

SEEKING SPECIES-SPECIFIC DRUG TARGETS IN MYCOBACTERIAL GENE
EXPRESSION

A Dissertation

by

KAILU YANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,
Committee Members,

Junjie Zhang
Hays S. Rye
Margaret E. Glasner
Matthew S. Sachs

Head of Department,

Gregory D. Reinhart

May 2018

Major Subject: Biochemistry

Copyright 2018 Kailu Yang

ABSTRACT

Gene expression in the human pathogen, *Mycobacterium tuberculosis* (*M. tuberculosis*), may possess species-specific structural features that can be targeted by drugs. Previous studies have identified, for example, several ribosomal RNA (rRNA) expansion segments and ribosomal proteins (rProtein) encoded by paralogous genes in the *M. tuberculosis* ribosome. To better understand the functions of these features and facilitate drug design, high-resolution structures are essential. Here we have used cryo-electron microscopy (cryo-EM) to solve the structures of RNA polymerase, the ribosome, and their higher order complexes.

The RNA polymerase structure we have solved is a core complex consisting of all five subunits. The structure indicates a new conformation of a bridging α -helix, termed “B2”, on the β subunit. In this new conformation, the B2 helix can potentially interact with rifampicin, a first-line drug against tuberculosis, suggesting exciting potential approach to design rifampicin derivatives. For the *M. tuberculosis* ribosome, we have solved multiple structures. These structures reveal that, upon joining of the large and small ribosomal subunits, a 100-nucleotide long expansion segment of the *M. tuberculosis* 23S rRNA, named H54a or the “handle”, switches from interacting with rRNA helix H68 and rProtein uL2 to interact with rProtein bS6, forming a new intersubunit bridge “B9”. In *M. tuberculosis* 70S, bridge B9 is mostly maintained, leading to correlated motions among the handle, the L1 stalk and the small subunit in the rotated and non-rotated states. Two new protein densities were discovered near the decoding center and the peptidyl transferase center, respectively. We have also characterized the high order complexes involved in gene expression, including ribosome dimers and expressome.

The biochemical and structural characterizations in this work provided a glimpse of the key components in the mycobacterial gene expression system. Most, if not all, unique features were unveiled, which should be not only further complemented with functional studies, but also taken into account for anti-tuberculosis drug development.

ACKNOWLEDGEMENTS

I am grateful to my advisor, Dr. Junjie Zhang, as well as my committee members, Drs. Hays S. Rye, Margaret E. Glasner, and Matthew S. Sachs, for their guidance and support throughout the course of this research. I thank them, as well as Karl Gorzelnik and Lunda Shen, for revising this dissertation.

I am also grateful to my collaborators, Drs. James C. Sacchettini, Xiaojun Li, Qingan Sun, and Xuelin Bian for their contributions to the *M. tuberculosis* projects, as well as Drs. Frank M. Raushel, and Zhongjie Ren for their contributions to the carbon-phosphorus lyase project.

Thanks also go to my friends and colleagues, and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my wife, parents and sisters, for their support, patience, and love.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professors Junjie Zhang, Hays S. Rye, and Margaret E. Glasner of the Department of Biochemistry and Biophysics, and Professor Matthew S. Sachs of the Department of Biology.

The *E. coli* carbon-phosphorus lyase complex, *M. tuberculosis* RNA polymerase, and *M. tuberculosis* ribosome were provided by Drs. Zhongjie Ren, Xiaojun Li, Qingan Sun, and Xuelin Bian. The linezolid analog was provided by Dr. Christoph M. Huwe. Jirapat Thongchol participated the data collection on the F20 microscope. Dr. Joanita Jakana directed, and Zhicheng Cui participated, the data collection on the JEM-3200FSC microscope. Jeng-Yih Chang, Ran Meng, and Lijun Duan participated the modeling of the *M. tuberculosis* ribosome. All of them, together with Drs. Frank M. Raushel, James C. Sacchettini and Junjie Zhang, participated the interpretation of results.

All other work conducted for the dissertation was completed by the student independently.

The work was supported by Welch Foundation Grants [A-1863, A-0015]; National Institutes of Health TB Structural Genomics Grant [P01AI095208]; National Institutes of Health Grants [P41GM103832, U24GM116787]; Department of Biochemistry and Biophysics Start-up Fund; Center for Phage Technology at Texas A&M University Start-up Fund.

TABLE OF CONTENTS

	Page
ABSTRACT.....	II
ACKNOWLEDGEMENTS.....	IV
CONTRIBUTORS AND FUNDING SOURCES	V
TABLE OF CONTENTS.....	VI
LIST OF FIGURES	IX
LIST OF TABLES.....	XI
CHAPTER I INTRODUCTION AND LITERATURE REVIEW.....	1
1.1 Tuberculosis remains a threat	1
1.2 <i>M. tuberculosis</i> is an uncommon bacterium	5
1.3 Drug targets in <i>M. tuberculosis</i>	8
1.4 Transcription in <i>M. tuberculosis</i> is a poorly-studied drug target.....	12
1.4.1 Transcription and RNAP.....	12
1.4.2 Bacterial transcription cycle	14
1.4.3 Bacterial transcription regulation.....	16
1.4.4 Bacterial transcription inhibition by small molecules	17
1.4.5 Previous studies on mycobacterial transcription.....	18
1.5 Gene translation in <i>M. tuberculosis</i> has species-specific features.....	19
1.5.1 Translation and ribosome.....	19
1.5.2 Bacterial translation cycle.....	20
1.5.3 Bacterial translation regulation	22
1.5.4 Bacterial translation inhibition by small molecules.....	22
1.5.5 Previous studies on mycobacterial translation.....	23
1.6 Gene expression in <i>M. tuberculosis</i> may involve different higher order complexes	25
1.7 Cryo-electron microscopy.....	27
CHAPTER II STRUCTURE OF THE RNA POLYMERASE FROM <i>M. TUBERCULOSIS</i>	34

	Page
2.1 Introduction.....	34
2.2 Materials and Methods.....	35
2.2.1 Expression and purification of <i>M. tuberculosis</i> RNAP and RbpA	35
2.2.2 Cryo-electron microscopy.....	36
2.2.3 Image processing	36
2.2.4 Molecular modeling.....	37
2.3 Results.....	40
2.3.1 Overview of the cryo-EM structure of <i>M. tuberculosis</i> RNAP core complex.....	40
2.3.2 The open conformation of the claw in the core complex.....	41
2.3.3 A unique conformation near the active center	43
2.4 Discussion.....	46
 CHAPTER III STRUCTURES OF THE RIBOSOME FROM <i>M. TUBERCULOSIS</i>	 47
3.1 Introduction.....	47
3.2 Materials and Methods.....	49
3.2.1 Purification of <i>M. tuberculosis</i> and <i>M. smegmatis</i> ribosomes	49
3.2.2 Cryo-electron microscopy.....	50
3.2.3 Image processing	51
3.2.4 Molecular modeling.....	53
3.2.5 Fitting the models for the forty conformations of the L1 stalk, the handle, and the 16S	55
3.2.6 Principal component analysis of the <i>M. tuberculosis</i> 70S structures.....	56
3.2.7 Figures and movies preparation	56
3.3 Results and Discussion	56
3.3.1 Architecture of the <i>M. tuberculosis</i> ribosomes.....	56
3.3.2 Interactions between handle and other regions in the <i>M. tuberculosis</i> ribosome	73
3.3.3 Correlated motions of the handle, the L1, and the SSU in the <i>M. tuberculosis</i> 70S	78
3.3.4 Inhibitors visualized in the <i>M. tuberculosis</i> ribosome	83
3.3.5 Two new protein densities close to the functional sites of the <i>M. tuberculosis</i> ribosome.....	88
3.4 Data Access.....	93
3.5 Acknowledgements.....	93
 CHAPTER IV STRUCTURES OF HIGHER ORDER COMPLEXES IN MYCOBACTERIAL GENE EXPRESSION.....	 94
4.1 Introduction.....	94
4.2 Materials and Methods.....	96
4.2.1 Measurement of <i>M. smegmatis</i> growth rate under optimal growth media	96
4.2.2 Purification of <i>M. smegmatis</i> polysome.....	96

	Page
4.2.3 Polysome profiling of <i>M. smegmatis</i> under different conditions.....	98
4.2.4 Preparation of <i>M. tuberculosis</i> expressome	98
4.2.5 Electron microscopy	98
4.2.6 Image processing	99
4.2.7 Molecular modeling.....	100
4.3 Results.....	101
4.3.1 <i>In silico</i> analysis of <i>M. tuberculosis</i> disome, 100S ribosome and expressome	101
4.3.2 Optimization of polysome purification from mycobacteria.....	104
4.3.3 Polysome profiling of <i>M. smegmatis</i> under different conditions.....	107
4.3.4 Cryo-EM analysis of the ribosome dimers	108
4.3.5 Preliminary studies of <i>M. tuberculosis</i> expressome	111
4.4 Discussion.....	112
CHAPTER V SUMMARY AND CONCLUSION	113
REFERENCES	118
APPENDIX I	156
APPENDIX II.....	195
APPENDIX III.....	215

LIST OF FIGURES

	Page
Figure 1.1. Schematic representation of bacterial gene expression with key machineries highlighted.....	11
Figure 1.2. Schematic representation of the vitrification method used in cryo-EM.....	30
Figure 2.1. Cryo-EM structure of <i>M. tuberculosis</i> RNAP core complex	39
Figure 2.2. Conformational difference between <i>M. tuberculosis</i> core complex and <i>T. aquaticus</i> core complex.....	41
Figure 2.3. Conformational change from <i>M. tuberculosis</i> core complex to <i>M. tuberculosis</i> RPo complex	43
Figure 2.4. The unique conformation of B2	45
Figure 3.1. Raw micrographs and resolution estimation	58
Figure 3.2. Cryo-EM structures of the <i>M. tuberculosis</i> 50S and 70S ribosomes	59
Figure 3.3. Cryo-EM density of the anticodon stem loop and the fragment of mRNA in the map of the <i>M. tuberculosis</i> 70S ribosome.....	60
Figure 3.4. Secondary structures of the <i>M. tuberculosis</i> rRNA.....	63
Figure 3.5. The rRNA kissing loop between H15 and H16a.....	67
Figure 3.6. Different lengths of H15 rRNA helices.....	68
Figure 3.7. Different conformations of the <i>M. tuberculosis</i> 70S.....	70
Figure 3.8. Five representative handle conformations.....	72
Figure 3.9. Interactions of the handle with other ribosomal components within the <i>M. tuberculosis</i> ribosomes.....	74
Figure 3.10. The conformation of the handle in the <i>M. tuberculosis</i> 50S specimen at 1 mM magnesium concentration.....	75
Figure 3.11. Superimposition of the <i>M. tuberculosis</i> 50S with the <i>E. coli</i> 70S initiation complex I (70S IC I).....	76
Figure 3.12. Correlated motions between the handle, SSU and the L1 stalk in the <i>M. tuberculosis</i> 70S ribosome	80

	Page
Figure 3.13. The correlation of the motions of the handle, the L1 stalk and the 16S.....	81
Figure 3.14. The handle in the rotated state is close to the mRNA exit	82
Figure 3.15. The comparison between linezolid and LZD-114	85
Figure 3.16. Unidentified proteins in the <i>M. tuberculosis</i> ribosomes discovered near the binding sites for two <i>M. tuberculosis</i> translation inhibitors.....	86
Figure 3.17. Comparison of the LZD-114 binding site in <i>M. tuberculosis</i> and mitochondrial ribosomes	87
Figure 3.18. Comparison of the binding sites of the mL41-like protein and the unknown protein in different ribosomes	89
Figure 3.19. The unknown protein is also present in <i>M. smegmatis</i> ribosome.....	91
Figure 4.1. <i>In silico</i> analysis of <i>M. tuberculosis</i> disome, 100S and expressome.....	103
Figure 4.2. Optimization of polysome purification from mycobacteria	106
Figure 4.3. Polysome profiling of <i>M. smegmatis</i> under different conditions	108
Figure 4.4. Cryo-EM analysis of the <i>M. tuberculosis</i> ribosome dimer.....	110
Figure 4.5. Preliminary cryo-EM study of <i>M. tuberculosis</i> expressome	111

LIST OF TABLES

	Page
Table 1.1. Major tuberculosis drugs and their targets.....	3
Table 3.1. Data collection and processing	52
Table 3.2. Model statistics for the final models of the <i>M. tuberculosis</i> 50S and the <i>M. tuberculosis</i> 70S	55
Table 3.3. Modeled regions of the rProteins and rRNAs.....	61
Table 3.4. Classification of <i>M. tuberculosis</i> 70S based on the region around the handle	71

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Recent advances of science and technology are substantial. For example, we can hear sounds from billions of years ago by detecting gravity waves; decipher the genetic code of any form of life on this planet relatively easily, cheaply, and quickly with next-generation sequencing; and even change these ciphers through advanced gene editing techniques. These achievements make us feel that humans, as the dominant life form on the planet, have the power to do whatever needs to be done for our advancement. However, the reality does not live up to our expectations so easily. Take the common cold as an example: it is caused by the simplest genetic entities, viruses comprised of just a few thousand nucleotides, yet annually causes outbreaks which can kill many people, particularly immuno-compromised individuals. For the common cold, recovery usually entails just suffering through the course of the infection, as there is little modern medicine can do to help the body remove the virus and speed up the recovery process. The common cold is a minor inconvenience, as healthy individuals can generally recover after a week of symptoms, but how about a disease that threatens one-quarter of the world's population and kills millions every single year?

1.1 Tuberculosis remains a threat

Tuberculosis (TB) is an infectious disease, caused by the bacterium *Mycobacterium tuberculosis* (*M. tuberculosis*) (Koch, 1882). TB has affected humans since the Neolithic era (Hershkovitz et al., 2008). It was not until 1882 that Robert Koch discovered that TB is caused by *M. tuberculosis* (Koch, 1882). In the first half of the 20th century, vaccines were developed

and have been widely used as a preventative measure (Luca and Mihaescu, 2013). On the other hand, diagnosing the disease through X-ray examinations has enabled better treatment of TB patients. In the middle of the 20th century, many efficient oral TB drugs were developed, including four first-line drugs: isoniazid (1952), pyrazinamide (1954), ethambutol (1961), and rifampicin (1963) (**Table 1.1**) (Zumla et al., 2013). All of these breakthroughs in pathology, prevention, diagnosis, and treatment have contributed to a dramatic decrease of in the number of deaths caused by TB (Daniel, 2006). However, even in the 21th century, TB is still among the top causes of death, according to the World Health Organization (WHO) (WHO, 2017). Overall, one-quarter of the world's population is infected by *M. tuberculosis* (Houben and Dodd, 2016). Most have a latent TB infection but do not show any symptoms and do not spread the disease. However, within these individuals there is a ~10% chance for them to develop an active TB infection (Vynnycky and Fine, 1997). In 2016, there were 10.4 million new TB cases, with children (under age 14) accounting for about 10% (WHO, 2017). At this active stage, the disease becomes contagious and one patient can transmit the disease to 10-15 people in a year through close contact. HIV-positive individuals are 20 to 30 times more susceptible to developing an active TB infection than those without HIV. If not treated properly, almost all HIV-positive individuals with TB will die from it, with ~40% of deaths in HIV patients due to TB. In 2016, TB caused 1.7 million human deaths, among which 0.4 million were HIV-positive, and 250,000 were children. Over 95% of deaths from TB occur in developing countries and almost two-thirds of the deaths come from just 7 countries: India, Indonesia, China, Philippines, Pakistan, Nigeria, and South Africa.

Table 1.1. Major tuberculosis drugs and their targets

First-line drugs	Target
Rifampicin	Gene expression (RNA polymerase)
Pyrazinamide	Gene expression (ribosome)
Isoniazid	Cell wall synthesis (Enoyl-[acyl-carrier-protein] reductase)
Ethambutol	Cell wall synthesis (Arabinosyl transferases)
Second-line drugs	Target
Streptomycin	Gene expression (ribosome)
Capreomycin	Gene expression (ribosome)
Kanamycin	Gene expression (ribosome)
Amikacin	Gene expression (ribosome)
Ethionamide	Cell wall synthesis (Enoyl-[acyl-carrier-protein] reductase)
Cycloserine	Cell wall synthesis (D-alanine racemase and ligase)
Ofloxacin	Gene replication (DNA topoisomerase)
Para-amino salicylic acid	Folate synthesis (Dihydropteroate synthase)

Diagnosis of TB has improved dramatically from the early days after *M. tuberculosis* was identified as the causative agent, but still remains a challenge, particularly in underdeveloped areas. Traditionally, an individual is suspected of having TB based on common symptoms, which include coughing, fever, night sweats, and weight loss. (Ryu, 2015). Then a medical professional must take a sputum sample from the patient's lung and conduct a smear microscopy test to diagnose TB. Basically, a doctor will look at the sputum under a microscope to look for the *M. tuberculosis* (Desikan, 2013). However, there is a 50% chance for this microscopy method to fail as greater than 10,000 organisms/ml are required for detection (Desikan, 2013). Even if *M. tuberculosis* is detected, no information about drug resistance for that particular infection can be

obtained without further expensive tests. Chest X-ray examinations are another popular method, since the main site of *M. tuberculosis* infections in TB patients is the lungs, but this method can only provide some clues as to the potential cause of infection and cannot be used to make an accurate diagnosis of the bacteria infecting the individual (Cudahy and Shenoi, 2016). Currently, the WHO recommends the rapid Xpert MTB/RIF® method, which takes only 2 hours (Detjen et al., 2015). This new method not only detects *M. tuberculosis* but also tests for resistance to the most important TB antibiotic, rifampicin. However, it is still challenging to diagnose multi-drug resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB). It is also difficult to diagnose TB in children, which accounts for 10% of total TB cases (WHO, 2017).

While vaccination of individuals against TB is an important approach, the commonly used BCG (bacille Calmette-Guerin) vaccine is not 100% effective, especially for TB infections in the lung (Colditz et al., 1994; Rodrigues et al., 2011). It also has less efficacy in adults than in children (Davenne and McShane, 2016). Typically the vaccination is given to children only when the detection of TB reaches 40 cases out of 100,000 children per year in a certain area. While the combination of TB and HIV is extremely lethal, it is hard to predict who will develop HIV and the need for vaccination in advance. On the other hand, in the most TB-prevalent countries, the healthcare infrastructure is inadequate and thus broad vaccinations against TB are not practical.

When modern medicine is available, the treatment of TB can be very successful. Around 53 million TB patients were cured between 2000 and 2016 (WHO, 2017). The treatment process involves a combination of the four TB first-line drugs, and takes about six months (Horsburgh Jr et al., 2015). During the treatment, patients need to be supervised and supported properly. However, again, this long treatment protocol is difficult to implement in underdeveloped

countries. Similar to any other infectious disease, *M. tuberculosis* will also develop drug resistance (Zaman et al., 2017). Due to the misuse of antibiotics, MDR-TB and XDR-TB have emerged. MDR-TB is defined as when *M. tuberculosis* is resistant to isoniazid and rifampicin, the most potent of the four first-line drugs (Gunther, 2014). In 2016, there were half a million MDR-TB patients, half of which are from India, China and the Russian Federation where antibiotics are overprescribed. MDR-TB is still curable using second-line drugs, but the treatment is toxic, expensive, and time-consuming (up to two years). Continuous infections can result if the drug regimen is not followed. In reality, only half of MDR-TB patients are successfully cured (WHO, 2017). If *M. tuberculosis* develops resistance to second-line drugs, then MDR-TB becomes the XDR-TB (Gunther, 2014). According to WHO, only 30% of XDR-TB patients are able to be successfully treated (WHO, 2017).

Although TB is preventable, detectable, and curable in most cases, the million plus annual human deaths is still a ghastly toll in this age of modern medicine. It is worthwhile to ask why the TB-causing pathogen, *M. tuberculosis*, is so successful, or at least, what differentiates *M. tuberculosis* from less successful pathogens and non-pathogenic bacteria.

1.2 *M. tuberculosis* is an uncommon bacterium

M. tuberculosis is a Gram-positive, acid-fast and GC-rich (65%) bacterium, within the *Actinobacteria*, belonging to the genus *Mycobacteria* (Cook et al., 2009). Morphologically, *M. tuberculosis* is similar to other mycobacteria, which are noted for their thick, waxy cell wall that contains peptidoglycan, arabinogalactan, and mycolic acids (Chiaradia et al., 2017). The innermost layer of the cell wall is composed of peptidoglycan which wraps around the cell membrane. Beyond the peptidoglycan layer is arabinogalactan, which is conjugated with mycolic

acids. Mycolic acids are long-carbon-chain (60-90 carbon atoms) fatty acids, which are responsible for the waxy cell wall, a defining feature of mycobacteria (Chiaradia et al., 2017; Daffe and Draper, 1998). The waxy cell wall forms a unique barrier between the inner metabolism and the outer environment. The mycobacterial cell wall further contributes to the characteristic persistence of *M. tuberculosis*, by inhibiting the diffusion of some drugs into the cell, thereby providing inherent drug resistance (Hett and Rubin, 2008).

Another feature of *M. tuberculosis*, and other mycobacteria, is their slow growth rate. The doubling time of *M. tuberculosis* under optimal growth conditions is around 20 hours (James et al., 2000), and that of its non-pathogenic close relative, *Mycobacterium smegmatis* (*M. smegmatis*), is around 2 hours (Stephan et al., 2005). This is in stark contrast to the 20 minute doubling time of *Escherichia coli* (*E. coli*) (Sezonov et al., 2007). Beyond these two species the growth rates for all other mycobacteria are also slow. It can take 2-40 days to see visible colonies of mycobacteria on an agar plate (Cook et al., 2009). Interestingly, the pathogenic *M. tuberculosis* and *Mycobacterium leprae*, the causative bacteria of leprosy, have only one ribosomal RNA operon (denoted as *rrn*; and for these two organisms as *rrnA*) in their genomes and grow slowly, while *M. smegmatis* and *Mycobacterium phlei* have two (*rrnA* and *rrnB*) and grow relatively faster (Cook et al., 2009). However, there also exist examples of mycobacteria that are slow growers which have two *rrn* operons and fast growers that have one *rrn* operon (Cook et al., 2009). Furthermore, inactivating either one of the two *rrn* operons in *M. smegmatis* does not affect the growth rate (Sander et al., 1996). Therefore, the number of *rrn* operons per genome does not determine the growth rate of mycobacteria. The reasons why mycobacteria grow slowly are unclear (Lewin and Sharbati-Tehrani, 2005). Slow growth rates, often suggest lower levels of metabolism, which decreases the efficacy of many drugs (Tuomanen et al., 1986).

There are many difficulties encountered when studying *M. tuberculosis* in research laboratories. Aside from the aforementioned pathogenic hazard and the slow growth rate, *M. tuberculosis* cells tend to clump together, causing resistance to traditional cell lysis techniques and hindering genetic manipulations (Cook et al., 2009). The most widely studied *M. tuberculosis* strain is H37Rv (KUBICA et al., 1972), with “H37” as the initial name given when it was isolated from a patient, and the “R” indicating a rough morphology and “v” virulent (Steenken Jr, 1935). The genome sequence of H37Rv was first published in 1998 (Cole et al., 1998). The genome has 4,411,529 base pairs with around 4,000 genes, a typical size for most bacteria. No mycobacteria sequenced so far have been found to have more than two *rrn* in their genomes, while other bacteria have many more, for instance *E. coli* has seven (Condon et al., 1995).

While H37Rv is the most widely studied *M. tuberculosis* strain, it is not easy to genetically manipulate. Therefore, *M. smegmatis*, a close relative of *M. tuberculosis*, was chosen as a model organism for mycobacteria. *M. smegmatis* does not generally infect humans and grows much faster than *M. tuberculosis*, both inherently advantageous to researchers aside from the fact that genetic manipulation is easier in *M. smegmatis* compared to *M. tuberculosis*. The commonly used *M. smegmatis* strain is MC² 155, due to its efficient plasmid transformation phenotype (Snapper et al., 1990). However, there has been some debate on the relevance of *M. smegmatis* as a model for *M. tuberculosis* (Barry, 2001a, b; Reyrat and Kahn, 2001; Tyagi and Sharma, 2002). Despite the conservation of housekeeping genes, *M. smegmatis* is universally found in soil and generally regarded as non-pathogenic, while *M. tuberculosis* infects humans and obligately resides in mammalian hosts (Montali et al., 2001). For conserved cellular processes, such as gene transcription and translation, the basic mechanisms are the same, and

even the same as in *E. coli*. However, for drug design purposes, the change of a single nucleotide may result in a different amino acid or RNA sidechain and thus potentially affect drug binding. For instance, while the first-line drug isoniazid inhibits both *M. tuberculosis* and *M. smegmatis*, *M. smegmatis* requires a higher minimum inhibitory concentration (MIC) to prevent growth (Mohan et al., 2015). There are few cases where the two species have identical sequences for common cellular targets. Moreover, genes required for virulence in *M. tuberculosis* may be absent in *M. smegmatis* (Tyagi and Sharma, 2002). Therefore, results obtained from *M. smegmatis* should generally be verified in *M. tuberculosis*. To avoid any misinterpretations, in this work *M. tuberculosis* H37Rv was studied, with *M. smegmatis* only used for optimization of polysome purification and subsequent structural characterization.

1.3 Drug targets in *M. tuberculosis*

There are two essential elements to consider when evaluating whether an antibiotic is considered good or not. The first is whether the drug is potent against a particular target organism. The potency of a drug requires that the drug target play an important role in cellular processes such as metabolism or replication. When this important target is inhibited by the drug, the bacteria will not survive. The second is whether the drug selectively kills the pathogen alone, rather than affecting the human host or beneficial bacteria. In other words, are there off target effects? The selectivity of a drug necessitates that the drug target be unique to the pathogen. This necessitates the target be distinct between the pathogen and humans, or non-pathogenic bacteria, *e.g.*, probiotic bacteria. Otherwise, the antibiotic will also inhibit the target in human cells with toxic side-effects, or kill off non-pathogenic bacteria which can displace other pathogens. In contrast to the large number of antibiotics, the number of drug targets is relatively limited. In

general, the antibiotics are classified, based on their targets, into the following four classes: those which target 1) cell wall biosynthesis, 2) gene replication, 3) gene expression, and 4) others, including folic acid metabolism, cell membranes (Yoneyama and Katsumata, 2006).

Cell walls, absent in human, are important for bacteria, because the intracellular pressure of a bacterium is often higher than the environment. Without protection from the cell wall, the cell membrane will undergo osmolysis and the cell will die. The most important component of the cell wall, for both Gram-negative and Gram-positive bacteria, is a layer of peptidoglycan consisting of covalently crosslinked glycan and peptides. The crosslinking process is catalyzed by transglycosidases and transpeptidases, the targets of β -lactam antibiotics, such as penicillin. Two out of the four TB first-line drugs, isoniazid and ethambutol, also target cell wall biosynthesis (**Table 1.1**) (Zumla et al., 2013). Isoniazid inhibits mycolic acid synthesis, while ethambutol inhibits arabinogalactan biosynthesis. Ethionamide, a second-line drug, also inhibits mycolic acid biosynthesis (Zumla et al., 2013).

DNA replication is essential for all domains of life. Chromosomes have to be unwound for chromosome replication to occur (Dorman, 2006). The enzymes which catalyze the unwinding process are topoisomerases (Drlica, 1990). There are two types of topoisomerases. Type I topoisomerases break only one strand of the double stranded DNA at a time, while type II break both strands at the same time. Antibiotics in the quinolone class target DNA gyrase, a type II topoisomerase (Champoux, 2001). Ofloxacin is a second-line TB drug and belongs to the fluoroquinolone family (Smythe and Rybak, 1989).

Gene expression can be divided into two mechanistically distinct steps, gene transcription and gene translation, both of which are essential (**Figure 1.1**). Transcription is the process of reading the genetic information from DNA and synthesizing a complementary RNA strand. The

enzymatic complex involved in this process is RNA polymerase (RNAP) (Murakami, 2015). In bacteria, there is only one type of RNAP, while, in eukaryotes, there are multiple types, each of which is responsible for synthesizing a distinct subset of RNA (Cramer et al., 2008). The differences between bacterial and eukaryotic RNAPs provides a great potential to target bacterial RNAP specifically, leading to less toxicity for humans. One of the four first-line TB drugs, rifampicin, inhibits RNAP by binding to its active site in the β subunit (Campbell et al., 2001). Most of the RNAs transcribed by the RNAP are involved in the next step, translation, where protein is synthesized. The mRNA serves as a code for each amino acid in the protein; each of the tRNAs correlates one amino acid according to the codon in the mRNA, and the rRNA constitutes the ribosome, the ribozyme for translation. Although all ribosomes perform the same function, translation of mRNA into proteins, ribosomes are very different between bacteria and eukaryotes, and many have species-specific features. One of the four first-line TB drugs, pyrazinamide, and four out of eight major second-line TB drugs target the translation process (**Table 1.1**) (Zumla et al., 2013). Indeed, translation is a major drug target (Poehlsgaard and Douthwaite, 2005), which may not be a surprise considering the complexity of the ribosome and multiple steps involved.

Folic acid metabolism is important for all living cells to synthesize a variety of other cellular components (Bermingham and Derrick, 2002). One feature that makes folic acid metabolism an attractive drug target is that humans get it from their diet while bacteria synthesize it. Among drugs that target this pathway are sulfa-based drugs that has been used since the 1930s, an example of which is para-aminosalicylic acid (PAS), a second-line TB drug, that inhibits dihydrofolate reductase (DHFR) by generating a hydroxyl dihydrofolate antimetabolite (Zheng et al., 2013).

In summary, while there are many antibiotics there are relatively few drug targets that are specific for bacteria. The gene expression process was selected as the target for this study due to the abundance of antibiotics and modes of action in this class. The general mechanisms of gene expression have been elucidated in other model bacteria, and it should be technically practical to further characterize the detailed features of RNAP and ribosomes in *M. tuberculosis*. The mechanisms and structural models deduced from studies in non-pathogenic species may not necessarily be exactly true in *M. tuberculosis*. The following three subsections will describe the key machineries involved in gene expression, including RNAP, ribosome, expressome and polysome (Figure 1.1).

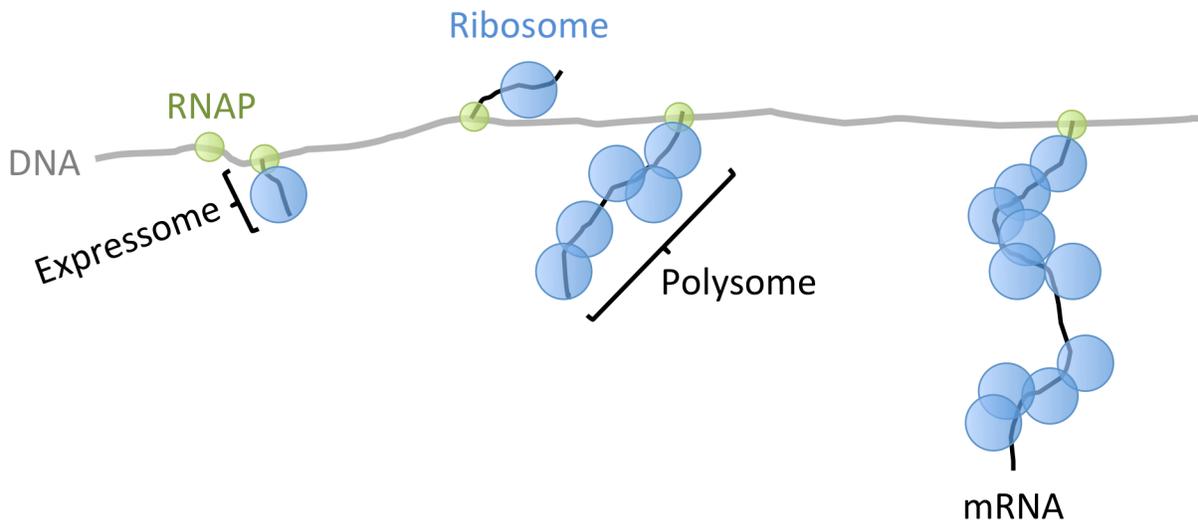


Figure 1.1. Schematic representation of bacterial gene expression with key machineries highlighted

This figure was drawn based on an electron microscopy image of an *E. coli* gene (Miller et al., 1970). Notably, the coupling between the two events, transcription by RNAP and translation by ribosome, is a unique feature of bacterial gene expression.

1.4 Transcription in *M. tuberculosis* is a poorly-studied drug target

1.4.1 Transcription and RNAP

Transcription is a chemical reaction in which four ribonucleoside triphosphates (ATP, UTP, CTP, GTP) are polymerized into an RNA chain, while the sequence of the addition of those four substrates are defined by a DNA template. It is also called DNA-directed synthesis of RNA, and the enzyme that catalyzes this reaction is called DNA-dependent RNA polymerase. By the 1950s, concept of RNA acting as an intermediate of the genetic information flow from DNA to protein was accepted but had not been demonstrated (Hurwitz, 2005). In 1955, an RNA-degrading enzyme was discovered and thought to catalyze the biosynthesis of RNA, *i.e.*, DNA-directed synthesis of RNA. However, it was later found the DNA template was not required by that enzyme and the reaction was reverse when performed *in vitro* (Kresge et al., 2006). It was not until the 1960s that the true DNA-dependent RNA polymerase, or RNAP, was discovered and isolated (Hurwitz, 2005). RNAP is found in all three domains of life and is also found in some viruses (Mirzakhanyan and Gershon, 2017). Since the basic formation of polynucleotides from all organisms is very similar, the catalytic mechanism is conserved (Kireeva et al., 2013).

A wealth of biochemical, biophysical and genetic information of RNAP has accumulated since its discovery. Understanding of bacterial transcription has been stimulated by emerging structural information since the late 1990s (Murakami, 2015). Similar to other big and complex macromolecules, the structural study of RNAP started with domains and subcomplexes. The first atomic structure solved was the C-terminal domain of the RNAP α subunit, or α CTD, from *E. coli*, determined by nuclear magnetic resonance spectroscopy (NMR) (PDB id: 1COO) (Jeon et al., 1995). The structure revealed a distinct folding of α CTD compared with other DNA-binding proteins. Together with mutagenesis studies, the structure explained the interaction of α CTD

with transcription activators and the upstream promoter (UP) element. The N-terminal domain of the RNAP α subunit, or α NTD, was later solved (PDB id: 1BDF) (Zhang and Darst, 1998). The structure showed the dimerization of one of the two subdomains of α NTD, which further serves as a platform for assembling other RNAP subunits. The two largest subunits, β and β' , feature large lineage-specific insertions (Lane and Darst, 2010). Many of these insertions are soluble domains and were solved by X-ray crystallography. For *M. tuberculosis*, the β subunit 1 and 2 domains were solved (Gulten and Sacchettini, 2013), which revealed its interaction with CarD, a transcription regulator. The structure provides important insights into how RNAP is regulated by CarD in *M. tuberculosis*. In 2001, the first complete structure of bacterial RNAP was published (Minakhin et al., 2001), revealing the architecture of $\alpha_2\beta\beta'\omega$. The shape of the complex structure is reminiscent of a crab claw, with the active site buried inside the center of the claw. Studies of eukaryotic and archaeal RNAPs confirmed the conserved overall shape, subunit organization and active site motifs, such as the bridge helix, trigger loop and switches (Cramer et al., 2001; Engel et al., 2013; Fernandez-Tornero et al., 2013; Hirata et al., 2008; Hoffmann et al., 2015).

All RNAPs catalyze RNA synthesis by the same two-metal-ion mechanism (Steitz and Steitz, 1993; Yang et al., 2006) but not all RNAPs are exactly the same. For example, in eukaryotes, it turns out that different types of RNAP synthesize different types of RNA. RNA polymerase I (RNAP I) synthesizes the 25S rRNA precursor; RNAP II synthesizes mRNAs; RNAP III synthesizes tRNAs and the 5S rRNA (Vannini and Cramer, 2012). These three RNAPs share a conserved 10-subunit core, but contain 14, 12, and 17 subunits, respectively (Vannini and Cramer, 2012). Some plants have RNAP IV and RNAP V to synthesize noncoding RNAs (Ream et al., 2014). In bacteria, there is only one type of RNAP. The core enzyme consists of 5 subunits: α (two copies), β , β' , and ω , all of which have homologs in eukaryotic RNAPs

(Murakami, 2015). The multi-subunit RNAPs from three domains of life are significantly different from the T7 phage-like single-subunit RNAP family, not only in the overall shape and domain organization, but also in the configuration of the active site. Single-subunit RNAP can directly recognize promoters, while the core enzyme of the multi-subunit RNAP cannot initiate the transcription without a σ factor (Murakami, 2015). The core enzyme and a σ factor form the holoenzyme, which initiates transcription by recognizing the promoter region in the DNA. The initiation is also assisted by activators (Busby and Ebright, 1999). Once transcription starts, the core enzyme can facilitate the elongation without a σ factor (Travers and Burgess, 1969).

1.4.2 Bacterial transcription cycle

The bacterial five-subunit RNAP represents the minimum form of RNAP across three domains of life (Murakami, 2015). Transcription, even in bacteria, is a complicated process. In one transcription cycle, one copy of RNA is synthesized, and many copies of RNA molecules are synthesized by repeating the transcription cycle. Generally, the transcription cycle can be divided into four steps: DNA template binding, RNA chain initiation, RNA chain elongation, and RNA chain termination.

To initiate transcription, the RNAP must bind the DNA template. Interestingly, the RNAP only binds to specific regions of the DNA template, *i.e.*, the promoter. The promoter is recognized by its specific σ factor (Paget, 2015). The binding specificity is determined by the conserved regions around -35 and -10 sites (Harley and Reynolds, 1987; Pribnow, 1975). The region around -10 position is the most conserved, usually with TATAAT consensus sequence. The region around -35 position usually has TTGACA, especially for efficient promoters. For highly expressed genes, the region between -40 and -60 positions has an AT-rich fragment,

termed the upstream promoter (UP) element (Estrem et al., 1998). The UP element interacts with the α CTD of RNAP (Ross et al., 1993). After the holoenzyme binds to the promoter, 12-14 base pairs of the DNA template, from the middle of the -10 region to the initiation site, will be deformed from double-stranded to single-stranded. This single-stranded region within the double stranded DNA resembles a bubble and thus is termed as transcription bubble (Robb et al., 2013). When the transcription bubble is observed, the DNA-holoenzyme complex, termed as an open promoter complex (RPO), becomes transcriptionally competent (Saecker et al., 2011). During this template binding process, the σ factor plays a critical role in recognizing specific promoters (Paget, 2015). Bacterial σ factors are homologous to the archaeal transcription factor B and the eukaryotic TFIIB (Burton and Burton, 2014). Based on the DNA sequence, the required σ factor can be different (Burton and Burton, 2014). In *E. coli*, there are 7 σ factors (Hillebrand et al., 2005). The primary or housekeeping one is σ^A or σ^{70} with a molecular weight of 70 kDa (Paget and Helmann, 2003). The σ^{70} recognizes essential genes and the σ^{70} family are universally expressed in all bacteria (Paget and Helmann, 2003; Sachdeva et al., 2010). The σ^{70} family has 4 domains, denoted as σ_1 , σ_2 , σ_3 , and σ_4 . σ_2 and σ_4 recognize the -10 and -35 elements, respectively, to facilitate the template binding (Murakami et al., 2002a). Another major family is the σ^{54} family which adopts a different structure and mechanism (Bonocora et al., 2015). The σ^{54} family has not been found in any GC-rich, Gram-positive bacteria (including *M. tuberculosis*) and cyanobacteria (Sachdeva et al., 2010). *M. tuberculosis* has 13 σ factors, almost twice as many as *E. coli*, while the sizes of their genomes are similar (Manganelli et al., 2004). Among obligate pathogens, *M. tuberculosis* has the highest number of σ factors / genome size ratio, which has been proposed to reflect the complexity of regulatory mechanisms (Rodrigue et al., 2006; Sachdeva et al., 2010).

The RNA chain initiation is a process where two free nucleotides dimerize into a dinucleotide (Travers, 1993). Usually the first nucleotide is a purine, *i.e.*, either A or G. After the dimerization, the 5'-triphosphate is retained in the product. Then, the RNA chain is elongated from 5' to 3'. During elongation, the DNA rotates to prevent the intertwining of DNA and RNA (Futcher, 1988; Harada et al., 2001). Once initiated, elongation runs rapidly at rates of 20-70 nucleotides per second (Maiuri et al., 2011). The second RNAP will bind to the promoter as long as there is enough room, which means a subsequent transcription cycle can start while the first cycle is proceeding (Miller et al., 1970). The transcription is finally terminated either intrinsically or in a Rho-dependent fashion (Ray-Soni et al., 2016). The intrinsic termination relies on a GC-rich stem followed by a U-tract (7-to-9-nucleotide U-rich sequence motif). The Rho factor is a hexameric protein, with a ring-like shape. The ring can be either closed (PDB id: 3ICE) (Thomsen and Berger, 2009) or open (PDB id: 1PVO) (Skordalakes and Berger, 2003) in different functional states.

1.4.3 Bacterial transcription regulation

Transcription is regulated in many different ways. First, anti- σ factor can block the formation of holoenzyme by interacting with σ factor (Paget, 2015). In *M. tuberculosis*, structures of several σ -anti- σ complexes have been solved (PDB ids: 4NQW, 3VEP, 3HUG) (Jaiswal et al., 2013; Shukla et al., 2014; Thakur et al., 2010), revealing the structural mechanisms of inactivation of σ factors in response to environmental changes. Second, many transcription factors can bind to the RNAP core complex to regulate its activity. For example, the catabolite activator protein (CAP) can bind both DNA and α CTD of RNAP, forming a CAP- α CTD-DNA complex to facilitate the initiation of transcription (PDB id: 1LB2) (Benoff et al.,

2002). Third, transcription factors from bacteriophages interact with host RNAP to inhibit host transcription and promote phage gene transcription instead. For example, the gp39 protein from the *Thermus thermophilus* (*T. thermophilus*) phage P23-45 binds to β -flap domain and σ_4 of RNAP and remodels the domain organization so that the binding with -35 region is hindered and the transcription of the -10/-35 class promoters is inhibited (Tagami et al., 2014). Fourth, repressors can bind to their corresponding operators to repress transcription. For example, the *lac* repressor can bind to the *lac* operon via three binding sites, O_1 , O_2 , and O_3 , resulting in DNA loop formation. Such a looped complex prevent RNAP from forming a productive initiation complex (Lewis et al., 1996).

1.4.4 Bacterial transcription inhibition by small molecules

Transcription can be inhibited by small molecules, some of which target the non-conserved regions in eukaryotic and bacterial RNAP. For example, the widely-used anti-tuberculosis drug, rifampicin, targets the active site at the center of bacterial RNAP. Several rifampicin-RNAP structures have been solved (PDB ids: 1YNN, 4KMU) (Campbell et al., 2005; Molodtsov et al., 2013). These structures revealed the details of the drug-target interaction, showing that the binding of the drug blocks the exit path of the newly-synthesized RNA. RNAP with bound rifampicin derivatives have also been structurally characterized. These derivatives generally have additional interactions with RNAP and have higher binding affinity. For example, benzoxazinorifamycin (bxRIF) has a C3'-tail that interacts with $\sigma_{3.2}$ loop (σ finger) and thus affects its conformation (PDB ids: 4KN4, 4KN7) (Molodtsov et al., 2013). This interaction provides an additional inhibition mechanism that appears to block the DNA template binding step.

Apart from rifampicin and its derivatives, several other inhibitors that have been visualized together with RNAP, including streptolydigin (PDB ids: 2PPB, 1ZYR, 2A6H) (Temiakov et al., 2005; Tuske et al., 2005; Vassylyev et al., 2007), myxopyronin (PDB ids: 3DXJ, 3EQL) (Belogurov et al., 2009; Mukhopadhyay et al., 2008), and salinamide (PDB id: 4MEX) (Degen et al., 2014). Despite disagreements as to the mechanisms of inhibition by some of these inhibitors (Belogurov et al., 2009; Mukhopadhyay et al., 2008), these structures revealed additional RNAP drug binding sites, such as the bridge helix, trigger loop, and switch. Interestingly, all of these binding sites reside in the catalytic region of RNAP, whereas the other major part of RNAP, the assembly platform region, has no identified drug binding sites (Werner and Grohmann, 2011).

1.4.5 Previous studies on mycobacterial transcription

Despite the large number of structures of RNAP from *E. coli*, *T. thermophilus*, *Thermus aquaticus* (*T. aquaticus*) (Murakami, 2015), only recently has a structure of a mycobacterial RNAP been reported (Hubin et al., 2017a; Lin et al., 2017). Nevertheless, accumulating data suggests that the transcription system of mycobacteria may have their own unique features. The RPo complex is less stable in mycobacteria than in *E. coli*, and it is stabilized by the CarD factor, which is absent in *E. coli* (Davis et al., 2015; Stallings and Glickman, 2011). A transcription factor, RbpA, is present in mycobacteria but absent in *E. coli* and *Thermus spp.* (Hubin et al., 2015; Tabib-Salazar et al., 2013). Moreover, mycobacterial RNAP exhibits better termination efficiency than *E. coli* RNAP, in the case of imperfect U-tracts (Czyz et al., 2014). Sequence-level analysis also revealed lineage-specific insertions in the large subunits (β and β') of RNAP (Lane and Darst, 2010). For example, the actinobacteria-specific β' In1 of *M. tuberculosis* is

different from that of *E. coli* or *T. thermophilus*. The recently published structures of RNAP from *M. tuberculosis* and *M. smegmatis* have identified this insertion, spanning from around residue 140 to residue 230 (Hubin et al., 2017a; Lin et al., 2017). Although the insertion point is the same as that in *Thermus spp.*, the structure of this insertion is different. In mycobacteria, 90-residue segment forms a two-helix bundle, while in *Thermus spp.*, the segment folds into a barrel. Interestingly, the charge distribution on the surface of the helix bundle is found to be conserved (Hubin et al., 2017b). In addition, the N-terminal of $\sigma^{A}_{1.2}$ (σ^{A}_N) is reported to be located between $\beta 2$ domain and the actinobacteria-specific $\beta'In1$. The spatially close organization of σ^{A}_N and $\beta'In1$ was proposed to block the DNA template binding (Hubin et al., 2017b), while in another study, it was proposed to trap the DNA template and thus stabilize the RPo (Lin et al., 2017).

1.5 Gene translation in *M. tuberculosis* has species-specific features

1.5.1 Translation and ribosome

Translation is the chemical reaction of peptide bond formation, by which the 20 canonical amino acids are polymerized into a polypeptide chain, based on the sequence of a mRNA (Green and Noller, 1997). Its direct substrates are aminoacylated tRNAs. Every tRNA has an anticodon stem loop (ASL) to pair with a particular codon in the mRNA, so that the sequence of the polypeptide is determined by the sequence of the mRNA. The tRNA and mRNA cannot perform translation without the ribosome. The ribosome is made of two subunits in all kingdoms of life, one large subunit (LSU) and one small subunit (SSU). The LSU contains the peptidyl transferase center (PTC) that consists exclusively of RNA, instead of protein (Moore and Steitz, 2011). The SSU contains the decoding center where the pairing between codon and anticodon is favored.

The LSU and SSU cooperate to translocate mRNA and tRNA. The mRNA is decoded from 5' to 3' and the polypeptide is synthesized from N-terminus to C-terminus. Despite the conservation of the ribosome, there are substantial differences between ribosomes from different species (Fox, 2010). In prokaryotes, the ribosome is called 70S and the two subunits are 50S and 30S, based on their sedimentation properties (“S” is the Svedberg unit). In eukaryotes, the ribosome is called 80S and the two subunits are 60S and 40S (Melnikov et al., 2012). The ribosomal factors involved are also very different between prokaryotes and eukaryotes. For example, there are only 3 canonical initiation factors in bacteria but at least 12 in eukaryotes (Jackson et al., 2010). In the following subsections, only the bacterial translation system will be discussed.

1.5.2 Bacterial translation cycle

Similar to gene transcription, translation can be initiated many times from the same mRNA template, *i.e.*, one polypeptide is synthesized during each translation cycle and the process will repeat many cycles. The translation cycle can be divided into 3 steps: chain initiation, chain elongation and chain termination.

Chain initiation is a process for forming a complex of fMet-tRNA_f^{Met}, mRNA and ribosome, with the help of initiation factors (Laursen et al., 2005). The fMet is the N-formylmethionine residue. On each ribosomal subunit, there are 3 major binding sites for tRNA: A, P, and E. First, the 30S, mRNA and fMet-tRNA_f^{Met} need to assemble correctly. The fMet-tRNA_f^{Met} binds to the P site. The mRNA needs to be properly aligned. The Shine-Dalgarno (SD) sequence (Shine and Dalgarno, 1973) of mRNA is base-paired with the anti-SD sequence of the 16S rRNA. Then the 50S subunit joins to form an initiation complex. During initiation, initiation factor 1 (IF1) was reported to anchor initiation factor 2 (IF2) and 3 (IF3) to enhance their

activities. IF2 helps to capture fMet-tRNA_f^{Met}. IF3 facilitates the binding of fMet-tRNA_f^{Met} to the P site (Hussain et al., 2016).

Chain elongation can be further divided into 3 substeps, decoding, transpeptidation, and translocation (Ramakrishnan, 2002). Once the initiation complex is formed, the ribosome needs to decode the second codon on the mRNA and select the corresponding aminoacyl tRNA. Direct binding of aminoacyl tRNA to the A site is often very slow. In cells, the aminoacyl tRNA forms a ternary complex with GTP and elongation factor EF-Tu, before binding to the A site, which is much faster than the direct binding (Clark and Nyborg, 1997; Nissen et al., 1996). After the ternary complex binds, GTP will be hydrolyzed, and GDP and EF-Tu will be released. Another elongation factor EF-Ts then replaces GDP with GTP (Andersen et al., 2003). The transpeptidation step occurs when the ester group of the peptidyl tRNA is attacked by the amino group nucleophile of the aminoacyl tRNA, resulting in the peptide bond formation on the aminoacyl tRNA (Rodnina et al., 2007). The entire transpeptidation process occurs in the PTC without aid from any protein, or other energy-carrier molecules. The reaction is driven by the energy stored in the ester group on the peptidyl tRNA. Therefore, the ribosome is a ribozyme (Rodnina et al., 2007). The ribosome catalyzes peptide bond formation by correctly positioning substrates within a constrained environment to accelerate the nucleophilic attack. After transpeptidation, the tRNA in the A site becomes a peptidyl tRNA, while the tRNA in the P site becomes unoccupied by amino acids. Then the P site tRNA needs to move to the E site, and the A site tRNA moves to the P site. At the same time, the mRNA also moves together with the tRNAs. Such a process, termed translocation, requires the help from another elongation factor, EF-G, which, together with GTP, binds near to the A site and pushes the A site tRNA to move to the P site (Brilot et al., 2013; Zhou et al., 2013).

The chain will elongate until it is terminated by encountering one of the three stop codons, UAA, UAG, and UGA (Brenner et al., 1967; Brenner et al., 1965). These three stop codons do not have a corresponding tRNA and can be recognized by class I release factors, RF-1 and RF-2 (Capecchi, 1967; Caskey et al., 1971; Vogel et al., 1969). RF-1 recognizes UAA and UAG, and RF-2 recognizes UAA and UGA. After the binding of RF-1 or RF-2, the peptidyl group can only be transferred to water, instead of aminoacyl tRNA, causing the release of the polypeptide. RF1 or RF2 dissociation from the ribosome, is then facilitated by the class II release factor, RF-3 (Freistroffer et al., 1997). Finally the ribosome is released from mRNA by EF-G and the ribosomal recycling factor (RRF) (Hirashima and Kaji, 1973).

1.5.3 Bacterial translation regulation

In bacteria, there are many ways to regulate translation. The ribosome modulation factor (RMF) and hibernation promoting factor (HPF) can dimerize the 70S ribosome into 100S, inhibiting translation (Yoshida and Wada, 2014). The ribosomal silencing factor S (RsfS) can also block the association of the two ribosomal subunits to inhibit translation (Hauser et al., 2012; Li et al., 2015). The structure of mRNA, which can be affected by temperature or binding with many other molecules, can also influence translation efficiency, particularly translation initiation efficiency (Kozak, 2005; Meyer, 2017).

1.5.4 Bacterial translation inhibition by small molecules

Almost every step of translation can be inhibited by chemical compounds (Wilson, 2014). To inhibit chain initiation, edeine, kasugamycin, pactamycin and thermorubin can act on the 30S subunit; orthosomycins avilamycin, evernimicin, and thiostrepton can act on the 50S subunit. For

example, kasugamycin binds to the P-site on the 30S, blocking the mRNA channel (Schuwirth et al., 2006). To inhibit chain elongation, streptomycin, tetracyclines and glycylyclines can inhibit aminoacyl tRNA delivery; blasticidin S, chloramphenicol, lincosamides, oxazolidinones, pleuromutilins, puromycin, streptogramin A and sparsomycin can inhibit peptide bond formation; capreomycin, viomycin, hygromycin B, neomycin, paromomycin, fusidic acid, spectinomycin, and thiostrepton can inhibit tRNA translocation; macrolides, streptogramin B and ketolides can inhibit nascent peptide elongation. Peptidyl-transferase inhibitors, such as blasticidin S, chloramphenicol, puromycin and sparsomycin, can inhibit termination. Fusidic acid can inhibit recycling.

Apart from these chemical compounds, some proline-rich antimicrobial peptides (PrAMPs), such as oncocin, can bind to the PTC and the nascent peptide tunnel to inhibit translation (Roy et al., 2015; Seefeldt et al., 2015).

1.5.5 Previous studies on mycobacterial translation

An unusual feature of *M. tuberculosis* is its capacity to persist, *i.e.*, avoid elimination from human host by immune system or antibiotics treatment (Gomez and McKinney, 2004). During persistence, *M. tuberculosis* remains in a non-replicating state, minimizing most metabolic activities including translation (Chao and Rubin, 2010). Genetic and biochemical studies have revealed that *M. tuberculosis* has incredibly sophisticated approaches to fine-tune translation, such as activating a dormancy survival regulator (DosR) regulon to control ribosome stability in hypoxic mycobacteria (Kumar et al., 2012; Trauner et al., 2012), using toxin-antitoxin pairs to regulate ribosome activity (Sala et al., 2014), or mistranslating to resist inhibitors (Javid et al., 2014). It has also been shown that leaderless mRNA translation is common and robust in

mycobacteria (Shell et al., 2015). These observations suggest the uniqueness of the mycobacteria translation system. However, the underlying structural mechanisms of these featured regulations in mycobacterial translation are not fully understood.

Sequence analyses (Cannone et al., 2002) have revealed that the *M. tuberculosis* ribosome contains species-specific structural features, such as extended or inserted rRNA helices, referred to as rRNA expansion segments. It has recently been shown that, in yeast ribosomes, such rRNA expansion segments could be important for the appropriate assembly and function of the ribosome (Gomez Ramos et al., 2016; Ramesh and Woolford, 2016). However, structurally characterizing these large rRNA expansion segments remains challenging, mainly due to their presence on the periphery of the ribosome, which allows for large conformational variability. Compared with other, better studied bacterial ribosomes (Eyal et al., 2015; Fischer et al., 2015; Sohmen et al., 2015; Wimberly et al., 2000), the *M. tuberculosis* ribosome has a 100-nucleotide rRNA expansion segment, referred to as H54a or the “handle,” in its 23S rRNA. The structure and function of the handle have not been clearly defined, though low resolution structures of *M. smegmatis* 70S ribosome and *M. tuberculosis* 50S ribosomal subunit have been determined (Li et al., 2015; Shasmal and Sengupta, 2012).

Several *M. tuberculosis* rProteins are significantly longer than their counterparts in model bacterial ribosomes, and there are several rProtein paralogs encoded by non-identical genes, which are expressed in response to varying physiological conditions (Cook et al., 2009). Moreover, the *M. tuberculosis* ribosome might have species-specific rProteins that have not been annotated based solely on sequence information. Therefore, the three-dimensional (3D) structure of the *M. tuberculosis* ribosome should reveal these unique features of the *M. tuberculosis* ribosome and allow their roles in translation to be elucidated.

1.6 Gene expression in *M. tuberculosis* may involve different higher order complexes

It has long been known that the two steps of gene expression, *i.e.*, transcription and translation, can be coupled in bacteria. Early evidence came from EM studies of *E. coli* cell lysates, prepared in a special way, called “Miller spread” (Miller et al., 1970). The EM images from this study showed that actively expressing genes have a typical ‘Christmas tree’ like shape. In this ‘tree’, the main trunk is the DNA and the branches are RNA decorated by many high contrast dots. The first dot from the trunk is presumably the RNA polymerase and all following dots are ribosomes. From these studies, a straightforward interpretation is that, under physiological conditions, bacterial gene expression may involve higher order complexes of RNAP and ribosomes. The first higher order complex is a RNAP-ribosome complex, or expressome (Kohler et al., 2017), and the second is a ribosome-ribosome complex, or polysome (Warner et al., 1962). Indeed, evidence has accumulated that suggests a physiological significance and mechanism of both forms. For example, *in vivo* measurement of the transcription and translation rates shows a correlation between these two events (Proshkin et al., 2010).

To explain the coupling between transcription and translation in bacteria, the highly conserved transcription factor NusG was proposed to physically link the RNAP with the ribosome (Burmann et al., 2010). Recently, the expressome of *E. coli* RNAP and 70S ribosome, connected only by mRNA and possibly glutaraldehyde, has been found to have a defined structure (Kohler et al., 2017). The exit site of mRNA on the RNAP was found to be directly interacting with the entry site of mRNA on the ribosome. Despite the discovery of this physiologically relevant configuration, two groups independently found that RNAP and 30S can

directly interact with each other in the absence of any tethering by mRNA or cross-linking (Demo et al., 2017; Fan et al., 2017). These observations showed that the exit site of mRNA on the RNAP is facing to the exit site of mRNA on the 30S. It is possible that both configurations exist and represent different stages of the coupling.

To study the spatial arrangement of polysomes, both negative staining EM and cryo-EM have been used to image polysomes from *E. coli* (Brandt et al., 2009), or wheat germ (Kopeina et al., 2008; Myasnikov et al., 2014), confirming that polysomes can have an ordered structure. Common characteristics can be found from these structures: 1) the mRNA is threading through multiple ribosomes in a helical manner; 2) the nascent peptides emerging from each ribosome have a large distance from each other. Such configuration maximizes the translation efficiency and minimizes the interference between nascent peptides.

By contrast with actively translating ribosomes, the 70S particle can also form hibernating dimers, or 100S, mediated by a “ribosome modulation factor” (RMF) in *E. coli* (Wada et al., 1990). The formation of hibernating 100S was later found to be related to cell growth rate and responsible for fast recovery from stationary phase when fresh media was added (Aiso et al., 2005; Yamagishi et al., 1993). Inhibition of translation by forming 100S is also widely found in other bacteria (Yoshida and Wada, 2014) and other related protein factors were discovered. In addition to dimerization by RMF, there is a separate dimerizing mechanism caused by hibernation promoting factors (HPFs), which can then be divided into short HPFs and long HPFs. The majority of bacteria only have long HPFs, including *T. thermophilus*, *Bacillus subtilis* (*B. subtilis*), and *M. tuberculosis*. Interestingly, 100S formation by long HPFs is not limited to the stationary phase growth, but spans all stages of bacterial growth. During exponential phase of cell growth, polysomes, including disomes, are also formed. Therefore,

during sucrose density gradient separation of the cell lysate, the fraction corresponding to the ribosome dimer fraction can be the disome or hibernating 100S.

1.7 Cryo-electron microscopy

To gain a clear mechanistic understanding of the unique features in *M. tuberculosis* gene expression, we believe that the structural study of the related molecular machineries, including RNAP, ribosome, and their higher order assemblies, is essential. There are three major techniques for structure determination: X-ray crystallography, NMR, and cryo-electron microscopy (cryo-EM). We have chosen cryo-EM as the major technique as it only requires a minimal amount of sample and also tolerates heterogeneity much better than the other two techniques.

A microscope works by magnifying the image of an object. In the traditional light microscope, the visible light can be distorted by optical lenses to form a magnified image. The construction of the electron microscope dates back to the early 20th century. Following the discovery of the electron by Joseph John Thomson in 1897, Louis de Broglie proposed the wave property of electrons and Hans Busch calculated that electrons can be focused by a magnetic field in the 1920s (Busch, 1926). Ernst Ruska and others built the first electron microscope in 1931. Since the wavelength of an electron is much shorter than that of visible light, the resolving power of electron microscopes is much better than light microscopes. However, imaging biological samples was soon found to be challenging. Biological molecules are often very sensitive to high energy electrons, which are sufficiently energetic that they can break chemical bonds and thus damage biological samples. Several methods were developed to overcome the radiation damage problem, including negative staining (Huxley and Zubay, 1961), and glucose

embedding (Henderson and Unwin, 1975). The negative staining agent is usually uranyl acetate, which strongly scatters electrons. The region occupied by the biological samples scatters electrons less. Therefore, an image can be formed with different contrast depending on the degree of staining. However, this method is limited to low resolutions where protein domains may be resolvable but secondary structures can not be visualized. In addition, negative stain preparation requires sample dehydration before imaging, while most biological molecules have to be in hydrated state to maintain their shape. The glucose embedding method utilized glucose to preserve the crystal array of protein samples, which resulted in an amazing subnanometer resolution structure of a membrane protein in the 1970s (Henderson and Unwin, 1975), in which α -helices could be distinguished from each other. Nevertheless, this method is little utilized, probably because of its dependence on crystallization and weak contrast.

The idea of “cryo” electron microscopy, *i.e.*, imaging samples at cryogenic temperatures, was demonstrated in 1974 by Kenneth Taylor and Robert Glaeser (Taylor and Glaeser, 1974). By keeping samples at liquid nitrogen temperature, molecular damage from the incident electron radiation could be minimized. Later on, Jacques Dubochet and colleagues invented a routine method for freezing samples by rapidly plunging them in liquid ethane, which has a large heat capacity, leading to a fast cooling rate. In this new method, the temperature of samples changes so rapidly that ice crystals cannot form and water is directly vitrified into a glass-like state (Dubochet et al., 1988). The elimination of ice crystal formation is important because ice formation can damage biological molecules while vitrification does not. In addition, compared to crystalline ice, vitrified water interferes more uniformly with incident electrons, thus resulting in significantly decreased level of background noise. It is noteworthy that the vitrification method has become very popular and is currently actively used in cryo-EM experiment worldwide. In

this freezing method (**Figure 1.2**), a few microliters of a sample are applied onto a supporting grid and excess water is removed using filter paper. Then the grid is rapidly plunged into liquid ethane, cooled by liquid nitrogen, and the sample layer is vitrified. A typical supporting grid has two layers, a copper grid layer and a holey carbon film layer. Such grids allow the formation of a thin, unsupported aqueous sample layer in the holes, as well as providing a relatively sturdy support. The unsupported sample layer, when imaged, results in images with minimal background noise.

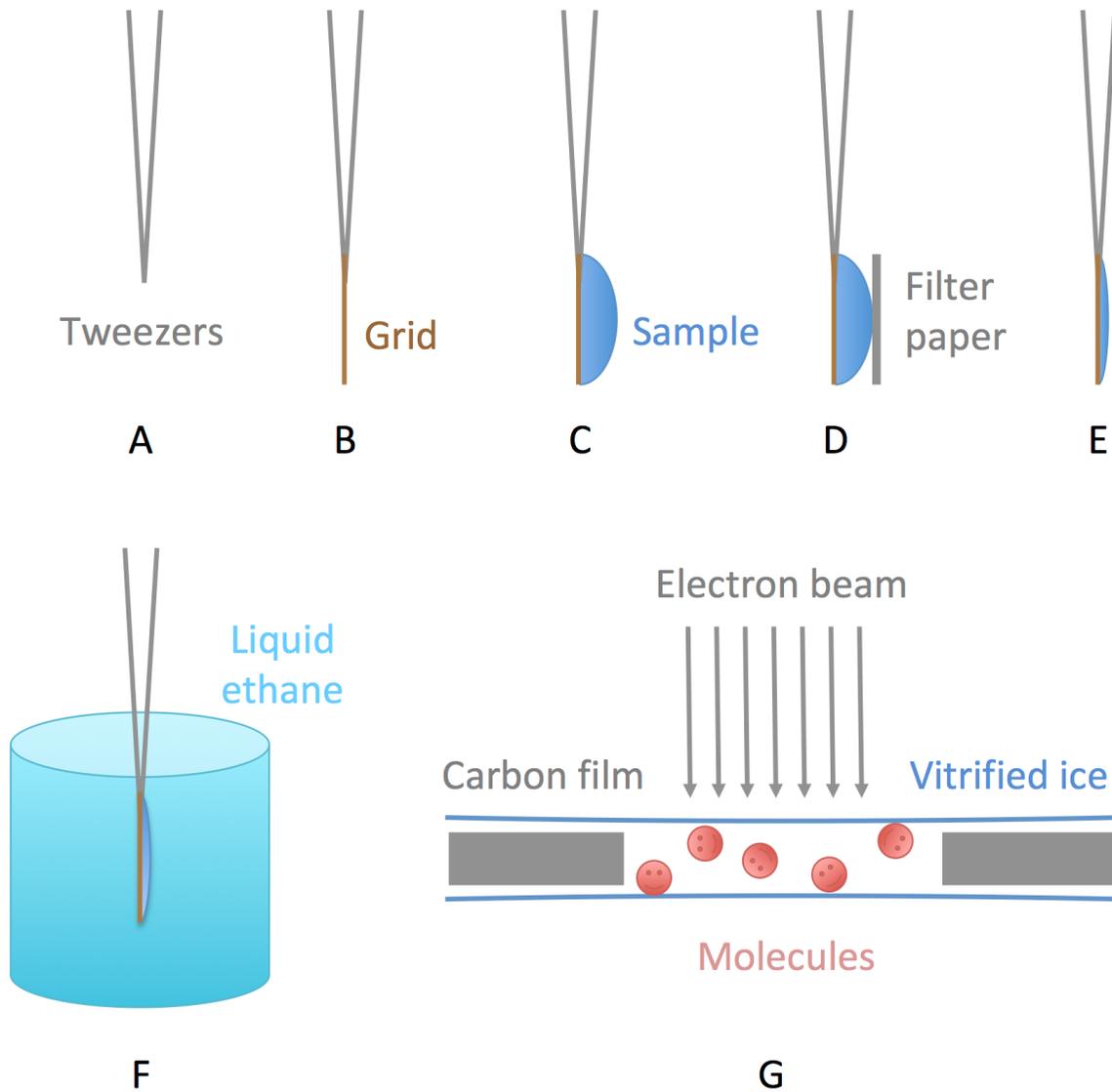


Figure 1.2. Schematic representation of the vitrification method used in cryo-EM Tweezers (A) are used to pick up a grid (B). A sample is applied to the grid (C), and then filter papers absorb excess sample (D), leaving a thin sample layer on the grid (E). Finally the grid is plunged into precooled liquid ethane (F), during which a thin layer of vitrified ice can form (G).

Apart from sample freezing, the quality of electron microscopes has also improved, in terms of beam, vacuum and specimen stage. Traditionally, tungsten or lanthanum hexaboride (LaB₆) filaments were used as the electron source, with subsequent development of the field

emission gun (FEG) (Crewe et al., 1968) that significantly improved coherence. The coherence of electrons includes two aspects, temporal coherence and spatial coherence. The temporal coherence refers to the distribution of electron energy. The narrower the distribution, the better the coherence. The spatial coherence refers to the similarity of the directions of incident electrons. Better coherence improves the phase contrast generated by defocusing. The commercially available FEI and JEOL microscopes provide an excellent vacuum and a stable specimen stage.

Camera technology has also improved, from photographic film to silica-based charge-coupled devices (CCD), and more recently to direct electron detectors. Traditional photographic film is very labor-intensive to use, involving development and scanning. This inconvenience was overcome by CCD cameras, however, at the cost of image quality. The direct electron detectors were developed based on the Complementary Metal Oxide Semiconductors (CMOS) technology (Clough et al., 2014). The new cameras can not only minimize the noise, but also record electrons quickly. The fast capture speed enables the recording of a movie during imaging, instead of a single frame. The movie can be analyzed to correct for specimen drift during imaging.

Apart from these breakthroughs in hardware, advancement in software also plays an important role in revolutionizing cryo-EM. Reconstructing a 3D structure from its 2D projections is not as easy as it first sounds, because one 2D projection is actually a mixture of different levels of a 3D object. The first 3D reconstruction was done in 1968 by David DeRosier and Aaron Klug (De Rosier and Klug, 1968). The object was a phage tail with helical symmetry. The helical symmetry ensures not only the full coverage of different views of the subunit, but also the relationship between these views. Later on, the idea of “single particle” reconstruction was

demonstrated by Joachim Frank and colleagues (Frank, 1975; Frank et al., 1978; Hoppe et al., 1974). In the single particle reconstruction method, the sample does not need to have any symmetry. The particles are expected to have identical shapes, at least to some extent, and have random orientations in the vitrified layer, so that different views of the particles can be obtained. If a large number of particles are imaged, then all orientations can be collected. Such a random-orientation dependent method is distinct from the tomography where a series of images of the same object were collected after tilting the specimen stage. Then it becomes a difficult problem to determine the orientation of the particle in a 2D image. The projection theorem states that the Fourier transform of the 2D image is a slice of the Fourier transform of the 3D object, and the Fourier transforms of the 2D images should have identical common lines in an assumed noise-free case (Crowther et al., 1970). Then the problem can be solved, in theory, by maximizing the cross-correlation between the central sections of the two 2D Fourier transforms. However, in reality, an EM image often has much more noise than signal for many reasons. Fortunately, averaging images after proper alignment can effectively increase the signal to noise ratio, though alignment errors will still exist. To overcome the alignment error problem, Frederick Sigworth introduced the maximum-likelihood algorithm to the EM field (Sigworth, 1998) and Sjors Scheres further implemented it in the Relion software (Scheres, 2012). In the maximum-likelihood approach, the particle in a 2D image is assumed to have all possible orientations, though in reality it only has one during imaging. Each orientation is weighted by its probability before integrating into the 3D map. This method turns out to be very successful and has become the state-of-the-art approach.

In the following chapters, I will present the cryo-EM study of the RNAP (chapter II), ribosome (chapter III) and expressome (chapter IV) from *M. tuberculosis*, as well as ribosome dimers (chapter IV) from *M. smegmatis*.

CHAPTER II

STRUCTURE OF THE RNA POLYMERASE FROM *M. TUBERCULOSIS*

2.1 Introduction

Transcription is an essential process, catalyzed by the RNA polymerase, or RNAP, to synthesize RNA based on a DNA template in all three domains of life. Accumulating data suggests that the transcription system of Mycobacteria may have their own unique features. The RNAP open promotor complex, or RPo, is less stable in Mycobacteria than in *E. coli*, and it is stabilized by the CarD factor, which is absent in *E. coli* (Davis et al., 2015; Stallings and Glickman, 2011). A transcription factor, RbpA, is present in Mycobacteria but absent in *E. coli* and *Thermus spp.* (Hubin et al., 2015; Tabib-Salazar et al., 2013). Moreover, mycobacterial RNAP exhibits better termination efficiency than *E. coli* RNAP, in the case of imperfect U-tracts (Czyz et al., 2014). Sequence-level analysis also revealed lineage-specific insertions in the large subunits (β and β') of RNAP (Lane and Darst, 2010). For example, the Actinobacteria-specific β' In1 of *M. tuberculosis* is different from that of *E. coli* or *T. thermophilus*.

Transcription is also the target of many drugs, including rifampicin, streptolydigin, myxopyronin, and salinamide. Interestingly, all these inhibitors bind to the catalytic region of RNAP. Particularly, rifampicin is a first-line tuberculosis drug that is frequently used to treat TB. Despite the importance of rifampicin and RNAP in TB therapy, the detailed mechanism of drug inhibition has long been understood based on structures of RNAP from non-pathogenic model organisms, instead of that from *M. tuberculosis*. Given that the sequences of RNAP subunits are not identical, it is possible that the understanding obtained from other species does not apply to *M. tuberculosis*. Therefore, solving the structure of RNAP from *M. tuberculosis* may advance

our understanding of drug inhibition. In this chapter, I will present our cryo-EM study of *M. tuberculosis* RNAP.

2.2 Materials and Methods

2.2.1 Expression and purification of *M. tuberculosis* RNAP and RbpA

M. tuberculosis RNAP and RbpA were purified by Dr. Qingan Sun. Briefly, genes of *M. tuberculosis* H37Rv β and β' subunits were cloned into a pET-Duet vector, α and ω in pACYC-DUET, σ^A in pET28b, RbpA in pET28b. A His-tag was fused to the N-terminus of the α subunit, and the N-terminus of the RbpA. All the plasmids for RNAP genes were co-transformed, or the plasmid for the RbpA gene was transformed, into *E. coli* BL21(DE3) cells. The cell pellet was resuspended in buffer A (50 mM Tris/HCl, 100 mM NaCl, 10% glycerol, pH 7.6), and passed through a microfluidizer. The lysate was loaded on the gravity column with Ni-NTA, washed with buffer A, and eluted with buffers containing different imidazole concentrations by mixing buffer A and buffer B (50 mM Tris/HCl, 100 mM NaCl, 10% glycerol, 500 mM imidazole, pH 7.6). The fractions containing RNAP, or RbpA, were pooled and dialyzed against low-salt buffer before further purification on a HiTrap Q high performance ion exchange columns (buffer C: 25 mM Tris/HCl, 50 mM NaCl, 5% glycerol, 0.1 mM EDTA, 5 mM DTT, pH 7.6; buffer D: 50 mM Tris/HCl, 1 M NaCl, 5% glycerol, 0.1 mM EDTA, 5 mM DTT, pH 8.0). At last, the purified *M. tuberculosis* RNAP, or RbpA, was polished, and the buffer was exchanged with Superdex 200 column (buffer E: 20 mM Tris/HCl, 200 mM NaCl, 2% glycerol, 5 mM DTT, pH 8). During buffer-exchange, the protein was concentrated to ~10 mg/mL. The sample was fractionated into small volumes and quickly frozen by liquid nitrogen before storage at -80°C.

2.2.2 Cryo-electron microscopy

The RNAP and RbpA were thawed on ice and mixed at either 1:1 or 1:4 molecular ratio. Both mixed samples were incubated at room temperature for 30 minutes. Then both samples were diluted to ~0.1 mg/mL with buffer E, and 3 μ L of the diluted sample was applied to a C-Flat 1.2/1.3 holey carbon grid, followed by plunge-freezing in Mark III Vitrobot (FEI company, Netherlands). The temperature and relative humidity were set at 4°C and 100%, respectively. The grids were screened and imaged under an FEI Tecnai TF20 cryo electron microscope with a field emission gun (FEI company, Netherlands) operated at 200 kV. Data were recorded on a Gatan K2 Summit (Gatan, Pleasanton CA) direct detection camera in the super-resolution electron counting mode. A nominal magnification of 19,000 x yielded a pixel size of 1.87 Å on the specimen. The beam intensity was adjusted to a dose rate of 10 electrons per pixel per second on the camera. An 82-frame movie stack was recorded, with a 0.2 second exposure per frame. In total, 1803 movie stacks were automatically collected using the automation software, SerialEM (Mastronarde, 2005), in several non-consecutive sessions.

2.2.3 Image processing

The 1803 movie stacks were aligned using MotionCor2 (Zheng et al., 2017), with customized automatic parameter optimization by p3motioncor2opt.py (See **APPENDIX III** for details). The movie stacks in which the drift was too fast were discarded at this step. The defocus value of each image stack was determined using GCTF (Zhang, 2016). Particles were automatically picked using Gautomatch (developed by Dr. Kai Zhang, MRC Laboratory of Molecular Biology in UK). After considerations in defocus value and image quality, the number of good micrographs was further screened to 866. From these 866 micrographs, 668,627 particles

were initially selected and used as input for the next classification step. 2D classification was performed in Relion2.0 (Kimanius et al., 2016). Then a rough 3D auto-refine was performed, with an *E. coli* RNAP structure (PDB id: 4IGC) (Murakami, 2013) as the initial model, to center the particles according to the x and y shifts after 3D refinement. After centering, duplicate particles were removed. After all the above screening steps, 364,201 particles were considered good and used as the input for the final refinement. The particles used for the final refinement were binned 1.5 times, which was selected considering both speed and the resolution limit. The overall resolution was estimated according to the gold-standard Fourier shell correlation (Scheres and Chen, 2012). 3D classification was also performed, which revealed little meaningful structural heterogeneity and resulted in no resolution improvement. The particles were also split into two datasets based on the molecular ratios (1:1 or 1:4) to do independent reconstructions and no significant difference was observed.

2.2.4 Molecular modeling

Initially, the molecular modeling was based on the crystal structure of the *T. aquaticus* RNAP core complex (PDB id: 1HQM) (Minakhin et al., 2001), because the *M. tuberculosis* RNAP was found to be a core complex and 1HQM was the only available high-resolution structure of bacterial RNAP core complexes. Later on, the crystal structures of *M. tuberculosis* RNAP were published (Lin et al., 2017) and the structure of the RPo (PDB id: 5UHA) was chosen as the starting model. Considering the low-resolution nature of the EM map, the MDFF software was chosen (Trabuco et al., 2008). The crystal structure was first rigid-body fit into the EM map based on the platform and the β pincer regions in UCSF Chimera (Pettersen et al., 2004). Then a normal MDFF run was performed with the following restraints: 1) secondary

structure restraints, 2) cis/trans peptide restraints, 3) chirality restraints, 4) domain restraints. The domain restraints were specified as follows: 1) the platform and β pincer regions ($\alpha_2\omega$, β residue 28-1033, β' residue 425-1205) were defined as one domain and the $C\alpha$ -trace was treated as a rigid-body; 2) the β' pincer region (β residue 1069-1153, β' residue 3-416, 1220-1281) was defined as one domain and the $C\alpha$ -trace was treated as a rigid-body; 3) the connecting region (β residue 1034-1069, β' residue 417-424, 1206-1219) was not restrained. Multiple rounds of MDFF run were performed and between each round, one particular region (β residue 1034-1069) was manually adjusted with caution to better fit the density.

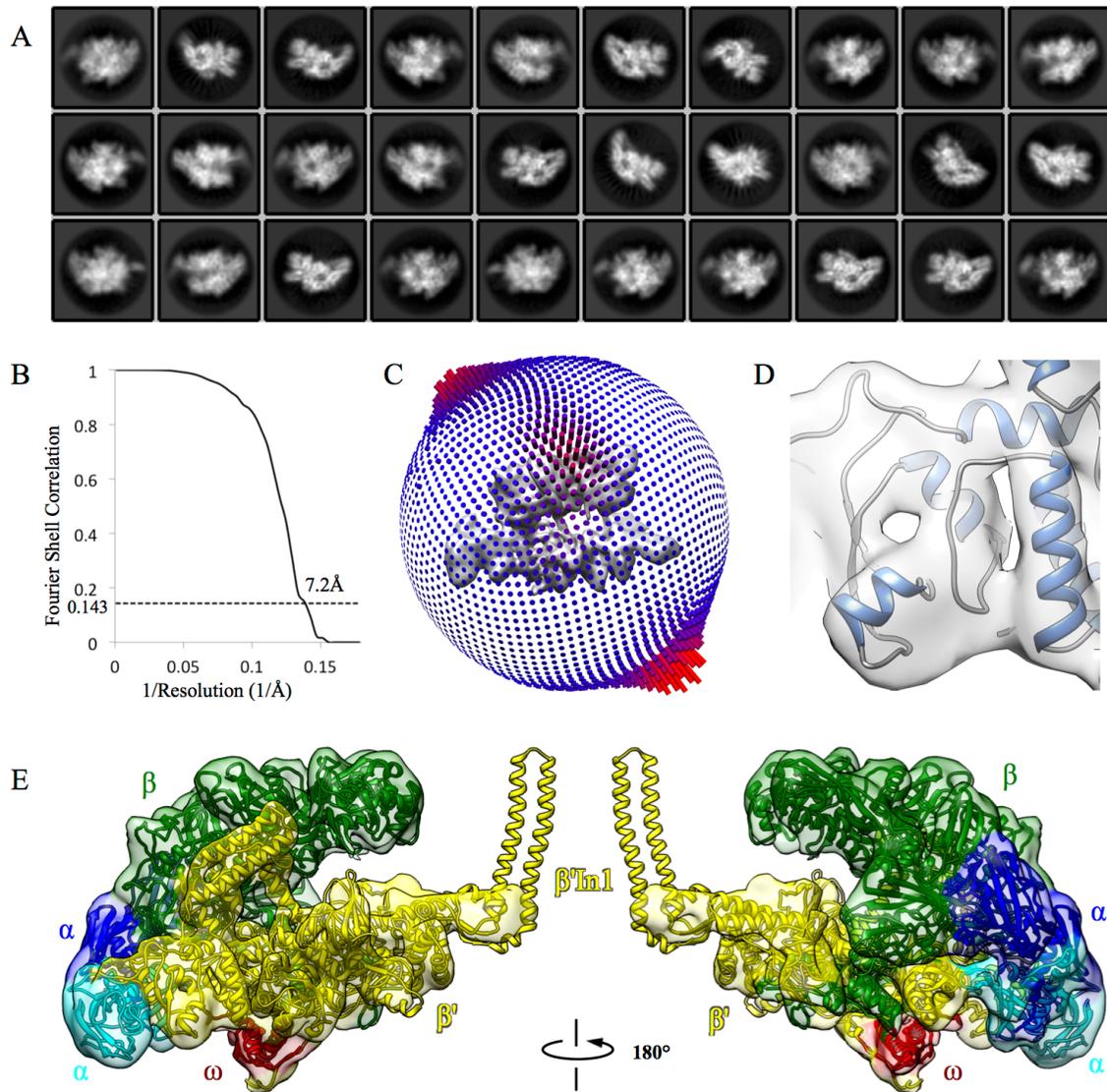


Figure 2.1. Cryo-EM structure of *M. tuberculosis* RNAP core complex

(A) Representative 2D class averages of the imaged particles.

(B) Fourier shell correlations of the cryo-EM reconstruction of *M. tuberculosis* RNAP core complex.

(C) Distribution of the particles orientations. The inner grey density is the core complex viewing through the channel between the two pincers. On the outer layer, each bar represent one orientation and its length (and color) corresponds to the number of particles in this orientation. Red bars represent preferred orientations.

(D) A representative view of the matching between the cryo-EM map and model. The densities of individual α -helices (light blue color) are discernible.

(E) Overview of the core complex. Both the model and map are colored by subunits. The two α subunits are colored by blue and cyan, respectively, and β green, β' yellow, and ω red. The two-helix bundle out of density is β' In1. No additional density is available to model σ^A or RbpA.

2.3 Results

2.3.1 Overview of the cryo-EM structure of *M. tuberculosis* RNAP core complex

The imaged particles from the 1:1 mixed samples were initially presumed to be the complex of the *M. tuberculosis* RNAP holoenzyme bound with RbpA. However, no density can be clearly assigned to RbpA or even the σ^A factor. Therefore, the molecular ratio of RNAP versus RbpA was increased to 1:4, but there was no improvement for the visibility of RbpA or σ^A . Finally the two datasets were merged together for data processing. The reference-free 2D classification showed clear shapes and multiple views of the RNAP particles (**Figure 2.1A**). The 3D refinement generated a density map at 7.2Å resolution, based on the gold-standard Fourier shell correlation (FSC, **Figure 2.1B**) (Scheres and Chen, 2012). The distribution of the particles orientations reviewed that the particles have two preferred orientations, but all other orientations are evenly distributed without any missing angles (**Figure 2.1C**). The protein secondary structure, α -helix, is clearly discernible as sausage-like density in the map (**Figure 2.1D**), confirming the consistency between the map quality and the claimed sub-nanometer resolution. The cryo-EM structure represents the core complex of the *M. tuberculosis* RNAP, in contrast to the initially expected holoenzyme-RbpA complex, because the density in the σ^A factor region is essentially missing and no density can be modeled into the RbpA either (**Figure 2.1E**). The RbpA is known to bind to RNAP via the σ_2 region of σ factors (Bortoluzzi et al., 2013; Tabib-Salazar et al., 2013), so missing RbpA, even at 1:4 molecular ratio, can be explained by the absence of the σ factor. All five subunits can be modeled into the density, while some peripheral flexible fragments are missing their density, including the Actinobacteria-specific β' In1. The absence of β' In1 indicates that it is highly flexible in the core complex.

2.3.2 The open conformation of the claw in the core complex

Despite the large number of available bacterial RNAP structures, currently there is only one high-resolution structure of the RNAP core complex, from *T. aquaticus*, solved by X-ray crystallography (PDB id: 1HQM) (Minakhin et al., 2001). Many comparisons have been made between this single structure with many others (Darst et al., 2002; Mukhopadhyay et al., 2008; Murakami et al., 2002b). Interestingly, the *M. tuberculosis* core complex solved by cryo-EM in this report shows slight conformational differences from the *T. aquaticus* core complex crystal structure (**Figure 2.2**). The differences are present in almost all domains around the channel between β and β' . The directions of the differences are not always the same for neighbouring domains. For example, all four conformationally different domains shift in opposite or perpendicular directions (**Figure 2.2A**), while the β' pincer region shown in Figure 2.2B has a uniformly right shift. The reasons underlying these differences are not clear; however, the β' In1 apparently has an effect on its neighboring domains.

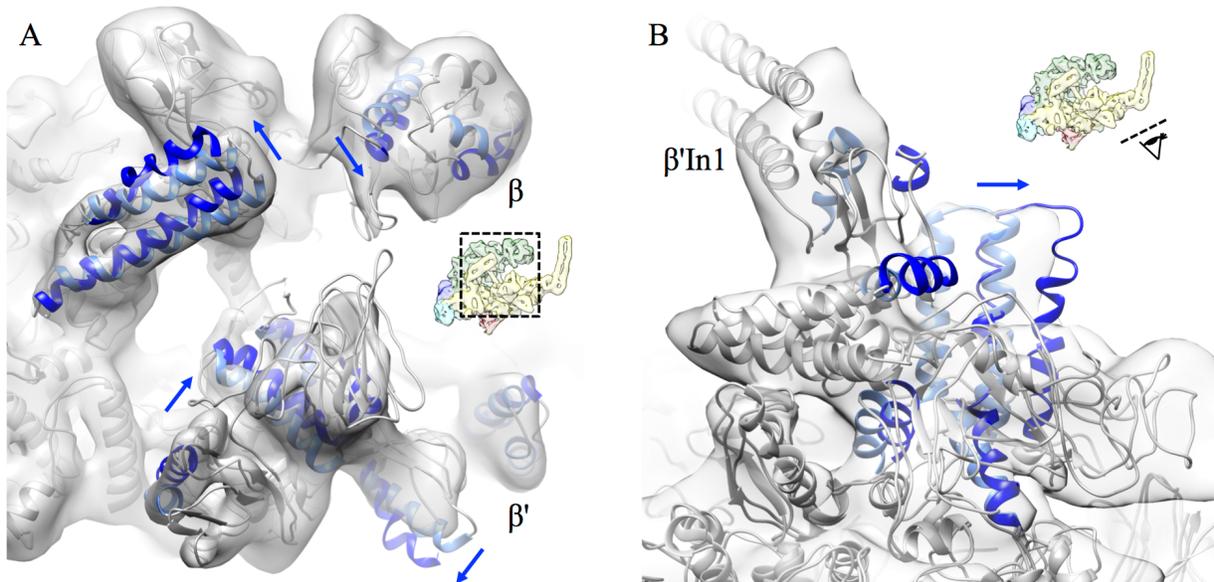


Figure 2.2. Conformational difference between *M. tuberculosis* core complex and *T. aquaticus* core complex

For clarity, only the α -helices in conformationally different regions are colored as light blue and blue for *M. tuberculosis* and *T. aquaticus*, respectively, while all others are colored as gray. The direction of the conformational difference is indicated by arrows.

(A) View through the channel between β and β' , as indicated by the box in the inset.

(B) View from beneath the β' , as indicated by the eye cartoon in the inset. The β' In1 is labeled.

The recently available structure of RNAP from *M. tuberculosis* in the RPo form provides a great chance for identifying interesting conformational changes between RPo and the core. Indeed, upon comparison with the *M. tuberculosis* RPo structure, the claw, or the β' pincer, shows an outward opening or rotation (**Figure 2.3**). The rotating angle is around $\sim 10^\circ$ with the switch region as the hinge, and the distal end travels $\sim 10\text{\AA}$. Although the overall direction of the opening motion is consistent with the previous comparisons between the *T. aquaticus* core complex and *E. coli* holoenzyme (Darst et al., 2002), or yeast RNAP (Mukhopadhyay et al., 2008), the degree of the rotation and the length of displacement decreases to around half. This minor difference may be caused by the comparison of RNAP structure from different species. The consistent direction of movement means that when the channel between the two pincers are empty, the opening of the channel is relatively large; when the channel is occupied by σ factors and DNA template, the opening becomes smaller. Apparently, the large opening is to accommodate the incoming σ factors and DNA template, while the small opening is to increase the binding affinity of DNA.

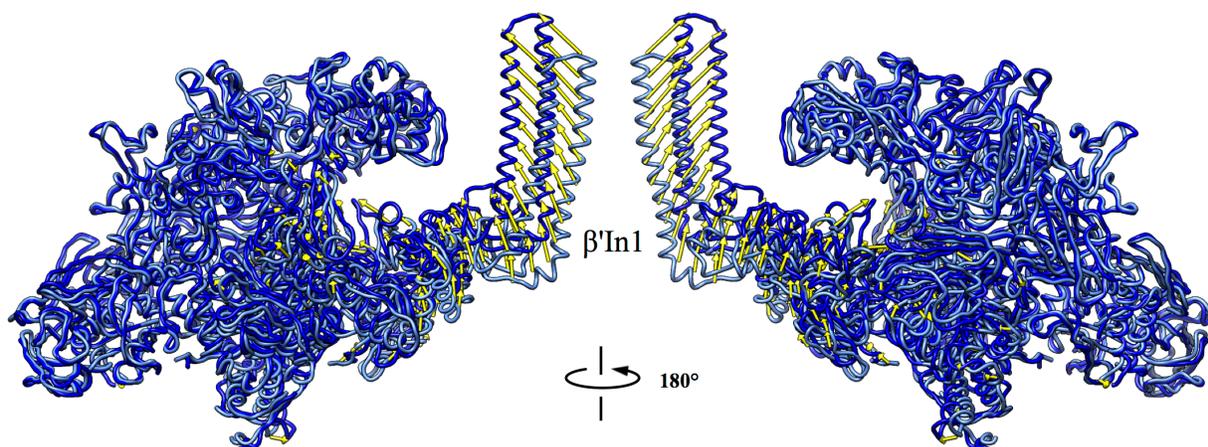


Figure 2.3. Conformational change from *M. tuberculosis* core complex to *M. tuberculosis* RPo complex

The RNAP is viewed from the same orientations as in Figure 2.1E. The RPo is superimposed onto the core based on the $\alpha_2\omega\beta$ region. The core complex is colored as light blue, and the RPo blue. The movements are indicated by yellow arrows. For clarity, four adjacent residues are treated as one group and the average distance of this group in the two structures is calculated. The length of an arrow is defined by the average distance, and the starting point of an arrow is positioned at the C α atom of the fourth residue in the four-residue group. In addition, arrows shorter than 3.5Å are not displayed.

2.3.3 A unique conformation near the active center

A common feature of multi-subunit RNAP is the so-called bridge helix. It is a long α -helix in the β' subunit that connects or bridges the two pincers. In *M. tuberculosis*, it spans from residue 846 to 880. Besides the bridge helix, there is another fragment from the β subunit that connects the two pincers (**Figure 2.4A**) and thus termed “bridge 2”, or “B2”, in analogy to the naming methods of the bridges in ribosomes. Unlike the bridge helix, B2 only has a very short α -helix. All other regions of B2 are loops, indicating the high flexibility of B2. In *M. tuberculosis*, B2 spans from residue 1033 to residue 1069. Direct docking of the *M. tuberculosis* RPo shows that the B2 is out of density; however, there is a nearby density that connects the two pincers and is not occupied by any models (**Figure 2.4B**). Therefore, it is tempting to model the B2 into this

empty density, since the empty density does have a typical sausage-like shape at subnanometer resolutions. Moreover, no alternative interpretations seem reasonable. First, the sausage-like density cannot belong to RbpA or σ , because the binding sites of RbpA or σ have been well characterized. Second, the density cannot belong to DNA, because no DNA was added and the binding of DNA requires σ . Third, the density cannot be ions, because density of ions are usually much smaller and invisible at the current resolution.

Assuming the density assignment to B2 is correct, B2 should shift $\sim 10\text{\AA}$ toward the active center. The new position of the B2 helix is occupied by the template strand of DNA in the RPo (**Figure 2.4C**). Usually it is common to observe conformational changes on the periphery of a complex, however, the conformational change of the B2 helix is within the center of the RNAP and is very close to the active center. Actually, in the new conformation, the rifampicin and the B2 backbone (side chains are not considered, due to the low resolution of the current EM structure) are as close as 5\AA (**Figure 2.4D**). However, no mutation on B2 helix was reported in rifampicin mutants, suggesting that the potential interaction between B2 helix and rifampicin is not critical for drug binding. It is possible to design rifampicin derivatives to lock B2 helix in this new conformation, so that the DNA template binding would be blocked (**Figure 2.4C**).

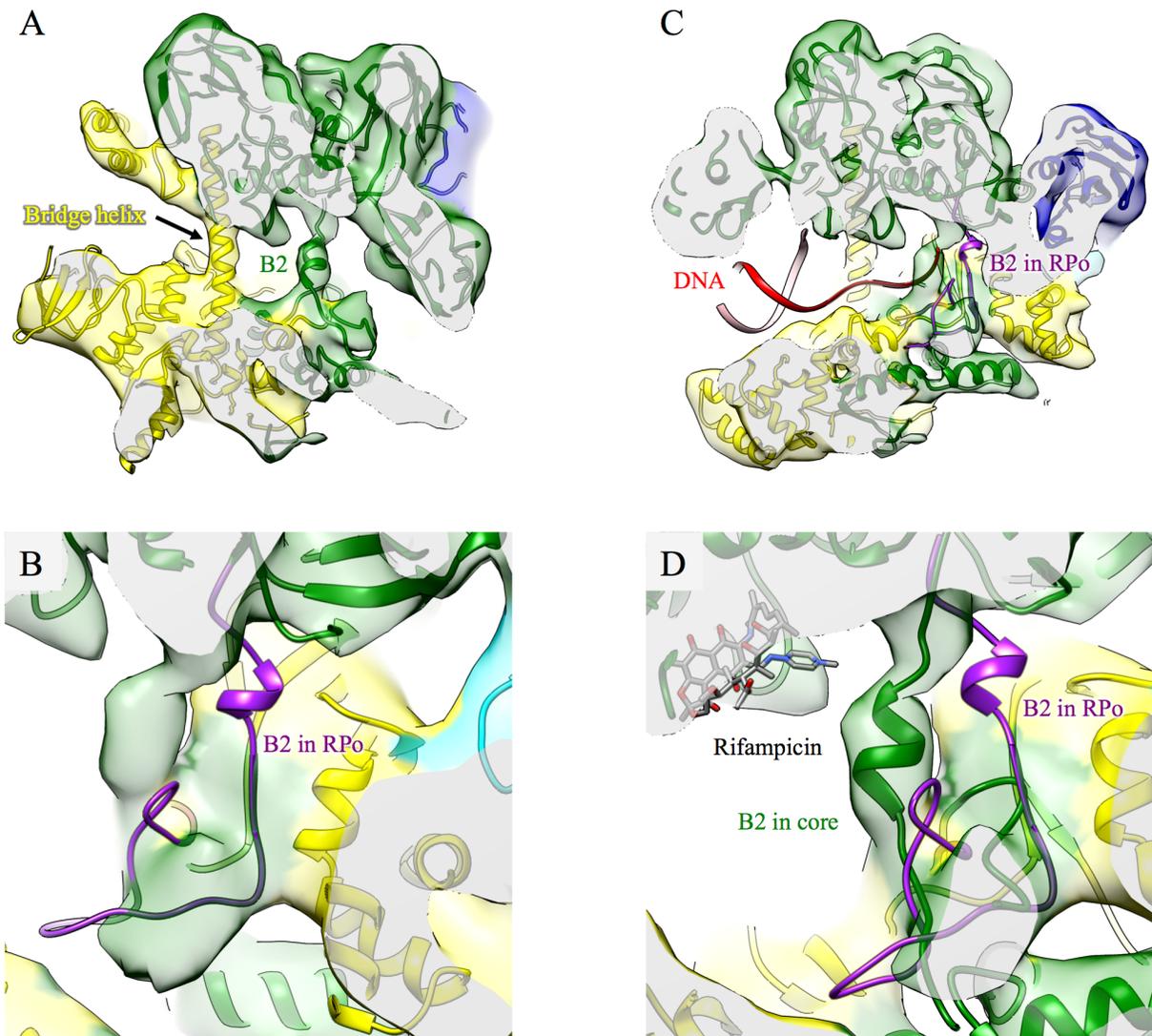


Figure 2.4. The unique conformation of B2

The color code of each subunit is the same as in Figure 2.1E, *i.e.*, α blue or cyan, β green, β' yellow.

(A) A cut-through view of the core complex, showing the two bridges, bridge helix and B2.

(B) Comparison between the EM map and the RPo model, to show the B2 in RPo (colored as purple) is out of density and there is empty density nearby.

(C) A cut-through view of the core complex with a DNA model (colored as red), to show the path of DNA. The longer strand is the template strand and its right end is the 3' terminus.

(D) Comparison between the core complex model and the RPo model, with rifampicin docked in its binding site, to show the close proximity between rifampicin and the B2 in core complex. The rifampicin is colored by elements, *i.e.*, carbon grey, nitrogen blue, oxygen red.

2.4 Discussion

We set out to solve the structure of *M. tuberculosis* RNAP holoenzyme in complex with RbpA. However, neither the σ factor nor the RbpA can be observed. In retrospect, this failure may be caused by the following reasons. First, the affinity between the core complex and the σ^A is low without the stabilization by either RbpA or a proper DNA promoter in *M. tuberculosis*. Second, to avoid the overlapping of particles in cryo-EM, the protein concentration is often diluted to a very low level, such as ~ 0.1 mg/mL, which inevitably dissociates low-affinity complexes.

Nevertheless, the core complex structure provides interesting insights into the unique features of *M. tuberculosis* RNAP. First, the high flexibility of β' In1 in the core complex and its relatively high stability in the RPo suggests that it is stabilized by the σ factor and/or the DNA template. Therefore, β' In1, σ , and DNA may mutually stabilize each other, supporting the notion that the role of β' In1 is to trap DNA and stabilize RPo (Lin et al., 2017), instead of blocking DNA binding (Hubin et al., 2017b). Second, the core complex structure confirmed a similar, yet not exactly the same, open conformation in *M. tuberculosis* to accommodate the incoming DNA. Third, the core complex structure indicated a new position of the B2 helix on β subunit, which provides possibilities to design rifampicin derivatives.

Finally, the current study highlights the therapeutic potential of the B2 helix, and further studies are definitely necessary. For example, the resolution of the EM structure under the current sample preparation condition, needs to be improved to at least 4\AA , so that the bulky side chains can be visualized and the entire backbone can be traced with certainty. Second, the structure and affinity relationship of B2 and rifampicin needs to be answered by mutagenesis experiments.

CHAPTER III

STRUCTURES OF THE RIBOSOME FROM *M. TUBERCULOSIS*

3.1 Introduction

An unusual feature of *M. tuberculosis* is its capacity to persist, *i.e.*, avoid elimination from the human host by the immune system or antibiotics treatment (Gomez and McKinney, 2004). During persistence, *M. tuberculosis* remains in a non-replicating state, minimizing most metabolic activities including translation (Chao and Rubin, 2010). Genetic and biochemical studies have revealed that *M. tuberculosis* has incredibly sophisticated approaches to fine-tune translation, such as activating a dormancy survival regulator (DosR) regulon to control ribosome stability in hypoxic Mycobacteria (Kumar et al., 2012; Trauner et al., 2012), using toxin-antitoxin pairs to regulate ribosome activity (Sala et al., 2014), or mistranslating to resist inhibitors (Javid et al., 2014). It has also been shown that leaderless mRNA translation is common and robust in Mycobacteria (Shell et al., 2015). These observations suggest the uniqueness of the Mycobacteria translation system. However, the underlying structural mechanisms of these featured regulations in mycobacterial translation are not fully understood.

Most of this chapter is reprinted with permission from “Structural insights into species-specific features of the ribosome from the human pathogen *Mycobacterium tuberculosis*.” by Kailu Yang, Jeng-Yih Chang, Zhicheng Cui, Xiaojun Li, Ran Meng, Lijun Duan, Jirapat Thongchol, Joanita Jakana, Christoph M. Huwe, James C. Sacchettini and Junjie Zhang, *Nucleic Acids Res* 2017 45(18), 10884-10894, Copyright 2017 by Oxford University Press.

Author contributions: XL purified the *M. tuberculosis* and *M. smegmatis* ribosomes with the assistance from ZC and measured the IC₅₀ values for linezolid and LZD-114 in an *in vitro* translation assay. KY collected the cryo-EM data of the *M. tuberculosis* ribosomes with the assistance of ZC, JT, and JJ. KY performed the data processing for the *M. tuberculosis* ribosomes. ZC collected and processed the cryo-EM data of the *M. smegmatis* ribosome. KY and JC built the ribosome models with the assistance of RM and LD. CH provided LZD-114. JCS and JZ aided in the analysis of the results.

Sequence analyses (Cannone et al., 2002) have revealed that the *M. tuberculosis* ribosome contains species-specific structural features, such as extended or inserted rRNA helices, referred to as rRNA expansion segments. It has recently been shown that, in yeast ribosomes, such rRNA expansion segments could be important for the appropriate assembly and function of the ribosome (Gomez Ramos et al., 2016; Ramesh and Woolford, 2016). However, structurally characterizing these large rRNA expansion segments remains challenging, mainly due to their presence on the periphery of the ribosome, which allows for large conformational variability. Compared with other, better studied bacterial ribosomes (Eyal et al., 2015; Fischer et al., 2015; Sohmen et al., 2015; Wimberly et al., 2000), the *M. tuberculosis* ribosome has a 100-nucleotide rRNA expansion segment, referred to as H54a or the “handle,” in its 23S rRNA. The structure and function of the handle have not been clearly defined, though low resolution structures of *M. smegmatis* 70S ribosome and *M. tuberculosis* 50S ribosomal subunit have been determined (Li et al., 2015; Shasmal and Sengupta, 2012).

Several *M. tuberculosis* rProteins are significantly longer than their counterparts in model bacterial ribosomes, and there are several rProtein paralogs encoded by non-identical genes, which are expressed in response to varying physiological conditions (Cook et al., 2009). Moreover, the *M. tuberculosis* ribosome might have species-specific rProteins that have not been annotated based solely on sequence information. Therefore, the three-dimensional (3D) structure of the *M. tuberculosis* ribosome should reveal these unique features of the *M. tuberculosis* ribosome and allow their roles in translation to be elucidated.

3.2 Materials and Methods

3.2.1 Purification of *M. tuberculosis* and *M. smegmatis* ribosomes

M. tuberculosis and *M. smegmatis* ribosomes were purified as previously described (Li et al., 2015). Briefly, MC²7000 *M. tuberculosis* or MC²155 *M. smegmatis* cells were grown in 7H9 medium supplemented with 0.5% glycerol, 0.05% Tween-80, and OADC (BD) at 37°C until they reached an OD₆₀₀ of around 1. All subsequent procedures were performed at 4°C. Harvested cells were lysed in a bead beater (BioSpec) in lysis buffer (20 mM Tris-HCl pH 7.5, 100 mM NH₄Cl, 10 mM MgCl₂, 0.5 mM EDTA, 6 mM β-mercaptoethanol). *M. tuberculosis* and *M. smegmatis* ribosomes were purified according to modified protocols. Cell lysates were clarified by centrifugation at 30,000 g for 1 hour. The supernatant was pelleted in sucrose cushion buffer (20 mM HEPES pH 7.5, 1.1 M sucrose, 10 mM MgCl₂, 0.5 M KCl, and 0.5 mM EDTA) at 40,000 rpm in a Beckman Type 45Ti rotor for 20 hours. The pellet was resuspended in a buffer containing 20 mM Tris-HCl pH 7.5, 1.5 M (NH₄)₂SO₄, 0.4 M KCl, and 10 mM MgCl₂. The suspension was then applied to a hydrophobic interaction column (Toyopearl Butyl-650S) and eluted with a reverse ionic strength gradient from 1.5 M to 0 M (NH₄)₂SO₄ in a buffer containing 20 mM Tris-HCl pH 7.5, 0.4 M KCl, and 10 mM MgCl₂. The eluted ribosome peak was changed to either a reassociation buffer (5 mM HEPES-NaOH, pH 7.5, 10 mM NH₄Cl, 50 mM KCl, 10 mM MgCl₂, and 6 mM β-mercaptoethanol) or dissociation buffer (20 mM Tris-HCl, pH 7.5, 2 mM MgCl₂, 150 mM NH₄Cl, 50 mM KCl, and 6 mM β-mercaptoethanol) then concentrated before applying to a 10%-40% linear sucrose gradient centrifuged in a Beckman SW28 rotor at 19,000 rpm for 19 hours. The 70S and 50S fractions were concentrated to about A₂₆₀ = 300 after removal of the sucrose.

3.2.2 Cryo-electron microscopy

In total five samples were prepared for cryo-EM, including *M. tuberculosis* 50S with 10 mM Mg²⁺, *M. tuberculosis* 50S with 1 mM Mg²⁺, *M. tuberculosis* 70S with capreomycin, *M. tuberculosis* 70S without capreomycin, and *M. smegmatis* 70S, for which the ribosome concentrations were diluted to 50 nM. *M. tuberculosis* 50S with 1 mM Mg²⁺ was diluted with the M1 dilution buffer (5 mM HEPES-Na, pH 7.5, 10 mM NH₄Cl, 50 mM KCl and 1 mM MgCl₂). The other samples were diluted with the M10 dilution buffer (5 mM HEPES-Na, pH 7.5, 10 mM NH₄Cl, 50 mM KCl and 10 mM MgCl₂). For *M. tuberculosis* 50S with 10 mM Mg²⁺, 10 μM LZD-114 was added and then incubated for 30 min at 25 °C. For *M. tuberculosis* 70S with capreomycin, 10 μM capreomycin was added and then incubated for 30 min at 25 °C. A total of 3 μL of the sample was applied to a C-Flat 1.2/1.3 holey carbon grid, at 16°C with 100% relative humidity, and vitrified using a Vitrobot Mark III (FEI company, The Netherlands).

The grids for *M. tuberculosis* 50S with 10 mM Mg²⁺ were imaged under a JEM-3200FSC transmission electron microscope operated at 300 kV. Data were recorded on a Gatan K2 Summit (Gatan, Pleasanton CA) direct detection camera in the super-resolution electron counting mode. A nominal magnification of 40,000 x was used, yielding a pixel size of 0.82 Å on the specimen. The beam intensity is adjusted to a dose rate of 8 electrons per pixel per second on the camera. A 30-frame movie stack was recorded for each exposure of 6 seconds, at 0.2 second per frame.

The grids for *M. tuberculosis* 70S with capreomycin, and *M. smegmatis* 70S were imaged under an FEI Tecnai TF20 cryo electron microscope with a field emission gun (FEI company, Netherlands) operated at 200 kV. Data were recorded on a Gatan K2 Summit (Gatan, Pleasanton CA) direct detection camera in the super-resolution electron counting mode. A nominal

magnification of 29,000 x yielded a pixel size of 1.25 Å on the specimen. The beam intensity was adjusted to a dose rate of 10 electrons per pixel per second on the camera. A 33-frame movie stack was recorded, with a 0.2 second exposure per frame. Additional data for *M. tuberculosis* 50S with 1 mM Mg²⁺ and *M. tuberculosis* 70S without capreomycin was collected using the similar procedure at a nominal magnification of 19,000 x, yielding a pixel size of 1.87 Å on the specimen. All related information for data collection was summarized in **Table 3.1**.

3.2.3 Image processing

The collected image stacks were aligned and summed using Unblur (Grant and Grigorieff, 2015). The defocus value of each image stack was determined using CTFFIND4 (Rohou and Grigorieff, 2015). Good image stacks were selected based on the results from both programs. From the summed micrographs of the good image stacks, particles were semi-automatically picked using the *Erase* and *Swarm* tools of *e2boxer.py* in EMAN2 (Tang et al., 2007). 2D and 3D classifications were performed in Relion1.4 (Bai et al., 2015) to get cleaner and more homogenous particles, then the particles were refined and polished (Scheres, 2014) to get the final reconstructions. Overall resolutions were estimated according to the gold-standard Fourier shell correlation (Scheres and Chen, 2012). Local resolutions were calculated using *blocres* from the Bsoft package (Heymann and Belnap, 2007). To further improve the map quality of the *M. tuberculosis* 70S SSU, masked classification with signal subtraction (Bai et al., 2015) followed by masked refinement was performed in Relion1.4.

Table 3.1. Data collection and processing

Sample	<i>Mtb</i> 50S		<i>Mtb</i> 70S		<i>Msm</i> 70S
	10 mM Mg ²⁺	1 mM Mg ²⁺	10 μ M capreomycin	No capreomycin	
Microscope	JEM-3200FSC	TF20	TF20	TF20	TF20
Voltage (kV)	300	200	200	200	200
Nominal magnification (x)	40,000	19,000	29,000	19,000	29,000
Pixel size (Å)	0.82	1.87	1.25	1.87	1.25
Electron dose rate (e ⁻ /pixel/s)	8	10	10	10	10
Number of frames	30	60	33	40	33
Total exposure time (s)	6	15	6.6	8	6.6
Total electron dose (e ⁻ /Å ²)	71	43	42	23	42
Number of particles	99,285	111,219	184,330	107,407	106,000
Map resolution (Å)	3.7	6.0	4.0	NA	7.1

To examine the conformational dynamics of the handle, the *M. tuberculosis* 70S particles collected at 19,000 x magnification were merged with the dataset of *M. tuberculosis* 70S collected at 29,000 x magnification. Then the merged dataset was analyzed using masked classification with signal subtraction (Bai et al., 2015) followed by unmasked refinement of the entire *M. tuberculosis* 70S for each class in Relion1.4. Briefly, the particles were downscaled by a factor of 4 and refined into one consensus density map. Then a mask around the handle region was generated. The projections of the consensus map (except the handle region) were subtracted

from the polished images using the orientation parameters from the consensus refinement, which generated new images of the handle alone. Then the new images were classified without alignment into forty good classes, each of which was then refined to get a final reconstruction. We examined the particle distribution over different classes and found no significant difference between the two datasets of *M. tuberculosis* 70S with or without the capreomycin.

3.2.4 Molecular modeling

The following modelling procedures were performed first in the highest resolution maps, *i.e.*, the LSU of the *M. tuberculosis* 50S and the locally refined SSU of the *M. tuberculosis* 70S. The sequences of *M. tuberculosis* 5S, 16S, and 23S rRNA were obtained through NCBI Gene with the Gene ID 2700459, 2700429, and 2700466, respectively. Template-based comparative modelling was performed using modeRNA (Rother et al., 2011). The *E. coli* 70S structure (PDB ID 5AFI) (Fischer et al., 2015) was chosen as the template. The sequence alignment of rRNA in bacteria was downloaded from the CRW website (Cannone et al., 2002). The rRNA expansion segments were manually built in Assemble2 (Jossinet and Westhof, 2005) with the aid of the secondary structures that were constructed based on the downloaded sequence alignment, and then refined into the density map using MDFF (Trabuco et al., 2008). The rRNA was refined in PHENIX (Adams et al., 2010), with restrictions of secondary structures. ERRASER (Chou et al., 2013) was also used to improve the RNA backbone geometry. The sequences of *M. tuberculosis* rProteins were obtained from the NCBI database. MODELLER (Eswar et al., 2007) was used to generate the initial homology models. Rosetta (DiMaio et al., 2015) was used to refine the homology model of each rProtein into the density map. The best model was selected from the Rosetta refinement based on both the geometry and the fitting scores. The rRNA and rProtein

models were merged into one model and subsequently refined using Rosetta and PHENIX. During the above modelling process, Coot (Emsley and Cowtan, 2004) was iteratively used to inspect and improve the local fitting.

To obtain a full model of the *M. tuberculosis* 70S ribosome, the models of the *M. tuberculosis* 50S and 30S from the above procedures were merged together with adjustment of the interface between the two subunits. For the handle in the *M. tuberculosis* 70S, we used MOSAICS-EM (Zhang et al., 2012) to flexibly fit the handle into the density of the dominant conformation, while ensuring its connectivity to the other part of the LSU. The 70S map has low local resolutions in some certain regions, including the tip of the handle and the periphery of the SSU. To prevent any improper refinement of high quality models into these poor quality regions, we added the density of the dominant handle conformation and the density of the locally refined SSU to the 70S map to refine the entire 70S model. Both models of the *M. tuberculosis* 50S and 70S were inspected and adjusted in Coot, and iteratively refined in Rosetta and PHENIX. The statistics, obtained by using PHENIX and MolProbity (Chen et al., 2010), of the refined models are listed in **Table 3.2**.

Table 3.2. Model statistics for the final models of the *M. tuberculosis* 50S and the *M. tuberculosis* 70S

Model	<i>Mtb</i> 50S	<i>Mtb</i> 70S
Bond RMSD	0.0067	0.0099
Angle RMSD	0.97	1.14
Molprobit score	1.66 (91 th percentile, N=27675, 0Å - 99Å)	1.81 (85 th percentile, N=27675, 0Å - 99Å)
Clashscore, all atoms	2.99 (98 th percentile, N=1784, all resolutions)	4.63 (95 th percentile, N=1784, all resolutions)
Poor rotamers (%)	0	0
Favored rotamers (%)	99.61	99.7
Ramachandran outliers (%)	0.6	0.71
Ramachandran favored (%)	89.2	89.18
Correct sugar puckers (%)	99.97	99.75
Good backbone conformations (%)	90.75	91.86

3.2.5 Fitting the models for the forty conformations of the L1 stalk, the handle, and the 16S

The Hierarchical Natural Move Monte Carlo or HNMMC as implemented in MOSAICS-EM (Zhang et al., 2012) was applied to refine the handle and the L1 stalk into all forty obtained EM density maps. The HNMMC allows the user to group atoms into regions with collective motions (Sim et al., 2012) to increase the efficiency of conformational sampling. For the forty conformations of the 16S, we first segmented out the densities for 16S only and then used MOSAICS-EM to refine the 16S model into its density.

3.2.6 Principal component analysis of the *M. tuberculosis* 70S structures

The principal component analysis was performed on forty models of the handle, the L1 stalk, and the 16S rRNA using the Bio3D package (Grant et al., 2006). Only the C4' atoms were used in calculating the principal components.

3.2.7 Figures and movies preparation

All figures and movies were made using UCSF Chimera (Pettersen et al., 2004) and XRNA (<http://rna.ucsc.edu/rnacenter/xrna/>).

3.3 Results and Discussion

3.3.1 Architecture of the *M. tuberculosis* ribosomes

The *M. tuberculosis* 70S ribosome consists of a large subunit (LSU or 50S) and a small subunit (SSU or 30S). For consistency throughout the text, we will use 50S and 70S to refer to the two ribosome specimens, and use LSU and SSU to refer to the two ribosomal subunits. *E.g.*, the *M. tuberculosis* 50S specimen contains only the LSU while the *M. tuberculosis* 70S specimen contains both the LSU and SSU. The *M. tuberculosis* LSU is composed of rRNA 23S, rRNA 5S, and about thirty rProteins, while the *M. tuberculosis* SSU is made of rRNA 16S and about twenty rProteins.

Using single-particle cryo-EM, structures of the purified *M. tuberculosis* 50S ribosomal subunit and 70S ribosome were determined to overall resolutions of 3.7 Å and 4.0 Å, respectively (**Table 3.1**, **Figure 3.1**). Cryo-EM densities in the core region of the *M. tuberculosis* ribosomes showed clear features of individual RNA bases (**Figure 3.2B**) as well as the bulky

side chains of the rProteins (**Figure 3.2C**). In the cryo-EM map of the *M. tuberculosis* 70S, the SSU exhibited a large degree of flexibility, leading to a lower local resolution. To improve the quality of the structure, densities of the SSU were extracted from the particle images for the entire *M. tuberculosis* 70S (Bai et al., 2015) and processed separately. This approach allowed us to resolve the SSU to an overall resolution of 4.5 Å (**Figure 3.1**). The models of *M. tuberculosis* 50S and 70S were built (see **Methods**) and the model statistics are summarized in **Table 3.2**.

The resultant structures of the *M. tuberculosis* 50S and 70S ribosomes showed that the overall architecture is conserved compared with other bacterial ribosomes, possessing all the structural landmarks, including the central protuberance, the L1 stalk, and the L7/L12 stalk base of the LSU, as well as the head, the beak, the body, the spur, and the platform of the SSU (**Figure 3.2D**). In the *M. tuberculosis* 70S structure, the density for the anticodon stem loop (ASL) of the P-site tRNA, along with the mRNA, is clearly visible, showing base pairing between the codon and the anticodon (**Figure 3.3**). We have resolved most of the rProteins and modelled them into the density with the exception of rProteins uL1, uL10, uL11, bL12, bS1, and uS2, due to the flexibility or absence of these proteins in our structure (**Table 3.3**).

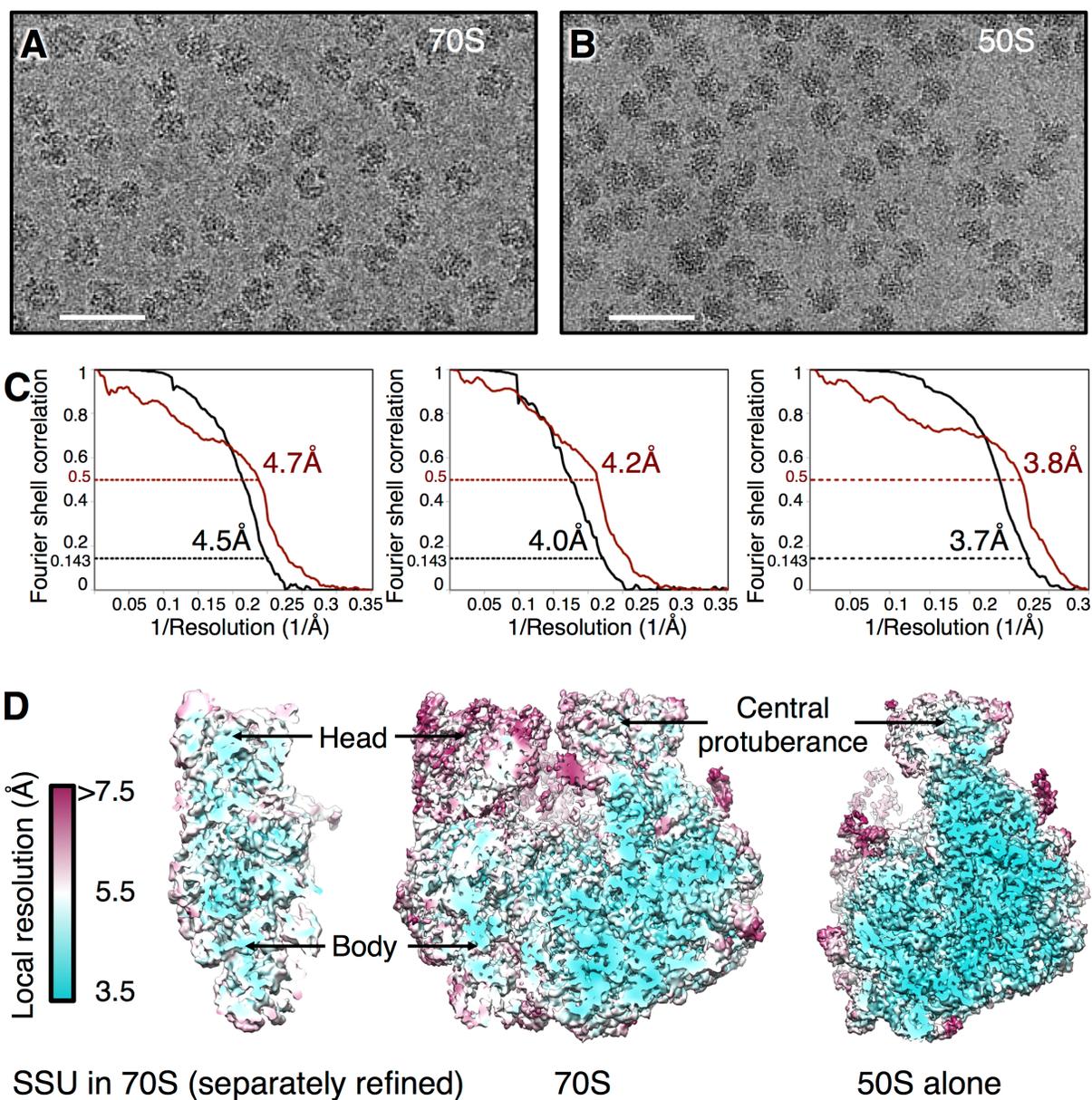


Figure 3.1. Raw micrographs and resolution estimation

(A) A representative raw micrograph of *M. tuberculosis* 70S with a defocus of 1.2 μm . The scale bar denotes 500 \AA .

(B) A representative raw micrograph of *M. tuberculosis* 50S with a defocus of 0.9 μm . The scale bar denotes 500 \AA .

(C) Fourier shell correlations (FSC) of the reconstructions and models for locally refined SSU, 70S and 50S alone, from left to right, respectively. The black curves denote the FSC between the two independent reconstructions from two half datasets. The red curves denote the FSC between the final maps of reconstructions and the final refined models.

(D) 3D density maps colored by local resolutions.

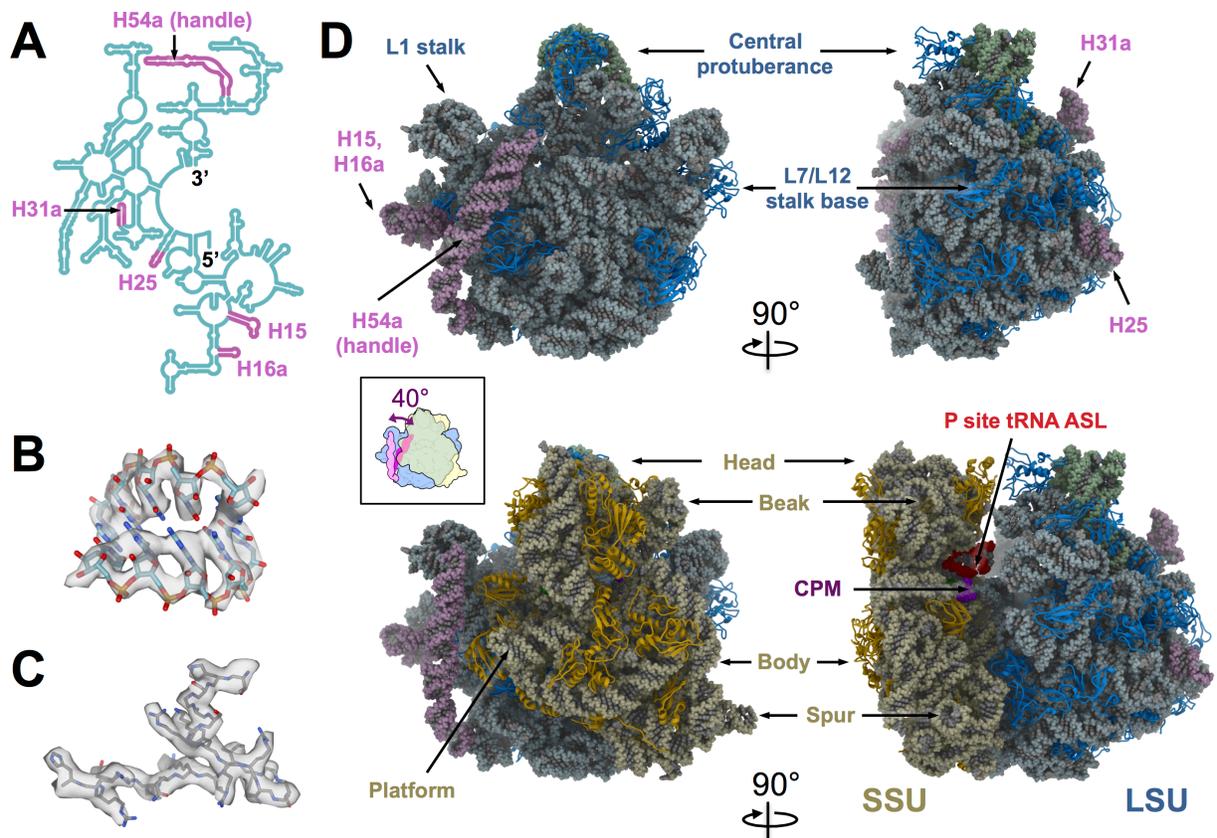


Figure 3.2. Cryo-EM structures of the *M. tuberculosis* 50S and 70S ribosomes
(A) Secondary structure of the 5' half of the *M. tuberculosis* 23S, showing the positions of the *M. tuberculosis* expansion segments (plum color) within the 23S rRNA.
(B) The model of a helix fragment of the 23S rRNA (residue 818–822 and residue 898–902) fits into the density from the *M. tuberculosis* 50S, showing individual RNA bases.
(C) The model of bL35 (residues 6–31) fits into the density from the *M. tuberculosis* 70S, showing bulky protein side chains.
(D) Overall structures of the *M. tuberculosis* 50S (top row) and 70S (bottom row) ribosomes viewing from the subunit interface (left column) and the L7/L12 stalk base (right column), respectively. Structural landmarks of the bacterial ribosome are labeled. Color schemes are dodger blue for LSU rProteins, light blue for 23S, plum for 23S rRNA expansion segments, green for 5S, gold for SSU rProteins, light yellow for 16S, purple for capreomycin (CPM) and red for the anticodon stem loop of the P-site tRNA. The cartoon in the inset box is an overlay of the 50S and 70S viewed from the SSU. The handle swings 40° counter-clockwise upon the association between SSU and LSU.

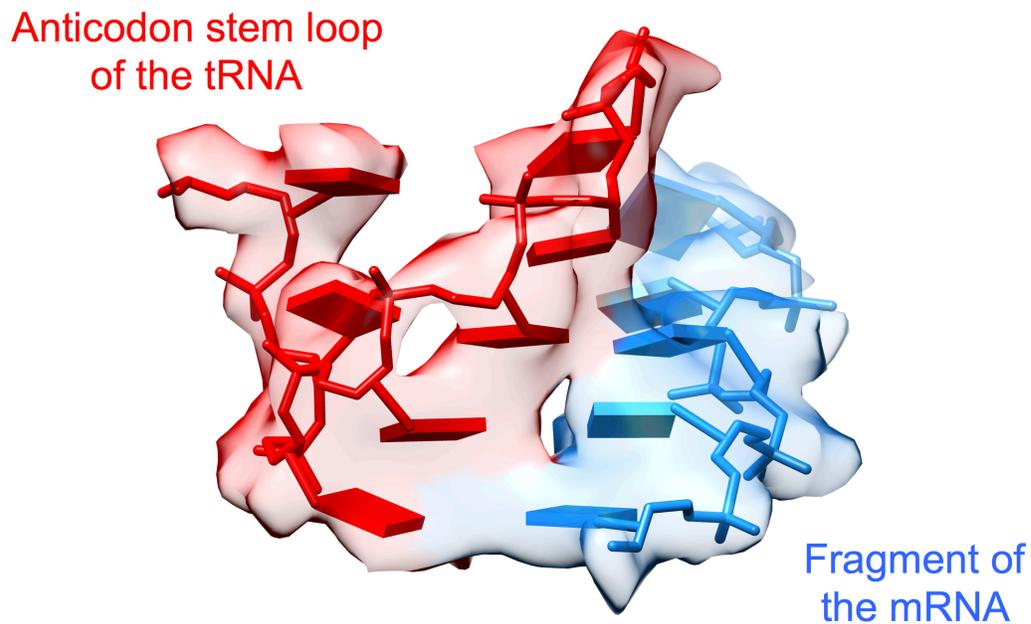


Figure 3.3. Cryo-EM density of the anticodon stem loop and the fragment of mRNA in the map of the *M. tuberculosis* 70S ribosome

The densities of the tRNA and the mRNA are in red and blue, respectively. Models of the anticodon stem loop and the mRNA were taken from the crystal structure (PDB ID 4WR6). Note that the anticodon stem loop and the mRNA can be of different base composition for each 70S particle. Therefore, the bases are shown as slabs without the identity of the bases.

Table 3.3. Modeled regions of the rProteins and rRNAs

New name	Old name	Rv number	Full length	Built residues	New name	Old name	Rv number	Full length	Built residues
uL1	L1	Rv0641	235	0	bL33	L33	Rv0634B	55	7-54
uL2	L2	Rv0704	280	2-273	bL34	L34	Rv3924c	47	4-45
uL3	L3	Rv0701	217	2-214	bL35	L35	Rv1642	64	2-63
uL4	L4	Rv0702	223	9-215	bL36	L36	Rv3461c	37	1-37
uL5	L5	Rv0716	187	12-181	bS1	S1	Rv1630	481	0
uL6	L6	Rv0719	179	3-176	uS2	S2	Rv2890c	287	0
bL9	L9	Rv0056	152	1-47	uS3	S3	Rv0707	274	2-208
uL10	L10	Rv0651	178	0	uS4	S4	Rv3458c	201	1-201
uL11	L11	Rv0640	142	0	uS5	S5	Rv0721	220	41-200
bL12	L7/L12	Rv0652	130	0	bS6	S6	Rv0053	96	34700
uL13	L13	Rv3443c	147	2-147	uS7	S7	Rv0683	156	2-153
uL14	L14	Rv0714	122	2-122	uS8	S8	Rv0718	132	3-132
uL15	L15	Rv0723	146	4-145	uS9	S9	Rv3442c	151	25-151
uL16	L16	Rv0708	138	2-135	uS10	S10	Rv0700	101	3-101
bL17	L17	Rv3456c	180	2-117	uS11	S11	Rv3459c	139	23-139
uL18	L18	Rv0720	122	7-122	uS12	S12	Rv0682	124	1-123
bL19	L19	Rv2904c	113	1-112	uS13	S13	Rv3460c	124	2-114
bL20	L20	Rv1643	129	2-123	uS14	S14	Rv0717	61	22341
bL21	L21	Rv2442c	104	5-102	uS15	S15	Rv2785c	89	32174
uL22	L22	Rv0706	197	9-121	bS16	S16	Rv2909c	162	31048
uL23	L23	Rv0703	100	3-100	uS17	S17	Rv0710	136	43-136
uL24	L24	Rv0715	105	1-46; 61-104	bS18	S18	Rv0055	84	20-82
bL25	L25	Rv1015c	215	6-182	uS19	S19	Rv0705	93	30376
bL27	L27	Rv2441c	86	12-85	bS20	S20	Rv2412	86	31533
bL28	L28	Rv2975A	64	1-63					
uL29	L29	Rv0709	77	4-68					
uL30	L30	Rv0722	65	2-60					
bL31	L31	Rv1298	80	1-45					
bL32	L32	Rv0979A	57	2-54					

rRNA name	Full length	Built residues
23S	3138	5-1012; 1023-3132
16S	1537	8-1526
5S	115	1-115

The secondary structures of all the *M. tuberculosis* rRNAs 23S, 16S, and 5S, were built based primarily on sequence alignment (Cannone et al., 2002)(**Figure 3.4**). All the rRNA expansion segments in the *M. tuberculosis* ribosomal LSU are located in the 5' half of the rRNA 23S (**Figure 3.2A**) and have all been identified in our *M. tuberculosis* ribosomal structures (**Figure 3.2D**). Expansion segments for H25 and H31a are located on the solvent exposed side of the LSU, away from the SSU, with H25 leaning towards the L7/L12 stalk base and H31a leaning towards the central protuberance. Expansion segments H15 and H16a are located on the solvent exposed side of the LSU, close to the L1 stalk, with their tips forming an RNA kissing loop (**Figure 3.5A**). From the Comparative RNA Website (CRW) (Cannone et al., 2002), we have identified expansion segments of H15 and H16a in the 23S from eight different bacteria (*M. tuberculosis*, *Mycobacterium leprae*, *Micrococcus luteus*, *Streptomyces ambofaciens*, *Thermotoga maritima*, *Leptospira interrogans*, *Borrelia burgdorferi*, and *Treponema pallidum*). Seven of the eight bacterial species have nucleotide bases at the tips of the H15 and H16a supporting Watson-Crick base pairing to form a kissing loop (**Figure 3.5B**). The only exception is *Treponema pallidum*, whose H15 is about 20 nucleotides shorter than the H15 in the other seven bacteria, and is therefore unable to form a kissing loop with H16a. The rRNA expansion segment of H15 from the *M. tuberculosis* ribosome interacts directly with the long α -helix of rProtein bL9, which connects the two globular domains at the N-terminal and C-terminal ends. This extended H15, stabilized by the kissing loop between H15 and H16a, may provide an additional structural anchor for binding the long connecting α -helix of bL9 (**Figure 3.6**).

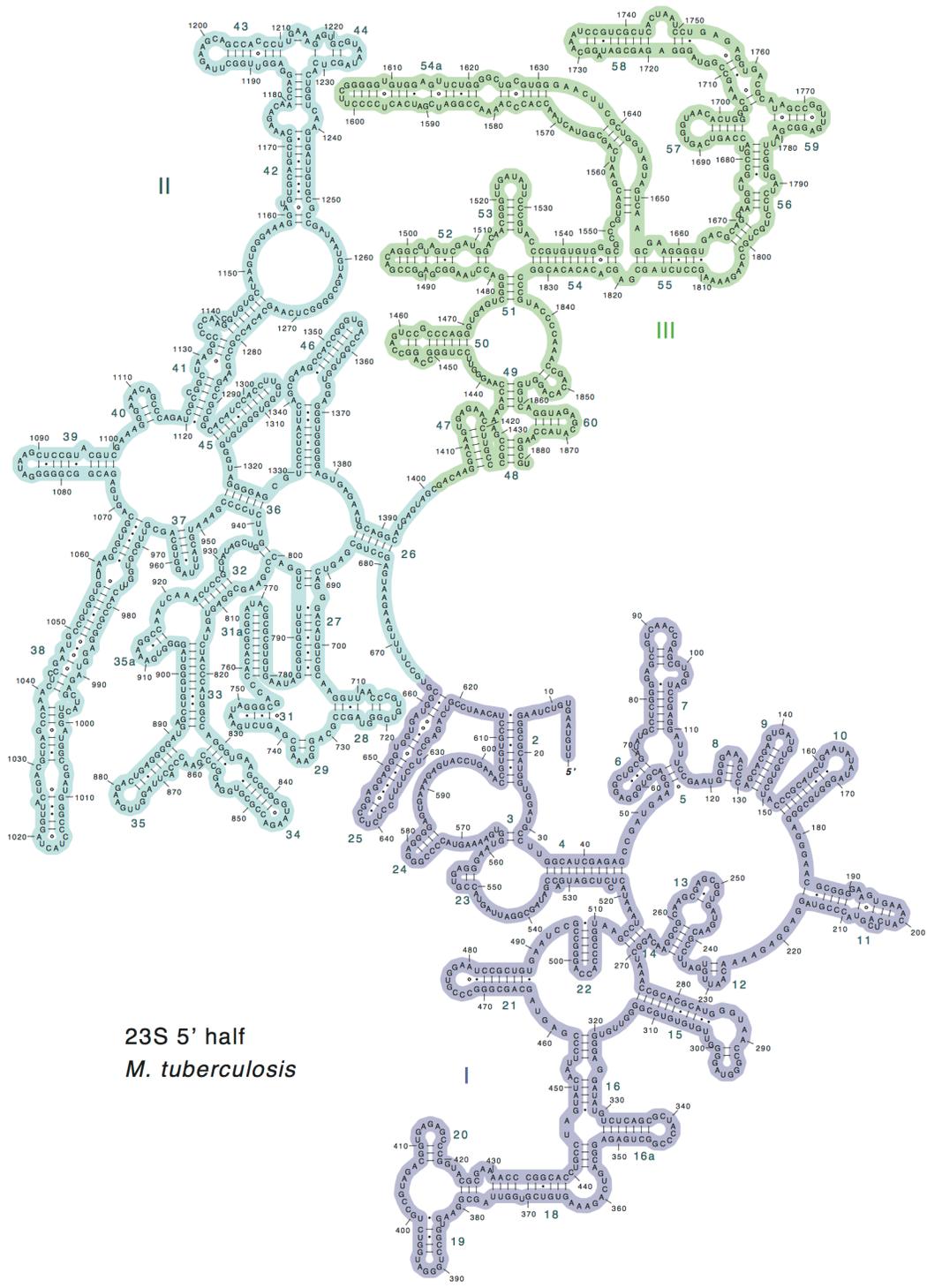


Figure 3.4. Secondary structures of the *M. tuberculosis* rRNA
(A) Secondary structure of 23S 5' half.

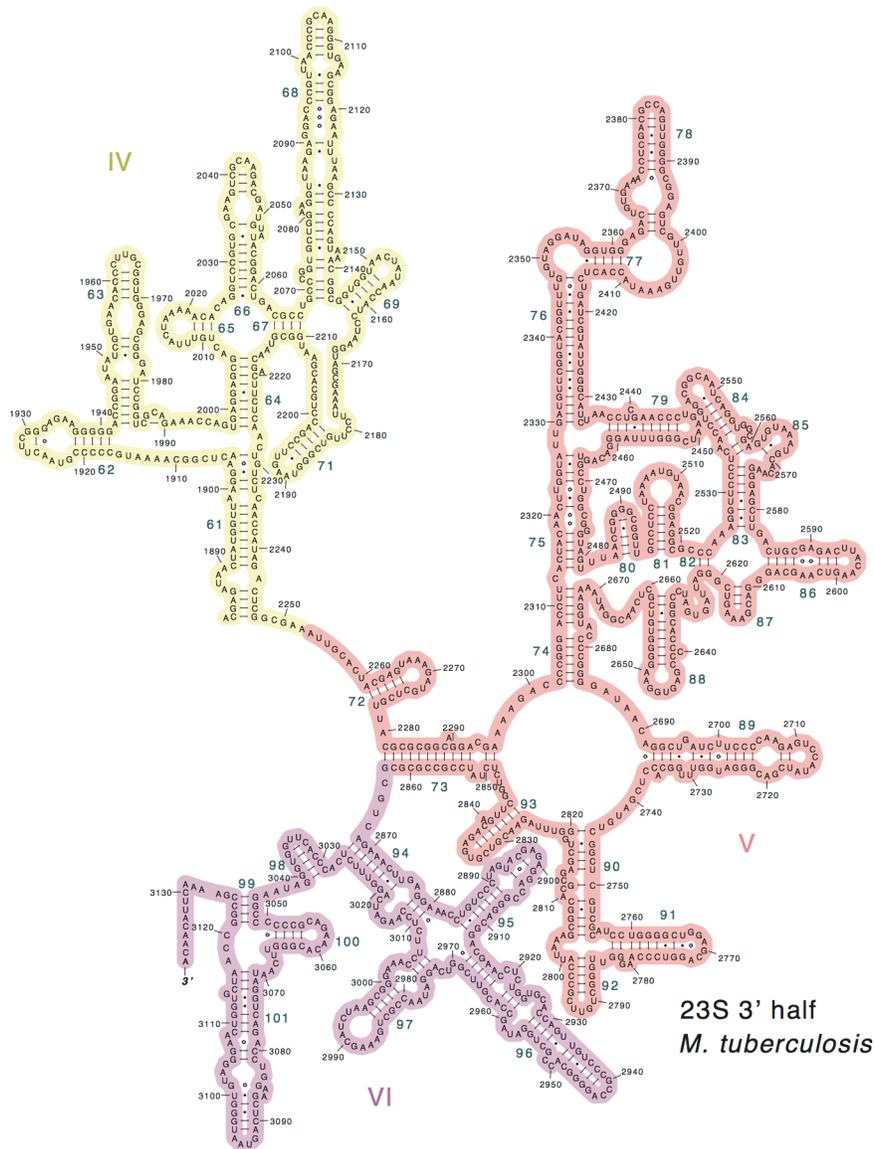


Figure 3. 4B The secondary structure of the *M. tuberculosis* rRNA 23S 3' half

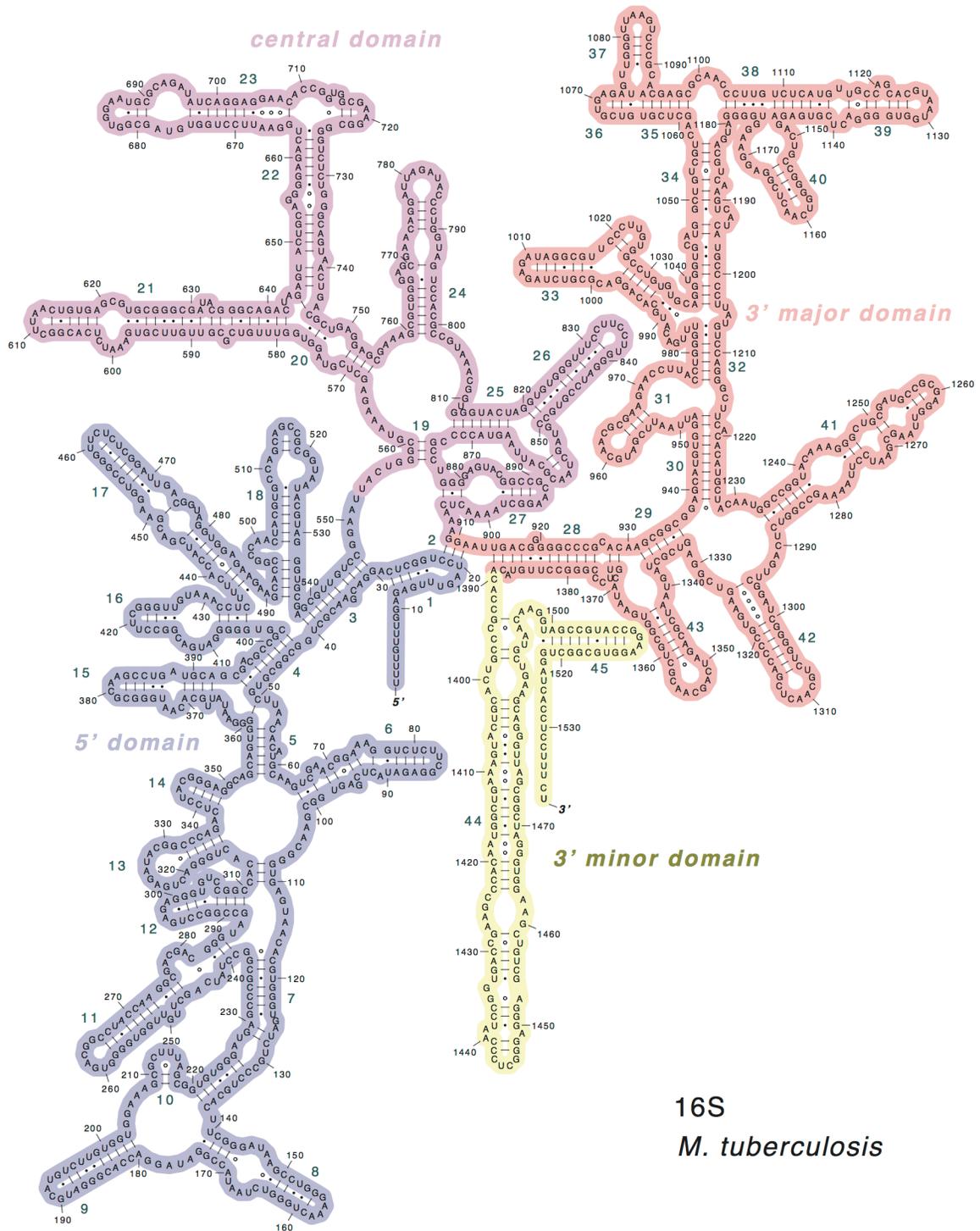


Figure 3. 4C The secondary structure of the *M. tuberculosis* rRNA 16S

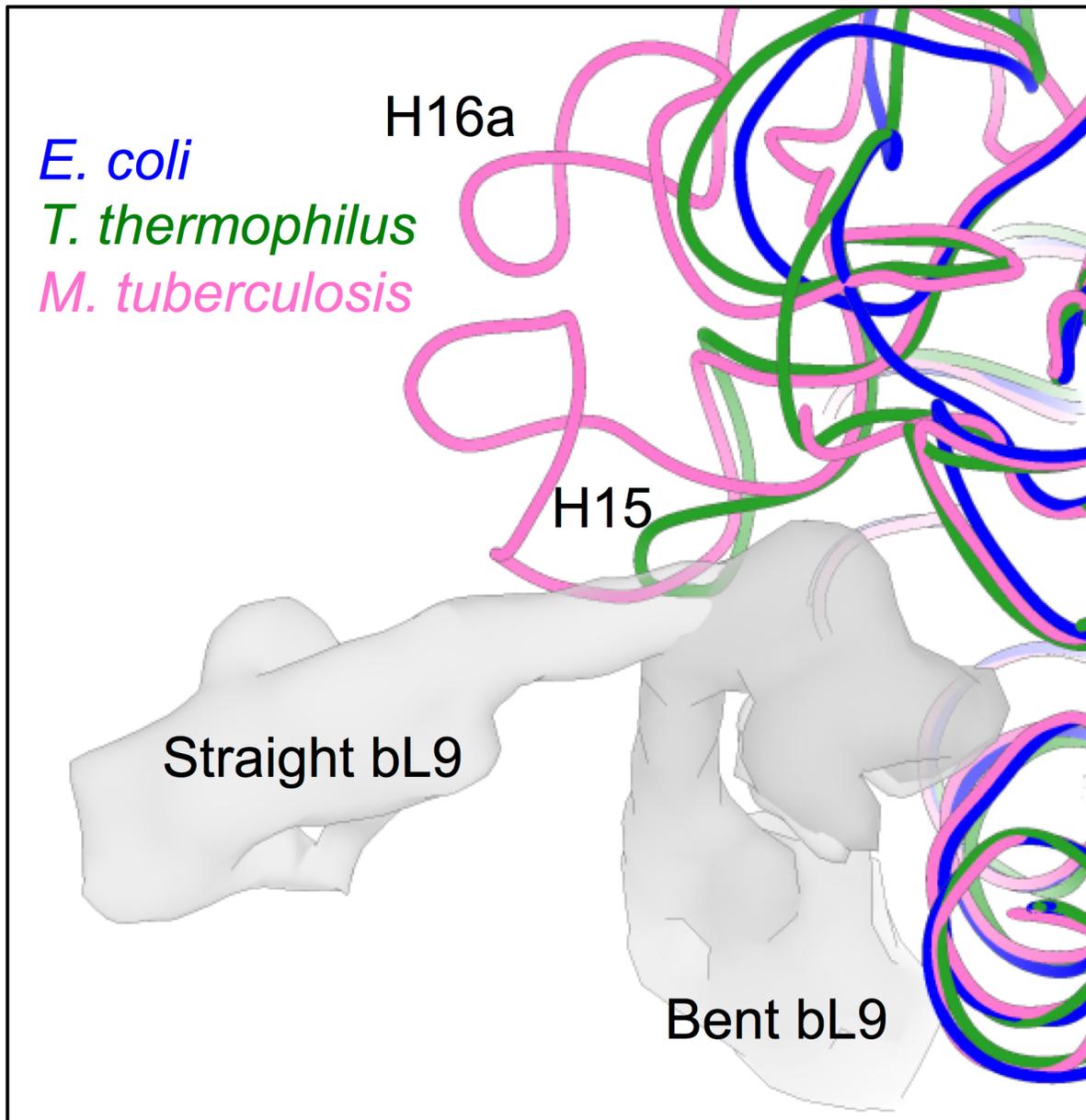


Figure 3.6. Different lengths of H15 rRNA helices

Color code: *E. coli*, blue; *T. thermophiles*, green; *M. tuberculosis*, pink.

The *E. coli* and *T. thermophilus* ribosome structures were from PDB ID 5AFI and 2J01, respectively. The bL9 C terminal domain is flexible and therefore shown as representative straight and bent conformations from available structures. The H15 in the *M. tuberculosis* ribosome is much longer than other bacterial ribosomes and forms kissing loop with H16a. The longer H15 interacts with the long connecting α -helix of the straight bL9.

The cryo-EM density for the 100-nucleotide long rRNA expansion segment, handle, is well defined in the *M. tuberculosis* 50S density map. However, in the 70S map, density for the tip of the handle was missing, suggesting significant levels of conformational variability among the handles in the 70S particles. To better resolve the structures of the handle in *M. tuberculosis* 70S, an unsupervised classification was performed locally on the handle to generate forty maps of the *M. tuberculosis* 70S ribosome, each showing more complete density of the handle in that particular conformation (**Figure 3.7**). The resolutions of these forty maps of the *M. tuberculosis* 70S range between 6.6 Å and 12.8 Å (**Table 3.4**). Five representative conformations of the handle in *M. tuberculosis* 70S are shown in **Figure 3.8**. The dominant conformation is conformation I, where the ribosome is in the non-rotated state and the handle is closer to the L1 stalk. In conformations I, II, and III, the densities of the handle and the bS6 are connected, while in conformations IV and V, they are disconnected.

The better densities of the handles obtained for the forty classes of the *M. tuberculosis* 70S, allowed us to compare it with its corresponding density in the *M. tuberculosis* 50S. The handle showed a remarkable conformational change between the *M. tuberculosis* 50S and the *M. tuberculosis* 70S structures. In the 50S, the handle is bent towards the central protuberance; the dominant conformation of the 70S shows the handle tilted towards the L1 stalk. The handle moves about 40 degrees from its primary orientation in the 50S to the predominant conformation in the 70S. This movement effectively translates into the tip traveling about 60 Å from one state to the other (inset of **Figure 3.2D**).

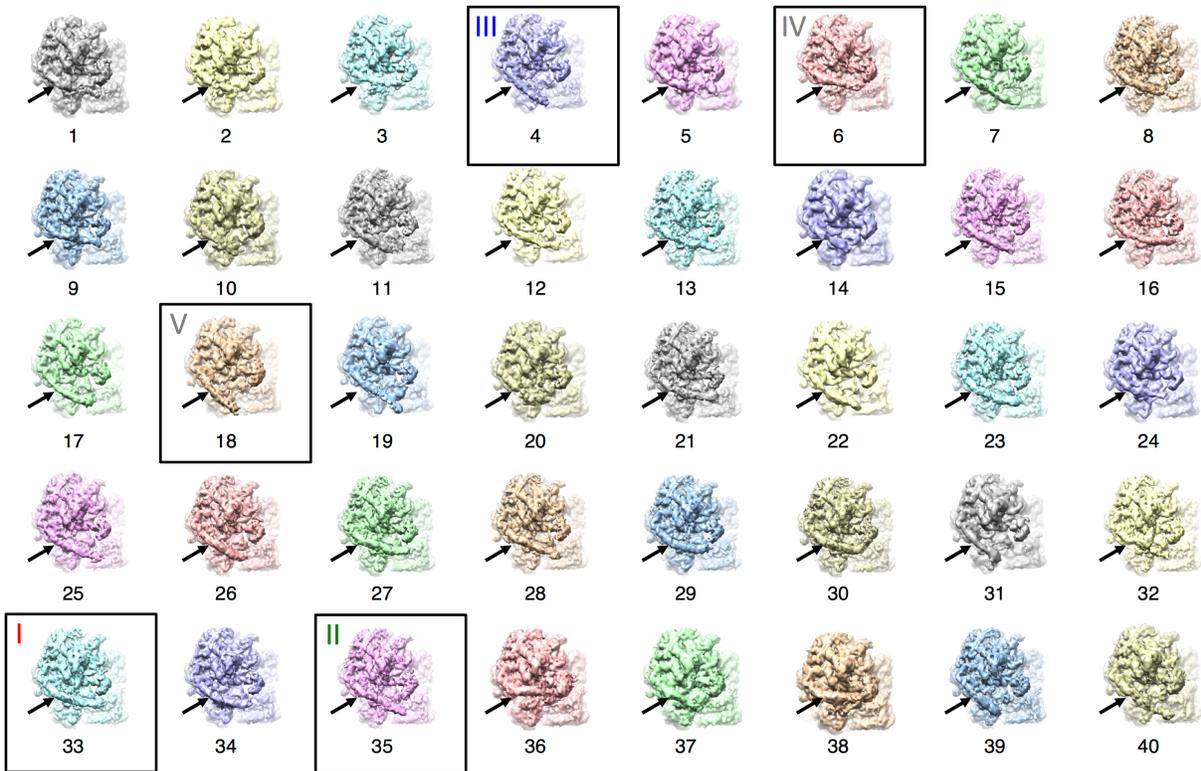


Figure 3.7. Different conformations of the *M. tuberculosis* 70S

The *M. tuberculosis* 70S particles were classified into forty good classes, based on the conformations of the handle. Arrows indicate the locations of the handles. The five representative conformations were selected (in the black boxes) and displayed in detail in Supplementary Figure 3.8. Map 33 is the most dominant conformation out of the forty maps.

Table 3.4. Classification of *M. tuberculosis* 70S based on the region around the handle
If the bridge B9 is maintained, the column B9 is recorded as “Y”; otherwise, “N”.

Map number	Number of particles	Percentage	Resolution (Å)	B9
1	6353	2.18%	10.3	Y
2	6647	2.28%	11.2	Y
3	6608	2.27%	10.3	N
4	6673	2.29%	10.3	N
5	5709	1.96%	11.2	N
6	6448	2.21%	10.3	N
7	5506	1.89%	11.2	Y
8	10711	3.67%	9.2	Y
9	5574	1.91%	10.3	N
10	4426	1.52%	10.9	Y
11	8348	2.86%	9.7	Y
12	4371	1.50%	11.2	Y
13	9379	3.21%	9.7	Y
14	2509	0.86%	12.8	Y
15	7808	2.68%	9.7	Y
16	8424	2.89%	9.7	N
17	4027	1.38%	11.2	N
18	4979	1.71%	10.9	N
19	4574	1.57%	10.9	Y
20	4772	1.64%	10.9	Y
21	9205	3.16%	10	Y
22	2472	0.85%	12.8	Y
23	9507	3.26%	9.2	Y
24	3319	1.14%	11.6	N
25	7550	2.59%	10	Y
26	7357	2.52%	10	Y
27	16509	5.66%	8.2	Y
28	10020	3.43%	9.2	Y
29	6773	2.32%	9.7	Y
30	13926	4.77%	8.8	Y
31	3752	1.29%	11.6	Y
32	9138	3.13%	9.4	Y
33	20029	6.87%	6.6	Y
34	4831	1.66%	10.9	Y
35	13879	4.76%	7.2	Y
36	6659	2.28%	10.6	N
37	5669	1.94%	10.6	Y
38	5042	1.73%	11.2	N
39	6924	2.37%	10.3	Y
40	5330	1.83%	10.9	Y

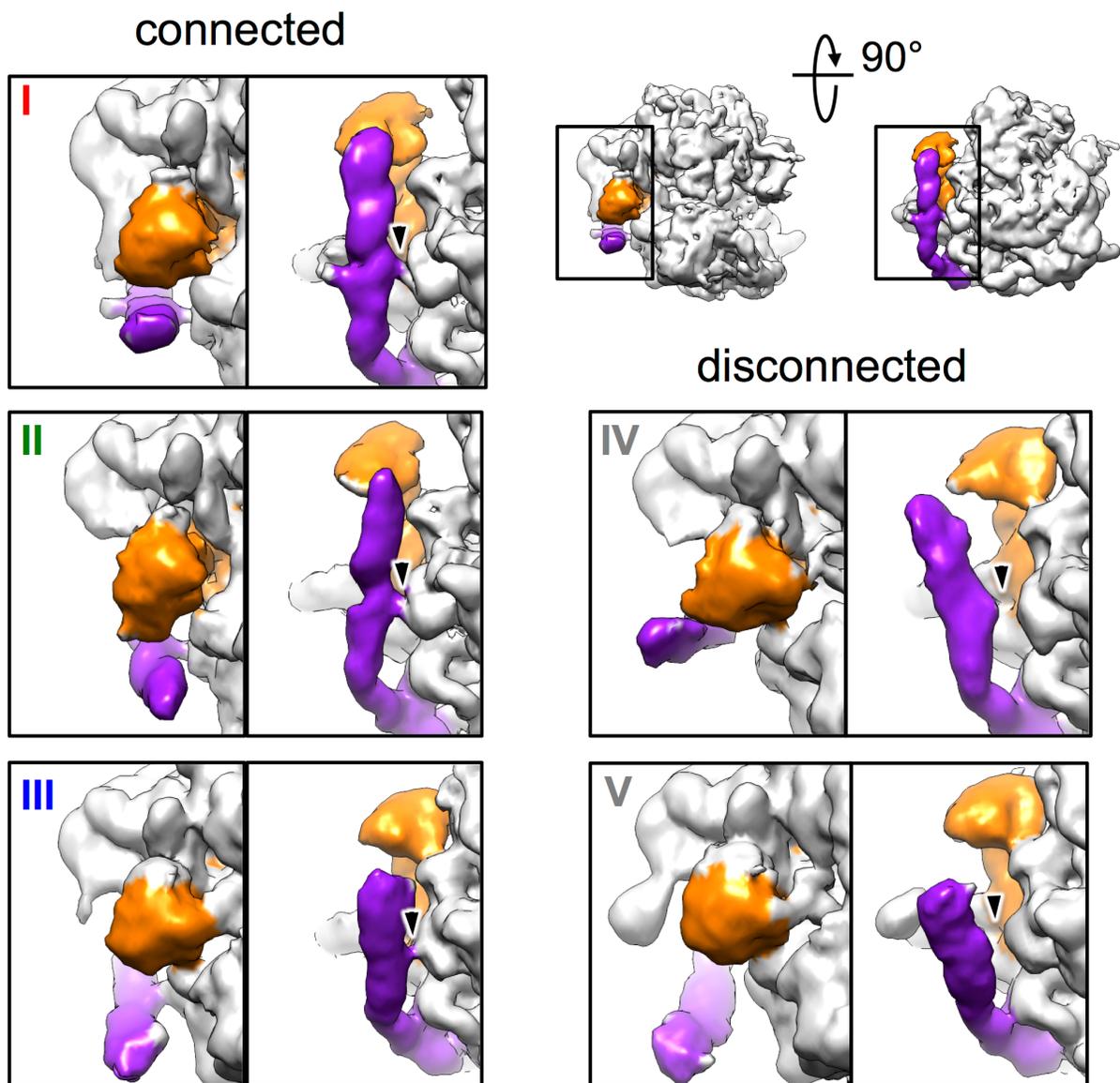


Figure 3.8. Five representative handle conformations
 The handle is colored purple and the L1 stalk is colored orange. For each conformation, two zoom-in views of the boxed regions of the *M. tuberculosis* 70S (labeled in the upper right corner) were shown. In conformation I, II and III, the densities of handle and bS6 are connected, while in conformation IV and V, they are disconnected.

3.3.2 Interactions between handle and other regions in the *M. tuberculosis* ribosome

In the *M. tuberculosis* 50S structure, three loops on the solvent exposed side of uL2 (amino acid residues 168 - 170, 134 - 136, 121 - 123; labeled with blue stars from top to bottom in **Figure 3.9A**) form interactions with a minor groove on the handle (nucleotide residues 1576 - 1580 and 1626 - 1630). There is another interaction (labeled by a black star in **Figure 3.9A**) between residue 1591A of the handle and a minor groove of helix H68 of the 23S (formed by nucleotides 2079 - 2081 and 2132 - 2134). The interactions of the handle with H68 (**Figure 3.9C**) and uL2 (**Figure 3.9D**) appear to be stable within the *M. tuberculosis* 50S structure and not an artifact of a high magnesium concentration (Ivanov et al., 2013), as reducing the magnesium concentration in the buffer to 1 mM does not affect these interactions (**Figure 3.10**). Interestingly, in the *M. tuberculosis* 70S model, the same minor groove of helix H68 interacts with an unpaired adenine (nucleotide residue 693A) on the 16S rRNA of the SSU, forming intersubunit bridge B7a.

In bacterial ribosomes, rProtein uL2 interacts with rRNA in the SSU to form the intersubunit bridge B7b (Liu and Fredrick, 2016). When we directly superimposed the 50S structure onto the *M. tuberculosis* 70S, the handle in the *M. tuberculosis* 50S specimen collides with both rProtein uS11 and helix h23 of the 16S rRNA in the 70S. Thus, the conformation of the handle seen in the 50S would not allow the direct association of the SSU with the LSU. Moreover, if the handle remained in the same conformation as in the 50S, the tip of the handle would be in close proximity to the initiator tRNA in the initiation complex (**Figure 3.11**) (Sprink et al., 2016). This position suggests the handle may play a role in coordinating translation initiation in the *M. tuberculosis* ribosome, such as the handle staying in this bent conformation in

the *M. tuberculosis* 50S to prevent the association of the 30S until the initiation complex is properly formed.

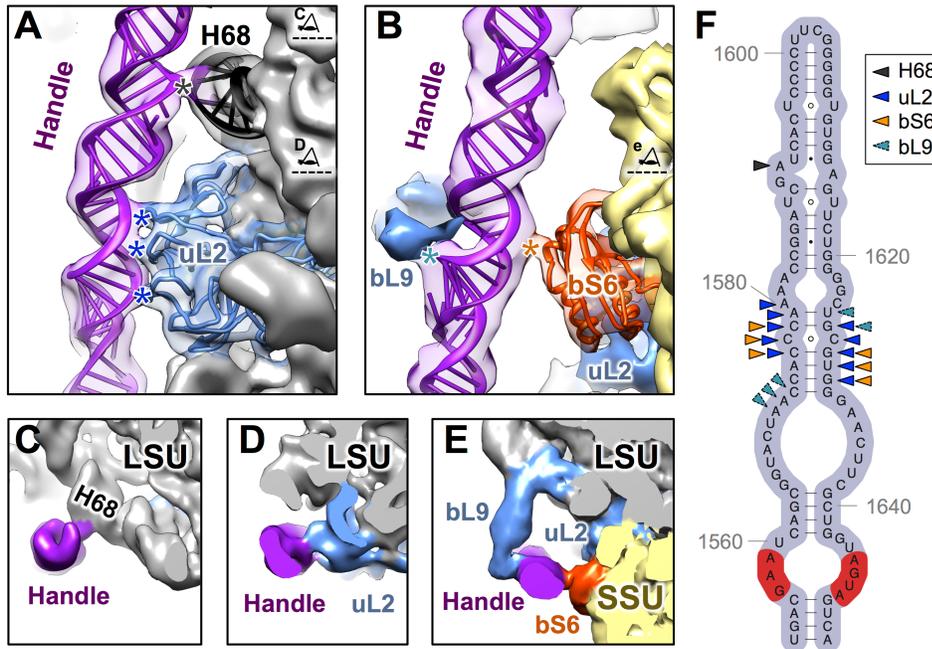


Figure 3.9. Interactions of the handle with other ribosomal components within the *M. tuberculosis* ribosomes

(A) Ribbon models showing where the handle (purple) interacts with rProtein uL2 (blue) and H68 (black) when in the 50S alone. The cryo-EM density of the 50S is low-pass filtered to 8Å resolution and overlaid onto the model. Black and blue stars indicate the sites of interactions from the handle to H68 and uL2, respectively. The two eye cartoons, top and bottom, label the cutting plane and viewing direction for Panels C and D, respectively.

(B) Ribbon models show the handle interacts with bS6 (orange red) in the 70S. The cryo-EM density for the predominant conformation of the 70S is low-pass filtered to 8Å resolution and overlaid onto the model. Densities of rProteins bL9 and uL2 are colored blue. Cyan and orange stars indicate the sites of interactions from the handle to bL9 and bS6, respectively. The eye cartoon labels the cutting plane and viewing direction of Panel E.

(C–E) Cryo-EM densities showing the handle interact with H68, uL2, bS6 and bL9. The 50S is labeled grey with uL2 and bL9 in blue while the handle is in purple. The 30S is in yellow with the rProtein bS6 in orange red.

(F) Secondary structure of the handle with arrow heads indicating the nucleotides that interact with H68 (black), uL2 (blue), bS6 (orange) and bL9 (cyan). The sequence, which can form the sarcin-ricin motif, is colored in the red background.

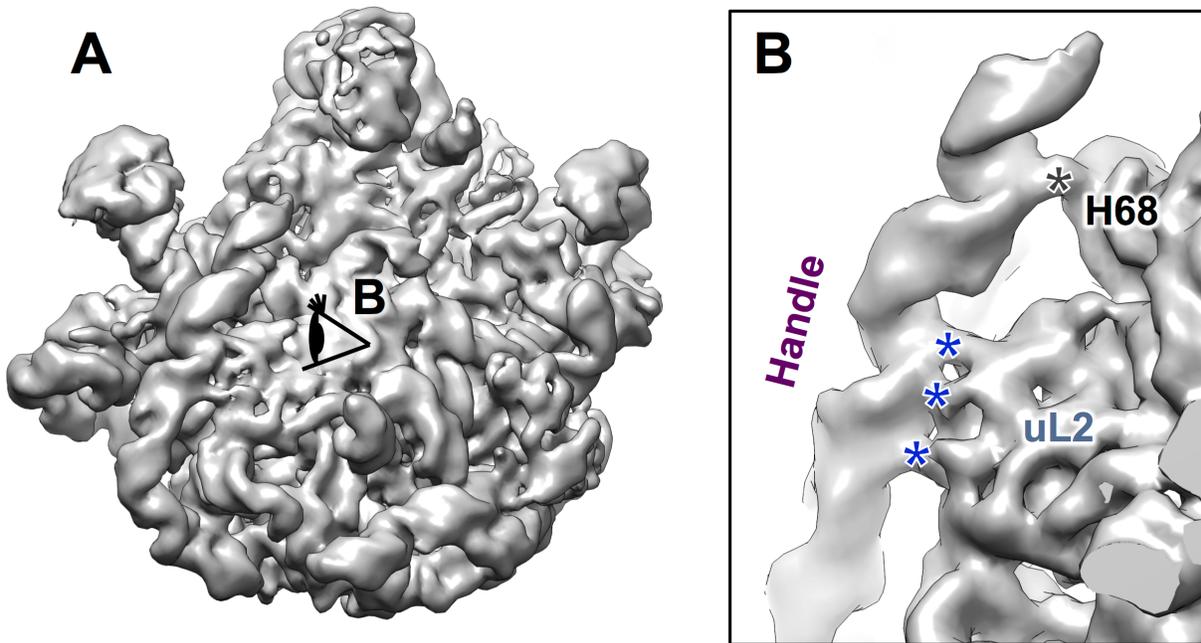


Figure 3.10. The conformation of the handle in the *M. tuberculosis* 50S specimen at 1 mM magnesium concentration
(A) The overview of the *M. tuberculosis* 50S at 1mM magnesium concentration. The eye cartoon indicates the viewing angle of Panel B.
(B) Zoom-in view of the interaction points of the handle with the uL2 and the H68 are shown as blue and black stars, respectively. The interactions have no considerable change when the magnesium concentration was lowered down to 1 mM.

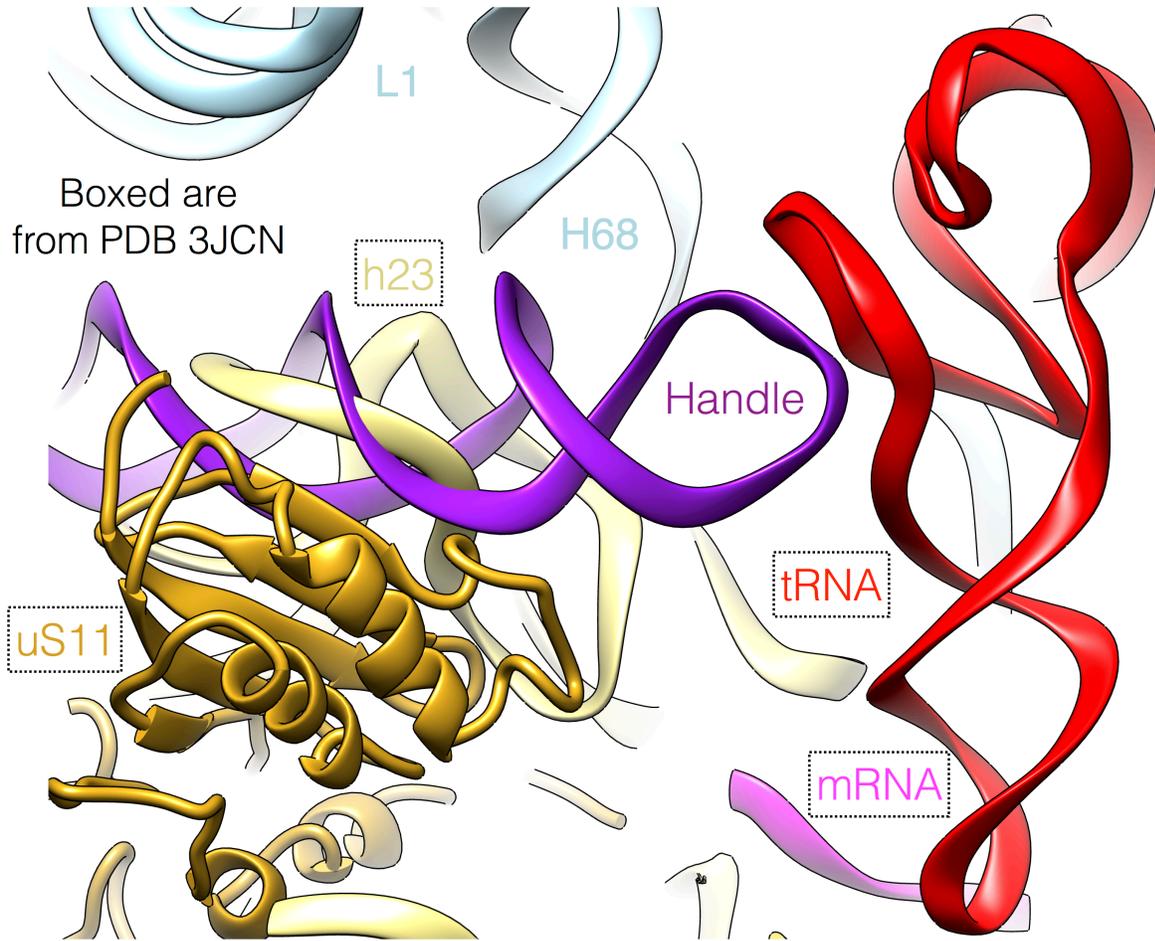


Figure 3.11. Superimposition of the *M. tuberculosis* 50S with the *E. coli* 70S initiation complex I (70S IC I)

The *E. coli* 70S IC I was taken from PDB ID 3JCN. The *M. tuberculosis* 50S was aligned onto the 70S IC I. The LSU of 70S IC I was hidden for simplicity (the *M. tuberculosis* 50S and the LSU of 70S IC I match well with each other). Apparently, the handle would collide with h23 and the initiator tRNA.

The 70S structure shows that when the subunits associate to form the mature 70S, the handle forms a new intersubunit bridge with rProtein bS6 (named B9, **Figure 3.9B, E**). Bridge B9 (labeled with an orange star in **Figure 3.9B**) comprises interactions between an α -helix from rProtein bS6 (amino acid residues 12 - 17) and the minor groove of the handle (residues 1576 - 1578 and 1628 - 1630). About 80% of the *M. tuberculosis* 70S particles have the B9 bridge maintained (**Table 3.4**).

In the most dominant conformation, the handle also interacts with rProtein bL9 (**Figure 3.9B, E**) with the C-terminal domain of bL9 inserted into a major groove of the handle (nucleotide residues 1571 - 1573 and 1625 - 1626, **Figure 3.9F**). In a recent high-resolution cryo-EM structure of the *E. coli* ribosome, *E. coli* bL9 was shown to interact with both uL2 and bS6 (Fischer et al., 2015). However, in the *M. tuberculosis* 70S ribosome, the handle can interact with the C-terminal domain of bL9 and may serve as a substitution for the interaction between bL9 and bS6 (**Figure 3.9E**). rProtein bL9 is highly flexible in the ribosome and has been shown to be involved in translational bypass of the T4 phage gene 60 (Adamski et al., 1996). Gene disruption of bL9 in *E. coli* increases the translational error rate when the cell is under stress, e.g., being treated with antibiotics (Naganathan et al., 2015). These observations, in combination with structural information from ribosomes and polysomes, led to the proposal that bL9 may slow the forward movement of ribosomes that trail transiently stalled ribosomes on the polysome. However, in *M. tuberculosis*, the extended RNA helices and the handle may further restrain the flexibility of bL9.

In particular, the handle contains a GAA-AGUA sequence (labeled red in **Figure 3.9F**), located at the bottom of this long RNA helix, to form a sarcin-ricin motif (or bulged-G motif) (Leontis and Westhof, 1998). The sarcin-ricin motif serves as a recognition site for RNAs and

proteins (Correll et al., 2003; Moore, 1999), which may aid in target binding (Ulyanov et al., 2006). Of note, many toxins, including α -sarcin (Ackerman et al., 1988), ricin A-chain (Endo and Tsurugi, 1988), Shiga toxin (Tesh and O'Brien, 1991) and VapC20 (Winther et al., 2013), bind to the sarcin-ricin motif in the sarcin-ricin loop (Szewczak et al., 1993) and cleave the nearby rRNA. The sarcin-ricin motif in the handle is exposed to solvent, as in the sarcin-ricin loop, thus accessible to its potential binding partners. It is yet to be discovered that such binding partners do exist or that a cleavage can occur in the *M. tuberculosis* handle, which may serve as a “switch” to control the potential translational modulation by the handle.

3.3.3 Correlated motions of the handle, the L1, and the SSU in the *M. tuberculosis* 70S

The *M. tuberculosis* 70S exhibits significant structural variation among the forty different density maps calculated from the 3D classification (see **Methods, Figure 3.7**), each representing a snapshot between the rotated and the non-rotated states of the *M. tuberculosis* 70S ribosome (**Figure 3.12A**) during the spontaneous intersubunit rotation (Dashti et al., 2014). Analysis of these states shows that the majority of the LSU remains unchanged. From the rotated to the non-rotated *M. tuberculosis* 70S with the LSU aligned, the SSU rotates clockwise when viewed from its solvent exposed side; the L1 stalk moves away from the tRNA binding sites and the handle moves towards the L1 stalk (**Figure 3.12B, C**). To further delineate motions in the *M. tuberculosis* 70S, we built structural models for each of the forty snapshots of the SSU, the L1 stalk, and the handle. Then we subjected them to principal component analysis, which revealed their dominant modes of motion (Grant et al., 2006). The first principal components of the motions for the SSU (represented by the rRNA 16S), the L1 stalk, and the handle, are the rotation of the SSU relative to the LSU, the tilt of the L1 stalk, and a large conformational

change of the handle between the mRNA exit and the L1 stalk. (**Figure 3.13A**). Each principal component 1 (PC1) represents 93%, 82%, and 68% of their total motions for the SSU, the L1 stalk and the handle, respectively. Therefore, each PC1 approximates their dominant structural variations. Projecting the forty conformations of the SSU and the L1 stalk onto their respective PC1s clearly shows a correlation between the major movements of the SSU and the L1 stalk (**Figure 3.13B**), which is consistent with the previous finding that the movement of the L1 stalk is directly linked to the rotation of the SSU (Mohan and Noller, 2017). Remarkably, in most of the *M. tuberculosis* 70S conformations, the major motion of the handle exhibits a strong correlation with the major movements of both the L1 stalk and the SSU (**Figure 3.13C, D**). When the SSU rotates clockwise and the L1 stalk moves away from the E-site, the handle moves close to the L1 stalk and vice versa. These concerted movements are consistent with a simple morphing between the rotated and non-rotated conformations of the *M. tuberculosis* 70S in **Figure 3.12**. Only about 20% of the *M. tuberculosis* 70S particles do not show such correlation of the handle to either the SSU or the L1 stalk. Closer inspection revealed that these 20% of the particles are missing the B9 bridge, resulting in less restrained movement of the handle.

The large conformational changes of the handle, which are correlated with the movement of the L1 stalk and SSU, may affect the translocation of the *M. tuberculosis* ribosome. For example, as the tRNA translocates between different sites on the ribosome, the SSU rotates and the L1 moves. As the handle moves around the exit sites of the mRNA and tRNA, it may coordinate the exit of the tRNA and the mRNA. Inspection of the different states of the *M. tuberculosis* 70S revealed that the tip of the handle is close to the mRNA exiting site in several rotated states. In fact it is less than 20 Å from the anti-Shine-Dalgarno sequence on the rRNA 16S in several conformations (**Figure 3.14**). The anti-Shine-Dalgarno sequence has been shown

to drive translational pausing and codon choice in bacteria (Li et al., 2012), and the close proximity of the handle to this region of the *M. tuberculosis* ribosome may well affect *M. tuberculosis* translation.

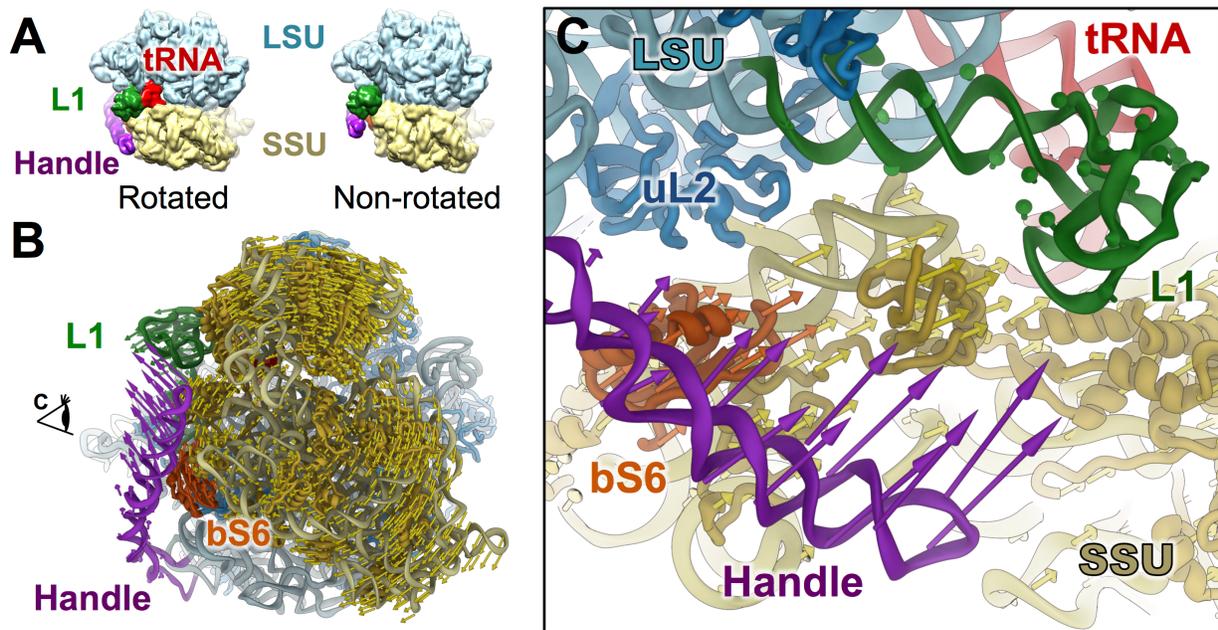


Figure 3.12. Correlated motions between the handle, SSU and the L1 stalk in the *M. tuberculosis* 70S ribosome
(A) *M. tuberculosis* 70S in the rotated (left) and non-rotated (right) states.
(B) Conformational difference between the rotated and non-rotated states of the *M. tuberculosis* 70S viewing from the solvent exposed side of the SSU. The SSU, L1 stalk, handle and bS6 on the SSU are colored yellow, green, purple and orange red, respectively. Colored arrows indicate the directions and amplitudes of the conformational differences for the handle (purple), the L1 stalk (green), the SSU (yellow) and the bS6 (orange red) on the SSU.
(C) Zoom-in view around the handle as viewed from the direction indicated by the eye cartoon in Panel B.

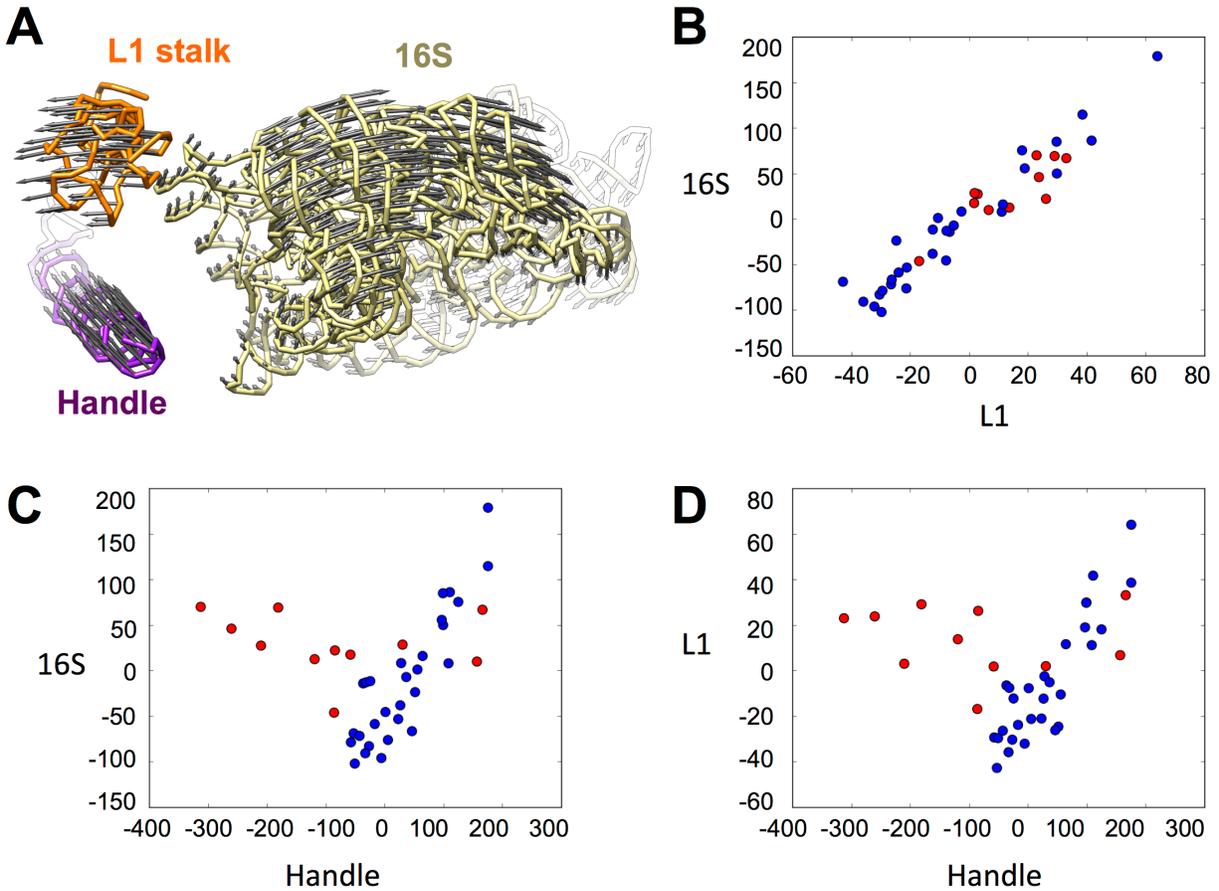


Figure 3.13. The correlation of the motions of the handle, the L1 stalk and the 16S
(A) Principal component analysis from forty conformations of the 70S ribosome with arrows indicating the principal component 1 (PC1) of motions for the handle (purple), the L1 stalk (orange) and the 16S rRNA (yellow), respectively. Each conformation of the forty *M. tuberculosis* 70S was projected onto the three PC1s for the handle, the L1 stalk and the 16S rRNA, and thus simplified as three parameters (the values in the X and Y axes in Panels B-D).

(B-D) Each conformation is plotted as a dot with respect to the parameters for L1 and 16S (Panel B), handle and 16S (Panel C), as well as handle and L1 (Panel D). Blue dots denote the conformations in which the densities of handle and bS6 are connected; red denotes the conformations in which the densities of handle and bS6 are disconnected.

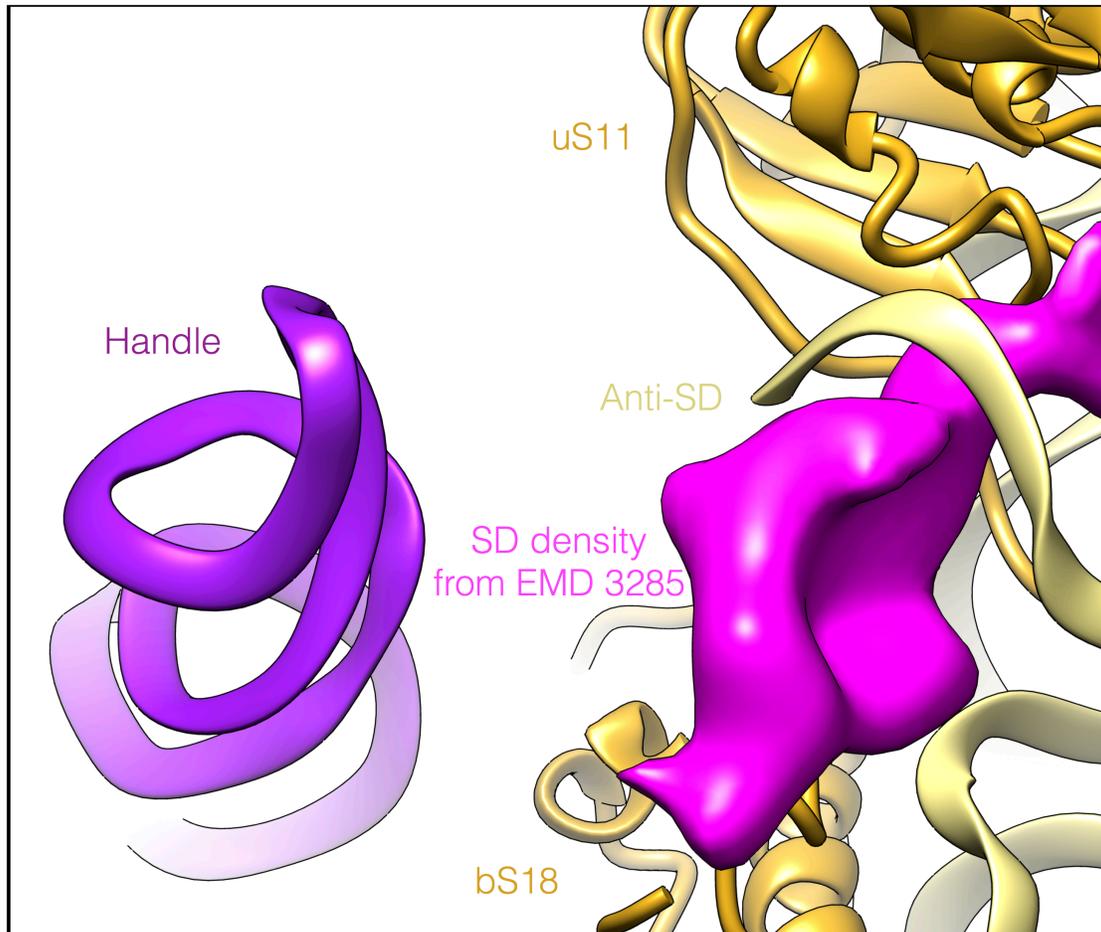


Figure 3.14. The handle in the rotated state is close to the mRNA exit
 The models of the handle and the SSU were built based on a representative map of the *M. tuberculosis* 70S in the rotated state. The density of the Shine-Dalgarno (SD) sequence of the mRNA was segmented from an *E. coli* ribosome density map (EMD 3285) to show where the mRNA exits. Since the tip of the handle and the mRNA are both flexible, it is possible that the tip of the handle may interfere with the mRNA.

Interestingly, in eukaryotic ribosomes, there is an rRNA expansion segment ES27, which undergoes conformational changes upon the rotation of the ribosomal subunits (Beckmann et al., 2001), similar to what have been observed for the handle of the *M. tuberculosis* ribosome. ES27 has been proposed to perform essential functions (Sweeney et al., 1994), such as coordinating nonribosomal factors or interfering with the nascent peptide around the peptide exit tunnel. The handle in *M. tuberculosis* may reach the exiting tRNA and mRNA and its potential functions are still under investigation. Future cryo-EM studies of the *M. tuberculosis* polysome may reveal potential interactions between the handle and the exiting mRNA.

3.3.4 Inhibitors visualized in the *M. tuberculosis* ribosome

The potent ribosomal inhibitors capreomycin and LZD-114 (a linezolid analog, **Figure 3.15**) were pre-incubated with the *M. tuberculosis* 70S and the *M. tuberculosis* 50S, respectively, prior to cryo-EM sample preparation. Capreomycin binds to the conserved decoding center, which is near helix h44 of the 16S (**Figure 3.16A, B**), in agreement with a crystal structure of capreomycin bound to the *Thermus thermophilus* ribosome (Stanley et al., 2010).

LZD-114 is about 20 times more potent against the *M. tuberculosis* ribosome than linezolid (the IC₅₀ of LZD-114 is 0.7 μM; the IC₅₀ of linezolid is 14 μM; **Figure 3.15**). The density of LZD-114 was observed in the peptidyl transferase center (PTC) of the *M. tuberculosis* 50S ribosomal subunit. It is bound in the same pocket and in a similar orientation to linezolid in other bacterial ribosomes (Eyal et al., 2015; Ippolito et al., 2008; Wilson et al., 2008). The improved potency of the LZD-114 may be due to the lack of a fluorine group in the B-ring and the substitution of the morpholine ring with a thiazole ring. The thiazole group is in close proximity to rRNA to form potential hydrogen bonds, *i.e.*, the sulfur in the thiazole ring and the

O2' of the U2744. Linezolid is not desirable as a TB treatment, because of an off-target inhibition of the human mitochondrial ribosome (mitoribosome) (Barnhill et al., 2012; Pacheu-Grau et al., 2013). A superimposition of the *M. tuberculosis* 70S and the human mitoribosome (**Figure 3.17**) (Amunts et al., 2015) showed that the binding site of LZD-114 is highly conserved, with only three residues difference between the two ribosomes within 10 Å of the LZD-114 binding site, namely A2296, C2745, and C2848 in the *M. tuberculosis* ribosome and G2721, U2994, and U3097 in the human mitoribosome. While the positions and orientations of these three nucleotide bases in the *M. tuberculosis* ribosome and the human mitoribosome are nearly identical, differences do exist, with one oxygen atom for each of the three bases (O6 in G2721, O4 in U2994 and U3097) of the human mitoribosome being replaced by a nitrogen atom (N6 in A2296, N4 in C2745 and C2848) in the *M. tuberculosis* ribosome, respectively. These small differences in the drug-binding pocket may be used to optimize linezolid analogs to avoid the off-target inhibition of the human mitoribosome.

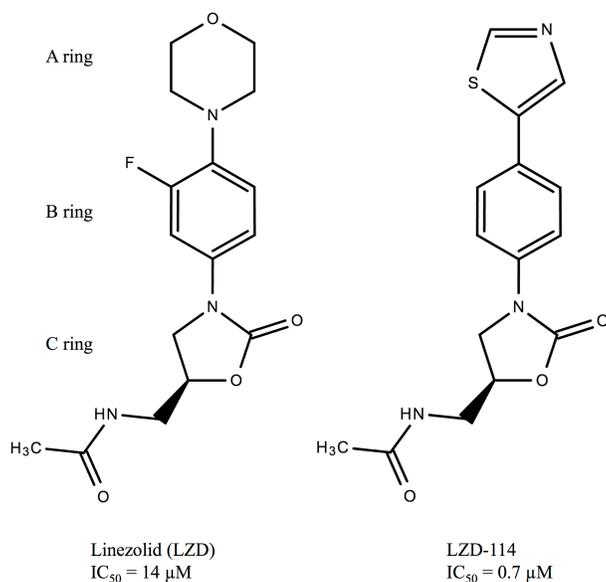


Figure 3.15. The comparison between linezolid and LZD-114
 Compared with linezolid, LZD-114 lacks a fluorine group in the B-ring, and the original morpholine ring is replaced by a thiazole ring in the A-ring. IC₅₀ denotes the half maximal inhibitory concentration. The values were measured in an in vitro *M. tuberculosis* translation system.

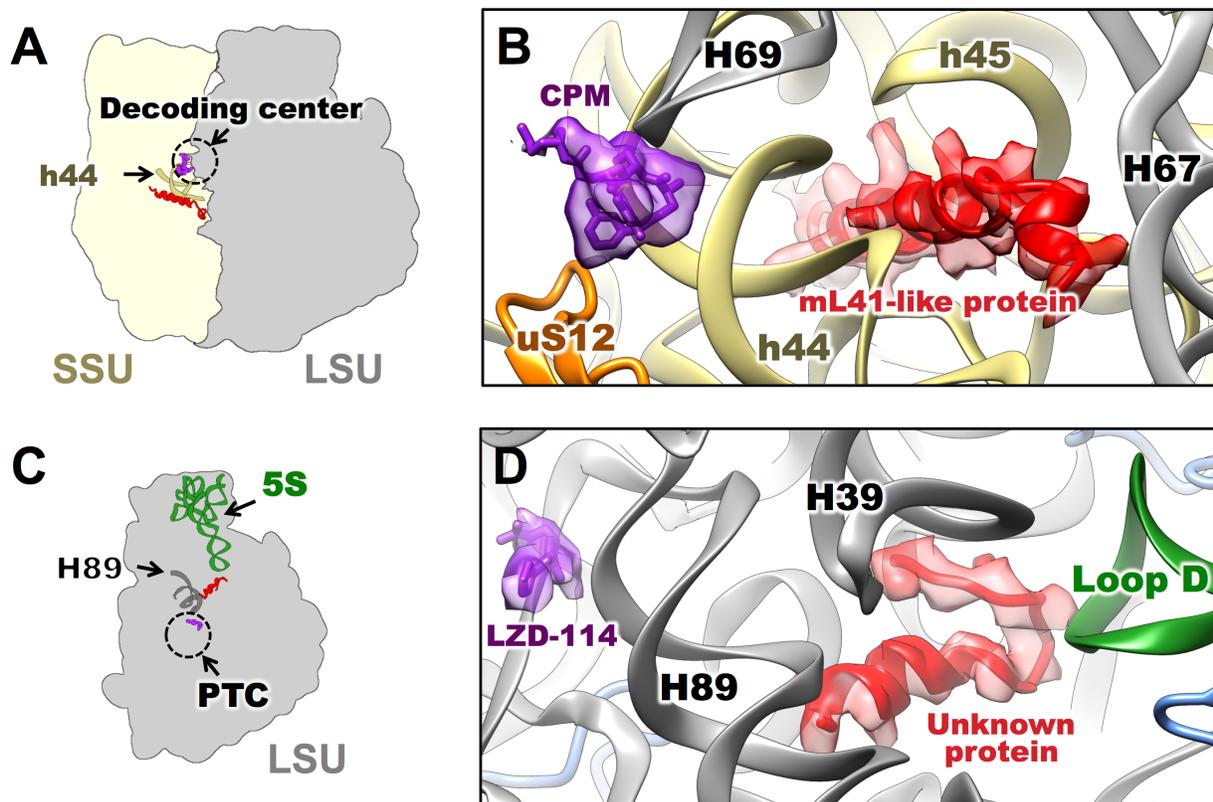


Figure 3.16. Unidentified proteins in the *M. tuberculosis* ribosomes discovered near the binding sites for two *M. tuberculosis* translation inhibitors

(A) Cartoon indicating the locations of the mL41-like protein (red), the antibiotic capreomycin (CPM, purple) in the structure of *M. tuberculosis* 70S incubated with capreomycin. Helix h44 of the 16S is in dark yellow. The dashed circle indicates the decoding center.

(B) Zoom-in view to show the models of CPM (purple) and polyaniline model of the mL41-like protein (red) and their neighboring environment. SSU protein uS12 (orange), helices h44, h45 of the 16S (yellow) and helices H67, H69 of the 23S (grey) are labeled.

(C) Cartoon indicating the locations of the unknown protein (red), the linezolid analog 114 (LZD-114, purple) in the structure of *M. tuberculosis* 50S incubated with LZD-114. The 5S is in green. The H89 is in dark gray. The dashed circle indicates the PTC.

(D) Zoom-in view to show the models of LZD-114 (purple) and polyaniline model of the unknown protein (red) and their neighboring environment. H39, H89 of the 23S (gray) and Loop D of the 5S (green) are labeled.

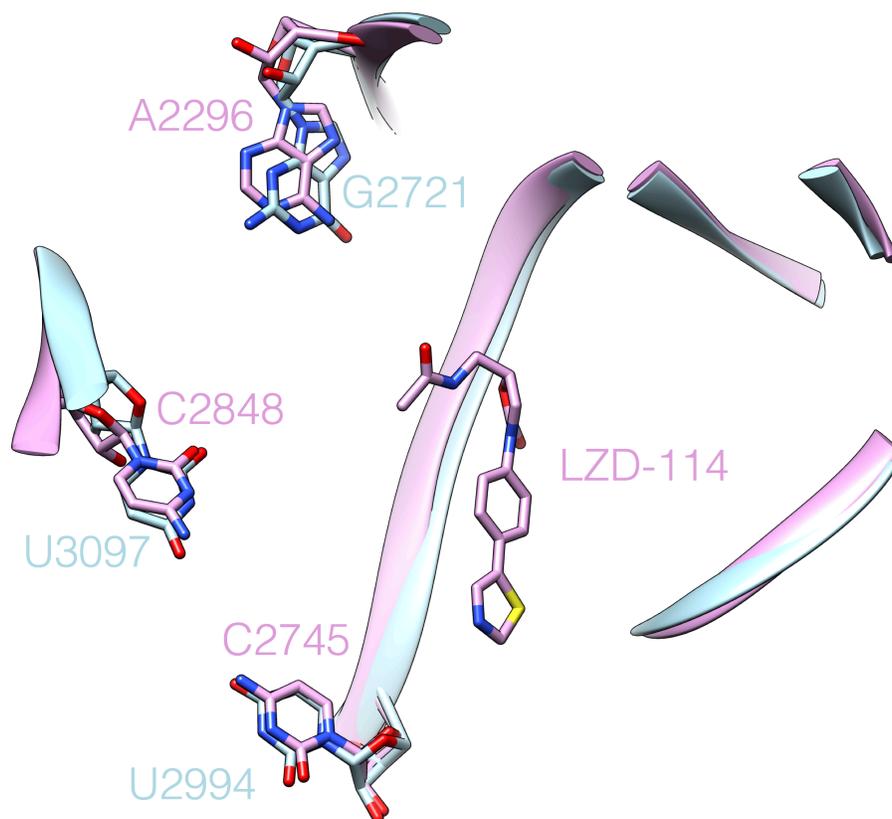


Figure 3.17. Comparison of the LZD-114 binding site in *M. tuberculosis* and mitochondrial ribosomes

The structure of the human mitochondrial ribosome (blue color, PDB ID 3J9M) is superimposed on the *M. tuberculosis* ribosome (red color) and only the residues, different between these two ribosomes, within 10 Å distance of LZD-114 are shown as stick models with their residue ID labeled.

3.3.5 Two new protein densities close to the functional sites of the *M. tuberculosis* ribosome

In the cryo-EM maps of the *M. tuberculosis* ribosomes, we observed densities for two distinct unknown proteins with clear α -helical geometries. One was positioned close to the decoding center where capreomycin binds, and the other close to the PTC near the LZD-114 binding site. Neither of these proteins has been observed in previously solved bacterial ribosome structures. Based on the full length and modelled sequences of nearby rProteins, these new structures do not correspond to any extensions of nearby rProteins.

The first protein, located close to the decoding center, is estimated, based on the density, to be about 30 amino acids long (**Figure 3.16A, B**). Interestingly, it has a very similar fold and is in the same location as that of eL41 and mL41, the rProteins found in eukaryotic cytosolic ribosomes and mitoribosomes, respectively (**Figure 3.18A-F**). The protein lies in a pocket formed by h44 and h45 from the 16S on the SSU and H67 and H69 from the 23S on the LSU, near the pivot point of intersubunit rotation (Ben-Shem et al., 2011). This pocket is highly conserved in ribosomes across species (Amunts et al., 2015; Ben-Shem et al., 2011; Cannone et al., 2002; Desai et al., 2017; Schlutzen et al., 2000; Wilson and Doudna Cate, 2012; Wimberly et al., 2000). While *E. coli* and other bacterial ribosomes have this pocket empty (Fischer et al., 2015), the ribosomes from *S. cerevisiae*, *H. sapiens*, *S. cerevisiae* mitochondria, and *H. sapiens* mitochondria all show an α -helical protein in this pocket (Amunts et al., 2015; Ben-Shem et al., 2011; Desai et al., 2017; Khatter et al., 2015), which overlaps with this “mL41-like protein” from *M. tuberculosis* (**Figure 3.18**).

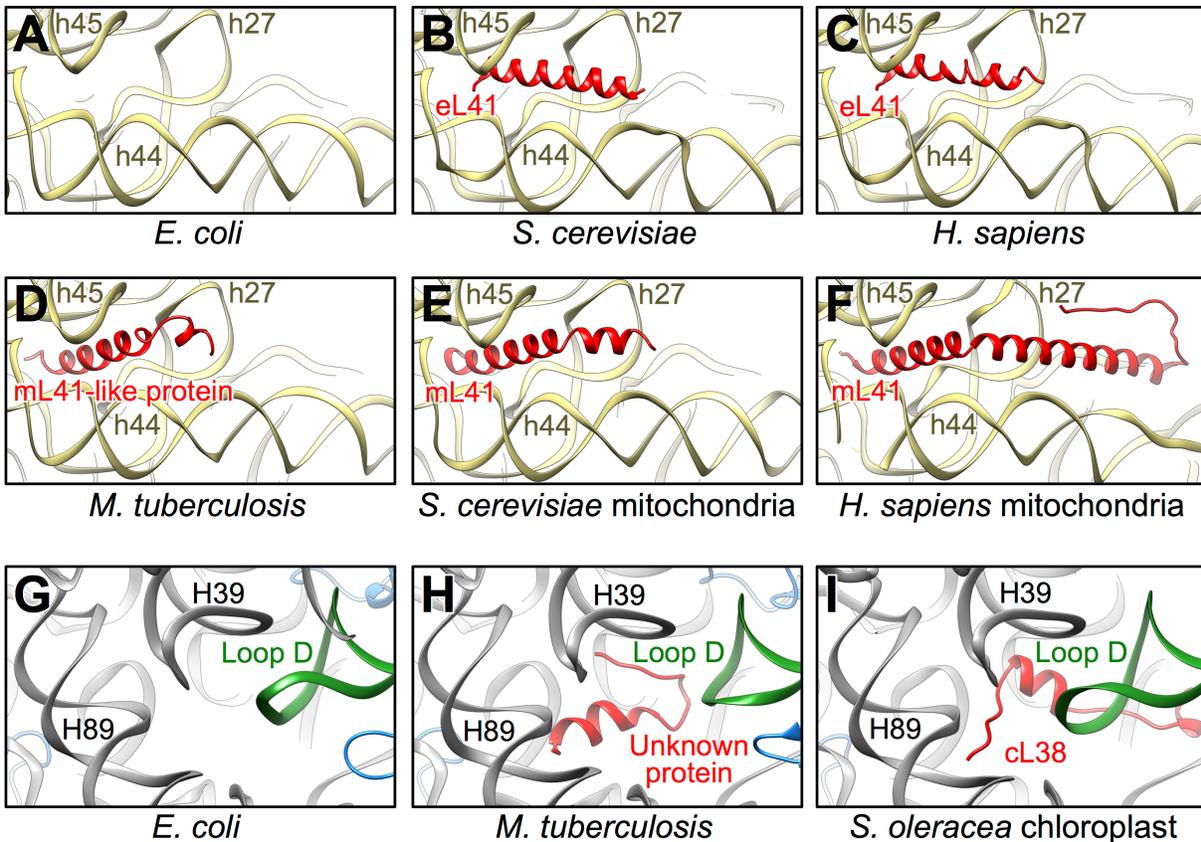


Figure 3.18. Comparison of the binding sites of the mL41-like protein and the unknown protein in different ribosomes

(A-F) Each of the six ribosomes from the labeled species was shown from the same viewing angle around the binding sites of eL41, mL41 and the mL41-like protein of the *M. tuberculosis* ribosome.

(G-I) Each of the three ribosomes from the labeled species was shown from the same viewing angle around the binding sites of cL38 and the unknown protein of the *M. tuberculosis* ribosome. The PDB ID for the *E. coli*, *S. cerevisiae*, *H. sapiens*, *S. cerevisiae* mitochondria, *H. sapiens* mitochondria and *S. oleracea* chloroplast are 5AFI, 4V88, 4UG0, 5MRC, 3J9M and 5MMM, respectively.

The second protein, close to the PTC, has not been identified in any bacterial, archaeal, yeast, or mammalian ribosome structures. It is an α -helix and coil loop that sits in a chamber formed by H39 and H89 of the 23S rRNA, and the loop D of the 5S rRNA. Building a polyalanine model into the density allowed us to estimate the unknown protein to be around 20 amino acids long. Related density is also observed in the 7.1 Å resolution cryo-EM structure of the ribosome from *M. smegmatis* (**Figure 3.19**). The pocket formed between H39 and H89 of rRNA 23S, and loop D of rRNA 5S is very similar among the 70S ribosomes from bacteria and chloroplasts. In the *E. coli* ribosome this pocket is empty, but both the *M. tuberculosis* and chloroplast ribosomes have a protein in this pocket (**Figure 3.17G-I**), that interacts with the loop D of rRNA 5S.

It has been proposed that the 5S rRNA is uniquely positioned to link all the functional centers of the ribosome. Previous studies have supported the hypothesis that the 5S acts as a physical transducer of information, facilitating communication between the different functional centers and coordinating multiple events catalyzed by the ribosome (Dinman, 2005). By interacting with loop D of the 5S, an rProtein in this location may stabilize the interactions between H38, H89, and the loop D of the 5S. The name of the corresponding protein in chloroplast ribosomes is cL38 (Bieri et al., 2017) or PSRP6 (Ahmed et al., 2016; Graf et al., 2016). The identity of this protein in *M. tuberculosis* is yet to be determined, but its orientation and length in the *M. tuberculosis* ribosome is different from cL38 (Ahmed et al., 2016; Bieri et al., 2017; Graf et al., 2016). Notably, one end of the helix of this protein is close to H89, a major component of the PTC.

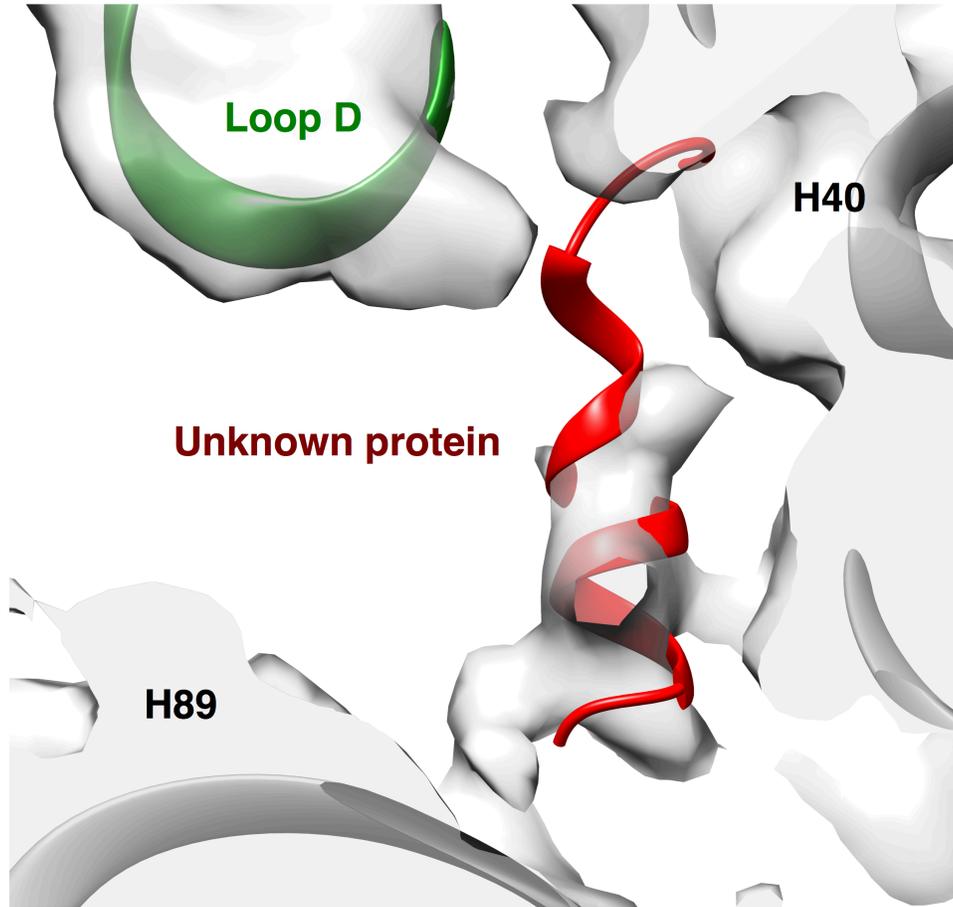


Figure 3.19. The unknown protein is also present in *M. smegmatis* ribosome. The model of *M. tuberculosis* LSU was fitted into the *M. smegmatis* ribosome. The helix of the unknown protein has a corresponding density that matches with each other.

In summary, we have shown the *M. tuberculosis* ribosome has several species-specific features, including the handle and two new proteins. In the *M. tuberculosis* 50S, the handle stays in a bent conformation, which blocks the association of the SSU. From the 50S to the 70S, the handle undergoes a large-scale conformational change to form intersubunit bridge B9. In the *M. tuberculosis* 70S, the handle shuttles between the L1 and the mRNA exiting site, which correlates with the rotation of the SSU.

While our manuscript was in review, a cryo-EM structure of the *M. smegmatis* 70S ribosome was reported (Hentschel et al., 2017). Despite the great interest of the unique handle conformation in 50S ribosomal subunit, the *M. smegmatis* 50S structure was not reported. Therefore, only the comparison of the 70S structures is possible. In the conserved core region, the *M. tuberculosis* 70S structure is in good agreement with the *M. smegmatis* 70S structure. However, in the peripheral region where most of the species-specific features exist, the *M. tuberculosis* 70S shows interesting difference from the *M. smegmatis* 70S structure. For example, the *M. smegmatis* paper reported an “altered” conformation of bL9, which was proposed as one possible effect of the kissing loop between H15 and H16a. However, such a conformation of bL9 in *M. smegmatis* may not be universal in mycobacteria, particularly in the human pathogen, *M. tuberculosis*. In our *M. tuberculosis* ribosome structure, the bL9 mainly exhibits a different conformation, which directly interact with the handle, even in the presence of a similar kissing loop (**Figure 3.9**). The fact that the same protein bL9 has distinct conformations in *M. tuberculosis* and *M. smegmatis*, further emphasizes the necessity and clinical relevance of performing structural studies using the exact pathogen, with respect to a closely-related non-pathogenic model organism.

3.4 Data Access

Cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under the accession numbers EMD-8641, 8645, 8646, 8647, 8648, and 8649 for *M. tuberculosis* 50S with LZD-114, *M. tuberculosis* 70S with capreomycin, the *M. tuberculosis* locally refined SSU, *M. smegmatis* 70S, the forty *M. tuberculosis* 70S conformations (as one entry) and *M. tuberculosis* 50S with 1 mM Mg²⁺, respectively. The atomic models of the *M. tuberculosis* 50S and 70S ribosomes have been deposited in the Protein Data Bank under the accession number 5V7Q and 5V93, respectively. Other data that support the findings of this study are available from the corresponding author upon request.

3.5 Acknowledgements

We acknowledge the Texas A&M High Performance Research Computing Center for providing the computational resources for the data processing. We thank Karl Gorzelnik and Tracey Musa for comments and editing of the paper. We thank Rhiju Das for the insightful discussion of the sarcin-ricin motif in the handle.

CHAPTER IV

STRUCTURES OF HIGHER ORDER COMPLEXES IN MYCOBACTERIAL GENE EXPRESSION

4.1 Introduction

Bacterial gene expression *in vivo* may involve higher order complexes of RNAP and ribosome, including the expressome (Kohler et al., 2017), polysome (Warner et al., 1962), and 100S ribosome (Wada et al., 1990). The expressome and polysome may contribute to the coupling between transcription and translation (Proshkin et al., 2010). The formation of hibernating 100S was found to be related to cell growth rate and responsible for fast recovery from stationary phase when fresh media was added (Aiso et al., 2005; Yamagishi et al., 1993).

Recently, the expressome of *E. coli* RNAP and 70S ribosome, connected only by mRNA and possibly glutaraldehyde, has been found to have a defined structure (Kohler et al., 2017). The exit site of mRNA on the RNAP was found to be directly interacting with the entry site of mRNA on the ribosome. Despite the discovery of this physiologically relevant configuration, two groups independently found that RNAP and the 30S subunit can directly interact with each other in the absence of any tethering by mRNA or cross-linking (Demo et al., 2017; Fan et al., 2017). These observations showed that the exit site of mRNA on the RNAP faces the exit site of mRNA on the 30S subunit. It is possible that both configurations exist and represent different stages of coupling transcription and translation.

To study the spatial arrangement of polysomes, both negative staining EM and cryo-EM have been used to image polysomes from *E. coli* (Brandt et al., 2009), or wheat germ (Kopeina et al., 2008; Myasnikov et al., 2014), confirming that polysomes can have an ordered structure.

Common characteristics can be found from these structures: 1) the mRNA threads through multiple ribosomes in a helical manner; 2) the nascent peptides emerging from each ribosome have a large distance from each other. Such configuration maximizes the translation efficiency and minimizes the interference between nascent peptides.

By contrast with actively translating ribosomes, the 70S particle can also form hibernating dimers, or 100S, mediated by a “ribosome modulation factor” (RMF) in *E. coli* (Wada et al., 1990). In addition to dimerization by RMF, there is a separate dimerizing mechanism caused by hibernation promoting factors (HPFs), which can then be divided into short HPFs and long HPFs. The majority of bacteria only have long HPFs, including *T. thermophilus*, *Bacillus subtilis* (*B. subtilis*), and *M. tuberculosis*. Interestingly, 100S formation by long HPFs is not limited to stationary phase growth, but spans all stages of bacterial growth. During exponential phase of cell growth, polysomes, including disomes, are also formed. Therefore, during sucrose density gradient separation of the cell lysate, the fraction corresponding to the ribosome dimer fraction can be the disome or hibernating 100S.

Since all these higher order complexes are only present in bacteria, there is a potential to design specific anti-tuberculosis drugs that do not affect humans. To achieve this goal, it is necessary to solve the 3D structures of these complexes and understand why their defined structures contribute to their functions. In this chapter, I will present the optimization of purifying *Mycobacterial* ribosome dimers and related cryo-EM studies.

4.2 Materials and Methods

4.2.1 Measurement of *M. smegmatis* growth rate under optimal growth media

A single colony of *M. smegmatis* 155² was picked after growth on a Middlebrook 7H10 agar plate, supplemented with Middlebrook OADC (Oleic acid, Albumin, Dextrose, Catalase) Enrichment. Then the single colony was inoculated into 2 mL optimal media A: Middlebrook 7H9 liquid media, supplemented with 0.5% (V/V) glycerol, 0.2% g/mL dextrose, 0.05% (V/V) Tween-80, 0.25 µg/mL malachite green. The growth condition was always at 37°C with shaking at 200 r.p.m. for liquid cultures throughout this chapter. After 24 - 36 h, the mid-log phase was roughly estimated based on the degree of turbidity. Then 1 mL liquid was used to inoculate 20 mL media A in a 125 mL flask. When the new culture reached the mid-log phase, 4 mL liquid was used to inoculate 0.5 L media A in a 2 L flask. Then at intervals of 1 or 2 hours, 1 mL liquid was removed from the flask and the apparent absorption (OD₆₀₀) was measured and recorded with a spectrophotometer. If the measured OD₆₀₀ was greater than 2, then the sample would be measured again after dilution, and a linear relationship between the OD₆₀₀ and the dilution was assumed.

4.2.2 Purification of *M. smegmatis* polysome

M. smegmatis 155² cells were cultured as described in 4.1.1. When the OD₆₀₀ of the culture reached ~0.5, the cells were harvested with or without fast cooling. In the fast cooling approach, a thick and 37°C-prewarmed glove was used to remove the flask from the shaker and pour the culture into a centrifuge bottle pre-packed with fresh ice. Once the culture reached 0°C, excess ice was removed and the centrifuge bottles were balanced before centrifugation at ~4,000 x g for 10 min at 4°C. The supernatant was decanted and then the pellet was washed with the

ribosome buffer A (20 mM Tris-HCl, 100 mM KCl, 10 mM MgCl₂, 6 mM 2-mercaptoethanol, pH 7.5). Then the cells were lysed using a French press at 20,000 psi. After lysis, the cell debris was removed by centrifugation at 10,000 x g for 10 min at 4°C. The cell lysate was then aliquoted into 1 mL tubes, fast cooled by liquid nitrogen and stored at -80°C.

Sucrose density gradient centrifugations were performed in two different ways. The first approach is the same as in **Chapter III**. Briefly, the sucrose density gradient was prepared by a “layering” method, *i.e.*, sequentially pumping 10%, 20%, 30%, 40% sucrose solutions (buffer A) into the bottom of a centrifuge tube. Then 1 mL cell lysate was carefully loaded onto the surface of the gradient. The centrifuge tubes were balanced and then centrifuged at 125,000 x g for 17 h at 4°C. The second approach uses a “thawing and freezing” method. The 40% and 15% sucrose solutions (buffer A) were first prepared. Then the two solutions were mixed at a 2:1 and 1:2 ratios, respectively. These four solutions were added to centrifuge tubes sequentially from highest density (40%) to lowest density (15%). After adding each layer of sucrose, the tube was frozen in liquid nitrogen for one minute before adding the next layer. Finally the tubes were stored at -80°C. The tubes were thawed at 4°C overnight before usage. Then a 1 mL aliquot of the cell lysate was carefully loaded onto the surface of the gradient. The tubes were balanced with buffer A and then centrifuged at 250,000 x g for 2h at 4°C.

After centrifugation, the solutions were pumped from bottom. The absorption was measured and recorded by the UV (260 nm) monitor from an AKTA chromatography system. The fractions corresponding to peaks of the UV absorption curve were collected. To concentrate the ribosomes and remove the sucrose, the sample was centrifuged using a protein concentrator with 30 kDa molecular weight cutoff. For “tandem” sucrose gradient centrifugation, a second centrifugation was performed after the concentration step.

4.2.3 Polysome profiling of *M. smegmatis* under different conditions

A total of 20 mL *M. smegmatis* 155² culture was obtained starting from a single colony, as described in 4.1.1. Four 2 L flasks each containing 0.5 L media A were each inoculated with 4 mL *M. smegmatis* culture. These flasks were placed in the same shaker at 37°C with shaking at 200 r.p.m. The culture from two flasks was used for polysome profiling at mid-log phase (OD₆₀₀ is ~0.5). Then, there are two remaining flasks. The culture from one of the two was transferred to a 0.5 L bottle. The bottle, full of culture, was sealed with parafilm and placed into the same shaker as the other remaining flask. After one day, the cells in the bottle was presumed to be under hypoxic conditions, while the cells in the flask was considered as in the stationary phase. Both of them were then subjected to the same polysome profiling procedure.

4.2.4 Preparation of *M. tuberculosis* expressome

The *M. tuberculosis* 30S was purified by Dr. Xiaojun Li using the same procedure described in Chapter III. The *M. tuberculosis* RNAP core was purified by Dr. Xuelin Bian using the same procedure described in Chapter II. The *M. tuberculosis* 30S was diluted to 20 uM, and RNAP to 80 uM. They were mixed at a 1:1 ratio before fractionation and storage at -80°C.

4.2.5 Electron microscopy

For negative stain EM, 4 µL *M. smegmatis* ribosome dimer fraction, or 4 µL 10 nM *M. tuberculosis* expressome, was applied to a glow discharged 200 copper mesh grid with continuous carbon film at room temperature. The sample was stained with 2% uranyl acetate.

The areas that showed negatively stained and mono-dispersed particles were imaged under a JEOL-1200 electron microscope or an FEI Tecnai F20 cryo electron microscope with a field emission gun (FEI company, Netherlands) operated at 200 kV. Images at different magnifications were recorded. Representative images shown in this chapter were all labeled with scale bars. For expressome, 267 micrographs were collected at a nominal magnification of 19,000 x with a pixel size of 1.87 Å on the specimen.

For cryo-EM, 3 µL of 100 nM *M. smegmatis* ribosome dimer, or *M. tuberculosis* expressome, was applied to a C-Flat 1.2/1.3 holey carbon grid, followed by plunge-freezing in a Mark III Vitrobot (FEI company, Netherlands). The temperature and relative humidity were set at 4°C and 100%, respectively. The grids were screened and imaged under an FEI Tecnai TF20 cryo electron microscope with a field emission gun (FEI company, Netherlands) operated at 200 kV. Data were recorded on a Gatan K2 Summit (Gatan, Pleasanton CA) direct detection camera in the super-resolution electron counting mode. A nominal magnification of 19,000 x yielded a pixel size of 1.87 Å on the specimen. The beam intensity was adjusted to a dose rate of 10 electrons per pixel per second on the camera. An 82-frame movie stack was recorded, with a 0.2 second exposure per frame. In total, 360 movie stacks for *M. smegmatis* ribosome dimer, and 218 movie stacks for *M. tuberculosis* expressome, were collected.

4.2.6 Image processing

The negative stain data of *M. tuberculosis* expressome was processed normally except that the CTF correction step was skipped due to the low resolution nature of negative stain. A total of 21,481 particles were considered as good after particle picking and screening in Relion 2.0 (Kimanius et al., 2016). The particles were classified into 3 classes with the *M. tuberculosis*

RNAP core (chapter II), 30S (chapter III) and a *M. tuberculosis* expressome model based on *E. coli* expressome (EMDB id: 7014) (Demo et al., 2017) as the three initial models.

The cryo-EM data of *M. tuberculosis* ribosome dimer was first processed by treating the particles as monosomes. In this case, the processing procedure is similar to previous methods (chapter II). The number of the initial particles is 90,452, out of which 35,107 particles were considered as good after 2D screening in Relion2.0 (Kimanius et al., 2016). These good particles were auto-refined, resulting in a reconstruction at 8.5 Å resolution, based on the gold-standard Fourier shell correlation (Scheres and Chen, 2012). To investigate the tRNA occupancy, these particles were analyzed using masked classification with signal subtraction (Bai et al., 2015) followed by unmasked refinement of the entire *M. tuberculosis* 70S for each class in Relion2.0 (Kimanius et al., 2016) (also see 3.1.3). In the 10 classes obtained, 3 classes have clear tRNA density. These 3 classes contain 32,430 particles, which were subsequently used for disome analysis. To remove the monosome, the images were binned extensively by a factor of 8, so that the 3D classification will be focused on the large difference between monosome and disome, instead of smaller differences, such as the ratcheting motion between the two subunits.

4.2.7 Molecular modeling

All the modeling in this chapter was done by rigid-body fitting in UCSF Chimera (Pettersen et al., 2004).

4.3 Results

4.3.1 *In silico* analysis of *M. tuberculosis* disome, 100S ribosome and expressome

Since the structures of *M. tuberculosis* RNAP (**chapter II**) and ribosome (**chapter III**) were solved, it became possible to computationally analyze *M. tuberculosis* disome, 100S ribosome and expressome, assuming the organizations of these higher order complexes are conserved to some extent.

For the disome, there are two structures of polysomes available, from *E. coli* (EMDB id: 1582) (Brandt et al., 2009) and wheat germ (EMDB id: 2790) (Myasnikov et al., 2014). The leading two ribosomes from each polysome were cut out and treated as a disome structure. The *M. tuberculosis* ribosome structure was rigid-body fit into each ribosome region of the disome structures. Two disome models were obtained, based on *E. coli* or wheat germ (**Figure 4.1 A, B**). The spatial organizations in these two models are similar, *i.e.*, the exit site of mRNA on the leading ribosome is facing towards the entry site of mRNA on the trailing ribosome. Such organization not only protects the fragile mRNA, but also maximizes the number of ribosomes on one mRNA molecule. Significant differences do exist. First, the distance between the mRNA entry and exit sites of adjacent ribosomes is larger in the wheat germ-based model (130Å) than in the *E. coli*-based model (60Å). Second, the trailing ribosome rotates relative to the leading ribosome in the wheat germ-based model, compared to the *E. coli*-based model. Obviously, the difference in distance is caused by the size difference between *E. coli* 70S ribosome and wheat germ 80S ribosome. The size of *M. tuberculosis* ribosome is more close to, but slightly larger than, the *E. coli* ribosome. Particularly, the unique handle of *M. tuberculosis* is in the contact region between the two ribosomes within the disome. Therefore, the *M. tuberculosis* disome is

likely to have a larger distance than *E. coli* disome, though the organization is expected to be similar.

For the 100S ribosome, there are several structures available, from *E. coli* (EMDB id: 1750) (Ortiz et al., 2010), *B. subtilis* (EMDB id: 3664) (Beckert et al., 2017), and *Staphylococcus aureus* (*S. aureus*) (EMDB ids: 3638, 3639) (Khusainov et al., 2017). Since the formation of the 100S ribosome is mediated by RMF in *E. coli*, which is distinct from long HPF in *B. subtilis*, *S. aureus*, and *M. tuberculosis*, the *E. coli* 100S structure was excluded for modeling. In addition, the *S. aureus* 100S was reported to have one “tight” and one “loose” conformations (the map quality of the tight conformation was better). The tight conformation of *S. aureus* 100S is more similar to the *B. subtilis* 100S structure and was chosen as a template. Following the same fitting strategy, two 100S models were obtained, based on *B. subtilis* or *S. aureus* tight conformation (**Figure 4.1C, D**). The spatial organization in these two models are similar, *i.e.*, the two ribosomes in the 100S are symmetric around the interaction site near the SSU. The physiological significance of such organization is still not clear, though it is known that the NTD of long HPF inhibits translation by blocking tRNA binding. The difference between the two models is the relative rotation between the two ribosomes within the 100S.

For the expressome, there are only two structures available, both from *E. coli*. One is a complex between 70S and RNAP, connected by mRNA and possibly cross-linking (EMDB id: 3580) (Kohler et al., 2017). The other is a complex between 30S and RNAP, stabilized by the interaction at the interface without any connecting molecules (EMDB id: 7014) (Demo et al., 2017). In addition, no consensus can be found between the two structures. Both structures were presumed to be correct and representing different functional states. Following the same strategy used in analyzing ribosome dimers, two expressome models were obtained (**Figure 4.1E, F**).

The RNAP binds to the mRNA entry site, while the other binds to the exit site. Such distinct difference stimulates the study of *M. tuberculosis* expressome.

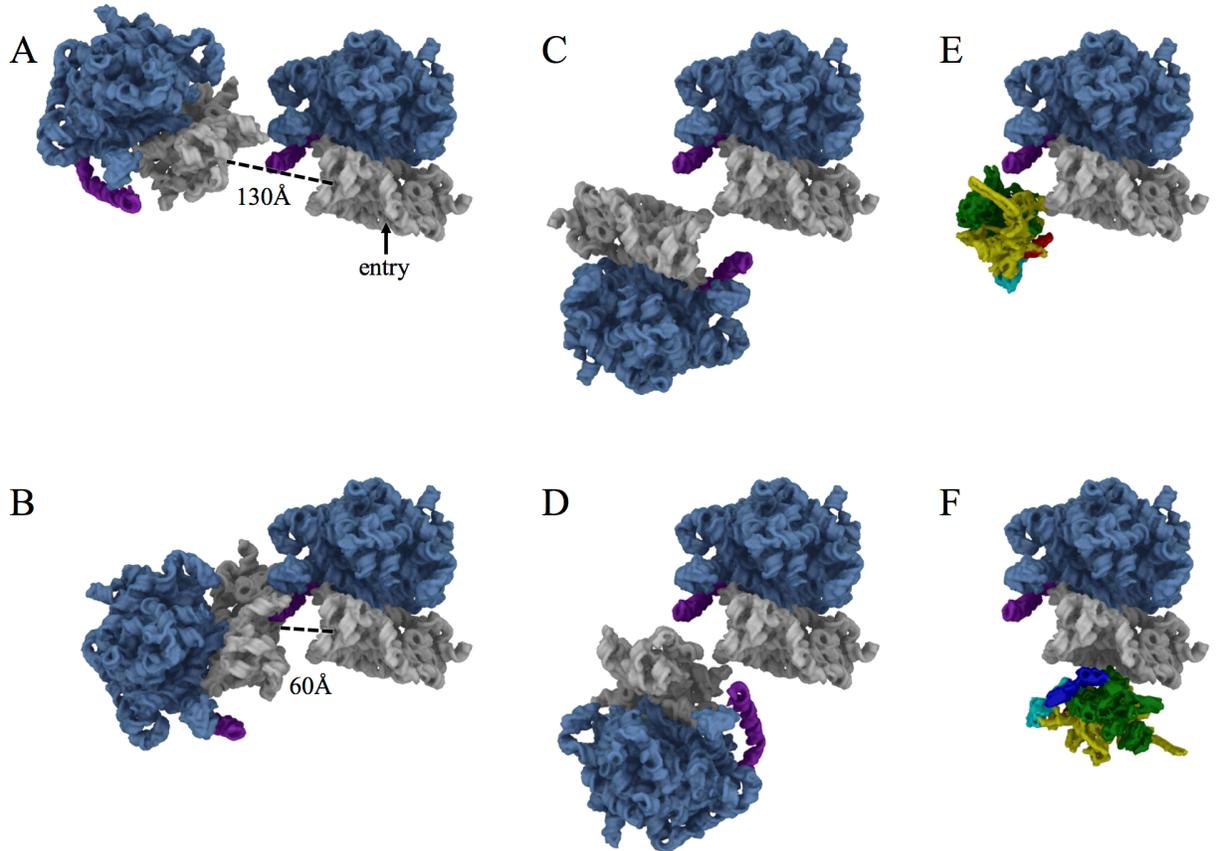


Figure 4.1. *In silico* analysis of *M. tuberculosis* disome, 100S and expressome

The ribosome is represented by the rRNAs for clarity, with the LSU colored as cornflower blue, SSU white, and the handle purple. The RNAP is colored the same as in Figure 2.1E, *i.e.*, α blue or cyan, β green, β' yellow, and ω red.

(A) A *M. tuberculosis* disome model based on wheat germ polysome (EMDB id: 2790).

(B) A *M. tuberculosis* disome model based on *E. coli* polysome (EMDB id: 1582).

(C) A *M. tuberculosis* 100S model based on *S. aureus* 100S (EMDB id: 3638).

(D) A *M. tuberculosis* 100S model based on *B. subtilis* 100S (EMDB id: 3664).

(E) A *M. tuberculosis* expressome model based on an *E. coli* expressome (EMDB id: 7014; 30S + RNAP).

(F) A *M. tuberculosis* expressome model based on another *E. coli* expressome (EMDB id: 3580; 70S + RNAP).

4.3.2 Optimization of polysome purification from mycobacteria

The purification of polysomes from mycobacteria may be hampered by the following two factors. The first is slow growth rate. Cell growth, a process of accumulating mass, is positively correlated with the total yield of translation. Polysomes are expected to occupy the translationally active mRNA pool, while monosomes are presumed to be translationally inactive or less productive (Heyer and Moore, 2016; Noll, 2008; Warner and Knopf, 2002). Thus the slow growing mycobacteria may not need too many polysomes. The second is the thick cell wall. The thick cell wall makes it more difficult to break down cells. The lysozyme method used in a popular bacterial polysome purification protocol (Qin and Fredrick, 2013) is not effective for mycobacteria cell lysis. Alternative methods of breaking down cells, including French press and Beadbeating, need harsher conditions to fully disrupt the mycobacterial cell wall, which may disrupt mRNA. Even after breakdown, the prevalence of cell wall debris may cause a significant level of contamination. To perform polysome profiling and get enough high-quality samples for EM studies, the optimization of polysome purification is necessary.

Presumably, the polysome is abundant in the exponential growth phase. So the growth curve of *M. smegmatis* 155² under an optimal condition was measured (**Figure 4.2A**). The growth curve showed a typical “S” shape for bacterial growth. The *M. smegmatis* entered the exponential phase around 10h after inoculation and the OD₆₀₀ was around 0.6. It kept growing at the maximum rate for at least 10h, after which the measurement may not be accurate because the OD₆₀₀ was assumed to be a linear relationship with dilution when it is larger than 2. Nevertheless, the OD₆₀₀ between 0.6 and 1 should be a good time to harvest the cells.

In initial trials of purifying polysomes, the method of purifying the 70S ribosome (chapter III) was used. The sedimentation profile did show a peak adjacent to the 70S peak

(**Figure 4.2B**). Samples corresponding to this peak were imaged under microscope. However, the particles are not ribosomes, based on the shape and size (**Figure 4.2C**). Instead, it may be the polyphosphate granules from mycobacteria (Ward et al., 2012). The later trial (see **Materials and Methods**), shown in **Figure 4.2D**, showed a much weaker dimer peak. Multiple peaks (disome, trisome, up to 6-mer) can be observed after lowering the input of cell lysate by 10 times (**Figure 4.2E**). The fraction of the dimer peak was imaged and validated as a dimer (**Figure 4.2G**). Indeed, they are ribosome dimers. In retrospect, this trial has the following improvements that might contribute to the success of obtaining ribosome dimers. First, the cell culture was quickly cooled by pouring it into a centrifuge bottle pre-packed with ice. Since the translation initiation takes much longer time than elongation, fast cooling stops both steps and minimizes the loss of polysome. Second, the preparation of the sucrose gradient was changed from “layering” to “thawing and freezing”. This means that when the centrifugation starts, the sucrose volume from “layering” is actually sucrose layers, while the one from “thawing and freezing” is presumably sucrose gradient. A better gradient results in higher resolution of separation. Third, the centrifugation parameters were change from 125,000 x g, 17h to 250,000 x g, 2h. Faster and shorter centrifugation is expected to separate the polysomes better. Despite the success in the visualization of multiple polysome peaks, the yield is tiny, around 1 to 2 orders of magnitude less than the 70S yield. More importantly, the purity of the dimer peak was expected to be low because of the contamination from the 70S peak. Therefore, a “tandem” sucrose gradient centrifugation (see **Materials and Methods**) was performed to separate the dimer and 70S better (**Figure 4.2F**).

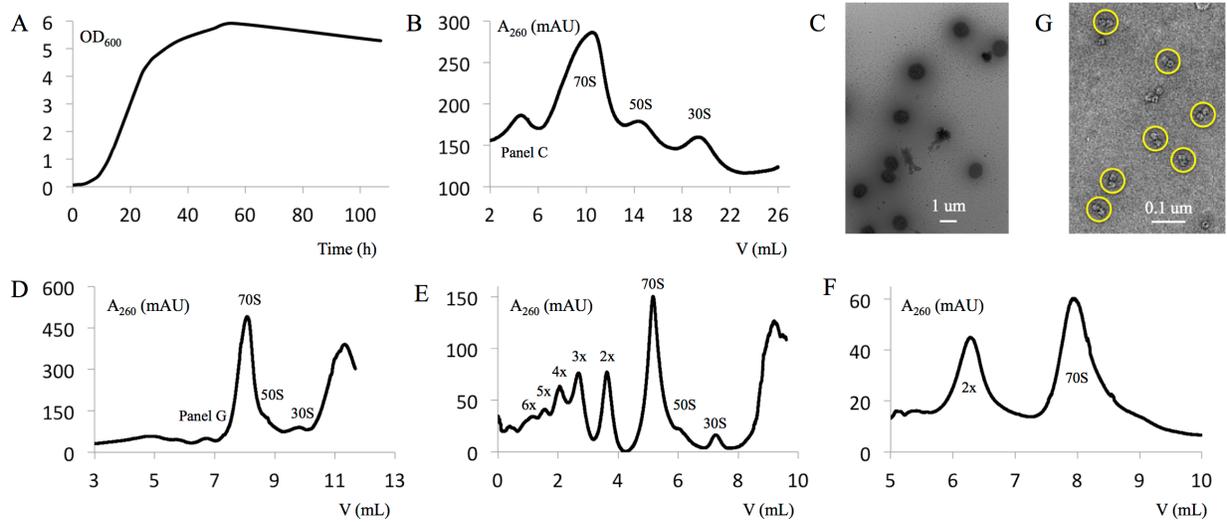


Figure 4.2. Optimization of polysome purification from mycobacteria
(A) Measurement of OD_{600} , monitoring the growth of *M. smegmatis* in liquid 7H9 media.
(B) Sedimentation profile of initial trials using sucrose density gradient centrifugation.
(C) A representative EM image of the assumed dimer peak in panel B.
(D) Sedimentation profile of sucrose density gradient centrifugation with improved procedures. The sucrose density gradient was prepared using the freezing-thawing method. The centrifugation parameters were change from 125,000 x g, 17h to 250,000 x g, 2h.
(E) Sedimentation profile from a repeat experiment of panel D with 10 times smaller input of cell lysate. The A_{260} values in the polysome region were multiplied by 10 to show the individual polysome peaks.
(F) Sedimentation profile from a tandem sucrose density gradient centrifugation.
(G) A representative EM image of the assumed dimer peak in panel D.

4.3.3 Polysome profiling of *M. smegmatis* under different conditions

One question remains from the previous purification of polysomes, *i.e.*, is the dimer an actively translating disome or hibernating 100S ribosome? One way to address this question is polysome profiling at different stages or conditions of cell growth. If the dimer peak corresponds to the 100S ribosome, then this peak will become larger and larger as time goes by, and eventually larger than the 70S peak (Akanuma et al., 2016). For this purpose and out of curiosity, I did the polysome profiling of *M. smegmatis* under three conditions (mid-log phase, stationary phase, and hypoxia condition; see **Materials and Methods**) (**Figure 4.3**). The dimer peak in the stationary phase was not larger than the one in the mid logarithmic phase, and disappeared in the hypoxia condition, suggesting this peak consists of actively translating disomes, instead of hibernating 100S ribosomes. In addition, the 50S peak was found to be significantly larger than the ones in mid logarithmic or stationary phase, despite the presence of considerable amount of 30S, suggesting the existence of a mechanism that inhibits translation by preventing the association of 50S and 30S, under hypoxia condition. For example, the ribosomal silencing factor S (RsfS) silences translation in the stationary phase using a similar mechanism (**Appendix I**) (Li et al., 2015).

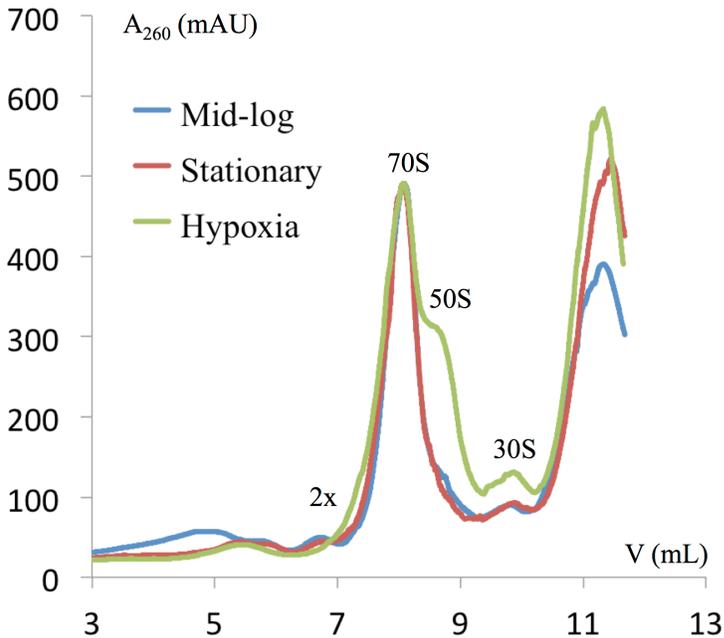


Figure 4.3. Polysome profiling of *M. smegmatis* under different conditions
The polysome profiles from mid-log, stationary, hypoxia conditions are colored as blue, red, and green, respectively. The three curves are aligned and normalized by the 70S peak.

4.3.4 Cryo-EM analysis of the ribosome dimers

Another way to figure out whether the dimer is disome or 100S is to directly solve its structure. If the dimer is disome, the ribosome must contain tRNAs, while the 100S does not have any tRNA. Therefore, I did cryo-EM imaging of the ribosome dimer fraction (**Figure 4.4A**). In the following tRNA occupancy analysis, three classes contain tRNA (**Figure 4.4B**). Interestingly, the number and conformation of tRNA are different in each class. The total number of particles is 35,107, out of which 13,774 (39%) particles have both A-site and P-site tRNAs, 10,803 (31%) particles have only P-site tRNA, 7,853 (22%) particles have both A-site and P-E-site tRNAs. This means that ~92% of the total particles contain at least one tRNA, suggesting that the dimer is actually actively translating disomes, rather than hibernating 100S ribosome.

Next, these 92% particles (that contain tRNA) were subject to disome analysis. To remove monosome, the particles were binned by 8 times and classified into 5 classes using a *M. tuberculosis* monosome as the initial model. Out of the 5 classes, the major (57%) class is a monosome, and another two classes are disomes, each accounting for ~11%. These two disome classes were found to be similar except that one is centered by the leading ribosome and the other by the trailing ribosome. The one centered by the leading ribosome is shown in **Figure 4.4C**. The map has two blobs, each of which can be fit with a ribosome structure. To compare with the models derived from structures of polysomes or 100S ribosomes from other species, a new model was built based on the disome map (**Figure 4.4D**). Interestingly, it is different from all other models. The distance between the exit site of mRNA on the leading ribosome and the entry site of mRNA on the trailing ribosome is 80Å, which is larger than the distance in the *E. coli* based model (60Å) and smaller than the distance in the wheat germ based model (130Å). The change of distance is expected as analyzed in **4.2.1**. Interestingly, the handle on the leading ribosome is not in the contact region between the two ribosomes as in the *E. coli* based model, suggesting the function of the handle is not related to polysomes in Mycobacteria. The new configuration of mycobacterial disome is closer, though not identical, to the 100S ribosomes. One advantage of this configuration is to avoid the tangling between the handle and the mRNA.

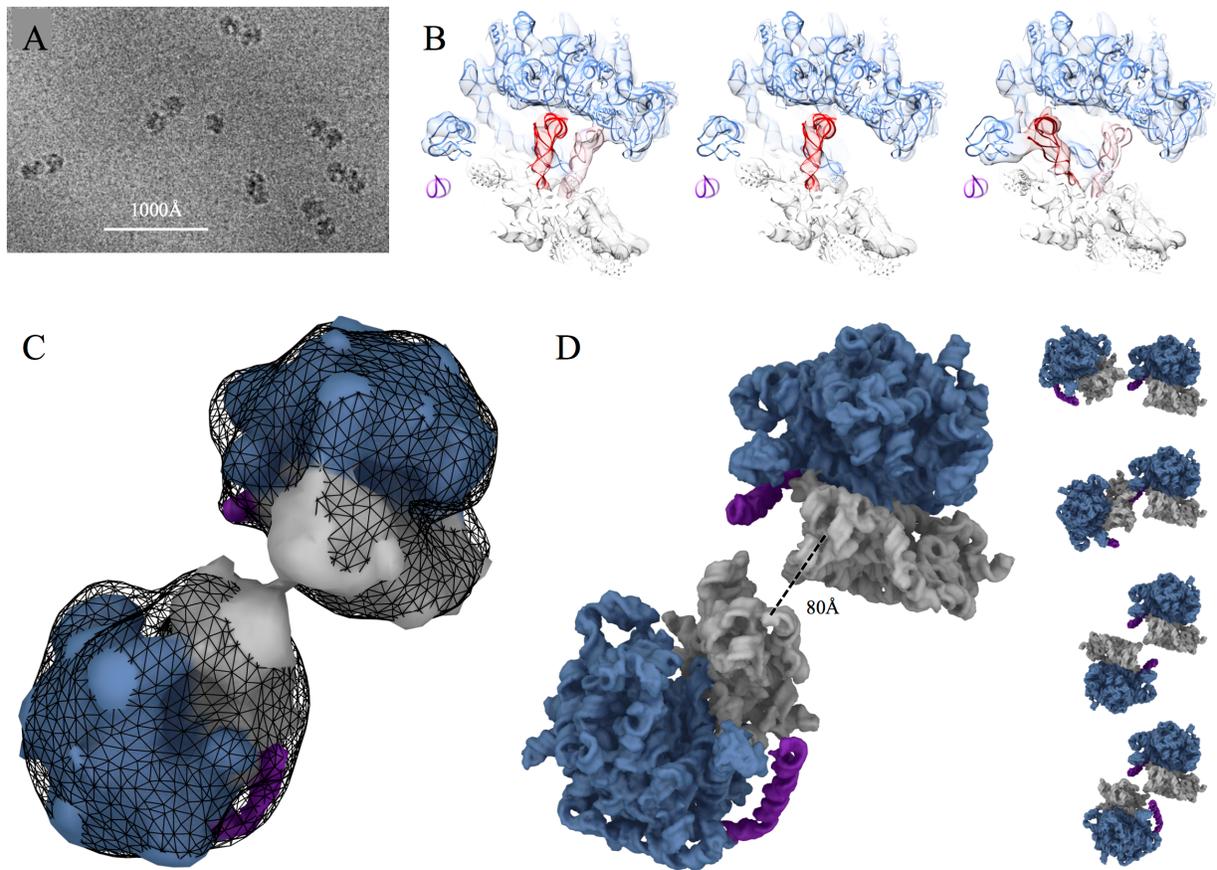


Figure 4.4. Cryo-EM analysis of the *M. tuberculosis* ribosome dimer

The color codes are similar to Figure 4.1, except that the reddish colors denote tRNA. A-site tRNA is colored pink; P-site tRNA red; P-E-site tRNA dark red.

A. Representative cryo-EM image of the *M. tuberculosis* ribosome dimer.

B. Three maps from the 3D classification showed clear tRNA densities. The map on the left has two tRNAs, A-site tRNA and P-site tRNA. The map in the middle has one tRNA, P-site tRNA. The map on the right has two tRNAs, A-site tRNA and P-E-site tRNA.

C. The two blobs from the disome reconstruction matches well with two ribosomes.

D. The *M. tuberculosis* disome model derived from the map from panel C is shown on the left. The four models from Figure 4.1 are shown on the right.

4.3.5 Preliminary studies of *M. tuberculosis* expressome

At the time of writing, the study of *M. tuberculosis* expressome is still underway and only some preliminary results have been obtained (**Figure 4.5**). The particles behave well under both negative-stain and cryo-EM. The main problem is that the RNAP appears to be much more abundant than the 30S particles.

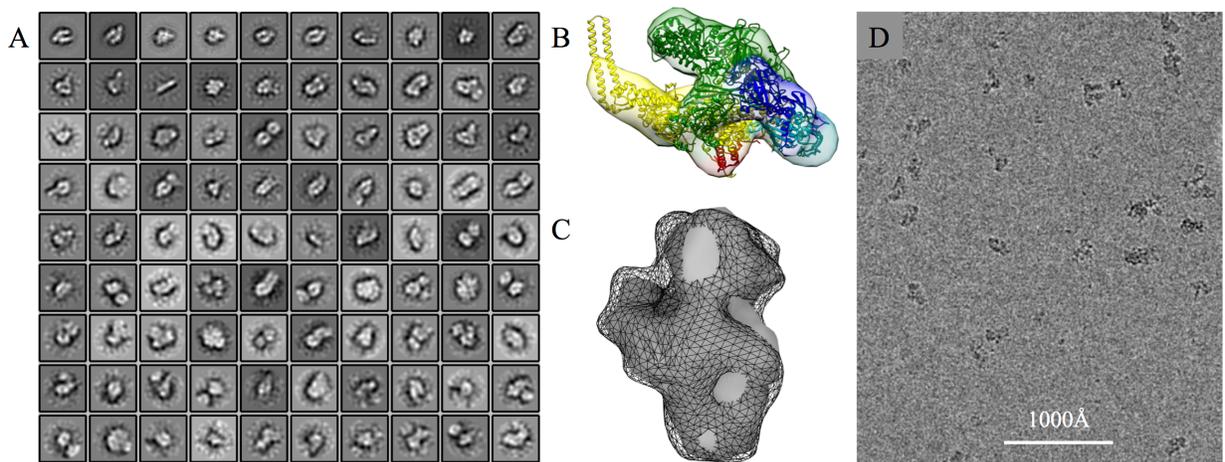


Figure 4.5. Preliminary cryo-EM study of *M. tuberculosis* expressome

A. Representative 2D class averages of the negative stain data. Most of them were actually the RNAP with a crab claw shape.

B. One representative class from 3D classification matches well with *M. tuberculosis* RNAP core. The subunits were colored the same as in Figure 2.1.

C. One representative class from 3D classification matches with *M. tuberculosis* 30S. Similar to Figure 4.4, the gray density is a map converted from *M. tuberculosis* 30S model, while the black mesh is the real map.

D. Representative image of cryo-EM data.

4.4 Discussion

To reconstitute the physiological scenario of gene expression in *M. tuberculosis*, we studied the higher order complexes involved in gene expression. Particularly, we have succeeded in the purification and cryo-EM imaging of actively translating disomes. The spatial organization of the two ribosomes in our disome structure is significantly different from the two published polysomes from other species. There might be three reasons for such difference. First, the difference is caused by the extra handle and kissing loop in mycobacterial ribosome. Second, the organization in polysome and disome may not be necessarily the same, *i.e.*, it is not appropriate to simplify the polysome as the disome. Third, our disome structure was purified in a “native” and actively translating condition, while the other two structures were inhibited by chloramphenicol and cycloheximide, respectively. Such strong inhibitors stall every ribosome and may cause a packing effect on the mRNA strand.

CHAPTER V

SUMMARY AND CONCLUSION

Tuberculosis remains a global health threat. To facilitate the development of anti-tuberculosis therapy, we studied gene expression, the major drug target, in *M. tuberculosis* and *M. smegmatis*. The key machineries of gene expression, including RNAP, ribosome, and their higher order complexes, have been visualized using the state-of-the-art structural approach, cryo-EM.

In the *M. tuberculosis* RNAP structure, the bridge B2 showed a potential conformational difference from the one in other species. Specifically, the B2 shifts about 10Å towards the active site and gets very close to the first-line TB drug, rifampicin. Future study should first focus on improving the resolution of the structure, especially around the B2 region, under the current buffer condition. A higher resolution cryo-EM map may not only confirm the new conformation, but also reveal the detailed interaction network around B2, which should explain the structural basis of the new conformation. The next question is what is the biological significance of the new conformation? Since the new conformation of B2 in the core complex overlaps with part of the template strand of DNA in RPo (**Figure 2.4C**), my hypothesis is that B2 precludes immature binding of DNA in the core complex and undergoes a conformational change, possibly facilitated by σ factors, to allow proper binding of DNA during or before the formation of RPo. Finally, the new conformation of B2 provides exciting potential to design rifampicin derivatives. These derivatives are expected to be more potent because 1) the additional interaction with the B2 may increase the binding affinity; 2) the derivatives may lock B2 in the new conformation, which blocks the binding of DNA.

The *M. tuberculosis* ribosome structures feature the long handle on the tRNA/mRNA exit side of the ribosome. The handle was found to be flexibly moving relative to the main body of the ribosome in 70S. To characterize a mobile fragment in cryo-EM, we have explored the following strategies. First, to solve a high resolution structure of the handle, we imaged the 50S ribosomal subunit where the handle is in the most stable state to our knowledge. Second, to get high quality density maps of the handle, we classified the particles based on the conformation of the handle, using the state-of-the-art signal subtraction approach. Third, to improve the accuracy of interpretation, we developed an RNA modeling method for low resolution cryo-EM maps that relies on identifying rigid structural motifs and sampling connecting regions between those motifs. Fourth, to characterize the movement, we used the statistical PCA analysis to derive the major motion of the handle, as well as other regions of the ribosome. It turned out that the major motion of the handle is correlated with that of the L1 stalk and the 30S subunit. Such correlation suggests that the handle may be involved in the tRNA and mRNA translocation. In addition, the comparison of the handles in 50S and 70S reveals a large conformational change of the handle, indicating its role in the joining of the two subunits. Two new ribosomal proteins were also found in the vicinity of the two important active sites: PTC and the decoding center.

Many questions about the *M. tuberculosis* ribosome remain to be answered. First, as the largest feature of the *M. tuberculosis* ribosome, does the handle have other functions, in addition to the suggested involvement in translation initiation and elongation? In this work, the startling discovery of the sarcin-ricin motif in the handle may provide a clue to the question. The sarcin-ricin motif, consisting of the GAA-AGUA sequence, is a very stable structure, mainly due to the large number of hydrogen-bonds per nucleotide (1.57 for the sarcin-ricin motif, larger than 1.50 for a stable GC helix) (Sripakdeevong et al., 2012). This motif is a recognition site for RNAs and

proteins (Correll et al., 2003; Moore, 1999), which may aid in target binding (Ulyanov et al., 2006). Of note, many toxins, including α -sarcin (Ackerman et al., 1988), ricin A-chain (Endo and Tsurugi, 1988), Shiga toxin (Tesh and O'Brien, 1991) and VapC20 (Winther et al., 2013), bind to the sarcin-ricin motif in the sarcin-ricin loop (Szewczak et al., 1993) and cleave the nearby rRNA. The sarcin-ricin motif in the handle is exposed to solvent, as in the sarcin-ricin loop, thus accessible to its potential binding partners. It is yet to be discovered that such binding partners do exist or that a cleavage can occur in the *M. tuberculosis* handle, which may serve as a “switch” to control the potential translational modulation by the handle. On the other hand, genetically deleting the handle may provide vital information on its function.

Second, what are the identities and functions of the two new proteins in the *M. tuberculosis* ribosome? Despite the importance of pointing out their existence and locations in the ribosome, it is still far from taking advantage of them in drug design, which will require learning their sequences. In this work, we named the new protein in the small subunit as mL41-like protein, because of its similar location and folding as the mL41 in the mitochondrial ribosome. We have also tried mass spectrometry but failed to obtain their sequences so far, probably due to their small sizes. Interestingly, these two proteins were claimed to be “identified by *de novo* structure interpretation” of a 3.3-Å resolution cryo-EM map (Hentschel et al., 2017). This *de novo* interpretation requires one to guess the identity of an amino acid based on its shape in the cryo-EM map. At 3.3-Å resolution, the shapes of two amino acids can be indistinguishable (Hryc et al., 2017) for the following reasons: 1) the difference between two amino acids can be as small as a carbon atom, which is far beyond the resolvability of a 3.3-Å resolution map; 2) negatively charged amino acids, such as aspartate and glutamate, usually show weak density beyond their C β -atoms, leading to a featureless bulge that resembles many other amino acids; 3)

cryo-EM maps contain significant level of noise, which can overwhelm the difference of amino acids. The problem of guessing amino acids based on shape is further complicated by the fact that a given amino acid can have many different shapes because of rotamers. Therefore, a rigorous verification of their “identified” sequences is necessary. There are three approaches to verify the sequences: 1) separate the new proteins by 2D gel before mass spectrometry; 2) add a fluorescent tag to the candidate sequences; 3) improving the resolution to $\sim 2\text{\AA}$. The functions of these two proteins can be studied by mutagenesis.

Third, how can the unique features in the *M. tuberculosis* ribosome be targeted for drug design, assuming the previous two questions have been addressed? One interesting direction, for instance, is to design a bipartite aptamer to lock the handle in the bent conformation, so that translation initiation is inhibited. The bipartite aptamer should have one end interacting with the handle and the other with the main body of the ribosome. RNA molecules in L-configuration may be great candidates, because of their intrinsic resistance to degradation by ribonucleases (Szczepanski and Joyce, 2013).

Higher order complexes of RNAP and ribosome may provide more understanding of the gene expression *in vivo*. We have explored the polysome purification and profiling in *M. smegmatis*. The ribosome dimer peak was further studied and confirmed to be actively translating disomes, rather than hibernating 100S. A low resolution reconstruction of the disome showed a reasonably different configuration from any ribosome dimers of other species. In this configuration, the handle is far from the interface between the two ribosomes, ruling out the possibility that the handle plays a role in polysomes as proposed by assuming the conservation between mycobacteria and *E. coli*. This work suggests that the actively translating disome is structurally flexible. Therefore, future study should consider using chloramphenicol or other

molecules to stabilize the disome before structural determination. In addition, gene expression under stress condition is another direction to explore. First, the mycobacterial 100S ribosome may be different from available 100S structures from other species, since the HPF in *M. tuberculosis* is longer. Second, as shown in this work (**Figure 4.3**), the 50S fraction is significantly higher in hypoxic condition than in normal conditions, indicating a difference of the 50S ribosomal subunit in hypoxic condition. Studying these differences may provide a better understanding of how *M. tuberculosis* can adapt to various growth conditions.

REFERENCES

- Ackerman, E.J., Saxena, S.K., and Ulbrich, N. (1988). Alpha-sarcin causes a specific cut in 28 S rRNA when microinjected into *Xenopus* oocytes. *J Biol Chem* *263*, 17076-17083.
- Adams, M.A., Luo, Y., Hove-Jensen, B., He, S.M., van Staalduinen, L.M., Zechel, D.L., and Jia, Z. (2008). Crystal structure of PhnH: an essential component of carbon-phosphorus lyase in *Escherichia coli*. *J Bacteriol* *190*, 1072-1083.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* *66*, 213-221.
- Adamski, F.M., Atkins, J.F., and Gesteland, R.F. (1996). Ribosomal protein L9 interactions with 23 S rRNA: the use of a translational bypass assay to study the effect of amino acid substitutions. *J Mol Biol* *261*, 357-371.
- Ahmed, T., Yin, Z., and Bhushan, S. (2016). Cryo-EM structure of the large subunit of the spinach chloroplast ribosome. *Sci Rep* *6*, 35793.
- Aiso, T., Yoshida, H., Wada, A., and Ohki, R. (2005). Modulation of mRNA stability participates in stationary-phase-specific expression of ribosome modulation factor. *J Bacteriol* *187*, 1951-1958.
- Akanuma, G., Kazo, Y., Tagami, K., Hiraoka, H., Yano, K., Suzuki, S., Hanai, R., Nanamiya, H., Kato-Yamada, Y., and Kawamura, F. (2016). Ribosome dimerization is essential for the efficient regrowth of *Bacillus subtilis*. *Microbiology* *162*, 448-458.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Ambudkar, S.V., Kim, I.W., Xia, D., and Sauna, Z.E. (2006). The A-loop, a novel conserved aromatic acid subdomain upstream of the Walker A motif in ABC transporters, is critical for ATP binding. *FEBS Lett* 580, 1049-1055.

Amunts, A., Brown, A., Toots, J., Scheres, S.H., and Ramakrishnan, V. (2015). Ribosome. The structure of the human mitochondrial ribosome. *Science* 348, 95-98.

Andersen, G.R., Nissen, P., and Nyborg, J. (2003). Elongation factors in protein biosynthesis. *Trends Biochem Sci* 28, 434-441.

Bai, X.C., Fernandez, I.S., McMullan, G., and Scheres, S.H. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* 2, e00461.

Bai, X.C., Rajendra, E., Yang, G., Shi, Y., and Scheres, S.H. (2015). Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* 4.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98, 10037-10041.

Bammes, B.E., Rochat, R.H., Jakana, J., Chen, D.H., and Chiu, W. (2012). Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *J Struct Biol* 177, 589-601.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905-920.

Barnhill, A.E., Brewer, M.T., and Carlson, S.A. (2012). Adverse effects of antimicrobials via predictable or idiosyncratic inhibition of host mitochondrial components. *Antimicrob Agents Chemother* 56, 4046-4051.

Barry, C.E. (2001a). *Mycobacterium smegmatis*: An absurd model for tuberculosis? *Trends Microbiol* 9, 473-474.

Barry, C.E., 3rd (2001b). Interpreting cell wall 'virulence factors' of *Mycobacterium tuberculosis*. *Trends Microbiol* 9, 237-241.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L., and Subramaniam, S. (2015). 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science* 348, 1147-1151.

Baugh, L.R., Hill, A.A., Brown, E.L., and Hunter, C.P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* 29, E29.

Beckert, B., Abdelshahid, M., Schafer, H., Steinchen, W., Arenz, S., Berninghausen, O., Beckmann, R., Bange, G., Turgay, K., and Wilson, D.N. (2017). Structure of the *Bacillus subtilis* hibernating 100S ribosome reveals the basis for 70S dimerization. *EMBO J* 36, 2061-2072.

Beckmann, R., Spahn, C.M., Eswar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J., and Blobel, G. (2001). Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* 107, 361-372.

Belogurov, G.A., Vassylyeva, M.N., Sevostyanova, A., Appleman, J.R., Xiang, A.X., Lira, R., Webber, S.E., Klyuyev, S., Nudler, E., Artsimovitch, I., *et al.* (2009). Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* 457, 332-335.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524-1529.

Benoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Berman, H.M., and Ebright, R.H. (2002). Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* 297, 1562-1566.

Berk, V., Zhang, W., Pai, R.D., and Cate, J.H. (2006). Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci U S A* 103, 15830-15834.

Bermingham, A., and Derrick, J.P. (2002). The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. *Bioessays* 24, 637-648.

Bieri, P., Leibundgut, M., Saurer, M., Boehringer, D., and Ban, N. (2017). The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J* 36, 475-486.

Bonocora, R.P., Smith, C., Lapierre, P., and Wade, J.T. (2015). Genome-Scale mapping of *Escherichia coli* σ 54 reveals widespread, conserved intragenic binding. *PLoS genetics* 11, e1005552.

Bortoluzzi, A., Muskett, F.W., Waters, L.C., Addis, P.W., Rieck, B., Munder, T., Schleier, S., Forti, F., Ghisotti, D., Carr, M.D., *et al.* (2013). Mycobacterium tuberculosis RNA polymerase-binding protein A (RbpA) and its interactions with sigma factors. *J Biol Chem* 288, 14438-14450.

Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell* 136, 261-271.

Brenner, S., Barnett, L., Katz, E.R., and Crick, F.H. (1967). UGA: a third nonsense triplet in the genetic code. *Nature* 213, 449-450.

Brenner, S., Stretton, A.O., and Kaplan, S. (1965). Genetic code: the 'nonsense' triplets for chain termination and their suppression. *Nature* 206, 994-998.

Brilot, A.F., Korostelev, A.A., Ermolenko, D.N., and Grigorieff, N. (2013). Structure of the ribosome with elongation factor G trapped in the pretranslocation state. *Proc Natl Acad Sci U S A* 110, 20994-20999.

Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E., and Rosch, P. (2010). A NusE:NusG complex links transcription and translation. *Science* 328, 501-504.

Burton, S.P., and Burton, Z.F. (2014). The sigma enigma: bacterial sigma factors, archaeal TFB and eukaryotic TFIIB are homologs. *Transcription* 5, e967599.

Busby, S., and Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). *J Mol Biol* 293, 199-213.

Busch, H. (1926). Berechnung der Bahn von Kathodenstrahlen im axialsymmetrischen elektromagnetischen Felde. *Annalen der Physik* 386, 974-993.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., *et al.* (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 433, 531-537.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 104, 901-912.

Campbell, E.A., Pavlova, O., Zenkin, N., Leon, F., Irschik, H., Jansen, R., Severinov, K., and Darst, S.A. (2005). Structural, functional, and genetic analysis of sorangicin inhibition of bacterial RNA polymerase. *EMBO J* 24, 674-682.

Campbell, M.G., Veesler, D., Cheng, A., Potter, C.S., and Carragher, B. (2015). 2.8 Å resolution reconstruction of the *Thermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. *Elife* 4.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., *et al.* (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3, 2.

Capecchi, M.R. (1967). Polypeptide chain termination in vitro: isolation of a release factor. *Proc Natl Acad Sci U S A* 58, 1144-1151.

Caskey, C.T., Beaudet, A.L., Scolnick, E.M., and Rosman, M. (1971). Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc Natl Acad Sci U S A* 68, 3163-3167.

Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* 70, 369-413.

Chao, M.C., and Rubin, E.J. (2010). Letting sleeping dogs lie: does dormancy play a role in tuberculosis? *Annu Rev Microbiol* 64, 293-311.

Chen, C.M., Ye, Q.Z., Zhu, Z.M., Wanner, B.L., and Walsh, C.T. (1990). Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the *phn* (*psiD*) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *J Biol Chem* 265, 4461-4471.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66, 12-21.

Chiaradia, L., Lefebvre, C., Parra, J., Marcoux, J., Burlet-Schiltz, O., Etienne, G., Tropis, M., and Daffe, M. (2017). Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane. *Sci Rep* 7, 12807.

Chou, F.C., Sripakdeevong, P., Dibrov, S.M., Hermann, T., and Das, R. (2013). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods* 10, 74-76.

Clark, B.F., and Nyborg, J. (1997). The ternary complex of EF-Tu and its role in protein biosynthesis. *Curr Opin Struct Biol* 7, 110-116.

Clough, R., Moldovan, G., and Kirkland, A. (2014). Direct detectors for electron microscopy. Paper presented at: *Journal of Physics: Conference Series* (IOP Publishing).

Colditz, G.A., Brewer, T.F., Berkey, C.S., Wilson, M.E., Burdick, E., Fineberg, H.V., and Mosteller, F. (1994). Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. *Jama* 271, 698-702.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544.

Condon, C., Liveris, D., Squires, C., Schwartz, I., and Squires, C.L. (1995). rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rrn* inactivation. *Journal of bacteriology* 177, 4152-4156.

Cook, G.M., Berney, M., Gebhard, S., Heinemann, M., Cox, R.A., Danilchanka, O., and Niederweis, M. (2009). Physiology of mycobacteria. *Adv Microb Physiol* 55, 81-182, 318-189.

Correll, C.C., Beneken, J., Plantinga, M.J., Lubbers, M., and Chan, Y.L. (2003). The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Res* 31, 6806-6818.

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., *et al.* (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 37, 337-352.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863-1876.

Crewe, A., Eggenberger, D., Wall, J., and Welter, L. (1968). Electron gun using a field emission source. *Review of Scientific Instruments* 39, 576-583.

Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J., and Klug, A. (1970). Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature* 226, 421-425.

Cudahy, P., and Shenoi, S.V. (2016). Diagnostics for pulmonary tuberculosis. *Postgrad Med J* 92, 187-193.

Czyz, A., Mooney, R.A., Iaconi, A., and Landick, R. (2014). Mycobacterial RNA Polymerase Requires a U-Tract at Intrinsic Terminators and Is Aided by NusG at Suboptimal Terminators. *Mbio* 5.

Daffe, M., and Draper, P. (1998). The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol* 39, 131-203.

Daniel, T.M. (2006). The history of tuberculosis. *Respir Med* 100, 1862-1870.

Darst, S.A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G., and Wriggers, W. (2002). Conformational flexibility of bacterial RNA polymerase. *Proc Natl Acad Sci U S A* 99, 4296-4301.

Das, R., and Baker, D. (2007). Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci U S A* 104, 14664-14669.

Dashti, A., Schwander, P., Langlois, R., Fung, R., Li, W., Hosseinizadeh, A., Liao, H.Y., Pallesen, J., Sharma, G., Stupina, V.A., *et al.* (2014). Trajectories of the ribosome as a Brownian nanomachine. *Proc Natl Acad Sci U S A* *111*, 17492-17497.

Davenne, T., and McShane, H. (2016). Why don't we have an effective tuberculosis vaccine yet? *Expert review of vaccines* *15*, 1009-1013.

Davidson, A.L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* *72*, 317-364, table of contents.

Davies, C., White, S.W., and Ramakrishnan, V. (1996). The crystal structure of ribosomal protein L14 reveals an important organizational component of the translational apparatus. *Structure* *4*, 55-66.

Davis, E., Chen, J., Leon, K., Darst, S.A., and Campbell, E.A. (2015). Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res* *43*, 433-445.

De Rosier, D.J., and Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. *Nature* *217*, 130-134.

Degen, D., Feng, Y., Zhang, Y., Ebricht, K.Y., Ebricht, Y.W., Gigliotti, M., Vahedian-Movahed, H., Mandal, S., Talaue, M., Connell, N., *et al.* (2014). Transcription inhibition by the depsipeptide antibiotic salinamide A. *Elife* *3*, e02451.

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A.B., Diaz-Avalos, R., Grigorieff, N., Nudler, E., and Korostelev, A.A. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* *6*.

Desai, N., Brown, A., Amunts, A., and Ramakrishnan, V. (2017). The structure of the yeast mitochondrial ribosome. *Science* *355*, 528-531.

Desikan, P. (2013). Sputum smear microscopy in tuberculosis: is it still relevant? *Indian J Med Res* *137*, 442-444.

Detjen, A.K., DiNardo, A.R., Leyden, J., Steingart, K.R., Menzies, D., Schiller, I., Dendukuri, N., and Mandalakas, A.M. (2015). Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. *Lancet Respir Med* *3*, 451-461.

Devi, S.K., Chichili, V.P., Jeyakanthan, J., Velmurugan, D., and Sivaraman, J. (2015). Structural basis for the hydrolysis of ATP by a nucleotide binding subunit of an amino acid ABC transporter from *Thermus thermophilus*. *J Struct Biol* *190*, 367-372.

DiMaio, F., Song, Y., Li, X., Brunner, M.J., Xu, C., Conticello, V., Egelman, E., Marlovits, T.C., Cheng, Y., and Baker, D. (2015). Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods* *12*, 361-365.

Dinman, J.D. (2005). 5S rRNA: Structure and Function from Head to Toe. *Int J Biomed Sci* *1*, 2-7.

Dorman, C.J. (2006). DNA supercoiling and bacterial gene expression. *Sci Prog* *89*, 151-166.

Drlica, K. (1990). Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet* *6*, 433-437.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowell, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* *21*, 129-228.

Elmlund, H., Elmlund, D., and Bengio, S. (2013). PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. *Structure* *21*, 1299-1306.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.

Endo, Y., and Tsurugi, K. (1988). The RNA N-glycosidase activity of ricin A-chain. The characteristics of the enzymatic activity of ricin A-chain with ribosomes and with rRNA. *J Biol Chem* 263, 8735-8739.

Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. *Nature* 502, 650-655.

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci U S A* 95, 9761-9766.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci Chapter 2, Unit 2 9*.

Eyal, Z., Matzov, D., Krupkin, M., Wekselman, I., Paukner, S., Zimmerman, E., Rozenberg, H., Bashan, A., and Yonath, A. (2015). Structural insights into species-specific features of the ribosome from the pathogen *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 112, E5805-5814.

Fan, H., Conn, A.B., Williams, P.B., Diggs, S., Hahm, J., Gamper, H.B., Jr., Hou, Y.M., O'Leary, S.E., Wang, Y., and Blaha, G.M. (2017). Transcription-translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res* 45, 11043-11055.

Fernandez-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M., Ruiz, F.M., Gruene, T., Legrand, P., Steuerwald, U., and Muller, C.W. (2013). Crystal structure of the 14-subunit RNA polymerase I. *Nature* 502, 644-649.

Fischer, N., Neumann, P., Konevega, A.L., Bock, L.V., Ficner, R., Rodnina, M.V., and Stark, H. (2015). Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* *520*, 567-570.

Flynn, J.L., and Chan, J. (2001). Tuberculosis: latency and reactivation. *Infect Immun* *69*, 4195-4201.

Fox, G.E. (2010). Origin and evolution of the ribosome. *Cold Spring Harb Perspect Biol* *2*, a003483.

Frank, J. (1975). Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* *1*, 159-162.

Frank, J., Goldfarb, W., Eisenberg, D., and Baker, T.S. (1978). Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy* *3*, 283-290.

Frank, J., Shimkin, B., and Dowse, H. (1981). Spider—A modular software system for electron image processing. *Ultramicroscopy* *6*, 343-357.

Freistroffer, D.V., Pavlov, M.Y., MacDougall, J., Buckingham, R.H., and Ehrenberg, M. (1997). Release factor RF3 in E.coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J* *16*, 4126-4133.

Fung, S., Nishimura, T., Sasarman, F., and Shoubridge, E.A. (2013). The conserved interaction of C7orf30 with MRPL14 promotes biogenesis of the mitochondrial large ribosomal subunit and mitochondrial translation. *Mol Biol Cell* *24*, 184-193.

Futcher, B. (1988). Supercoiling and transcription, or vice versa? *Trends Genet* *4*, 271-272.

Galagan, J.E., Sisk, P., Stolte, C., Weiner, B., Koehrsen, M., Wymore, F., Reddy, T.B., Zucker, J.D., Engels, R., Gellesch, M., *et al.* (2010). TB database 2010: overview and update. *Tuberculosis (Edinb)* *90*, 225-235.

Gao, Y.G., Selmer, M., Dunham, C.M., Weixlbaumer, A., Kelley, A.C., and Ramakrishnan, V. (2009). The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* 326, 694-699.

Gartmann, M., Blau, M., Armache, J.P., Mielke, T., Topf, M., and Beckmann, R. (2010). Mechanism of eIF6-mediated inhibition of ribosomal subunit joining. *J Biol Chem* 285, 14848-14851.

Gomez, J.E., and McKinney, J.D. (2004). M. tuberculosis persistence, latency, and drug tolerance. *Tuberculosis (Edinb)* 84, 29-44.

Gomez Ramos, L.M., Smeekens, J.M., Kovacs, N.A., Bowman, J.C., Wartell, R.M., Wu, R., and Williams, L.D. (2016). Yeast rRNA Expansion Segments: Folding and Function. *J Mol Biol* 428, 4048-4059.

Graf, M., Arenz, S., Huter, P., Donhofer, A., Novacek, J., and Wilson, D.N. (2016). Cryo-EM structure of the spinach chloroplast ribosome reveals the location of plastid-specific ribosomal proteins and extensions. *Nucleic Acids Res.*

Grant, B.J., Rodrigues, A.P., ElSawy, K.M., McCammon, J.A., and Caves, L.S. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22, 2695-2696.

Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife* 4, e06980.

Green, R., and Noller, H.F. (1997). Ribosomes and translation. *Annual review of biochemistry* 66, 679-716.

Guex, N., Peitsch, M.C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30 Suppl 1, S162-173.

Gulten, G., and Sacchettini, J.C. (2013). Structure of the Mtb CarD/RNAP beta-lobes complex reveals the molecular basis of interaction and presents a distinct DNA-binding domain for Mtb CarD. *Structure* *21*, 1859-1869.

Gunther, G. (2014). Multidrug-resistant and extensively drug-resistant tuberculosis: a review of current concepts and future challenges. *Clin Med (Lond)* *14*, 279-285.

Harada, Y., Ohara, O., Takatsuki, A., Itoh, H., Shimamoto, N., and Kinoshita, K., Jr. (2001). Direct observation of DNA rotation during transcription by Escherichia coli RNA polymerase. *Nature* *409*, 113-115.

Harley, C.B., and Reynolds, R.P. (1987). Analysis of E. coli promoter sequences. *Nucleic Acids Res* *15*, 2343-2361.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* *107*, 679-688.

Harries, A.D., and Dye, C. (2006). Tuberculosis. *Ann Trop Med Parasitol* *100*, 415-431.

Hauser, R., Pech, M., Kijek, J., Yamamoto, H., Titz, B., Naeve, F., Tovchigrechko, A., Yamamoto, K., Szaflarski, W., Takeuchi, N., *et al.* (2012). RsfA (YbeB) proteins are conserved ribosomal silencing factors. *PLoS Genet* *8*, e1002815.

Henderson, R., and Unwin, P.N. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* *257*, 28-32.

Hentschel, J., Burnside, C., Mignot, I., Leibundgut, M., Boehringer, D., and Ban, N. (2017). The Complete Structure of the Mycobacterium smegmatis 70S Ribosome. *Cell Rep* *20*, 149-160.

HersHKovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., *et al.* (2008). Detection and molecular

characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* 3, e3426.

Hett, E.C., and Rubin, E.J. (2008). Bacterial growth and cell division: a mycobacterial perspective. *Microbiol Mol Biol Rev* 72, 126-156, table of contents.

Heyer, E.E., and Moore, M.J. (2016). Redefining the Translational Status of 80S Monosomes. *Cell* 164, 757-769.

Heymann, J.B., and Belnap, D.M. (2007). Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* 157, 3-18.

Hillebrand, A., Wurm, R., Menzel, A., and Wagner, R. (2005). The seven *E. coli* ribosomal RNA operon upstream regulatory regions differ in structure and transcription factor binding efficiencies. *Biol Chem* 386, 523-534.

Hirashima, A., and Kaji, A. (1973). Role of elongation factor G and a protein factor on the release of ribosomes from messenger ribonucleic acid. *J Biol Chem* 248, 7580-7587.

Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451, 851-854.

Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W.J., Sachse, C., and Muller, C.W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. *Nature* 528, 231-236.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. *J Struct Biol* 157, 47-55.

Hoppe, W., Gassmann, J., Hunsmann, N., Schramm, H.J., and Sturm, M. (1974). Three-dimensional reconstruction of individual negatively stained yeast fatty-acid synthetase molecules from tilt series in the electron microscope. *Hoppe Seylers Z Physiol Chem* 355, 1483-1487.

Horsburgh Jr, C.R., Barry III, C.E., and Lange, C. (2015). Treatment of tuberculosis. *New England Journal of Medicine* 373, 2149-2160.

Houben, R.M., and Dodd, P.J. (2016). The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLoS medicine* 13, e1002152.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008, 420747.

Hryc, C.F., Chen, D.H., Afonine, P.V., Jakana, J., Wang, Z., Haase-Pettingell, C., Jiang, W., Adams, P.D., King, J.A., Schmid, M.F., *et al.* (2017). Accurate model annotation of a near-atomic resolution cryo-EM map. *Proc Natl Acad Sci U S A* 114, 3103-3108.

Hubin, E.A., Fay, A., Xu, C., Bean, J.M., Saecker, R.M., Glickman, M.S., Darst, S.A., and Campbell, E.A. (2017a). Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *Elife* 6.

Hubin, E.A., Lilic, M., Darst, S.A., and Campbell, E.A. (2017b). Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. *Nat Commun* 8, 16072.

Hubin, E.A., Tabib-Salazar, A., Humphrey, L.J., Flack, J.E., Olinares, P.D., Darst, S.A., Campbell, E.A., and Paget, M.S. (2015). Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proc Natl Acad Sci U S A* 112, 7171-7176.

Hurwitz, J. (2005). The discovery of RNA polymerase. *J Biol Chem* 280, 42477-42485.

Hussain, T., Llacer, J.L., Wimberly, B.T., Kieft, J.S., and Ramakrishnan, V. (2016). Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell* *167*, 133-144 e113.

Huxley, H.E., and Zubay, G. (1961). Preferential staining of nucleic acid-containing structures for electron microscopy. *J Biophys Biochem Cytol* *11*, 273-296.

Ippolito, J.A., Kanyo, Z.F., Wang, D., Franceschi, F.J., Moore, P.B., Steitz, T.A., and Duffy, E.M. (2008). Crystal structure of the oxazolidinone antibiotic linezolid bound to the 50S ribosomal subunit. *J Med Chem* *51*, 3353-3356.

Ivanov, A.V., Malygin, A.A., and Karpova, G.G. (2013). [Mg²⁺ ions affect the structure of the central domain of the 18S rRNA in the vicinity of the ribosomal protein S13 binding site]. *Mol Biol (Mosk)* *47*, 157-166.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* *11*, 113-127.

Jaiswal, R.K., Prabha, T.S., Manjeera, G., and Gopal, B. (2013). Mycobacterium tuberculosis RsdA provides a conformational rationale for selective regulation of sigma-factor activity by proteolysis. *Nucleic Acids Res* *41*, 3414-3423.

James, B.W., Williams, A., and Marsh, P.D. (2000). The physiology and pathogenicity of Mycobacterium tuberculosis grown under controlled conditions in a defined medium. *J Appl Microbiol* *88*, 669-677.

Javid, B., Sorrentino, F., Toosky, M., Zheng, W., Pinkham, J.T., Jain, N., Pan, M., Deighan, P., and Rubin, E.J. (2014). Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. *Proc Natl Acad Sci U S A* *111*, 1132-1137.

Jeon, Y.H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., and Kyogoku, Y. (1995). Solution structure of the activator contact domain of the RNA polymerase alpha subunit. *Science* 270, 1495-1497.

Jiang, J., Pentelute, B.L., Collier, R.J., and Zhou, Z.H. (2015). Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* 521, 545-549.

Jiang, M., Sullivan, S.M., Walker, A.K., Strahler, J.R., Andrews, P.C., and Maddock, J.R. (2007). Identification of novel *Escherichia coli* ribosome-associated proteins using isobaric tags and multidimensional protein identification techniques. *J Bacteriol* 189, 3434-3444.

Jochimsen, B., Lolle, S., McSorley, F.R., Nabi, M., Stougaard, J., Zechel, D.L., and Hove-Jensen, B. (2011). Five phosphonate operon gene products as components of a multi-subunit complex of the carbon-phosphorus lyase pathway. *Proc Natl Acad Sci U S A* 108, 11393-11398.

Jossinet, F., and Westhof, E. (2005). Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* 21, 3320-3321.

Kamat, S.S., Williams, H.J., Dangott, L.J., Chakrabarti, M., and Raushel, F.M. (2013). The catalytic mechanism for aerobic formation of methane by bacteria. *Nature* 497, 132-136.

Kamat, S.S., Williams, H.J., and Raushel, F.M. (2011). Intermediates in the transformation of phosphonates to phosphate by bacteria. *Nature* 480, 570-573.

Khatter, H., Myasnikov, A.G., Natchiar, S.K., and Klaholz, B.P. (2015). Structure of the human 80S ribosome. *Nature* 520, 640-645.

Khusainov, I., Vicens, Q., Ayupov, R., Usachev, K., Myasnikov, A., Simonetti, A., Validov, S., Kieffer, B., Yusupova, G., Yusupov, M., *et al.* (2017). Structures and dynamics of hibernating ribosomes from *Staphylococcus aureus* mediated by intermolecular interactions of HPF. *EMBO J* 36, 2073-2087.

Kim, J., Wu, S., Tomasiak, T.M., Mergel, C., Winter, M.B., Stiller, S.B., Robles-Colmanares, Y., Stroud, R.M., Tampe, R., Craik, C.S., *et al.* (2015). Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter. *Nature* 517, 396-400.

Kimanius, D., Forsberg, B.O., Scheres, S.H., and Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* 5.

Kireeva, M.L., Kashlev, M., and Burton, Z.F. (2013). RNA polymerase structure, function, regulation, dynamics, fidelity, and roles in gene expression (ACS Publications).

Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* 334, 941-948.

Koch, A., Mizrahi, V., and Warner, D.F. (2014). The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? *Emerg Microbes Infect* 3, e17.

Koch, R. (1882). Die Aetiologie der Tuberculose. *Berliner Klinische Wochenzeitschrift* 19, 239–253.

Kohler, R., Mooney, R.A., Mills, D.J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science* 356, 194-197.

Kopeina, G.S., Afonina, Z.A., Gromova, K.V., Shirokov, V.A., Vasiliev, V.D., and Spirin, A.S. (2008). Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. *Nucleic Acids Res* 36, 2476-2488.

Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F. (2006). Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* 126, 1065-1077.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361, 13-37.

Kresge, N., Simoni, R.D., and Hill, R.L. (2006). The discovery and isolation of RNA polymerase by Jerard Hurwitz. *Journal of Biological Chemistry* 281, e12-e12.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.

KUBICA, G.P., KIM, T.H., and DUNBAR, F.P. (1972). Designation of Strain H37Rv as the Neotype of *Mycobacterium tuberculosis*. *International Journal of Systematic and Evolutionary Microbiology* 22, 99-106.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 11, 63-65.

Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science* 343, 1443-1444.

Kumar, A., Majid, M., Kunisch, R., Rani, P.S., Qureshi, I.A., Lewin, A., Hasnain, S.E., and Ahmed, N. (2012). *Mycobacterium tuberculosis* DosR regulon gene Rv0079 encodes a putative, 'dormancy associated translation inhibitor (DATIN)'. *PLoS One* 7, e38709.

Kwan, C.K., and Ernst, J.D. (2011). HIV and tuberculosis: a deadly human syndemic. *Clin Microbiol Rev* 24, 351-376.

Lane, W.J., and Darst, S.A. (2010). Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J Mol Biol* 395, 671-685.

Laursen, B.S., Sorensen, H.P., Mortensen, K.K., and Sperling-Petersen, H.U. (2005). Initiation of protein synthesis in bacteria. *Microbiol Mol Biol Rev* 69, 101-123.

Leontis, N.B., and Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol* 283, 571-583.

Lewin, A., and Sharbati-Tehrani, S. (2005). [Slow growth rate of mycobacteria. Possible reasons and significance for their pathogenicity]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 48, 1390-1399.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271, 1247-1254.

Li, G.W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538-541.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10, 584-590.

Li, X., Sun, Q., Jiang, C., Yang, K., Hung, L.W., Zhang, J., and Sacchettini, J.C. (2015). Structure of Ribosomal Silencing Factor Bound to Mycobacterium tuberculosis Ribosome. *Structure* 23, 1858-1865.

Liang, B., Li, Z., Jenni, S., Rahmeh, A.A., Morin, B.M., Grant, T., Grigorieff, N., Harrison, S.C., and Whelan, S.P.J. (2015). Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell* 162, 314-327.

Lin, W., Mandal, S., Degen, D., Liu, Y., Ebright, Y.W., Li, S., Feng, Y., Zhang, Y., Mandal, S., Jiang, Y., *et al.* (2017). Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition. *Mol Cell* 66, 169-179 e168.

Liu, Q., and Fredrick, K. (2016). Intersubunit Bridges of the Bacterial Ribosome. *J Mol Biol* 428, 2146-2164.

Locher, K.P., Lee, A.T., and Rees, D.C. (2002). The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* 296, 1091-1098.

Lu, P., Bai, X.C., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H.W., *et al.* (2014). Three-dimensional structure of human gamma-secretase. *Nature* 512, 166-170.

Luca, S., and Mihaescu, T. (2013). History of BCG Vaccine. *Maedica (Buchar)* 8, 53-58.

Lyumkis, D., Brilot, A.F., Theobald, D.L., and Grigorieff, N. (2013). Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol* 183, 377-388.

Maiuri, P., Knezevich, A., De Marco, A., Mazza, D., Kula, A., McNally, J.G., and Marcello, A. (2011). Fast transcription rates of RNA polymerase II in human cells. *EMBO Rep* 12, 1280-1285.

Manganelli, R., Provvedi, R., Rodrigue, S., Beaucher, J., Gaudreau, L., and Smith, I. (2004). Sigma factors and global gene regulation in *Mycobacterium tuberculosis*. *J Bacteriol* 186, 895-902.

Mastrorade, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* 152, 36-51.

McGrath, J.W., Chin, J.P., and Quinn, J.P. (2013). Organophosphonates revealed: new insights into the microbial metabolism of ancient molecules. *Nat Rev Microbiol* 11, 412-419.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* 19, 560-567.

Metcalf, W.W., and Wanner, B.L. (1993a). Evidence for a fourteen-gene, *phnC* to *phnP* locus for phosphonate metabolism in *Escherichia coli*. *Gene* 129, 27-32.

Metcalf, W.W., and Wanner, B.L. (1993b). Mutational analysis of an *Escherichia coli* fourteen-gene operon for phosphonate degradation, using Tnp_{phoA}' elements. *J Bacteriol* 175, 3430-3442.

Meyer, M.M. (2017). The role of mRNA structure in bacterial translational regulation. *Wiley Interdiscip Rev RNA* 8.

Miller, O.L., Jr., Hamkalo, B.A., and Thomas, C.A., Jr. (1970). Visualization of bacterial genes in action. *Science* 169, 392-395.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* 98, 892-897.

Mirzakhanyan, Y., and Gershon, P.D. (2017). Multisubunit DNA-Dependent RNA Polymerases from Vaccinia Virus and Other Nucleocytoplasmic Large-DNA Viruses: Impressions from the Age of Structure. *Microbiol Mol Biol Rev* 81.

Mohan, A., Padiadpu, J., Baloni, P., and Chandra, N. (2015). Complete Genome Sequences of a *Mycobacterium smegmatis* Laboratory Strain (MC2 155) and Isoniazid-Resistant (4XR1/R2) Mutant Strains. *Genome Announc* 3.

Mohan, S., and Noller, H.F. (2017). Recurring RNA structural motifs underlie the mechanics of L1 stalk movement. *Nat Commun* 8, 14285.

Molodtsov, V., Nawarathne, I.N., Scharf, N.T., Kirchhoff, P.D., Showalter, H.D., Garcia, G.A., and Murakami, K.S. (2013). X-ray crystal structures of the *Escherichia coli* RNA polymerase in complex with benzoxazinorifamycins. *J Med Chem* 56, 4758-4763.

Montali, R.J., Mikota, S.K., and Cheng, L.I. (2001). *Mycobacterium tuberculosis* in zoo and wildlife species. *Rev Sci Tech* 20, 291-303.

Moore, P.B. (1999). Structural motifs in RNA. *Annu Rev Biochem* 68, 287-300.

Moore, P.B. (2012). How should we think about the ribosome? *Annu Rev Biophys* 41, 1-19.

Moore, P.B., and Steitz, T.A. (2011). The roles of RNA in the synthesis of protein. *Cold Spring Harb Perspect Biol* 3, a003780.

Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., Sarafianos, S., Tuske, S., Patel, J., Jansen, R., *et al.* (2008). The RNA polymerase "switch region" is a target for inhibitors. *Cell* 135, 295-307.

Murakami, K.S. (2013). X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J Biol Chem* 288, 9126-9134.

Murakami, K.S. (2015). Structural biology of bacterial RNA polymerase. *Biomolecules* 5, 848-864.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296, 1285-1290.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296, 1280-1284.

Myasnikov, A.G., Afonina, Z.A., Menetret, J.F., Shirokov, V.A., Spirin, A.S., and Klaholz, B.P. (2014). The molecular structure of the left-handed supra-molecular helix of eukaryotic polyribosomes. *Nat Commun* 5, 5294.

Naganathan, A., Wood, M.P., and Moore, S.D. (2015). The large ribosomal subunit protein L9 enables the growth of EF-P deficient cells and enhances small subunit maturation. *PLoS One* 10, e0120060.

Nissen, P., Kjeldgaard, M., Thirup, S., Clark, B.F., and Nyborg, J. (1996). The ternary complex of aminoacylated tRNA and EF-Tu-GTP. Recognition of a bond and a fold. *Biochimie* 78, 921-933.

Noens, E.E., Williams, C., Anandhakrishnan, M., Poulsen, C., Ehebauer, M.T., and Wilmanns, M. (2011). Improved mycobacterial protein production using a *Mycobacterium smegmatis* groEL1DeltaC expression strain. *BMC Biotechnol* 11, 27.

Noll, H. (2008). The discovery of polyribosomes. *Bioessays* 30, 1220-1234.

Noll, M., Hapke, B., Schreier, M.H., and Noll, H. (1973). Structural dynamics of bacterial ribosomes. I. Characterization of vacant couples and their relation to complexed ribosomes. *J Mol Biol* 75, 281-294.

Nooren, I.M., and Thornton, J.M. (2003). Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol* 325, 991-1018.

Ortiz, J.O., Brandt, F., Matias, V.R., Sennels, L., Rappsilber, J., Scheres, S.H., Eibauer, M., Hartl, F.U., and Baumeister, W. (2010). Structure of hibernating ribosomes studied by cryoelectron tomography in vitro and in situ. *J Cell Biol* 190, 613-621.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276, 307-326.

Pacheu-Grau, D., Gomez-Duran, A., Iglesias, E., Lopez-Gallardo, E., Montoya, J., and Ruiz-Pesini, E. (2013). Mitochondrial antibiograms in personalized medicine. *Hum Mol Genet* 22, 1132-1139.

Paget, M.S. (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* 5, 1245-1265.

Paget, M.S., and Helmann, J.D. (2003). The sigma70 family of sigma factors. *Genome Biol* 4, 203.

Petrov, A.S., Bernier, C.R., Hsiao, C., Norris, A.M., Kovacs, N.A., Waterbury, C.C., Stepanov, V.G., Harvey, S.C., Fox, G.E., Wartell, R.M., *et al.* (2014). Evolution of the ribosome at atomic resolution. *Proc Natl Acad Sci U S A* 111, 10251-10256.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W., and Gossard, D.C. (2010). Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J Struct Biol* 170, 427-438.

Poehlsgaard, J., and Douthwaite, S. (2005). The bacterial ribosome as a target for antibiotics. *Nat Rev Microbiol* 3, 870-881.

Polikanov, Y.S., Blaha, G.M., and Steitz, T.A. (2012). How hibernation factors RMF, HPF, and YfiA turn off protein synthesis. *Science* 336, 915-918.

Pribnow, D. (1975). Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A* 72, 784-788.

Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504-508.

Qin, D., and Fredrick, K. (2013). Analysis of polysomes from bacteria. *Methods Enzymol* 530, 159-172.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557-572.

- Ramesh, M., and Woolford, J.L., Jr. (2016). Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. *RNA* 22, 1153-1162.
- Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem* 85, 319-347.
- Ream, T.S., Haag, J.R., and Pikaard, C.S. (2014). Plant multisubunit RNA polymerases IV and V. In *Nucleic Acid Polymerases* (Springer), pp. 289-308.
- Reddy, T.B., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H., *et al.* (2009). TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res* 37, D499-508.
- Ren, Z., Ranganathan, S., Zinnel, N.F., Russell, W.K., Russell, D.H., and Raushel, F.M. (2015). Subunit Interactions within the Carbon-Phosphorus Lyase Complex from *Escherichia coli*. *Biochemistry* 54, 3400-3411.
- Reyrat, J.M., and Kahn, D. (2001). *Mycobacterium smegmatis*: an absurd model for tuberculosis? *Trends Microbiol* 9, 472-474.
- Robb, N.C., Cordes, T., Hwang, L.C., Gryte, K., Duchi, D., Craggs, T.D., Santoso, Y., Weiss, S., Ebright, R.H., and Kapanidis, A.N. (2013). The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J Mol Biol* 425, 875-885.
- Rodnina, M.V., Beringer, M., and Wintermeyer, W. (2007). How ribosomes make peptide bonds. *Trends in biochemical sciences* 32, 20-26.
- Rodrigue, S., Provvedi, R., Jacques, P.E., Gaudreau, L., and Manganelli, R. (2006). The sigma factors of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev* 30, 926-941.

Rodrigues, L.C., Mangtani, P., and Abubakar, I. (2011). How does the level of BCG vaccine protection against tuberculosis fall over time? *BMJ: British Medical Journal (Online)* 343.

Rohou, A., and Grigorieff, N. (2015). CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* 192, 216-221.

Rorbach, J., Gammage, P.A., and Minczuk, M. (2012). C7orf30 is necessary for biogenesis of the large subunit of the mitochondrial ribosome. *Nucleic Acids Res* 40, 4097-4109.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262, 1407-1413.

Rother, M., Rother, K., Puton, T., and Bujnicki, J.M. (2011). ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* 39, 4007-4022.

Roy, R.N., Lomakin, I.B., Gagnon, M.G., and Steitz, T.A. (2015). The mechanism of inhibition of protein synthesis by the proline-rich peptide oncocin. *Nat Struct Mol Biol* 22, 466-469.

Russell, D.G. (2007). Who puts the tubercle in tuberculosis? *Nat Rev Microbiol* 5, 39-47.

Ryu, Y.J. (2015). Diagnosis of pulmonary tuberculosis: recent advances and diagnostic algorithms. *Tuberc Respir Dis (Seoul)* 78, 64-71.

Sachdeva, P., Misra, R., Tyagi, A.K., and Singh, Y. (2010). The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *FEBS J* 277, 605-626.

Saecker, R.M., Record, M.T., Jr., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol* 412, 754-771.

Sala, A., Bordes, P., and Genevaux, P. (2014). Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins (Basel)* 6, 1002-1020.

Sander, P., Prammananan, T., and Bottger, E.C. (1996). Introducing mutations into a chromosomal rRNA gene using a genetically modified eubacterial host with a single rRNA operon. *Mol Microbiol* 22, 841-848.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180, 519-530.

Scheres, S.H. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* 3, e03665.

Scheres, S.H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9, 853-854.

Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., *et al.* (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* 102, 615-623.

Schmitt, L., Benabdelhak, H., Blight, M.A., Holland, I.B., and Stubbs, M.T. (2003). Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. *J Mol Biol* 330, 333-342.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827-834.

Schuwirth, B.S., Day, J.M., Hau, C.W., Janssen, G.R., Dahlberg, A.E., Cate, J.H., and Vila-Sanjurjo, A. (2006). Structural analysis of kasugamycin inhibition of translation. *Nat Struct Mol Biol* 13, 879-886.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31, 3381-3385.

Sczepanski, J.T., and Joyce, G.F. (2013). Binding of a structured D-RNA molecule by an L-RNA aptamer. *J Am Chem Soc* *135*, 13290-13293.

Seefeldt, A.C., Nguyen, F., Antunes, S., Perebaskine, N., Graf, M., Arenz, S., Inampudi, K.K., Douat, C., Guichard, G., Wilson, D.N., *et al.* (2015). The proline-rich antimicrobial peptide Onc112 inhibits translation by blocking and destabilizing the initiation complex. *Nat Struct Mol Biol* *22*, 470-475.

Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* *313*, 1935-1942.

Seweryn, P., Van, L.B., Kjeldgaard, M., Russo, C.J., Passmore, L.A., Hove-Jensen, B., Jochimsen, B., and Brodersen, D.E. (2015). Structural insights into the bacterial carbon-phosphorus lyase machinery. *Nature* *525*, 68-72.

Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *J Bacteriol* *189*, 8746-8749.

Shasmal, M., and Sengupta, J. (2012). Structural diversity in bacterial ribosomes: mycobacterial 70S ribosome structure reveals novel features. *PloS one* *7*, e31742.

Shell, S.S., Wang, J., Lapierre, P., Mir, M., Chase, M.R., Pyle, M.M., Gawande, R., Ahmad, R., Sarracino, D.A., Ioerger, T.R., *et al.* (2015). Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genet* *11*, e1005641.

Shine, J., and Dalgarno, L. (1973). Occurrence of heat-dissociable ribosomal RNA in insects: the presence of three polynucleotide chains in 26 S RNA from cultured *Aedes aegypti* cells. *Journal of molecular biology* *75*, 57-72.

Shukla, J., Gupta, R., Thakur, K.G., Gokhale, R., and Gopal, B. (2014). Structural basis for the redox sensitivity of the Mycobacterium tuberculosis SigK-RskA sigma-anti-sigma complex. *Acta Crystallogr D Biol Crystallogr* *70*, 1026-1036.

Sigworth, F.J. (1998). A maximum-likelihood approach to single-particle image refinement. *J Struct Biol* *122*, 328-339.

Sim, A.Y., Levitt, M., and Minary, P. (2012). Modeling and design by hierarchical natural moves. *Proc Natl Acad Sci U S A* *109*, 2890-2895.

Skordalakes, E., and Berger, J.M. (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* *114*, 135-146.

Smythe, M.A., and Rybak, M.J. (1989). Ofloxacin: A Review. *DICP* *23*, 839-846.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. *Mol Microbiol* *4*, 1911-1919.

Sohmen, D., Chiba, S., Shimokawa-Chiba, N., Innis, C.A., Berninghausen, O., Beckmann, R., Ito, K., and Wilson, D.N. (2015). Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling. *Nat Commun* *6*, 6941.

Sprink, T., Ramrath, D.J., Yamamoto, H., Yamamoto, K., Loerke, J., Ismer, J., Hildebrand, P.W., Scheerer, P., Burger, J., Mielke, T., *et al.* (2016). Structures of ribosome-bound initiation factor 2 reveal the mechanism of subunit association. *Sci Adv* *2*, e1501502.

Sripakdeevong, P., Beauchamp, K., and Das, R. (2012). Why Can't We Predict RNA Structure At Atomic Resolution? In *RNA 3D Structure Analysis and Prediction* (Springer), pp. 43-65.

Stallings, C.L., and Glickman, M.S. (2011). CarD: a new RNA polymerase modulator in mycobacteria. *Transcription* *2*, 15-18.

Stanley, R.E., Blaha, G., Grodzicki, R.L., Strickler, M.D., and Steitz, T.A. (2010). The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nat Struct Mol Biol* *17*, 289-293.

Steenken Jr, W. (1935). Lysis of tubercle bacilli in vitro. *Proceedings of the Society for Experimental Biology and Medicine* *33*, 253-255.

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* *90*, 6498-6502.

Stephan, J., Bender, J., Wolschendorf, F., Hoffmann, C., Roth, E., Mailander, C., Engelhardt, H., and Niederweis, M. (2005). The growth rate of *Mycobacterium smegmatis* depends on sufficient porin-mediated influx of nutrients. *Mol Microbiol* *58*, 714-730.

Swartz, J.R., Jewett, M.C., and Woodrow, K.A. (2004). Cell-free protein synthesis with prokaryotic combined transcription-translation. *Methods Mol Biol* *267*, 169-182.

Sweeney, R., Chen, L., and Yao, M.C. (1994). An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. *Mol Cell Biol* *14*, 4203-4215.

Szewczak, A.A., Moore, P.B., Chang, Y.L., and Wool, I.G. (1993). The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc Natl Acad Sci U S A* *90*, 9581-9585.

Tabib-Salazar, A., Liu, B., Doughty, P., Lewis, R.A., Ghosh, S., Parsy, M.L., Simpson, P.J., O'Dwyer, K., Matthews, S.J., and Paget, M.S. (2013). The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase. *Nucleic Acids Res* *41*, 5679-5691.

Tagami, S., Sekine, S., Minakhin, L., Esyunina, D., Akasaka, R., Shirouzu, M., Kulbachinskiy, A., Severinov, K., and Yokoyama, S. (2014). Structural basis for promoter specificity switching of RNA polymerase by a phage factor. *Genes Dev* *28*, 521-531.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 157, 38-46.

Taylor, K.A., and Glaeser, R.M. (1974). Electron diffraction of frozen, hydrated protein crystals. *Science* 186, 1036-1037.

Temiaikov, D., Zenkin, N., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., Kashkina, E., Savkina, M., Zorov, S., Nikiforov, V., Igarashi, N., *et al.* (2005). Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol Cell* 19, 655-666.

Ternan, N.G., Mc Grath, J.W., Mc Mullan, G., and Quinn, J.P. (1998). Organophosphonates: occurrence, synthesis and biodegradation by microorganisms. *World Journal of Microbiology and Biotechnology* 14, 635-647.

Tesh, V.L., and O'Brien, A.D. (1991). The pathogenic mechanisms of Shiga toxin and the Shiga-like toxins. *Mol Microbiol* 5, 1817-1822.

Thakur, K.G., Praveena, T., and Gopal, B. (2010). Structural and biochemical bases for the redox sensitivity of Mycobacterium tuberculosis RslA. *J Mol Biol* 397, 1199-1208.

Thomsen, N.D., and Berger, J.M. (2009). Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* 139, 523-534.

Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16, 673-683.

Trauner, A., Loughheed, K.E., Bennett, M.H., Hingley-Wilson, S.M., and Williams, H.D. (2012). The dormancy regulator DosR controls ribosome stability in hypoxic mycobacteria. *J Biol Chem* 287, 24053-24063.

Travers, A. (1993). The mechanism of RNA chain initiation. In *DNA-Protein Interactions* (Springer), pp. 87-108.

Travers, A.A., and Burgessrr (1969). Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222, 537-540.

Tuomanen, E., Cozens, R., Tosch, W., Zak, O., and Tomasz, A. (1986). The rate of killing of *Escherichia coli* by beta-lactam antibiotics is strictly proportional to the rate of bacterial growth. *J Gen Microbiol* 132, 1297-1304.

Tuske, S., Sarafianos, S.G., Wang, X., Hudson, B., Sineva, E., Mukhopadhyay, J., Birktoft, J.J., Leroy, O., Ismail, S., Clark, A.D., Jr., *et al.* (2005). Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* 122, 541-552.

Tyagi, J.S., and Sharma, D. (2002). *Mycobacterium smegmatis* and tuberculosis. *Trends Microbiol* 10, 68-69.

Ulyanov, N.B., Mujeeb, A., Du, Z., Tonelli, M., Parslow, T.G., and James, T.L. (2006). NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J Biol Chem* 281, 16168-16177.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45, 439-446.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448, 163-168.

Vilcheze, C., Baughn, A.D., Tufariello, J., Leung, L.W., Kuo, M., Basler, C.F., Alland, D., Sacchettini, J.C., Freundlich, J.S., and Jacobs, W.R., Jr. (2011). Novel inhibitors of InhA efficiently kill *Mycobacterium tuberculosis* under aerobic and anaerobic conditions. *Antimicrob Agents Chemother* 55, 3889-3898.

Vogel, Z., Zamir, A., and Elson, D. (1969). The possible involvement of peptidyl transferase in the termination step of protein biosynthesis. *Biochemistry* 8, 5161-5168.

Vynnycky, E., and Fine, P.E. (1997). The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 119, 183-201.

Wackett, L.P., Shames, S.L., Venditti, C.P., and Walsh, C.T. (1987). Bacterial carbon-phosphorus lyase: products, rates, and regulation of phosphonic and phosphinic acid metabolism. *J Bacteriol* 169, 710-717.

Wada, A., Yamazaki, Y., Fujita, N., and Ishihama, A. (1990). Structure and probable genetic location of a "ribosome modulation factor" associated with 100S ribosomes in stationary-phase *Escherichia coli* cells. *Proc Natl Acad Sci U S A* 87, 2657-2661.

Wanschers, B.F., Szklarczyk, R., Pajak, A., van den Brand, M.A., Gloerich, J., Rodenburg, R.J., Lightowers, R.N., Nijtmans, L.G., and Huynen, M.A. (2012). C7orf30 specifically associates with the large subunit of the mitochondrial ribosome and is involved in translation. *Nucleic Acids Res* 40, 4040-4051.

Ward, S.K., Heintz, J.A., Albrecht, R.M., and Talaat, A.M. (2012). Single-cell elemental analysis of bacteria: quantitative analysis of polyphosphates in *Mycobacterium tuberculosis*. *Front Cell Infect Microbiol* 2, 63.

Warner, J.R., and Knopf, P.M. (2002). The discovery of polyribosomes. *Trends Biochem Sci* 27, 376-380.

Warner, J.R., Rich, A., and Hall, C.E. (1962). Electron Microscope Studies of Ribosomal Clusters Synthesizing Hemoglobin. *Science* 138, 1399-1403.

Wayne, L.G., and Sohaskey, C.D. (2001). Nonreplicating persistence of *mycobacterium tuberculosis*. *Annu Rev Microbiol* 55, 139-163.

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9, 85-98.

WHO (2017). Global tuberculosis report 2017.

Wilson, D.N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol* 12, 35-48.

Wilson, D.N., and Doudna Cate, J.H. (2012). The structure and function of the eukaryotic ribosome. *Cold Spring Harb Perspect Biol* 4.

Wilson, D.N., Schluenzen, F., Harms, J.M., Starosta, A.L., Connell, S.R., and Fucini, P. (2008). The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning. *Proc Natl Acad Sci U S A* 105, 13339-13344.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327-339.

Winther, K.S., Brodersen, D.E., Brown, A.K., and Gerdes, K. (2013). VapC20 of *Mycobacterium tuberculosis* cleaves the sarcin-ricin loop of 23S rRNA. *Nat Commun* 4, 2796.

Yamagishi, M., Matsushima, H., Wada, A., Sakagami, M., Fujita, N., and Ishihama, A. (1993). Regulation of the *Escherichia coli* *rmf* gene encoding the ribosome modulation factor: growth phase- and growth rate-dependent control. *EMBO J* 12, 625-630.

Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell* 22, 5-13.

Yoneyama, H., and Katsumata, R. (2006). Antibiotic resistance in bacteria and its future for novel antibiotic development. *Biosci Biotechnol Biochem* 70, 1060-1075.

Yoshida, H., and Wada, A. (2014). The 100S ribosome: ribosomal hibernation induced by stress. *Wiley Interdiscip Rev RNA* 5, 723-732.

Yu, X., Jiang, J., Sun, J., and Zhou, Z.H. (2015). A putative ATPase mediates RNA transcription and capping in a dsRNA virus. *Elife* 4, e07901.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883-896.

Zaman, S.B., Hussain, M.A., Nye, R., Mehta, V., Mamun, K.T., and Hossain, N. (2017). A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* 9, e1403.

Zhang, G., and Darst, S.A. (1998). Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain. *Science* 281, 262-266.

Zhang, J., Minary, P., and Levitt, M. (2012). Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. *Proc Natl Acad Sci U S A* 109, 9845-9850.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol* 193, 1-12.

Zheng, J., Rubin, E.J., Bifani, P., Mathys, V., Lim, V., Au, M., Jang, J., Nam, J., Dick, T., Walker, J.R., *et al.* (2013). para-Aminosalicylic acid is a prodrug targeting dihydrofolate reductase in Mycobacterium tuberculosis. *J Biol Chem* 288, 23447-23456.

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.

Zhou, J., Lancaster, L., Donohue, J.P., and Noller, H.F. (2013). Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation. *Science* 340, 1236086.

Zumla, A., Nahid, P., and Cole, S.T. (2013). Advances in the development of new tuberculosis drugs and treatment regimens. *Nat Rev Drug Discov* 12, 388-404.

APPENDIX I

RIBOSOMAL SILENCING FACTOR RSFS

A1.1 Introduction

A wealth of structural and biochemical studies on the ribosome published over the last two decades has helped to create a deeper understanding of the mechanism of translation (Ban et al., 2000; Gao et al., 2009; Harms et al., 2001; Korostelev et al., 2006; Moore, 2012; Schuwirth et al., 2005; Selmer et al., 2006; Yusupov et al., 2001). Ribosome structures have provided molecular details for the machinery of translation as well as information on how drugs bind to the ribosome and interfere with protein synthesis (Wilson, 2014). In bacteria, the functional ribosome consists of two subunits, a 50S and a 30S. The 50S subunit contains 23S rRNA, 5S rRNA and over 30 proteins (35 in *M. tuberculosis*), while the 30S subunit contains a 16S rRNA and over 20 proteins (22 in *M. tuberculosis*). Given their fundamental importance to cell viability, it is not surprising that more than half of all clinically prescribed antibiotics target the ribosome (Wilson, 2014). However, the detailed mechanisms underlying the regulation of translation are, in some cases, poorly understood. This is especially true for the pathogenic bacterium *M. tuberculosis*, and it is most likely due to its complex growth characteristics,

Most of this appendix is reprinted with permission from “Structure of Ribosomal Silencing Factor Bound to Mycobacterium tuberculosis Ribosome.” by Xiaojun Li, Qingan Sun, Cai Jiang, Kailu Yang, Li-Wei Hung, Junjie Zhang, James C. Sacchettini, *Structure* 2015 23(10), 1858-1865, Copyright 2015 by Elsevier Inc.

Author contributions: XL and CJ purified *M. tuberculosis* ribosome and crystallized *M. tuberculosis* RsfS. CJ developed the function assay based on the GFP signal. LH collected and analyzed the X-ray diffraction data and QS solved the crystal structure of RsfS. JZ and KY solved the cryo-EM structures and interpreted the binding between the RsfS and the *M. tuberculosis* 50S ribosome. JS and JZ are the principal investigators.

including the ability to enter a non-replicating state in which protein synthesis is presumed to be greatly diminished.

Tuberculosis (TB) remains a major global health problem and is linked to high morbidity and mortality worldwide (Harries and Dye, 2006; Koch et al., 2014; Russell, 2007). The World Health Organization estimates that one-third of the global population is infected with latent *M. tuberculosis* while more than eight million people develop active TB. Roughly one million people die from the disease annually (<http://www.who.int/mediacentre/factsheets/fs104/en/>) and co-infection with HIV markedly increases mortality associated with TB (Kwan and Ernst, 2011).

The success of *M. tuberculosis* as a pathogen is largely attributed to its ability to persist in host tissues (Flynn and Chan, 2001; Wayne and Sohaskey, 2001). In chronic or persistent infections, *M. tuberculosis* shows reduced growth, which is thought to be related to the host's immune system or to the lack of susceptibility to antibiotics (Harries and Dye, 2006; Wayne and Sohaskey, 2001). Persistent *M. tuberculosis* are thought to exist a non-replicating state, equivalent to dormancy, in which many metabolic processes, including protein synthesis, are assumed to be greatly reduced in order to conserve cellular resources (Kumar et al., 2012; Trauner et al., 2012). As a consequence, drug treatment must be dramatically extended (Gomez and McKinney, 2004). *M. tuberculosis*'s emergence from dormancy is not understood, but it can lead to the reactivation of the disease.

A recently characterized protein from *Escherichia coli*, named ribosomal silencing factor during starvation or stationary phase (RsfS), has been shown to slow down or block translation entirely (Hauser et al., 2012). It appears to play an important role in the maintenance of sustainable energy levels during nutritional shortages. Disruption of the *rsfS* gene slows the adaptation from rich to poor media, and impairs the viability of the cells during the stationary

phase. Previous mutagenesis studies in *E. coli* suggest that RsfS binds to the 50S large ribosomal subunit (Jiang et al., 2007), preventing the normal association of the 50S and 30S into a functional 70S complex (Hauser et al., 2012). In order to improve our understanding of the molecular details of the inhibition of translation by RsfS, biochemical and structural studies have been completed on the 50S ribosome from *M. tuberculosis*. The crystal structure of *M. tuberculosis* RsfS and the cryo-electron microscopy (cryo-EM) structure of the *M. tuberculosis* 50S, with and without *M. tuberculosis* RsfS bound, provide a detailed understanding of the molecular basis of RsfS function.

A1.2 Materials and Methods

A1.2.1 Expression and purification of *M. tuberculosis* RsfS and *E. coli* RsfS

M. tuberculosis *rsfS* wild-type and site-directed mutant (Y102A, E74A) were cloned into the NdeI and HindIII sites of p1602-dest (Life Technologies) vector. The vectors encoded a C-terminal His₆-tagged RsfS and that was transformed into *M. smegmatis* cells MC²4517. Colonies containing the plasmid were selected by hygromycin. For large-scale production of recombinant proteins, cells were grown in 6 l of 7H9 broth to a cell density (OD₆₀₀) of 0.8, and then induced by 0.2% acetamide at 37°C for 8 hr. The *M. tuberculosis* RsfS protein was purified according to a modified protocol (Noens et al., 2011). The culture was centrifuged, separated from the pellet, and the cell pellet was lysed using a French press in lysis buffer composed of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, and 10 mM imidazole. The suspension was centrifuged at 30,000 × *g* for 45 min at 4°C and its supernatant was applied to a nickel chromatography column (GE Healthcare). The column was washed extensively and the overexpressed RsfS protein was eluted using elution buffer (lysis buffer with 250 mM imidazole [pH 7.5]). Next, size-exclusion

chromatography was performed using a Superdex 75 column (GE Healthcare) in a buffer containing 20 mM Tris-Cl (pH 7.5) and 150 mM NaCl and the protein profile was compared with protein molecular size standards. The Y102A RsfS mutant was only used for crystallization. Both wild-type RsfS and mutant E74A were used in the ribosome function assay and the nickel pull-down assay, and only wild-type RsfS was used for cryo-EM structure determination.

E. coli *rsfS* wild-type was cloned into the NdeI and XhoI sites of pET22b (Novagen) vector. The final construct encoded a C-terminal His₆-tagged RsfS and was transformed into *E. coli* strain BL21 (DE3). The colonies containing plasmid were selected by ampicillin. For large-scale production, cells were grown in 6 l of LB broth up to a cell density (OD₆₀₀) of 0.8 and then induced by 0.5 mM imidazole at 37°C for 3 hr. The *E. coli* RsfS protein was purified with the same method as *M. tuberculosis* RsfS.

A1.2.2 Purification of *M. tuberculosis* 70S and 50S

M. tuberculosis cells MC²7000 (Vilcheze et al., 2011) were grown in 7H9 medium supplemented with 10% oleic albumin dextrose catalase (BD), 0.5% glycerol, 0.05% Tween-80, and 50 µg/ml pantothenic acid at 37°C until an OD₆₀₀ of 1.0. The following procedures were performed at 4°C. Harvested cells were lysed in a bead beater (BioSpec) in lysis buffer (20 mM Tris-HCl [pH 7.5], 100 mM NH₄Cl, 10 mM MgCl₂, 0.5 mM EDTA, 6 mM 2-mercaptoethanol). *M. tuberculosis* ribosome 70S and 50S were purified according to modified protocols (Noll et al., 1973; Selmer et al., 2006). Cell lysate was clarified by centrifugation at 30,000 × *g* for 1 hr. The supernatant was pelleted in sucrose cushion buffer (20 mM HEPES [pH 7.5], 1.1 M sucrose, 10 mM MgCl₂, 0.5 M KCl, and 0.5 mM EDTA) at 40,000 rpm in a Beckman Type 45Ti rotor for 20 hr. The pellet was resuspended in the buffer containing 20 mM Tris-HCl (pH 7.5), 1.5 M

(NH₄)₂SO₄, 0.4 M KCl, and 10 mM MgCl₂. The suspension was then applied to a hydrophobic interaction column (Toyopearl Butyl-650S) and eluted with a reverse ionic strength gradient from 1.5 M to 0 M (NH₄)₂SO₄ in the buffer containing 20 mM Tris-HCl (pH 7.5), 0.4 M KCl, and 10 mM MgCl₂. The eluted ribosome peak was changed to re-association buffer (5 mM HEPES-NaOH [pH 7.5], 10 mM NH₄Cl, 50 mM KCl, 10 mM MgCl₂, and 6 mM 2-mercaptoethanol) or dissociation buffer (20 mM Tris-HCl [pH 7.5], 2 mM MgCl₂, 150 mM NH₄Cl, 50 mM KCl, and 6 mM 2-mercaptoethanol) and concentrated before loading on top of a 10%–40% linear sucrose gradient centrifuged in a Beckman SW28 rotor at 19,000 rpm for 19 hr. The 70S and 50S fractions were concentrated to about A₂₆₀ = 300 after removal of the sucrose.

A1.2.3 Ribosome functional assay

The assay used to measure in vitro ribosome activity relied on the production of GFP in an *M. tuberculosis*-based cell-free system. *M. tuberculosis* S100 cell-free extract was prepared from *M. tuberculosis* MC²7000. The supernatant was pelleted in the buffer containing 20 mM HEPES (pH 7.5), 1.1 M sucrose, 10 mM MgCl₂, 0.5 M KCl, and 0.5 mM EDTA to remove endogenous ribosome. The assay was carried out in a 96-well plate, which involved incubating the necessary substrates with ribosome-free cell extract in an incubator plate reader. The standard reaction mixture contained 2 mM each of the 20 amino acids, 33 mM phosphoenolpyruvate, 0.33 mM nicotinamide adenine dinucleotide, 0.26 mM coenzyme A, 2 µl of 0.85 µM purified *M. tuberculosis* ribosome, 200 ng of GFP mRNA, and 24 µl of S100 ribosome-free *M. tuberculosis* cell-free extract in certain buffer (Swartz et al., 2004). GFP mRNA was prepared from an in vitro transcription assay (Baugh et al., 2001). Either *M. tuberculosis* RsfS or *E. coli* RsfS was added to the reaction to final concentrations of 0.14 µM and 0.28 µM. The final concentrations of

streptomycin and chloramphenicol were 0.7 μM and 0.14 μM , respectively. The total volume of the assay was 100 μl and samples were incubated at 37°C for 40 hr.

A1.2.4 Crystallization and structure determination of *M. tuberculosis* RsfS

Nine different single-point mutants were made to hydrophobic amino acids that were predicted to be on the surface. Only one mutant protein of Y102A yielded diffraction-quality crystals. The RsfS mutant Y102A was concentrated to ~40 mg/ml and mixed with an equal volume of 0.1 M sodium cacodylate (pH 6.5), 0.2 M magnesium acetate, and 30% (+/-)-2-methyl-2,4-pentanediol. Crystals were produced by vapor diffusion in sitting-drop trays at 20°C and were directly harvested from the drop, flash frozen, and stored in liquid N₂. Diffraction data were collected to 2.1 Å at the Advanced Light Source synchrotron at Lawrence Berkeley National Laboratory and were processed by HKL2000 (Otwinowski and Minor, 1997) (**Table A1.1**). The structure was solved by molecular replacement with a truncated poly-Ala model (Ala7-Ala103) derived from a *B. halodurans* ortholog (PDB: 2O5A) as the search model in AutoMR of PHENIX (Adams et al., 2010). After initial refinement, the side chains were rebuilt in AutoBuild of PHENIX. Then iterative cycles of manual rebuilding in COOT (Emsley et al., 2010) and PHENIX refinement led to the final model. Simulated annealing was applied in the early stages of refinement.

Table A1. 1 Data Collection and Refinement Statistics of *M. tuberculosis* RsfS Crystal Structure

<i>Mtb</i> RsfS	
Data Collection	
Space group	P1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	50.98, 50.98, 64.55
α , β , γ (°)	110.27, 96.17, 110.59
Resolution (Å)	46.06–2.1 (2.14–2.1) ^a
<i>R</i> _{sym} or <i>R</i> _{merge}	0.086 (0.263)
<i>I</i> / <i>σ</i>	6.3 (3.1)
Completeness (%)	93.3 (82.7)
Redundancy	1.7 (1.6)
Refinement	
Resolution (Å)	46.06–2.1
No. of reflections	49,832
<i>R</i> _{work} / <i>R</i> _{free}	0.21/0.26
No. of atoms	3,922
Protein	3,522
Ligand/ion	22
Water	378
<i>B</i> factors	24.8
Protein	24.1
Ligand/ion	19.3
Water	31.6
Rmsd	
Bond lengths (Å)	0.005
Bond angles (°)	0.89

A1.2.5 Cryo-EM sample preparation

RsfS was added to purified 50S ribosome in a molar ratio of 15:1. This mixture was further diluted with 5 mM HEPES-Na (pH 7.5), 10 mM NH₄Cl, 50 mM KCl, and 10 mM MgCl₂ to final concentrations of 50S at 0.36 μM and RsfS at 5.4 μM. This sample was then applied onto a 200 mesh R2/2 Quantifoil grid (Quantifoil Micro Tools). The grid was previously glow discharged. After applying the sample, the grid was blotted and rapidly frozen in liquid ethane using a Vitrobot (FEI), and then stored in liquid nitrogen before imaging.

A1.2.6 Electron microscopy, image processing, map segmentation, and visualization

The grid was imaged on an FEI Tecnai F20 with a field emission gun operated at 200 kV (FEI). One hundred and sixty-five micrographs were recorded at an effective magnification of 81,081 \times , on a Gatan 4k \times 4k charge-coupled device camera (Gatan) with a final image pixel size of 1.85 Å.

Each micrograph was carefully examined for drift and astigmatism. One hundred and fifty-four micrographs with a defocus range of 1–2.5 μm were used for further processing. We carefully boxed the raw particles using EMAN2 (Tang et al., 2007) and manually removed (1) ice contamination, (2) particles that touched each other, and (3) particles that were on the carbon. This gave us 73,103 particles. We then used the unsupervised 3D classification in Relion 1.3 to classify the particles into ten classes and removed seven bad classes of particles, leaving a total of 42,138 “clean” particles (**Figure A1.1**, step 1). We then processed these selected particles to generate a 3D map in Relion 1.3. This map (named Initial Map 1) already shows relatively weak extra density for RsfS next to the L14 protein. Next, we manually erased the extra density of the RsfS in UCSF Chimera (Pettersen et al., 2004) and generated Initial Map 2. Both the 3D initial

maps were subjected to additional low-pass filtering with a Gaussian radius of 30 Å before applied for the supervised classification to separate the raw particles into two classes, 21,287 particles for RsfS-free and 20,851 particles for RsfS bound. This 30-Å radius filtering removed the high-resolution features and minimized the artifacts by manually removing the weak RsfS density, but still showed a noticeable difference in the RsfS binding site. Finally, we swapped the two reference maps to refine against the separated raw particles. The refined maps showed consistent results, with or without RsfS density, even with the contrary reference maps (**Figure A1.1**). This validated the correct separation of the raw particles and ruled out the possibility of model bias. The final resolutions were 9.3 Å for RsfS-free 50S and 9.1 Å for the RsfS-bound 50S (**Figure A1.2**). The local resolution of RsfS in the RsfS-bound state was ~9 Å according to the ResMap result. Map segmentation was done in UCSF Chimera with the reference from an *E. coli* 50S ribosome PDB structure (PDB: 2I2V; (Berk et al., 2006)). Figures of the maps and models were produced with UCSF Chimera.

A1.2.7 Molecular modeling, docking and flexible fitting of RsfS within the density map

The *M. tuberculosis* RsfS monomer from the RsfS crystal structure was first roughly docked onto the homology model of *M. tuberculosis* L14, which was built with the SWISS-MODEL server (Schwede et al., 2003) in the cryo-EM density map. To avoid the initial model bias in the refinement, the complex of L14 and RsfS was diversified into 1,000 initial models in the following two steps: (1) arbitrarily displacing RsfS away from L14 within a hemisphere of 10-Å radius; (2) randomly applying a rotation on the RsfS with the azimuthal angle between 0° and 360°, an altitude angle between 0° and 180°, and phi angle between 0° and 360°. All the 1,000 initial models were then refined independently with MOSAICS-EM (Zhang et al., 2012)

using the cryo-EM density map as a constraint. The best-fit model to the EM density map was further refined using the real-space refinement routine in PHENIX to optimize the protein stereochemistry (Adams et al., 2010).

A1.2.8 Accession codes

Coordinates and structure factors of the *M. tuberculosis* RsfS-50S complex have been deposited in the Protein Data Bank under accession code 4WCW. The cryo-EM maps of the *M. tuberculosis* 50S subunit with and without RsfS have been deposited in the Electron Microscopy Databank with accession code EMD-6177 and EMD-6178, respectively.

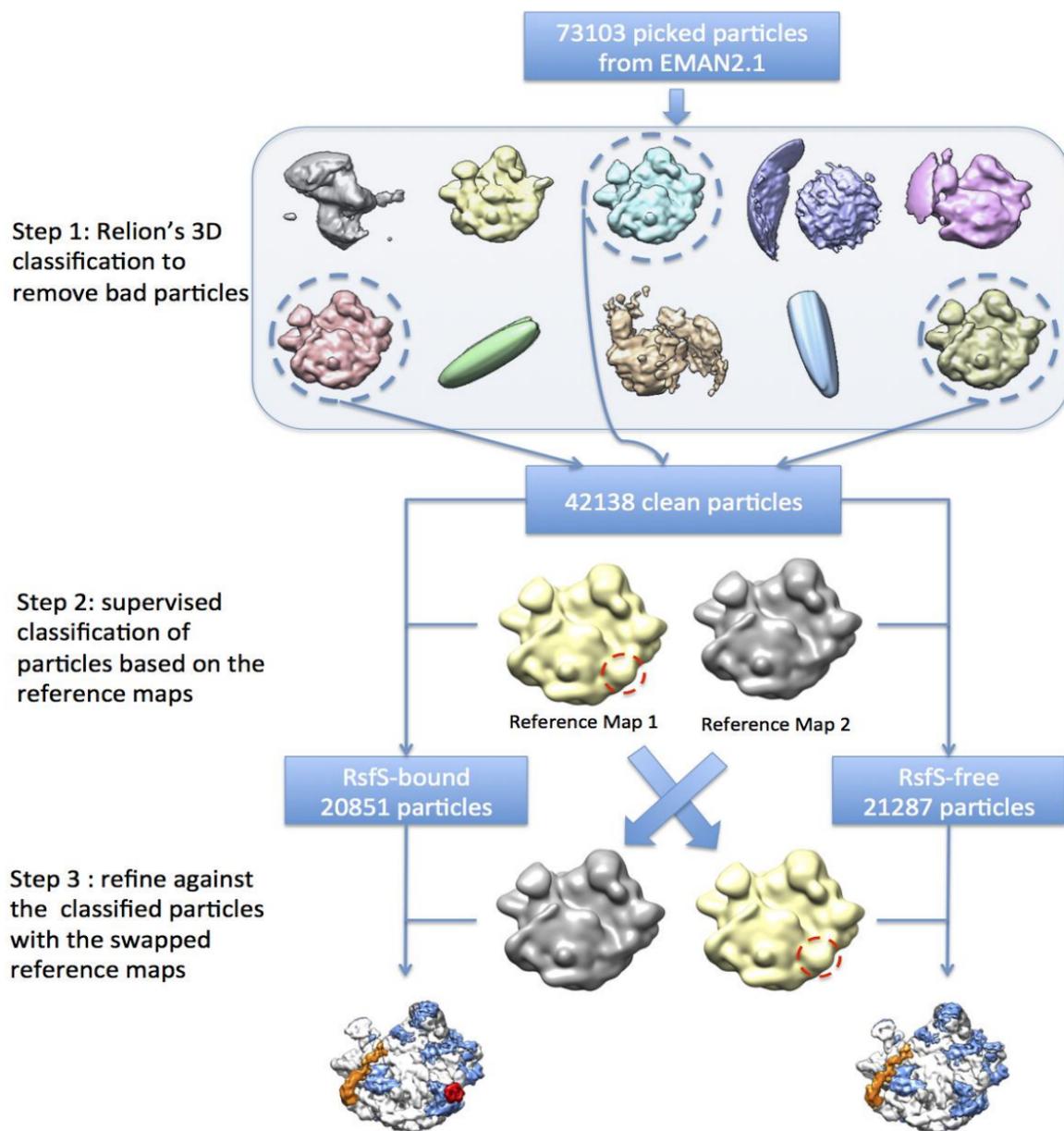


Figure A1. 1, related to Figure A1.3

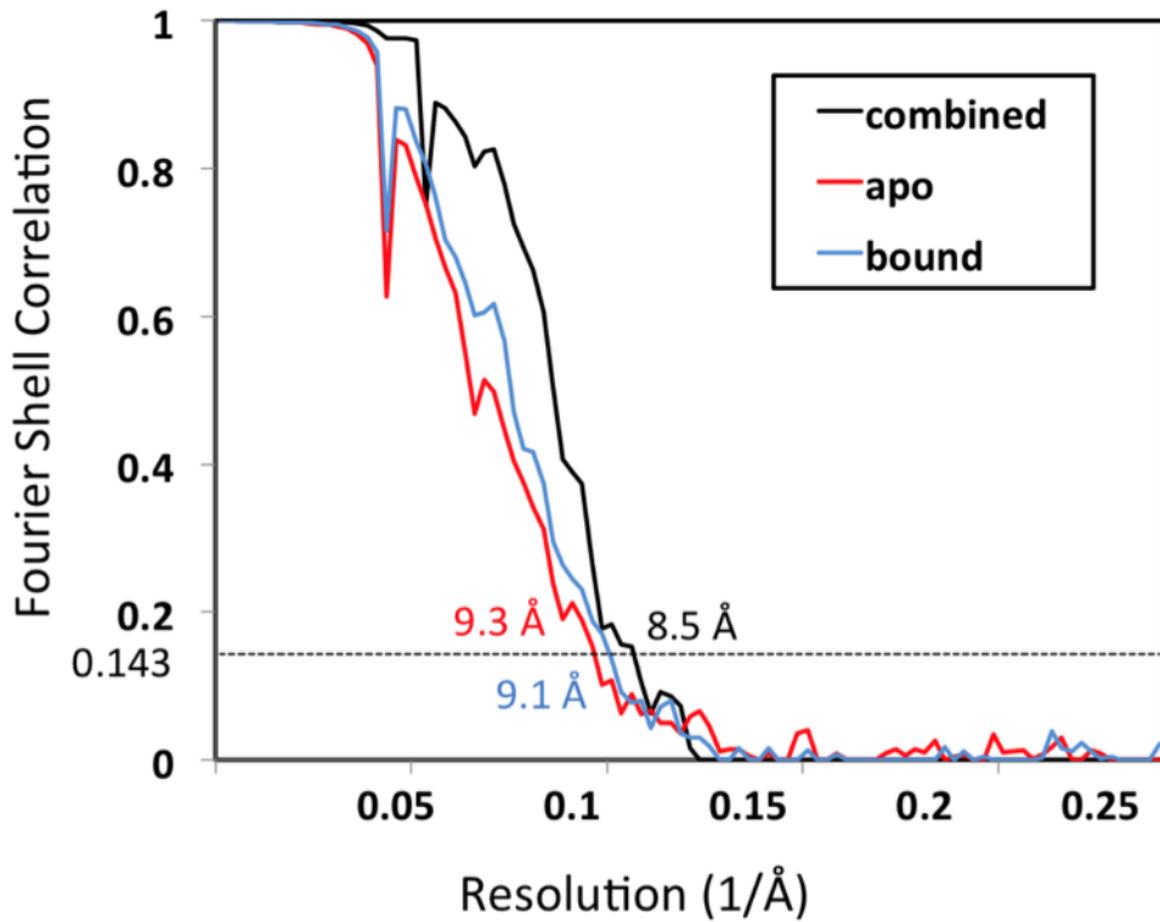


Figure A1. 2, related to Figure A1.3

Fourier shell correlations (FSC) for the three reconstructions of the *M. tuberculosis* 50S ribosome calculated using the gold-standard criterion. The resolutions are 8.5Å (when using all the particles), 9.1Å (RsfS-bound state) and 9.3Å (RsfS-free state).

A1.3 Results

A1.3.1 *M. tuberculosis* RsfS inhibits translation by preventing 70S ribosome association

We developed a translation assay that uses *M. tuberculosis* ribosome to translate green fluorescent protein (GFP) mRNA in an *M. tuberculosis*-derived cell-free extract. When purified recombinant *M. tuberculosis* RsfS was added to our cell-free assay, we observed a significant reduction in translation of GFP. Specifically, RsfS (0.28 μM) was able to block GFP synthesis by the ribosome (0.14 μM) by 83 percent, a level comparable to 0.7 μM streptomycin and 0.14 μM chloramphenicol (**Figure A1.3A**). Interestingly, when *M. tuberculosis* 50S and 30S ribosome were mixed with 15-fold molar excess of RsfS (4.2 μM final concentration) and applied to a sucrose-gradient (**Figure A1.3B**), only 10% of the large subunit was in the 70S form. However, in the absence of *M. tuberculosis* RsfS, more than 70% of 50S subunits were associated into the 70S ribosome. However, when the same concentration of RsfS was incubated with pre-formed 70S ribosome for 2 hrs at room temperature, no significant dissociation of 70S was observed. These results indicated that while RsfS was able to block the association of 50S and 30S subunits, it was unable to dissociate the 70S ribosome (Hauser et al., 2012).

M. tuberculosis RsfS shows selectivity for the 50S subunit, as demonstrated using a pull-down assay with His-tagged RsfS. In this assay, 6.7 μM (0.1 mg/ml) RsfS was mixed with purified 1.2 μM 70S (2.9 mg/ml), 50S (1.9 mg/ml) and 30S (1.0 mg/ml) fractions from a sucrose gradient. After incubation for 1 hr at 4°C, nickel affinity beads were used to pull-down RsfS. Only the 50S subunit associated with RsfS (**Figure A1.3C**).

A1.3.2 Crystal structure of *M. tuberculosis* RsfS

Crystals of full length *M. tuberculosis* RsfS could not be obtained; therefore we resorted to screening single point mutants of RsfS in order to produce diffraction-quality crystals. Nine distinct single-point mutants were made to hydrophobic amino acids that were predicted to be on the surface. Only one mutant protein of Y102A yielded diffraction-quality crystals. *M. tuberculosis* RsfS Y102A crystallized in the P1 space group with four *M. tuberculosis* RsfS molecules (referred to as A, B, C, D) in the crystal's asymmetric unit (ASU) (**Figure A1.4**). The structure was solved by molecular replacement using the *Bacillus halodurans* homolog (PDB: 2O5A) as the search model. The structure was refined with diffraction data to 2.1-Å resolution (**Table A1.1**). The *R* factor of the final model was 21.0% ($R_{\text{free}} = 26.0\%$) with good stereochemistry and 97.0% of the amino acids were in the preferred regions of the Ramachandran plot. The C-terminal 8-13 amino acids were disordered and showed either weak or no corresponding electron density.

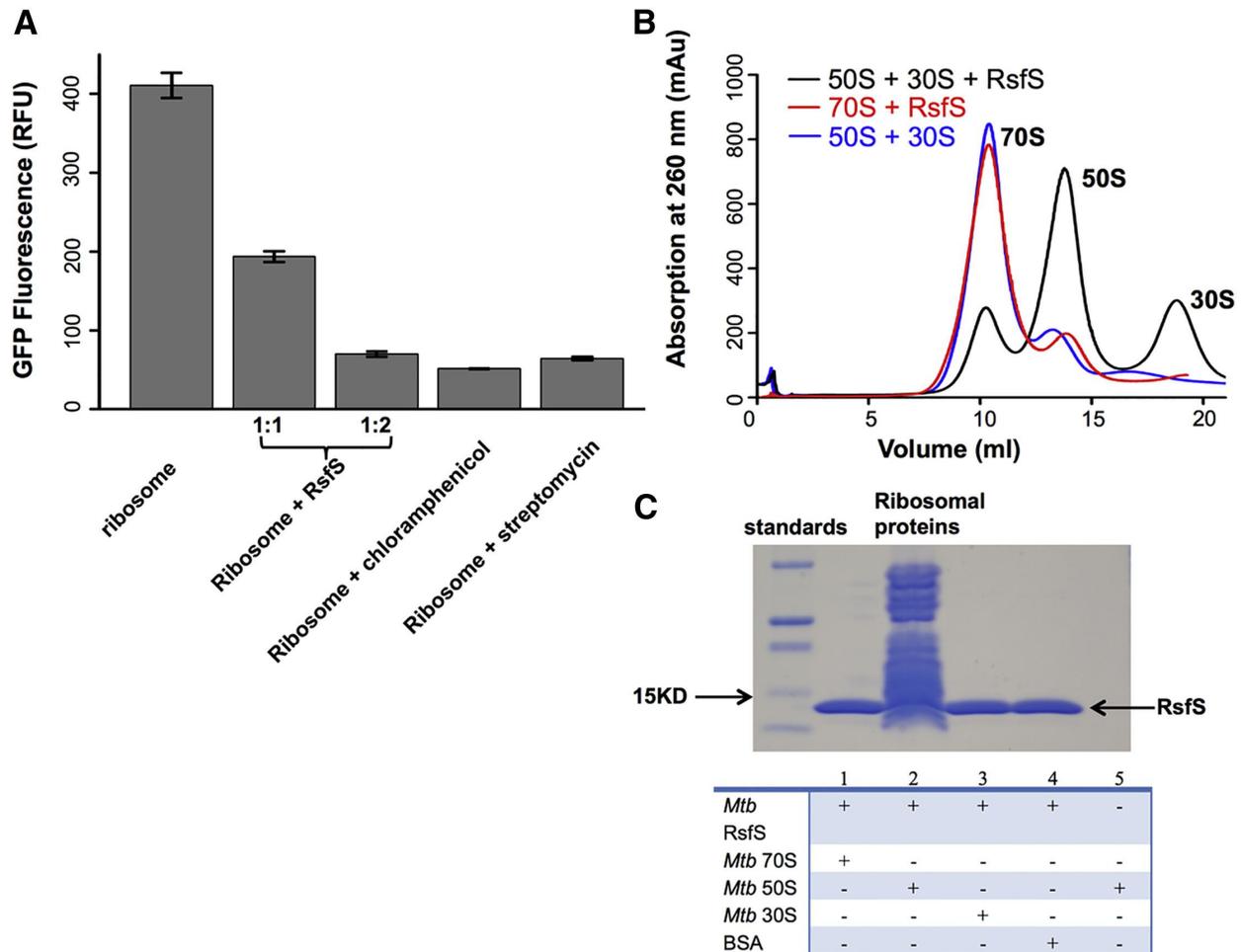


Figure A1. 3 RsfS Prevents Ribosomal Subunit Association Resulting in Inhibition of Protein Translation

(A) GFP translation by *M. tuberculosis* ribosome. The assay was performed with *M. tuberculosis* ribosome 70S at 0.14 μ M and *M. tuberculosis* RsfS at 0.14 μ M and 0.28 μ M. Chloramphenicol at 0.14 μ M and streptomycin at 0.7 μ M provided approximately the same inhibition. Data are represented as means \pm SEM.

(B) Sucrose gradient separation of 70S ribosome association.

(C) The results of the nickel pull-down assay of *M. tuberculosis* ribosome using His₆-tagged RsfS. RsfS was mixed with purified 70S, 50S, and 30S, respectively, and incubated for 1 hr before the nickel beads were added. The beads were washed and the eluted fractions were loaded on SDS polyacrylamide gel. Only ribosome 50S can be pulled down by RsfS.

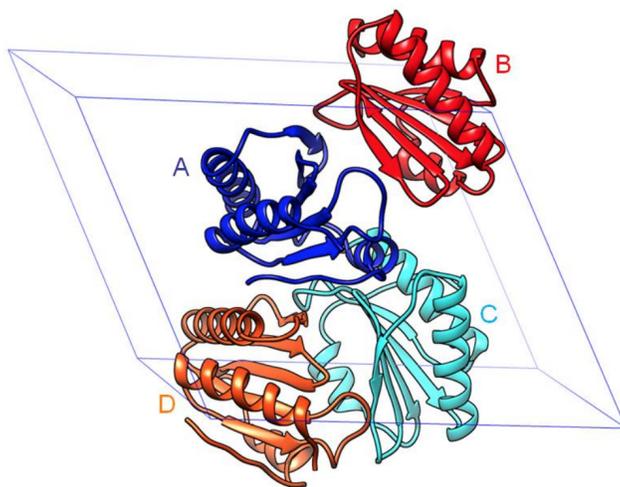


Figure A1. 4, related to Figure A1.5

Four *M. tuberculosis* RsfS molecules in one asymmetric unit of crystal lattice Chain A had 117 residues visible, Chain B had 113 residues, and Chains C & D had 118 residues.

RsfS adopts the $\alpha 1\text{-}\beta 1\text{-}\beta 2\text{-}\alpha 2\text{-}\beta 3\text{-}\beta 4\text{-}\beta 5\text{-}\alpha 3$ fold, where five β strands form one β sheet. The first two long α helices ($\alpha 1$ and $\alpha 2$) reside on one side of the β sheet, while the short $\alpha 3$ helix with a long C-terminal tail resides on the edge of β sheet close to $\beta 1$ (**Figure A1.5A**). This is a well-conserved domain referred to as DUF143 (Fung et al., 2013).

Alignment of the sequence of *M. tuberculosis* RsfS with other bacterial orthologs indicated relatively good sequence similarity for the portion of RsfS that is well ordered in the crystal (**Figure A1.6**). Superimposition of the three ortholog structures (PDB: 2O5A for *B. halodurans* RsfS, PDB: 3UPS for *Zymomonas mobilis* RsfS, and PDB: 2ID1 for *Chromobacterium violaceum* RsfS) onto *M. tuberculosis* RsfS showed that the overall structure is well conserved from the N terminus to the end of $\beta 5$ (**Figure A1.7**), corresponding to His95 in *M. tuberculosis* RsfS. The mutation Y102A did not appear to change the overall structure of *M. tuberculosis* RsfS when compared to the crystal structures of the other orthologs. The $C\alpha$ root-

mean squared difference (rmsd) values (for residues 6-95 in *M. tuberculosis* RsfS) are 1.6 Å, 1.9 Å, and 2.2 Å for 2ID1, 3UPS, and 2O5A, respectively. The lack of similarity in the structure of the C termini implies that they are not the critical determinant for inhibition of the ribosome. Although *E. coli* RsfS has an overall sequence of 25% with *M. tuberculosis* RsfS, the C-terminal region is much shorter (15 residues) than *M. tuberculosis* RsfS (32 residues) and shares only 16% sequence identity in this region compared with *M. tuberculosis* RsfS. However, full-length recombinant *E. coli* RsfS showed approximately the same level of inhibition as *M. tuberculosis* RsfS in the *M. tuberculosis* ribosome cell-free translation assay (80% inhibition of *E. coli* RsfS and 84% inhibition of *M. tuberculosis* RsfS on *M. tuberculosis* translation assay). This supported the notion that the C-terminal extensions of these orthologs were not critical components in the binding to ribosome.

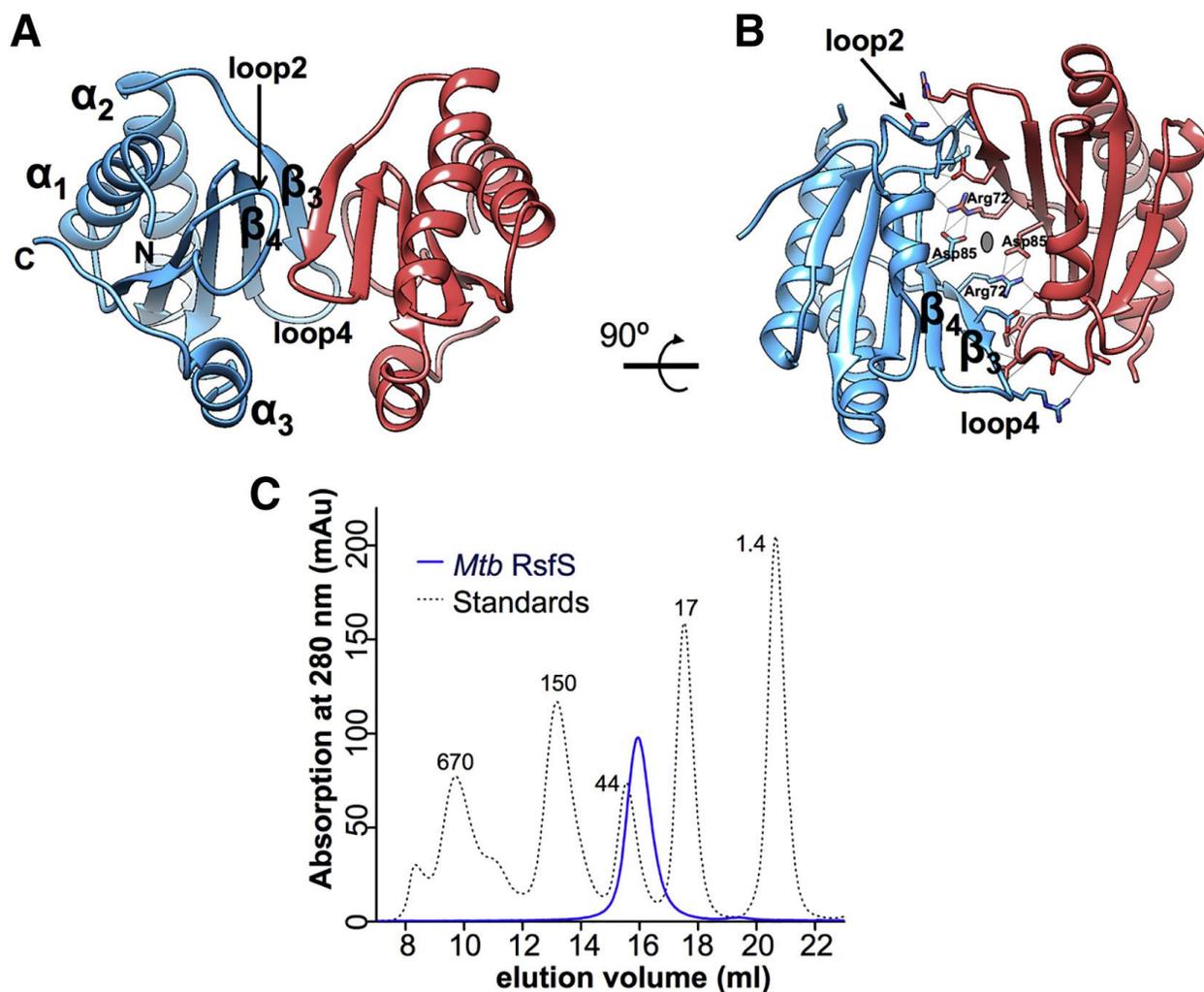


Figure A1.5 *M. tuberculosis* RsfS Structure

(A) *M. tuberculosis* RsfS dimer.

(B) RsfS dimer rotated 90° from the view in (A). The two RsfS molecules are related by ~180° rotation around the pseudo-dyad axis in the center of dimer. The residues involved in hydrogen bonding are highlighted as sticks and the hydrogen bonds are shown as black lines (see also Figures A1.4, A1.6, A1.7, and A1.10).

(C) Size-exclusion chromatography (Superdex 75) of *M. tuberculosis* RsfS showed a solution dimer in agreement with the crystal structure. The standard curve is shown as a dashed line (see also Figures A1.8 and A1.9).

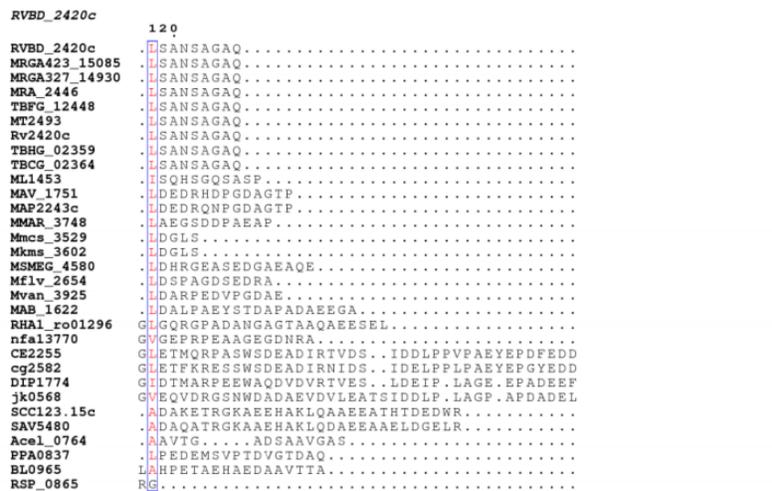
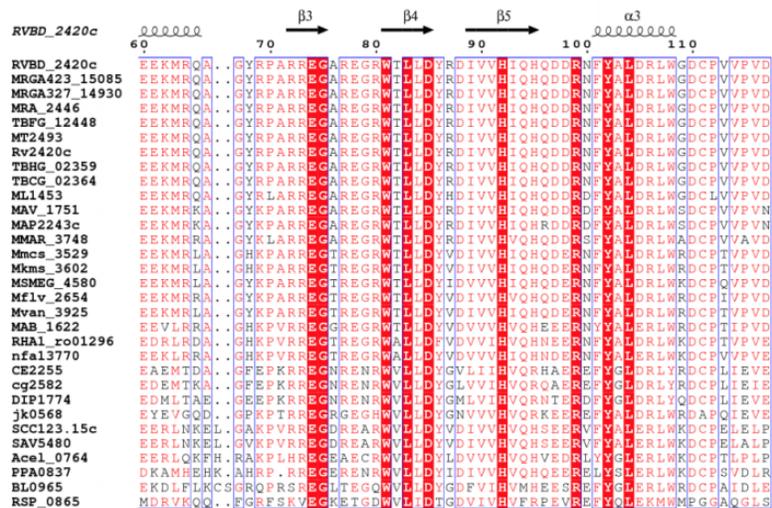
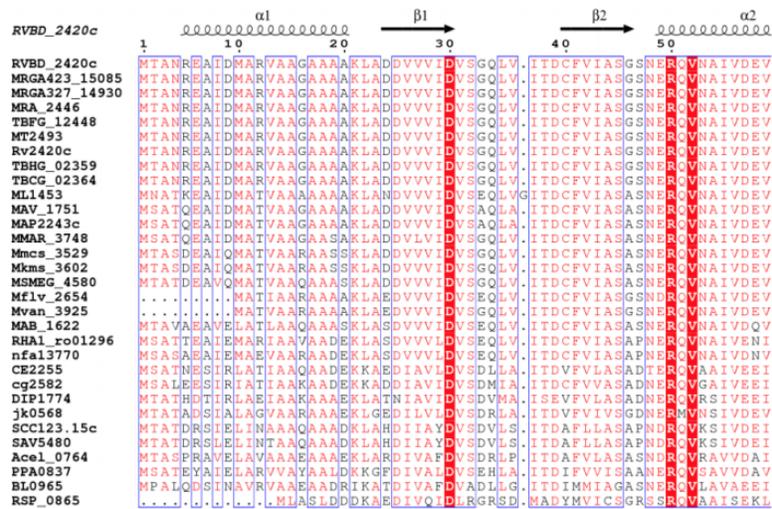


Figure A1. 6, related to Figure A1.5. Multiple sequence alignment of *M. tuberculosis* RsfS (*RVBD_2420c*) with iojap-like orthologs in TBDB database.

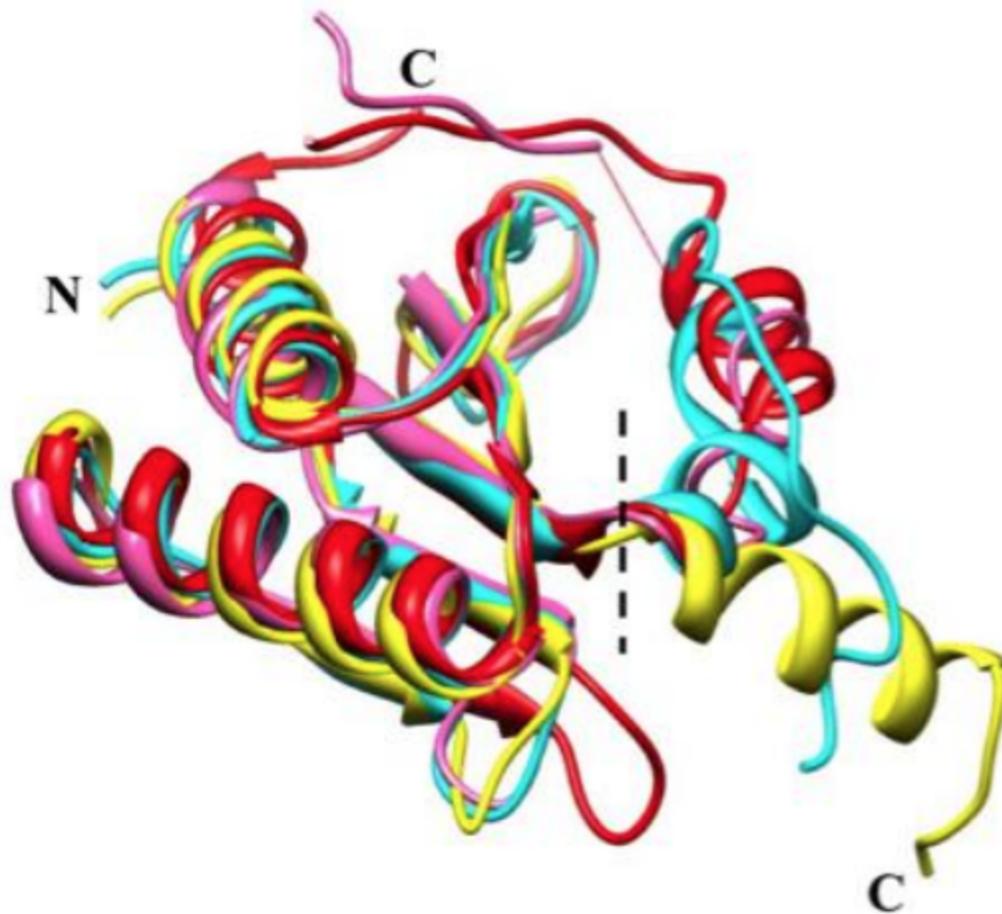


Figure A1. 7, related to Figure A1.5

Superimposition of the three ortholog structures (2ID1, 3UPS, and 2O5A) onto *M. tuberculosis* RsfS showed that the overall structure is well conserved from the N terminus to the end of $\beta 5$ (Red: *M. tuberculosis* RsfS; cyan: 2ID1; Yellow: 3UPS; Pink: 2O5A). The C terminal helices or unstructured regions appear to be flexible.

A1.3.3 *M. tuberculosis* RsfS forms a dimer in both crystal and solution

The structures of the four copies of the *M. tuberculosis* RsfS in the crystal's ASU are very similar (Thr2 to Pro112), except for the last six residues of C termini. The four subunits are packed into two nearly identical dimers (dimers AB and CD) each with quasi two-fold rotational symmetry for the subunits. The buried surface areas for the two dimers are 1,822 Å² and 1,706 Å², respectively (analyzed by PISA; (Krissinel and Henrick, 2007)), and this represents around 15% of the total surface of each dimer. A high percentage of buried surface area and the conservation of the packing of both dimers in the asymmetric unit indicated that the dimer observed in the crystal was equivalent to what is observed in solution (Krissinel and Henrick, 2007; Nooren and Thornton, 2003). RsfS also eluted from a size-exclusion chromatography (**Figure A1.5C**) as a dimer based on the calculated molecular weight of 29 kDa (an RsfS subunit is 15 kDa) (**Figure A1.8**). The RsfS dimer is also consistent with native gel electrophoresis and glutaraldehyde crosslinking experiments (**Figure A1.9**).

The dimer interface is predominately formed through interactions between the side-chains of amino acids contained within β3 (Arg72-Gly75) and β4 (Trp81- Asp85) from opposing subunits. In addition, loop2 (Val31-Asp39) and loop4 (Ala76-Arg80) from each subunit contribute residues that are located at the dimer interface (**Figure A1.5B**). Indeed, the two subunits appear to be stitched together through an extensive network of 22 intermolecular hydrogen bonds and electrostatic interactions (**Figure A1.10**). The two dimers are very well conserved with only small conformational differences in loop4 and the adjacent N-terminus of helix α2 (rmsd between A and B chain for the first 111 Cα is 0.88Å; that between C and D is 0.90Å). In contrast, the rmsd between chains A and C is only 0.17 Å, and that between chains B

and D is 0.37 Å. In the crystal lattice, contacts between subunits A and C of each dimer are very close to these loops and it is likely that these crystal contacts influence the observed differences.

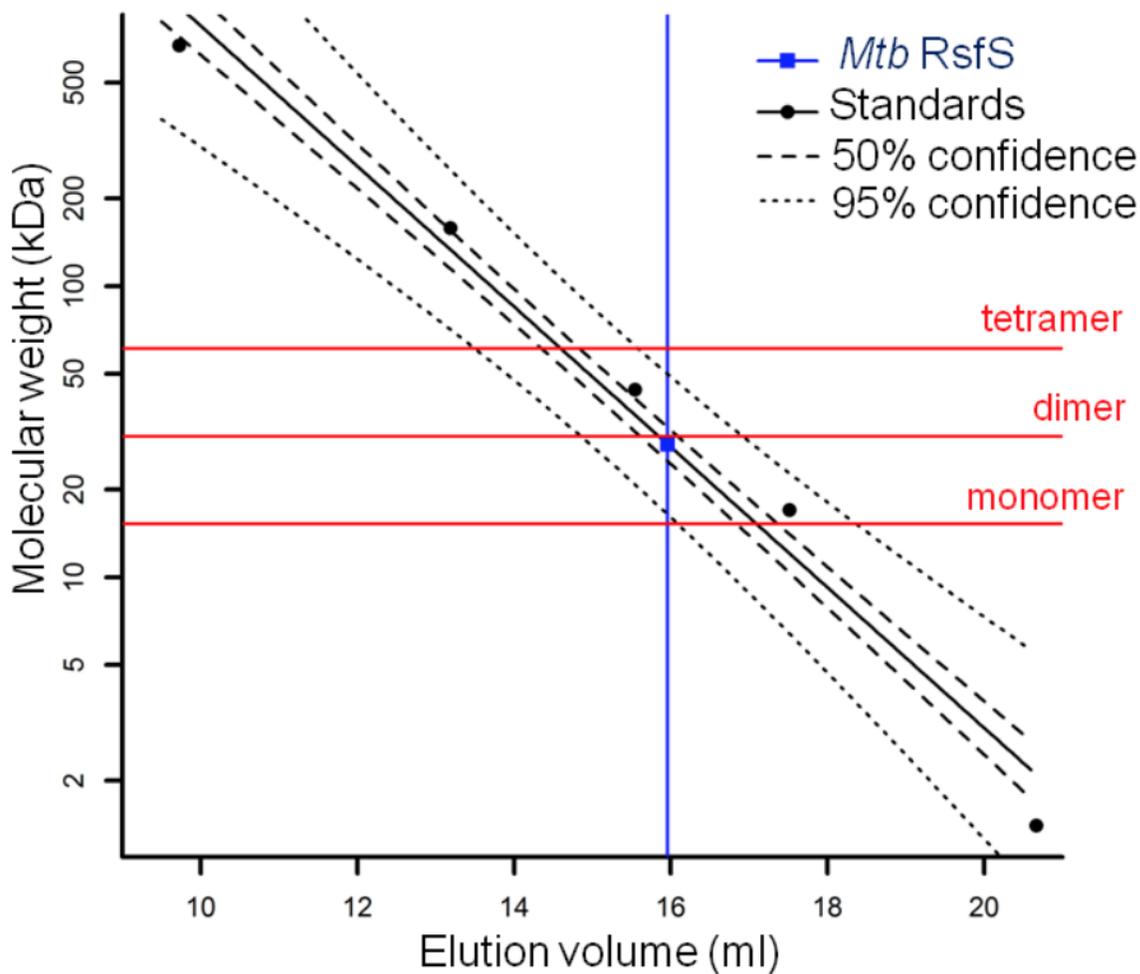


Figure A1. 8, related to Figure A1.5. Calculation of *M. tuberculosis* RsfS size based on size-exclusion chromatography

The black solid line is the linear regression curve through the standards

($\log_{10}MW = -0.24Ve + 5.3$, MW: molecular weight of standards, Ve : elution volume of standard peaks; $F=144.4$; $df=1, 3$; $p=0.00124$; $R^2=0.980$). 50% confidence and 95% confidence lines are shown in long-dash and short-dash, respectively. *M. tuberculosis* RsfS elution peak is at 15.96 ml, labeled in the blue line. The theoretical molecular weight of *M. tuberculosis* RsfS quaternary structures (monomer, dimer and tetramer) are labeled in red lines.

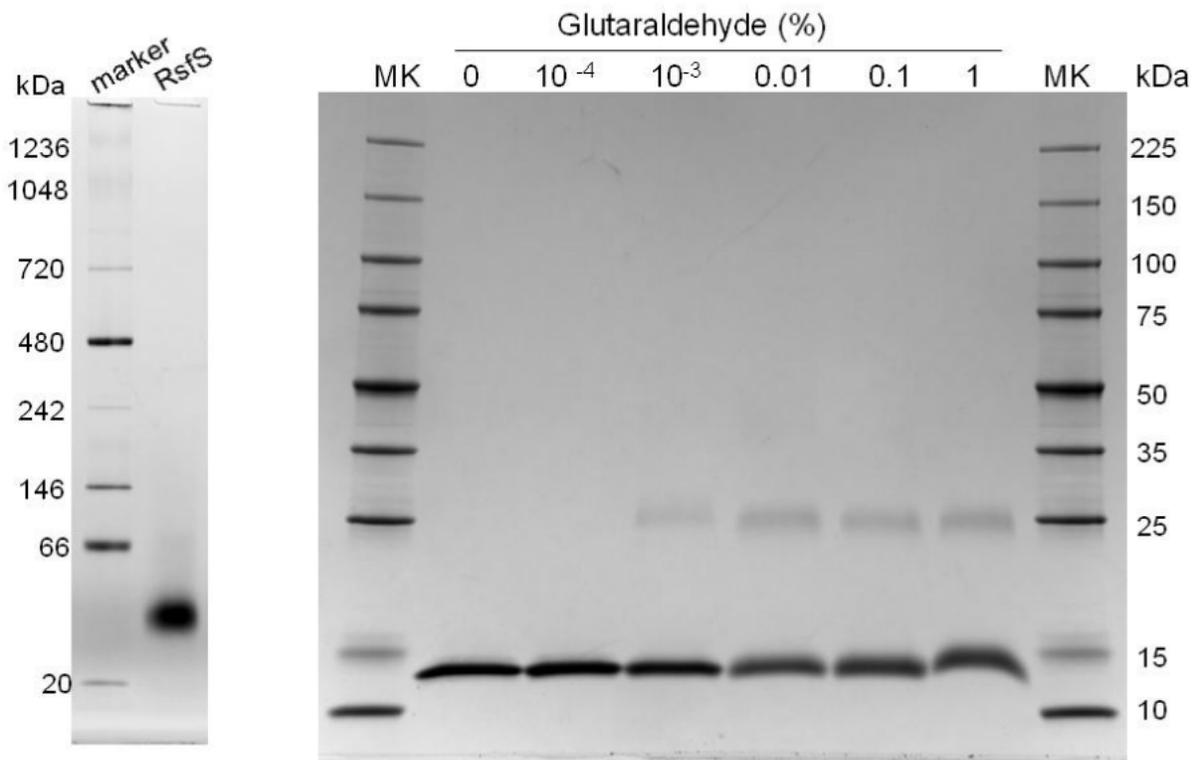


Figure A1. 9, related to Figure A1.5
Blue native gel (left) and Glutaraldehyde cross-linking PAGE (right) of *M. tuberculosis* RsfS

When the three homologs deposited in the PDB were compared to the *M. tuberculosis* RsfS crystal structures, they all showed significant differences in their quaternary structures. *Z. mobilis* RsfS (PDB: 3UPS) forms the clearest dimer of the three with a buried surface area of 2,840 Å², which represents 28% of the total surface of the dimer. *B. halodurans* RsfS (PDB: 2O5A) forms an apparently less stable dimer with a buried surface area of 1,070 Å², which represents 9% of the total surface of the dimer. The analysis of the crystal packing for *C. violaceum* RsfS (PDB: 2ID1) does not reveal any higher level oligomerization, suggesting it is a monomer in solution (Krissinel and Henrick, 2007). Comparison of *M. tuberculosis* RsfS dimer with *Z. mobilis* and *B. halodurans* RsfS dimers indicated that the intra-dimer interfaces occur at different regions for each of the three proteins. For the *Z. mobilis* RsfS dimer, the interactions are primarily between α helices ($\alpha 2$ and $\alpha 3$), loop 2 (connecting $\beta 1$ and $\beta 2$), and the two bends (connecting $\alpha 2$ and $\beta 3$ or $\beta 4$ and $\beta 5$). In the *B. halodurans* RsfS dimer, the interactions are between β strands ($\beta 3$ and $\beta 4$) and loops (loop 1 connects $\beta 1$ and $\beta 2$; loop 4 connects $\beta 5$ and $\alpha 3$). Although the dimers organize differently between the three proteins, the interfaces of the *Z. mobilis* and *B. halodurans* dimers are similarly dominated by hydrogen bonding and electrostatic interactions (**Figure A1.10**).

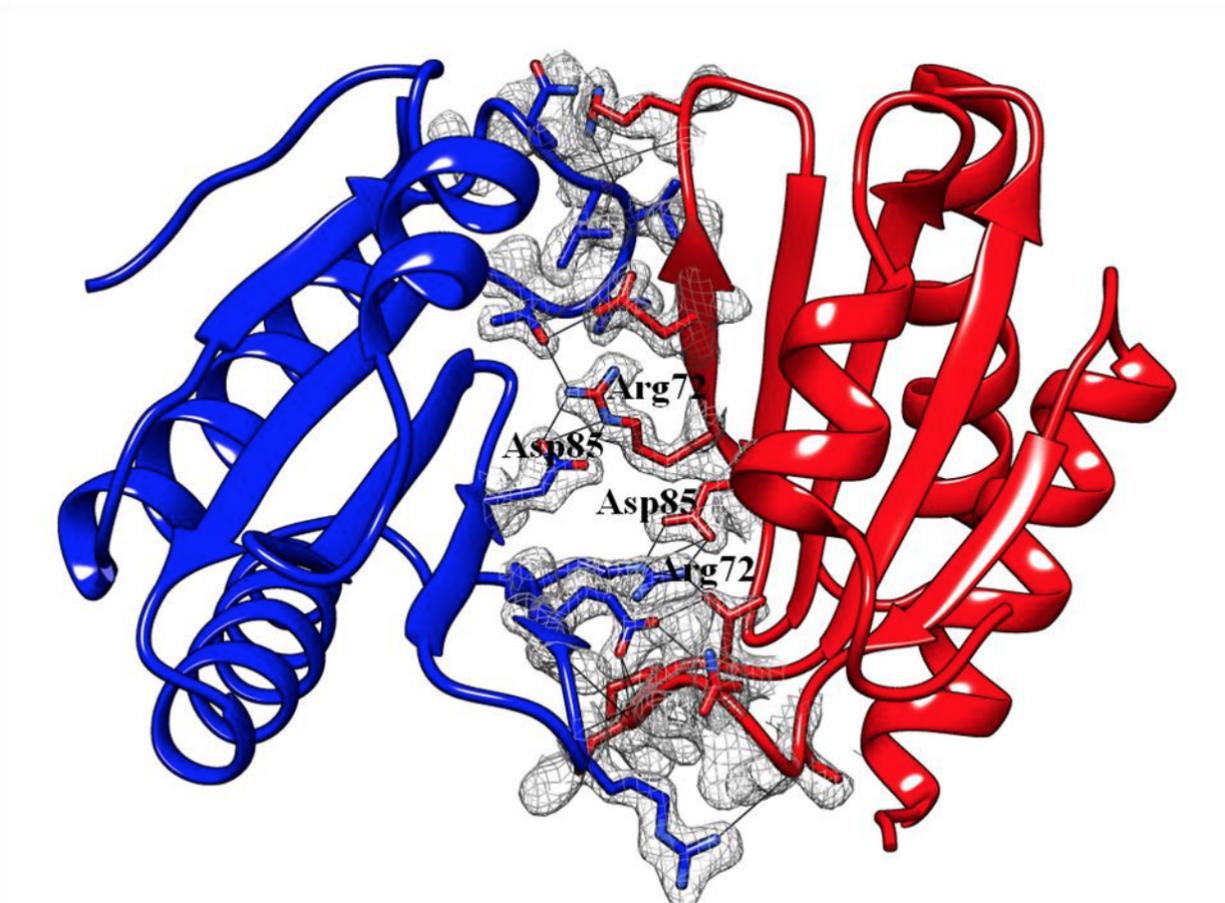


Figure A1. 10, related to Figure A1.5. Hydrogen bonds and salt bridges in *M. tuberculosis* RsfS dimer

Chain A was shown in blue, Chain B in red. The hydrogen bonds involved in the dimeric interface are shown in black line. The salt bridges between Arg72 and Asp85 are labeled. The residues involved in hydrogen bonding are highlighted as stick, and the $2F_o-F_c$ electron density map around these residues are shown in mesh at 1.5σ level.

A1.3.4 Cryo-EM structures of *M. tuberculosis* 50S

We used single particle cryo-electron microscopy (EM) to visualize the purified 50S *M. tuberculosis* ribosome mixed with recombinant *M. tuberculosis* RsfS at a molar ratio of 1:15. A density map was generated from 42,138 screened particles (**Figure A1.11**) and was calculated to be at 8.5-Å resolution using the gold standard Fourier shell correlation (Scheres and Chen, 2012)(**Figure A1.2**). We observed weak density for RsfS, suggesting that despite the 15-fold molar excess of RsfS to ribosome, we may have had a mixture of RsfS-bound and RsfS-free 50S particles. A modified supervised classification was used to classify the 42,138 particles into the RsfS-bound (20,851 particles) and RsfS-free (21,287 particles) states, from which two density maps were reconstructed (**Figure A1.1**). After the classification, the RsfS density was strong in the RsfS-bound state and clearly showed the secondary structures of the RsfS. The calculated final resolutions were 9.3 Å for the 50S subunit alone and 9.1 Å for the RsfS-bound 50S (**Figures A1.12A–A1.12D; Figure A1.2**). To rule out the possibility that the density of RsfS was due to the reference bias, we performed extra steps, described in the Materials and Methods, to prove the reliability of the classification.

At 8–9 Å, the cryo-EM density map of the core regions of the *M. tuberculosis* 50S ribosome agree well with the crystal structure of the *E. coli* 50S ribosome (Berk et al., 2006). The density was observed for the common structural motifs associated with the ribosome structures including the body, the stalk base (SB) of the L7/L12 arm near the A site (entry for the aminoacyl tRNA), the L1 stalk near the E site (exit site of the uncharged tRNA), and the central protuberance (CP) found between the L1 and L7/L12 stalks (**Figure A1.12**). It is known that the two peripheral stalks (L1 and L7/L12) of the 50S subunit are intrinsically dynamic. In fact, the L1 stalk of the *M. tuberculosis* structure has relatively weaker density and the density for L7/L12

was completely missing. Density maps for the ribosomal proteins L9 and L11 were missing, probably due to their loose association with the 50S subunit. **Figure A1.13** shows the local resolution of our density map calculated from ResMap (Kucukelbir et al., 2014). Most regions of our 50S density maps have a resolution better than 9 Å.

The most distinct structural feature of the *M. tuberculosis* 50S ribosome is a 107 nucleotide-long RNA helical extension of H54a (golden density in **Figures A1.12** and **A1.14**) that ends close to the L1 stalk and the E site. Density attributable to the nucleotide (nt) sequence extensions of rRNA (23S) was also observed in *M. tuberculosis* 50S subunit (**Figure A1.15**). Helix 15 and helix 16a are ~40 and 20 nt long in *M. tuberculosis*, observed in the density close to the base of the L1 stalk. Helix 31a, which is 25 nt longer in *M. tuberculosis* compared with *E. coli*, is visible at the solvent side of the 50S close to the CP. Helix 25 shows the greatest length variability among the three phylogenetic domains (Petrov et al., 2014). It is a short stem loop in *E. coli*, an ~80 nt bent helix in Archaea, and is longer as one progresses to higher organisms (876 nt in humans). *M. tuberculosis* H25 is 42 nt long, which is 15 nt longer than its counterpart in *E. coli*.

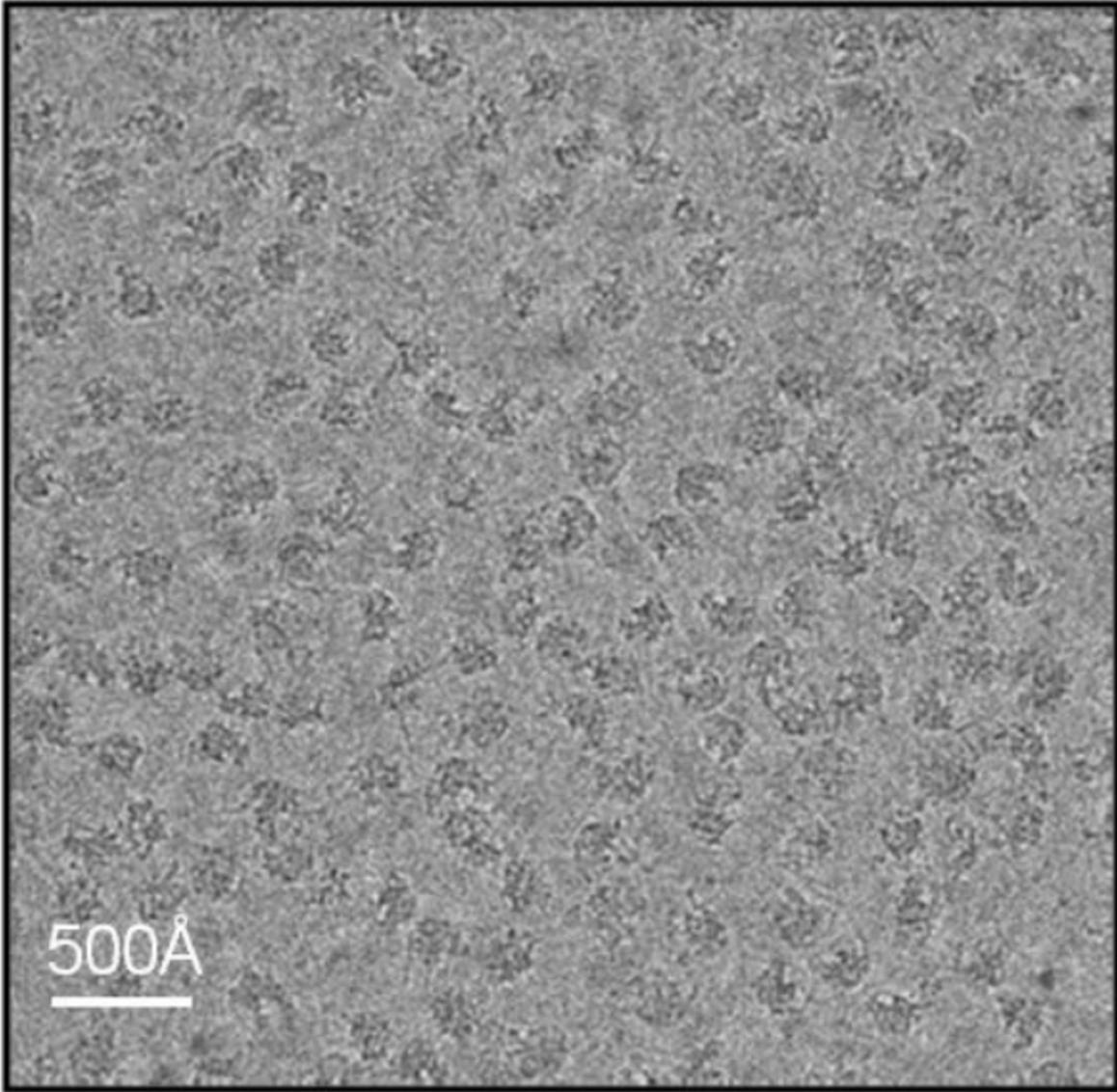


Figure A1. 11, related to Figure A1.12. Cryo-EM image of *M. tuberculosis* 50S + RsfS

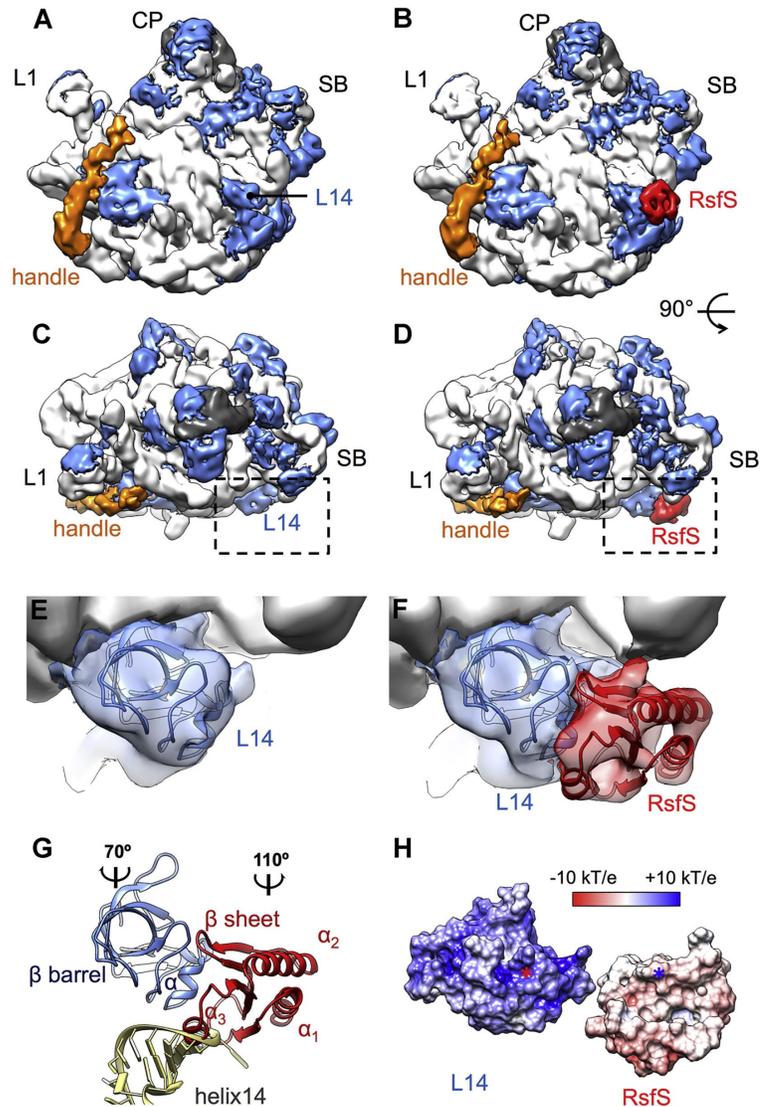


Figure A1.12 Cryo-EM Maps of *M. tuberculosis* Ribosome 50S in Its RsfS-free and RsfS-Bound States

The cryo-EM density maps of 50S ribosome (A) without RsfS and (B) RsfS-bound *M. tuberculosis* 50S are shown in “crown view” with their densities colored (23S RNA, light gray; 5S RNA, dark gray; ribosomal proteins, blue; RsfS, red; handle, gold).

(C–F) The “front view” of the density map made by rotating the maps 90° along the x axis in (A) and (B). Central protuberance (CP), stalk base (SB), and L1 protuberance (L1) are labeled accordingly. The dashed square regions in (C) and (D) show maps and models of the RsfS-free 50S (E) and RsfS-bound 50S (F). The cryo-EM density maps for 23S RNA are in gray, L14 in blue, and RsfS in red.

(G) The interacting surfaces between L14 and RsfS.

(H) L14 has a net positively charged side while the RsfS is negatively charged on the interface (see also Figures A1.1, A1.2, A1.11, A1.13, A1.14, A1.15, and A1.16).

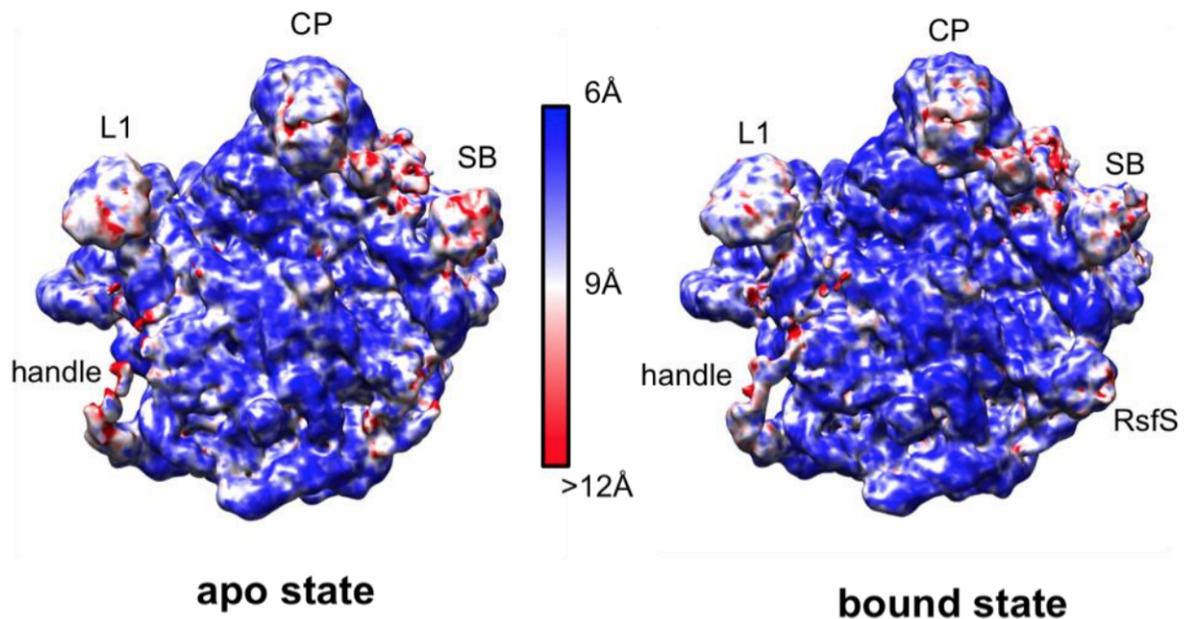


Figure A1. 13, related to Figure A1.12

Local resolutions of the reconstructions of the *M. tuberculosis* 50S ribosomes in the RsfS-free state and RsfS-bound states showing most regions of the maps have a resolution better than 9 Å (color ranges from blue 6 Å to red 12 Å or greater).

A1.3.5 A single subunit of RsfS directly binds to the L14 protein on the *M. tuberculosis* 50S

Cryo-EM density corresponding to a single RsfS protein was clearly identified on the surface of the 50S, interacting directly with the L14 protein. L14 is composed of a five-stranded β barrel, a C-terminal loop region that contains two small α helices, and a β ribbon that projects from the β barrel (Davies et al., 1996). The primary structures of L14 from *M. tuberculosis*, *E. coli*, *Thermus thermophilus*, and *Haloarcula marismortui* are highly conserved. *M. tuberculosis* L14 has between 66% and 78% sequence identities with these orthologs and, as expected, all three of the L14 crystal structures fit nicely into the density map. Our homology model of *M. tuberculosis* L14 was built using SWISS-MODEL (Schwede et al., 2003) based on the *E. coli* L14 structure and the resulting model was fit into the *M. tuberculosis* density map.

Given that RsfS is a dimer in solution, we were surprised to find that the cryo-EM density for the bound RsfS was only large enough to accommodate a single protein subunit (**Figures A1.12E and A1.12F**). In order to get the structure of the complex, the refined crystal structure of a single subunit of RsfS and the homology model of L14 were optimized to fit their cryo-EM densities by starting from random initial orientations of the RsfS relative to the L14. The two models were refined into density using the real-space refinement routine in PHENIX (Adams et al., 2010). The final refined model of the complex between L14 and RsfS had a cross-correlation score of 0.9 with its cryo-EM density map. Visual inspection showed very good agreement of the secondary structural components of the RsfS crystal structure with the EM density.

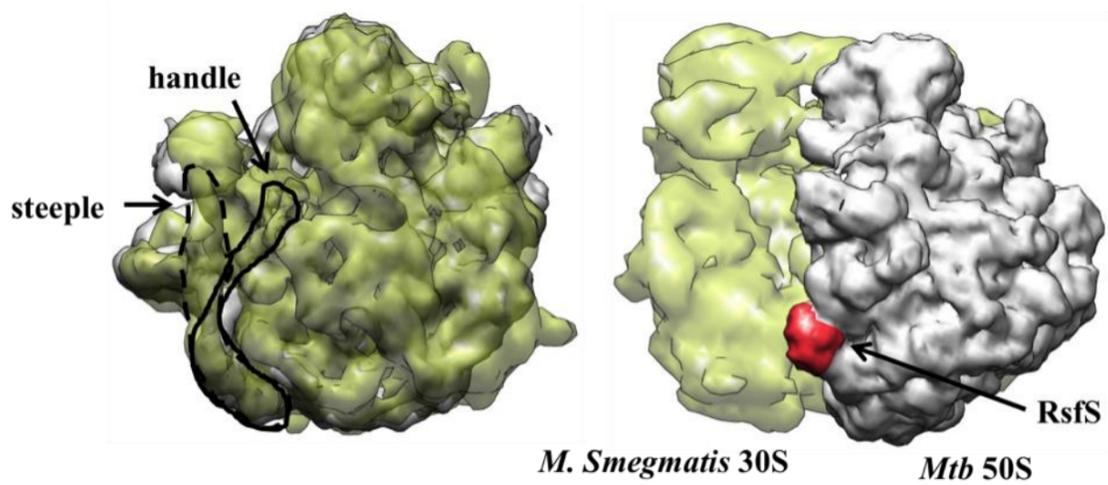


Figure A1. 14, related to Figure A1.12

“Crown-view” (Left) and Side view (Right) of *M. tuberculosis* 50S superimposed with *M. smegmatis* 70S cryo-EM density map. The “handle” in *M. tuberculosis* and “steeple” in *M. smegmatis* have been outlined and are overlapped on the base.

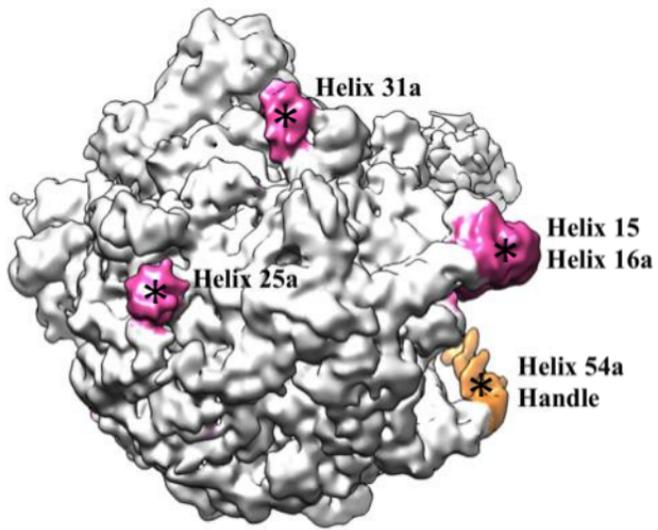


Figure A1. 15, related to Figure A1.12
Extra densities marked with * for nucleotide extensions.

Most of the atoms of RsfS that form the dimer interface are also at the interface of RsfS with L14. This is in agreement with the observation of a single subunit of RsfS dimer bound to L14. The structure of the complex indicates that the two small C-terminal α helices (Arg104 to Leu117) of L14 are interacting with the RsfS β sheet (β 1, Val26-Asp30; β 2, Cys40-Gly46; β 3, Arg72-Gly75; β 4, Trp81-Asp85; β 5, Ile89-His95), loop 2 (Val31-Asp39), and the C-terminal α 3 (Phe101-Gly109) of RsfS (**Figure A1.12G**). In fact, the buried surface area between L14 and RsfS is 2,212 Å², which represents about 20% of the total surface area of the RsfS-L14 complex (Krissinel and Henrick, 2007), slightly more than the buried surface area observed in the dimer. While it is not possible to assign hydrogen bonds at this resolution, the binding interface of the RsfS-L14 contains complementary electrostatic surface potentials as well as many potential H-bond donors and acceptors. At the interface of the two proteins, L14 has a net positive electrostatic surface while the RsfS interface has a negative potential (**Figure A1.12H**), indicating that H bonds and electrostatic forces are likely the dominant interactions between the RsfS and the L14. We have made 14 mutations to residues on RsfS that are common between the interface of the dimer and the RsfS-L14 complex. Only one mutation, E74A, provided soluble recombinant protein. The mutant was 64% less active in the inhibition for the *M. tuberculosis* translation assay. In addition, we found this mutant could not pull down 50S to the same degree as wild-type (**Figure A1.16**), suggesting that there was a significant reduction in affinity. However, the mutant protein was still a dimer in solution.

We compared our cryo-EM *M. tuberculosis* RsfS-50S structure to a published *E. coli* RsfS-50S homology model (Hauser et al., 2012). The model was constructed based on alanine scanning mutagenesis. Both *M. tuberculosis* RsfS and the modeled *E. coli* ortholog interact with L14 through their C-terminal helix α 3 and loop 2, as well as the β sheet. However, there is a

relatively large difference in the position and orientation of RsfS bound to L14. The position of the *E. coli* model of RsfS on L14 was rotated about 172° compared with the *M. tuberculosis* cryo-EM structure. While we cannot rule out the possibility that *E. coli* RsfS could bind differently, *M. tuberculosis* RsfS and L14 share 25% and 66% sequence identities with their *E. coli* counterparts, respectively, and therefore one would expect the binding to be conserved.

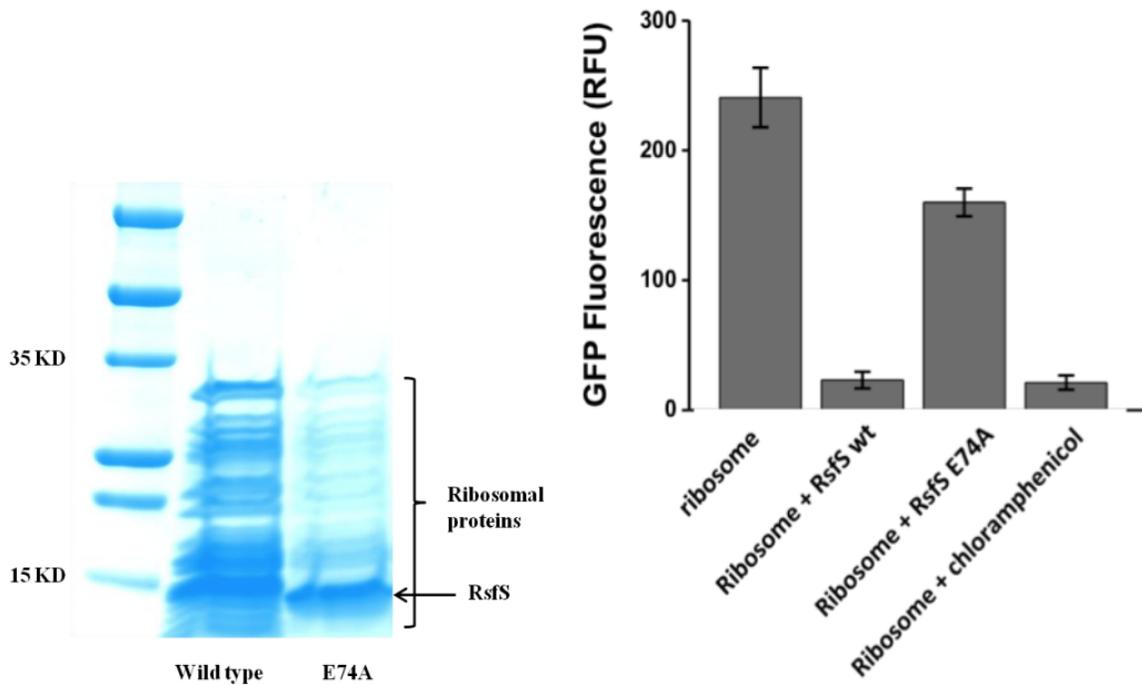


Figure A1. 16, related to Figure A1.12

(Left) Nickel pull down – assay of RsfS wild type and mutant E74A with *M. tuberculosis* ribosome.

(Right) GFP translation assay of RsfS wild type and E74A. Data are represented as mean \pm SEM.

A1.4 Discussion

A1.4.1 *M. tuberculosis* RsfS regulates translation by binding to 50S

The mechanism by which RsfS is able to slow translation when cells transition to the stationary phase is not well understood. Transcriptome analysis in *E. coli* and *M. tuberculosis* shows that RsfS mRNA levels are not significantly altered during bacterial growth (Hruz et al., 2008). Indeed, RsfS mRNA levels remained relatively constant from the early log phase through the stationary phase, a time when one would expect RsfS levels to increase so that translation would be slowed. The gene expression profiling results indicated that mRNA levels of RsfS and other ribosomal proteins are consistent (Galagan et al., 2010; Reddy et al., 2009). This suggests that RsfS is likely to be regulated at the protein level.

RsfS inhibits translation by directly binding to the L14 protein of the 50S ribosome at a site that, in the functional 70S, is occupied by helix 14 of the rRNA 16S in the 30S subunit. The overlap between the RsfS binding site and that of helix 14 is relatively small, only about 30 Å², compared with the extremely large buried surface area of the 50S with the 30S. The structures are consistent with the observation that RsfS binding is sufficient to compete with the 30S subunit. However, RsfS was unable to disassociate preformed 70S, as we only observed binding to the free 50S subunit. Yet when RsfS was added to an *M. tuberculosis* cell-free translation assay, it was able to significantly decrease translation, indicating that the monomer was present in the cell-free translation conditions. These results suggest that RsfS does not interfere with normal ribosomal functions during the elongation phase but it has the potential to block the formation of the functional 70S ribosome and to inhibit mRNA translation, both of which are consistent with RsfS's role as a regulator of translation.

The crystal structure and biochemical analysis clearly show that RsfS is a dimer in solution, and the EM density indicates that a single subunit of the dimer binds to L14 of the 50S subunit. It is highly unlikely that the second subunit of the dimer is bound and not visible due to flexibility, because most of the RsfS dimer interface observed in the crystal structure directly interacts with L14 in the inhibited RsfS-50S complex. Therefore, dissociation of the RsfS dimer must occur before binding to the L14.

A relatively large cavity is found adjacent to the interface of the RsfS dimer. On the periphery, the cavity contains the side chains of mostly polar and charged residues. It is tempting to speculate that binding of a molecule might be responsible for dimer dissociation. Numerous attempts using pull-down experiments have neither identified a molecule bound to RsfS nor conditions where RsfS dimer dissociates. However, other groups have reported that RsfS from *E. coli* interacted with several hypothetical proteins, such as yehL, yehQ, yihU, and yjcF in *E. coli* (Butland et al., 2005). *M. tuberculosis* does not have any identifiable orthologs to any of these proteins, again suggesting that *M. tuberculosis* dimer dissociation may be a regulatory event.

E. coli has three other proteins that have been implicated in the regulation of the ribosome in the bacterial transition to the stationary phase (Polikanov et al., 2012). Ribosome modulation factor (RMF) and hibernation promoting factor (HPF) are thought to act by inducing dimerization of the ribosome into a 100S particle, which is incapable of translation. Protein Y (PY) has been shown to reverse this ribosome dimerization, although the resulting 70S ribosome appeared to be inactive. RMF and HPF are thought to induce dimerization by binding to the mRNA and tRNA binding sites on 30S subunits. PY is a paralog of HPF and its binding site overlaps with that of HPF and part of the RMF. We searched for the presence of RMF, HPF, and

PY by BLASTP (Altschul et al., 1990) in *M. tuberculosis* using *E. coli* counterpart sequences and found that only one hypothetical protein (Rv3241c) shares homology with HPF.

A1.4.2 RsfS may have multiple roles in *M. tuberculosis*

It is possible that RsfS serves other roles in the bacterium aside from simply silencing the ribosome. Interestingly, the RsfS homolog in humans, C7orf30, has been shown to participate in the assembly and stability of the large subunit of the mitochondrial ribosome. Inactivation of C7orf30 using RNAi leads to instability and an assembly defect in the large subunit, which results in reduced mitochondrial translation (Fung et al., 2013; Rorbach et al., 2012; Wanschers et al., 2012).

The initiation factor eIF6 is highly conserved from yeast to mammals. It binds to RpL23, the L14 counterpart in yeast. Binding of eIF6 to the large 60S subunit inhibits subunit joining and thus prevents translation initiation (Gartmann et al., 2010; Klinge et al., 2011). The mechanism of anti-association factor eIF6 is likely to be similar to that of RsfS. However, this protein shares no structural similarity with RsfS. While eIF6 consists of five β -sheets that form a barrel to cap RpL23, RsfS contains only one β sheet. It is possible that RsfS is involved in the assembly of the ribosome and prevents the association of a premature 50S.

Acknowledgments

This work was supported by Welch foundation grants A-0015 (J.C.S.), A-1863 (J.Z.) and National Institutes of Health (NIH) TB Structural genomics grant P01AI095208. We would like to thank the staffs at the beam line 5.0.2 managed by the Berkeley Center for Structural Biology (BCSB) at the Advanced Light Source (ALS) for technical support. The BCSB is supported in

part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. J.Z. is grateful to Michael Levitt and Roger Kornberg at Stanford University for their support on the cryo-EM experiments and to BioX3 at Stanford University for the initial cryo-EM data processing. J.Z. would like to acknowledge the Texas A&M Supercomputing Facility for providing computing resource, Center for Phage Technology and the Department of Biochemistry and Biophysics at Texas A&M University for providing startup funding. We thank Jeng-Yih Chang for the preparation of Figure A1.12H and Dr. Matthew Sachs for carefully editing the paper.

APPENDIX II

CARBON-PHOSPHORUS LYASE COMPLEX

A2.1 Introduction

Phosphorus is essential for life. Currently, this element is most abundant as phosphoric acid and phosphate esters (Wackett et al., 1987). However, certain organisms can acquire needed phosphorus under conditions of low phosphate directly from organophosphonate compounds, a class of compounds that contain a carbon-phosphorus (C-P) bond that must be enzymatically cleaved to phosphate (McGrath et al., 2013). Phosphonates have been proposed as a major source of phosphorus on the prebiotic earth and large quantities of phosphonates are discharged annually into the environment as agricultural herbicides and industrial detergents (Ternan et al., 1998). In bacteria such as *Escherichia coli*, the C-P bond in unactivated organophosphonates can be enzymatically cleaved to form phosphate and an alkane by the multi-subunit C-P lyase complex (Chen et al., 1990). In *E. coli*, the C-P lyase complex is encoded by the 14 genes (*phnCDEFGHIJKLMNOP*) contained within the *phn* operon (Chen et al., 1990; Metcalf and Wanner, 1993a, b). Previous genetic and biochemical studies have demonstrated that seven of the genes in this operon (*phnGHIJKLM*) are critical for the expression of proteins that are required for the enzymatic cleavage of the C-P bond during the transformation of phosphonates to phosphate (Metcalf and Wanner, 1993b). The individual subunits of the *E. coli*

Most of this appendix is reprinted with permission from “Structures of the Carbon-Phosphorus Lyase Complex Reveal the Binding Mode of the NBD-like PhnK” by Kailu Yang, Zhongjie Ren, Frank M. Raushel, Junjie Zhang, *Structure* 2015 24(1), 37-42, Copyright 2016 by Elsevier Inc. Author contributions: ZR purified the sample. KY solved the structures. KY, ZR, FMR, and JZ designed the experiments, interpreted the results, and wrote the paper. FMR and JZ are the principal investigators.

C-P lyase complex that are essential for the enzymatic cleavage of the C-P bond in methylphosphonate have been reconstituted and characterized *in vitro* (Kamat et al., 2013; Kamat et al., 2011). The series of enzymatic steps catalyzed by the C-P lyase complex is illustrated in **Figure A2.1A**. The initial reaction of the C-P lyase pathway is catalyzed by a nucleotide phosphorylase, PhnI, in the presence of PhnG, PhnH and PhnL in which ATP and methylphosphonate are converted to adenine and α -D-ribose-1-methylphosphonate-5-triphosphate (RPnTP). The RPnTP is hydrolyzed by the phosphohydrolase PhnM to produce pyrophosphate and 5-phosphoribosyl-1-phosphonate (PRPn). PRPn serves as a substrate for the cleavage of the C-P bond by PhnJ through a SAM-dependent radical-based reaction, transforming PRPn to the production of 5-phosphoribosyl-1, 2-cyclic phosphate (PRcP) and methane (or corresponding alkane).

Fragments of the C-P lyase complex containing PhnG₂I₂, PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K can be expressed and purified in high yield (Jochimsen et al., 2011). We have used mass spectrometry and H/D exchange methods to construct a low-resolution interaction map of the PhnG₂H₂I₂J₂K complex that illustrates how the individual subunits associate with one another to form larger multi-subunit complexes (Ren et al., 2015). Interestingly, the demonstrated stoichiometry indicates that one copy of PhnK binds to the dimeric core complex (PhnG₂H₂I₂J₂). The X-ray crystal structure of the core complex PhnG₂H₂I₂J₂ has been recently reported (Seweryn et al., 2015). It revealed an intertwined network of subunits PhnG, PhnH, PhnI and PhnJ with self-homologies. However, how PhnK binds to the core complex remains unclear.

From sequence analysis, PhnK is homologous to the nucleotide-binding domain (NBD) of ATP-binding cassette (ABC) transporters. PhnK has all the NBD motifs: Walker A, Walker B, ABC signature, A-loop, D loop, Q-loop, and switch H-loop. The role of PhnK in the C-P lyase

pathway is unclear: PhnK is apparently not required for any of the essential reactions of the C-P lyase pathway in vitro (Kamat et al., 2011). However, the *E. coli* $\Delta phnK$ strain is phosphonate growth deficient (Metcalf and Wanner, 1993b). PhnK was previously mapped in a groove close to the two-fold symmetry axis of the core complex $PhnG_2H_2I_2J_2$ based on a low-resolution density map from negative stain electron microscopy (negative stain EM) (Seweryn et al., 2015). The ABC signature was suggested to bind to the core complex based on the assumption that conserved regions comprise the interaction surface. However, this model for the binding of PhnK needs to be validated. Moreover, the following questions still remain unanswered: (1) Why is there only one copy of PhnK bound to a dimeric $PhnG_2H_2I_2J_2$ core complex? (2) How does PhnK bind to the core complex? (3) What is the effect on the core complex after PhnK binds? A higher-resolution structure of $PhnG_2H_2I_2J_2K$ is clearly needed. Unfortunately, attempts to crystalize the $PhnG_2H_2I_2J_2K$ complex have thus far failed.

Recent advances in single-particle cryo-electron microscopy (cryo-EM) have enabled the structure determination of large macromolecular complexes, usually with a molecular weight of more than 400 kDa, to atomic resolutions (Bartesaghi et al., 2015; Campbell et al., 2015; Fischer et al., 2015; Grant and Grigorieff, 2015; Jiang et al., 2015; Yu et al., 2015). This breakthrough is mostly attributed to the development of the instrumentation needed to acquire better images (Bai et al., 2013; Bammes et al., 2012; Fischer et al., 2015; Li et al., 2013), and the software to more reliably process and validate the results (Frank et al., 1981; Hohn et al., 2007; Lyumkis et al., 2013; Scheres, 2012; Tang et al., 2007), especially the use of statistical approach for processing images of particles with conformational and compositional heterogeneity (Lyumkis et al., 2013; Scheres, 2012). Such resolution revolution (Kuhlbrandt, 2014) in cryo-EM is expanding to the structure determination of small (~200 kDa in size), nonsymmetrical macromolecular complexes,

with a few examples being determined to near-atomic or subnanometer resolutions (Kim et al., 2015; Liang et al., 2015; Lu et al., 2014). In this study, we used single-particle cryo-EM to determine two respective structures of the PhnG₂H₂I₂J₂K (248 kDa) and PhnG₂H₂I₂J₂ (220 kDa) complexes at 7.8Å resolution from a sample containing both of them. The interaction between the monomeric PhnK and the dimeric PhnG₂H₂I₂J₂ was revealed.

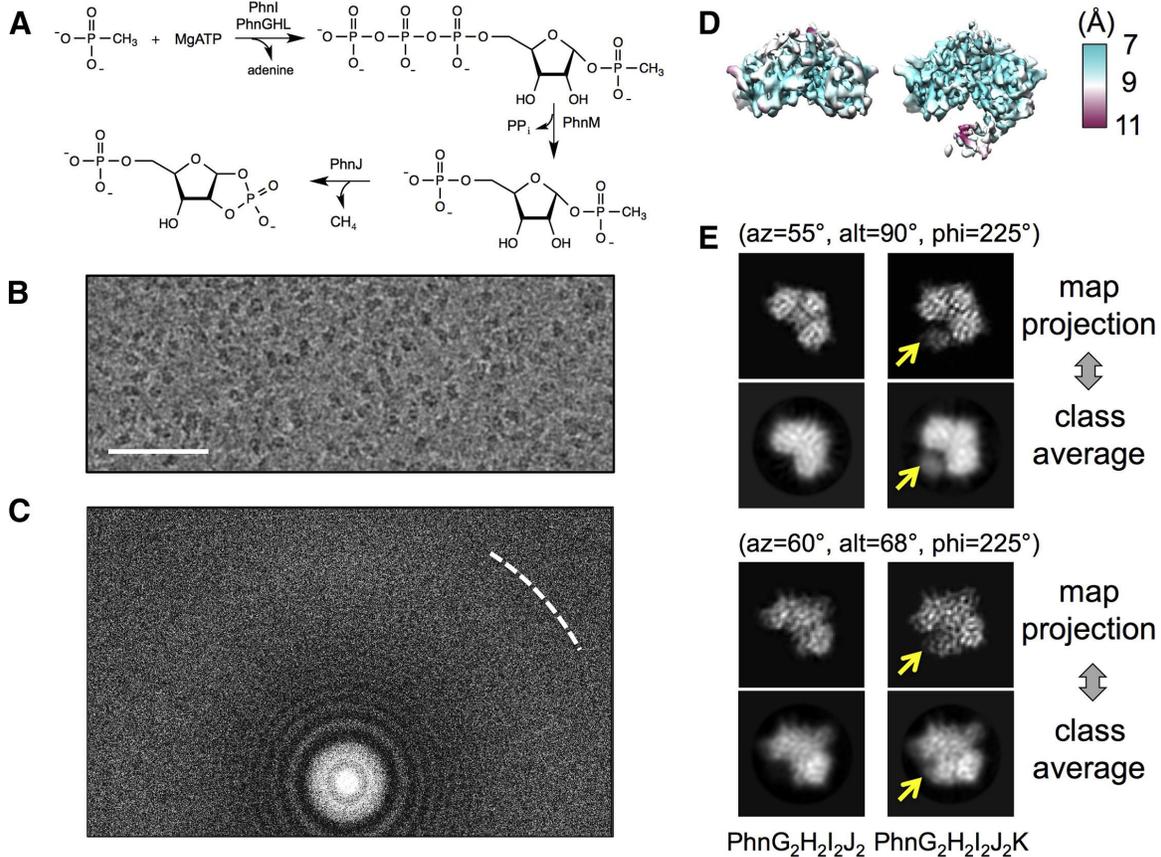


Figure A2.1 Cryo-EM Analysis Reveals the Co-existence of PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K

(A) The series of the enzymatic steps of C-P lyase.

(B) A region of a representative micrograph showing the particles of PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K with a defocus of $-1.2 \mu\text{m}$. The scale bar denotes 500 \AA .

(C) Power spectrum of the same micrograph in (B). The white dashed curve shows the Nyquist at 3 \AA .

(D) 3D density maps of PhnG₂H₂I₂J₂ (left) and PhnG₂H₂I₂J₂K (right). The maps were colored based on the local resolutions.

(E) Projections of the two maps compared with the reference-free class averages in two representative orientations. The Euler angles were based on the EMAN2 convention.

Yellow arrows indicate the location of PhnK. See also Figure A2.2.

A2.2 Materials and Methods

A2.2.1 Protein purification

The plasmids for the expression of PhnG₂H₂I₂J₂K from the C-P lyase complex of *E. coli* were cloned, transformed and expressed as previously described (Ren et al., 2015). The cells were resuspended (4:1, v/w) in binding buffer (50 mM HEPES [pH 8.5], 150 mM NaCl, 20 mM imidazole and 2 mM tris (2-carboxyethyl) phosphine) containing 0.10 mg/mL PMSF and 0.1% protease cocktail for polyhistidine-tagged protein purification. The cells were lysed by sonication and the supernatant solution was separated from the insoluble debris by centrifugation at 10,000 x g for 20 minutes at 4°C. The pellet was discarded and the supernatant solution was applied to a 5-mL HisTrap (GE Healthcare) column, which was pre-equilibrated with 10 column volumes of binding buffer. The proteins were eluted with a gradient of elution buffer (0.5 M imidazole in binding buffer). The eluted fractions were pooled and applied to a High Load 26/60 Superdex 200 prep grade gel filtration column (GE Healthcare), which was previously equilibrated with running buffer (binding buffer without imidazole). The fractions were pooled and analyzed by SDS-PAGE.

A2.2.2 Electron microscopy

PhnG₂H₂I₂J₂K was diluted to 0.1 mg/mL with the dilution buffer (50 mM HEPES [pH 8.5], 150 mM NaCl, 2 mM TCEP). 3 µL of the sample was applied to a C-Flat 1.2/1.3 holey carbon grid at 16°C with 100% relative humidity and vitrified using a Vitrobot Mark III (FEI). The thin-ice areas that were expected to show clearly visible and mono-dispersed particles were imaged under an FEI Tecnai F20 cryo electron-microscope with a field emission gun (FEI) operated at 200 kV. Data were recorded on a Gatan K2 Summit electron-counting direct

detection camera (Gatan) in electron counting mode (Li et al., 2013). A nominal magnification of 25,000x was used, yielding a pixel size of 1.5 Å. The beam intensity was adjusted to a dose rate of ~10 electrons per pixel per second on the camera. A 25-frame or 50-frame movie stack was recorded for each picture, with 0.2 s per frame for a total exposure time of 5 s or 10 s, respectively.

A2.2.3 Image Processing

967 collected image stacks were iteratively aligned and filtered based on the electron dose using Unblur (Grant and Grigorieff, 2015). 694 sum images were selected according to the visibility of particles and power spectra. For particle picking, the sum images were first low-pass filtered to 15 Å to visualize particles unambiguously. Then the Erase tool of e2boxer.py in EMAN2 (Tang et al., 2007) was used to erase bad areas that have junk or have no particles. Finally the Swarm tool of e2boxer.py was used to semi-automatically pick 492,889 particles using a box size of 120 pixels, during which the threshold for picking was interactively changed so that the false-positive picking was minimized. The coordinates were imported into Relion (Scheres, 2012) to extract the particles from the original sum images with a downscaling factor of 1, 2 or 4. Reference-free 2D classification was performed in Relion to remove bad particles that cannot average well with each other, which generated a dataset of 294,203 particles. We selected 3,956 particles with good contrast to generate an initial model using the PRIME routine in the SIMPLE package (Elmlund et al., 2013), which was low-pass filtered to 60 Å and used for the first round of unsupervised 3D classification in Relion. The 3D classification and refinement were performed hierarchically (**Figure A2.2**). Briefly, in the first two rounds of 3D classification, particles were separated into three classes based on its quality and the presence or

absence of PhnK. 50,061 good particles from one class with PhnK were selected and further refined with C1 symmetry to produce the density map of PhnG₂H₂I₂J₂K. 86,391 particles without PhnK were subjected to an additional round of 3D classification to separate into three classes, which yielded 23,861 good particles in one class. These 23,861 particles were further refined with C2 symmetry to produce the density map of PhnG₂H₂I₂J₂. The particle images were downsampled by 4 times in the first two rounds of the 3D classification and downsampled by 2 times in the later 3D classification and refinement. Refinement against the particles without downscaling, did not improve the final resolution of the maps. Blocres of the Bsoft package (Heymann and Belnap, 2007) was used to estimate the local resolution of the cryo-EM maps. Our resolution is limited to 7.8Å, which may be due to the flexible nature of the specimen.

A2.2.4 Molecular modeling and map segmentation

Fitting of models into the density map is done in UCSF Chimera (Pettersen et al., 2004). The correct handedness of the density maps of PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K were initially determined by docking the crystal structure of PhnH (PDB: 2FSU) (Adams et al., 2008) into both the original maps and the maps with the flipped handedness. The crystal structure of PhnH fitted into the flipped maps with all the secondary structures matching, confirming that the original maps should have their handedness flipped. A homology model of PhnK was built using SWISS Model (Guex et al., 2009) based on a NBD (PDB: 4U00) (Devi et al., 2015). The location of PhnK was determined by the difference map between PhnG₂H₂I₂J₂K and PhnG₂H₂I₂J₂. Moreover, the homology model of PhnK fitted the difference map with all of the secondary structures in place, confirming the correct location of PhnK in PhnG₂H₂I₂J₂K. Segmentation of PhnG₂H₂I₂J₂ core was done using Segger (Pintilie et al., 2010) in UCSF Chimera based on the

crystal structure of PhnG₂H₂I₂J₂ (PDB: 4XB6). The electrostatic surface of the models was calculated using the APBS module (Baker et al., 2001) in UCSF Chimera.

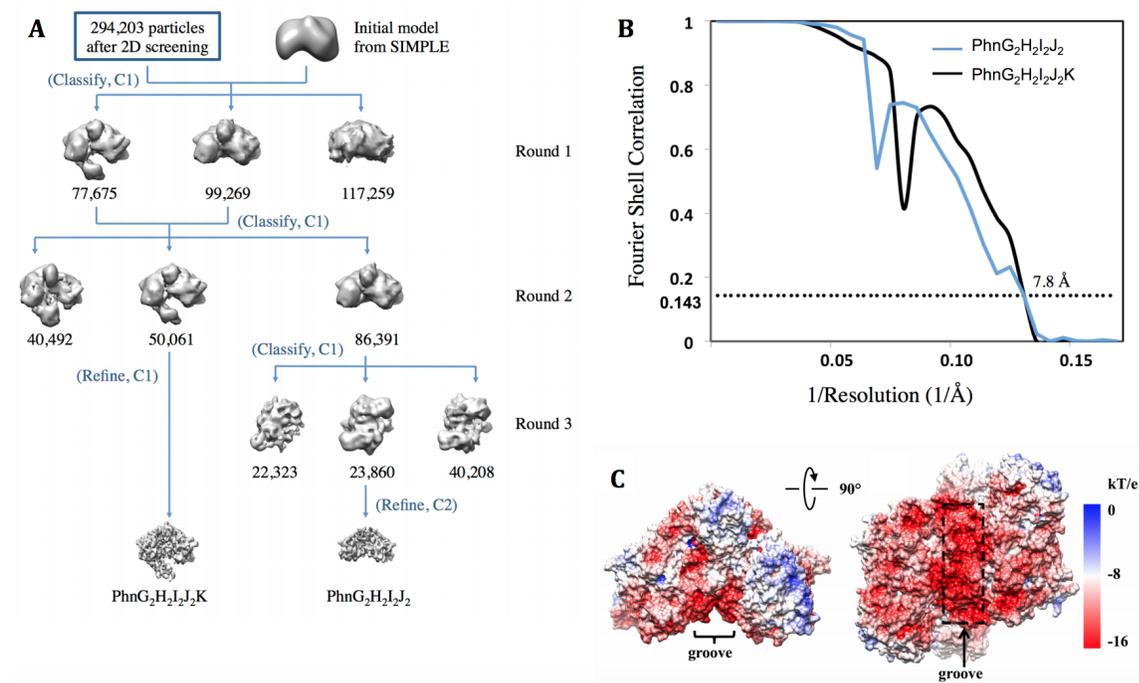


Figure A2. 2, related to Figures A2.1 and A2.3

(A) Hierarchical unsupervised classification of particles in Relion. “(Classify, C1)” denotes unsupervised 3D classification without any symmetry constraints. “(Refine, C1)” denotes auto-refinement without any symmetry constraints. “(Refine, C2)” denotes auto-refinement with C2 symmetry constraint. Initial models for all classifications or refinements were low-pass filtered to 60Å. In round 1 and round 2, particles were downsampled by 4 times and separated mainly by the presence or absence of PhnK. In round 3, particles were downsampled by 2 times. All maps are shown with the original handedness and in the original order of the output. Refinement against the particles without downscaling, did not improve the final resolution of the maps.

(B) Fourier shell correlations of the reconstructions of PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K.

(C) Electrostatic surface of PhnG₂H₂I₂J₂ shows a negatively charged groove. The electrostatic surface is calculated based on the PDB structure 4XB6.

A2.2.5 Accession Numbers

The cryo-EM maps have been deposited in the EM Databank with accession codes EMD-6410 (PhnG₂H₂I₂J₂K) and EMD-6411 (PhnG₂H₂I₂J₂).

A2.3 Results

A2.3.1 Co-existence of PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K

The genes expressing PhnG, PhnH, PhnI, PhnJ and PhnK were cloned into a pET-28b plasmid with a polyhistidine tag on the N-terminus of PhnG as previously described (Jochimsen et al., 2011; Ren et al., 2015). The expressed proteins were isolated using a nickel column and then further purified by passage through a gel filtration column. The purified complexes were vitrified and imaged (Materials and Methods). The collected movie stacks, from an electron-counting direct detection camera (Li et al., 2013), showed clear particles and strong power spectra after aligning the frames within each movie (Grant and Grigorieff, 2015) (**Figures A2.1B** and **A2.1C**). A total of 492,889 particles were semi-automatically picked, and subsequently classified with reference-free two-dimensional (2D) classification and unsupervised three-dimensional (3D) classification in Relion (Scheres, 2012) (**Figure A2.2**). Two populations of particles were identified and reconstructed into two density maps, which correspond to the PhnG₂H₂I₂J₂ (core complex) and PhnG₂H₂I₂J₂K (core complex bound with PhnK), respectively. These two maps were similar except for an extra lobe of density in one of them (**Figure A2.1D**). Differences in the extra density were also observed in the reference-free 2D class averages calculated from all particles (**Figure A2.1E**). This extra density was later confirmed as the PhnK subunit of the PhnG₂H₂I₂J₂K complex. Further refinements against the separated particles yielded the cryo-EM density maps of PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K, both at 7.8Å resolution,

based on the gold-standard Fourier shell correlation (FSC, **Figure A2.2**) (Scheres and Chen, 2012). Local resolution was estimated, showing most regions of the maps were at a resolution range between 7 and 8 Å (**Figure A2.1D**). At this resolution, the protein secondary structures were clearly visible.

The existence of a complex missing PhnK, in the purified sample expressed from a plasmid containing the *phnGHIJK* genes, is consistent with previous native gel electrophoresis and mass spectrometry experiments (Jochimsen et al., 2011), suggesting the loose binding of PhnK to the core complex PhnG₂H₂I₂J₂. The fact that PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K co-exist in the same sample apparently hindered previous trials for the crystallization of PhnG₂H₂I₂J₂K (Seweryn et al., 2015).

A2.3.2 Architecture of PhnG₂H₂I₂J₂ with or without PhnK bound

The 3D density maps of the PhnG₂H₂I₂J₂ and PhnG₂H₂I₂J₂K complexes measure 110 Å x 70 Å x 100 Å and 110 Å x 110 Å x 100 Å in dimension, respectively (**Figure A2.3**). The common part of these two complexes is the PhnG₂H₂I₂J₂ core. The core retains the same architecture after PhnK binding. In both density maps, the two copies of PhnG, PhnH, PhnI and PhnJ are arranged in a head-to-tail fashion in a sequence of H-J-I-G as labeled in **Figure A2.3B**. While the cryo-EM maps were independently determined and not biased by the recently determined crystal structure of PhnG₂H₂I₂J₂ (Materials and Methods), the crystal structure fits well into the cryo-EM density map of the core complex PhnG₂H₂I₂J₂ with matching secondary structures (**Figure A2.3C**).

A2.3.3 Two identical binding sites for PhnK on PhnG₂H₂I₂J₂

The binding site for PhnK is located on the α -helix 6 (residue 147-158) and a nearby loop (residue 227-230) of PhnJ. We term this loop “chock-loop” as it may stabilize the binding of PhnK like a chock. Two PhnJ subunits in the core complex provide two identical binding sites for PhnK (purple polygons in **Figure A2.3B**). These two binding sites are approximately 50 Å away from each other. Noticeably, between these two binding sites lies a negatively charged groove (**Figure A2.2**), which was previously proposed as the PhnK-binding region based on a low-resolution density map of PhnG₂H₂I₂J₂K from negative stain EM and protein sequence conservation (Seweryn et al., 2015). However, the entire groove is exposed to solvent and not in contact with PhnK in our cryo-EM map of PhnG₂H₂I₂J₂K.

A2.3.4 One copy of PhnK, through its helical domain, binds to PhnG₂H₂I₂J₂

Even though the core complex provides two identical binding sites for PhnK, only one copy of PhnK was found to bind (**Figures A2.3D, A2.3E, A2.3F and A2.4**). PhnK is homologous to the conserved NBD of proteins in the ABC transporter family. Like NBDs, PhnK consists of a helical domain and a RecA-like domain. It has all of the sequence motifs (Walker A, Walker B, ABC signature, A-loop, D-loop, Q-loop and switch H-loop) of the NBD (**Figure A2.4A**) (Ambudkar et al., 2006; Davidson et al., 2008). A sequence alignment of PhnK with other structurally characterized NBDs is presented in **Figure A2.5**. NBDs are known to interact with a transmembrane domain or protein (TMD) through its helical domain (Schmitt et al., 2003). The L-loop of the TMD inserts into a cavity around the helical domain of the NBD (**Figure A2.6**), which forms the major interaction between NBD and TMD (Locher et al., 2002). In contrast, PhnK, an NBD-like protein, interacts with the cytoplasmic protein complex

PhnG₂H₂I₂J₂. This interaction is via a helix-turn-helix motif (residues 100-120 containing the α -helices 3 and 4) within the helical domain of PhnK to bind to the α helix 6 and chock loop of PhnJ (**Figures A2.4A and A2.4C**). When PhnK binds to PhnJ, the cavity of PhnK is completely exposed to solvent. The chock loop of PhnJ may instead provide additional support to block the movement of PhnK (**Figure A2.6**). The binding interface between PhnK and PhnJ shows complementary electrostatic surface potentials. At the interface, PhnK exhibits a net positive charge while PhnJ exhibits a negative charge (**Figure A2.4D**).

Attempts to dock a second copy of PhnK, in the same conformation as the first one, to the empty binding site in the PhnG₂H₂I₂J₂K, resulted in considerable collision between their RecA-like domains, which prevents the simultaneous binding of two PhnK subunits.

A2.3.5 Binding of PhnK exposes the active site residue Gly32

The universally conserved Gly32 of PhnJ was proposed to generate a stable glycy radical enzyme that supports multiple turnovers without further SAM consumption (Kamat et al., 2013).

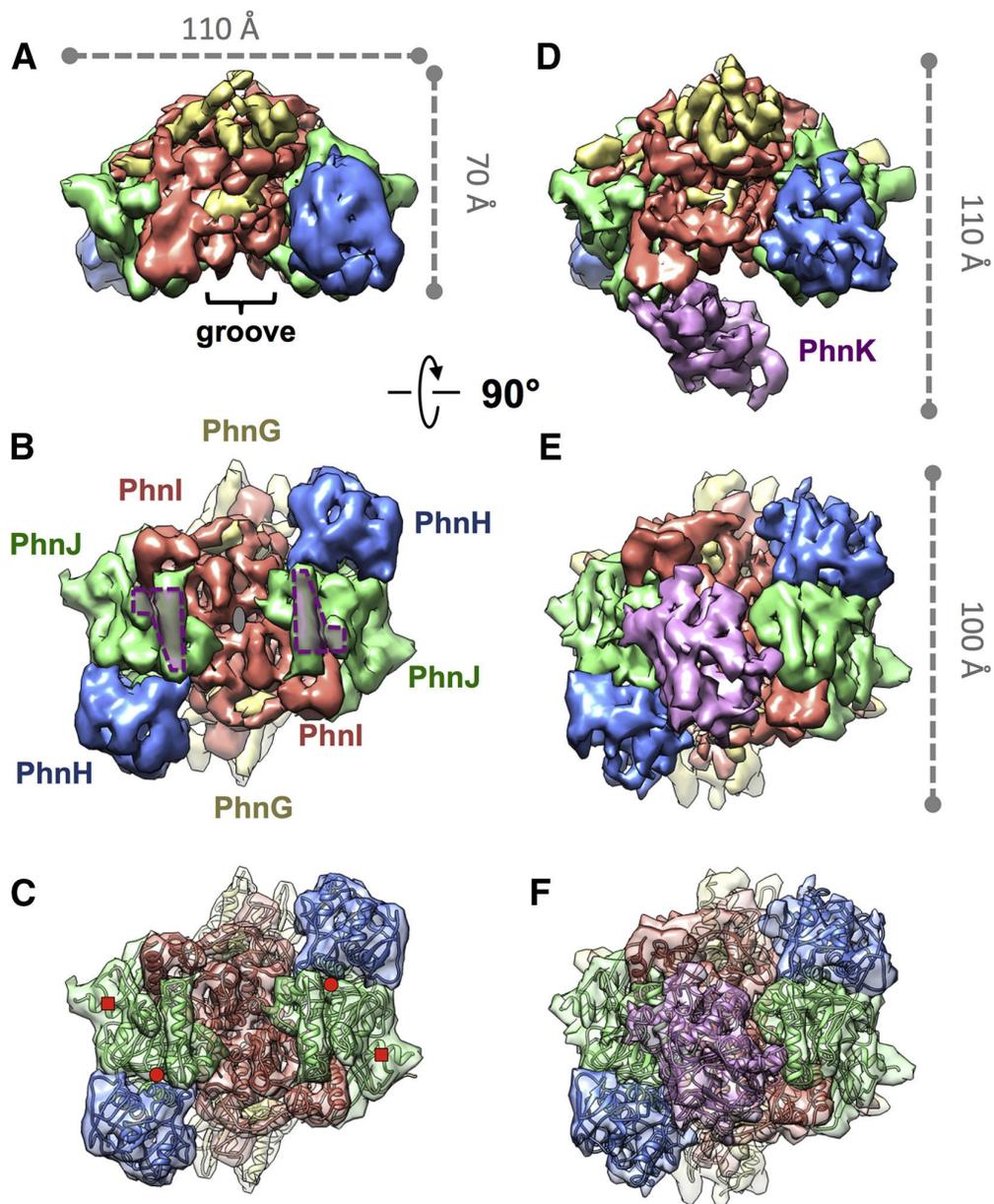


Figure A2.3 Architecture of PhnG₂H₂I₂J₂ without or with PhnK Bound

(A and B) Front and bottom view of PhnG₂H₂I₂J₂. In (A), the groove on the core complex PhnG₂H₂I₂J₂ is labeled. In (B), the gray ellipse denotes the two-fold symmetry axis, and the two purple polygons denote the two identical binding sites for PhnK.

(C) The crystal structure of PhnG₂H₂I₂J₂ fitted into the cryo-EM map of PhnG₂H₂I₂J₂ in the same orientation as (B). The red circles and squares denote the positions of Gly32 and Cys272 of PhnJ, respectively.

(D and E) Front and bottom view of PhnG₂H₂I₂J₂K in the same orientations as (A) and (B), respectively.

(F) The crystal structure of PhnG₂H₂I₂J₂ and the homology model of PhnK fitted into the cryo-EM map of PhnG₂H₂I₂J₂K in the same orientation as (E). See also Figure A2.2.

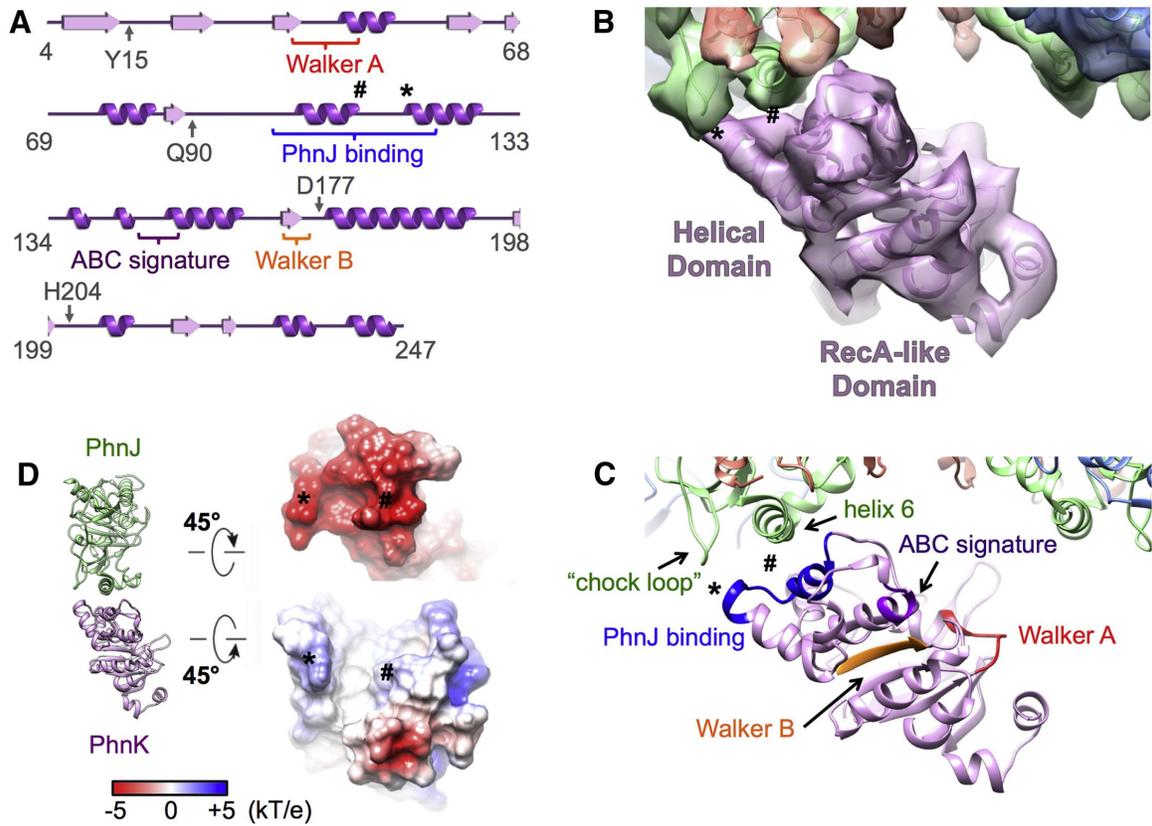


Figure A2.4 PhnK is an NBD-like Protein and Interacts with PhnJ
(A) Secondary structure of PhnK with the NBD motifs labeled. The symbols # and * label the sites for the interactions between PhnK and PhnJ.
(B) One PhnK binds to the core complex with the models fitted in the map.
(C) The same view as (B) with the NBD motifs color labeled.
(D) Electrostatic surface potentials at the interface of PhnJ and PhnK. See also Figures A2.5 and A2.6.

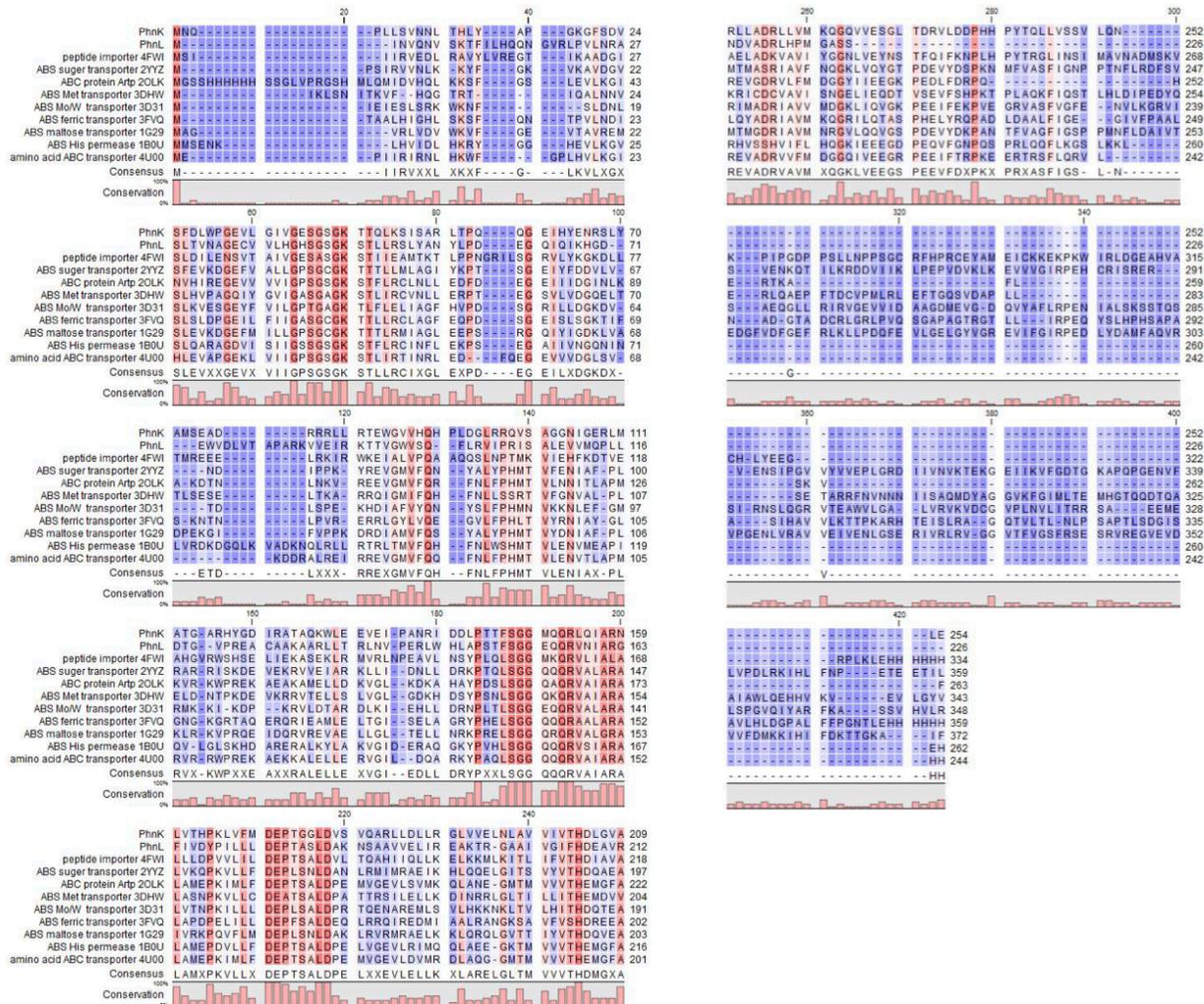


Figure A2.5, related to Figure A2.4
Sequence alignment of PhnK, PhnL with NBDs in the ABC transporter family using CLC sequence viewer (CLC bio, QIAGEN).

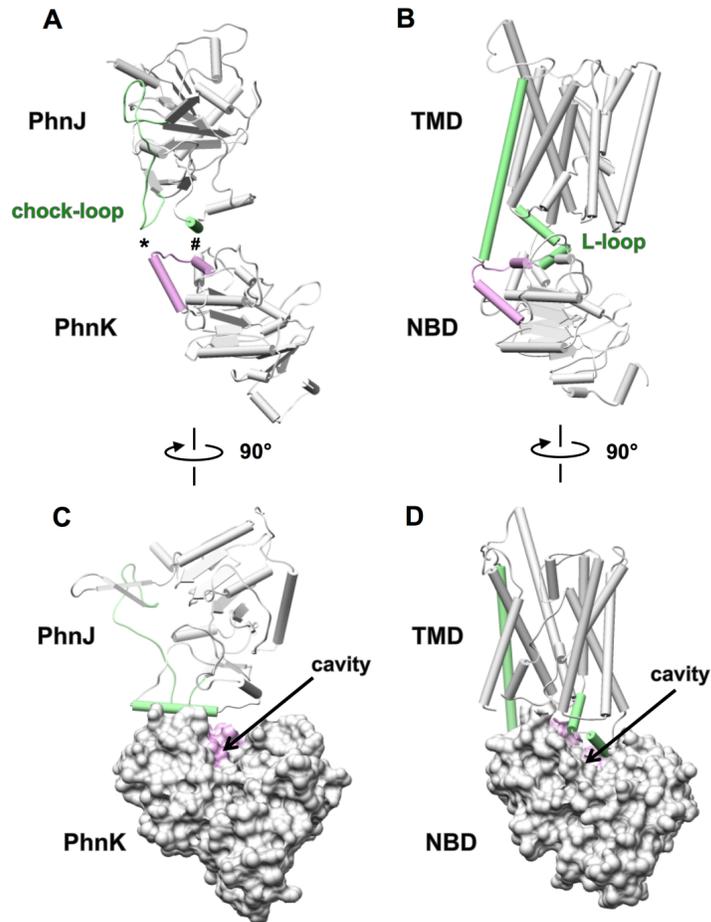


Figure A2. 6, related to Figure A2.4

The interaction mode between PhnK and PhnJ (A, C) was compared with the interaction between a classical NBD and the corresponding TMD (B, D; PDB id: 1L7V). Purple color: Helices 3 and 4 of PhnK interacting with PhnJ, and their corresponding helices in the NBD. Green color: the chock-loop and Helix 6 of PhnJ, the L-loop and a helix (at the similar position of the chock-loop) of TMD. PhnK and NBD in (C) and (D) are in the surface representation to show the cavity for L-loop binding in ABC transporters.

In the core complex, the two Gly32, each from one subunit of PhnJ, are symmetrically located on the two sides of the complex. Both of these glycine residues are buried in PhnJ near the interface between PhnJ and PhnH. In the map of PhnG₂H₂I₂J₂K, the density burying Gly32 of PhnJ shows striking difference between the two copies of PhnJ (**Figure A2.7**). On the side where there is no PhnK bound (apo side), the density burying Gly32 is stronger than the corresponding density on the side where there is a PhnK bound (bound side), indicating the interaction between PhnJ and PhnH in the apo side is tighter than in the bound side. On the apo side, the residue Gly32 is hidden, as in our cryo-EM density map of PhnG₂H₂I₂J₂ (**Figure A2.7A**). On the bound side, the α helices 5 and 6 of PhnJ move away from PhnH, disrupting the molecular interaction between PhnJ and PhnH to expose the active site residue Gly32 (**Figure A2.7B**). Further comparison with the crystal structure of PhnG₂H₂I₂J₂ is shown in **Figure A2.8**. On the apo side, the density of α helices 5 and 6 of PhnJ matches well with the crystal structure, while on the bound side, the density of α helices 5 and 6 of PhnJ is shifted relative to the crystal structure. This shift indicates there is a structural rearrangement in this region to reveal the active site underneath.

The glycy radical formed at Gly32 was proposed to abstract the hydrogen of Cys272 to form a cysteine radical during PhnJ catalysis (Kamat et al., 2013). In contrast, (Seweryn et al., 2015) made the startling discovery that Gly32 of PhnJ is more than 30 Å away from Cys272 in the crystal structure of the PhnG₂H₂I₂J₂ complex (**Figure A2.3C**). Here, in our cryo-EM structure of PhnG₂H₂I₂J₂ with PhnK bound, the overall structure of PhnJ remains similar to the crystal structure with Gly32 still ~30 Å away from Cys272. If the proposed reaction mechanism for C-P lyase, which involves the direct interaction between Cys272 and Gly32, is correct, then

substantial structural rearrangements must occur transiently and/or after the binding of substrates or other protein subunits to this complex.

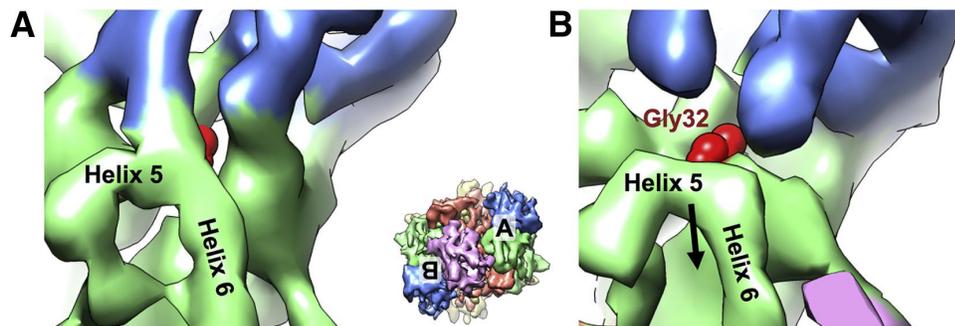


Figure A2. 7 Binding of PhnK Causes Gly32 of PhnJ to Be More Exposed

(A) On the side without PhnK bound (apo side), helix 5 of PhnJ connects PhnH. (B) On the side with PhnK bound (bound side), the densities for helix 5 and helix 6 of PhnJ move away from PhnH to expose the Gly32 of PhnJ. The density map of PhnG₂H₂I₂J₂K is rendered at 4.5 σ above the mean and colored coded as in Figure A2.3. The red sphere model denotes Gly32 of PhnJ. Inset shows the map of PhnG₂H₂I₂J₂K in the same orientation as Figure A2.3E with the locations of (A) and (B) labeled. See also Figure A2.8.

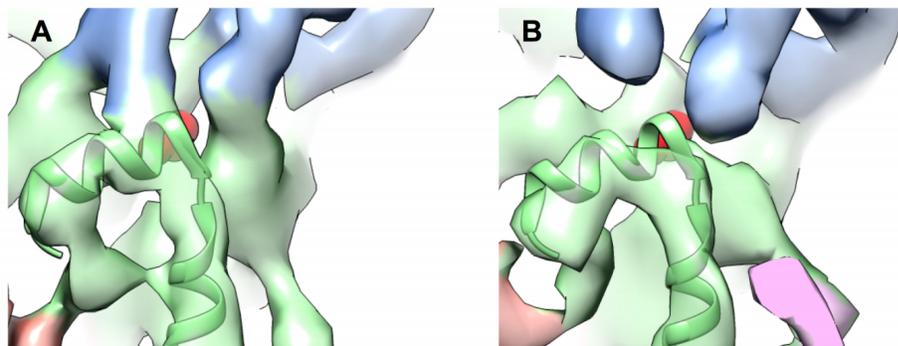


Figure A2. 8, related to Figure A2.7

Comparison between our density map of PhnG₂H₂I₂J₂K and α -helices 5 and 6 of PhnJ in the crystal structure (PDB id: 4XB6) for the PhnK apo side (A) and bound side (B), respectively. The color code and orientation are the same as Figure A2.7.

A2.4 Discussion

We have demonstrated the unique binding mode of the NBD-like PhnK to the C-P lyase core complex PhnG₂H₂I₂J₂. Even though there exist two identical binding sites for PhnK on the core complex, only one PhnK can bind at a time, leading to a more exposed active site residue Gly32 of PhnJ on the PhnK bound side. Such rearrangement around the active site may facilitate the delivery of the substrate or release of the reaction product.

PhnK has all the structural motifs to bind the ATP molecule. It might help recruit the ATP as the substrate for the C-P lyase reaction. While the two-fold symmetrical C-P lyase core complex PhnG₂H₂I₂J₂ potentially provides two equivalent sets of active sites for two series of the catalytic reactions, the loose and monomeric binding of PhnK suggest that it may alternate between the two binding sites on the core complex to deliver ATPs and facilitate the two series of reactions in a staggered time shift. Further biochemical and structural studies are needed to test these hypotheses.

The structural information provided here will help a better understanding of the enzymatic reaction mechanism for phosphonates utilization, especially the function of the NBD-like PhnK. Besides, our results clearly demonstrate the power of cryo-EM for the structural study of small (~200 kDa) asymmetrical multi-subunit protein complexes with compositional heterogeneity.

APPENDIX III

CUSTOMIZED AUTOMATIC TOOLS FOR 3D ELECTRON MICROSCOPY

To process the massive data in 3D Electron Microscopy (3DEM), many state-of-the-art software are freely available, such as EMAN2 (Tang et al., 2007) and Relion (Scheres, 2012). However, there may be substantial need for in-house programs or even small scripts with the following functionalities. First, the processing strategy should be customized or personalized for one's own application. Second, the entire data processing pipeline should be as automatic as possible, *ie.*, with minimal human intervene. Third, the processing results should be as good as possible, *eg.*, with novel ideas to process the data. It should serve as a summary of the best procedures of processing EM data to the best of one's knowledge. To satisfy the above need, I write and maintain a package, initially named Personalized Automatic Tools for 3D Electron Microscopy (PAT3DEM), that contains Python scripts and libraries covering many aspects of the data processing. To have the version control, I also regularly upload the codes to the Github website (<https://github.com/emkailu/pat3dem>).

The installation of this package is simple. First download the package from the above Github link. Then uncompress the downloaded file and goto the uncompressed folder using a terminal. Finally run `“./INSTALL.sh”`. The prerequisite is that the default Python is able to import the EMAN2 module.

The usage of this package is uniform and self explanatory. I will use two examples to explain it.

Example 1: p3starcoord.py

If you type this script in a terminal, a brief description of its functionality will show up as “Output the coordinates from star files. Origin offsets will be considered as float.” It also say “Please run 'p3starcoord.py -h' for detailed options.” Following this instruction, the detailed options will show up. For example, “-b BOX, --box BOX specify a box size (in pixel) for output, by default -1 (output .star only)” means that you can use either “-b” or “--box” to specify a box size, and if you do not specify anything, the default value, -1, will be used and no box files will be output. Do notice that this script will consider the offset values contained in the input star files. If you do not want to consider the offset values, you can use p3starscale.py to reset them as 0. If you do use it, it is suggested to run an additional script, p3starcoordcheck.py, to remove duplicate particles.

Example 2: p3rnarosettaem.py

This script is described as “Mimic MOSAICS-EM. Use Rosetta *de novo* RNA sampling to replace MOSAICS.” In such cases, if you found this script useful in your publications, you may consider to cite the MOSAICS-EM and Rosetta *de novo* RNA sampling papers (Das and Baker, 2007; Zhang et al., 2012), though there is no need to cite this script.