in the many steps in their preparation. The situation turns out to be more complicated. As is illustrated in **Figure 3**, the quality of images of the streptavidin crystals can vary within one grid, as well as from one grid to another. Furthermore, when there is variation within one grid, there seem to be local areas that are good and others that are bad, as is suggested by there being sequential "runs" with similar image quality.

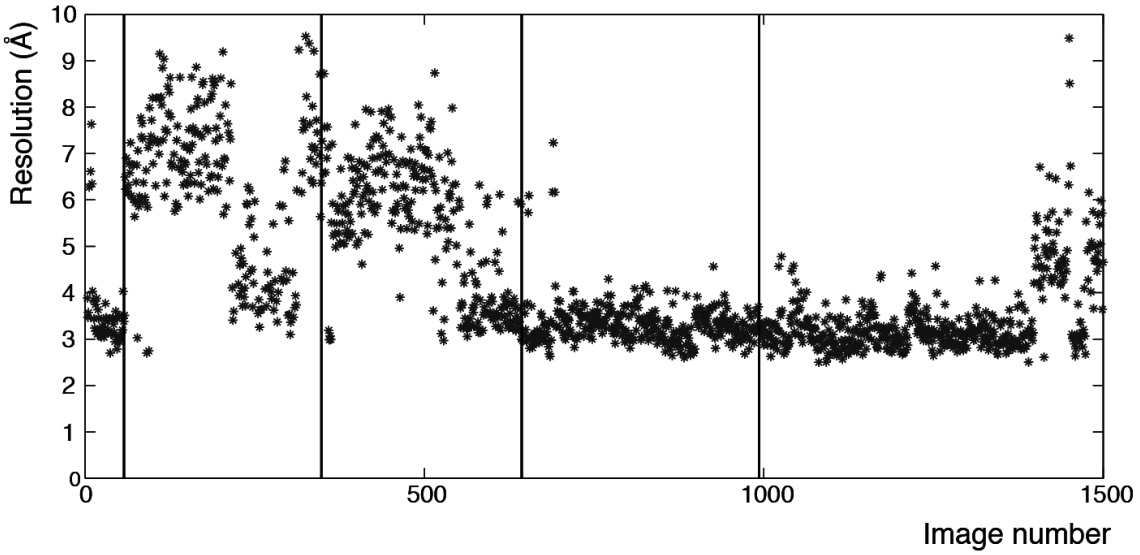


Figure 3. Plot of the estimated resolution of the SA crystal in a given image as a function of the sequential number given to that image within the full data set. Vertical lines indicate boundaries between images that were recorded from different grids.

In addition, the expectation that the resolution of the SA crystals would scale with the resolution of the Thon rings is also incorrect. Instead, as is shown in **Figure 4**, a 2-D scatter plot of SA-image quality vs Thon-ring quality again exhibits a bimodal distribution. It is informative that the images lying within the banana-shaped data-cloud do, in fact, go to somewhat higher resolution when the Thon rings are also better, and, conversely, the resolution values for the SA images within this group are somewhat worse when the Thon rings are not as good. On the other hand, images within the second cloud of data points, for which the resolution of the SA lattice is poor, mainly exhibit quite good Thon rings.

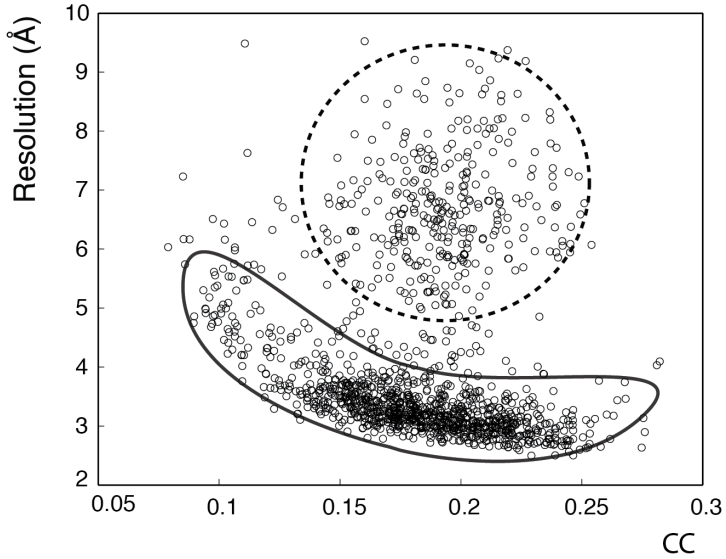


Figure 4. Scatter plot of the estimated resolution of the SA crystal versus the cross-correlation between the continuous Fourier transform of the image and a model of the CTF for the defocus value of that image. CTF fitting, and the corresponding values of the cross-correlation value in the resolution zone between 8 Å and 4 Å, were obtained by

using CTFFIND4[15]. The plot shows a bimodal distribution, with two clouds indicated schematically by dotted and solid outlines for "bad" and "good" SA crystals, respectively.

We next investigated whether the data-quality for ribosome particles would correlate with the data-quality for the SA crystals to which they were bound. To do this, we created two subsets of particles, depending upon where their images fell with respect to the two clouds in the scatter plot shown in **Figure 4**. More specifically, one group consisted of 22,697 particles for which the resolution for the SA crystals was 5 Å or worse, and the second group consisted of the same number of particles for which the resolution for the SA crystals was 4 Å or better. We then obtained refined density maps for the two groups. As is shown in **Figure 5**, the resolution of the map was only about 9.5 Å for particles in the first group, i.e. ones boxed from images in which the resolution of the SA lattice was 5 Å or worse. On the other hand, the resolution obtained for the second group was quite good (4.0 Å), considering the relatively small number of particles that was used to obtain the density map.

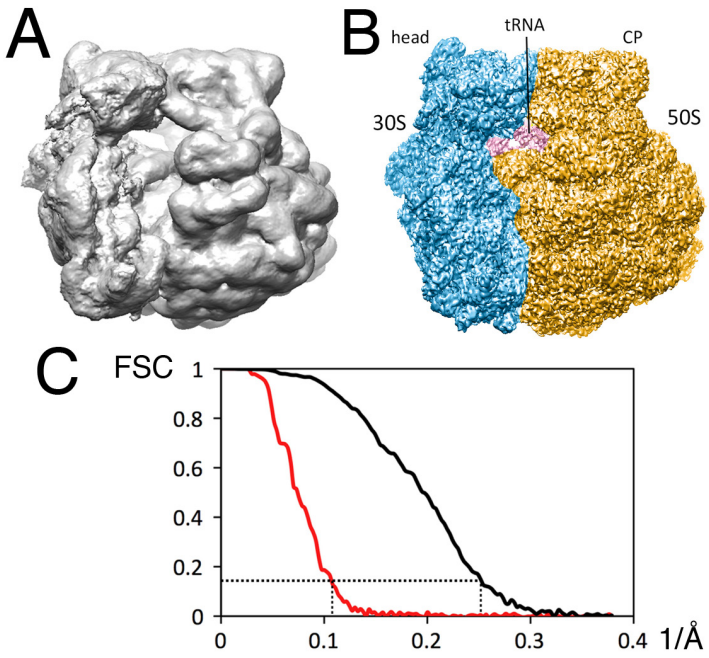


Figure 5. Result obtained when performing refinement using the same number of particles boxed from good images and from bad images, respectively. (**A**) The density map for 22,697 ribosomes boxed from bad images. (**B**) The density map for the same number (22,697) of ribosomes boxed from good images. Head, head domain of the 30S subunit; CP, central protuberance of the 50S subunit. (**C**) The "gold standard" FSC curves for the two maps shown in panels (**A**) and (**B**) respectively; the red curve is for ribosomes boxed from bad images and the black curve is for ribosomes boxed from good images. The point where the value of the FSC curve is 0.143 is indicated by the dashed lines.

Next, we determined the resolution obtained when using the full set of ribosome particles boxed from "good" images. To do this, we first edited that data set by performing 3 rounds of 2-D classification, deleting, in each round, mainly those particles for which the class averages showed no significant evidence that they contained 70S ribosomes. This process reduced the number of particles from 61,646 to 57,826. We then used RELION to calculate a refined (single-class) density map. As is shown in **Figure 6**, the resolution of the map, determined by the "gold standard" FSC, was 3.6 Å. This resolution was the same, to two significant figures, for the refined

map obtained from the unedited data set of 61,646 particles, confirming that the candidate particles removed by editing contained little value of interest.

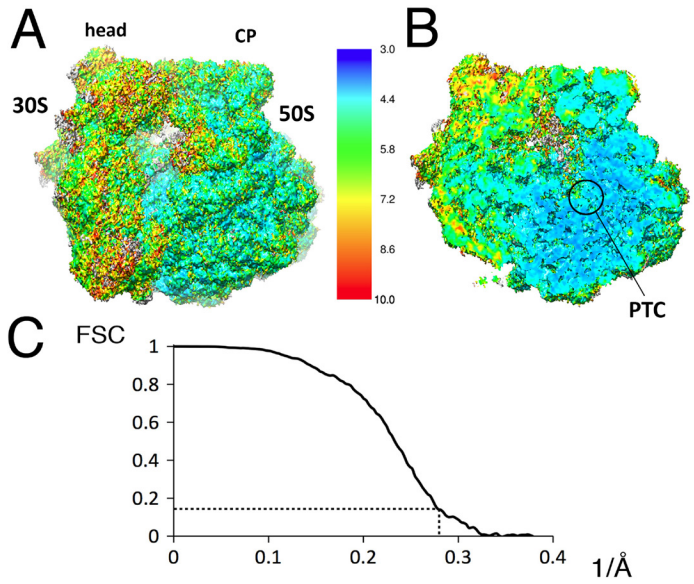


Figure 6. Map obtained when using ribosomes picked from good SA images. The particles were treated as being in a single class, i.e. 3-D classification was not used to address the expected conformational heterogeneity. As stated in Methods, a mask surrounding the 50S subunit was used during refinement. ResMap[19] version 1.1.4 was used to compute the local resolution. (**A**) Refined map for particles in the edited set of good images. Head, head domain of the 30S subunit; CP, central protuberance of the 50S subunit. (**B**) A cutaway of the local-resolution representation of the density map, for a plane that includes the peptidyl transferase center (PTC) of the 50S subunit. (**C**) The corresponding "gold-standard" FSC curve for the map shown in (**A**) and (**B**). The point where the value of the FSC curve is 0.143 is indicated by the dashed lines.

Encouraged by the relatively high resolution obtained, we performed 3-D classification of particles in the edited data set. We classified these particles into only three classes because the size of the data set was relatively small. As is shown in **Figure 7**, this produced two major classes consisting of 70S ribosome particles and a third, minor class consisting mainly of 50S particles. Refinement of the two major classes resulted in structures at resolutions of 4.3 Å (from 14,848 particles) and 3.8 Å (from 35,937 particles), respectively.

Comparison of the density maps produced by particles in the two major classes shows the familiar intersubunit "rotation" of the 30S particle relative to the 50S particle[20]. This figure was prepared by first using Chimera[21] to align the maps in panels (A) and (B) such that their 50S densities overlapped well with one another. Each 70S map was then compared to the X-ray crystal structures available as PDB accession numbers 4V9D and 4YBB. Map A was found to correlate best with the rotated P/E state in 4V9D assembly I, while map B correlates best with the unrotated state in 4YBB assembly 2.

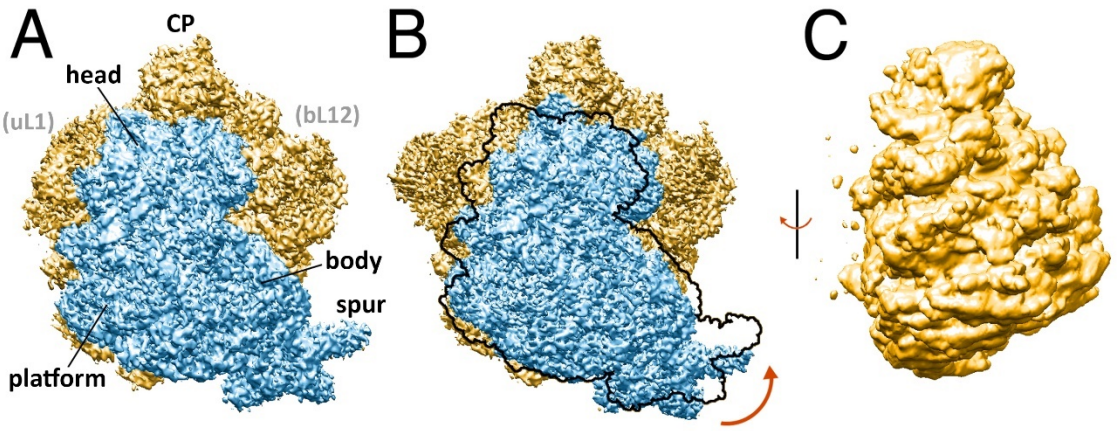


Figure 7. Results obtained after using 3-D classification to sort particles, which were all obtained from "good" images, into three smaller but more homogeneous subsets. (**A**) Density map for 14,846 particles assigned to one of the classes, representing an unrotated state of the ribosome. Features of the ribosome are indicated, including the location of uL 1 and bL 12 of the 50S subunit, which are not visible in the present map due to disorder. (**B**) Density map for 35,937 particles assigned to a second class**.** The outline of the subvolume of the 30S subunit, in the position occupied in panel A, is superimposed on the density map shown in B. The well-known rotation of the 30S subunit[20] is indicated by the curved arrow. (**C**) Density map for 7,041 particles in the third class shows that it consists mainly of 50S particles. The resolution after 3D refinement was 4.3 Å, 3.8 Å and 8.2Å in panels (**A**), (**B**), and (**C**), respectively.

DISCUSSION

Our analysis of the quality of cryo-EM images of ribosome particles prepared on SA-affinity grids is based on three quantitative metrics. One metric is the estimated resolution limit for the image of the SA-crystal. A second metric is based on the Thon rings seen in the Fourier transforms of images. Parenthetically, we believe that the thin carbon film that was evaporated onto the lipid-tail side of the SA crystals should make the main contribution to the Thon rings. The third metric is the resolution of the three-dimensional reconstruction obtained for biotinylated 70S ribosomes that were bound to the affinity grids.

Discarding poor-quality images reduces the computational burden involved during refinement while, at the same time, it is expected to improve the final result. As the data in **Figure 4** demonstrate, however, it is not sufficient to use just the Thon rings alone to select images with the best quality. Instead, significant value is also added by discarding images in which the diffraction peaks from the SA crystals fail to go to high resolution. The criterion of SA image quality thus complements other reasons for discarding images such as "high drift, low signal, heavy ice contamination, or very thin ice", mentioned by Loveland et al.[22].

The bimodal distribution that we found for the resolution of SA crystals – see **Figure 2** and **Figure 4** – indicates that something other than excessive amounts of inelastic scattering can be a strong factor affecting image quality. The broad peak in **Figure 2**, centered at a resolution of ~6.5 Å, obviously comprises mainly those images that form a diffuse, circular cloud in **Figure 4**. Importantly, as already mentioned, these same images tend to be quite good when one looks at the Thon rings, indicating that the ice was relatively thin for all of them.

Turning attention first to the banana-shaped cloud in **Figure 4**, the resolution values of the SA images does become slightly worse as the quality of Thon rings become worse, as might be expected. We hypothesize that the simultaneous loss of resolution for SA crystals and for the Thon rings may well be due, in this case, to increasing ice thickness. It is possible to test this hypothesis in the future by recording images with a microscope that is equipped with a zero-loss energy filter,

because the number of electron counts in the image would be measurably less, the thicker the specimen.

Focusing attention next on the bimodal distribution seen in **Figure 4**, it is significant that the ribosome particles appear to be damaged when the images of the SA crystals are poor, as is shown by the lower resolution of the density map that is obtained from such particles (compare **Figure 5A** to **Figure 5B**). One possible explanation is that the sample in such images – both ribosomes and streptavidin – was either partially dehydrated or in some other way contacted by the air-water interface after blotting. It is puzzling, however, that images in which the sample is good and those in which it is bad can both occur over a wide range of vitrified-specimen thicknesses, as inferred from the variable quality of the Thon rings for specimens with very similar thicknesses of evaporated carbon. A possible explanation is that the sample was spared from being damaged by a surfactant monolayer at the air-water interface in the good images, whereas in "bad images" the surfactant – if any – was too dilute to be effective. The direction in which we are thinking is explained further in the Supporting Material. A diagram has been included to illustrate how complex we believe the situation might be when the specimen becomes too thin. Since this hypothesis is complicated and speculative, however, it will not be easy to test it. A second possible explanation is based on the shear forces that occur as liquid is blotted away from the grid. Since the maximum shear rate is likely to vary in magnitude at different locations on a grid, it is worth considering whether mechanical forces could damage both the ribosomes and the SA crystals in some locations but not in others.

An alternative explanation notes that the background contributed by the SA layer must be less perfectly subtracted when the crystals are disordered. The question that then arises is whether this extra amount of background might cause alignment and assignment of Euler angles to be less accurate, thereby causing the refined density map for 22,697 particles to extend to only 9.5 Å. In order to test this alternative, we boxed 22,697 areas of background from images in which the SA was poor, and added these background-boxes, one at a time, to ribosome particles boxed previously from images in which the SA component went to a resolution of 4 Å or better. These intentionally corrupted images of ribosome particles were then used to obtain a refined density map, which was found to extend to a resolution of 4.8 Å. This value of resolution is only slightly worse than the value of 4.0 Å obtained from the same, limited number of particles before corrupting the images with additional noise. This result seems to rule that out the hypothesis that added background, resulting from imperfect subtraction of the SA component, is responsible for the significantly lower resolution obtained from images in which the Thon rings are excellent but the SA component is not.

Finally, it is still to be determined whether the carbon-backed version of SA affinity grids can be used for high-resolution imaging of macromolecular particles that are significantly smaller than ribosome particles. While the SA "background" can be subtracted from images by Fourier filtering, the noisy signal due to the carbon film is likely to eventually interfere with boxing particles and with alignment and assignment of Euler angles as the particle size decreases. The same limitation applies to making cryo-specimens directly on continuous carbon films, of course. We thus point out that the option may exist to use single-layer graphene to "back" SA monolayer crystals. An attractive approach to do so, not yet tested, would be to grow SA crystals on the surface of small wells, as described by Wang and Sigworth[11], and pick these up with grids already covered by graphene. While picking up monolayer crystals with evaporated carbon films does not perform very well[23], graphene may be more hydrophobic than evaporated carbon, and thus it may perform better. What we have tried, in preliminary experiments, is to pick up lipid monolayers on

graphene coated grids, and then use the on-grid method to grow SA crystals. The lipid transfer appeared to work well, as did the binding of SA to biotinylated lipid. Nevertheless, only small crystalline domains were obtained, possibly because the fluidity of lipids adsorbed to graphene may be much less than that of lipids spanning open holes in holey carbon films. An alternative approach may be to reduce the thickness of the evaporated carbon to as little as 1 nm, which has been shown to still be strong enough to span micrometer-sized holes[24].

ACKNOWLEDGMENTS

This work was supported in part by NIH grants P01 GM051487 and R01 GM065050, and by the National Science Foundation Graduate Research Fellowship Program under Grant number 1106400. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco (supported by NIH grant P41 GM103311).

SUPPORTING MATERIAL

A monolayer of surfactant, shown schematically as a green, dot-and-dashed line in **Figure S1**, is expected to stabilize thin aqueous specimens, working to prevent the "dewetting" of the streptavidin monolayer crystal in "random" locations that would otherwise occur when the aqueous film becomes thinner than ~100 nm. The reader is referred to Glaeser et al.[25] for a critique of the underlying physics of wetting, thinning and dewetting. In brief, van der Waals forces are the fundamental reason why liquid films thinner than ~100 nm are unstable against dewetting at "random" points and concomitant thickening in adjacent areas.

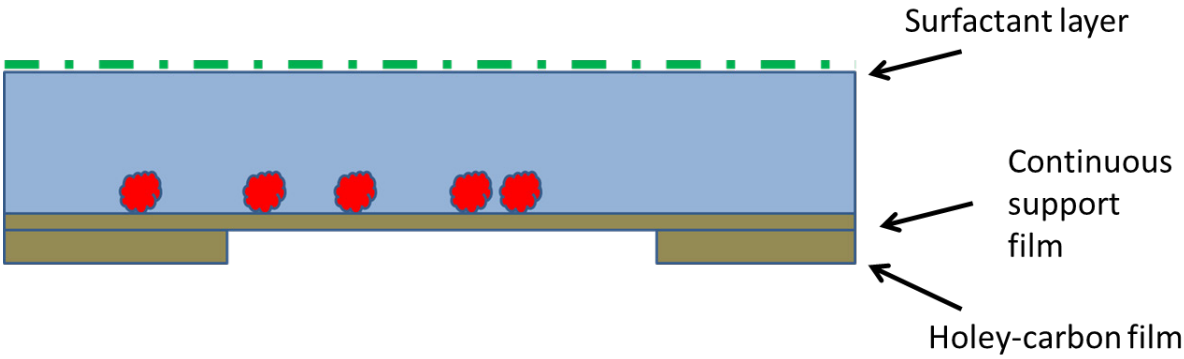


Figure S1. Diagram illustrating how we picture the situation in which particles, shown as red "balls", are adsorbed to a support film. Most of the aqueous buffer is imagined to have been blotted away. The remaining thin film of buffer (continuous blue) is shown here to have an unspecified surfactant layer (dot-and-dash green) at the air-water interface.

Loss of water due to evaporation can also thin the aqueous sample, of course, but the possibility exists that this will not occur – for example, if the specimen temperature is at the dew point of the ambient atmosphere. In any case, a layer of surfactant is expected to inhibit the dewetting of hydrophilic support films as the specimen becomes very thin, just as it inhibits the rupture of soap bubbles or the collapse of foams.

As the uniform aqueous film shown in **Figure S1** becomes thinner and thinner, the surfactant layer will eventually come within two or three nanometers of the surface of the adsorbed particles. At this point the tendency of polar groups to remain hydrated[26] is expected to produce a disjoining pressure that further acts to resist dewetting of the particles. We suggest that the air-water interface will no longer remain flat at that point. Instead it will sag and effectively "wrap" around the particles as thinning continues. Indeed, it is seen from time to time in cryo-EM images that the ice adjacent to particles is thicker than it is at a slightly greater distance from the particles. In other words, a "meniscus" of vitreous ice can sometimes form around individual particles, analogous to a "meniscus" of negative stain.

Our suggestion is that the protein structure may be "spared from being damaged" if a 2 nm or 3 nm thick hydration layer remains between the particle and the hypothesized surfactant layer. Again, the concept is that, due to hydration forces, the surfactant prevents the particle from being dewetted.

Conversely, protein structures are at greater risk in the absence of a surfactant layer

If there is too little, or perhaps even no surfactant on the air-water interface, it is possible that dewetting may occur quite easily, thus leaving many areas of the grid that are either too thick or that have dried out completely. It is common, of course, that at least some areas of a grid are judged to be completely dried out, while – on the same grid – others areas are found to be just the right thickness.

Even if dewetting does not occur, direct contact between a particle and a pristine air-water interface will occur when the thickness of a uniform film of buffer is equal to, or slightly less than the diameter of the particle. In this case the reader should imagine that the air-water interface shown in **Figure S1** is in direct contact with the tops of the red particles. Such direct contact is expected to be dangerous for protein molecules, because the energy landscape for denaturation is very different at the interface from what it is in bulk water, as has been discussed recently[27].

As a final comment, we note that the structural consequences of bringing proteins into contact with a pristine air-water interface are expected to be affected by the value of the ambient relative humidity. This is because a greater fraction of the structurally important "water of hydration" is more likely to be retained by proteins, even after bulk-phase dewetting has occurred, if the ambient relative humidity is high.

REFERENCES

1. Han, B.-G. *et al.* Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. *J. Struct. Biol.* **195**, 238–244 (2016).
2. Benjamin, C. J. *et al.* Nonfouling NTA-PEG-Based TEM Grid Coatings for Selective Capture of Histidine-Tagged Protein Targets from Cell Lysates. *Langmuir* (2016). doi:10.1021/acs.langmuir.5b03445
3. Kelly, D. F., Dukovski, D. & Walz, T. A practical guide to the use of monolayer purification and affinity grids. in *Methods in Enzymology* (2010). doi:10.1016/S0076-6879(10)81004-3
4. Kelly, D. F., Abeyrathne, P. D., Dukovski, D. & Walz, T. The Affinity Grid: A Pre-fabricated EM Grid for Monolayer Purification. *J. Mol. Biol.* (2008). doi:10.1016/j.jmb.2008.07.023
5. Kubalek, E. W., Le Grice, S. F. J. & Brown, P. O. Two-dimensional crystallization of histidine-tagged, HIV-1 reverse transcriptase promoted by a novel nickel-chelating lipid. *J.*

Struct. Biol. (1994). doi:10.1006/jsbi.1994.1039
6. Llaguno, M. C. *et al.* Chemically functionalized carbon films for single molecule imaging. *J. Struct. Biol.* (2014). doi:10.1016/j.jsb.2014.01.006
7. Yu, G., Li, K. & Jiang, W. Antibody-based affinity cryo-EM grid. *Methods* (2016). doi:10.1016/j.ymeth.2016.01.010
8. Crucifix, C., Uhring, M. & Schultz, P. Immobilization of biotinylated DNA on 2-D streptavidin crystals. *J. Struct. Biol.* (2004). doi:10.1016/j.jsb.2004.02.001
9. Wang, L., Ounjai, P. & Sigworth, F. J. Streptavidin crystals as nanostructured supports and image-calibration references for cryo-EM data collection. *J. Struct. Biol.* (2008). doi:10.1016/j.jsb.2008.07.008
10. Han, B. G. *et al.* Electron microscopy of biotinylated protein complexes bound to streptavidin monolayer crystals. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.04.025
11. Wang, L. & Sigworth, F. J. Liposomes on a streptavidin crystal. A system to study membrane proteins by cryo-EM. in *Methods in Enzymology* (2010). doi:10.1016/S0076-6879(10)81007-9
12. Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* (2005). doi:10.1016/j.jsb.2005.03.010
13. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* (2013). doi:10.1038/nmeth.2472
14. Gipson, B., Zeng, X., Zhang, Z. Y. & Stahlberg, H. 2dx-User-friendly image processing for 2D crystals. *J. Struct. Biol.* (2007). doi:10.1016/j.jsb.2006.07.020
15. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* (2015). doi:10.1016/j.jsb.2015.08.008
16. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2012.09.006
17. Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. Structure of purple membrane from halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy* (1986). doi:10.1016/0304-3991(86)90203-2
18. Glaeser, R. M., Downing, K., DeRosier, D., Chiu, W. & Frank, J. *Electron crystallography of biological macromolecules*. (Oxford University Press, 2007).
19. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* (2014). doi:10.1038/nmeth.2727
20. Frank, J. & Agrawal, R. K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* (2000). doi:10.1038/35018597
21. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* (2004). doi:10.1002/jcc.20084
22. Loveland, A. B. *et al.* Ribosome.RelA structures reveal the mechanism of stringent response activation. *Elife* (2016). doi:10.7554/eLife.17029
23. Kubalek, E. W., Kornberg, R. D. & Darst, S. A. Improved transfer of two-dimensional crystals from the air/water interface to specimen support grids for high-resolution analysis by electron microscopy. *Ultramicroscopy* (1991). doi:10.1016/0304-3991(91)90082-H
24. Williams, R. C. & Glaeser, R. M. Ultrathin carbon support films for electron microscopy. *Science (80-.).* (1972). doi:10.1126/science.175.4025.1000
25. Glaeser, R. M. *et al.* Factors that Influence the Formation and Stability of Thin, Cryo-EM Specimens. *Biophysical Journal* (2016). doi:10.1016/j.bpj.2015.07.050

26. Parsegian, V. A. & Zemb, T. Hydration forces: Observations, explanations, expectations, questions. *Current Opinion in Colloid and Interface Science* (2011). doi:10.1016/j.cocis.2011.06.010
27. Glaeser, R. M. & Han, B.-G. Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys. Reports* (2017). doi:10.1007/s41048-016-0026-3

Chapter 4: Defects in the assembly of ribosomes selected for β-amino acid incorporation

Reproduced with permission from: Ward FR*, Watson ZL*, Ad O, Schepartz A, Cate JHD. Defects in the Assembly of Ribosomes Selected for β-Amino Acid Incorporation. Biochemistry. 2019 Nov 12;58(45):4494-4504. doi: 10.1021/acs.biochem.9b00746. Epub 2019 Oct 28. PMID: 31607123.

* These authors contributed equally

ABSTRACT

Ribosome engineering has emerged as a promising field in synthetic biology, particularly concerning the production of new sequence-defined polymers. Mutant ribosomes have been developed that improve the incorporation of several non-standard monomers including D-amino acids, dipeptides, and β-amino acids into polypeptide chains. However, there remains little mechanistic understanding of how these ribosomes catalyze incorporation of these new substrates. Here we probed the properties of a mutant ribosome – P7A7 – evolved for better *in vivo* β-amino acid incorporation through *in vitro* biochemistry and cryo-electron microscopy. Although P7A7 is a functional ribosome *in vivo*, it is inactive *in vitro*, and assembles poorly into 70S ribosome complexes. Structural characterization revealed large regions of disorder in the peptidyl transferase center and nearby features, suggesting a defect in assembly. Comparison of RNA helix and ribosomal protein occupancy with other assembly intermediates revealed that P7A7 is stalled at a late stage in ribosome assembly, explaining its weak activity. These results highlight the importance of ensuring efficient ribosome assembly during ribosome engineering towards new catalytic abilities.

INTRODUCTION

The ribosome is a large molecular machine capable of directing the polymerization of a distinct sequence of amino acids by decoding a messenger RNA (mRNA) template. Its ability to accurately and efficiently select and incorporate the correct monomer from a pool of over 20 substrates makes the ribosome one the most versatile machines for polymer synthesis. No other known methods, from solid-state chemistry to bulk polymerization, can achieve both the specificity and yield of ribosome-catalyzed synthesis, particularly for products longer than ~50 monomers in length.[1] The ribosome has thus become a target of engineering efforts to eventually develop a ribosome-based platform for the synthesis of new classes of sequence-defined polymers. The large obstacle to these engineering attempts is the several-billion-year optimization of the ribosome, and the rest of the translational apparatus, toward efficient and selective α-amino acid polymerization. New sequence-defined polymerization chemistries will likely require repurposing a suite of enzymes and nucleic acids, including tRNAs,[2,3] aminoacyl tRNA synthetases (aaRSs),[4,5] elongation factor Tu (EF-Tu),[6,7] in addition to the ribosome itself. [8–12]

The catalytic core of the ribosome, the peptidyl transferase center (PTC), induces the attack of an A-site aminoacyl-tRNA on the ester bond of the P-site peptidyl-tRNA, transferring and extending the growing polypeptide chain. The PTC is thought to encourage proper polymerization

through several mechanisms including orienting the substrates,[13] favoring productive acid-base chemistry,[14] and protecting the peptidyl-tRNA from dead-end hydrolysis.[15] Like many other enzymes, the PTC undergoes a substantial activating rearrangement upon substrate binding that can accommodate a wide range of monomer side chains. However, the ribosome only weakly catalyzes polymerization of substrates with different backbone chemistries, including peptoids,[16] N-methyl amino acids,[17–19] D-amino acids,[20] aromatic foldamers,[21,22] aramids,[23] malonates,[23] and β^3-amino acids.[24]

β^3-amino acids (referred to here as β-amino acids) are a useful model substrate for novel ribosome polymerization reactions as they introduce new challenges for catalysis while closely resembling the natural α-amino substrates. The extra methylene between the alpha carbon and carboxylic acid groups adds more bulk, rotational degrees of freedom, and diminishes reactivity, [25] but the nucleophile remains an amino group, as in α-amino acids. The extra methylene group confers advantages to β-amino acid oligomers as peptidomimetics, as they form stable folds, display reduced immunogenicity, and are more resistant to proteolysis.[26–28] Importantly, despite reacting poorly in the PTC, under certain conditions β-amino acids can be consecutively polymerized by the ribosome *in vitro*.[3] Their similarities to natural monomers have made β-amino acids an attractive substrate target for ribosome engineering.

Attempts to engineer new polymerization chemistry into the ribosome have understandably focused on mutations to the PTC. Aggressive mutations of 10+ bases in the PTC have generated ribosomes that can better translate D-amino acids,[12] dipeptide substrates,[29] polyproline motifs,[8] and β-amino acids.[9,10,30–32] These mutations have been proposed to solve the space, orientation, and reactivity problems inherent to many backbone-modified substrates, but no mechanistic studies of these improved ribosomes have been published. Indeed, most of these mutant ribosomes have only been tested *in vivo* or in lysate-based cell free systems with wild-type ribosomes present. Although these ribosomes represent exciting steps toward new sequence-defined polymerization capabilities, a detailed understanding of how PTC mutations help widen the substrate scope of the ribosome is still lacking. In particular, the PTC is highly conserved across all domains of life and large-scale mutations may have knock-on effects that limit ribosome utility. Therefore, these engineering problems must always be evaluated with a discerning focus on the mechanistic implications of such changes, in order to better understand how these PTC alterations affect ribosome activity.

Here were present structural and biochemical characterization of a β-amino acid translating ribosome, P7A7, discovered via an *in vivo* selection for β-puromycin incorporation.[9] The P7A7 PTC carries twelve mutations over two regions near the tRNA A site and exit tunnel. We show that purified P7A7 ribosomes are inactive during *in vitro* translation and do not form stable 70S complexes. A cryo-electron microscopy (cryo-EM) structure of P7A7 50S ribosomal subunits reveals substantial disordering of the PTC and nearby inter-subunit bridge helices when compared with an equivalent wild-type (WT) 50S structure. Analysis of the P7A7 map reveals a depletion of late-assembling ribosomal proteins, which was confirmed using tandem mass tag (TMT) relative quantitation. When compared with existing studies of 50S ribosomal subunit assembly, P7A7 appears trapped as a late assembly intermediate, explaining its poor activity. Our results suggest that the radical PTC mutations often seen in engineered ribosomes may have unintended effects on ribosome assembly and stability that limit the utility of many of these variants.

METHODS

Plasmids and Cloning

All PCR reactions were performed using a Q5 DNA polymerase kit (NEB) and manufacturer recommended concentrations of primers, dNTPs, and enzyme. Mach1 cells (ThermoFisher) were used for all rRNA plasmid cloning and expression. See Table S2 for primer sequences. Plasmids pKK3535-P7A7 (Tet^R), encoding the P7A7 23S rRNA, and pLK35[33] (Amp^R) were used for cloning and rRNA expression. Plasmid pLK35-WT-MS2 was generated by whole-plasmid PCR amplification of pLK35 and blunt end ligation using primers h98_MS2_F/R (MS2 tag fragments in capitals). Plasmid pLK35-P7A7-MS2 was generated by Gibson assembly of a PCR fragment from pKK3535-P7A7 with primers P7A7_ex_F/R and a fragment from pLK35-WT-MS2 using the corresponding complementary Gibson primers. pKK3535-040329-MS2 was created via iterative whole-plasmid PCR and blunt end ligation using primers h98_MS2_F/R and then primers 040329_F/R.

Polysome profiling and rRNA quantitation

Polysome profiling was performed as described[34] using Mach1 cells expressing the desired 23S rRNA-expressing plasmid. Fractions were collected and extracted twice with 1 volume phenol-chloroform (pH 4.5), washed with 1 volume chloroform and precipitated with 1/10th vol. 3M sodium acetate (pH 5.2) and 2.5 vol. ethanol at -20˚C overnight. Precipitated RNA was washed with 300 µL cold 70% ethanol and resuspended in water. Ribosomal RNA was reverse transcribed using Superscript III reverse transcriptase (ThermoFisher) with up to 2 µg of input rRNA and primer MS2_quant_R (Table S2). RNA was hydrolyzed with 10 M NaOH at 37 ˚C for 30 min and cDNA was precipitated as before with sodium acetate and ethanol. The cDNA pellet was washed with 70% ethanol and resuspended in water, and concentrations were normalized to 1 ng/µL. Quantification of the ratio of MS2-tagged rRNA to WT rRNA was done via semi-quantitative PCR with Q5 DNA polymerase (NEB). Primers MS2_quant_F/R bind outside the MS2 tagged region of 23S cDNA, generating a 146 base pair (bp) band for MS2-tagged molecules and a 114 bp band for WT. Each sample was amplified twice with varying input cDNA amounts or cycle numbers to verify linear PCR amplification. Most samples were optimized around 0.2 ng input cDNA and 8x-12x cycles for a 10 µL PCR reaction. PCR products were run on a 15% acrylamide Tris-Borate-EDTA minigel and stained with SybrSAFE DNA stain. Gel images were quantitated using ImageJ.[35]

Untagged WT Ribosome Purification

WT 50S/30S ribosomal subunits were prepared as described previously[36] except the 50S/30S peak was isolated alone and subjected to an additional spin over a 15-40% sucrose gradient in buffer C (20 mM Tris-HCl pH 7.5, 60 mM NH_4Cl, 6 mM $MgCl_2$, 0.5 mM EDTA, 2 mM DTT). Fractions were concentrated and sucrose was removed using 100 kDa cutoff spin filters (millipore). The purified subunits were frozen in aliquots at -80 °C.

MS2 tagged WT, P7A7 and 040329 Ribosome Crude Purification

E. coli Mach1 cells (ThermoFisher) were transformed with plasmid pLK35-WT-MS2, pLK35-P7A7-MS2, or pKK3535-040329-MS2. Overnight cultures of transformants were diluted 1:100 into 3L LB + 100 µg/mL ampicillin (WT and P7A7) or 50 µg/mL tetracycline (040329). Ribosomes were then prepared as described[36] up through crude ribosome pelleting. The pellet was

resuspended in ~2mL of buffer A (20 mM Tris-HCl pH 7.5, 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, 2 mM DTT) and stored at 4 °C until MS2-tag affinity purification.

MBP-MS2 purification

A pMAL-c2 plasmid (NEB) encoding N-terminally 6xHis tagged MBP-MS2 (Gift from Nadège Liaud, UC-Berkeley) was transformed into BL21 Codon+ RIL chemically competent cells (Agilent). An overnight culture was used to inoculate 3L of ZYM-5052 auto-inducing media (Studier) with 100 µg/mL ampicillin and grown for 16 h at 37 °C. Cells were cooled and all subsequent steps carried out at 4 ˚C. After pelleting at 4000 *g*, cells were washed and resuspended in 130 mL lysis buffer (20 mM HEPES pH 7.5, 250 mM KCl, 10 mM imidazole, 1 protease inhibitor tablet (Pierce) per 50 mL, 2 mM 2-mercaptoethanol) before lysis by sonication. Lysate was clarified by centrifugation (48,000 *g*, 30 min) and filtered through a 0.2 µM filter. MBP-MS2 was initially purified using an Akta fast protein liquid chromatography (FPLC) system (GE) with a 5 mL HisTrap column (GE). The column was washed with 10 column volumes (CV) of lysis buffer and eluted with a 10 CV linear gradient of lysis buffer with 10-500 mM imidazole. Protein containing fractions were combined and dialyzed against a low salt MS2 buffer (MS2-LS) (20 mM HEPES pH 7.5, 20 mM KCl, 1 mM EDTA, 2 mM 2-mercaptoethanol) in 10 kDa cutoff dialysis cassettes overnight. MBP-MS2 was then purified again on an FPLC using a 5 mL Heparin column (GE). Batches of 50-100 mg protein (estimated by A280, ε=83310 $M^{-1}cm^{-1}$) were bound to the column, washed with 5 CV of MS2-LS, and eluted with a linear gradient of MS2-LS with 20 mM-1 M KCl. MBP-MS2 fractions were combined, glycerol added to 10%, and stored in aliquots at -80 ˚C.

MS2-tagged ribosome purification

All steps were performed at 4 °C. Crude ribosomes (>60 mg) were diluted to ~15 mg/mL in buffer A. MBP-MS2 (10 mg) was diluted to 0.5 mg/mL in MS2-150 buffer (20 mM HEPES pH 7.5, 150 mM KCl, 1 mM EDTA, 2 mM 2-mercaptoethanol) and loaded onto a 5 mL MBPtrap column (GE). The column was washed with 5 CV buffer A, loaded with crude ribosomes, and then attached to an FPLC. Ribosomes were purified with the following program: 5 CV wash with buffer A, 5 CV wash with buffer A with 250 mM NH_4Cl, 10 CV elution with a linear gradient of buffer A with 0-10 mM maltose. Ribosome containing fractions were concentrated and washed with buffer A in 100 kDa cutoff spin filters. Concentration was estimated with the conversion 1 nM 70S ribosomes = 24 A260. MS2-purified ribosomes were stored at -80 ˚C.

For the short peptide *in vitro* translations, P7A7 ribosomes were additionally purified at small scale using the same method over ~100 µL (packed volume) amylose resin. Ribosomal purity was assayed using semi-quantitative RT-PCR as described in the polysome profiling assay.

Nanoluciferase in vitro translation

In vitro translation (IVT) reactions used the PURExpress Δribosome kit (NEB) and a nanoluciferase reporter plasmid (Promega). IVTs used 0.5-1 µM purified ribosomal subunits, 10 ng/µL reporter plasmid, and all other components as per the manufacturer's recommendation. Reactions (3 µL) were incubated for 60 min at 37 ˚C and quantified with the Nano-glo assay system (Promega).

Short peptide translation

Formation of acyl-tRNAs used for protein synthesis. Synthesis, purification, and aminoacylation of $tRNA^{Val}$ was carried out using the protocol described previously.[23] Specifically, 10 µL of 250 µM Flexizyme (eFx)[37] was added to 10 µL of either 200 mM HEPES with 200 mM KCl (pH 7.5, for α-Phenylalanine) or 500 mM HEPES (pH 7.5 for β-Phenylalanine) and 10 µL of 250 µM $tRNA^{Val}$. The samples were incubated at 95 ºC for 2 min and allowed to reach room temperature in 5 min. 60 µL of 1 M magnesium chloride was then added, followed by 10 µL of a DMSO solution of each cyanomethyl ester amino acid variant (50 mM). Reactions were incubated at 4 ºC for 60 h. The reactions were quenched by addition of sodium acetate (pH 5.2) to a final concentration of 300 mM and ethanol was added to a final volume of 70% (*v/v*). The samples were then incubated at -80 ºC for 1 h and the RNA was pelleted by centrifugation at 21,300 *g* for 30 min at 4 ºC. The supernatant was removed and the pellet was washed with 500 µL of 70% (*v/v*) ethanol (stored at -20 ºC). The sample was then centrifuged at 21,300 *g* for 7 min at 4 ºC and the supernatant was removed. The pellet was air-dried for 2-5 min either at room temperature or on ice. When used immediately, the pellet was resuspended in 1 mM sodium acetate (pH 5.2). If used at a later date, the pellet was stored dry at -80 ºC and resuspended in 1 mM sodium acetate (pH 5.2) before use.

In vitro translation reactions. Templates for *in vitro* translation of short peptides containing a FLAG tag fMet-Val-Phe-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (MVFDYKDDDDK, fMVF-Flag) were generated as described previously.[23] *In vitro* transcription/translation of (fMVF-Flag was carried out using the combination of PureExpress (ΔtRNA, Δaa (E6840S)) and PureExpress (Δribosome (E3313S)) kits (NEB) with the following modifications. To generate the fMVF-Flag WT peptide reactions contained (25 µL): Solution A (ΔtRNA, Δaa kit), 330 µM methionine, 330 µM valine, solution containing 330 µM tyrosine, 330 µM phenylalanine, and 330 µM lysine, 280 µM aspartic acid (pH 7), tRNA solution (ΔtRNA, Δaa kit), Factors Mix (Δribosome kit), 0.8 µM purified 30S subunits, 0.8 µM purified 50S subunits (either WT or P7A7), 500 ng dsDNA template, and water (to 25 µL). When using $tRNA^{Val}$ (50 µM) charged using eFx, valine was omitted from the reaction mixture. The reactions were then incubated for 6 h at 37 ºC. The reactions were quenched by placing them on ice and adding 25 µL of dilution buffer (10 mM magnesium acetate (Sigma-Aldrich) and 100 mM sodium chloride (Sigma Aldrich)). To remove the proteins and majority of nucleic acid macromolecules, 5 µL of Ni-NTA (Qiagen) slurry was added and the solution was incubated with light agitation at 4 ºC for 50 min. The Ni-NTA resin was removed by centrifugation at 21,300 *g* for 10 min at 4 ºC. The supernatant was then frozen at -80 ºC for 5 min and centrifuged once more at 21,300 *g* for 10 min at 4 ºC. The supernatant was analyzed on a Poroshell 120 EC-C18 column (2.7 µm, 3.0 × 50 mm, 45 ºC, Agilent) using a linear gradient from 5 to 55% acetonitrile over 6.5 min with 0.1% formic acid as the aqueous mobile phase after an initial hold at 95% 0.1% formic acid for 0.5 min (0.6 mL/min) using a 1290 Infinity II UHPLC (G7120AR, Agilent). Peptides were identified using LC-HRMS as described previously.[23]

Magnesium-dependent 70S complex formation

10 pmol WT or P7A7 50S subunits were mixed with 20 pmol WT 30S subunits in 20 µL of buffer C, described above, with 10 mM, 15 mM or 20 mM $MgCl_2$ and incubated for 30 min at 37 ˚C. Reactions were spun over 15-30% sucrose gradients made in buffer C with appropriate $MgCl_2$ concentrations at 178,000 *g* (SW-41 rotor, Beckman Coulter) for 3 hours. A254 traces were measured with an ISCO gradient fractionation system.

EM sample preparation

Samples were deposited onto glow-discharged 300 mesh Quantifoil UltraAuFoil R1.2/1.3 grids with an additional top layer of continuous amorphous carbon floated on. The sample was initially incubated on the grid for approximately 1 minute, after which excess sample was washed off in a buffer containing 20 mM Tris-HCl, pH 7.5, 60 mM NH_4Cl, 6 mM $MgCl_2$, 0.5 mM EDTA, 2mM DTT. A Vitrobot Mark IV was used for plunge-freezing with the settings: 20 °C, 100% humidity, blot force 8, blot time 3 seconds. Inside the Vitrobot, the grid was first side-blotted with filter paper to remove the majority of the solvent, followed by depositing 1.2 μL of buffer prior to blotting and plunging into liquid ethane.

EM Data Collection

Images were collected on an FEI Titan Krios electron microscope operated at 300 keV and with a GIF energy filter. The images were collected on a GATAN K2 Summit camera in super-resolution mode. Magnification was set to 215,000x for a pixel size 0.56 Å (0.28 Å super-resolution size). We collected images using a focal pair approach that we have since determined to have no advantage over a conventional defocus ramp. Briefly, two movies were recorded of each area, the first targeting a defocus level roughly around -0.3 μm (P7A7) or -0.5 μm (WT) with a total dose of 10 e-/ Å2, and the second with a target defocus around -3.5 μm (P7A7) or -3.6 μm (WT) and total dose 20 e-/ Å2. SerialEM[38] was used for automated data collection and Focus[39] was used for real-time monitoring of the data.

Image processing

For all movies, motion correction was performed with MotionCor2.[40] CTF estimation was performed with CTFFind4,[41] and poorly fit micrographs were discarded based upon visual inspection. For ease of working with the focal pair data, far-focused movies were first processed to determine particle position and rough orientation before moving to the near-focused movies for high resolution structure determination.

For P7A7, templates for automatic picking were generated in RELION 2.0[42] by first picking using a Gaussian blob, then classifying the results and choosing the best classes for templates. Three rounds of 2D classification were performed on 92,647 template-picked particles, reducing the number to 79,268 particles. Ab-initio reconstruction with 3 classes was performed in cryoSPARC v1,[43] resulting in one class that resembled a normal 50S subunit and one class that appeared to be completely missing the central protuberance (CP). These two classes were separately submitted for 3D auto-refinement in RELION. The latter structure (with 21,319 particles from the far-focused data) did not refine to high enough resolution to pursue further. The class with an intact CP was subject to a round of 3D classification without alignment. Three out of four classes were pooled and refined to 4.17 Å from far-focused data. The resulting particle positions and orientations were then applied to near-focus movies with custom python scripts, and those particles were refined starting from local angular searches to 3.11 Å with 37,609 particles.

For WT 50S, all processing was done in RELION 2.0.[42] 109,802 particles were auto-picked from far-focused micrographs with a Gaussian blob. Four rounds of 2D classification were performed to yield 99,356 particles for an initial 3D auto-refinement. Processing was shifted to near-focused micrographs as described for P7A7. 3D classification without alignment was performed after an initial near-focused refinement from local searches. The best class, containing 84,372 particles, was refined again for a final resolution of 3.20 Å. Coordinates for the WT 50S

subunit based on PDB entry 4YBB[44] were docked into the density as a rigid body. Molecular graphics were created using ChimeraX.[45]

Cryo-EM map occupancy calculation

Map occupancy at each feature was calculated using a previously developed algorithm.[46] A suitable contour level for each map was chosen after amplitude scaling and resampling by normalizing to the volume of uL4, a known early assembling ribosomal protein. Complete-linkage clustering was used to group structures by occupancy similarity. Results matched well with visual inspection of map features.

LC-MS/MS and TMT-based quantification

Peptide preparation. Ribosomal proteins were precipitated in triplicate from WT-MS2 or P7A7-MS2 purified 50S subunits with 20% trichloroacetic acid at 4 °C for 1 hr. Protein pellets were washed 3x with 500 µL 0.01 M HCl in 90% acetone and dried. Protein was digested and TMT labeled using the TMTduplex isobaric mass tagging kit (ThermoFisher).

Mass spectrometry. Mass spectrometry was performed by the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley. Peptides were analyzed on a ThermoFisher Orbitrap Fusion Lumos Tribrid mass spectrometry system equipped with an Easy nLC 1200 ultrahigh-pressure liquid chromatography system interfaced via a Nanospray Flex nanoelectrospray source. Samples were injected on a C18 reverse phase column (25 cm x 75 µm packed with ReprosilPur C18 AQ 1.9 µm particles). Peptides were separated by a gradient from 5 to 32% acetonitrile in 0.02% heptafluorobutyric acid over 120 min at a flow rate of 300 nL/min. Spectra were continuously acquired in a data-dependent manner throughout the gradient, acquiring a full scan in the Orbitrap (at 120,000 resolution with an AGC target of 400,000 and a maximum injection time of 50 ms) followed by 10 MS/MS scans on the most abundant ions in 3 s in the dual linear ion trap (turbo scan type with an intensity threshold of 5000, CID collision energy of 35%, AGC target of 10,000, maximum injection time of 30 ms, and isolation width of 0.7 m/z). Singly and unassigned charge states were rejected. Dynamic exclusion was enabled with a repeat count of 1, an exclusion duration of 20 s, and an exclusion mass width of ±10 ppm. Data was collected using the MS3 method[47] for obtaining TMT tag ratios with MS3 scans collected in the orbitrap at a resolution of 60,000, HCD collision energy of 65% and a scan range of 100-500.

Data analysis. Protein identification and quantification were done with IntegratedProteomics Pipeline (IP2, Integrated Proteomics Applications, Inc. San Diego, CA) using ProLuCID/Sequest, DTASelect2 and Census [3,4,5,6].[48–50] Tandem mass spectra were extracted into ms1, ms2 and ms3 files from raw files using RawExtractor[51] and were searched against the *E. coli* protein database plus sequences of common contaminants, concatenated to a decoy database in which the sequence for each entry in the original database was reversed.[52] All searches were parallelized and searched on the VJC proteomics cluster. Search space included all fully tryptic peptide candidates with no missed cleavage restrictions. Carbamidomethylation (+57.02146) of cysteine was considered a static modification; TMT tag masses, as given in the TMT kit product sheet, were also considered static modifications. We required 1 peptide per protein and both tryptic termini for each peptide identification. The ProLuCID search results were assembled and filtered using the DTASelect[49] program with a peptide false discovery rate (FDR) of 0.001 for single peptides and a peptide FDR of 0.005 for additional peptide s for the same

protein. Under such filtering conditions, the estimated false discovery rate was zero for the dataset used. Quantitative analysis on MS3-based MultiNotch TMT data was analyzed with Census 2 in IP2 platform.[53] As TMT reagents are not 100% pure, we referred to the ThermoFisher Scientific TMT product data sheet to obtain purity values for each tag and normalized reporter ion intensities. While identification reports best hit for each peptide, Census extracted all PSMs that can be harnessed to increase accuracy from reporter ion intensity variance. Extracted reporter ions were further normalized by using total intensity in each channel to correct sample amount error.

To account for purity differences between ribosomal protein samples, two additional normalization methods were attempted. First, intensities were normalized to the total 50S ribosomal protein signal for each respective mass tag. Second, intensities were normalized to the averaged signals of early assembling proteins uL4, uL13, bL20, uL22, and uL24. These proteins are expected to be stoichiometrically present in both mature WT 50S and assembly-stalled P7A7 subunits. These two methods produced similar results and the latter was chosen for further analysis. Reported proteins were identified in all three relative quantitation experiments with the exception of L35, which was seen in two.

RESULTS

Prior analysis of the P7A7 ribosome and its precursor, 040329, was limited to *in vivo* or lysate-based characterization in the presence of wild-type ribosomes.[9,10] The complexity of cell lysate-based translation and the presence of background WT ribosomes, which can be active in β-amino acid incorporation,[3,24] makes it challenging to accurately understand P7A7's altered substrate scope, necessitating the development of a pure P7A7 translation system. Initial attempts at creating *rrn* knockout 'Squires' strains carrying a P7A7-23S rRNA encoding variant of the *rrnb* operon on a plasmid failed, indicating that P7A7 cannot support cellular growth on its own.[54] While the 12 PTC mutations of P7A7 likely diminish its efficiency at natural protein synthesis, it had been observed to synthesize full-length dihydrofolate reductase *in vivo*, suggesting a level of translational competency amenable to study. We assayed the levels of P7A7 50S ribosomal subunits in polysome fractions from mid-log phase *E. coli* Mach1 cells using a semi-quantitative reverse transcription-PCR (RT-PCR) assay, recognizing an MS2 RNA tag grafted onto helix 98 of the 23S rRNA.[55] P7A7 ribosomes are depleted, but not absent, in both 70S and polysome fractions (**Figure 1A**; **Figure S1A**), indicating that P7A7 50S subunits are deficient in subunit association**.** To remove background WT ribosomes and develop an isolated P7A7 translation system, we purified P7A7 50S subunits using affinity chromatography of MBP-phage MS2 coat protein fusion bound to a tag on helix 98 (**Figure S1B**).

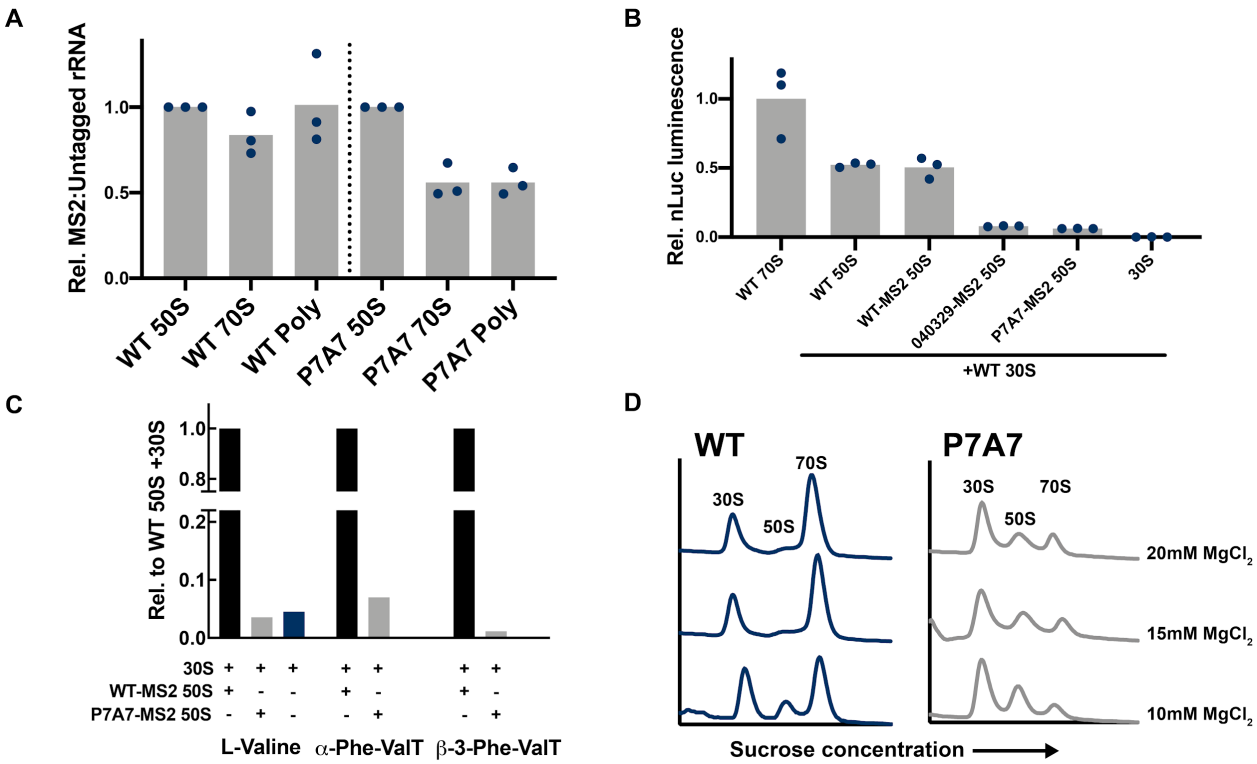


Figure 1. WT and P7A7 ribosome activity. (**A**) WT-MS2:WT or P7A7-MS2:WT rRNA ratios in different polysome profile fractions, normalized to 50S ratio per replicate. Relative quantities of rRNA measured by RT-PCR over the MS2-tag region and quantification of bands resolved by polyacrylamide gel electrophoresis. Bars represent mean values. (**B**) Nanoluciferase expression measured from PURExpress Δ-ribosome *in vitro* translation reactions with different purified ribosomes. Values reported relative to WT 70S. '50S' samples represent individually purified 50S and 30S subunits combined at initiation of reaction. Bars represent means. (**C**) LC-MS monitored MXFDYKDDDDK peptide synthesis from PURExpress Δ-ribosome-tRNA-amino-acid *in vitro* translation reactions with varying purified ribosomes. The tRNA for position X was either Val-tRNA$_{val}$, Phe-tRNA$_{val}$ (flexizyme charged), or β-3-Phe-tRNA$_{val}$ (flexizyme charged). Single replicates are shown for each condition. (**D**) *In vitro* apo-ribosome association at differing Mg^{2+} concentrations, measured by sucrose gradient centrifugation and fractionation for WT 50S + WT 30S and P7A7 50S + WT 30S ribosomes. Representative gradients are shown.

We tested the activity of purified P7A7 50S subunits with WT 30S subunits in the PURExpress Δ-ribosome *in vitro* transcription-translation system.[56] Translation of a standard α-amino acid nanoluciferase reporter with P7A7 was very low relative to untagged and MS2-tagged WT 50S subunits (**Figure 1B**), especially considering observable WT contamination in MS2-tag purifications (**Figure S1B**). Similar results were found with the earlier-identified 040329 ribosome (**Figure 1B**). While *in vitro* translation is known to be far less efficient than seen *in vivo*, the minute activity of purified P7A7 ribosomes was unexpected in light of P7A7's DHFR synthesis[9] and polysome occupancy *in vivo*. After additional purification of P7A7 50S subunits to remove WT contamination, short FLAG-derived α- and β-amino acid containing peptides could not be detected above background levels (**Figure 1C**; **Figure S2**). Because subunit association is a necessary step during translation initiation, we compared the magnesium-dependent subunit association of P7A7 and WT ribosomes. Apo-P7A7 ribosomes require high concentrations of magnesium to form 70S complexes, suggesting a defect in association that could impair translation (**Figure 1D**).

To better understand the basis for P7A7's low activity, we solved the cryo-EM structure of its large subunit, alongside a WT 50S subunit as a control. Images were collected on a 300 kV

Titan Krios microscope with a K2 direct electron detector, and processed with RELION[42] and cryoSPARC.[43] Global resolutions achieved for the two maps were 3.11 Å and 3.20 Å, respectively (**Figure S3**). Intriguingly, although both maps have substantial regions of the 50S subunit at high resolution, large portions of the ribosome that are resolved in the WT structure are absent in the P7A7 map (**Figure 2**). Heterogeneous ab initio reconstructions of the P7A7 50S subunit revealed two classes of particles that represent the large subunit with two different types of disorder. The major class lacks density for ribosomal proteins uL6, uL10, uL11, uL16, bL33, and bL35 (excluding proteins bL31 and bL36 which are known to disassociate during ribosome isolation),[57] as well as multiple stretches of the 23S rRNA comprising the PTC, helices H44, H71, H89-H93, and the end of H39. The smaller class further lacks the central protuberance entirely (**Figure S4** for workflow).

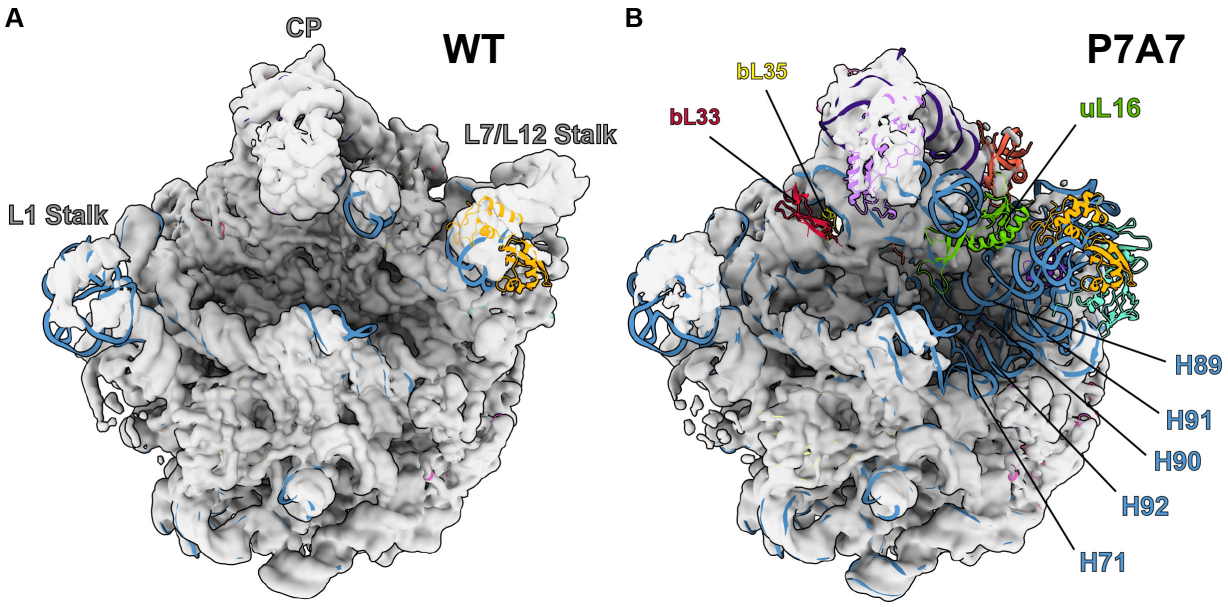


Figure 2. Comparison of WT 50S (A) and P7A7 (B) large subunit cryo-EM maps. Maps are low-pass filtered to 8 Å for clarity of structural features, and 50S coordinates from PDB 4YBB[44] are docked to identify missing components. The central protuberance (CP), L1 Stalk, and L7/L12 Stalk are identified on the WT structure, and features confirmed to be missing from P7A7 by combined cryo-EM and LC-MS/MS analysis are labeled on the P7A7 structure.

Because cryo-EM often cannot resolve whether features of the ribosome are missing due to their absence from the complex or due to flexibility, we also performed relative quantitation of ribosomal proteins (RPs) using liquid chromatography with tandem mass tag (TMT) spectrometry (LC-MS/MS), with TMT labeling on WT and P7A7 50S subunits.[58] Normalized ratios of P7A7:WT RP levels generally matched the structural observations. P7A7 is depleted of RPs uL16, bL33, and bL35 (**Figure 3**). We were able to detect stoichiometric levels of RPs uL6, uL10, and uL11, whose cryo-EM density is often weak due to motion in the large subunit arms.[46,57]

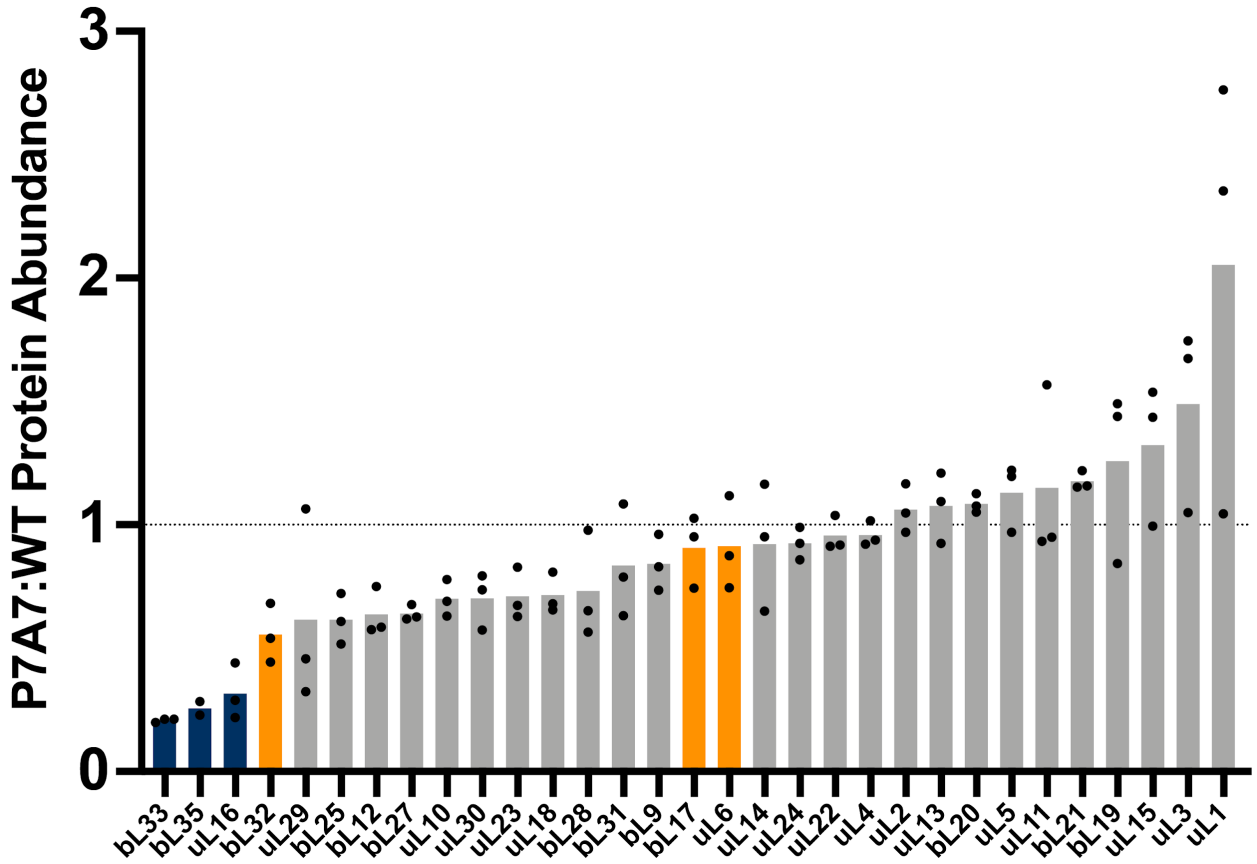


Figure 3. 50S ribosomal protein ratios detected by LC-MS/MS and TMT relative quantitation between P7A7 and WT ribosomes. Bars represent means and are colored by absence in final assembly states (E4 and E5) of Davis and Tan, *et al*. (orange) or absence in both Davis and Tan, *et al.* and Nikolay, *et al.* states (IV and V) (blue).[46,57]

In the vicinity of the P7A7 mutations in the PTC, we observe a stark contrast between well-ordered regions seen in the WT structure, with features like base stacking clearly resolved, and adjacent regions where the density is mainly noise (**Figure 4**). Comparison of these disordered regions in the P7A7 50S subunit map with the corresponding region in the WT PTC shows several interactions that would be perturbed by some of the mutations (**Figure 5**),[44] and may be responsible for the structural defects we observe. In particular, two base triples–the first between C2063, A2450, and C2501, and the second between C2507, G2553, and G2582– would not be able to form because of the C-A and C-U mutations at positions 2063 and 2507, respectively. Other disrupted interactions include the G2057-C2611 and G2505-C2610 base pairs (**Figure 5**).

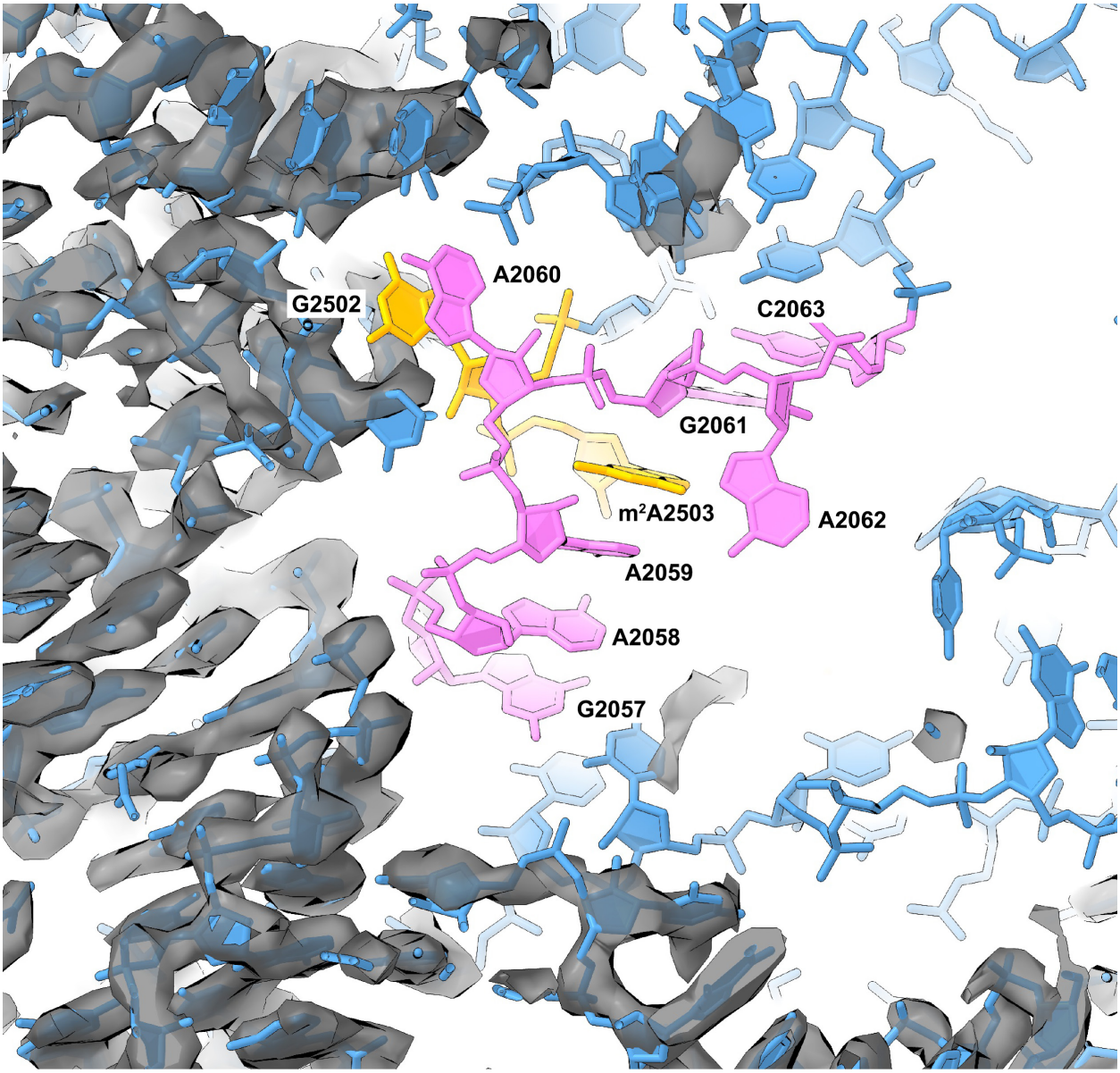


Figure 4. Disorder in the PTC of P7A7 50S subunits. Model of the wild-type 23S rRNA in the vicinity of the PTC superimposed on the map of the P7A7 50S subunit. Regions that are well ordered include helices H26, H35, and several loops from domain II. Mutated PTC residues 2057-2063 and 2502-2503 in P7A7 are depicted in pink and gold, respectively. The "hide dust" feature in ChimeraX[45] was used to remove noise for clarity.

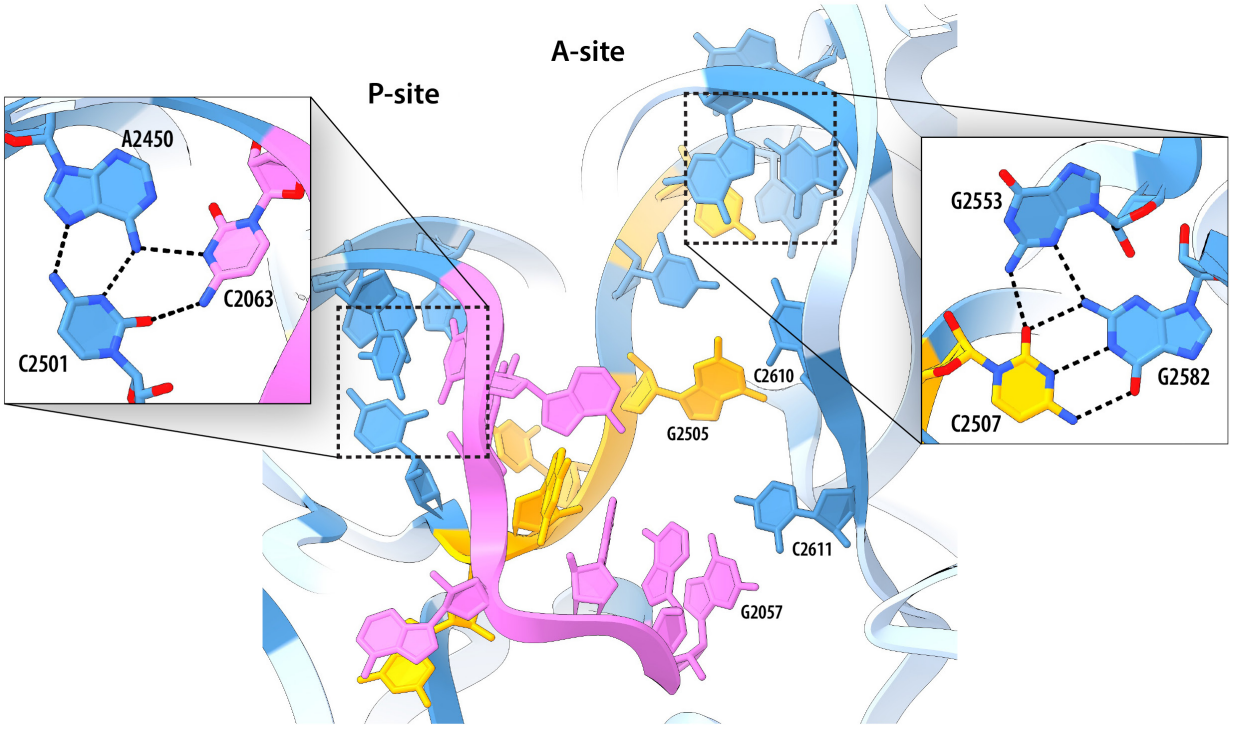


Figure 5. WT 23S rRNA residues directly affected by P7A7 mutations. Coordinates shown are from PDB 4YBB.[44] Mutated residues 2057-2063 and 2502-2505, 2507 in P7A7 are depicted in pink and gold, respectively. Unaltered bases within 3 Å of mutations are shown in darker blue, and those further away are lighter blue. Close-ups highlight base triples that would be disrupted by mutations, and two disrupted base pairs are also labeled.

Low-pass filtering of the P7A7 50S subunit map allowed for some interpretation of broader features of its structure that show poor connectivity and are more difficult to discern in the high-resolution map. Notably, there appears to be some helical RNA density spanning the PTC, which does not fit with the WT model (**Figure 6**). This connectivity is strikingly similar to the non-native PTC organization seen in late intermediates of the 50S assembly pathway both *in vivo* and *vitro*.[46,57] While ribosome biogenesis is a complex process, a general order of rRNA helix formation and ribosomal protein association is known.[59,60] Despite multiple assembly routes, resulting in a spectrum of assembly intermediates, the PTC is usually observed as the last-assembling 50S motif.

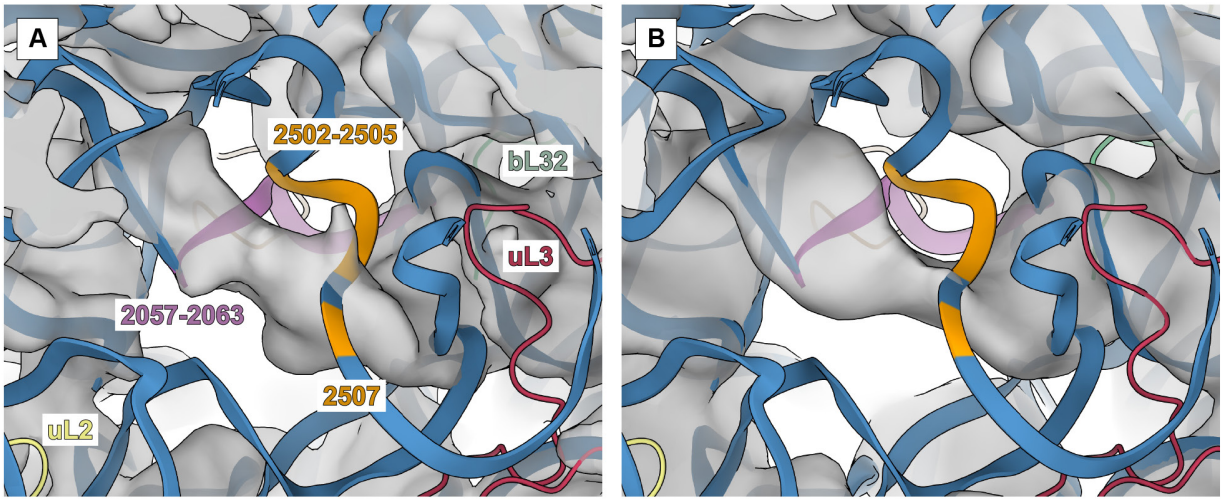


Figure 6. Comparison of PTC in the P7A7 map (A) and an assembly intermediate, E4, determined by Davis and Tan, *et al*. (B). Both maps show helical density that does not appear in fully assembled WT structures. P7A7 mutations regions 2057-2063 and 2502-2505, 2507 are colored in pink and orange, respectively. Other 23S rRNA regions are in blue, and nearby proteins uL2, uL3, and bL32 are yellow, red, and green, respectively.

To further explore the relationship of the observed P7A7 structure with assembly pathways, we carried out a detailed tabulation of structural features across the P7A7 and WT 50S subunit EM maps and compared them to known assembly intermediates, including states IV and V from characterization of *in vitro* reconstitution[57] and states E4 and E5 from *in vivo* intermediates caused by bL17 knockdown.[46] Using a previously developed calculation method[46] we generated a comprehensive list of the cryo-EM occupancies for every ribosomal protein and rRNA helix across all 6 structures (**Figure 7A**). Hierarchical clustering grouped P7A7 alongside other PTC-unfolded states, with P7A7 most closely aligned to E4. However, ribosomal protein occupancy, particularly the absence of bL17 and bL32 from E4, revealed P7A7 as a distinct intermediate. Our TMT quantitation results (**Figure 3**) correlate well with these analyses. Nevertheless, the large-scale organization of several intermediates is quite similar despite the different methods of assembly perturbation, suggesting that a PTC folding defect is a common roadblock in 50S ribosomal subunit maturation. Interestingly, careful analysis of P7A7's migration on a sucrose gradient reveals that it sediments slower than mature WT 50S subunits, similar to the immature 48S particle seen during *in vitro* reconstitution experiments[57] (**Figure 7B**). Together, our structure, mass spectrometry, and biochemical results suggests that P7A7's minimal activity *in vitro* stems primarily from errors in assembly, leaving the ribosome without a mature PTC.

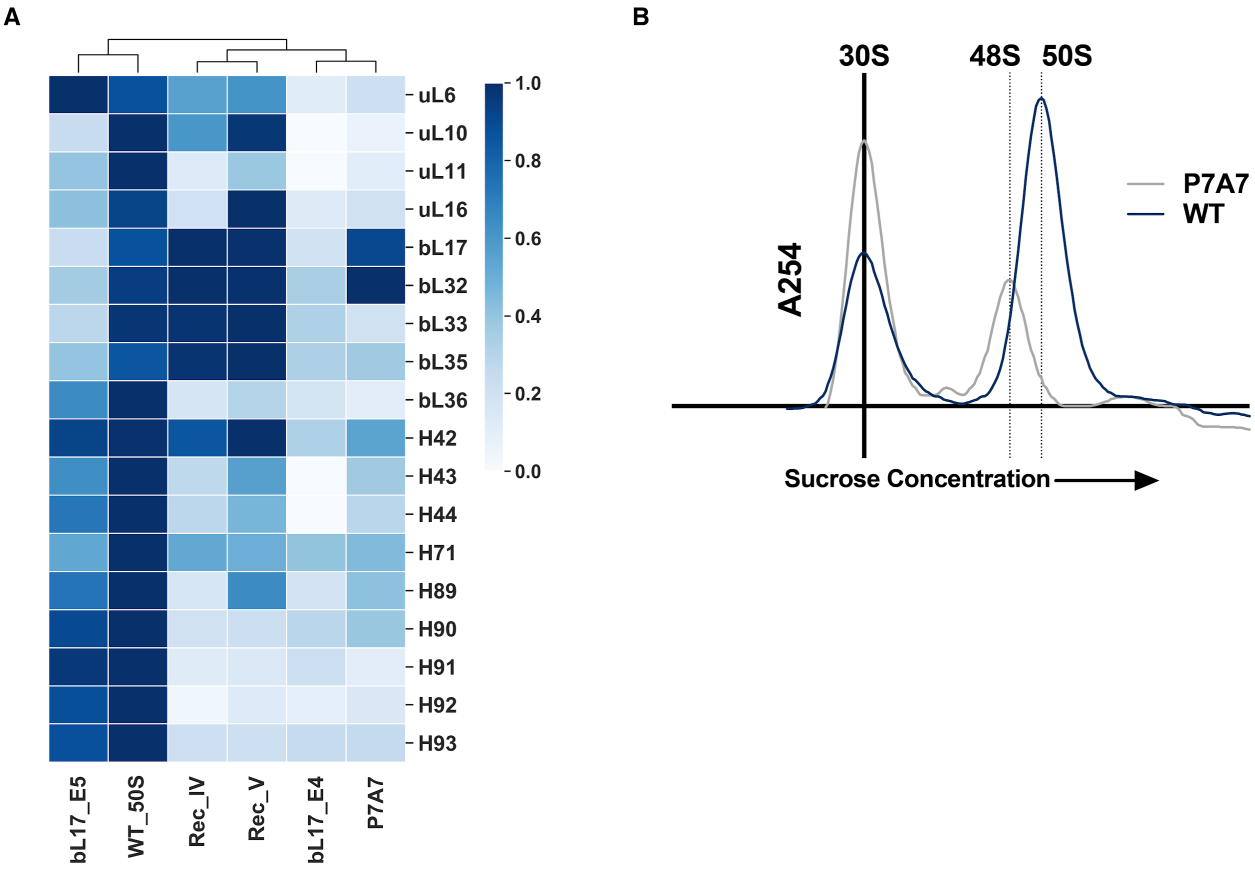


Figure 7. Analysis of P7A7 Assembly. (**A**) Heatmap of 50S subunit rRNA helices and protein occupancy across cryo-EM structures of several assembly intermediates, using the method of Davis and Tan, *et al.*[46] The heatmap is hierarchically clustered across assembly intermediates. Universally present features have been removed for clarity. (**B**) Sucrose gradient fractionation of MS2-tag purified WT and P7A7 50S subunits run with WT 30S subunits as a standard. The 254 nm absorbance traces are shown, and the estimated sedimentation coefficient for P7A7 is marked.

DISCUSSION

Ribosome engineering holds promise to expand efficient template-directed polymer synthesis beyond traditional peptide chemistry, unlocking entirely new classes of materials. Toward this end, there have been several attempts at mutating the ribosome PTC to catalyze the synthesis of difficult sequences or accept new substrates, demonstrating some level of success in expanding the chemistry allowed by the ribosome. Many of these ribosomes, including P7A7 studied here, carry a large number of mutations to highly conserved bases in the PTC and nearby exit tunnel. These base mutations are necessary to overcome the strong optimization of the WT PTC toward natural protein synthesis. However, conservation of nucleotides within the PTC is likely not for catalytic ability alone. Ribosome assembly, and particularly the folding of the protein-sparse PTC, is also dependent on rRNA sequence, presenting a multi-parameter optimization problem when engineering ribosomes. While there is evidence that the ribosome possesses an innate tolerance for mutations to the PTC, even to highly conserved bases,[61] many engineered ribosomes likely possess assembly defects that limit their functionality, even if they perform better than WT ribosomes in certain challenging polymerization reactions. In the case of P7A7 ribosomes, the large number of mutations exceeds the plasticity of the PTC. Further confounding engineering efforts, RNA in general shows less structural stability *in vitro*,[62] possibly

exacerbating assembly defects to the point of complete inactivity. These challenges have prevented detailed study of altered-substrate ribosomes *in vitro*, an important step in understanding how new polymerization chemistries are accessed in a remodeled PTC.

A major confounding factor hampering mechanistic analysis of ribosome activity is the large difference between *in vitro* environments and the cellular milieu. Among the chemical and physical differences between these environments, crowding and divalent salt concentration are commonly implicated for their large effects on RNA structure.[62] Success in improving conditions for RNA folding by mimicking a crowded macromolecular environment (up to 300 mg/mL[63]) provides evidence for the importance of excluded volume effects.[64–66] Notably, *in vitro* translation reactions have been improved by optimizing molecular crowding, reducing agents, and ionic concentrations.[67–69] Nevertheless, even these reconstituted *in vitro* translation systems are several orders of magnitude less complex than the bacterial cytosol. Missing are many small molecules and proteins that can interact with RNA to ensure functional folding. Even cell-lysate based *in vitro* translation systems used to assay mutant ribosomes still include molecular crowding agents.[10,29–31] Additionally, the use of cell lysates suffers from the inability to control the steps of translation critical for determining the activity of mutant or WT ribosomes. For example, WT ribosomes weakly incorporate β-amino acids in cell lysate-based experiments, but are functional in fully reconstituted *in vitro* translation systems with high EF-Tu concentrations, optimized tRNAs, and without competing termination reactions.[3,24]

There are many plausible hypotheses that explain the effect of PTC mutations on ribosome substrate specificity, including opening up space for bulkier groups (as with β-amino acids and dipeptide substrates), properly orienting flexible substrates (such as β-amino acids), and rearranging suboptimal nucleophiles (i.e. poly-proline stretches, D-amino acids). These mechanistic questions will remain challenging without interpretable structures of these engineered ribosomes and a detailed understanding of conformational coupling of ribosome activities. The general disorder of P7A7's PTC (**Figure 4**), in contrast with its *in vivo* activity (**Figure 1A**), suggests that disruption of rigid structures in the PTC may be a contributing mechanism to its acceptance of β-amino acids. It is possible that the PTC of P7A7 is more tolerant of a larger β-substrate only because its flexibility allows an accommodating rearrangement. Analogous promiscuity has been observed in protein enzymes, where flexibility in the active site is correlated with wider substrate specificity.[70–72] Furthermore, the ribosome is often described as an "entropy trap," in that proper substrate positioning is the primary contributor to highly efficient catalysis.[13] The main barrier to β-amino acid polymerization on the ribosome would then be due to improper geometric constraints for β-substrates and not their altered reactivity. An induced-fit mechanism of PTC catalysis has been proposed to be the source of the ribosome's side-chain promiscuity,[15,73] albeit with less disorder than observed in P7A7. As the mutations in P7A7 lead to a less ordered, more flexible PTC fold, we can propose a model in which the additional flexibility confers additional substrate promiscuity. However, this comes at a large cost to overall catalytic efficiency.

Stable RNA folding is particularly important for the PTC, which has a high RNA:protein ratio and folds late in the 50S subunit maturation process.[74–76] With no ribosomal proteins to chaperone or stabilize assembly, the PTC is likely more sensitive to both mutational and environmental disruptions. High magnesium concentrations are often used to stabilize RNA conformations *in vitro*, presumably acting as a stand-in for missing components of the cellular environment. However, *in vivo* RNA structure probing has shown that magnesium can be a poor substitute.[66] Components of the cytosol may keep the highly mutated P7A7 somewhat functional, but extraction to an *in vitro* environment leads to a nonfunctional PTC. Of particular note are two

chaperones, ObgE and EngA, that are proposed to interact with the PTC during assembly and may help direct PTC folding away from inactive intermediates, such as the non-native fold seen in P7A7 (**Figure 5**).[57,77,78] Thus, due to an unstable PTC, P7A7 ribosomes are not suited for detailed *in vitro* study of improved β-amino acid polymerization reactions by the ribosome. Future efforts toward an understanding could include optimizing a cytosol-mimicking buffer for purification and analysis or mapping *in vivo* RNA structure of P7A7 ribosomes to assess whether it possesses a more stable PTC in cells.

The assembly issues with P7A7 suggest that future ribosome engineering efforts may need to consider three layers of optimization. First, the ribosome must be directed toward a novel function, be it antibiotic resistance, translation efficiency, or new polymerization chemistry. Secondly, the engineered ribosome must interact safely with the rest of the cell, avoiding dominant lethal phenotypes by inhibition of overall translation. Finally, ribosomes must productively assemble and remain stable when deployed for synthesis or study. These three goals may often be in conflict, with the necessary mutation of highly conserved rRNA bases towards new ribosome function threatening to derail cell survival or ribosome assembly. To solve part of this problem, much effort has gone into developing orthogonal ribosomes that do not interfere with normal translation. Mutations to anti-Shine-Dalgarno sequences[79–81] and tethering of ribosomal subunits[82,83] have created a platform for powerful ribosome engineering, but there is as of yet no way to select for efficient ribosome assembly. Indeed, some of these engineered orthogonal ribosomes are similarly affected by assembly defects.[11] Although P7A7 was the top hit in a screen for β-amino acid incorporation[9] and there was little reason to believe it harbored significant assembly defects from initial *in vivo* assays, our results highlight that assembly problems as a result of aggressive mutations can go undetected in traditional cell-based experiments. Future efforts directed to incorporate assembly into selections for new ribosome function could open new opportunities for engineering robust translation systems for sequence-defined polymer synthesis.

ACKNOWLEDGMENTS

We thank Bong-Gyoon Han, Dan Toso, and Paul Tobias for assistance with cryo-EM sample preparation and data collection, as well as Lori Kohlstaedt and Carl Ward for mass spectrometry advice. Mass spectrometry was performed by the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory, QB3 Institute, UC Berkeley. The cryo-EM work reported herein was supported by the Center for Genetically Encoded Materials, an NSF Center for Chemical Innovation (NSF CHE-1740549) and by an NSF predoctoral fellowship to Z.W. (1106400). The biochemical studies were supported by the NIH (GM R01-114454). O.A. was supported in part by Agilent Technologies as an Agilent Fellow.

SUPPLEMENTARY INFORMATION

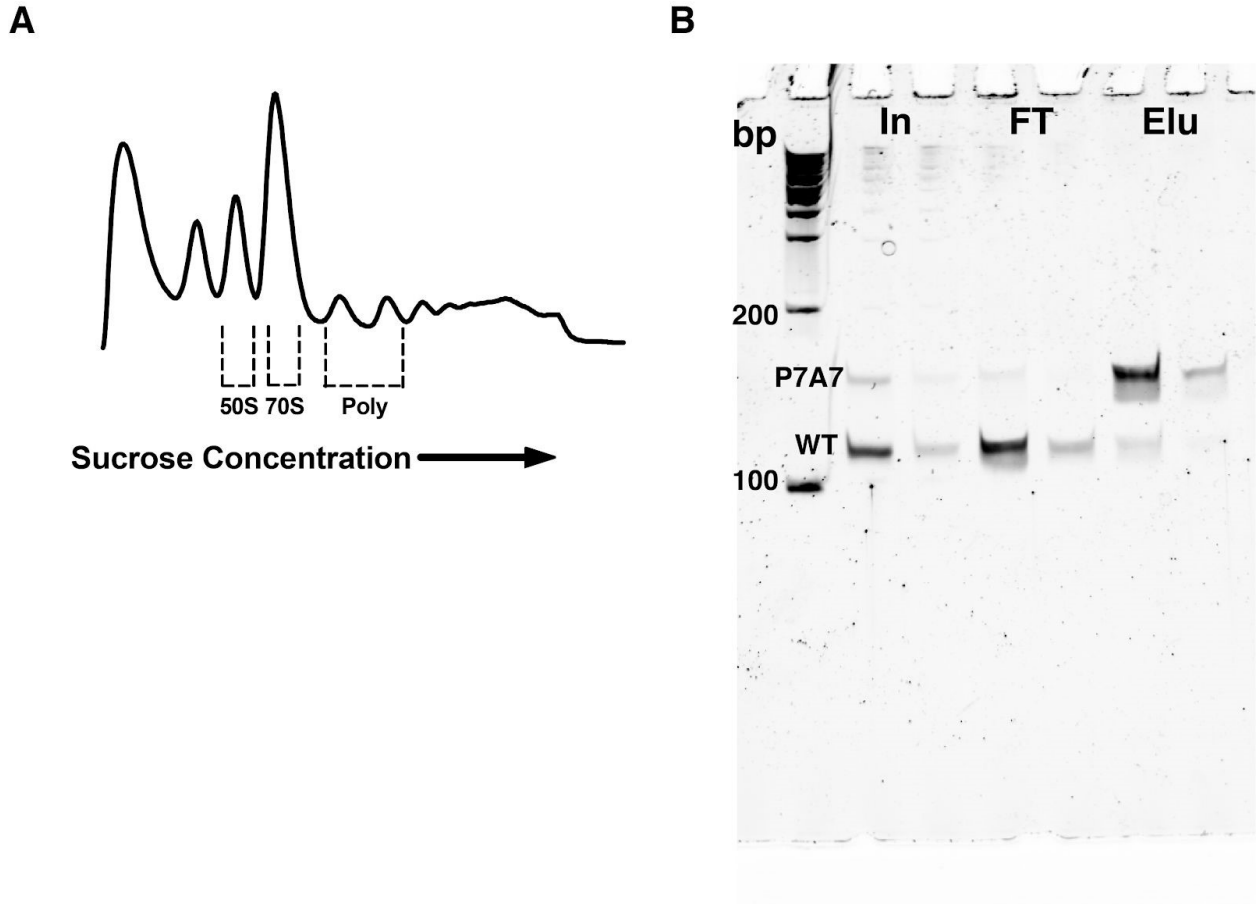


Figure S1. Distribution of P7A7 50S subunits in polysomes and P7A7 purification. (**A**) Representative sucrose gradient fractionation of polysomes showing fractions collected for analysis in **Figure 1A**. The 254 nm absorbance trace is shown. (**B**) P7A7 purification assayed by semi-quantitative RT-PCR over the MS2-tag of isolated rRNA. Input (In), flow through (FT), and elution (Elu) fractions analyzed via two PCRs for each sample, with 0.2 ng and 0.04 ng of input cDNA for the stronger and weaker lanes, respectively.

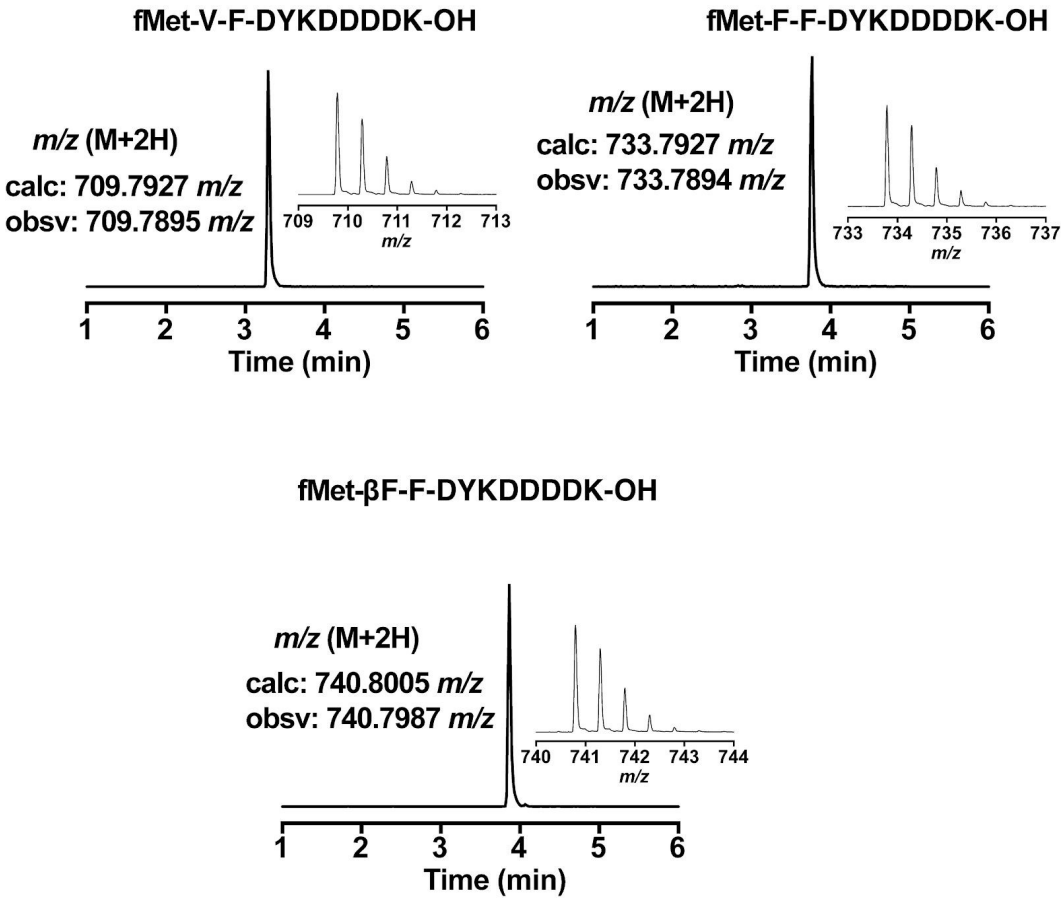


Figure S2. Extracted ion chromatograms and mass spectra of FLAG-derived peptides. Liquid chromatography and mass spectrometry results for the *in vitro* translation experiments in **Figure 1C**.

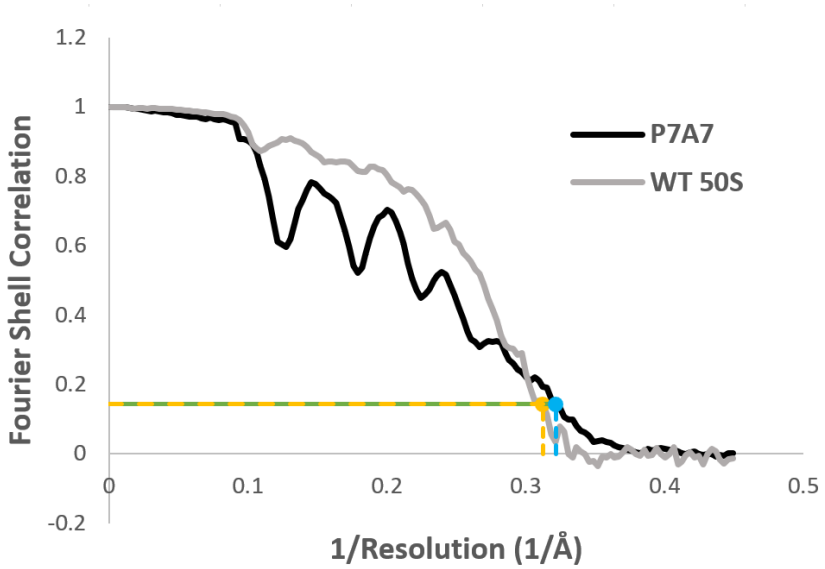


Figure S3. Fourier shell correlation (FSC) curves for P7A7 and WT 50S cryo-EM maps. Oscillations in the FSC for P7A7 are an artifact of the focal pair method of data collection, caused by zeroes in the CTF at the predominant defocus value. The same effect is not observed for the WT 50S subunit map due to different targeting of the near-focus value during the data collection session. The gold-standard FSC cutoff value for global resolution (0.143) is marked in gold for WT and blue for P7A7.

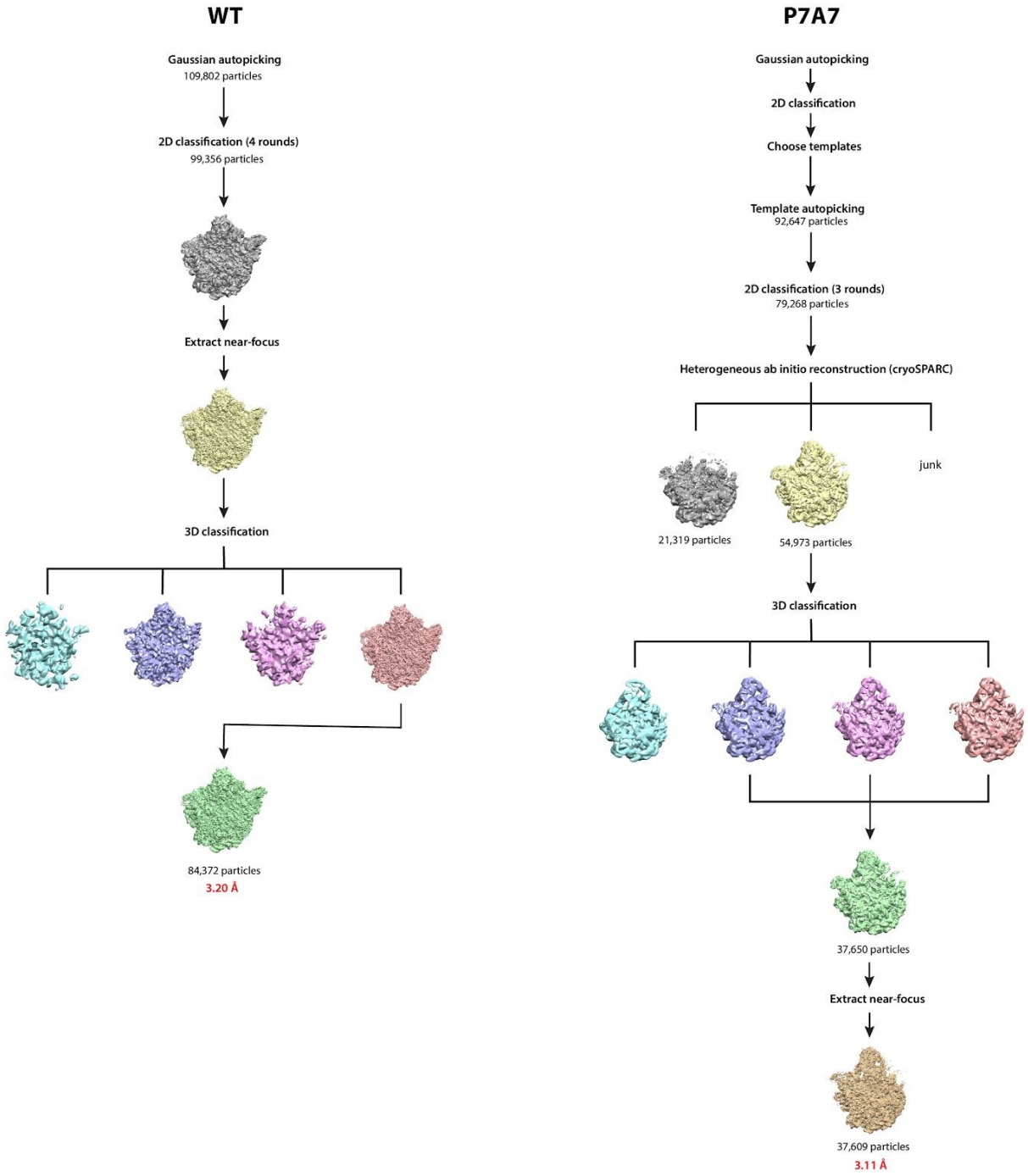


Figure S4. Cryo-EM data processing workflow for P7A7 and WT 50S subunits. Most steps were completed in RELION, otherwise in cryoSPARC where noted. The smaller class of P7A7 subunits lacking the CP is shown in grey.

Table S1. Cryo-EM data collection and refinement information.

	P7A7 large subunit	WT 50S
Microscope	FEI Titan Krios	
Accelerating Voltage (kV)	300	
Detector	GATAN K2 Summit	
Spherical Aberration (mm)	2.7	
Magnification	215,000×	
Defocus targets (μm)	-0.3 and -3.5	-0.5 and -3.6
Micrographs (defocus pairs)	1402	1442
Particles picked	92,647	109,802
Particles refined	37,609	84,372
Resolution achieved (Å)	3.11	3.20

Table S2: Primers used in this study.

h98_MS2_F	CACCCATGTttttgatcagggtcctgaaggaacgttgaag
h98_MS2_R	ATCCTCATGTaaatgatcagggtcagggagaactcatctcg
P7A7_ex_F	aagatgcagtgtacccgcggc Complementary Gibson primer: gccgcgggtacactgcatctt
P7A7_ex_R	gaactgtctcacgacgttctaaaccc Complementary Gibson primer: gggtttagaacgtcgtgagacagttc
MS2_quant_F	cttgccccgagatgagttctccc
MS2_quant_R	gtaccggttagctcaacgcatcgct
040329_F	tggcagggctcatcacatcctggggc
040329_R	gaggtgccaaacaccgcc

REFERENCES

1. Knight, A. S., Zhou, E. Y., Francis, M. B., and Zuckermann, R. N. (2015) Sequence Programmable Peptoid Polymers for Diverse Materials Applications. *Adv. Mater. 27*, 5665–5691.
2. Reynolds, N. M., Vargas-Rodriguez, O., Söll, D., and Crnković, A. (2017) The central role of tRNA in genetic code expansion. *Biochim. Biophys. Acta Gen. Subj. 1861*, 3001–3008.
3. Katoh, T., and Suga, H. (2018) Ribosomal Incorporation of Consecutive β-Amino Acids. *J.*

Am. Chem. Soc. 140, 12159–12167.
4. Vargas-Rodriguez, O., Sevostyanova, A., Söll, D., and Crnković, A. (2018) Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr. Opin. Chem. Biol. 46*, 115–122.
5. Iqbal, E. S., Dods, K. K., and Hartman, M. C. T. (2018) Ribosomal incorporation of backbone modified amino acids via an editing-deficient aminoacyl-tRNA synthetase. *Org. Biomol. Chem. 16*, 1073–1078.
6. Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T., and Sisido, M. (2007) Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc. 129*, 14458–14462.
7. Haruna, K.-I., Alkazemi, M. H., Liu, Y., Söll, D., and Englert, M. (2014) Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic Acids Res. 42*, 9976–9983.
8. Schmied, W. H., Tnimov, Z., Uttamapinant, C., Rae, C. D., Fried, S. D., and Chin, J. W. (2018) Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature 564*, 444–448.
9. Melo Czekster, C., Robertson, W. E., Walker, A. S., Söll, D., and Schepartz, A. (2016) In Vivo Biosynthesis of a β-Amino Acid-Containing Protein. *J. Am. Chem. Soc. 138*, 5194–5197.
10. Dedkova, L. M., Fahmi, N. E., Paul, R., del Rosario, M., Zhang, L., Chen, S., Feder, G., and Hecht, S. M. (2012) β-Puromycin Selection of Modified Ribosomes for in Vitro Incorporation of β-Amino Acids. *Biochemistry 51*, 401–415.
11. Aleksashin, N. A., Leppik, M., Hockenberry, A. J., Klepacki, D., Vázquez-Laslop, N., Jewett, M. C., Remme, J., and Mankin, A. S. (2019) Assembly and functionality of the ribosome with tethered subunits. *Nat. Commun. 10*, 930 1–13.
12. Dedkova, L. M., Fahmi, N. E., Golovine, S. Y., and Hecht, S. M. (2003) Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc. 125*, 6616–6617.
13. Sievers, A., Beringer, M., Rodnina, M. V., and Wolfenden, R. (2004) The ribosome as an entropy trap. *Proc. Natl. Acad. Sci. U. S. A. 101*, 7897–7901.
14. Polikanov, Y. S., Steitz, T. A., and Innis, C. A. (2014) A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat. Struct. Mol. Biol. 21*, 787–793.
15. Martin Schmeing, T., Huang, K. S., Strobel, S. A., and Steitz, T. A. (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature 438*, 520–524.
16. Kawakami, T., Murakami, H., and Suga, H. (2008) Ribosomal Synthesis of Polypeptoids and Peptoid−Peptide Hybrids. *J. Am. Chem. Soc. 130*, 16861–16863.
17. Zhang, B., Tan, Z., Dickson, L. G., Nalam, M. N. L., Cornish, V. W., and Forster, A. C. (2007) Specificity of Translation for N-Alkyl Amino Acids. *J. Am. Chem. Soc. 129*, 11316–11317.
18. Wang, J., Kwiatkowski, M., Pavlov, M. Y., Ehrenberg, M., and Forster, A. C. (2014) Peptide Formation by N-Methyl Amino Acids in Translation Is Hastened by Higher pH and tRNAPro. *ACS Chem. Biol. 9*, 1303–1311.
19. Subtelny, A. O., Hartman, M. C. T., and Szostak, J. W. (2008) Ribosomal Synthesis of N-Methyl Peptides. *J. Am. Chem. Soc. 130*, 6131–6136.
20. Englander, M. T., Avins, J. L., Fleisher, R. C., Liu, B., Effraim, P. R., Wang, J., Schulten,

K., Leyh, T. S., Gonzalez, R. L., Jr, and Cornish, V. W. (2015) The ribosome can discriminate the chirality of amino acids within its peptidyl-transferase center. *Proc. Natl. Acad. Sci. U. S. A. 112*, 6038–6043.
21. Rogers, J. M., Kwon, S., Dawson, S. J., Mandal, P. K., Suga, H., and Huc, I. (2018) Ribosomal synthesis and folding of peptide-helical aromatic foldamer hybrids. *Nat. Chem. 10*, 405–412.
22. Tsiamantas, C., Kwon, S., Douat, C., Huc, I., and Suga, H. (2019) Optimizing aromatic oligoamide foldamer side-chains for ribosomal translation initiation. *Chem. Commun. 55*, 7366–7369.
23. Ad, O., Hoffman, K. S., Cairns, A. G., Featherston, A. L., Miller, S. J., Söll, D., and Schepartz, A. (2019) Translation of Diverse Aramid- and 1,3-Dicarbonyl-peptides by Wild Type Ribosomes in Vitro. *ACS Cent. Sci. 5*, 1289–1294.
24. Fujino, T., Goto, Y., Suga, H., and Murakami, H. (2016) Ribosomal Synthesis of Peptides with Multiple β-Amino Acids. *J. Am. Chem. Soc. 138*, 1962–1969.
25. Edsall, J. T., and Blanchard, M. H. (1933) The Activity Ratio of Zwitterions and Uncharged Molecules in Ampholyte Solutions. The Dissociation Constants of Amino Acid Esters. *J. Am. Chem. Soc. 55*, 2337–2353.
26. Geueke, B., Heck, T., Limbach, M., Nesatyy, V., Seebach, D., and Kohler, H.-P. E. (2006) Bacterial β-peptidyl aminopeptidases with unique substrate specificities for β-oligopeptides and mixed β,α-oligopeptides. *FEBS J. 273*, 5261–5272.
27. Guichard, G., Zerbib, A., Le Gal, F. A., Hoebeke, J., Connan, F., Choppin, J., Briand, J. P., and Guillet, J. G. (2000) Melanoma peptide MART-1(27-35) analogues with enhanced binding capacity to the human class I histocompatibility molecule HLA-A2 by introduction of a beta-amino acid residue: implications for recognition by tumor-infiltrating lymphocytes. *J. Med. Chem. 43*, 3803–3808.
28. Daniels, D. S., Petersson, E. J., Qiu, J. X., and Schepartz, A. (2007) High-resolution structure of a beta-peptide bundle. *J. Am. Chem. Soc. 129*, 1532–1533.
29. Maini, R., Dedkova, L. M., Paul, R., Madathil, M. M., Chowdhury, S. R., Chen, S., and Hecht, S. M. (2015) Ribosome-Mediated Incorporation of Dipeptides and Dipeptide Analogues into Proteins in Vitro. *J. Am. Chem. Soc. 137*, 11206–11209.
30. Maini, R., Nguyen, D. T., Chen, S., Dedkova, L. M., Chowdhury, S. R., Alcala-Torano, R., and Hecht, S. M. (2013) Incorporation of β-amino acids into dihydrofolate reductase by ribosomes having modifications in the peptidyltransferase center. *Bioorg. Med. Chem. 21*, 1088–1096.
31. Maini, R., Chowdhury, S. R., Dedkova, L. M., Roy, B., Daskalova, S. M., Paul, R., Chen, S., and Hecht, S. M. (2015) Protein Synthesis with Ribosomes Selected for the Incorporation of β-Amino Acids. *Biochemistry 54*, 3694–3706.
32. Dedkova, L. M., and Hecht, S. M. (2019) Expanding the Scope of Protein Synthesis Using Modified Ribosomes. *J. Am. Chem. Soc. 141*, 6430–6447.
33. Douthwaite, S., Powers, T., Lee, J. Y., and Noller, H. F. (1989) Defining the Structural Requirements for a Helix in 23 S Ribosomal RNA that Confers Erythromycin Resistance. *J. Mol. Biol. 209*, 655–665.
34. Qin, D., and Fredrick, K. (2013) Chapter Eight - Analysis of Polysomes from Bacteria, in *Methods in Enzymology* (Lorsch, J., Ed.), pp 159–172. Academic Press.
35. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods 9*, 671–675.

36. Travin, D. Y., Watson, Z. L., Metelev, M., Ward, F. R., Osterman, I. A., Khven, I. M., Khabibullina, N. F., Serebryakova, M., Mergaert, P., Polikanov, Y. S., Cate, J. H. D., and Severinov, K. (2019, July 18) Phazolicin – a Novel Thiazole/Oxazole-Modified Peptide Inhibiting the Bacterial Ribosome in a Species-Specific Way. *bioRxiv*.
37. Goto, Y., Katoh, T., and Suga, H. (2011) Flexizymes for genetic code reprogramming. *Nat. Protoc. 6*, 779–790.
38. Mastronarde, D. N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol. 152*, 36–51.
39. Biyani, N., Righetto, R. D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K. N., and Stahlberg, H. (2017) Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol. 198*, 124–133.
40. Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods 14*, 331–332.
41. Rohou, A., and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol. 192*, 216–221.
42. Scheres, S. H. W. (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol. 180*, 519–530.
43. 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.
44. 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.
45. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci. 27*, 14–25.
46. Davis, J. H., Tan, Y. Z., Carragher, B., Potter, C. S., Lyumkis, D., and Williamson, J. R. (2016) Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell 167*, 1610–1622.e15.
47. Ting, L., Rad, R., Gygi, S. P., and Haas, W. (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods 8*, 937–940.
48. Xu, T., Venable, J. D., Park, S. K., Cociorva, D., and III. Yates, J. R. (2006) ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program. *Mol. Cell. Proteomics 5*, S174–S174.
49. Cociorva, D., L Tabb, D., and Yates, J. R. (2007) Validation of tandem mass spectrometry database search results using DTASelect. *Curr. Protoc. Bioinformatics Chapter 13*, Unit 13.4 1–14.
50. Park, S. K., Venable, J. D., Xu, T., and Yates, J. R., 3rd. (2008) A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat. Methods 5*, 319–322.
51. McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J., Venable, J., Graumann, J., Johnson, J. R., Cociorva, D., and Yates, J. R., 3rd. (2004) MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun. Mass Spectrom. 18*, 2162–2168.
52. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry

(LC/LC−MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *J. Proteome Res. 2*, 43–50.
53. Park, S. K. R., Aslanian, A., McClatchy, D. B., Han, X., Shah, H., Singh, M., Rauniyar, N., Moresco, J. J., Pinto, A. F. M., Diedrich, J. K., Delahunty, C., and Yates, J. R., 3rd. (2014) Census 2: isobaric labeling data analysis. *Bioinformatics 30*, 2208–2209.
54. Asai, T., Zaporojets, D., Squires, C., and Squires, C. L. (1999) An Escherichia coli strain with all chromosomal rRNA operons inactivated: Complete exchange of rRNA genes between bacteria. *Proc. Natl. Acad. Sci. U. S. A. 96*, 1971–1976.
55. Youngman, E. M., and Green, R. (2005) Affinity purification of in vivo-assembled ribosomes for in vitro biochemical analysis. *Methods 36*, 305–312.
56. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol. 19*, 751–755.
57. Nikolay, R., Hilal, T., Qin, B., Mielke, T., Bürger, J., Loerke, J., Textoris-Taube, K., Nierhaus, K. H., and Spahn, C. M. T. (2018) Structural Visualization of the Formation and Activation of the 50S Ribosomal Subunit during In Vitro Reconstitution. *Mol. Cell 70*, 881–893.e3.
58. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A. K. A., and Hamon, C. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem. 75*, 1895–1904.
59. Chen, S. S., Sperling, E., Silverman, J. M., Davis, J. H., and Williamson, J. R. (2012) Measuring the dynamics of E. coli ribosome biogenesis using pulse-labeling and quantitative mass spectrometry. *Mol. Biosyst. 8*, 3325–3334.
60. Herold, M., and Nierhaus, K. H. (1987) Incorporation of six additional proteins to complete the assembly map of the 50 S subunit from Escherichia coli ribosomes. *J. Biol. Chem. 262*, 8826–8833.
61. Rakauskaitė, R., and Dinman, J. D. (2011-5) Mutations of highly conserved bases in the peptidyltransferase center induce compensatory rearrangements in yeast ribosomes. *RNA 17*, 855–864.
62. Leamy, K. A., Assmann, S. M., Mathews, D. H., and Bevilacqua, P. C. (2016/ed) Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biophys. 49*, 1–26.
63. Zimmerman, S. B., and Trach, S. O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol. 222*, 599–620.
64. Kilburn, D., Roh, J. H., Guo, L., Briber, R. M., and Woodson, S. A. (2010) Molecular Crowding Stabilizes Folded RNA Structure by the Excluded Volume Effect. *J. Am. Chem. Soc. 132*, 8690–8696.
65. Strulson, C. A., Yennawar, N. H., Rambo, R. P., and Bevilacqua, P. C. (2013) Molecular Crowding Favors Reactivity of a Human Ribozyme Under Physiological Ionic Conditions. *Biochemistry 52*, 8187–8197.
66. Tyrrell, J., McGinnis, J. L., Weeks, K. M., and Pielak, G. J. (2013) The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry 52*, 8777–8785.
67. Jelenc, P. C., and Kurland, C. G. (1979-7) Nucleoside triphosphate regeneration decreases the frequency of translation errors. *Proc. Natl. Acad. Sci. U. S. A. 76*, 3174–3178.
68. Jewett, M. C., and Swartz, J. R. (2004) Mimicking the Escherichia coli cytoplasmic

environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng. 86*, 19–26.
69. Fritz, B. R., Jamil, O. K., and Jewett, M. C. (2015) Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction. *Nucleic Acids Res. 43*, 4774–4784.
70. Gatti-Lafranconi, P., and Hollfelder, F. (2013) Flexibility and Reactivity in Promiscuous Enzymes. *Chembiochem 14*, 285–292.
71. Skopalík, J., Anzenbacher, P., and Otyepka, M. (2008) Flexibility of Human Cytochromes P450: Molecular Dynamics Reveals Differences between CYPs 3A4, 2C9, and 2A6, which Correlate with Their Substrate Preferences. *J. Phys. Chem. B 112*, 8165–8173.
72. Ekroos, M., and Sjogren, T. (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proceedings of the National Academy of Sciences 103*, 13682–13687.
73. Lehmann, J. (2017) Induced fit of the peptidyl-transferase center of the ribosome and conformational freedom of the esterified amino acids. *RNA 23*, 229–239.
74. Polacek, N., and Mankin, A. S. (2005) The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit. Rev. Biochem. Mol. Biol. 40*, 285–311.
75. Li, N., Chen, Y., Guo, Q., Zhang, Y., Yuan, Y., Ma, C., Deng, H., Lei, J., and Gao, N. (2013) Cryo-EM structures of the late-stage assembly intermediates of the bacterial 50S ribosomal subunit. *Nucleic Acids Res. 41*, 7073–7083.
76. Jomaa, A., Jain, N., Davis, J. H., Williamson, J. R., Britton, R. A., and Ortega, J. (2014) Functional domains of the 50S subunit mature late in the assembly process. *Nucleic Acids Res. 42*, 3419–3435.
77. Feng, B., Mandava, C. S., Guo, Q., Wang, J., Cao, W., Li, N., Zhang, Y., Zhang, Y., Wang, Z., Wu, J., Sanyal, S., Lei, J., and Gao, N. (2014) Structural and Functional Insights into the Mode of Action of a Universally Conserved Obg GTPase. *PLoS Biol. 12*, e1001866 1–14.
78. Zhang, X., Yan, K., Zhang, Y., Li, N., Ma, C., Li, Z., Zhang, Y., Feng, B., Liu, J., Sun, Y., Xu, Y., Lei, J., and Gao, N. (2014) Structural insights into the function of a unique tandem GTPase EngA in bacterial ribosome assembly. *Nucleic Acids Res. 42*, 13430–13439.
79. Hui, A., Jhurani, P., and de Boer, H. A. (1987) Directing ribosomes to a single mRNA species: a method to study ribosomal RNA mutations and their effects on translation of a single messenger in Escherichia coli. *Methods Enzymol. 153*, 432–452.
80. Rackham, O., and Chin, J. W. (2005) A network of orthogonal ribosome·mRNA pairs. *Nat. Chem. Biol. 1*, 159–166.
81. Chubiz, L. M., and Rao, C. V. (2008) Computational design of orthogonal ribosomes. *Nucleic Acids Res. 36*, 4038–4046.
82. Orelle, C., Carlson, E. D., Szal, T., Florin, T., Jewett, M. C., and Mankin, A. S. (2015) Protein synthesis by ribosomes with tethered subunits. *Nature 524*, 119–124.
83. Fried, S. D., Schmied, W. H., Uttamapinant, C., and Chin, J. W. (2015) Ribosome Subunit Stapling for Orthogonal Translation in E. coli. *Angew. Chem. Int. Ed Engl. 54*, 12791–12794.

Chapter 5: Collection of cryo-EM movies in focal pairs as a strategy for preservation of high-resolution information

INTRODUCTION

Alongside hardware and software advances in cryo-electron microscopy (cryo-EM), the protocol for data collection is another candidate to explore for improvement in data quality, outside the wet lab. In addition to their beam sensitivity, biological macromolecules are weak phase objects in phase-contrast imaging, which require some under-focus ("defocus") to create contrast at low spatial frequencies[1]. As defocus is increased, the contrast transfer function (CTF) oscillates more rapidly at higher spatial frequencies (**Eqn. 1**) and is more strongly attenuated by the envelope function, making it more difficult to recover this information[2]. Typically, the strategy to get around these limitations is to collect images with a range of defocus values that allow for the particles to be seen by eye and to ensure adequate coverage of higher-frequency areas in Fourier space. However, there may be advantages to finding ways to make close-to-focus images wieldier for processing.

$$CTF = \sin\left(2\pi\left[\frac{C_s}{4}\lambda^3 s^4 - \frac{\Delta Z}{2}\lambda s^2\right]\right)$$

Equation 1. Formula for the phase contrast transfer function (CTF). C_s is the spherical aberration coefficient, ΔZ is the defocus, λ is the electron wavelength, and s is the spatial frequency.

Besides the universal goal of high resolution, there are some cases where low contrast is a particular concern, especially for relatively small particle size (< 100 kDa), or for samples that require detergents or other contrast-limiting agents for stabilization. A prominent existing approach is the use of a phase plate to introduce an additional phase shift to the scattered beam after it hits the sample[3,4]. Phase plates can be useful for overcoming low contrast at or near focus, however, the installation, calibration, and cost of a phase plate are non-trivial; additionally, enthusiasm for the most well-established types of phase plates may be waning as a greater number of small protein structures are able to be determined without the use of one[5]. Newer laser phase plate technology is emerging, but is still in development[6].

Here, an alternative approach to conventional cryo-EM data collection was investigated wherein movies were collected twice for each acquisition target: an initial exposure close to focus, followed by a second exposure further from focus. The first movie captures high-resolution information from the specimen before it suffers too much damage from the beam, while the second movie serves to produce images with high contrast at low frequencies. The processing strategy begins with the second movie, which allows for ease of working with the data in early stages where images are generally binned and/or low-pass filtered regardless. This is sufficient for preliminary "cleaning" of the set of particle images and initial assignment of Euler angles for the map reconstruction, which then provides a starting point from which the close-to-focus movies can be further processed and refined to high resolution.

The concept of focal pairs (or a focal series) is not new[7]. A paper by Ludtke and Chiu from 2003 describes a method to combine the useful information from the pairs into single, merged images[8]. More recently, the method has also been used for cryo-electron tomography (cryo-ET), where thickness of the sample as it is tilted attenuates resolution more severely than in single-particle work[9]. However, these publications either predated the use of or did not use direct electron detectors, and therefore suffered from the generally poorer image quality available at the time (and in the case of tomography, the resolution goals are much lower to begin with). Indeed, this was before atomic resolution with single-particle cryo-EM was remotely possible. Atomic resolution is now becoming a more realistic goal with direct detectors and other advances, bringing a new relevance to optics-based techniques focused on harvesting the highest-resolution information. Focal pairs were collected by Rickgauer, Grigorieff, and Denk in a 2017 publication describing the use of high-resolution projection matching with tomograms[10], which was inspiration for the work described here.

The ribosome was used here as a test case for focal pairs. At the time of data collection (a period spanning late 2017-2018), direct detectors had been commonly in use for some time, but ribosome structures were still only beginning to break much beyond the 3-Å barrier[11,12]. As a specimen that stands to reap the potential benefits of focal pairs, the ribosome does not suffer from inherently reduced contrast, but the well-ordered nature of the large subunit makes it an attractive choice for testing out new methods meant to push high resolution. Additionally, because higher amounts of defocus result in greater delocalization of high-resolution information in the micrographs, a reduction of this effect in close-to-focus images was expected to be particularly important for a large particle like the ribosome, where the particle image size can become computationally expensive.

METHODS

In the development of the focal pairs protocol, a total of six datasets of either 70S ribosomes or large ribosomal subunits from *Escherichia coli* were collected on which to apply the approach, three of which are discussed here. Two of these datasets were collected as a part of another project, described more fully in Chapter 4. All data collection was performed on a Titan Krios microscope at 300 kV accelerating voltage and equipped with a GIF energy filter and K2 camera. For the specific purpose of investigating focal pairs an apo 70S ribosome dataset was collected, targeting -0.3 μm defocus (with somewhat poor accuracy, as discussed below) for the first (10-frame) movie, with a total dose of 10 $e^-/Å^2$, and -3.5 μm defocus for the second (20-frame) movie, with a dose of 20 $e^-/Å^2$. A control ("conventional") apo 70S ribosome dataset was also collected with a more traditional defocus ramp from -0.4 to -1.5 μm and a total dose of 30 $e^-/Å^2$ split over 30 frames (to equal the combined dose for the near- and far-focused movies). The pixel size was 0.56 Å in the linear dimension. Movies for these two datasets were collected on the same grid, during the same session. Automated data collection was performed with SerialEM[13]. The salient results for these, as well as another conventional 70S dataset that was collected and processed around the same time (described further in Appendix 1), are summarized in **Table 1**.

The biochemical preparation and cryo-EM grid preparation for the apo 70S ribosomes was as described for the sample in Appendix 1, without the addition of phazolicin to the complex. The general computational workflow for focal pairs is described in Methods, Chapter 4, where this method was used to obtain the P7A7 structure and the wildtype 50S subunit structure. The main difference for processing the apo 70S focal pair dataset, compared to those in Chapter 4, was that

additional steps were taken to perform Bayesian polishing[14] and per-particle CTF correction in RELION 3[15]. Also, a large subunit (50S) mask was used for focused refinement.

The control apo 70S dataset was processed briefly as follows, in the beta version of RELION 3[15]: movies were aligned with MotionCor2[16] and CTF values were estimated with CTFFind4[17]. Particles were picked initially with a gaussian blob, then 2D classes of these particles were used as templates to re-pick particles from the micrographs. The template-picked particles were subject to 2D classification with 4× binning of the particle images. Because of the small pixel size and computational limitations, further steps were carried out with 2× binned particle images (1.12 Å/pixel). Particle orientations were assigned with 3D auto-refine. As the sample contained apo ribosomes, it was expected that there would be a mixture of ribosomes in the rotated and non-rotated states. Therefore, 3D classification without alignment (4 classes) was used to obtain a subset of more conformationally homogeneous particles. The largest and most well-resolved class was selected for further processing, comprising 11,575 particles. CTF refinement and Bayesian polishing were then performed before focused refinement on the 50S subunit.

RESULTS AND DISCUSSION

The results for the 70S focal pair and 70S control datasets are among those summarized in **Table 1**, and the "gold-standard" Fourier shell correlation (GS-FSC) curves and overall structures are shown in **Figure 1**. Similar resolution was achieved for the two sets of data, which were collected on the same grid. Surprisingly, this was achieved with only 11,575 particles for the control dataset, compared to 70,786 for the focal pairs (using only close-to-focus images). Moreover, truncating the set of particles for the focal pair dataset to match the control severely reduced resolution, to ~11 Å (data not shown). Taken together with the resolutions and particle numbers of the additional focal pair and conventional datasets summarized in **Table 1**, the results indicate overall that the use of focal pairs has no obvious benefits, and may even have detrimental effects on the result. Grid-to-grid variability is a significant challenge in cryo-EM and is most certainly at play here, and in fact, the apo 70S ribosome datasets point to significant square-to-square variation as well (discussed further below). Therefore, whether or not focal pairs may have any advantage at all is not strictly disproven based on resolution and particle number from these experiments alone. However, it is clear that in view of the many other factors at play between grids and microscope sessions, any potential advantages of focal pairs would be negligible if they exist.

Table 1. Summary of results for select focal pair and conventional datasets.

Sample	Collection method	GS-FSC resolution (Å)	No. particles in final reconstruction
Apo 70S ribosome*	Focal pairs	3.00	70,786
Wild type 50S subunit	Focal pairs	3.20	84,372
P7A7 large subunit	Focal pairs	3.11	37,609
Apo 70S ribosome*	Conventional	2.94	11,575
70S ribosome with PHZ	Conventional	2.87	65,393

*Denotes datasets collected on the same grid for evaluation of focal pairs vs. conventional defocus ramp.

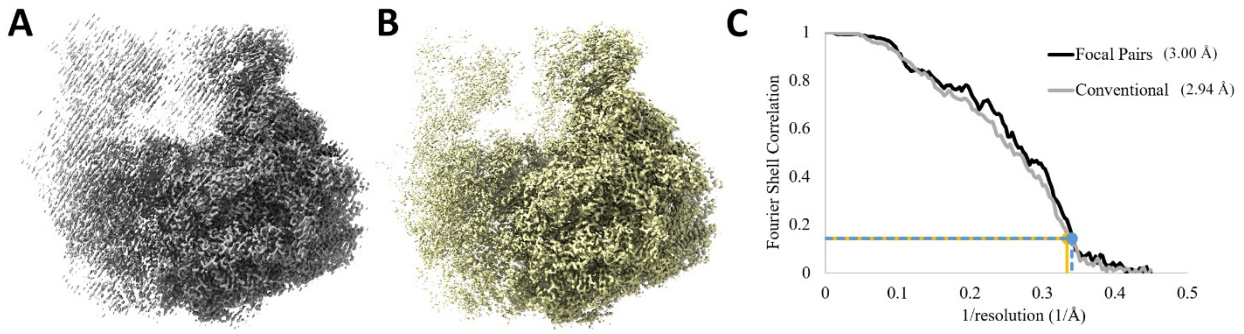


Figure 1. Overview of 70S ribosome structures obtained with (**A**) focal pairs and (**B**) conventional defocus ramp. The small subunit in both cases is expectedly poorly resolved after focused refinement on the large subunit. (**C**) FSC curves for focal pairs and conventional structure. Yellow and Blue lines indicate resolution at the gold-standard FSC cutoff value of 0.143 for the conventional and focal pair datasets, respectively.

Aside from the failure to demonstrate gains from collecting focal pairs, disadvantages were apparent. The most obvious of these is the time cost by taking an additional movie for each acquisition target. A challenge in implementing this approach was that the auto-focusing script sometimes struggled with the close-to-focus condition, possibly because of the gold foil support. This complication affected the target defocus (in terms of both accuracy and precision) to a varying degree for different microscope sessions. Histograms of the defocus values for three focal pair datasets (**Figure 2**) show the variability in the defocus targeting for the close-to-focus movies; while the same defocus was desired in each case, there was a failure to achieve this.

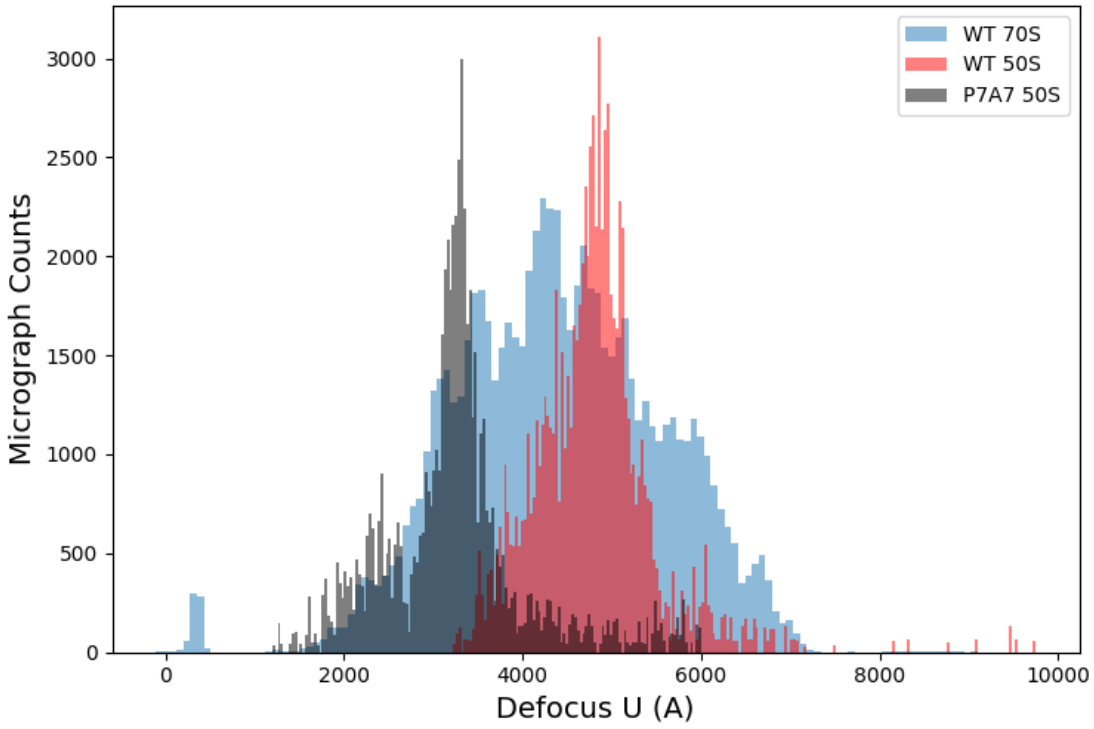


Figure 2. Histogram of defocus values for three different focal pair datasets. The peak of the distribution for the P7A7 dataset, which was overall closest-to-focus on average, occurs at about 330 nm. For the wild type (WT) 50S distribution, the peak occurs at around 480 nm.

Another drawback came from successfully aiming too precisely for a single target defocus value, which is readily apparent in the FSC curve obtained for the P7A7 map (**Figure S3, Chapter 4**). It was anticipated that being as close to focus as possible would make it easier to recover high-resolution information due to the lessened attenuation of signal at lower defocus values, perhaps also aided by the presence of fewer CTF crossover points (zeros) in the range of expected map resolution. However, it turned out that in this small range of defocus values there was not enough coverage of spatial frequencies around the zeros to avoid artifacts. It is notable that for the wild type (WT) 50S structure, although the precision of defocus targeting was similar, the same effects were not present. This is likely because at a higher median defocus value, the small variations between micrographs were enough to sufficiently shift the more closely-spaced CTF zeros and maxima as to perform corrections at all spatial frequencies. The P7A7 FSC curve is reproduced in **Figure 3** along with a simulated CTF at the peak in its defocus distribution from **Figure 2**. Panel B shows a simulated CTF at the same peak for the WT 50S structure, highlighting the smaller spacing of CTF crossovers.

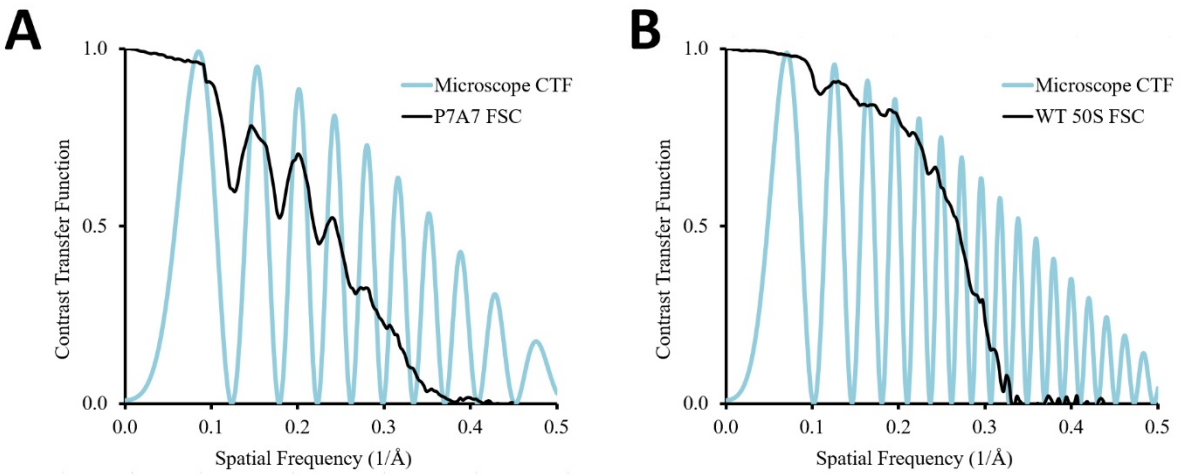


Figure 3. Simulated CTF curves for the experimental imaging conditions with a defocus value of 330 nm (**A**) and 480 nm (**B**), chosen to coincide with the peaks of the distribution of defocus values observed for the P7A7 large subunit structure and the WT 50S subunit structure, respectively. The curves are shown alongside the corresponding FSC curves for P7A7 and WT 50S. The P7A7 curve displayed oscillating features not observed for the WT 50S (Chapter 4).

While the case for focal pairs remained unconvincing after several attempts, the 70S control dataset was notable for the small number of particles needed to achieve 3 Å resolution, and for the fact that the quality of the particles seemed to be significantly higher than those collected as focal pairs from the same grid. To further examine this, the number of particles contributing to the final reconstruction were counted for each micrograph and divided by the number of particles originally picked (**Figure 4**). In principle, poor-quality particles are removed during early stages in processing, and the proportion of particles that go on to the final 3D refinement should generally reflect the quality of that micrograph. Notable stratification is observed between four stretches of the conventional micrographs, corresponding to four different squares on the grid, indicating that square-to-square variation is significant. In this case, the chosen squares looked very similar at the overview level of magnification, and only upon closer inspection of the holes was it clear that for some squares, the vast majority of holes were dry or otherwise compromised in quality. While ice thickness was not measured directly, the focal pair data was collected on squares that appeared to

have slightly thicker ice on average, and did not suffer from such variability in proportion of particles used.

The possibility of thicker ice for the focal pairs overall is generally supported by comparison of estimated maximum resolution of Thon rings in the images (**Figure 5**), where it is expected that a thicker ice layer will attenuate signal more rapidly. However, many micrographs have similar values to those in the conventional dataset that led to high-quality particles, suggesting that the focal pair method itself is indeed less efficient at capturing high-resolution information. It is possible that the low dose combined with the close-to-focus condition has affected alignment of particles, but this is unlikely as the established practice of using subsets of early frames to visualize the effects of radiation over time has been used to high resolution[18,19]. Another possibility is that fewer CTF crossovers have led to less accurate CTF estimations, but further investigation would be needed.

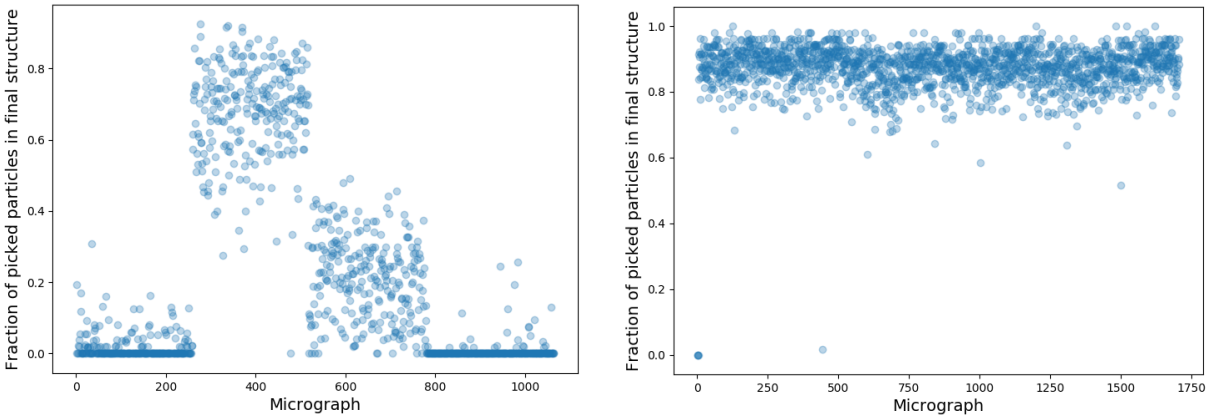


Figure 4. Particle count analysis for the conventional 70S control dataset (left) and focal pairs dataset (right). The number of particles in the final reconstruction divided by the number of particles originally autopicked from each micrograph is displayed as a point.

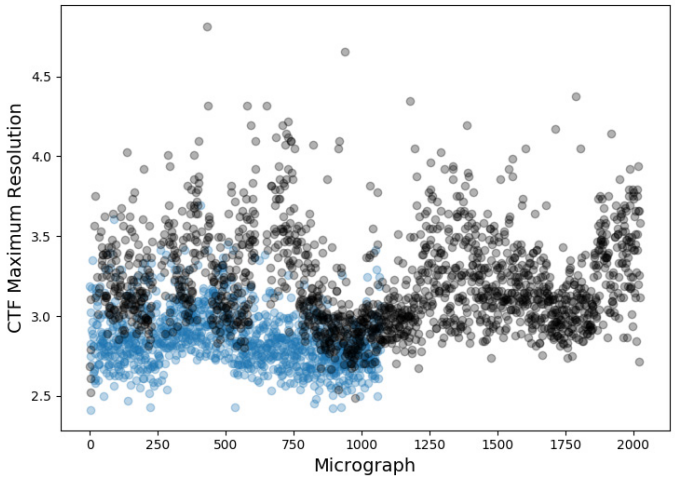


Figure 5. Estimated maximum resolution of the CTF (visible Thon rings) in the control (blue) and focal pair (black) datasets. This was determined for a total dose of 10 $e^-/Å^2$ in both cases. The overall lower resolution observed for focal pairs is consistent with slightly thicker ice.

ACKNOWLEDGMENTS
I thank Dan Toso and Paul Tobias for assistance with cryo-EM data collection and Bob Glaeser for guidance on imaging conditions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1106400.

REFERENCES
1. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
2. Cheng, Y. Single-particle Cryo-EM at crystallographic resolution. *Cell* (2015). doi:10.1016/j.cell.2015.03.049
3. Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1418377111
4. Danev, R. & Baumeister, W. Cryo-EM single particle analysis with the volta phase plate. *Elife* (2016). doi:10.7554/eLife.13046
5. Herzik, M. A., Wu, M. & Lander, G. C. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* (2019). doi:10.1038/s41467-019-08991-8
6. Schwartz, O. *et al.* Laser phase plate for transmission electron microscopy. *Nat. Methods* (2019). doi:10.1038/s41592-019-0552-2
7. Schiske, P. Image reconstruction by means of focus series 1 . *J. Microsc.* (2002). doi:10.1046/j.1365-2818.2002.01042.x
8. Ludtke, S. J. & Chiu, W. Focal pair merging for contrast enhancement of single particles. *J. Struct. Biol.* (2003). doi:10.1016/j.jsb.2003.10.012
9. Kudryashev, M., Stahlberg, H. & Castaño-Díez, D. Assessing the benefits of focal pair cryo-electron tomography. *J. Struct. Biol.* (2012). doi:10.1016/j.jsb.2011.10.004
10. Rickgauer, J. P., Grigorieff, N. & Denk, W. Single-protein detection in crowded molecular environments in cryo-EM images. *Elife* (2017). doi:10.7554/eLife.25648
11. Shalev-Benami, M. *et al.* 2.8-Å Cryo-EM Structure of the Large Ribosomal Subunit from the Eukaryotic Parasite Leishmania. *Cell Rep.* (2016). doi:10.1016/j.celrep.2016.06.014
12. Shalev-Benami, M. *et al.* Atomic resolution snapshot of Leishmania ribosome inhibition by the aminoglycoside paromomycin. *Nat. Commun.* (2017). doi:10.1038/s41467-017-01664-4
13. Mastronarde, D. N. SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. in *Microscopy and Microanalysis* (2003). doi:10.1017/s1431927603445911
14. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* (2019). doi:10.1107/S205225251801463X
15. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* (2018). doi:10.7554/eLife.42166
16. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331 (2017).
17. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from

electron micrographs. *J. Struct. Biol.* (2015). doi:10.1016/j.jsb.2015.08.008
18. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
19. Watson, Z. L. *et al.* Structure of the bacterial ribosome at 2 Å resolution. *Elife* (2020). doi:10.7554/ELIFE.60482

Chapter 6: Structure of the bacterial ribosome at 2 Å resolution

Reproduced with permission from: Watson ZL, Ward FR, Méheust R, Ad O, Schepartz A, Banfield JF, Cate JH. eLife. 2020 Sep 14;9:e60482. doi: 10.7554/eLife.60482. PMID: 32924932; PMCID: PMC7550191.

ABSTRACT

Using cryo-electron microscopy (cryo-EM), we determined the structure of the *Escherichia coli* 70S ribosome with a global resolution of 2.0 Å. The maps reveal unambiguous positioning of protein and RNA residues, their detailed chemical interactions, and chemical modifications. Notable features include the first examples of isopeptide and thioamide backbone substitutions in ribosomal proteins, the former likely conserved in all domains of life. The maps also reveal extensive solvation of the small (30S) ribosomal subunit, and interactions with A-site and P-site tRNAs, mRNA, and the antibiotic paromomycin. The maps and models of the bacterial ribosome presented here now allow a deeper phylogenetic analysis of ribosomal components including structural conservation to the level of solvation. The high quality of the maps should enable future structural analysis of the chemical basis for translation and aid the development of robust tools for cryo-EM structure modeling and refinement.

INTRODUCTION

The ribosome performs the crucial task of translating the genetic code into proteins and varies in size from 2.3 MDa to over 4 MDa across the three domains of life[1]. Polypeptide synthesis occurs in the peptidyl transferase center (PTC), where the ribosome acts primarily as an "entropic trap" for peptide bond formation[2]. To carry out the highly coordinated process of translation, the ribosome orchestrates the binding and readout of messenger RNA (mRNA) and transfer RNAs (tRNAs), coupled with a multitude of interactions between the small and large ribosomal subunits and a host of translation factors. These molecular interactions are accompanied by a wide range of conformational dynamics that contribute to translation accuracy and speed[3; 4–6]. Because of the ribosome's essential role in supporting life, it is naturally the target of a plurality of antibiotics with diverse mechanisms of action[7]. The ribosome also plays a unique role in our ability to study the vast array of RNA secondary and tertiary structural motifs found in nature, as well as RNA-protein interactions. Although X-ray crystallography has been central in revealing the molecular basis of many steps in translation, the resolution of available X-ray crystal structures of the ribosome in key functional states remains too low to provide accurate models of non-covalent bonding, i.e. hydrogen bonding, van der Waals contacts, and ionic interactions. Furthermore, development of small molecule drugs such as antibiotics is hampered at the typical resolution of available X-ray crystal structures of the ribosome (~3 Å)[7,8]. Thus, understanding the molecular interactions in the ribosome in chemical detail would provide a foundation for biochemical and biophysical approaches that probe ribosome function and aid antibiotic discovery.

The ribosome has been an ideal target for cryo-EM since the early days of single-particle reconstruction methods, as its large size, many functional states, and multiple binding partners lead to conformational heterogeneity that make it challenging for X-ray crystallography[9]. Previous

high-resolution structures of the bacterial ribosome include X-ray crystal structures of the *E. coli* ribosome at 2.4 Å (2.1 Å by CC ½)[10] and *T. thermophilus* ribosome at 2.3 Å[11], and cryo-EM reconstructions at 2.1-2.3 Å[12–14]. While resolution is well-defined for crystallography, which measures information in Fourier space[15], it is more difficult to assign a metric to global resolution of cryo-EM structures. Fourier Shell Correlation (FSC) thresholds are widely used for this purpose[16,17]. However, the "gold-standard" FSC (GS-FSC) most commonly reported, wherein random halves of the particles are refined independently and correlation of the two half-maps is calculated as a function of spatial frequency, is more precisely a measure of self-consistency of the data[18–20]. Reporting of resolution is further complicated by variations in map refinement protocols and post-processing by the user, as well as a lack of strict standards for deposition of these data[20]. A separate measure, the map-to-model FSC, compares the cryo-EM experimental map to the structural model derived from it[20–22]. In this work we use both metrics to evaluate our results, but bring attention to the use of the map-to-model FSC criterion as it is less commonly used but more directly reports on the quality and utility of the atomic model.

Here we have determined the structure of the *E. coli* 70S ribosome to a global resolution of 2.0 Å, with higher resolution up to 1.8 Å in the best-resolved core regions of the 50S subunit. We highlight well-resolved features of the map with particular relevance to ribosomal function, including contacts to mRNA and tRNA substrates, a detailed description of the aminoglycoside antibiotic paromomycin bound in the mRNA decoding center, and interactions between the ribosomal subunits. We also describe solvation and ion positions, as well as features of post-transcriptional modifications and post-translational modifications seen here for the first time. Discovery of these chemical modifications, as well as a new RNA-interacting motif found in protein bS21, provide the basis for addressing phylogenetic conservation of ribosomal protein structure and clues towards the role of a protein of unknown function. These results open new avenues for studies of the chemistry of translation and should aid future development of tools for refining structural models into cryo-EM maps.

RESULTS

Overall map quality

We determined the structure of the *Escherichia coli* 70S ribosome in the classical (non-rotated) state with mRNA and tRNAs bound in the Aminoacyl-tRNA and Peptidyl-tRNA sites (A site and P site, respectively). Partial density for Exit-site (E-site) tRNA is also visible in the maps, particularly for the 3'-terminal C75 and A76 nucleotides. Our final maps were generated from two 70S ribosome complexes that were formed separately using P-site $tRNA^{fMet}$ that differed only by being charged with two different non-amino acid monomers (see Methods). In both complexes, we used the same mRNA and A-site Val-$tRNA^{Val}$. Both complexes yielded structures in the same functional state, with similar occupancy of the tRNAs. As neither A-site nor P-site tRNA 3'-CCA-ends were resolved in the individual cryo-EM maps, we merged the two datasets for the final reconstructions (**Figure 1–figure supplement 1**). After Ewald sphere correction in RELION[23], global resolution of the entire complex reached 1.98 Å resolution by GS-FSC, with local resolution reaching 1.8 Å (**Figure 1A**, **Figure 1–figure supplement 2, Table 1**). The global resolution reached 2.04 Å using the map-to-model FSC criterion (**Figure 1–figure supplement 2**). We also used focused refinement of the large (50S) and small (30S) ribosomal subunits, and further focused refinements of smaller regions that are known to be conformationally flexible, to enhance their resolution (**Figure 1–figure supplement 3**)[24]. In particular, focused refinement of the 30S subunit

improved its map quality substantially, along with its immediate contacts to the 50S subunit and the mRNA and tRNA anticodon stem-loops. Additional focused refinement of the Central Protuberance (CP) in the 50S subunit, the 30S head domain, and 30S platform aided in model building and refinement. In the following descriptions, maps of specific ribosomal subunits or domains refer to the focused-refined maps. Details of the resolutions obtained are given in **Figure 1–figure supplements 2-3** and **Tables 2-3**.

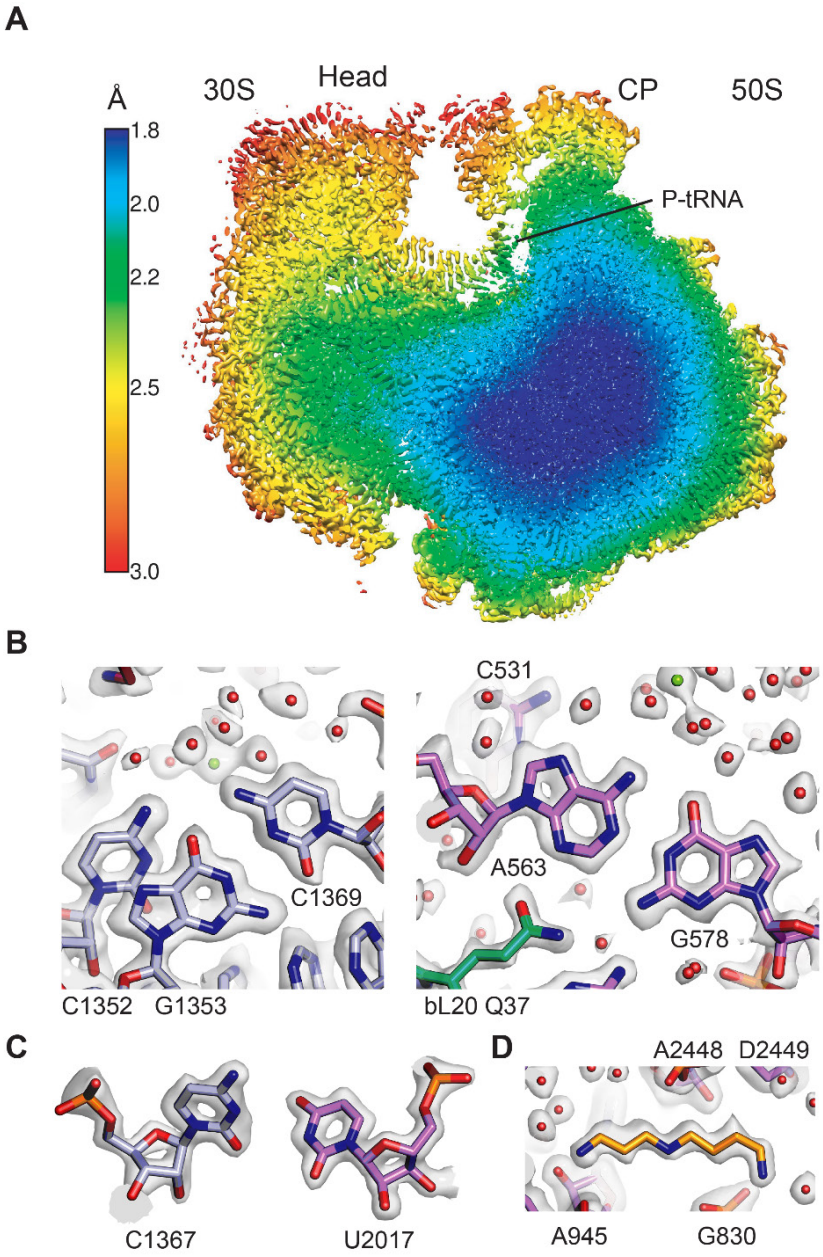


Figure 1. Overall structure of the 70S ribosome and cryo-EM map quality. (**A**) Cutaway view through the local resolution map of the 70S ribosome reconstruction. (**B**) Base pair density in the cores of the 30S (left) and 50S (right) ribosomal subunits. Examples demonstrate overall high resolution of base pairs and nearby solvation and Mg^{2+} sites. B-factors of -15 $Å^2$ and -10 $Å^2$ were applied to the RELION post-processed 50S subunit- and 30S subunit head-focused maps, respectively. (**C**) Nucleotide riboses in the core of 30S subunit (left) and 50S subunit (right). A B-factor of -10 $Å^2$ was applied to 30S density after post-processing. (**D**) Cryo-EM density of the 50S subunit showing the polyamine spermidine.

Table 1. Data collection and processing.

Magnification	109,160×
Voltage (kV)	300
Spherical aberration (mm)	2.7
Electron exposure (e^{-}/Å^2)	39.89
Defocus range (μm)	-0.6/ -1.5
Pixel size (Å)	0.7118
Symmetry imposed	C1
Initial particle images (no.)	874,943
Final particle images (no.)	307,495
Map resolution (Å)	2.02
Map resolution with Ewald correction (Å)	1.98
FSC threshold (gold-standard)	0.143

The high resolution indicated by the FSC curves is supported by several visual features observable in the maps, including holes in many aromatic rings and riboses, as well as ring puckers, directionality of non-bridging phosphate oxygens in ribosomal RNA, and numerous well-resolved ions, water molecules, and small molecules (**Figure 1B-D**, **Figure 1–figure supplement 4**). The maps also reveal known post-transcriptional and post-translational modifications of ribosomal RNA and proteins in detail (**Figure 1–figure supplement 5**, **Figure 1–figure supplement 6**), many of which are as previously described[10,11,13,25]. As seen in other ribosome structures, elements of the ribosome at the periphery are less ordered, including the uL1 arm, the GTPase activating center, bL12 proteins and central portion of the A-site finger in the 50S subunit (23S rRNA helix H38), as well as the periphery of the 30S subunit head domain and spur (16S rRNA helix h6). The resolution of the maps for the elbow and acceptor ends of the P-site and A-site tRNAs is also relatively low, likely as a result of poor accommodation of the unnatural substrates used.

High-resolution structural features of the 30S ribosomal subunit

The 30S ribosomal subunit is highly dynamic in carrying out its role in the translation cycle[3]. Previously published structures demonstrate that even in defined conformational states of the ribosome, the 30S subunit exhibits more flexibility than the core of the 50S subunit. Furthermore, the mass of the 50S subunit dominates alignments in cryo-EM reconstructions of the 70S ribosome. Solvation of the 30S subunit has not been extensively modeled, again owing to the fact that it is generally more flexible and less well-resolved than the 50S subunit, even in available high-resolution structures[10,11]. Using focused-refined maps of the 30S subunit, we achieved the resolution necessary for in-depth chemical analysis of key contacts to mRNA and tRNAs, the 50S subunit, and to generate more complete models of 30S subunit components, Mg^{2+} ion positions and solvation (**Figure 2 - figure supplement 1**).

mRNA and tRNA interactions with the 30S ribosomal subunit

The 30S ribosomal subunit controls interactions of tRNAs with mRNA and helps maintain the mRNA in the proper reading frame. The present structure reveals the interactions of the 30S subunit with mRNA and tRNA in the A and P sites including solvation. The tight binding of tRNA to the P site[26] is reflected in extensive direct contacts between the tRNA anticodon stem loop (ASL) and both the 30S subunit head and platform domains (**Figure 2A-D**). The C-terminal

residues Lys129 and Arg130 of protein uS9, which are important for translational fidelity[27], form ionic and hydrogen bonding interactions with the nucleotides in the U-turn motif of the P-site ASL, and Arg130 stacks with the base of U33 (**Figure 2D**). However, the C-terminal tails of ribosomal proteins uS13 and uS19, which come from the 30S subunit head domain and are lysine-rich, are not visible in the map, suggesting they do not make specific contacts with the tRNA when the ribosome is in the unrotated state. Direct interactions between A-site tRNA and the 30S subunit are highly localized to the top of helices h44 and h18 in 16S rRNA and to helix H69 in 23S rRNA (i.e. 16S rRNA nucleotides G530, A1492, and A1493, and 23S rRNA nucleotide A1913). By contrast, contacts between the A-site tRNA and 30S subunit head domain are entirely solvent mediated (**Figure 2E**) apart from nucleotide C1054, possibly reflecting the weaker binding of tRNA to the A site[26] and the need for A-site tRNA conformational dynamics during mRNA decoding[28].

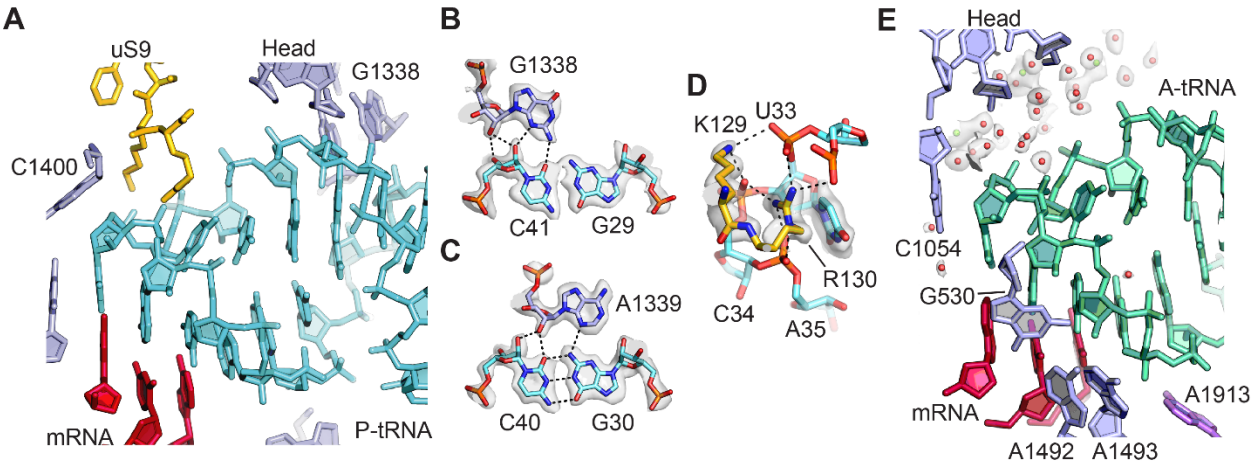


Figure 2. tRNA binding to the 30S ribosomal subunit. (**A**) Overall view of P-site tRNA anticodon stem-loop (ASL, cyan), mRNA (red), 16S rRNA nucleotides (light purple), and uS9 residues (gold). (**B**) Interactions between 30S subunit head nucleotide G1338 with P-tRNA ASL. (**C**) Interactions between 30S subunit head nucleotide A1339 with P-site ASL. (**D**) Interactions between P-tRNA ASL and protein uS9. Arg130 is observed stacking with nucleotide U33 of the ASL and forming hydrogen bonds with backbone phosphate groups. (**E**) Solvation of A-site tRNA near the 30S subunit head domain. A-site tRNA ASL in green, 16S rRNA in light purple, mRNA in purple-red. Water oxygen atoms in red spheres and Mg^{2+} in green spheres. Maps shown in panels **B**-**E** are from 30S subunit head-focused refinement.

Contacts of protein bS21 with the 30S subunit head domain

Protein bS21, which resides near the path of mRNA on the 30S subunit platform[29–31], is essential in *E. coli*[32,33]. Although partial structural models of bS21 have been determined[10,25], disorder in this region has precluded modeling of the C-terminus. In the present maps, low-pass filtering provides clear evidence for the conformation of the entire protein chain, including 13 amino acids at the C-terminus that extend to the base of the 30S subunit head domain (**Figure 3A**). Most of the C-terminal residues are found in an alpha-helical conformation near the Shine-Dalgarno helix formed between the 3' end of 16S rRNA and the mRNA ribosome binding site[34], while the C-terminal arginine-leucine-tyrosine (RLY) motif makes close contacts with 16S rRNA helix h37 and nucleotide A1167 (**Figure 3B**). The arginine and leucine residues pack in the minor groove of helix h37, and the terminal tyrosine stacks on A1167. Multiple sequence alignment of bS21 sequences from distinct bacterial phyla revealed that the RLY C-terminal motif is conserved in bS21 sequences of the Gammaproteobacteria phylum while in Betaproteobacteria, the sibling

group of Gammaproteobacteria, bS21 sequences possess a lysine-leucine-tyrosine (KLY) C-terminal motif instead (**Figure 3C**). Interestingly, such C-terminal extensions (RLY or KLY) are absent in other bacterial phyla (**Figure 3–figure supplement 1**). Recently, putative homologues of bS21 were identified in huge bacteriophages, which were shown to harbor genes encoding components of the translational machinery[35]. Inspection of sequence alignments reveals that some phage S21 homologues also contain KLY-like motifs (**Figure 3C**). The presence of phage S21 homologues containing KLY-like motifs is consistent with the host range of these phages (**Figure 3C, Supplementary Table 1**). Other phages with predicted hosts lacking the C-terminal motif in bS21 encode S21 homologues that also lack the motif (**Supplementary Table 1**). Although the C-terminal region of bS21 resides near the Shine-Dalgarno helix, examination of ribosome binding site consensus sequences in these bacterial clades and the predicted ribosome binding sites in the associated phages do not reveal obvious similarities (**Supplementary File 1**).

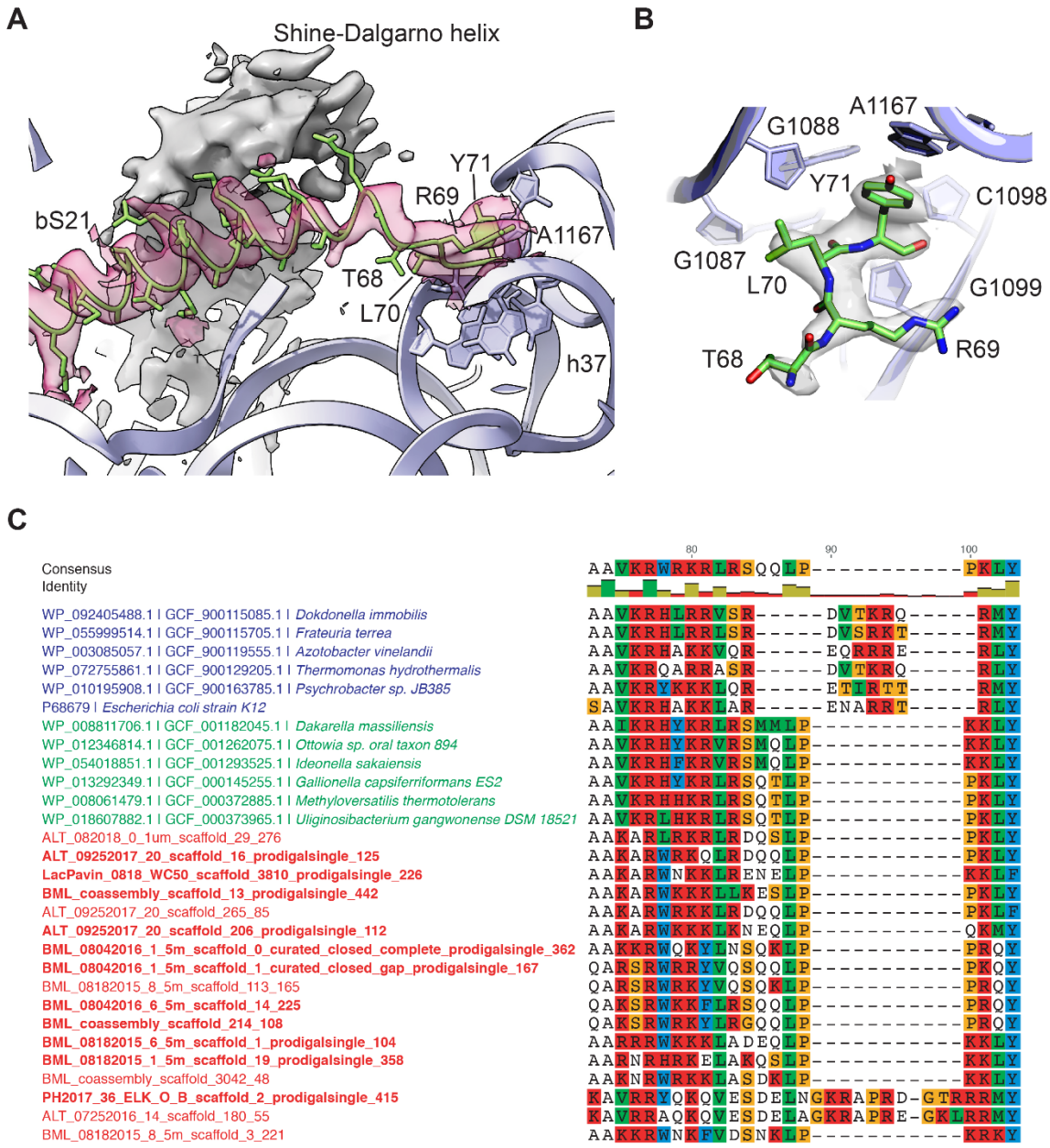


Figure 3. Protein bS21 interactions with 30S ribosomal subunit head domain. (**A**) bS21 C-terminal structure in the 30S subunit head-focused map, with Shine-Dalgarno helical density shown in grey and density for bS21 in rose. Low-pass filtering to 3.5 Å resolution was applied to clarify helical density. 16S rRNA shown in light purple ribbon and bS21 shown in light green. 16S rRNA bases that interact with RLY motif are shown in stick representation. (**B**) Closeup of RLY motif of bS21 and contacts with 16S rRNA h37 and A1167. (**C**) Protein bS21 sequence alignment near the C-terminus, along with associated phage S21 sequences. Gammaproteobacteria (blue), Betaproteobacteria (green), and phage (red) are shown.

Post-translational and post-transcriptional modifications in the 30S subunit

Ribosomal protein uS11 is a central component of the 30S subunit platform domain, and assembles cooperatively with ribosomal proteins uS6 and uS18[36], preceding the binding of bS21 late in 30S subunit maturation[29,31]. Protein uS11 makes intimate contact with 16S rRNA residues in a 3-helix junction that forms part of the 30S subunit E site and stabilizes the 16S rRNA that forms the platform component of the P site[36]. Remarkably, inspection of the 30S subunit map

uncovered a previously unmodeled isoaspartyl residue in protein uS11, at the encoded residue N119 (**Figure 4**) (see Methods), marking the first identified protein backbone modification in the ribosome. While it has been known that this modification can exist in uS11[37], its functional significance has remained unclear, and prior structures did not have the resolution to pinpoint its exact location. Conversion of asparagine to isoaspartate inserts an additional methylene group into the backbone (the Cβ position) and generates a methylcarboxylate side chain[38]. In the present map, the shapes of the backbone density and proximal residues in the chain clearly reveal the presence of the additional methylene, allowing it flexibility to pack closely with the contacting rRNA nucleotides (**Figure 4A**, **4B**). A 30S-focused reconstruction using only early movie frames, in which damage to carboxylates would not be as severe[39], shows improved density for the IAS sidechain, albeit at slightly lower resolution overall (**Figure 4–figure supplement 1**).

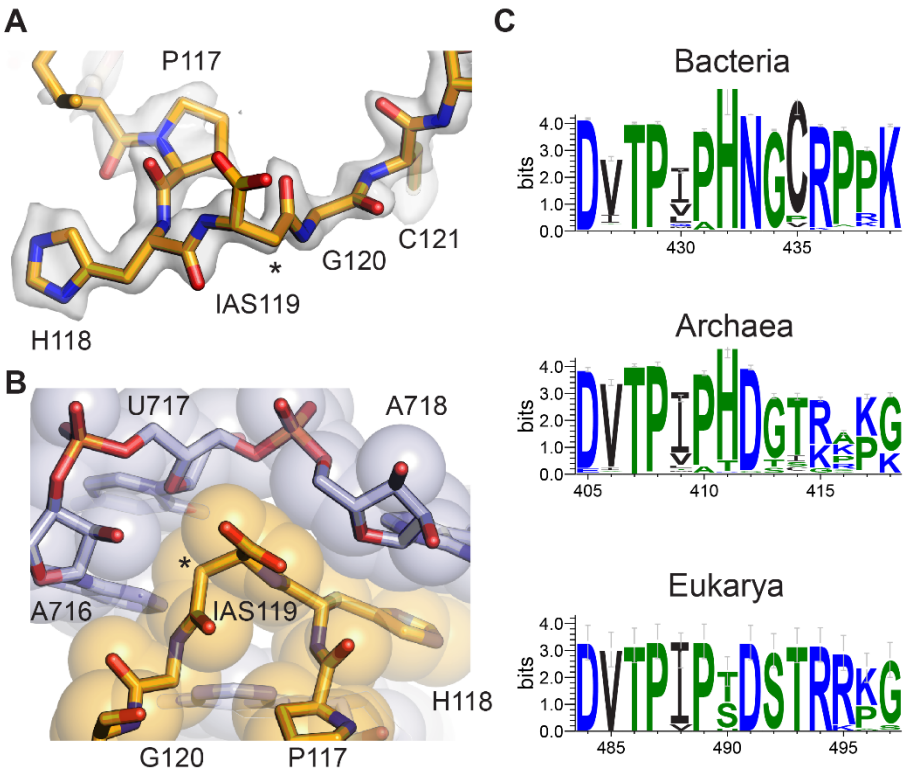


Figure 4. Isoaspartyl residue in protein uS11. (**A**) Model of isoAsp at residue position 119 in uS11, with nearby residues and cryo-EM density from the 30S subunit platform-focused refinement. Weak density for the carboxylate is consistent with effects of damage from the electron beam. The asterisk indicates the position of the additional backbone methylene group. (**B**) Shape complementarity between uS11 and 16S rRNA nucleotides surrounding IsoAsp119. 16S rRNA is shown in light purple and uS11 in orange, with atomistic coloring for the stick model. (**C**) Sequence logos of conserved amino acids spanning the putative isoAsp residue in all three domains of life.

Investigation of residues flanking the isoaspartate in uS11 reveal near-universal conservation in bacteria, chloroplasts and mitochondria (**Figure 4C**, **Figure 4–figure supplement 2**), suggesting that the isoaspartate may contribute to 30S subunit assembly or stability. Consistent with this idea, the isoaspartate allows high shape complementarity including van der Waals contacts and hydrogen bonds between this region of uS11 and the 16S rRNA it contacts, which involves three consecutive purine-purine base pairs in bacteria (**Supplementary Table 2**), and a change in rRNA helical direction that is capped by stacking of histidine 118 in uS11 on a conserved

purine (A718 in *E. coli*) (**Figure 4B**, **Supplementary Table 2**). Strikingly, the sequence motif in bacterial uS11 is also conserved in a domain-specific manner in archaea and eukaryotes (**Figure 4C**, **Figure 4–figure supplement 1**), as are the rRNA residues in close proximity to the predicted isoAsp (**Supplementary Table 2**). Remodeling the isoAsp motifs in maps from recently-published cryo-EM reconstructions of an archaeal 30S ribosomal subunit complex at 2.8 Å resolution[40] and a yeast 80S ribosome complex at 2.6 Å resolution[41] shows that isoaspartate also seems to be present in these organisms based on residue-level correlation between map and model (**Figure 4–figure supplement 2**). Taken together, the phylogenetic data and structural data indicate that the isoaspartate in uS11 is nearly universally conserved, highlighting its likely important role in ribosome assembly and function.

The *E. coli* 30S ribosomal subunit has eleven post-transcriptionally modified nucleotides in 16S rRNA, all of which can be seen in the present maps or inferred from hydrogen bonding patterns in the cases of many pseudouridines. Interestingly, two methylated nucleotides–m^7G527 and $m^6{}_2$A1519–appear not to be fully modified, based on the density at ~2.1 Å resolution (**Figure 1–figure supplement 5**). In the map, m^7G527 appears partially methylated, and $m^6{}_2$A1519 lacks one of the two methyl groups. Loss of methylation at m^7G527, which is located near the mRNA decoding site, has been shown to confer low-level streptomycin resistance[42], and possibly neomycin resistance in some cases[43]. The position of the methyl group is located in a pocket formed with ribosomal protein uS12, adjacent to the post-translationally modified Asp89, β-methylthio-Asp[44,45]. The β-methylthio-Asp also has weak density for the β-methylthio group suggesting it is also hypomodified in the present structure (**Figure 1–figure supplement 6**). Notably, loss of m^7G527 methylation is synergistic with mutations in uS12 that lead to high-level streptomycin resistance[42,46]. Loss of m^7G527 methylation would remove a positive charge and open a cavity adjacent to uS12, which may contribute to resistance by shifting the equilibrium of 30S subunit conformational states to an "open" form that is thought to be hyperaccurate with respect to mRNA decoding[6,47,48].

Within the 30S subunit platform near the P site, the two dimethylated adenosines–$m^6{}_2$A1518 and $m^6{}_2$A1519–have also been connected to antibiotic resistance. Although impacting 30S subunit assembly[49] and ribosome function[50], loss of methylation of these nucleotides also leads to kasugamycin resistance[51]. By contrast, bacteria lacking KsgA, the methyltransferase responsible for dimethylation of both nucleotides, become highly susceptible to other antibiotics including aminoglycosides and macrolides[52–54]. In the present structure, $m^6{}_2$A1519 is singly-methylated whereas $m^6{}_2$A1518 is fully methylated (**Figure 1–figure supplement 5**). KsgA fully methylates both nucleotides in *in vitro* biochemical conditions[55], but the methylation status of fully-assembled 30S subunits *in vivo* has not been determined. The loss of a single methylation of $m^6{}_2$A1519, observed here for the first time, could be a mechanism for conferring low-level antibiotic resistance to some antibiotics without appreciably affecting 30S subunit assembly or leading to sensitivity to other classes of antibiotics, a hypothesis that could be tested in the future.

Paromomycin binding in the mRNA decoding site

Aminoglycoside antibiotics (AGAs) are a widely-used class of drugs targeting the mRNA decoding site (A site) of the ribosome, making them an important focus for continued development against antibiotic resistance[56]. Paromomycin, a 4,6-disubstituted 2-deoxystreptamine AGA (**Figure 5A**), is one of the best studied structurally. Structures include paromomycin bound to an oligoribonucleotide analog of the A site at 2.5 Å resolution[57], to the small subunit of the *T. thermophilus* ribosome at 2.5 Å resolution[58], and to the full 70S *T. thermophilus* ribosome at 2.8

Å resolution[59]. While the overall conformation of rings I-III of paromomycin is modeled in largely the same way (**Figure 5B**), the resolution of previous structures did not allow for unambiguous interpretation of ring IV, and only the oligoribonucleotide structural model of the decoding site includes some water molecules in the drug's vicinity[57]. Although ring IV remains the least ordered of the four rings in the present structure, the cryo-EM map of the focus-refined 30S subunit allows high-resolution modeling of the entire molecule and the surrounding solvation for the first time (**Figure 5C**, **5D**). The conformation of paromomycin in the oligonucleotide structure[57] agrees most closely with the current structure, with ring IV adopting a chair conformation with the same axial and equatorial positioning of exocyclic functional groups. However, the N6''' group in the present structure points in the opposite direction and forms multiple contacts with the backbone phosphate groups of G1489 and U1490 (**Figure 5D**). Tilting of ring IV in the present model also positions N2''' and O3''' to make contacts with the G1405 and A1406 phosphate groups, respectively. The paromomycin models in the previous structures of the 30S subunit and 70S ribosome differ further by modeling ring IV in the alternative chair conformation, which also breaks the contacts observed here. We do not see paromomycin bound in H69 of the 50S ribosomal subunit, a second known AGA binding site, consistent with prior work indicating that binding of aminoglycosides to H69 may be favored in intermediate states of ribosomal subunit rotation[60,61].

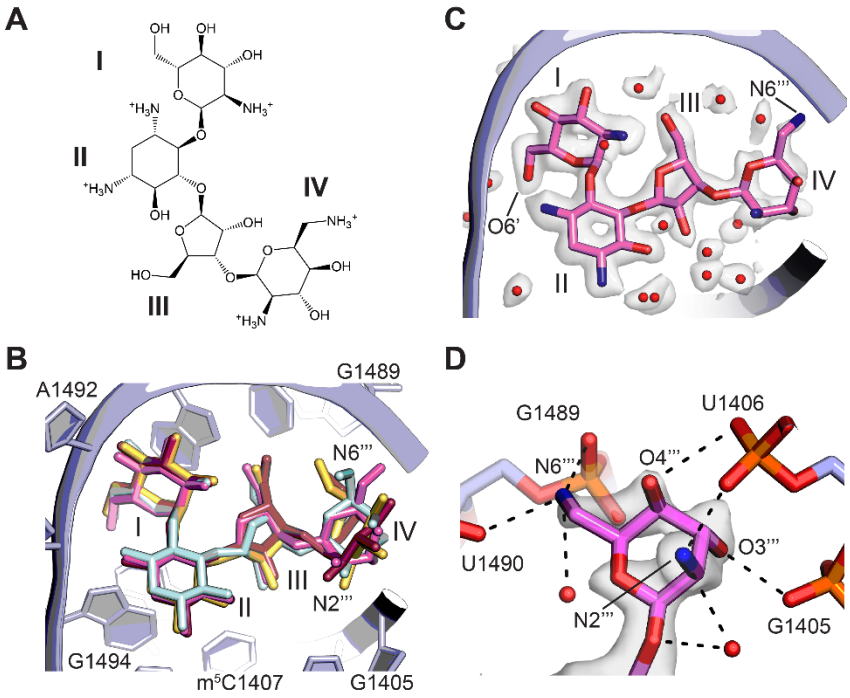


Figure 5. Binding of paromomycin to the mRNA decoding site in the 30S subunit. (**A**) Chemical structure of paromomycin (PAR) with ring numbering. (**B**) Comparison of paromomycin conformations in different structures. Paromomycin from three prior structural models [57–59], shown in yellow, dark pink, and light blue, respectively, superimposed with the present model of paromomycin, shown in pink. The binding pocket formed by 16S rRNA is shown in light purple. (**C**) Overall positioning of PAR within the binding site including solvation. (**D**) Paromomycin ring IV contacts to the phosphate backbone in 16S rRNA helix h44. Dashed lines denote contacts within hydrogen-bonding distance. Map was blurred with a B-factor of 10 Å^2.

Ribosomal subunit interface

The ribosome undergoes large conformational changes within and between the ribosomal subunits during translation, necessitating a complex set of interactions that maintain ribosome

function. Contacts at the periphery of the subunit interface have been less resolved in many structures, likely due to motions within the ribosome populations. Additionally, some key regions involved in these contacts are too conformationally flexible to resolve in structures of the isolated subunits. In the present structure of the unrotated state of the ribosome, with tRNAs positioned in the A site and P site, improvement of maps of the individual ribosomal subunits and smaller domains within the subunits help to define these contacts more clearly. Helix H69 of the 50S subunit, which is mostly disordered in the isolated subunit, becomes better defined once the intact ribosome is formed. The 23S rRNA stem-loop closed by H69 is intimately connected to the 30S subunit at the end of 16S rRNA helix h44 near the mRNA decoding site and tRNA binding sites in the ribosome. During mRNA decoding, the RNA loop closing H69 rearranges to form specific interactions with the A-site tRNA[59]. The stem of H69 also compresses as the 30S subunit rotates during mRNA and tRNA translocation, thereby maintaining contacts between the 30S and 50S subunits[62]. In the present reconstructions, helix H69 seems conformationally more aligned to the 30S subunit than the 50S subunit, as the cryo-EM density for H69 is much better defined in the map of the 30S subunit compared to the map of the 50S subunit.

The loop comprising 23S rRNA nucleotides G713-A718 closing helix H34 forms an additional bridge between the 50S and 30S subunits and is also known to be dynamic in its position[62]. In the present reconstructions this bridge is also better defined in the map of the 30S subunit compared to the 50S subunit, placing the more highly conserved arginine Arg88 in uS15 in direct contact with the RNA backbone of the H34 stem-loop rather than the less conserved Arg89 (**Figure 6A**). An additional conformationally dynamic contact between the 30S and 50S subunits involves the A-site finger (ASF, helix H38 in 23S rRNA), which is known to modulate mRNA and tRNA translocation[63]. In the present model, although the central helical region of the ASF is only visible in low-resolution maps, loop nucleotide C888 which stacks on uS13 residues Met81 and Arg82 in the 30S subunit head domain is clearly defined (**Figure 6B**). Maps of the 30S subunit head domain and Central Protuberance of the 50S subunit also reveal clearer density defining the unrotated-state contacts between uS13 in the small subunit, uL5 in the large subunit, and bL31 (bL31A in the present structure) which spans the two ribosomal subunits.

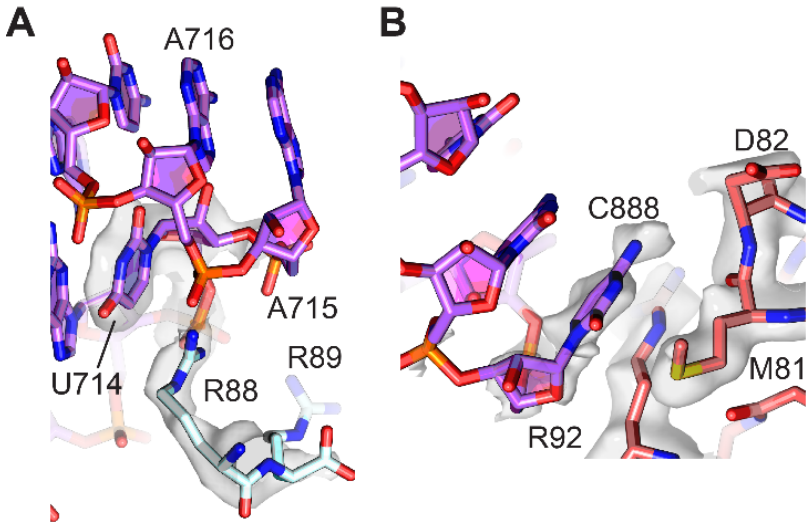


Figure 6. Peripheral contacts between the 30S and 50S subunits. (**A**) Interaction of the C-terminus of uS15 (light blue) with 23S rRNA nucleotides 713-715 (purple). The 30S subunit cryo-EM map is shown with a B-factor of 20 Å^2 applied. (**B**) Interaction between uS13 (salmon) and the A-Site Finger hairpin loop nucleotide C888 in the 50S subunit (purple). A B-factor of 10 Å^2 was applied to the head-focused map.

High-resolution structural features of the 50S ribosomal subunit

The core of the 50S subunit is the most rigid part of the ribosome, which has enabled it to be modeled to higher resolution than the 30S subunit, historically and in the present structure. In the present 70S ribosome and 50S subunit reconstructions, which have global map-to-model resolutions of 2.04 Å and 1.90 Å respectively, the resolution of the core of the 50S subunit reaches 1.8 Å (**Figure 1**, **Figure 1–figure supplement 3A**, **Table 2**), revealing unprecedented structural details of 23S rRNA, ribosomal proteins, ions and solvation (**Figure 1–figure supplement 4**, **Figure 1–figure supplement 5**, **Figure 7–figure supplement 2**). The resulting maps are also superior in the level of detail when compared to maps previously obtained by X-ray crystallography[10,11], which aided in improving models of high-resolution chemical features like backbone dihedrals in much of the rRNA, non-canonical base pairs and triples, arginine side chain rotamers, and glycines in conformationally constrained RNA-protein contacts. The density also enabled modeling of thousands of water molecules, dozens of magnesium ions, and polyamines (**Figure 1–figure supplement 4**, **Figure 7–figure supplement 2**, **Table 2**). The present model now even allows for comparison of ribosome phylogenetic conservation to the level of solvent positioning. For example, water molecules and ions with conserved positions in the Peptidyl Transferase Center can be seen in comparisons of different bacterial and archaeal ribosome structures, even when solvation was not included in the deposited models[11,14,64] (**Figure 7**). The Central Protuberance (CP) of the 50S subunit, which contacts the P-site tRNA and the head of the 30S subunit, is dynamic, but is well resolved with focused refinement, here reaching a resolution of 2.13 Å by GS-FSC and 2.26 Å in map-to-model FSC comparisons (**Figure 1–figure supplement 3**). The improved resolution of the CP aided in modeling ribosomal proteins uL5 and bL31A, as well as the CP contact to P-site tRNA.

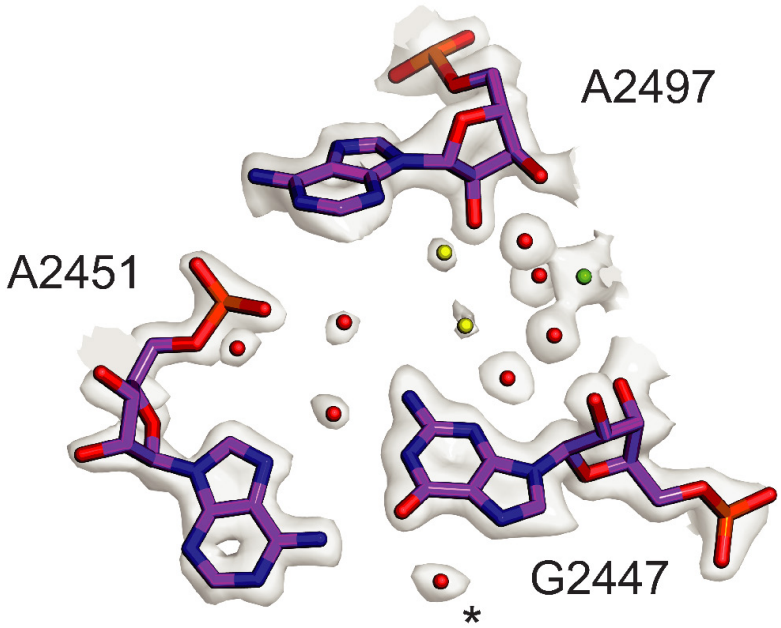


Figure 7. Conserved solvation in the PTC in the 50S ribosomal subunit. Comparison of solvation in the PTC near *E. coli* nucleotide G2447 to that in phylogenetically diverse 50S subunits. Solvent molecules conserved in bacterial ribosomes from *E. coli*, *S. aureus*, and *T. thermophilus*[11,14] and in the archaeal 50S subunit from *H. marismortui*[64] are colored red. Water molecules conserved in 3 of 4 structures are colored yellow. Mg^{2+} is shown in green. Asterisk (*) denotes density modeled as K^{+} in the *H. marismortui* 50S subunit structure.

Table 2. Model resolutions for subunits and domains.

Model	Without Ewald correction (Å)	Ewald sphere corrected (Å)	Map sharpening B-factor for Ewald (Å^2)
30S subunit	2.15	2.11	-25.7
30S subunit head domain	2.09	2.01	-19.7
30S subunit platform	2.12	2.08	-21.8
50S subunit	1.92	1.9	-25.1
50S subunit Central Protuberance	2.28	2.26	-21.5
70S ribosome	2.06	2.04	-29.5

*Map-vs-model FSC with threshold = 0.5

Table 3. Model refinement statistics.

Model component	70S ribosome	30S subunit	50S subunit
Model resolution, Ewald-corrected map (Å)	2.04	2.11	1.9
FSC threshold (map-vs.-model)	0.5	0.5	0.5
Map sharpening *B* factor (Å^2)	-29.5	-25.7	-25.1
Model composition			
non-hydrogen atoms	149356	54550	91592
Mg^{2+} ions	309	93	218
Zn^{2+} ions	2	0	2
polyamines	17	2	15
waters	7248	2413	4835
ligands (paromomycin)	1	1	0
B factors (Å^2)			
RNA	23.83	28.09	20.9
protein	24.42	28.91	20.95
waters	20.66	17.94	22.02
other	29.29	20.35	33.12
R.m.s. deviations from ideal values			
Bond (Å)	0.006	0.005	0.006
Angle (°)	0.952	0.838	0.997
Molprobity all-atom clash score	7.34	7.12	7.02
Ramachandran plot			
Favored (%)	96	95.66	96.26
Allowed (%)	3.87	4.17	3.65
Outliers (%)	0.13	0.17	0.1
RNA validation			
Angles outliers (%)	0.02	0.009	0.02
Sugar pucker outliers (%)	0.46	0.39	0.39
Average suiteness	0.579	0.586	0.583

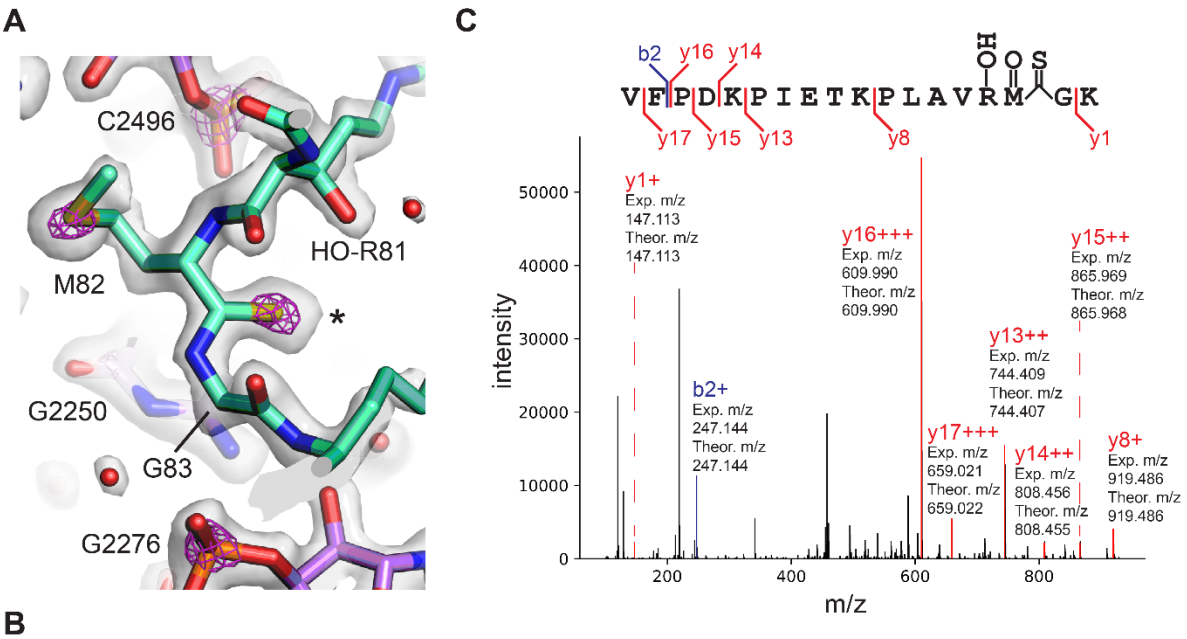
Breaks in RNA backbone EM density

Despite the overall high resolution of the 50S subunit core, there are a number of regions where the RNA backbone density shows relatively poor connectivity (**Figure 1–figure supplement 7**). In some cases, linkages between the ribose and phosphate groups become weakly visible or broken, with the phosphate group having an overall more rounded appearance than in cases where density is strong throughout the backbone. In some cases, breaks in the ribose ring are observed. Our initial impression was that this may be indicative of damage from the electron beam. However, reconstructions using the first 2 or 3 frames (corresponding to the first ~2-3 $e^-/Å^2$ in the exposure) show similar patterns of weak or broken density in these regions (**Figure 1–figure supplement 7**). This suggests that RNA conformational flexibility rather than radiation damage may be responsible for the broken density.

Backbone modification in ribosomal protein uL16

Details of post-transcriptional and post-translational modifications are also clear in the 50S subunit maps (**Figure 1 - figure supplement 5**). Surprisingly, the post-translationally modified β-hydroxyarginine at position 81 in uL16[65] is followed by unexplained density consistent with a thiopeptide bond between Met82 and Gly83 (**Figure 8A**). Adjusting the contour level of the map shows map density for the modified atom similar to that of the sulfur in the adjacent methionine and nearby phosphorus atoms in the RNA backbone, in contrast to neighboring peptide oxygen atoms. In the other cryo-EM maps of the 50S subunit[12,13], the density for the sulfur in the thioamide is not visible or barely visible (**Figure 8–figure supplement 1**). Notably, the mass for *E. coli* uL16 has been shown to be 15328.1 Da and drops to 15312.1 Da with loss of Arg81 hydroxylation[65]. However, this mass is still +30.9 more than the encoded sequence (15281.2 Da, Uniprot P0ADY7). In *E. coli* uL16 is also N-terminally methylated[66], leaving 16 mass units unaccounted for, consistent with the thiopeptide we observe in the cryo-EM map. We examined a high-resolution mass spectrometry bottom-up proteomics dataset[67] to find additional evidence supporting the interpretation of the cryo-EM map as a thiopeptide. Several uL16 peptides were found across multiple experiments that matched the expected mass shift closer to that of a thiopeptide's O to S conversion (+15.9772 Da) than an oxidation (+15.9949), a common modification with a similar mass shift (**Figure 8B**). Fragmentation spectra localized the mass shift near the Met82-Gly83 bond, further supporting the presence of a thiopeptide (**Figure 8C**). Taken together, the cryo-EM map of the 50S subunit and mass spectrometry data support the model of a thiopeptide between Met82 and Gly83 in *E. coli* uL16.

The enzymes that might be responsible for insertion of the thioamide in uL16 remain to be identified. *E. coli* encodes the prototypical YcaO enzyme, which can form thiopeptides, but for which no substrate is known[68]. A phylogenetic tree of YcaO family members shows a clear break separating YcaO proteins associated with secondary metabolism into a major branch (**Figure 8–figure supplement 2A**). A sub-grouping in the other major branch includes YcaO family members within Gammaproteobacteria (**Figure 8–figure supplement 2A**). Examination of genes in close proximity to YcaO across Gammaproteobacteria reveals three genes that form the *focA-pfl* operon involved in the anaerobic metabolism of *E. coli*[69] (**Figure 8–figure supplement 2B**). The combination of its unknown substrate in *E. coli*, ability to catalyze thioamidiation in other species, and syntenic conservation in Gammaproteobacteria identify YcaO as a primary candidate for uL16 thioamidation.



B

Peptide	Experimental Mass (Da)	Theoretical Mass (Da)	Mass Diff. (ppm)
VFPDKPITEKPLAVR[O]M[O,O->S]GK	2073.1106	2073.1113	-0.3377
VFPDKPITEKPLAVR[O]	1724.9823	1724.9823	0.0
VFPDKPITEKPLAVR	1708.9865	1708.9874	-0.5266
VFPDKPITEKPLAVRM[O,O->S]GK	2057.1174	2057.1165	0.4375
VFPDKPITEKPLAVR[O]M[O,O]GK (hypothetical)	2073.1106	2073.1290	**-8.8756**

Figure 8. Thioamide modification in protein uL16. (**A**) Structural model of thioamide between Met82 and Gly83 in uL16 (mint), with cryo-EM density contoured at two levels to highlight sulfur and phosphorus atoms. The lower contour level is shown as a grey surface and the higher contour level is shown as fuchsia mesh. 23S rRNA is shown in purple. Asterisk marks the position of the sulfur in the thiocarbonyl. (**B**) LC-MS/MS data supporting the presence of a thioamide bond between M82 and G83 of uL16[67]. Shown are selected uL16 peptides with designated modifications found in the spectral search and their associated experimental masses, theoretical masses, and mass differences. All peptides were found in multiple fractions and replicates of the experiment. The final row shows a hypothetical peptide identical to the first row, except carrying an oxidation modification instead of O to S replacement. (**C**) Annotated fragmentation spectra from the LC-MS/MS experiment showing a uL16 peptide with a thioamide bond. Peptide is assigned modifications of: oxidation on M, oxidation on R, and a thiopeptide between M and G. Fragmentation ions are annotated with experimental and theoretical m/z ratios.

DISCUSSION

High resolution cryo-EM maps are now on the cusp of matching or exceeding the quality of those generated by X-ray crystallography, opening the door to deeper understanding of the chemistry governing structure-function relationships and uncovering new biological phenomena. Questions about the ribosome, which is composed of the two most abundant classes of biological macromolecules and essential for life, reach across a diverse range of inquiry. Structural information about ribosomal components can have implications ranging from fundamental chemistry to mechanisms underlying translation and evolutionary trends across domains of life. For example, in our cryo-EM reconstructions, we observed a surprising level of detail about modifications to nucleobases and proteins that could not be seen in prior X-ray crystallographic

structures. The most unexpected of these is the presence of two previously unknown post-translational modifications in the backbones of ribosomal proteins, which would be otherwise difficult to confirm without highly targeted analytical chemical approaches. Precise information about binding of antibiotics, protein-RNA contacts, and solvation are additional examples of what can be interrogated at this resolution. Beyond purely structural insights, these findings generate new questions about protein synthesis, ribosome assembly, and antibiotic action and resistance mechanisms, providing a foundation for future experiments.

The remarkable finding of a thioamide modification in protein uL16, only the second such example in a protein[70], is a perfect example of the power of working at <2 Å resolution. The difference in bond length of a thiocarbonyl compared to a typical peptide carbonyl is ~0.4 Å, with otherwise unchanged geometry, and is too subtle to identify at lower resolution (**Figure 8–figure supplement 1**). Moreover, the sulfur density is not as pronounced in maps at lower resolution. Analysis of previously published mass spectrometry data with sufficient mass accuracy to differentiate O to S modifications from a more common +O oxidation event[67] (**Figure 8**) corroborates the finding. The possible role for the thiopeptide linkage in the *E. coli* ribosome, which is located near the PTC and involves contact between the thiocarbonyl sulfur atom and the hydroxyl group in the β-hydroxyarginine at position 81 in uL16, remains to be shown. The mechanism by which its formation is catalyzed also remains an open question. One candidate enzyme for this purpose is *E. coli* protein YcaO, an enzyme known to carry out thioamidation and other amide transformations[68]. Although this enzyme has been annotated as possibly participating with RimO in modification of uS12 Asp89[71], genetic evidence for a specific YcaO function is lacking. For example, *E. coli* lacking YcaO are cold-sensitive, and have phenotypes most similar in pattern to those observed with knockout of UspG, universal stress protein 12[72]. Furthermore, knockout of YcaO has phenotypes uncorrelated with those of knockout of YcfD, the β-hydroxylase for Arg81 in uL16 adjacent to the thioamide[72]. *YcaO*-like genes in Gammaproteobacteria genomes colocalize with the *focA-pfl* operon, a common set of genes involved in anaerobic and formate metabolism[69] (**Figure 8–figure supplement 1**). Since the ribosomes used here were obtained from aerobically grown cultures and the *focA-Pfl* operon is transcribed independently of the *YcaO* gene in *E. coli*[73] it is likely that the *YcaO* gene and the *focA-Pfl* operon encode proteins with unrelated functions. Interestingly, the clear phylogenetic separation between the *YcaO* gene in Gammaproteobacteria and the *YcaO* genes known to be involved in secondary metabolism in the phylogenetic tree suggests that, if YcaO is responsible for uL16 thioamidation, this modification may only be conserved in Gammaproteobacteria.

The maps of the 30S subunit, resolved to a slightly lower resolution of ~2.0-2.1 Å (**Figure 1–figure supplement 3**, **Table 2**), enabled the identification of the only known isopeptide bond in a ribosomal protein, an isoAsp at position 119 in uS11. While isoaspartyl residues have been hypothesized to mainly be a form of protein damage requiring repair, previous work identified the existence of isoAsp in uS11 at near stoichiometric levels, suggesting it might be functionally important[37]. Certain hotspots in protein sequences are known to be especially prone to isoaspartate formation[38], including Asn-Gly, as encoded in nearly all bacterial uS11 sequences (**Figure 4C**). However, the half-life of the rearrangement is on the timescale of days[74,75]. In archaea and eukaryotes, formation of the isoaspartate would require dehydration of the encoded aspartate, which occurs even more slowly than deamidation of asparagine[74]. Importantly, the residue following the aspartate is nearly always serine in eukaryotes, and is enriched for glycine, serine and threonine in archaea (**Figure 4C**, **Figure 4–figure supplement 1**), consistent with the higher rates of dehydration that occur when aspartate is followed by glycine and serine in peptide

models[74]. These results suggest that the isoAsp modification may be nearly universally conserved in all domains of life. Concordant with this hypothesis, isoAsp modeling provides a better fit to cryo-EM maps of uS11 in archaeal and eukaryotic ribosomes (**Figure 4–figure supplement 2**). Although it is possible that isoaspartate formation could be accelerated in specific structural contexts[38], it is not clear if the isoAsp modification in uS11 occurs spontaneously or requires an enzyme to catalyze the reaction. *O*-methyl-transferase enzymes have been identified that install a β-peptide in a lanthipeptide[76] or serve a quality control function to remove spontaneously formed isoaspartates[37]. Deamidases that catalyze isoAsp formation from asparagine are not well described in the literature, although examples have been identified in viral pathogens, possibly repurposing host glutamine amidotransferases[77]. Future work will be needed to identify the mechanisms by which the isoAsp in uS11 is generated in cells. Its biological significance, whether in 30S subunit assembly or other steps in translation, also remains to be defined.

The resolution achieved here also has great potential for better informing structure activity relationships in future antibiotic research, particularly because the ribosome is so commonly targeted. For example, we were able to identify hypomodified bases in 16S rRNA (m^7G527 and $m^6{}_2$A1519) and possible hypomodification of Asp89 (β-methylthio-Asp) in uS12 (**Figure 1–figure supplement 5**, **Figure 1–figure supplement 6**). These hypomodifications could confer resistance to kasugamycin and streptomycin antibiotics in some cases. Furthermore, we were also able to see more clearly the predominant position of paromomycin ring IV in the decoding site of the 30S subunit (**Figure 5**). The proposed primary role of ring IV has been to increase the positive charge of the drug to promote binding[78], in line with its ambiguous modeling in previous structures[57–59]. While ring IV's features in the current map are weaker relative to those of rings I-III, we were able to identify interactions of ring IV with surrounding 16S rRNA nucleotides and ordered solvent molecules that were not previously modeled. Importantly, the observed interactions between the N6''' amino group and the phosphate backbone of nucleotides G1489-U1490 in particular are likely responsible for known susceptibility of PAR to N6''' modification[79]. While the same loss of interactions is expected for neomycin, which differs from paromomycin only by the presence of a 6'-hydroxy rather than a 6'-amine in ring I, the penalty of modifying N6''' in neomycin is likely compensated for by the extra positive charge and stronger hydrogen bonding observed with neomycin ring I[79]. The level of detail into modes of aminoglycoside binding that can now be obtained using cryo-EM thus should aid the use of chemical biology to advance AGA development.

The cryo-EM maps of the 30S subunit also revealed new structural information about protein bS21 at lower resolution, particularly at its C terminus. The location of bS21 near the ribosome binding site suggests it may play a role in translation initiation. The conservation of the RLY (or KLY) motif and its contacts to the 30S subunit head domain also suggests bS21 may have a role in modulating conformational dynamics of the head domain relative to the body and platform of the 30S subunit. Rearrangements of the 30S subunit head domain are seen in every stage of the translation cycle[4]. Although we could align putative S21 homologues from huge phages[35] with specific bacterial clades, and show that many also possess KLY-like motifs, there were no clear relationships between the predicted consensus ribosome binding sites in these bacteria and these phages. It is possible that bS21 and the phage homologues interact with nearby mRNA sequences 5' of the Shine-Dalgarno helix, affecting translation initiation in this way. Taken together, the structural and phylogenetic information on bS21 and the phage S21 homologues raise new questions about their role in translation and the phage life cycle, i.e. whether they contribute to specialized translation and/or help phage evade bacterial defenses.

The rotameric nature of nucleic acid backbones has historically been a challenge for modeling the sugar-phosphate conformation, in contrast to the generally well-ordered bases[80]. Ribose puckers, for example, are directly visualized only at better than ~2 Å resolution, but significantly affect the remaining backbone dihedrals[81]. Much work has been done to simplify the multidimensional problem of modeling RNA conformers given the scarcity of high resolution RNA structures[82]. While some areas of the present structure show backbone details very clearly (**Figure 1C**), some level of disorder is observed in the conformations of many other residues (**Figure 1 - figure supplement 7**). Our initial impression was that this might be due to radiation damage. However, reconstructions with the first 2-3 frames of the exposure reveal similar breaks, and sometimes new breaks, in the EM density. Previous work has shown that at this dose, amino acid residues well known to be highly susceptible to radiation damage should be better preserved[83]. Furthermore, it is known that nucleic acids tend to be more resilient to X-ray damage compared to the most beam-sensitive moieties in proteins[84]. Because the same general trends in specific radiation damage seem to hold for cryo-EM[83], and noting that the global resolutions of the low-dose 70S reconstructions remain resolved to ~2.1-2.2 Å, the persistence of the broken density at low doses is more likely to be a result of disorder in the backbone. Our observation that poorer connectivity in the RNA backbone seems to be more common in regions lacking close contacts to other regions of the structure tracks with this conclusion, while a minority of cases where only a single bond in the ribose appears to be clearly broken are more of a puzzle. Closer investigation of these features may reveal more quantitative information about nucleotide rotamer preferences.

With regard to resolution, in this work we have supplemented the reporting of the "gold-standard" FSC with map-to-model FSC curves for our maps and for comparisons to previous work. Although the map-to-model FSC metric has been described for some time, it is not routinely used in the ribosome field[6,12–14,40,41]. Acknowledging that there is no substitute for visual inspection of the map to determine its quality, it is necessary to also consider which metrics are useful on the scale of the questions being answered. Sub-Ångstrom differences in resolution as reported by half-map FSCs have significant bearing on chemical interactions at face value but may lack usefulness if map correlation with the final atomic model is not to a similar resolution. For example, maps from recent cryo-EM reconstructions of the bacterial 50S ribosomal subunit report resolutions of ~2.1 Å-2.3 Å, but the deposited models reach global resolutions of ~2.3 Å-2.5 Å by the map-to-model FSC criterion[12–14] (see Methods) (**Figure 7–figure supplement 1**). Notably, for the 2.1 Å map of the *E. coli* 50S subunit based on half-map FSC values[12], the map-to-model FSC fit of our 50S subunit model to that map has a higher resolution (2.07 Å), compared to the deposited model (2.29 Å, PDB entry 6XZ7) (**Figure 7–figure supplement 1**). Thus, although the half-map FSC tells us something about the best model one might achieve, the map-to-model FSC captures new information that lies in how the model was generated and refined. Additionally, while map-to-model FSC calculations carry intrinsic bias from the model's dependence on the map, model refinement procedures leverage well-defined chemical properties (i.e. bond lengths, angles, dihedrals, and steric restraints) that are entirely independent of the map and should ensure realism. This is perhaps a reason why the map-to-model FSC appears in recent work more focused on methods and tool development[85–87].

Aside from global high resolution, the conformational heterogeneity of the ribosome also calls attention to the tools used for working with complexes that display variable resolution. Methods for refining heterogeneous maps have proliferated, including multi-body refinement[88] and 3D variability analysis[89] among others, but ways to work with and create a model from many maps of the same complex have not yet been standardized and require substantial manual

intervention. For example, we built and refined regions of the ribosome separately into focus-refined maps, using real-space refinement in Chimera[90] and Coot[91] to "repair" the breakpoints between model segments. Creation of composite maps from multiple refinements also suffers from imperfect stitching between refinements of distinct domains, and in our own composite map we observe that the highest resolution components degrade somewhat in the process. We also note that B-factor refinement in phenix.real_space_refinement is still under development, for example only allowing grouped B-factor refinement for nucleotides and amino acids. B-factor refinement in phenix.real_space_refine also results in unrealistic values for parts of the model, for example by exhibiting coupling to the global B-factor applied to the map used in the refinement. Other strategies for deriving an analogue to the B-factor suitable for cryo-EM are still in development[92].

Moreover, as it tends to be most convenient to develop tools with the highest resolution possible, it is common for new methods to utilize very high-resolution maps of apoferritin as a standard, which is highly symmetric and well-ordered. Specimens that do not share these characteristics may require new tools that move beyond these assumptions. Modeling tools for RNA also generally lag those for proteins. Here, we were able to use the maps and the highly solvated nature of RNA secondary and tertiary structure to address parameterization of solvent modeling in PHENIX (phenix.douse)[22]. For the present 70S map, the half-map FSC $\geq$ 0.97 up to 3.3 Å resolution, which is estimated to represent a theoretical correlation with a "perfect" map up to 0.99[86]. Structural information with certainty up to so-called "near-atomic" resolution has potential use in benchmarking newer tools and may specifically make our results valuable in addressing issues with focused or multi-body refinement. This structure also has potential use for aiding the future development of *de novo* RNA modeling tools, which are historically less developed compared to similar tools for proteins, and often rely on information generated from lower-resolution RNA structures[93]. Finally, our micrographs uploaded to the Electron Microscopy Public Image Archive (EMPIAR)[94] should serve as a resource for ribosome structural biologists and the wider cryo-EM community to build on the present results.

METHODS

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
strain, strain background	*Escherichia coli* MRE600	Given by Arto Pulk, UC Berkeley	ATCC #29417, NCTC #8164	Strain with low ribonuclease activity
other	300 mesh R1.2/1.3 UltraAuFoil grids	Electron Microscopy Sciences	Q350AR13 A	
software, algorithm	SerialEM	[95]	RRID:SCR_017293	
software, algorithm	MotionCor2	[96]	RRID:SCR_016499	
software, algorithm	CTFFind4	[97]	RRID:SCR_016732	
software, algorithm	RELION	[23]	Version 3&3.1 RRID:SCR_016274	
software, algorithm	cryoSPARC	[98]	Version 2 RRID:SCR_016501	
software, algorithm	Chimera	[90]	RRID:SCR_004097	
software, algorithm	PHENIX	[22]	RRID:SCR_014224	
software, algorithm	Coot	[91]	RRID:SCR_014222	

Biochemical preparation

E. coli 70S ribosome purification[99] and tRNA synthesis, purification, and charging[100] were performed as previously described. Briefly, 70S ribosomes were purified from *E. coli* MRE600 cells using sucrose gradients to isolate 30S and 50S ribosomal subunits, followed by subunit

reassociation and a second round of sucrose gradient purification. Transfer RNAs were transcribed from PCR DNA templates using T7 RNA polymerase, and purified by phenol-chloroform extraction, ethanol precipitation, and column desalting. Flexizyme ribozymes were used to charge the P-site $tRNA^{fMet}$ with either pentafluorobenzoic acid or malonate methyl ester and the A-site $tRNA^{Val}$ with valine[101]. Ribosome-mRNA-tRNA complexes were formed non-enzymatically by incubating 10 μM P-site tRNA, 10 μM mRNA, and 100 μM paromomycin with 1 μM ribosomes for 15 min at 37 °C in Buffer AC (20 mM Tris pH 7.5, 100 mM NH_4Cl, 15 $MgCl_2$, 0.5 mM EDTA, 2 mM DTT, 2 mM spermidine, 0.05 mM spermine). Then, 10 μM A-site tRNA was added and the sample was incubated for an additional 15 min at 37 °C. Complexes were held at 4 °C and diluted to 100 nM ribosome concentration in the same buffer immediately before grid preparation.

The mRNA of sequence 5'-GUAUAA**GGAGG**UAAAA*AUGGUA*UAACUA-3' was chemically synthesized (IDT) and was resuspended in water without further purification. The Shine-Dalgarno sequence is shown in bold, and the Met-Val codons are in italics.

Cryo-EM sample preparation

300 mesh R1.2/1.3 UltraAuFoil grids from Quantifoil with an additional amorphous carbon support layer were glow discharged in a Pelco sputter coater. 4 μL of each sample was deposited onto grids and incubated for 1 minute, then washed in Buffer AC with 20 mM NH_4Cl rather than 100 mM NH_4Cl. Grids were plunge-frozen in liquid ethane using a Vitrobot Mark IV with settings: 4 °C, 100% humidity, blot force 6, blot time 3.

Data Acquisition

Movies were collected on a 300-kV Titan Krios microscope with a GIF energy filter and Gatan K3 camera. Super-resolution pixel size was 0.355 Å, for a physical pixel size of 0.71 Å. SerialEM[95] was used to correct astigmatism, perform coma-free alignment, and automate data collection. Movies were collected with the defocus range -0.6 to -1.5 μm and the total dose was 39.89 $e^-/Å^2$ split over 40 frames. One movie was collected for each hole, with image shift used to collect a series of 3×3 holes for faster data collection[102], and stage shift used to move to the center hole. Based on the 1.2/1.3 grid hole specification, this should correspond to a maximum image shift of ~1.8 μm, although the true image shift used was not measured. The beam size was chosen such that its diameter was slightly larger than that of the hole, i.e. > 1.2 μm, although we have observed variation in the actual hole size compared to the manufacturer specifications.

Image processing

Datasets of 70S ribosome complexes with the two differently charged P-site tRNAs were initially processed separately. Movies were motion-corrected with dose weighting and binned to the recorded physical pixel size (0.71 Å) within RELION 3.0[103] using MotionCor2[96]. CTF estimation was done with CTFFind4[97], and micrographs with poor CTF fit as determined by visual inspection were rejected. Particles were auto-picked with RELION's Laplacian-of-Gaussian method. 2D classification was performed in RELION, and 4x binned particles were used for all classification steps. Particles were separated into 3D classes in cryoSPARC heterogeneous refinement[98], using an initial model generated from PDB 1VY4 with A- and P-site tRNAs[104] low-pass filtered to the default 20 Å resolution, and keeping particles that were sorted into well-resolved 70S ribosome classes. Particles were migrated back to RELION to generate an initial 3D-refined volume (reference low-pass filtered to 60 Å) on which to perform masked 3D classification without alignment to further sort particles based on A-site tRNA occupancy. CTF Refinement and

Bayesian polishing were performed in RELION 3.1 before pooling the two datasets together, with nine optics groups defined based on the 3×3 groups for image shift-based data collection. The resulting 70S ribosome reconstruction was used as input for focused refinements of the 50S and 30S subunits. We used rigid-body docked coordinates for the 70S ribosome, individual ribosomal subunits or domains (30S subunit head, 50S subunit Central Protuberance) to define the boundaries of the map regions to be used in the focused refinements. Focused refinement of the Central Protuberance was performed starting from the 50S subunit-focused refinement reconstruction, and head- and platform-focused refinements started from the 30S subunit focused refinement reconstruction. Ewald sphere correction, as implemented in RELION 3.1 with the single side-band correction[23,105], provided some additional improvements in resolution (**Tables 1-2**).

In addition to using the 40-frame movies, we used the first 3 frames corresponding to a ~3 electron/Å^2 dose to calculate 3D reconstructions, including focused refinements of the 30S and 50S subunits. The focused-refined map of the 30S subunit had a resolution of 2.45 Å by the map-to-model FSC metric. These maps were used to examine the density for the isoAsp in uS11, which lacked clear density for the side chain in maps reconstructed from the full 40-frame movies. We also used the maps from the initial 3 frames to examine connectivity in ribose density, to determine if there is visual evidence for the impact of electron damage.

Modeling

The previous high-resolution structure of the *E. coli* 70S ribosome[10] was used as a starting model. We used the "Fit to Map" function in Chimera[90] to calibrate the magnification of the cryo-EM map of the 50S ribosomal subunit generated here to maximize correlation, resulting in a pixel size of 0.7118 Å rather than the recorded 0.71 Å. Focused-refined maps were transformed into the frame of reference of the 70S ribosome for modeling and refinement, using the "Fit to Map" function in Chimera, and resampling the maps on the 70S ribosome grid. The 50S and 30S subunits were refined separately into their respective focused-refined maps using PHENIX real-space refinement (RSR)[22]. Protein and rRNA chains were visually inspected in Coot[91] and manually adjusted where residues did not fit well into the density, making use of B-factor blurred maps where needed to interpret regions of lower resolution. Focused-refined maps on smaller regions were used to make further manual adjustments to the model, alternating with PHENIX RSR. Some parts of the 50S subunit, including H69, H34 and the tip of the A-site finger, were modeled based on the 30S subunit focused-refined map. The A-site and P-site tRNAs were modeled as follows: anticodon stem-loops, 30S subunit focused-refined map; P-site tRNA body, 50S subunit focused-refined map, with a B-factor of 20 Å^2 applied; A-site tRNA body, 30S subunit focused-refined map and 50S subunit focused-refined map with B-factors of 20 Å^2 applied; tRNA-ACCA 3' ends, 50S subunit focused-refined map with B-factors of 20-30 Å^2 applied. Alignments of uS15 were generated using BLAST[106] with the *E. coli* sequence as reference. The model for bL31A (*E. coli* gene rpmE) was manually built into the CP and 30S subunit head domain focused-refined maps before refinement in PHENIX.

A model for paromomycin was manually docked into the 30S subunit focused-refined map, followed by real-space refinement in Coot and in PHENIX. Comparisons to prior paromomycin structural models (PDB codes 1J7T, 2VQE, and 4V51)[57–59] used least-squares superposition of paromomycin in Coot. Although ring IV is in different conformations in the various paromomycin models, the least-squares superposition is dominated by rings I–III, which are in nearly identical conformations across models.

Ribosome solvation including water molecules, magnesium ions, and polyamines was modeled using a combination of PHENIX (phenix.douse) and manual inspection. The phenix.douse feature was run separately on individual focused-refined maps, and the resulting solvent models were combined into the final 30S subunit and 50S subunit models. Due to the fact that the solvent conditions used here contained ammonium ions and no potassium, no effort was made to systematically identify monovalent ion positions. The numbers of various solvent molecules are given in **Table 2**.

Along with the individual maps used for model building and refinement, we have also generated a composite map of the 70S ribosome from the focused-refined maps for deposition to the PDB and EMDB for ease of use (however, experimental maps are recommended for examination of high-resolution features). We made the composite map using the "Fit in map" and vop commands in Chimera. First, we aligned the unmasked focus-refined maps with the 70S ribosome map using the "Fit in map" tool. We then used the "vop resample" command to transform these aligned maps to the 70S ribosome grid. After the resampling step, we recorded the map standard deviations as reported in the "Volume Mean, SD, RMS" tool. Then, we added the maps sequentially using "vop add" followed by rescaling the intermediate maps to the starting standard deviation using the "vop scale" command.

Modeling of isoAsp residues in uS11

Initial real-space refinement of the 30S subunit against the focused-refined map using PHENIX resulted in a single chiral volume inversion involving the backbone of N119 in ribosomal protein uS11, indicating that the L-amino acid was being forced into a D-amino acid chirality, as reported by phenix.real_space_refine. Of the 10,564 chiral centers in the 30S subunit model, the Cα of N119 had an energy residual nearly two orders of magnitude larger than the next highest deviation. Inspection of the map in this region revealed clear placement for carbonyl oxygens in the backbone, and extra density consistent with an inserted methylene group, as expected for isoAsp. The model of isoAsp at this position was refined into the cryo-EM map using PHENIX RSR, which resolved the stereochemical problem with the Cα chiral center. IsoAsp was also built and refined into models of archaeal and eukaryotic uS11 based on cryo-EM maps of an archaeal 30S ribosomal subunit complex (PDB 6TMF)[40] and a yeast 80S ribosome complex (PDB 6T4Q)[41]. These models were refined using PHENIX RSR, and real-space correlations by residue calculated using phenix.model_map_cc.

Phylogenetic analysis of uS11 and its rRNA contacts

All archaeal genomes were downloaded from the NCBI genome database (2,618 archaeal genomes, last accessed September 2018). Due to the high number of bacterial genomes available in the NCBI genome database, only one bacterial genome per genus (2,552 bacterial genomes) was randomly chosen based on the taxonomy provided by the NCBI (last accessed in December 2017). The eukaryotic dataset comprises nuclear, mitochondrial and chloroplast genomes of 10 organisms (*Homo sapiens, Drosophila melanogaster, Saccharomyces cerevisiae, Acanthamoeba castellanii, Arabidopsis thaliana, Chlamydomonas reinhardtii, Phaeodactylum tricornutum, Emiliania huxleyi, Paramecium aurelia* and *Naegleria gruberi*).

Genome completeness and contamination were estimated based on the presence of single-copy genes (SCGs) as described in[107]. Only genomes with completeness >70% and contamination <10% (based on duplicated copies of the SCGs) were kept and were further de-replicated using

dRep at 95% average nucleotide identity (version v2.0.5)[108]. The most complete genome per cluster was used in downstream analyses.

Ribosomal uS11 genes were detected based on matches to the uS11 Pfam domain (PF00411)[109] using hmmsearch with an E-value below 0.001[110]. Amino acid sequences were aligned using the MAFFT software (version v7.453)[111]. Alignment was further trimmed using Trimal (version 1.4.22) (--gappyout option)[112]. Tree reconstruction was performed using IQ-TREE (version 1.6.12)[113], using ModelFinder[114] to select the best model of evolution, and with 1000 ultrafast bootstrap[115]. The tree was visualized with iTol (version 4)[116] and logos were made using the weblogo server[117].

16S and 18S rRNA genes were identified from the prokaryotic and eukaryotic genomes using the method based on hidden Markov model (HMM) searches using the cmsearch program from the Infernal package[118] and fully described in[119]. The sequences were aligned using the MAFFT software.

Phylogenetic analysis of bS21 and phage S21 homologues

S21 sequences were retrieved from the huge phage database described in[35]. Cd-hit was run on the set of S21 sequences to reduce the redundancies[120] (default parameters) (version 4.8.1). Non redundant sequences were used as a query against the database of prokaryotic genomes used for uS11 above using BLASTP (version 2.10.0+) (e-value 1e-20)[121]. Alignment and tree reconstruction were performed as described for uS11 except that we did not perform the alignment trimming step.

Phylogenetic analysis of YcaO genes

Similarly to uS11, the YcaO sequences were identified in prokaryotic genomes based on its PFAM accession (PF02624)[109] using hmmsearch with an E-value below 0.001[110]. Amino acid sequences were aligned using the MAFFT software (version v7.453)[111]. Alignment was further trimmed using Trimal (version 1.4.22) (--gappyout option)[112]. Tree reconstruction was performed using IQ-TREE (version 1.6.12)[113], using ModelFinder[114] to select the best model of evolution, and with 1000 ultrafast bootstraps[115]. The tree was visualized with iTol (version 4)[116]. The three genes downstream and upstream of each *YcaO* gene were identified and annotated using the PFAM[109] and the Kegg[122] databases.

Map-to-model FSC calculations

Masks for each map were generated in two ways. First, to calculate map-to-model FSC curves for comparisons of the present models with the cryo-EM maps generated here, we used masked maps generated by RELION during postprocessing (citation). The effective global resolution of a given map is given at the FSC cutoff of 0.5 in **Table 2** and **Figure 1–figure supplements 2-3**. Second, we used refined PDB coordinates for the 70S ribosome, individual ribosomal subunits or domains (30S subunit head, 50S subunit Central Protuberance) for comparisons to the 70S ribosome map or focused-refined maps, and to previously published maps and structural models. Masks for each map were generated in Chimera[90] using the relevant PDB coordinates as follows. A 10 Å resolution map from the coordinates was calculated using molmap, and the surface defined at 1 standard deviation was used to mask the high-resolution map. For the present models and maps, the effective global resolution of a given map using this second approach was similar or slightly lower than that using the approach in RELION (within a few hundredths of an Å).

Map-to-model comparisons for other 50S subunit reconstructions (emd_20353, emd_10077)

For the recent *E. coli* 50S subunit structure[13], we used Chimera to resize the deposited map (emd_20353) to match the dimensions of the maps presented here. Briefly, our atomic coordinates for the 50S subunit were used with the "Fit to Map" function and the voxel size of the deposited map was calibrated to maximize correlation. The resulting voxel size changed from 0.822 Å to 0.8275 Å in linear dimension. After rescaling the deposited map, we used phenix.model_map_cc to compare the map with rescaled atomic coordinates deposited in the PDB (6PJ6) or to the present 50S model, yielding a map-to-model FSC of 0.5 at ~2.5 Å. Similar comparisons of the structure of the *Staphylococcus aureus* 50S subunit to the deposited map (PDB 6S0Z, emd_10077)[14] yielded a map-to-model FSC of 0.5 at 2.43 Å, accounting for a change in voxel linear dimension from 1.067 Å to 1.052 Å. For comparisons to the map and model deposited by[12] (EMD-10655 and PDB 6XZ7), we removed tRNA bodies, the L1 arm and the GTPase-Associated-Center coordinates from 6XZ7 since these regions are disordered or missing in the deposited 50S subunit reconstruction.

Analysis of uL16 mass spectrometry datasets

Previously published *E. coli* tryptic peptide mass spectrometry (MS/MS) raw data was used for the analysis[67] (MassIVE accession: MSV000081144). Peptide searches were performed with MSFragger[123] using the default parameters for a closed search with the following exceptions: additional variable modifications were specified on residues R (hydroxylation, Δ mass: 15.9949) and M (thioamide, Δ mass: 15.9772), maximum modifications per peptide set to four, and multiple modifications on a residue were allowed. Spectra were searched against a database of all *E. coli* proteins plus common contaminants concatenated to a decoy database with all original sequences reversed. Results were analyzed using TPP[124] and Skyline[125].

Figure preparation

Cryo-EM maps were supersampled in Coot for smoothness. Figure panels showing structural models were prepared using Pymol (Schrödinger) and ChimeraX[126]. Sequence logo figures made with WebLogo 3.7.4[117]. Phylogenetic trees were visualized with iTol (version 4)[116] and multiple alignments were visualized with geneious 9.0.5 (https://www.geneious.com).

Data deposition

Ribosome coordinates have been deposited in the Protein Data Bank (entry **7K00**), maps in the EM Database (entries **EMD-22586**, **EMD-22607**, **EMD-22614**, **EMD-22632**, **EMD-22635**, **EMD-22636**, and **EMD-22637** for the 70S ribosome composite map, 70S ribosome, 50S subunit, 30S subunit, 30S subunit head, 30S subunit platform, and 50S subunit CP maps, respectively), and raw movies in EMPIAR (entry **EMPIAR-10509**).

ACKNOWLEDGMENTS

We thank Dan Toso and Paul Tobias for assistance with cryo-EM data collection, Kyle Hoffman and Dieter Söll for supplying tRNAs, Andrew Cairns and Aaron Featherstone for monomer synthesis, Pavel Afonine for discussions and help with phenix.douse, Douglas Mitchell for discussions on the thioamide linkage, and Dieter Söll and Nikolay Aleksashin for comments on the manuscript. This work was funded primarily by the Center for Genetically Encoded

Materials (NSF No. CHE-2021739) with additional contributions from NIH No. GM R01-114454 (F.W. support). O.A. was supported in part by Agilent Technologies as an Agilent Fellow. R.M. and J.F.B. were supported by the Innovative Genomics Institute at Berkeley and the Chan Zuckerberg Biohub.

SUPPLEMENTARY INFORMATION

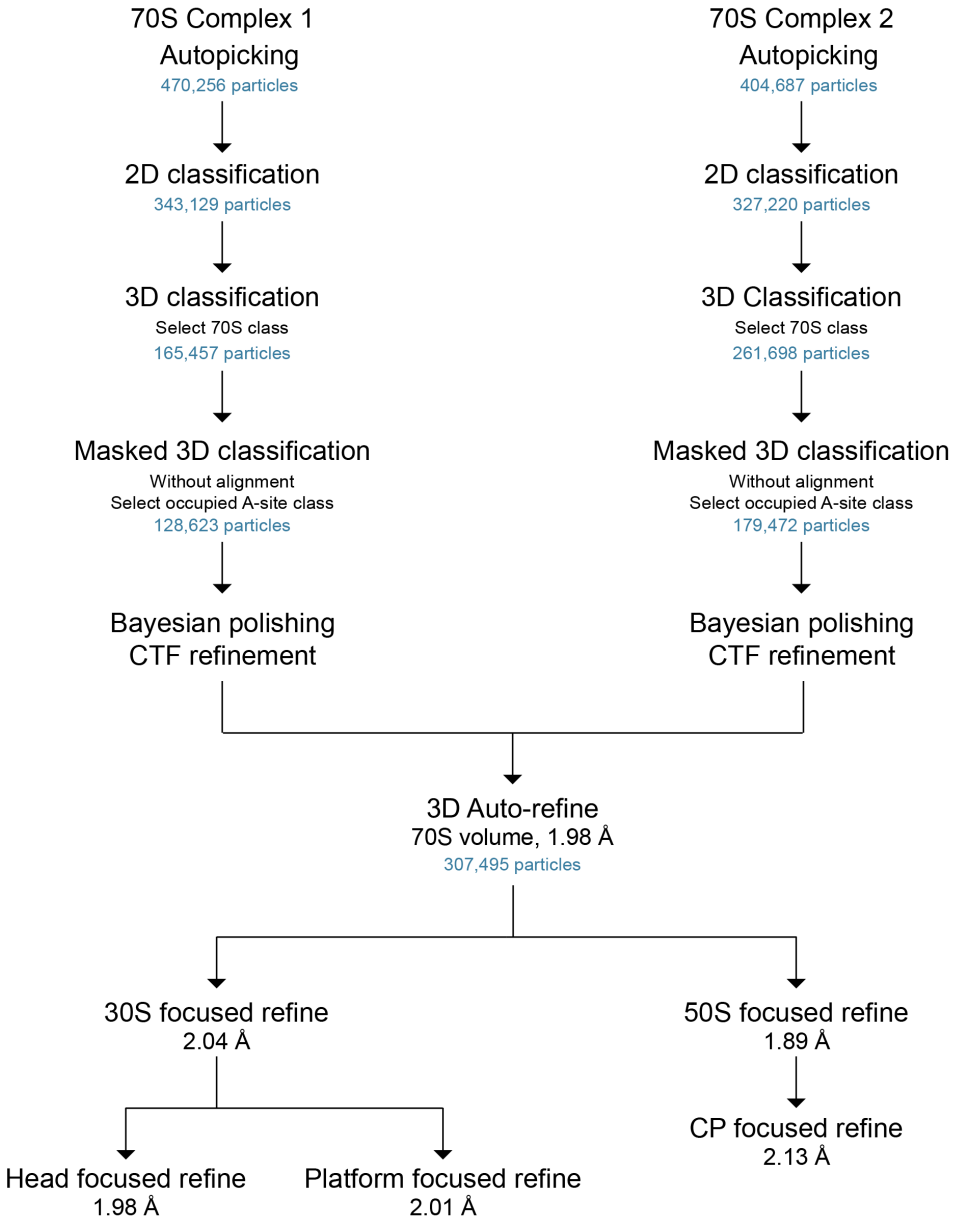


Figure 1–figure supplement 1. General scheme of cryo-EM data processing workflow. The number of particles at each stage are shown, as well as the final resolutions as defined by half-map FSCs (cutoff of 0.143).

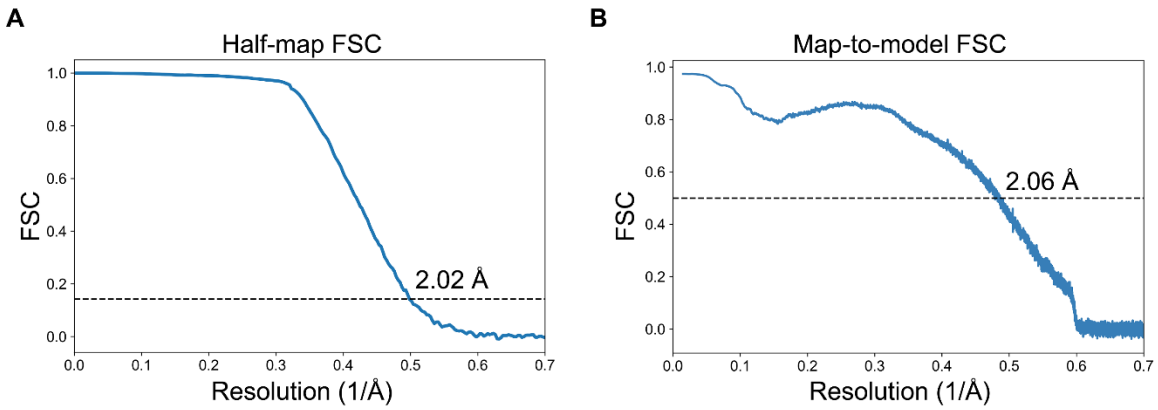


Figure 1–figure supplement 2. Fourier Shell Correlations for cryo-EM maps of the 70S ribosome. (**A**) Half-map FSC curve for the 70S ribosome is shown. The "gold-standard" cutoff value for resolution (0.143) is indicated. (**B**) The map-to-model FSC curve for the 70S ribosome, with an overall resolution indicated at an FSC value of 0.5.

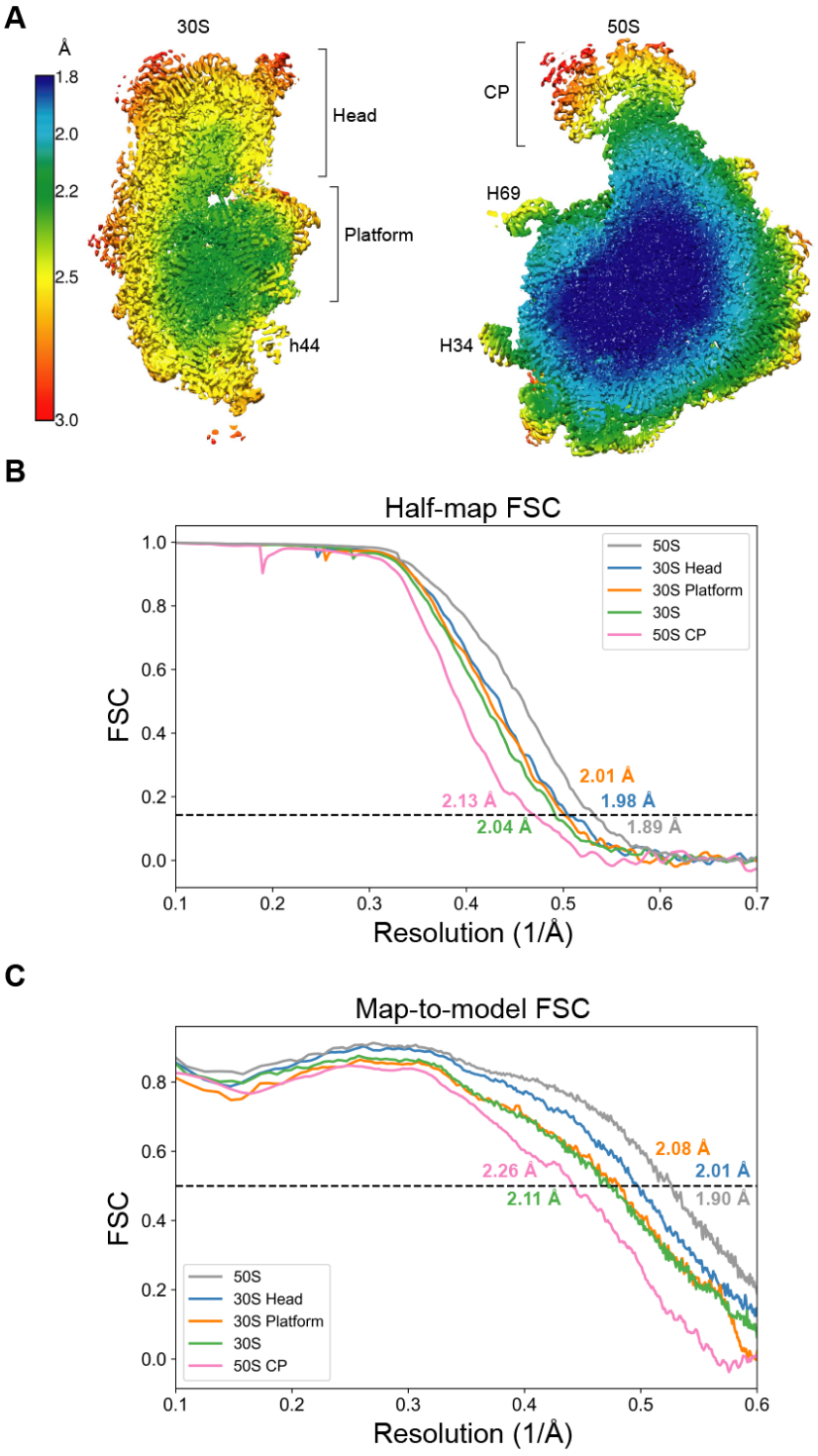


Figure 1–figure supplement 3. Resolution of maps of the 30S and 50S ribosomal subunits. (**A**) Local resolution of the 30S and 50S subunits. Brackets indicate regions masked for further focused refinement. 16S rRNA helix h44 and 23S rRNA helices H34 and H69 are also labeled. (**B**) Half-map FSC curves for focused-refined maps of the 50S subunit, 30S subunit, 50S Central Protuberance, 30S subunit head domain, and 30S subunit platform are shown. The "gold-standard" cutoff value for resolution (0.143) is indicated in each graph. (**C**) Map-to-model FSC curves for the focused-refined maps of the 50S subunit, 30S subunit, 50S Central Protuberance, 30S subunit head domain, and 30S subunit platform are shown.

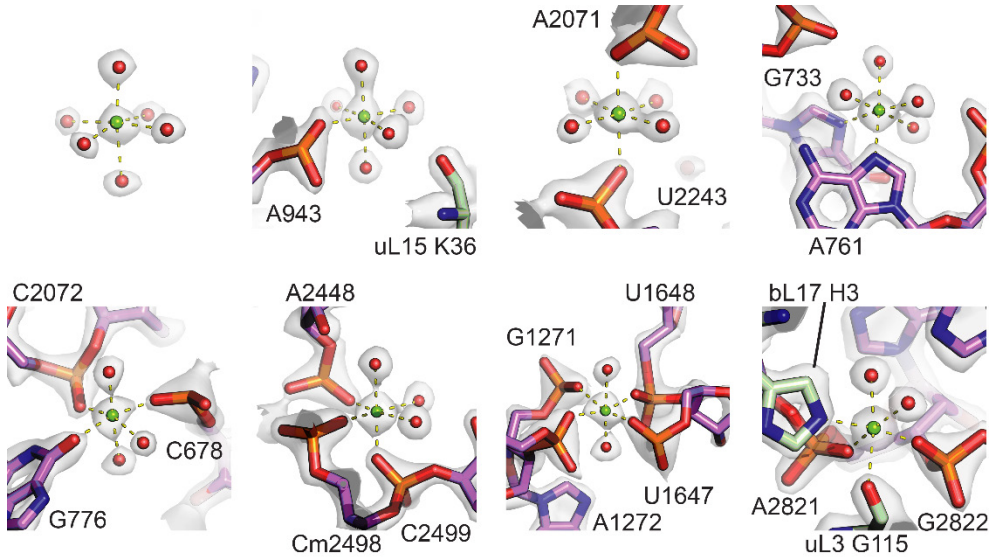


Figure 1–figure supplement 4. Gallery of Mg^{2+} coordination states observed in the 50S subunit. Examples include magnesium ions with different levels of direct coordination to the rRNA. The fully-hydrated Mg^{2+} is located near 23S rRNA nucleotide A973.

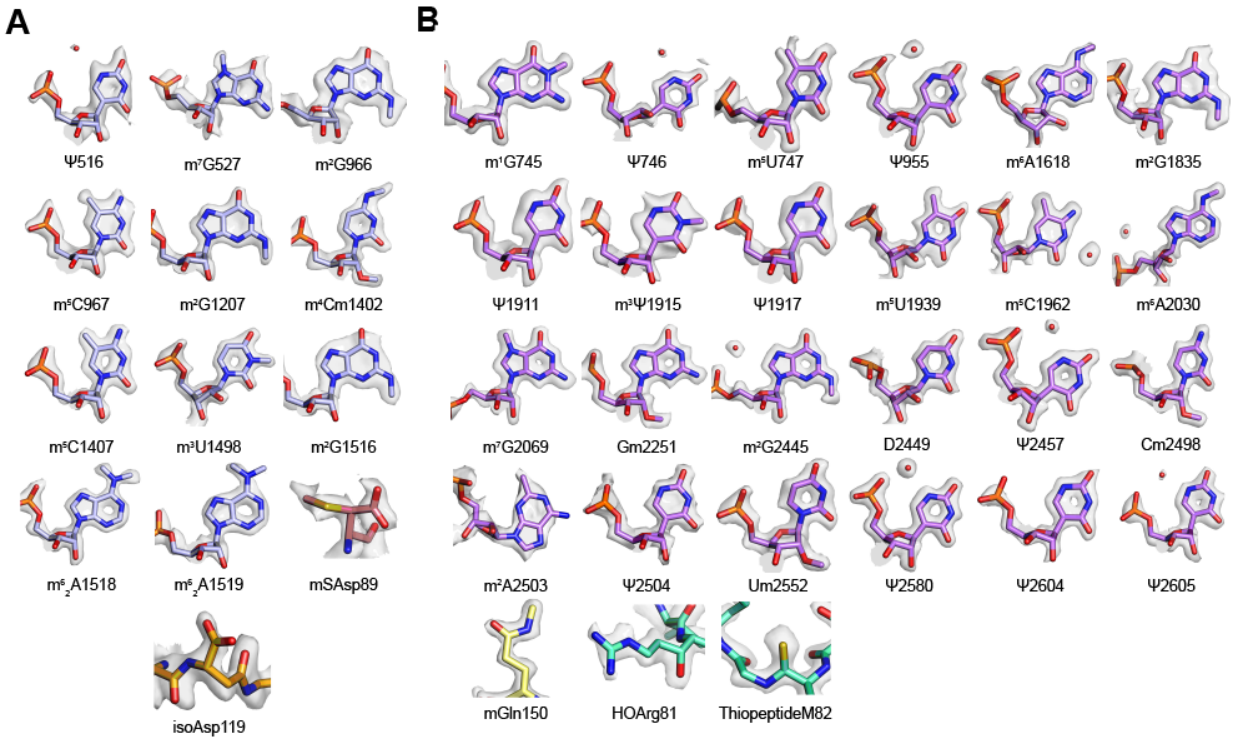


Figure 1–figure supplement 5. Gallery of post-transcriptionally modified nucleotides and post-translationally modified amino acids. (**A**) Post-transcriptionally modified nucleotides and amino acids in the 30S subunit. The density for m^7G527 and mSAsp89 is contoured at a lower level to show the presence of the modifications. (**B**) Post-transcriptionally modified nucleotides and amino acids in the 50S subunit. In panel **A**, nucleotides m^7G527 and $m^6{}_2$A1519 appear hypomodified, when compared to m^7G2069 in 23S rRNA, and when compared to other methylated nucleobases and $m^6{}_2$A1518 in 16S rRNA.

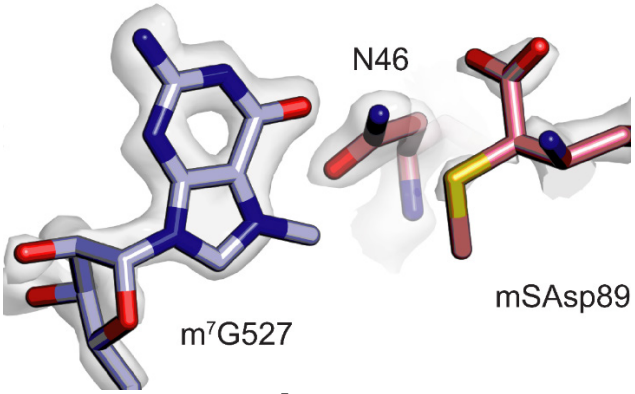


Figure 1–figure supplement 6. Close proximity of m^{7}G527 in 16S rRNA and β-methylthio-Asp89 in uS12. Both m^{7}G527 (light purple) and β-methylthio-Asp89 in uS12 (pink) appear hypomodified based on the cryo-EM map contour level required to enclose adjacent atoms and residues.

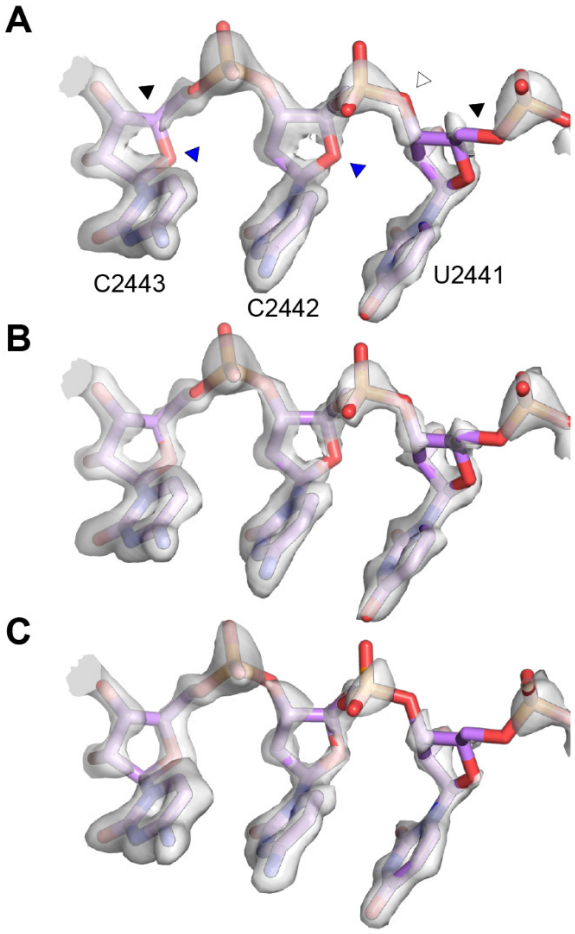


Figure 1–figure supplement 7. **Weak RNA backbone density in the 50S subunit.** (**A**) Map around bases C2443, C2442, and U2441 of the 50S in a reconstruction from frames 1-2 of the exposure, corresponding to the first ~2 electrons per Å^2. The same bases are shown for the first 3 frames (~3 electrons per Å^2) (**B**) and the final 50S-focused, Ewald sphere-corrected map with all frames included (~40 electrons per Å^2, with dose weighting) (**C**). Some features persist in all 3 reconstructions, for instance the break between C3' and C4' in C2443 or between C5' and O5' in U2441 (indicated with black triangles). Some features become less connected with longer exposure (e.g. O3' in C2441, indicated with white triangle), while others appear improved with longer exposure (e.g. O4' in C2443 and C2442, indicated with blue triangles).

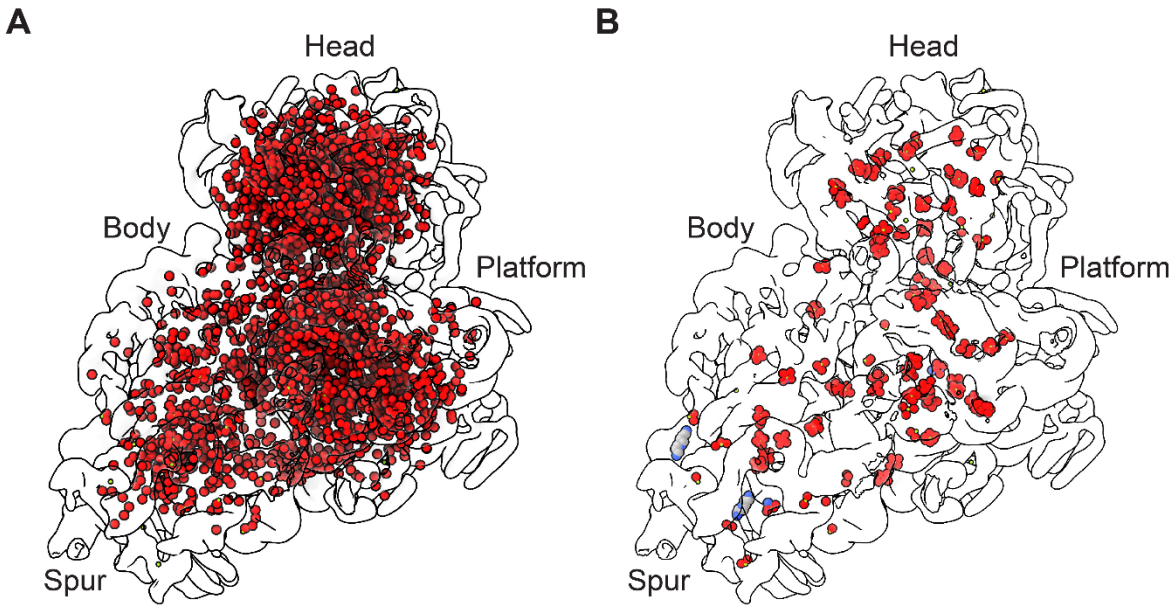


Figure 2–figure supplement 1. Solvation of the 30S ribosomal subunit. (**A**) Solvation of the 30S subunit with solvent oxygen atoms shown as red spheres. (**B**) Polyamine (grey carbon) and Mg^{2+} (green) sites in the 30S subunit, shown with metal-coordinating atoms (oxygen in red, nitrogen in blue). Cryo-EM density is shown with a low-pass filter of 8 Å.

Figure 3–figure supplement 1. Phage S21 and bacterial host bS21 sequence alignments. Sequence alignments of phage S21 homologues (in red) and associated host bacterial clades from (**A**) Bacteroidetes (orange), (**B**) The Candidate Phyla Radiation (magenta), (**C**) Betaproteobacteria (green) and Gammaproteobacteria (blue), (**D**) Spirochetes (blue-green) and (**E**) Firmicutes (black). Phage sequences in bold indicate phages with a predicted bacterial host.

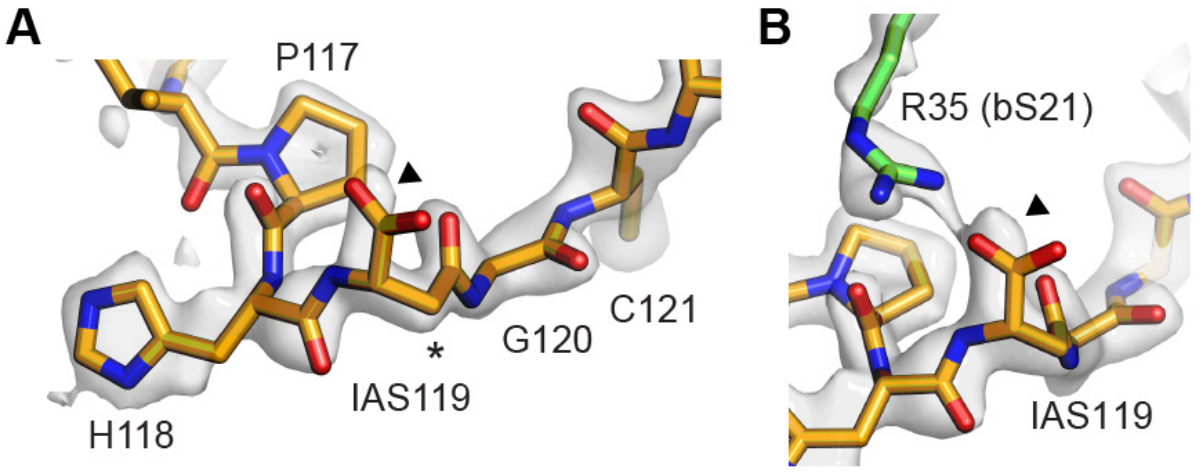


Figure 4–figure supplement 1. Cryo-EM density for uS11 based on early movie frames. Frames 1-3 of the acquired movies were used to calculate a 30S focused-refined cryo-EM map. (**A**) Model of isoaspartate 119 in uS11 and neighboring amino acids is shown in the density. The arrow points to the isoaspartate sidechain density, and the asterisk indicates the position of the additional backbone methylene group. Map-to-model resolution for the 30S subunit in this map was 2.45 Å and GS-FSC resolution was 2.26 Å. (**B**) Alternative view of isoaspartate sidechain in uS11 (orange) showing interaction with R35 of protein bS21 (green).

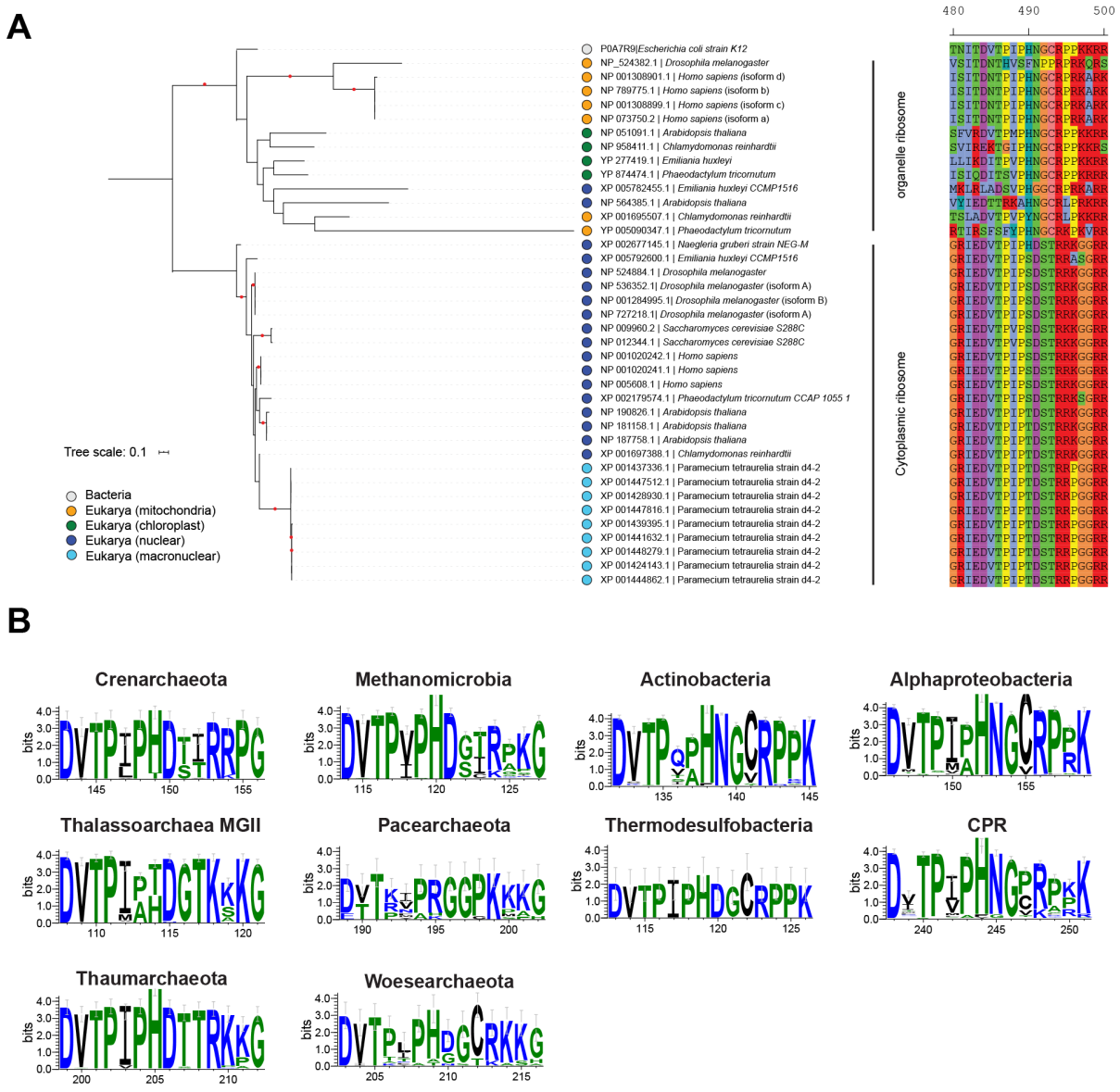


Figure 4–figure supplement 2. Conservation of residues near the isoAsp residue in uS11 homologues. (**A**) Phylogenetic tree of the uS11 ribosomal proteins in Eukaryotes, including both cytoplasmic and organelle examples, along with *Escherichia coli* and the amino acids around the PHNG motif. The maximum likelihood tree was constructed under an LG+G4 model of evolution. Nodes with bootstrap values ≥ 85 are indicated by red circles. Scale bar indicates the average substitutions per site. (**B**) Sequence conservation in uS11 homologues in different bacterial and archaeal phyla.

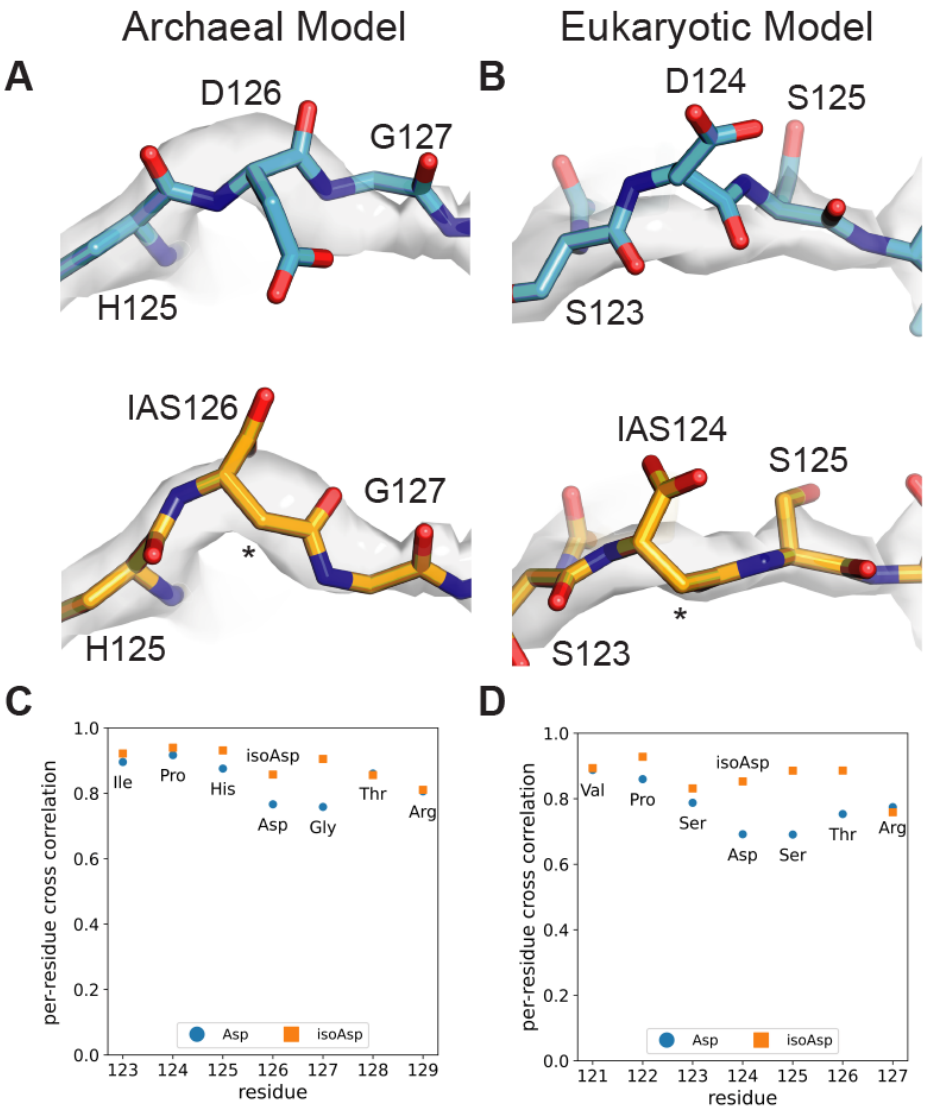


Figure 4–figure supplement 3. Structural models for isoaspartate in archaeal and eukaryotic ribosomes. Comparisons of published uS11 models (blue) in the archaeal[40] (**A**) and eukaryotic[41] ribosomes (**B**) with models incorporating the IsoAsp modification (yellow-orange), real-space refined into the corresponding published maps. (**C**) Real-space correlations of models refined into the archaeal 30S subunit cryo-EM map, on a per-residue basis. (**D**) Real-space correlations of models refined into the eukaryotic ribosome cryo-EM map.

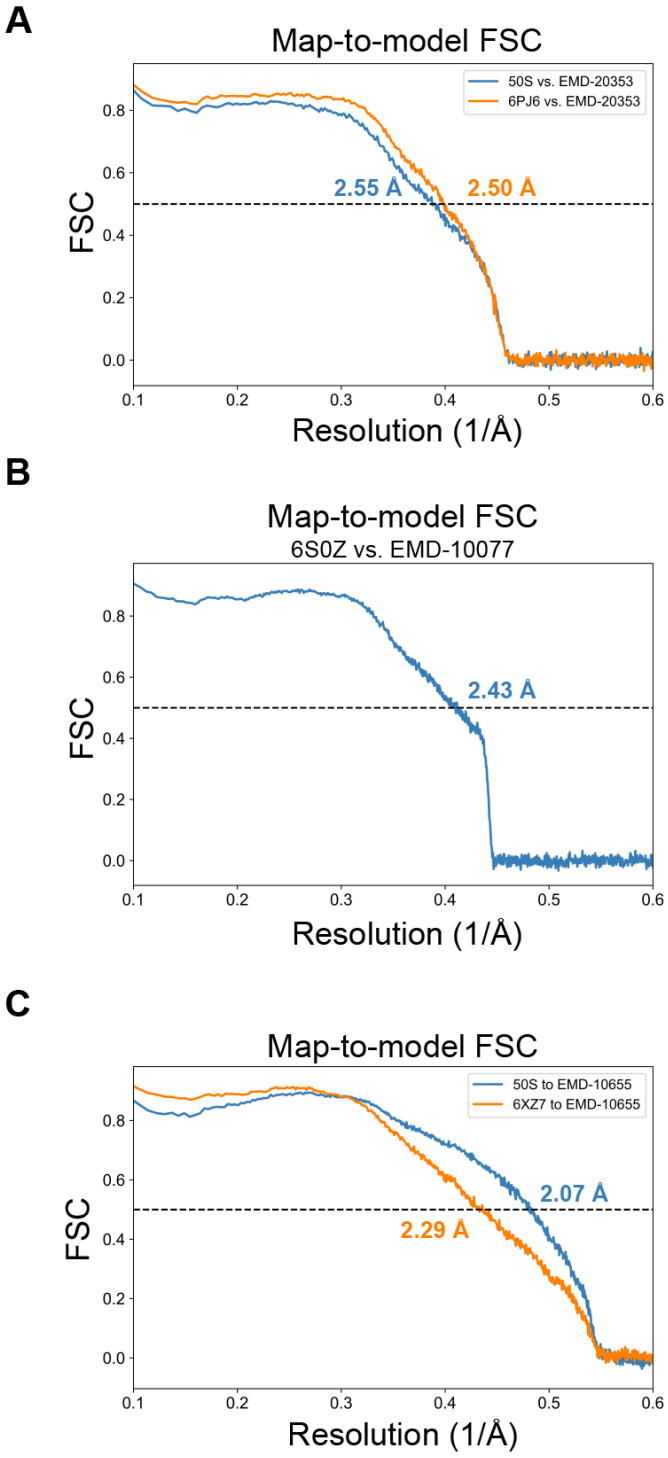


Figure 7–figure supplement 1. Map-to-model resolution estimates for deposited 50S subunit structures. (**A**) Map-to-model FSC curves calculated with the map from[13] (EMD-20353) against the associated model (orange; PDB: 6PJ6) and the 50S model presented here (blue). (**B**) Map-to-model FSC curves calculated for the map and model from[14] (EMD-10077; PDB: 6S0Z). (**C**) Map-to-model FSC curves calculated with the map from[12] (EMD-10655) against the associated model (orange; PDB: 6XZ7) and the 50S model presented here (blue).

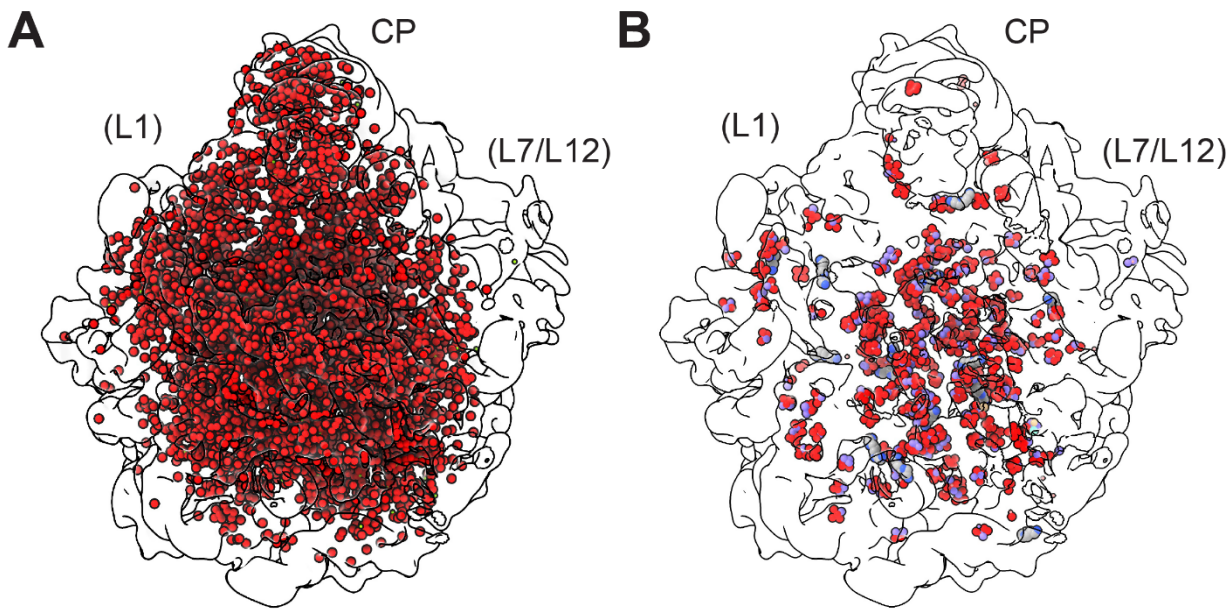


Figure 7–figure supplement 2. Solvation in the 50S ribosomal subunit. (**A**) Water molecules modeled in the 50S subunit are shown as red spheres (oxygen atoms) in outline of the map. View is from the 30S subunit interface side, with approximate locations of uL1 and bL12 (disordered in the structure) shown. (**B**) Polyamine (grey carbons) and Mg^{2+} (green) sites in the 50S subunit, shown with metal-coordinating atoms (oxygen in red, nitrogen in blue, and carbons in violet, ivory, cyan). Cryo-EM density is shown with a low-pass filter of 8 Å.

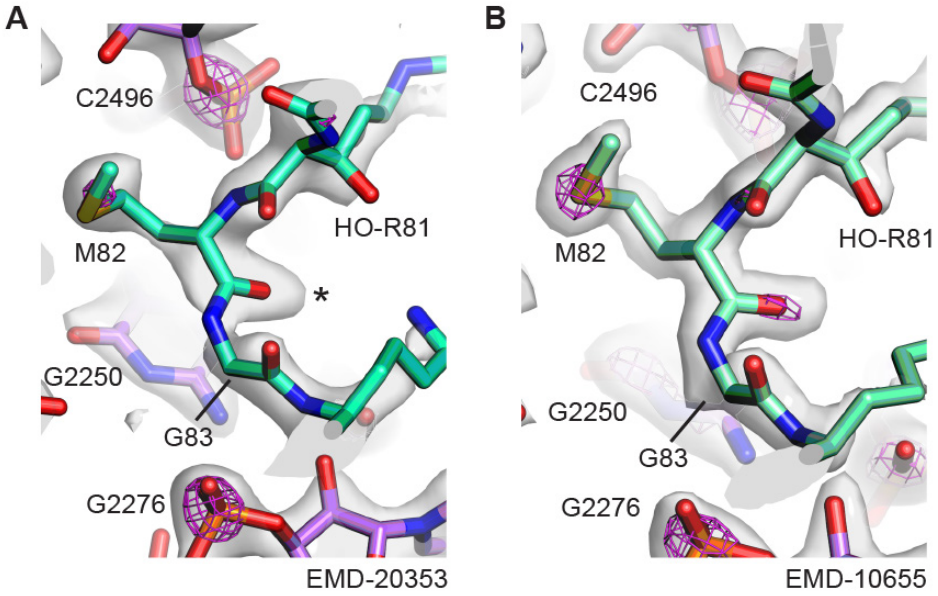


Figure 8–figure supplement 1. Thioamide density in other *E. coli* 50S ribosomal subunit cryo-EM reconstructions. (**A**) Structural model of thioamide between Met82 and Gly83 in uL16 (mint), with cryo-EM density of map EMD-20353 contoured at two levels to highlight sulfur and phosphorus atoms. The lower contour level is shown as a grey surface and the higher contour level is shown as fuchsia mesh. 23S rRNA is shown in purple. The asterisk highlights the position of the sulfur in the thioamide. (**B**) Structural model of thioamide between Met82 and Gly83 in uL16 (mint), with cryo-EM density of map EMD-10655 contoured at two levels to highlight sulfur and phosphorus atoms, as in panel (**A**).

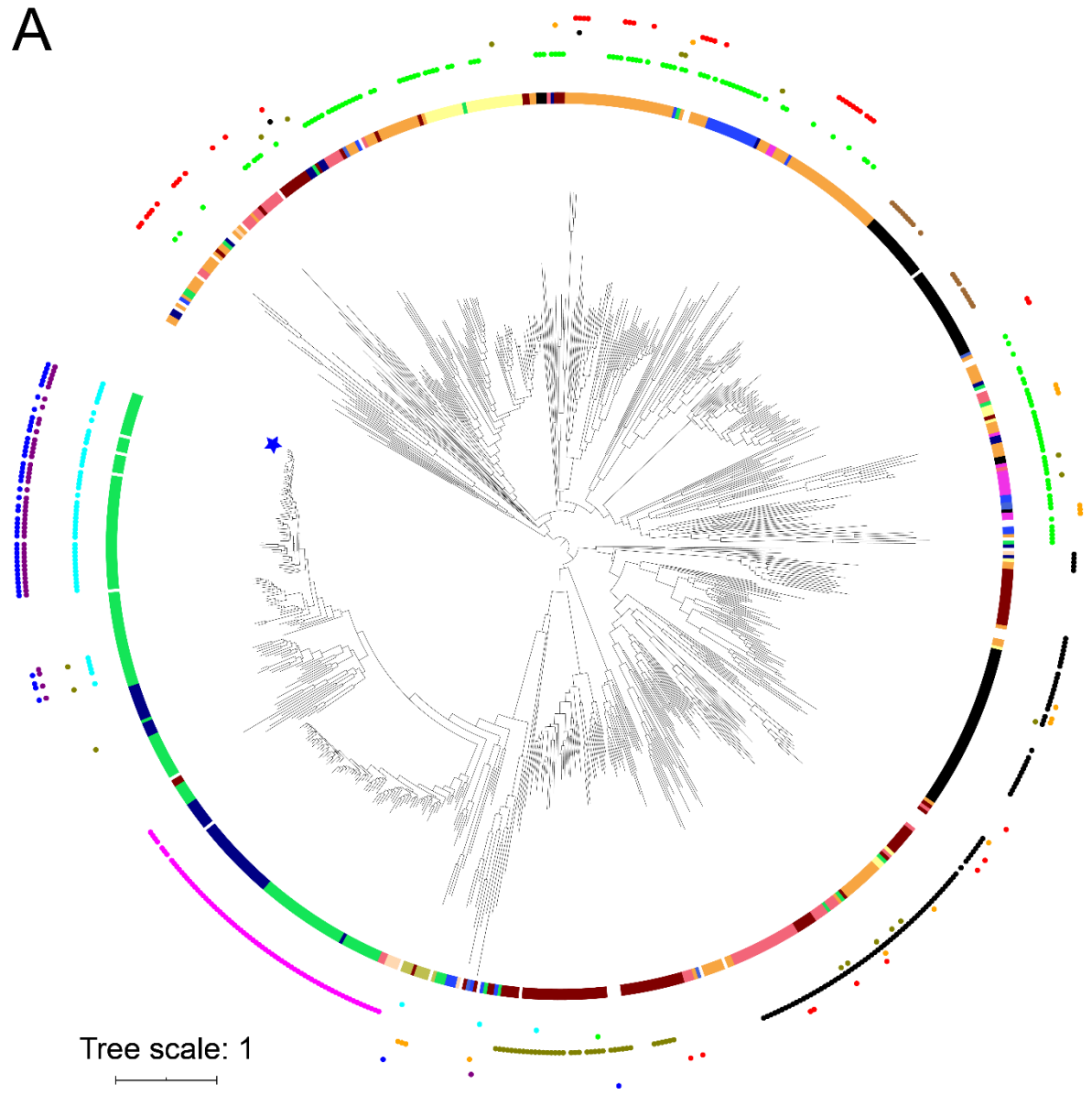

A
Tree scale: 1
Gene annotations
ThiF (PF00899)
TfuA (PF07812)
DUF63 (PF01889)
TPR repeats
focA (K06212)
OsmC (PF02566) + YcaO (PF02624)
pflA, pflC, pflE (K04069)
YcaO (PF02624)
Nitroreductase (PF00881)
E2.3.1.54, pflD (K00656)
Genome taxonomy
Betaproteobacteria
Gammaproteobacteria
Alphaproteobacteria
Deltaproteobacteria
Epsilonproteobacteria
Bacteroidetes
Firmicutes
Unknown
Actinobacteria
Cyanobacteria
Euryarchaeota
CPR
B
pflA
pflB
focA
YcaO
YcaP

Figure 8–figure supplement 2. Phylogeny of YcaO family members. (**A**) Phylogenetic tree of YcaO family members. The maximum likelihood tree was constructed under an LG+F+G4 model of evolution. Scale bar indicates the average substitutions per site. The inner circle shows organismal distribution, while the outer circle shows neighboring gene patterns. Star indicates location of *Escherichia coli* str K-12 substr MG1655. (**B**) Genes neighboring YcaO in *E. coli* K-12 substr MG1655.

The following supplementary files are available at https://elifesciences.org/articles/60482

Supplementary File 1. Phages encoding S21 homologs. Tabs include phages encoding S21 homologs with predicted bacterial hosts, along with ribosome binding sites for the phages, Betaproteobacteria, Firmicutes, CPR bacteria, Spirochaetes, and Bacteroidetes.

Supplementary File 2. Phylogenetic analysis of rRNA contacts near the uS11 isoAsp residue. Tabs include 16S base pair statistics for prokaryotes, bacteria, archaea, 16S rRNA genome information for prokaryotes, 18S base pair statistics for eukaryotes, 18S rRNA genome information for eukaryotes, and nucleotide statistics for position 718. All 16S rRNA base pairs and position 718 are with *E. coli* numbering. 18S rRNA base pairs are with *S. cerevisiae* numbering.

REFERENCES

1. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* 19, 560–567 (2012).
2. Rodnina, M. V. The ribosome as a versatile catalyst: reactions at the peptidyl transferase center. *Curr. Opin. Struct. Biol.* 23, 595–602 (2013).
3. Munro, J. B., Sanbonmatsu, K. Y., Spahn, C. M. T. & Blanchard, S. C. Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.* 34, 390–400 (2009).
4. Javed, A. & Orlova, E. V. Unravelling Ribosome Function Through Structural Studies. *Subcell. Biochem.* 93, 53–81 (2019).
5. Morse, J. C. *et al.* Elongation factor-Tu can repetitively engage aminoacyl-tRNA within the ribosome during the proofreading stage of tRNA selection. *Proc. Natl. Acad. Sci. U. S. A.* 117, 3610–3620 (2020).
6. Loveland, A. B., Demo, G. & Korostelev, A. A. Cryo-EM of elongating ribosome with EF-Tu•GTP elucidates tRNA proofreading. *Nature* (2020) doi:10.1038/s41586-020-2447-x.
7. Arenz, S. & Wilson, D. N. Bacterial Protein Synthesis as a Target for Antibiotic Inhibition. *Cold Spring Harb. Perspect. Med.* 6, (2016).
8. Yusupova, G. & Yusupov, M. Crystal structure of eukaryotic ribosome and its complexes with inhibitors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, (2017).
9. Frank, J. Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nat. Protoc.* 12, 209–212 (2017).
10. Noeske, J. *et al.* High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.* 22, 336–341 (2015).
11. Polikanov, Y. S., Melnikov, S. V., Söll, D. & Steitz, T. A. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.* 22, 342–344 (2015).
12. Pichkur, E. B. *et al.* Insights into the improved macrolide inhibitory activity from the high-

resolution cryo-EM structure of dirithromycin bound to the E. coli 70S ribosome. *RNA* 26, 715–723 (2020).
13. Stojković, V. *et al.* Assessment of the nucleotide modifications in the high-resolution cryo-electron microscopy structure of the Escherichia coli 50S subunit. *Nucleic Acids Res.* 48, 2723–2732 (2020).
14. Halfon, Y. *et al.* Exit tunnel modulation as resistance mechanism of S. aureus erythromycin resistant mutant. *Sci. Rep.* 9, 11460 (2019).
15. Karplus, P. A. & Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* 34, 60–68 (2015).
16. Frank, J. & Al-Ali, L. Signal-to-noise ratio of electron micrographs obtained by cross correlation. *Nature* 256, 376–379 (1975).
17. Harauz, G. & van Heel, M. Exact filters for general geometry three dimensional reconstruction. *Optik* 73, 146–156 (1986).
18. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745 (2003).
19. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* 20, 205–214 (2012).
20. Subramaniam, S., Earl, L. A., Falconieri, V., Milne, J. L. & Egelman, E. H. Resolution advances in cryo-EM enable application to drug discovery. *Curr. Opin. Struct. Biol.* 41, 194–202 (2016).
21. DiMaio, F., Zhang, J., Chiu, W. & Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein Sci.* 22, 865–868 (2013).
22. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 75, 861–877 (2019).
23. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* 7, (2018).
24. von Loeffelholz, O. *et al.* Focused classification and refinement in high-resolution cryo-EM structural analysis of ribosome complexes. *Curr. Opin. Struct. Biol.* 46, 140–148 (2017).
25. Fischer, N. *et al.* Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* 520, 567–570 (2015).
26. Lill, R., Robertson, J. M. & Wintermeyer, W. Affinities of tRNA binding sites of ribosomes from Escherichia coli. *Biochemistry* 25, 3245–3255 (1986).
27. Arora, S., Bhamidimarri, S. P., Weber, M. H. W. & Varshney, U. Role of the ribosomal P-site elements of m^2G966, m^5C967, and the S9 C-terminal tail in maintenance of the reading frame during translational elongation in Escherichia coli. *J. Bacteriol.* 195, 3524–3530 (2013).
28. Rodnina, M. V., Fischer, N., Maracci, C. & Stark, H. Ribosome dynamics during decoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372, (2017).
29. Held, W. A., Mizushima, S. & Nomura, M. Reconstitution of Escherichia coli 30 S ribosomal subunits from purified molecular components. *J. Biol. Chem.* 248, 5720–5730 (1973).
30. Marzi, S. *et al.* Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell* 130, 1019–1031 (2007).
31. Sashital, D. G. *et al.* A combined quantitative mass spectrometry and electron microscopy

analysis of ribosomal 30S subunit assembly in E. coli. *Elife* 3, (2014).
32. Bubunenko, M., Baker, T. & Court, D. L. Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in Escherichia coli. *J. Bacteriol.* 189, 2844–2853 (2007).
33. Goodall, E. C. A. *et al.* The Essential Genome of Escherichia coli K-12. *MBio* 9, (2018).
34. Shine, J. & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* 254, 34–38 (1975).
35. Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* 578, 425–431 (2020).
36. Stern, S., Powers, T., Changchien, L. M. & Noller, H. F. Interaction of ribosomal proteins S5, S6, S11, S12, S18 and S21 with 16 S rRNA. *J. Mol. Biol.* 201, 683–695 (1988).
37. David, C. L., Keener, J. & Aswad, D. W. Isoaspartate in ribosomal protein S11 of Escherichia coli. *J. Bacteriol.* 181, 2872–2877 (1999).
38. Reissner, K. J. & Aswad, D. W. Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals? *Cell. Mol. Life Sci.* 60, 1281–1295 (2003).
39. Marques, M. A., Purdy, M. D. & Yeager, M. CryoEM maps are full of potential. *Curr. Opin. Struct. Biol.* 58, 214–223 (2019).
40. Nürenberg-Goloub, E. *et al.* Molecular analysis of the ribosome recycling factor ABCE1 bound to the 30S post-splitting complex. *EMBO J.* 39, e103788 (2020).
41. Tesina, P. *et al.* Molecular mechanism of translational stalling by inhibitory codon combinations and poly(A) tracts. *EMBO J.* 39, e103365 (2020).
42. Okamoto, S. *et al.* Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol. Microbiol.* 63, 1096–1106 (2007).
43. Mikheil, D. M., Shippy, D. C., Eakley, N. M., Okwumabua, O. E. & Fadl, A. A. Deletion of gene encoding methyltransferase (gidB) confers high-level antimicrobial resistance in Salmonella. *J. Antibiot.* 65, 185–192 (2012).
44. Kowalak, J. A. & Walsh, K. A. Beta-methylthio-aspartic acid: identification of a novel posttranslational modification in ribosomal protein S12 from Escherichia coli. *Protein Sci.* 5, 1625–1632 (1996).
45. Anton, B. P. *et al.* RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1826–1831 (2008).
46. Benítez-Páez, A., Cárdenas-Brito, S., Corredor, M., Villarroya, M. & Armengod, M. E. Impairing methylations at ribosome RNA, a point mutation-dependent strategy for aminoglycoside resistance: the rsmG case. *Biomedica* 34 Suppl 1, 41–49 (2014).
47. Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* 111, 721–732 (2002).
48. Zaher, H. S. & Green, R. Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection. *Mol. Cell* 39, 110–120 (2010).
49. Connolly, K., Rife, J. P. & Culver, G. Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Mol. Microbiol.* 70, 1062–1075 (2008).
50. Sharma, H. & Anand, B. Ribosome assembly defects subvert initiation Factor3 mediated scrutiny of bona fide start signal. *Nucleic Acids Res.* 47, 11368–11386 (2019).
51. Ochi, K. *et al.* Inactivation of KsgA, a 16S rRNA methyltransferase, causes vigorous emergence of mutants with high-level kasugamycin resistance. *Antimicrob. Agents*

Chemother. 53, 193–201 (2009).
52. Zou, J. *et al.* Studies on Aminoglycoside Susceptibility Identify a Novel Function of KsgA To Secure Translational Fidelity during Antibiotic Stress. *Antimicrob. Agents Chemother.* 62, (2018).
53. Phunpruch, S. *et al.* A role for 16S rRNA dimethyltransferase (ksgA) in intrinsic clarithromycin resistance in Mycobacterium tuberculosis. *Int. J. Antimicrob. Agents* 41, 548–551 (2013).
54. O'Farrell, H. C. & Rife, J. P. Staphylococcus aureus and Escherichia coli have disparate dependences on KsgA for growth and ribosome biogenesis. *BMC Microbiol.* 12, 244 (2012).
55. O'Farrell, H. C., Musayev, F. N., Scarsdale, J. N. & Rife, J. P. Control of substrate specificity by a single active site residue of the KsgA methyltransferase. *Biochemistry* 51, 466–474 (2012).
56. Sati, G. C. *et al.* Modification at the 2'-Position of the 4,5-Series of 2-Deoxystreptamine Aminoglycoside Antibiotics To Resist Aminoglycoside Modifying Enzymes and Increase Ribosomal Target Selectivity. *ACS Infect Dis* 5, 1718–1730 (2019).
57. Vicens, Q. & Westhof, E. Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* 9, 647–658 (2001).
58. Kurata, S. *et al.* Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U.G wobble pairing during decoding. *J. Biol. Chem.* 283, 18801–18811 (2008).
59. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942 (2006).
60. Wang, L. *et al.* Allosteric control of the ribosome by small-molecule antibiotics. *Nat. Struct. Mol. Biol.* 19, 957–963 (2012).
61. Wasserman, M. R. *et al.* Chemically related 4,5-linked aminoglycoside antibiotics drive subunit rotation in opposite directions. *Nat. Commun.* 6, 7896 (2015).
62. Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 332, 981–984 (2011).
63. Komoda, T. *et al.* The A-site finger in 23 S rRNA acts as a functional attenuator for translocation. *J. Biol. Chem.* 281, 32303–32309 (2006).
64. Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A. & Steitz, T. A. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol. Cell* 20, 437–448 (2005).
65. Ge, W. *et al.* Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat. Chem. Biol.* 8, 960–962 (2012).
66. Brosius, J. & Chen, R. The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett.* 68, 105–109 (1976).
67. Dai, Y. *et al.* Elucidating Escherichia coli Proteoform Families Using Intact-Mass Proteomics and a Global PTM Discovery Database. *J. Proteome Res.* 16, 4156–4165 (2017).
68. Burkhart, B. J., Schwalen, C. J., Mann, G., Naismith, J. H. & Mitchell, D. A. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem. Rev.* 117, 5389–5456 (2017).
69. Sawers, G. & Suppmann, B. Anaerobic induction of pyruvate formate-lyase gene expression is mediated by the ArcA and FNR proteins. *J. Bacteriol.* 174, 3474–3478 (1992).
70. Mahanta, N., Szantai-Kis, D. M., Petersson, E. J. & Mitchell, D. A. Biosynthesis and

Chemical Applications of Thioamides. *ACS Chem. Biol.* 14, 142–163 (2019).
71. Strader, M. B. *et al.* A proteomic and transcriptomic approach reveals new insight into beta-methylthiolation of Escherichia coli ribosomal protein S12. *Mol. Cell. Proteomics* 10, M110.005199 (2011).
72. Nichols, R. J. *et al.* Phenotypic landscape of a bacterial cell. *Cell* 144, 143–156 (2011).
73. Sawers, R. G. Evidence for novel processing of the anaerobically inducible dicistronic focA-pfl mRNA transcript in Escherichia coli. *Mol. Microbiol.* 58, 1441–1453 (2005).
74. Stephenson, R. C. & Clarke, S. Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J. Biol. Chem.* 264, 6164–6170 (1989).
75. Robinson, N. E. & Robinson, A. B. Molecular clocks. *Proc. Natl. Acad. Sci. U. S. A.* 98, 944–949 (2001).
76. Acedo, J. Z. *et al.* O-Methyltransferase-Mediated Incorporation of a β-Amino Acid in Lanthipeptides. *J. Am. Chem. Soc.* 141, 16790–16801 (2019).
77. Zhao, J., Li, J., Xu, S. & Feng, P. Emerging Roles of Protein Deamidation in Innate Immune Signaling. *J. Virol.* 90, 4262–4268 (2016).
78. Hobbie, S. N. *et al.* Binding of neomycin-class aminoglycoside antibiotics to mutant ribosomes with alterations in the A site of 16S rRNA. *Antimicrob. Agents Chemother.* 50, 1489–1496 (2006).
79. Sati, G. C. *et al.* N6', N6''', and O4' Modifications to Neomycin Affect Ribosomal Selectivity without Compromising Antibacterial Activity. *ACS Infect Dis* 3, 368–377 (2017).
80. Murray, L. J. W., Arendall, W. B., 3rd, Richardson, D. C. & Richardson, J. S. RNA backbone is rotameric. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13904–13909 (2003).
81. Richardson, J. S., Williams, C. J., Videau, L. L., Chen, V. B. & Richardson, D. C. Assessment of detailed conformations suggests strategies for improving cryoEM models: Helix at lower resolution, ensembles, pre-refinement fixups, and validation at multi-residue length scale. *J. Struct. Biol.* 204, 301–312 (2018).
82. Richardson, J. S. *et al.* RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* 14, 465–481 (2008).
83. Hattne, J. *et al.* Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* 26, 759–766.e4 (2018).
84. Bury, C. S. *et al.* RNA protects a nucleoprotein complex against radiation damage. *Acta Crystallogr D Struct Biol* 72, 648–657 (2016).
85. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *bioRxiv* 2020.05.22.110189 (2020) doi:10.1101/2020.05.22.110189.
86. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *bioRxiv* 845032 (2020) doi:10.1101/845032.
87. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V., Adams, P. D. & Read, R. J. Density modification of cryo-EM maps. *bioRxiv* 2020.05.13.091884 (2020) doi:10.1101/2020.05.13.091884.
88. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* 7, (2018).
89. Punjani, A. & Fleet, D. J. 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *bioRxiv*

2020.04.08.032466 (2020) doi:10.1101/2020.04.08.032466.
90. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).
91. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* 29, 1069–1078 (2020).
92. Zhang, K., Pintilie, G. D., Li, S., Schmid, M. F. & Chiu, W. Resolving Individual-Atom of Protein Complex using Commonly Available 300-kV Cryo-electron Microscopes. doi:10.1101/2020.08.19.256909.
93. Watkins, A. M. & Das, R. FARFAR2: Improved de novo Rosetta prediction of complex global RNA folds. *bioRxiv* 764449 (2019) doi:10.1101/764449.
94. Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* 13, 387–388 (2016).
95. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* 16, 471–477 (2019).
96. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332 (2017).
97. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221 (2015).
98. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14, 290–296 (2017).
99. Travin, D. Y. *et al.* Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition. *Nat. Commun.* 10, 4563 (2019).
100. Ad, O. *et al.* Translation of Diverse Aramid- and 1,3-Dicarbonyl-peptides by Wild Type Ribosomes in Vitro. *ACS Cent Sci* 5, 1289–1294 (2019).
101. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* 6, 779–790 (2011).
102. Cheng, A. *et al.* High resolution single particle cryo-electron microscopy using beam-image shift. *J. Struct. Biol.* 204, 270–275 (2018).
103. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530 (2012).
104. Polikanov, Y. S., Steitz, T. A. & Innis, C. A. A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat. Struct. Mol. Biol.* 21, 787–793 (2014).
105. Russo, C. J. & Henderson, R. Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy* 187, 26–33 (2018).
106. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402 (1997).
107. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* vol. 7 (2016).
108. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868 (2017).
109. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* 40, D290–301 (2012).

110. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* 14, 755–763 (1998).
111. Katoh, K. & Standley, D. M. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32, 1933–1942 (2016).
112. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973 (2009).
113. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274 (2015).
114. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589 (2017).
115. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* vol. 35 518–522 (2018).
116. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259 (2019).
117. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190 (2004).
118. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25, 1335–1337 (2009).
119. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* 523, 208–211 (2015).
120. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152 (2012).
121. Altschul, S. F. *et al.* Lipman. DJ 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389 (1997).
122. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–62 (2016).
123. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* 14, 513–520 (2017).
124. Deutsch, E. W. *et al.* Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics Clin. Appl.* 9, 745–754 (2015).
125. Pino, L. K. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* 39, 229–244 (2020).
126. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* 27, 14–25 (2018).

Chapter 7: Conclusion

Conclusion and future directions

The ribosome structural biology projects contained in Chapters 2-6 and Appendix 1 have progressed in step with cryo-EM's "resolution revolution." They reflect the leaps the field has made while making varied contributions to ribosome science, including ribosomal chemistry, antibiotics, phylogenetic conservation of structural motifs, and assembly defects associated with mutating the ribosome.

Chapter 4 has brought attention to the importance of considering structure and ribosomal assembly in addition to functional tests when mutating the ribosome, especially when *in vitro* activity is important. There is interest in pushing the ribosome to take on new substrates for a variety of applications, but the further we stray from native chemistry, the more important this is likely to be. Better-assembled mutant ribosomes, when they become available, will allow us to obtain structures that will be key for deepening our understanding of how they are able to perform new tasks. Advances towards broadening the ribosome's ability to create a range of sequence-defined polymers then have the potential to be impactful in areas ranging from medicine to materials science.

In Chapter 6, the resolution achieved has led to a much-improved model of the *E. coli* ribosome over previous crystal structures, which has also shone a light on what new tools may be useful for the community in terms of model building. (Relatedly, in Appendix 1, unusual protein backbone modifications were challenging to accommodate with programs typically used for modeling in structural biology.) Because of the highly solvated nature of folded RNAs, the 2 Å ribosome has already led to the implementation of automated water building into cryo-EM map density in PHENIX[13,14]. However, additional tools to aid in modeling with maps that have highly heterogeneous local resolution, as well as better automated recognition of backbone oddities such as the one that led to the discovery of the isoaspartate in uS11, could facilitate future discoveries. RNA rotamers also stand out as an obvious feature that could be further explored to understand more about basic RNA structure, with the high-resolution maps revealing interesting local variations in connectivity and clarity of features for the RNA backbone vs. the bases. Finally, new structural elements visible in the map bring up multiple questions relating to phylogenetics and mechanisms for the introduction of chemical modifications.

In regards to cryo-EM, an exclusive focus on the ribosome and the collection of over a dozen ribosome datasets in the past six years has been illuminating for what remains to hinder reliable high resolution once the complication of biochemical preparation (or significant variations in sample grid preparation technique) is largely taken out of the picture. Hardware advances continue to push resolution boundaries[1,2], but the increasing variety of results using more widespread technology, including the 2 Å *E. coli* ribosome described here, raises the question of how we might achieve high resolution more routinely without new accessories. For example, recent publications indicate that 200 kV "screening" microscopes are capable of producing high-quality datasets, comparable to what is achievable with 300 kV microscopes more commonly used for high resolution[3,4]. Another example is described in Chapter 5, where I have processed a 70S ribosome dataset to 3.0 Å resolution from < 12,000 particles. Learning from successes like these as they increase in abundance may hold clues for what set them apart.

Reproducibility at the level of sample preparation is one major challenge to tackle. Particle quality as determined by this is likely to require many solutions given that the properties of different samples create unique demands, but solutions to certain problems like ice quality and protection from the air-water interface are potentially solvable with broader means. Current strategies under investigation include graphene or graphene oxide as a sample support and automated sample deposition with self-wicking grids[5–7]. In principle, streamlining of sample preparation is one way to ensure that one is likely to (almost) always be collecting high-quality data in the first place, rather than taking on the degree of uncertainty that we do now.

The way the microscope session itself is conducted is another important stage to consider, which currently requires expert intervention for careful screening. Chapter 5 has described an attempt to revive the use of focal pairs, which is less-so geared towards screening but nevertheless intended to improve data quality intrinsic to the images. However, in contrast to more successful use of focal series in TEM for materials science[8], the evidence was against the use of focal pairs here. For quality judgement during the session, on-the-fly processing is being pursued[9,10] and is useful inasmuch as it takes advantage of the time consumed by data collection to perform initial necessary tasks like motion correction and CTF estimation. Progress is being made towards fine-tuning these processes well enough to begin classifying particles and determining preliminary structures as the session proceeds. The current utility of real-time metrics, however, is mostly for detecting obviously bad imaging areas as opposed to identifying ones with well-preserved high-resolution information, and even this requires somewhat active monitoring by the user. Real-time processing of diffraction spots from streptavidin grids could offer a more reliable way to assess particle quality on-the-fly in light of the analysis from Chapter 3, but this has yet to be implemented and is of course a specialized solution.

It would be advantageous in the more general case to be able to queue good areas for collection to begin with, rather than having to course-correct once collection has started. My general observation is that for my best datasets, there is often a significant amount of variability in quality from square to square within a grid. Analysis of streptavidin crystal quality in Chapter 3 has pointed to this idea. Some preliminary analysis of one dataset in Chapter 5 has also supported this observation and guided the screening process for my most recent structures, including the 2 Å ribosome in Chapter 6. Sometimes features like an apparent meniscus around the perimeter of a grid square are used to gauge whether or not the ice thickness is likely to be good, in the absence of rigorous measurements. Recent cryo-SEM work has shown that in fact, this meniscus can be on the back side of the grid and does not necessarily reflect ice thickness on the front side[11], which is likely a key variable for high resolution. Personal experience also supports the idea that a grid square with sharply defined edges in transmission images (i.e. no visible meniscus) may well contain particles of particularly high quality, although attention and extra time is required to be sure these squares are not dry. Future work could focus on how to most efficiently recognize features signaling good imaging areas, perhaps aided by automatic measurement of ice thickness[12], or by some other means.

REFERENCES

1. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
2. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* (2020).
3. Herzik, M. A., Wu, M. & Lander, G. C. High-resolution structure determination of sub-100

kDa complexes using conventional cryo-EM. *Nat. Commun.* (2019). doi:10.1038/s41467-019-08991-8
4. Wu, M., Lander, G. C. & Herzik, M. A. Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *J. Struct. Biol. X* (2020). doi:10.1016/j.yjsbx.2020.100020
5. Wang, F. *et al.* General and robust covalently linked graphene oxide affinity grids for high-resolution cryo-EM. *Proc. Natl. Acad. Sci. U. S. A.* (2020). doi:10.1073/pnas.2009707117
6. Dandey, V. P. *et al.* Spotiton: New features and applications. *J. Struct. Biol.* (2018). doi:10.1016/j.jsb.2018.01.002
7. Wei, H. *et al.* Optimizing 'Self-Wicking' Nanowire Grids HHS Public Access. *J Struct Biol* (2018).
8. O'Keefe, M. A. *et al.* Sub-Ångstrom high-resolution transmission electron microscopy at 300keV. *Ultramicroscopy* **89**, 215–241 (2001).
9. Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat. Protoc.* (2019). doi:10.1038/s41596-018-0084-8
10. Stabrin, M. *et al.* TranSPHIRE: Automated and feedback-optimized on-the-fly processing for cryo-EM. *bioRxiv* 2020.06.16.155275 (2020). doi:10.1101/2020.06.16.155275
11. Armstrong, M. *et al.* Microscale Fluid Behavior during Cryo-EM Sample Blotting. *Biophys. J.* (2020). doi:10.1016/j.bpj.2019.12.017
12. Rice, W. J. *et al.* Routine determination of ice thickness for cryo-EM grids. *J. Struct. Biol.* (2018). doi:10.1016/j.jsb.2018.06.007
13. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909052925
14. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* (2019). doi:10.1107/S2059798319011471

Appendix 1: Selected passages from "Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition"

Reproduced in part with permission from: Travin DY, Watson ZL, Metelev M, Ward FR, Osterman IA, Khven IM, Khabibullina NF, Serebryakova M, Mergaert P, Polikanov YS, Cate JHD, Severinov K. Nat Commun. 2019;10(1):4563. Published 2019 Oct 8. doi:10.1038/s41467-019-12589-5

Figure numbering is kept as in the published text.

RESULTS

Phazolicin obstructs the nascent peptide exit tunnel (NPET) of the ribosome

Protein synthesis inhibitors can interfere with the activity of the ribosome or any of the various other enzymes associated with protein production (translation factors, aminoacyl-tRNA synthetases, etc.). Following the analogy with Klebsazolicin[1] (KLB), we proposed that Phazolicin (PHZ) acts upon the bacterial ribosome. To determine the exact mechanism of translation inhibition by PHZ we attempted to crystallize the 70S ribosome from *Thermus thermophilus* (*Tth*) in complex with PHZ, messenger RNA (mRNA), and transfer RNAs (tRNAs), as previously done with KLB[1], and several other translation inhibitors[2–9]. Despite numerous attempts, we were not able to observe any positive electron density peaks corresponding to the ribosome-bound PHZ anywhere within the *Tth* 70S ribosome. Moreover, in vitro translation inhibition experiments using *E. coli* (*Eco*) 70S or hybrid 70S (*Eco* 30S subunit + *Tth* 50S subunit) demonstrated that PHZ does not target *Tth* ribosome (**Supplementary Figure 4**). However, the same data also suggested that PHZ does target the large ribosomal subunit of the *E. coli* ribosome.

Since the crystallization-based approach was unsuccessful, we switched to cryo-EM to determine the structure of PHZ bound to the sensitive *Eco* 70S ribosome. The obtained charge density map, characterized by the overall resolution of 2.87 Å according to the "gold-standard" Fourier shell correlation (FSC) method (**Supplementary Figure 5A**), revealed PHZ bound to the ribosome (**Figure 3A**). The high resolution and excellent quality of the map in the PHZ-binding site (**Supplementary Figure 5B**), allowed us not only to fit its atomic model (residues 2–23; **Figure 3B**) but also to confirm the positions of seven azoles in the structure of the modified peptide proposed according to the MS–MS analysis. While the results of tandem mass spectrometry indicate another oxazole present at the position 24, we left Ser23 unmodified in the model since Ser24 was not visible in the map due to its poor quality in that region (**Figure 3A**).

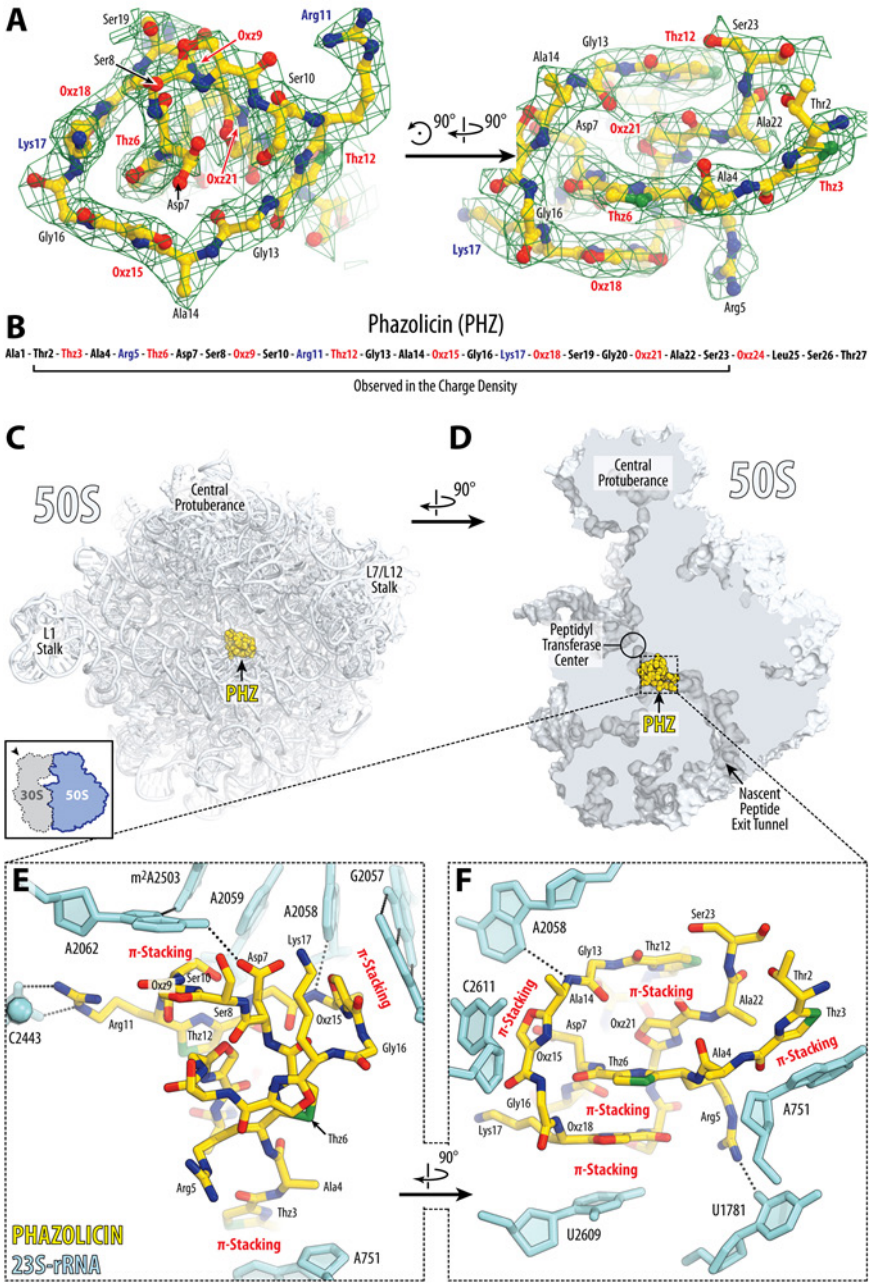


Figure 3. The structure of PHZ in complex with the bacterial ribosome. (**A**) Cryo-EM map of PHZ in complex with the *E. coli* 70S ribosome (green mesh). The fitted model of the compound is displayed in its respective charge density viewed from two different perspectives. The map is contoured at 2.5σ. Carbon atoms are colored yellow, nitrogen atoms are blue, oxygen atoms are red, and sulfur atoms are green. (**B**) Schematic diagram showing that only residues 2–23 of the ribosome-bound PHZ molecule are visible in the cryo-EM map. (**C**, **D**) Overview of the PHZ-binding site (yellow) on the *E. coli* large ribosomal subunit (light blue) viewed from two different perspectives. In c, the 50S subunit is viewed from the inter-subunit interface (30S subunit is removed for clarity) as indicated by the inset. The view in **D** is from the cytoplasm onto the A site. (**E**, **F**) Close-up views of the PHZ-binding site in the ribosome exit tunnel. *E. coli* numbering of the nucleotides in the 23S rRNA is used. Potential H-bond interactions are indicated with dotted lines.

Similar to KLB, linear modified PHZ peptide forms a compact globule that binds in the upper part of the NPET of the 50S ribosomal subunit where it extensively interacts with the 23S

rRNA and ribosomal proteins (**Figure 3C**, **D**). In comparison to antibiotics of smaller molecular weight, such as erythromycin (ERY), which binds at the same location but only partially occludes the NPET (**Supplementary Figure 6A, B**), binding of both KLB and PHZ results in nearly complete obstruction of the NPET (**Supplementary Figure 6C, D**). A prominent feature of the ribosome-bound PHZ molecule is its complex intramolecular interactions that involves both face-to-face and edge-to-face π–π stacking of Thz12, Oxz21, Thz6, and Oxz18, along with the nucleobase U2609 from the 23S rRNA (**Figure 3F**; **Supplementary Figure 7D**). Moreover, we observed two major types of interactions between the PHZ and the nucleotides of the 23S rRNA. First, three azoles (Thz3, Oxz15, and Oxz18) are involved in π–π stacking with nucleotides A751, C2611, and U2609, respectively (**Figure 3E**, **F**). Second, two positively charged side chains of Arg5 and Arg11, as well as hydrophilic side chains of Asp7 and Ser8, form additional stabilizing hydrogen bonds (H-bonds) with nucleobases and backbone phosphate groups of the 23S RNA (**Figure 3E**, **F**; **Supplementary Figure 7A–C**). The presence of intramolecular π–π stacking and positively charged side chains, as well as the absence of the N-terminal amidine cycle, make the mode of PHZ binding significantly different from that of KLB. Unlike KLB, the N-terminal part of PHZ is not critical for binding to the ribosome, which is further supported by the observation that naturally occurring longer forms with an alternative leader cleavage site (A-PHZ and TA-PHZ) demonstrate the same level of in vitro translation inhibition activity as PHZ (**Supplementary Figure 8**).

Interestingly, the PHZ-binding site overlaps with those of several other classes of antibiotics, such as macrolides and type B streptogramins. Comparison of the structures of ribosome-bound PHZ and KLB reveals that PHZ binds further away than KLB from the peptidyl transferase center (PTC) of the ribosome. We superimposed our new structure of the ribosome-bound PHZ (**Figure 4A**) or the previously published KLB (**Figure 4B**) with the cryo-EM structure of the 70S ribosome containing ErmBL nascent peptide chain connected to the P-site tRNA[10]. Apparently, there is more space between the ribosome-bound drug and the PTC in the case of PHZ in comparison to KLB. This suggests that PHZ should, in principle, allow for a few more amino acids to be incorporated into the nascent polypeptide chain before it encounters the NPET-bound PHZ molecule and translation stalls. To test this hypothesis, we performed a toe-printing (primer-extension inhibition) assay, which allows identification of drug-induced ribosome stalling on a given mRNA template with a single-nucleotide precision[11]. Because PHZ causes strong induction (stalling of the ribosomes) on Dualrep2 reporter used in our in vivo bioactivity test (**Figure 2A**), we have chosen the corresponding mRNA (*trpL*-2Ala) as a template for our toe-printing assay. Addition of PHZ to the PURExpress cell-free transcription–translation system programmed with *trpL*-2Ala mRNA resulted in dose-dependent ribosome stalling at the seventh codon of mRNA (**Figure 4C**, lanes 4–7), while the addition of KLB induced major stalling at the third and the fifth codons (**Figure 4C**, lane 3) corroborating our hypothesis. Moreover, these data indicate that PHZ is an inhibitor of the elongation step during protein synthesis, as is KLB. In the absence of any inhibitors in our toe-printing reactions, ribosomes stall at codons 9–10 due to a specific secondary structure in the mRNA (**Figure 4C**, lane 8).

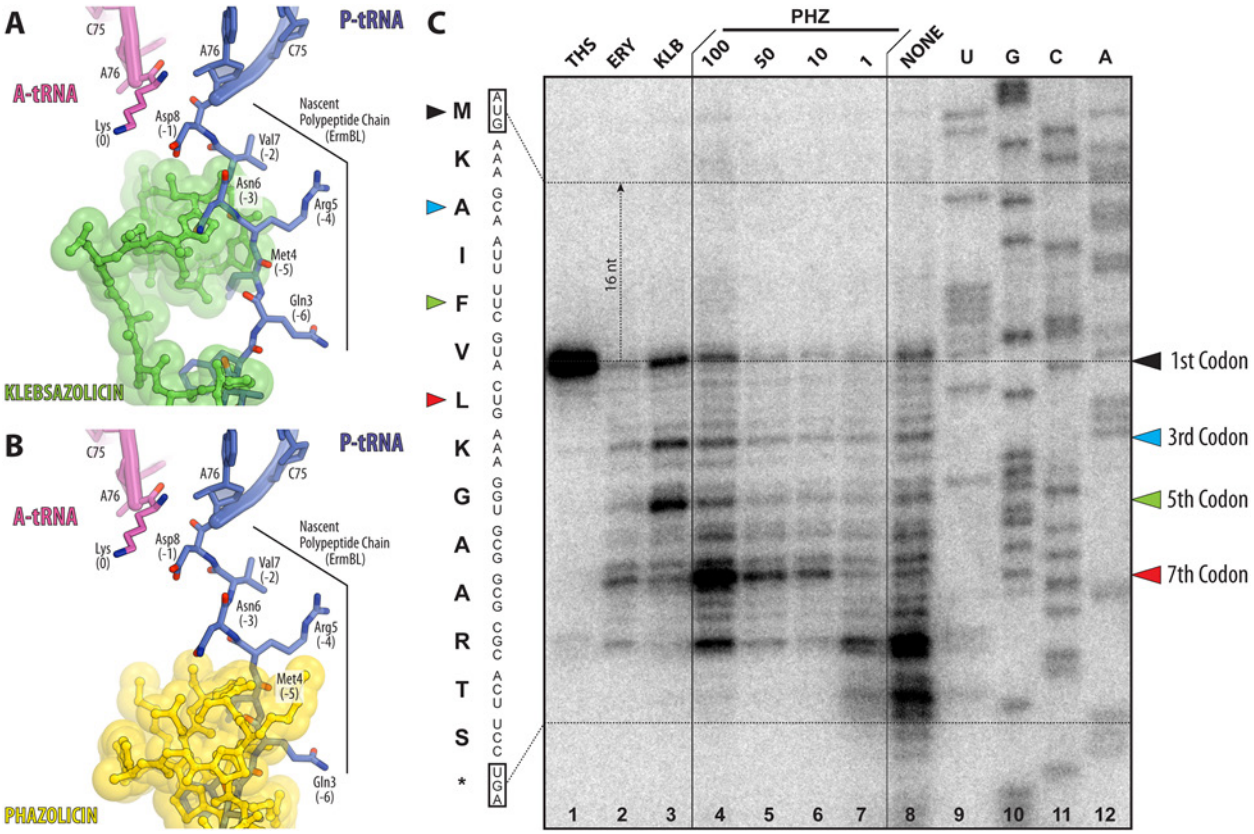


Figure 4. PHZ inhibits the elongation step during protein synthesis. (**A**, **B**) In silico modeling of the nascent polypeptide chain in the ribosome exit tunnel in the presence of klebsazolicin (**A**, PDB entry 5W4K[1]) or phazolicin (**B**, current structure) using the cryo-EM structure of the 70S ribosome with ErmBL peptide bound to the P-site tRNA (PDB entry 5JTE[10]). (**C**) Ribosome stalling by PHZ on *trpL*-2Ala mRNA in comparison with other translation inhibitors (thiostrepton, THS; erythromycin, ERY; klebsazolicin, KLB), as revealed by reverse-transcription primer-extension inhibition (toe-printing) assay in a cell-free translation system. *trpL*-2Ala mRNA nucleotide sequence and the corresponding amino acid sequence are shown on the left. Colored triangles show ribosome stalling at 1st, 3rd, 5th, and 7th codons. Note that owing to the large size of the ribosome, the reverse transcriptase stops at the nucleotide + 16 relative to the codon located in the P-site.

Species-specificity of PHZ action

The PHZ-binding site is adjacent to the NPET constriction point—the narrowest part of the NPET formed by the extended loops of the ribosomal proteins uL4 and uL22 (**Figure 5A**, **b**). The N-terminal and C-terminal parts of the PHZ molecule are located between the uL4 and uL22 loops (**Figure 5C**) and occupy more space in this part of the NPET than the C-terminal part of KLB (**Figure 5D**). The amino acid sequences of uL4 and uL22 proteins including their loop regions vary between distinct bacteria (**Figure 5E**). Superpositioning of our *Eco* 70S ribosome structure with bound PHZ with the structure of *Tth* 70S reveals that residues His69 of the *Tth* protein uL4 sterically clashes with the PHZ molecule (**Figure 5C**) rationalizing why PHZ does not bind to or act upon the *Tth* ribosome. Because the same residue in *Eco* ribosome is represented by a less bulky Gly64 (uL4), we hypothesized that the differences in the uL4 loop sequence are responsible for the inability of PHZ to bind in the NPET and, hence, to inhibit the translation by *Tth* ribosome. To test our hypothesis, we expressed a plasmid-borne copy of the *S. meliloti rplD* gene encoding for the uL4 protein carrying the *Tth*-like G68H substitution (*rplD*^{G68H}) and showed that the loop structure of *Tth* uL4 confers resistance to PHZ and allows growth of otherwise sensitive strain on the medium with 8×MIC of the antibiotic (**Figure 5F**). The expression of

rplD^{K65A} mutant, which carries an alanine substitution of the *Eco* Arg61-equivalent (**Figure 5E**) that forms H-bond with the PHZ molecule in our structure (**Figure 5C**), does not confer resistance to PHZ. This is likely due to multiple contacts, which the PHZ molecule forms with the ribosome, and the loss of one H-bond does not affect the overall binding. The same superpositioning of the *Eco* and *Tth* ribosome structures also revealed that residue Arg90 of the *Tth* protein uL22 sterically clashes with the PHZ molecule (**Figure 5C**), while the comparable in size-equivalent residue Lys90 in the *Eco* protein uL22 does not clash with the PHZ and is repositioned to the side relative to the ribosome-bound PHZ to avoid a similar clash. Unlike His69^{Tth}, which is located in a confined pocket and cannot re-adjust its position in response to PHZ binding, residues Arg90^{Tth} and Lys90^{Eco} have more space around and should be able to easily readjust their positions when PHZ is present to avoid a possible clash. Indeed, the substitution of Lys90 in the loop of protein uL22 with a larger Arg residue (*rplV*^{K90R}) does not render any resistance to PHZ (**Figure 5F**). Thus, the fine structure of the PHZ-binding site (formed in part by the loop regions of ribosomal proteins uL4 and uL22) together with the repertoire of peptide transporters involved in the uptake contribute to the specificity of the antibiotic action.

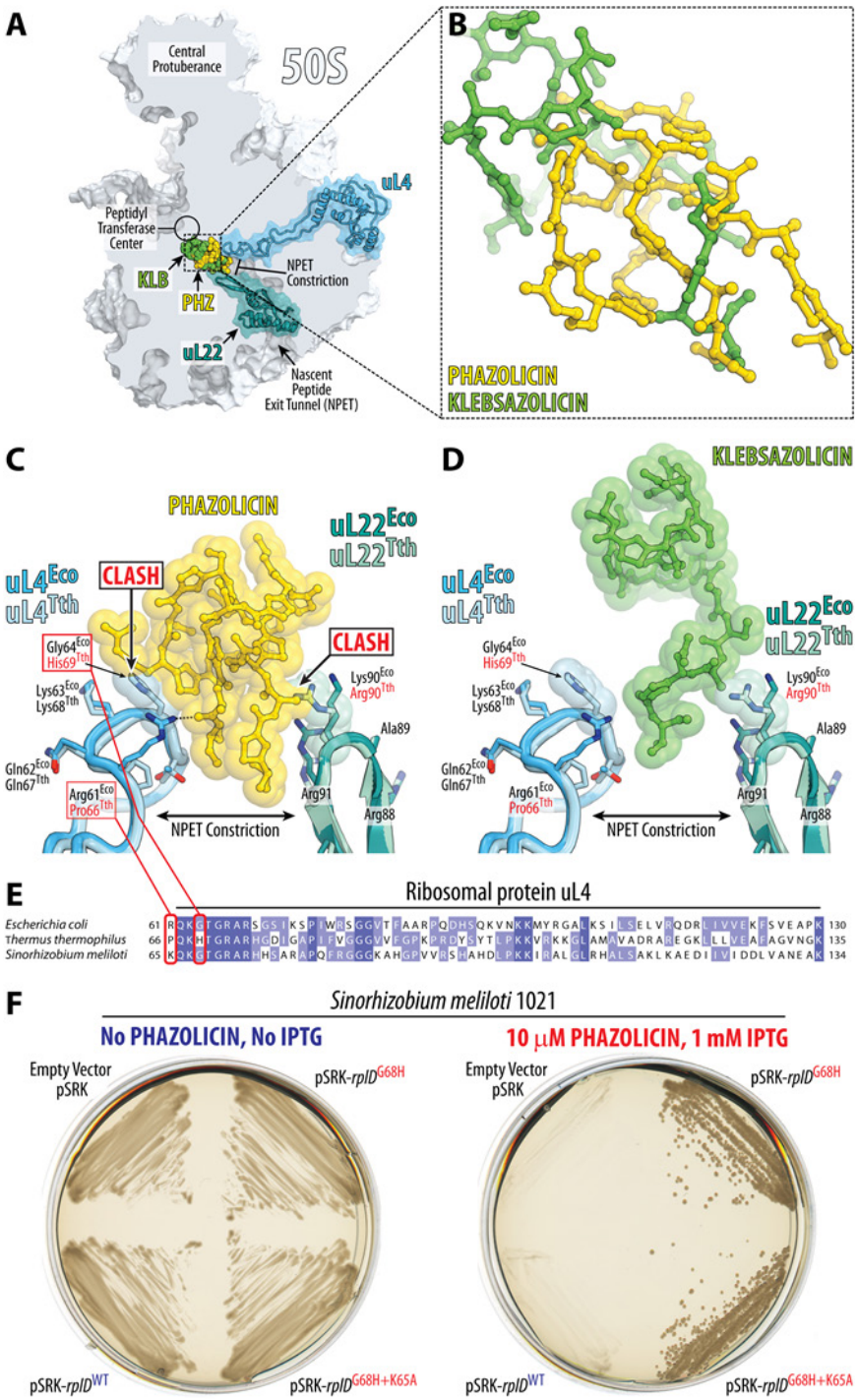


Figure 5. The loop of ribosomal protein uL4 defines species-specificity of PHZ action. (**A**) Comparison of PHZ (yellow) and KLB (green) binding sites in the ribosome exit tunnel. The 50S subunit is shown in light blue. Ribosomal proteins uL4 and uL22 are highlighted in blue and teal, respectively. (**B**) Superposition of the ribosome structures with bound KLB (PDB entry 5W4K[1]) and PHZ based on the alignment of the 23S rRNAs. Ribosome parts are omitted for clarity. Note that KLB binds closer to the PTC than PHZ. (**C**, **D**) Close-up view of PHZ (**C**) and KLB (**D**) interactions with the loops of proteins uL4 and uL22 forming the NPET constriction site. Superposition of the loops of proteins from *E. coli* and *T. thermophilus* is shown. Note that steric clashes occur only with the side chains in the loops of proteins uL4/uL22 from *T. thermophilus*. (**E**) Multiple sequence alignment of the homologous parts of the ribosomal protein uL4 from several bacterial species. Red boxes highlight the amino acid residues of uL4, which either clash with the PHZ (His69 in *Tth*) or interact with it in our structure (Arg61 in *Eco*). (**F**) Growth of *S. meliloti* Sm1021 and

its derivatives expressing plasmid-borne copies of WT or mutant rplD or rplV genes (encoding ribosomal proteins uL4 and uL22, respectively) on agar plates without PHZ (left panel) or with 8 μM PHZ (right panel). Note that only overexpression of *Tth*-like version of uL4 protein in *S. meliloti* confers resistance to PHZ.

DISCUSSION

The newly-discovered antibiotic PHZ represents the second known example of a linear azole-containing peptide with a known mechanism of action and mode of binding to its intracellular target, the ribosome. Although PHZ and the previously described LAP klebsazolicin both inhibit the elongation step of prokaryotic translation through the obstruction of the NPET of the large ribosomal subunit, the mode of binding of PHZ to the bacterial ribosome is significantly different from that of KLB. Unlike KLB, which does not appear to have extensive intramolecular interactions, four out of eight azole rings of PHZ form continuous π–π-stacking system (**Supplementary Figure 5D**), which apparently confers the rigidity and stability of the antibiotic in its binding site. Moreover, PHZ contains several positively charged side chains, which provide additional anchoring on the ribosome through the electrostatic interactions with the negatively charged 23S rRNA (**Figure 3E**), and which were absent in KLB. Finally, PHZ appears to be the first ribosome-targeting antibiotic with a pronounced bacterial species-specific mode of action, which is directly related to its mode of interaction with the molecular target. PHZ can bind to the ribosomes and act upon those species whose ribosomal protein uL4 contains glycine (or non-bulky side chain amino acid residues) in the position equivalent to Gly64 in *E. coli* that is located at the tip of the extended loop of this protein. The residue equivalent to Gly64 of the *E. coli* protein uL4 is not conserved in other bacterial species. The majority of bacteria (including *E. coli*) possess glycine at this position (**Figure 5E**). However, others (e.g., *T. thermophilus*) contain histidine that renders them naturally resistant to this antibiotic due to a steric hindrance with PHZ molecule. There are many ribosome-targeting antibiotics that, as PHZ (see **Table 1**), work on one group of bacteria and do not work on the others, however, this phenomenon is usually not referred to as "species-specificity of action" because it is due to the ability of a drug to get inside the bacterial cell rather than a differential activity on its target.

MATERIALS AND METHODS

Cryo-EM structure determination

Ribosome preparation: *E. coli* strain MRE600 overnight cultures were diluted 1:100 into 3 L of LB. After growth to mid-log phase (OD600 = 0.6), cells were cooled on ice for 3 min, pelleted, and washed with buffer A (20 mM Tris–HCl pH 7.5, 100 mM NH4Cl, 10 mM MgCl2, 0.5 mM EDTA, 2 mM DTT). Cells were resuspended in 100 ml buffer AS (buffer A with 0.15 M sucrose) and lysed with two passes through an Emulsiflex-C3 homogenizer at 18,000 psi. Lysate was clarified by centrifugation for 30 min at 18,000 rpm in a JA-20 rotor (Beckman-Coulter). The supernatant was layered over a sucrose cushion of 24 ml buffer B (buffer A with 500 mM NH4Cl) with 0.5 M sucrose and 17 ml buffer C (20 mM Tris–HCl pH 7.5, 60 mM NH4Cl, 6 mM MgCl2, 0.5 mM EDTA, 2 mM DTT) with 0.7 M sucrose in ti-45 tubes (Beckman-Coulter). Ribosomes were pelleted in a Ti-45 rotor at 27,000 rpm for 16 h at 4 °C. Crude ribosome pellets were resuspended in disassociation buffer (buffer C with 1 mM MgCl2) and clarified via centrifugation at 15,000 × g for 10 min. The cleared supernatant was layered over 15–40% sucrose gradients in disassociation buffer and spun 28,000 rpm for 16 h at 4 °C in a SW-32 rotor (Beckman-Coulter).

Gradients were fractionated with an ISCO fractionation system and 30S + 50S peaks were combined. The combined subunits were concentrated in 15 ml Millipore 100k molecular weight cutoff spin filter and buffer exchanged with reassociation buffer (buffer C with 10 mM MgCl2). Ribosomes were incubated 45 min at 37 °C, layered onto 15–40% sucrose gradients in reassociation buffer, and spun 27,000 rpm for 15 h at 4 °C in an SW-32 rotor. Gradients were fractionated as before, 70S peaks were collected, and ribosomes were concentrated and washed with buffer C in an Amicon stirred cell filtration system using 100k cutoff filters. Ribosomes were stored in aliquots at −80 °C.

Cryo-EM sample preparation: Roughly 4–6 μM PHZ was incubated with 100 nM ribosomes in RC buffer (50 mM Hepes pH 7.5, 150 mM KOAc, 12 mM MgOAc, 7 mM β-mercaptoethanol) for 30 min. at 37 °C. Complexes were deposited in 4-μl aliquots on 300 mesh Quantifoil UltraAuFoil R1.2/1.3 grids that were topped with a layer of amorphous continuous carbon and glow-discharged for 15 s with a Pelco SC-6 sputter coater. After incubation period of about a minute, the excess sample was washed off on drops of RC–LS buffer (RC with 25 mM KOAc). Samples were blotted and plunge-frozen in liquid ethane using an FEI Vitrobot with blot force 6, humidity 100%, and 20 or 4 °C depending on the freezing session.

Data collection: Images were collected in two separate sessions on an FEI Titan Krios Microscope operated at 300 keV with a GIF energy filter and GATAN Summit K2 direct electron detector in super-resolution mode. Images were collected at ×215,000 magnification for a pixel size of 0.56 Å (0.28 Å super-resolution). Automated movie collection was performed with SerialEM[12] over the defocus range −0.6 to −2.0 μm, and Focus software[13] was used to monitor data collection. The total dose was 29.97 $e^{-}/Å^2$ for session 1 and 29.83 $e^{-}/Å^2$ for session 2, each over 30 frames.

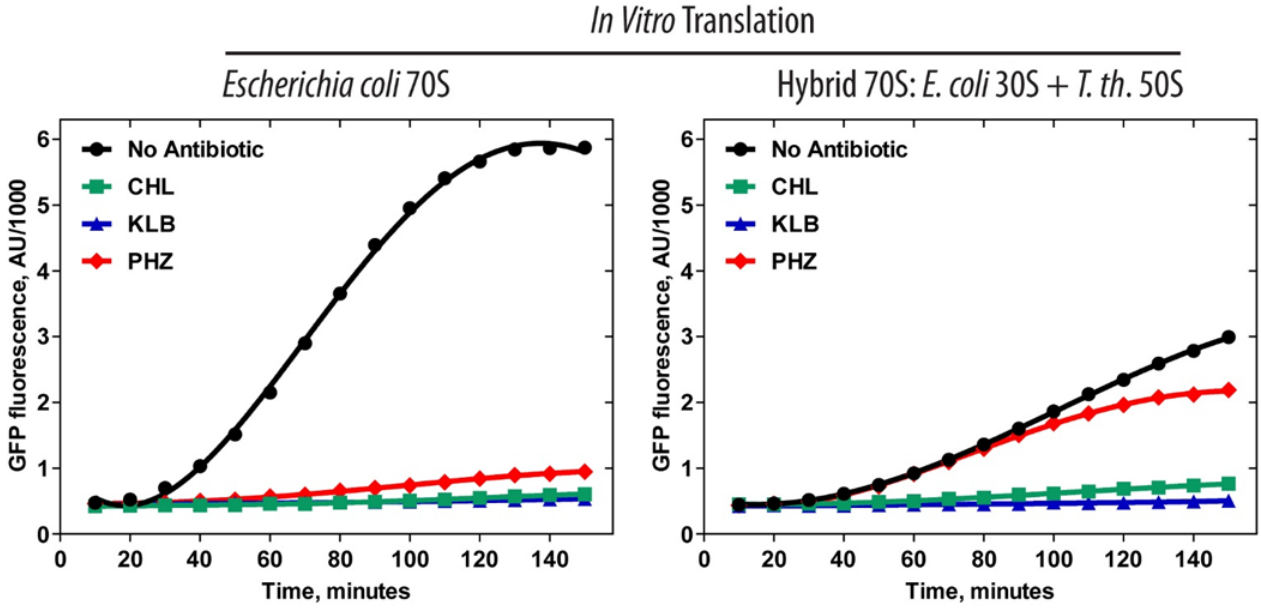
Data processing: Cryo-EM data processing was done using RELION 3.0 software[14] (**Supplementary Figure 9**; **Supplementary Table 2**). The two datasets were processed separately until the final round of map refinement, for which the best particles from both sets were combined. Movies were motion-corrected and dose-weighted using RELION's motion correction algorithm. Contrast transfer function (CTF) parameters were estimated using CTFFind4[15] and micrographs with poor CTF fit, as determined by visual inspection, were sorted out. Particles were picked automatically with the Laplacian-of-Gaussian method and subjected to several rounds of careful 2D class-based cleaning before an initial round of 3D auto-refine. CTF refinement with per-particle defocus and beam tilt correction was performed, and particles were further sorted through 3D classification without alignment. Another round of 3D auto-refine was performed on the best 3D classes with focused refinement on the 50S subunit. This was followed by one round of Bayesian polishing[16] and another round of CTF refinement, which were each followed by another round of 3D auto-refine with a mask around the 50S subunit. Particles from the two sessions were then pooled for the final round of 3D auto-refine. A charge density map was calculated in Chimera[17] as previously described[18]. Post-processing was also performed in RELION to generate a FSC curve. All refinement procedures were run on 2× binned particles because of the small pixel size.

Model building: Assembly #2 from the PDB entry 4YBB[19] was used as a starting ribosome model. Only the 50S subunit was subjected to real-space refinement in PHENIX[20], because charge density corresponding to the 30S subunit was relatively poor after focused refinement on the 50S, where PHZ binds. Ions and water molecules were removed from the starting model prior to real-space refinement. The atomic model for the PHZ was built as follows: first, baton building in COOT[21] was used to trace the main chain; next, the resulting polyalanine chain was mutated to the

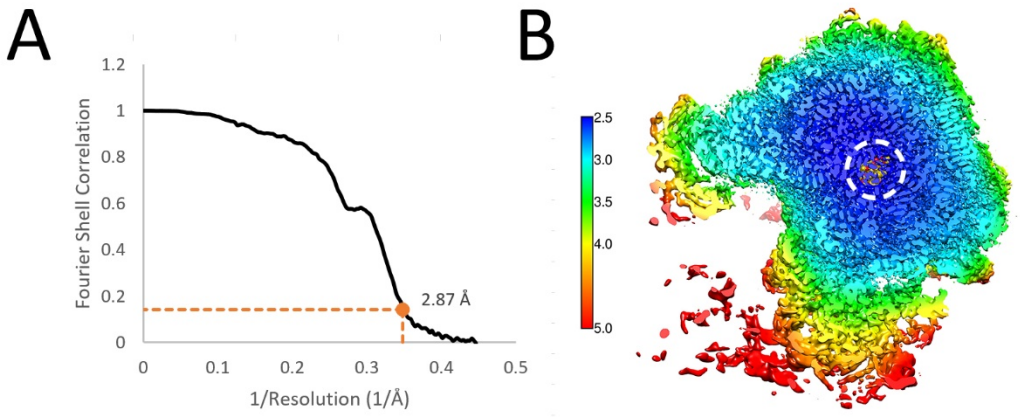
unmodified PHZ sequence after visual search for the characteristic features in the map; next, oxazole and thiazole modifications were built in Avogadro[22] version 1.2.0 (http://avogadro.cc) based on the density. Finally, CIF restraints for the oxazole and thiazole moieties were created based on published structural data[23].
All figures showing atomic models were generated using the PyMol software (www.pymol.org). USCF Chimera was used for EM density maps shown in **Figs. S5** and **S9**.

ACKNOWLEDGMENTS
We thank Dr. Viktor Zgoda for the invaluable help with high-resolution spectra and ESI–MS–MS spectra acquisition and Dr. Esperanza Martinez-Romero who provided the strain Rhizobium sp. Pop5. We thank all members of the K.S., J.H.D.C., and Y.S.P. laboratories for important discussions and critical feedback. We also want to thank Dr. Dmitry Ghilarov and Dr. Svetlana Dubiley for critical reading of the manuscript and valuable suggestions. This work was supported by RSF grant number 19-14-00266 to Svetlana Dubiley, RSF grant number 18-44-04005 to I.A.O. (used for the toe-printing assays and in vivo (pDualrep2 reporter) and in vitro activity tests), by Illinois State startup funds [to Y.S.P.], National Institutes of Health [R01-GM132302 and R21-AI137584 to Y.S.P., R01-GM114454 to J.H.D.C. and F.R.W.] and by the National Science Foundation Graduate Research Fellowship Program under Grant number 1106400 [to Z.L.W.]. Access to MALDI-MS facilities was through the framework of the Lomonosov Moscow State University Development Program PNR 5.13.

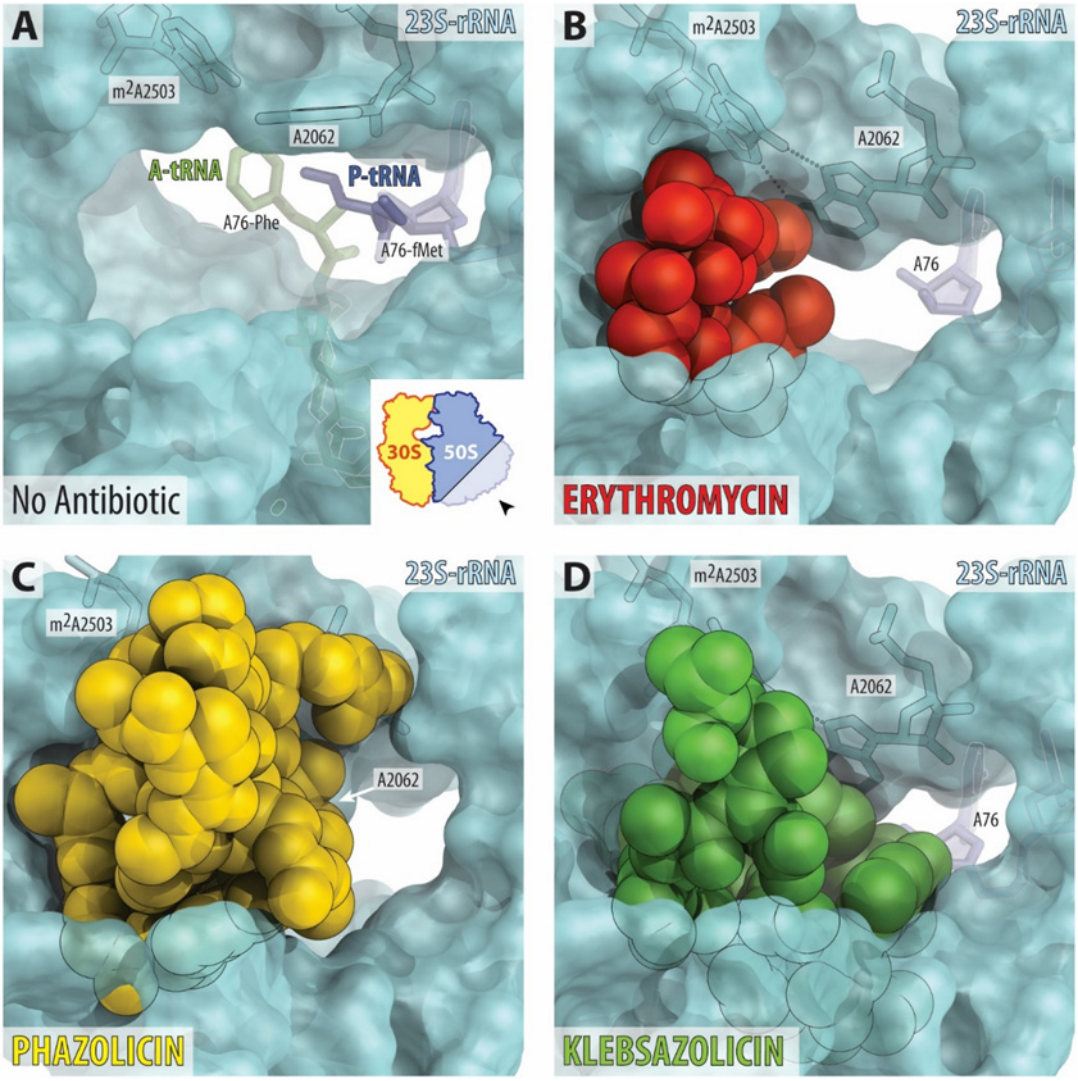
SUPPLEMENTARY MATERIAL



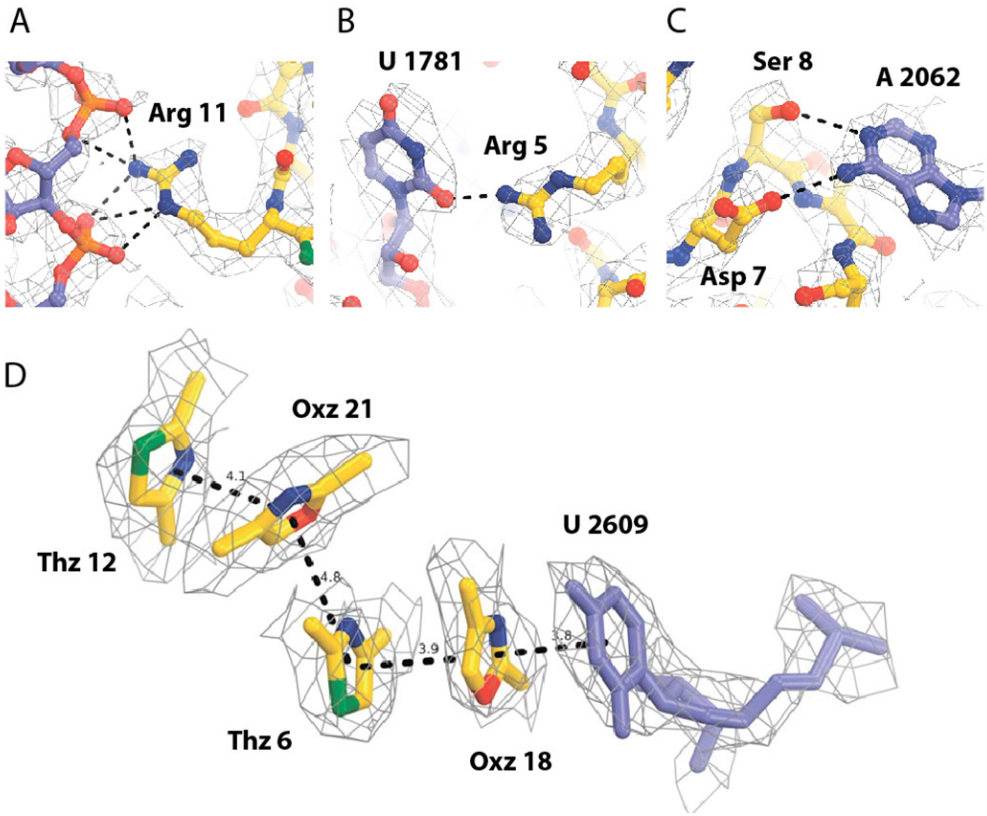
Supplementary Figure 4. Effects of phazolicin (PHZ), chloramphenicol (CHL), and klebsazolicin (KLB) on in vitro synthesis of superfolder green fluorescent protein (sfGFP) using PURExpress system supplied with either the *E. coli* 70S WT ribosomes or hybrid ribosomes (*E. coli* 30S + *T. thermophilus* 50S). CHL and KLB were used as positive controls. Final concentrations for all antibiotics were 50 μM. Note, that the overall rate of protein synthesis by the hybrid ribosome is ~2-fold lower than that of the WT 70S (black curves).



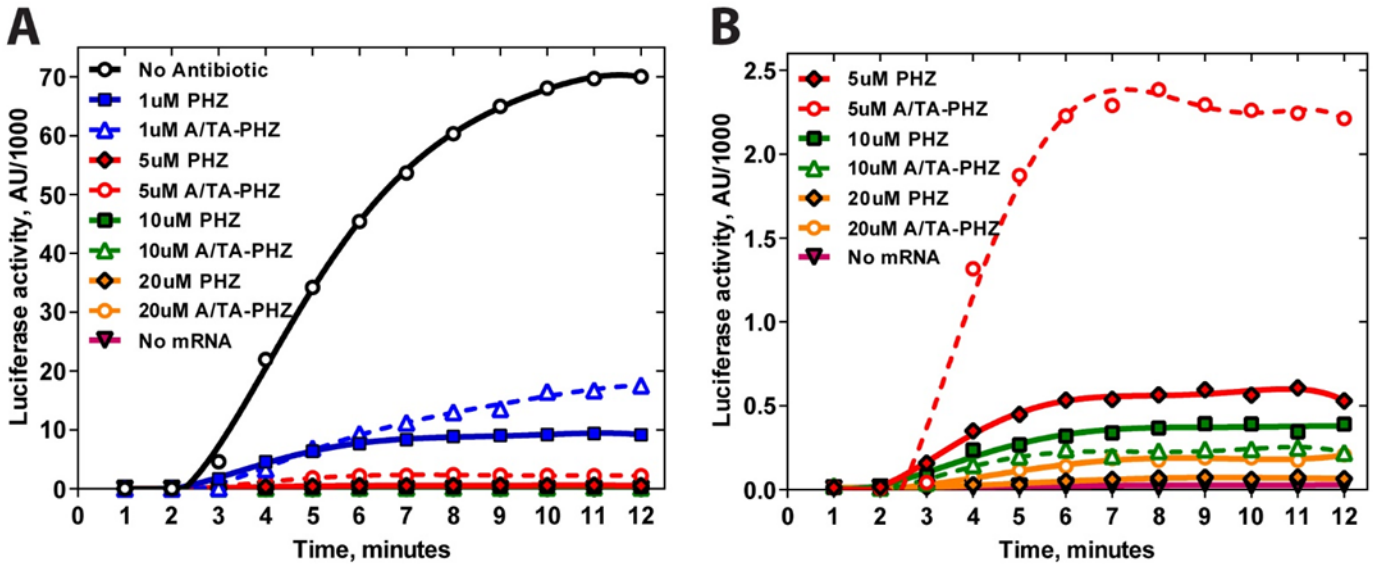
Supplementary Figure 5. Fourier shell correlation (FSC) curve for the 70S-PHZ cryo-EM structure. (**A**) Based on the "gold-standard FSC" cutoff value 0.143, the overall resolution for the large ribosomal subunit in the 70S ribosome is 2.87 Å. (**B**) Slice through the 50S subunit showing local resolution limits. Note that the region of PHZ binding (indicated with the dashed white circle) is characterized by the highest local resolution. Color scale bar on the left is in Ångstroms. 30S ribosomal subunit is not shown and was not included in the final model because after the focused refinement on the 50S subunit the local resolution and connectivity of the map for the 30S subunit were poor.



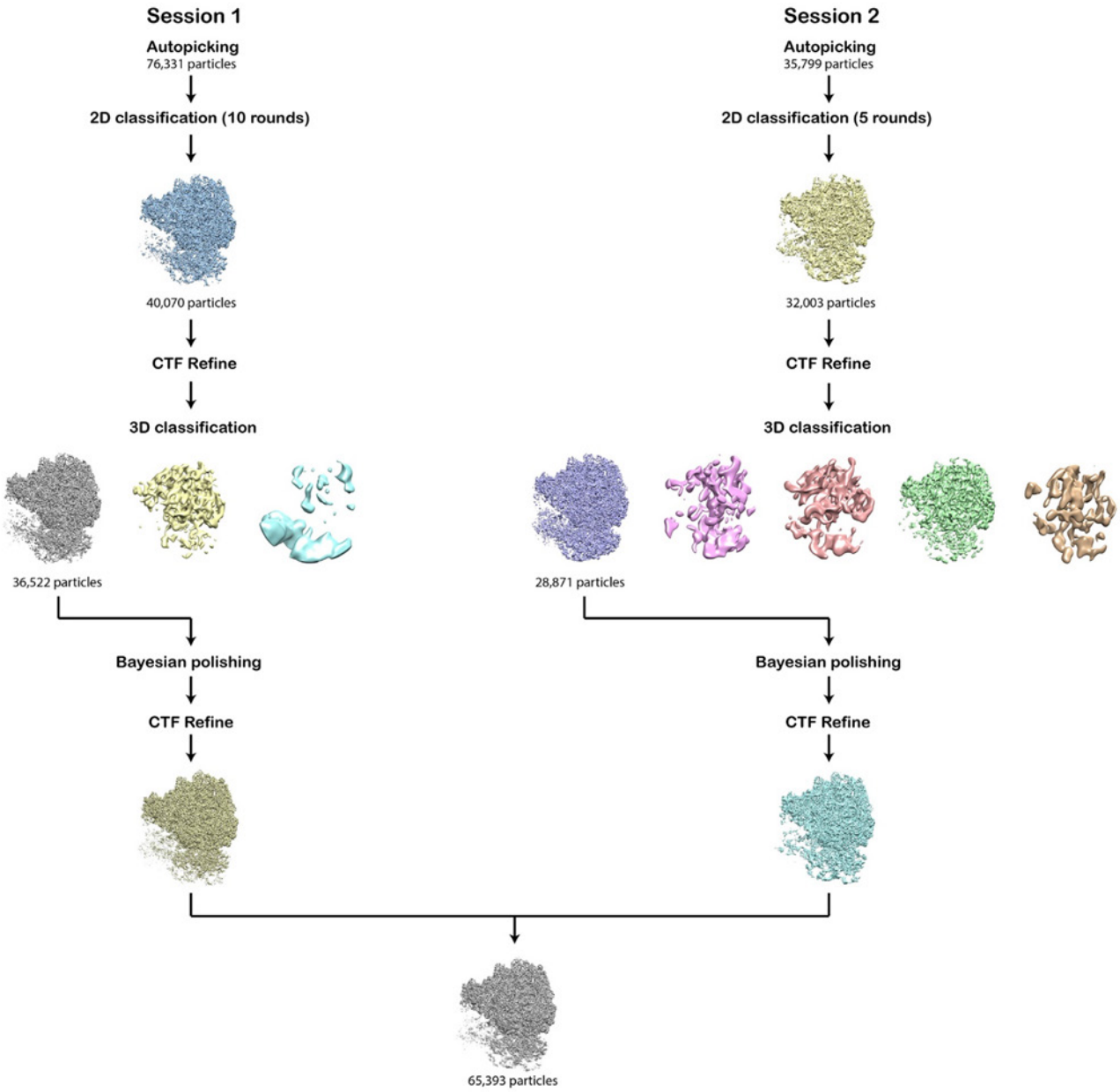
Supplementary Figure 6. Occlusion of the nascent peptide exit tunnel by antibiotics. (**A**) The lumen of the nascent peptide exit tunnel of the drug-free 70S ribosome (PDB entry 4Y4P[24]). The view is from the wide-open part of the tunnel onto the PTC as indicated by the inset. Nucleotide A76 of the P-site tRNA is shown in dark blue. A-site tRNA is not visible in this view, however, the location of its aminoacylated moiety is shown. Note that nucleotide A2062 of the 23S rRNA is pointed toward the viewer and is not involved in Hoogsteen-edge base pairing with the nucleotide m^2A2503 in the wall of NPET. (**B**, **C**, **D**) Occlusion of the nascent peptide exit tunnel by ERY (**B**), PHZ (**C**), and KLB (**D**). Structures of ERY and KLB are from PDB entries 6ND6[6] and 5W4K[1], respectively. Note that, unlike ERY and similar to KLB, PHZ almost completely occludes the lumen of the peptide exit tunnel. Also note that binding of ERY, PHZ, or KLB causes characteristic rotation of nucleotide A2062 by more than ninety degrees away from the viewer to form Hoogsteen base-pair with the m^2A2503 of the 23S rRNA.



Supplementary Figure 7. PHZ contacts with the ribosome. (**A-C**) Close-up views of the individual PHZ contacts with the nucleotides of the 23S rRNA. Hydrogen bonds are shown as dashed lines. (**D**) Close-up view of the intramolecular π-π-stacking system and its interaction with the nucleobase U2609 of the 23S rRNA. The directions of face-to-face and edge-to-face π-π stacking interactions are shown as dashed lines. Distances between the planes of aromatic heterocycles are shown in ångstroms (Å).



Supplementary Figure 8. Comparison of the in vitro translation inhibition activity of PHZ and its naturally occurring longer forms with additional amino acids at the N-terminus (A-PHZ and TA-PHZ). Graphs corresponding to PHZ are shown as solid lines. Dashed lines correspond to the mixture of long forms of PHZ (A/TA-PHZ, HPLC peak 4, see Figure S1A). (**B**) is a close-up view of (**A**) in the upper range of antibiotic concentrations (5-20 μM).



Supplementary Figure 9. Flowchart showing the steps of cryo-EM data processing in RELION.

Supplementary Table 2. Cryo-EM data collection, refinement and validation statistics.

	50S and PHZ (EMDB-20638) (PDB 6U48)
Data collection and processing	
Magnification	215,000×
Voltage (kV)	300
Electron exposure (e–/Å^2)	30
Defocus range (μm)	-0.6 to -2.0
Pixel size (Å)	0.5568 (processed at 1.1136)
Symmetry imposed	none
Initial particle images (no.)	112,130
Final particle images (no.)	65,393
Map resolution (Å)	2.87
FSC threshold	0.143
Map resolution range (Å)	2.5-7
Refinement	
Initial model used (PDB code)	4YBB
Model resolution (Å)	2.7
FSC threshold	0.143
Map sharpening *B* factor (Å^2)	n/a*
Model composition	
Non-hydrogen atoms	90950
Protein residues	3356
Ligands	1
B factors (Å^2)	
Protein	33.20
Ligand	14.14
Nucleic Acid	52.39
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.946
Validation	
MolProbity score	1.58
Clashscore	3.46
Poor rotamers (%)	0.77
Ramachandran plot	
Favored (%)	93.26
Allowed (%)	6.32
Disallowed (%)	0.43

*Input charge density map for refinement was prepared as described by Wang et al.[18].

REFERENCES

1. Metelev, M. *et al.* Klebsazolicin inhibits 70S ribosome by obstructing the peptide exit tunnel. *Nat. Chem. Biol.* (2017). doi:10.1038/nchembio.2462
2. Polikanov, Y. S. *et al.* Amicoumacin A Inhibits Translation by Stabilizing mRNA Interaction with the Ribosome. *Mol. Cell* (2014). doi:10.1016/j.molcel.2014.09.020
3. Almutairi, M. M. *et al.* Co-produced natural ketolides methymycin and pikromycin inhibit bacterial growth by preventing synthesis of a limited number of proteins. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx673
4. Osterman, I. A. *et al.* Madumycin II inhibits peptide bond formation by forcing the peptidyl transferase center into an inactive state. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx413
5. Khabibullina, N. F. *et al.* Structure of dirithromycin bound to the bacterial ribosome suggests new ways for rational improvement of macrolides. *Antimicrob. Agents Chemother.* (2019). doi:10.1128/AAC.02266-18
6. Svetlov, M. S. *et al.* High-resolution crystal structures of ribosome-bound chloramphenicol and erythromycin provide the ultimate basis for their competition. *RNA* (2019). doi:10.1261/rna.069260.118
7. Tereshchenkov, A. G. *et al.* Binding and Action of Amino Acid Analogs of Chloramphenicol upon the Bacterial Ribosome. *J. Mol. Biol.* (2018). doi:10.1016/j.jmb.2018.01.016
8. Polikanov, Y. S. *et al.* Distinct tRNA Accommodation Intermediates Observed on the Ribosome with the Antibiotics Hygromycin A and A201A. *Mol. Cell* (2015). doi:10.1016/j.molcel.2015.04.014
9. Pantel, L. *et al.* Odilorhabdins, Antibacterial Agents that Cause Miscoding by Binding at a New Ribosomal Site. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.03.001
10. Arenz, S. *et al.* A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest. *Nat. Commun.* (2016). doi:10.1038/ncomms12026
11. Orelle, C. *et al.* Tools for characterizing bacterial protein synthesis inhibitors. *Antimicrob. Agents Chemother.* (2013). doi:10.1128/AAC.01673-13
12. Mastronarde, D. N. SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. in *Microscopy and Microanalysis* (2003). doi:10.1017/s1431927603445911
13. Biyani, N. *et al.* Focus: The interface between data collection and data processing in cryo-EM. *J. Struct. Biol.* (2017). doi:10.1016/j.jsb.2017.03.007
14. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* (2018). doi:10.7554/eLife.42166
15. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* (2015). doi:10.1016/j.jsb.2015.08.008
16. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* (2019). doi:10.1107/S205225251801463X
17. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* (2004). doi:10.1002/jcc.20084
18. Wang, J. Experimental charge density from electron microscopic maps. *Protein Sci.* (2017). doi:10.1002/pro.3198

19. Noeske, J. *et al.* High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. &Amp; Mol. Biol.* **22**, 336 (2015).
20. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444909052925
21. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2010). doi:10.1107/S0907444910007493
22. Hanwell, M. D. *et al.* Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* (2012). doi:10.1186/1758-2946-4-17
23. You, S.-L. & Kelly, J. W. The total synthesis of bistratamides F–I. *Tetrahedron* **61**, 241–249 (2005).
24. Polikanov, Y. S., Melnikov, S. V., Söll, D. & Steitz, T. A. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.* (2015). doi:10.1038/nsmb.2992