Structure Characterization of the 70S–BipA Complex Using Novel Methods of Single-Particle Cryo-Electron Microscopy

Danny Nam Ho

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

Structure Characterization of the 70S–BipA Complex Using Novel Methods of Single Particle Cryo-Electron Microscopy

Danny Nam Ho

Diseases caused by pathogenic bacteria continue to be major health concerns. For example, it is estimated that in the year 2000 typhoid fever caused over 21,000,000 illnesses and ~200,000 deaths (Crump et al., 2004). The disease is caused by *S. typhi,* a closely-related serotype of *S. typhiumurium*, the salmonella strain in which BipA was first identified. The CDC estimated that in 2013, multidrug resistant bacteria caused over 2 million infections in the United States, ending in more than 23,000 deaths (CDC, 2013). This number is set to rise as more bacteria become resilient to the collection of conventional antibiotics. The increasing number of multidrug resistant bacterial strains necessitates the development of new antimicrobial drugs.

The protein BipA is an attractive target for drug research. The *bipA* gene is ubiquitous in eubacteria and lower eukaryotes such as protozoa, but is absent from higher-order eukaryotes such as humans. Studies have shown that BipA participates in a variety of stress-related pathways and its expression is paramount to a bacteria's ability to adapt to adverse environment conditions. Because the protein is essential for bacterial survival, BipA presents a major vulnerability of pathogenic bacteria. Additionally, BipA's only known binding partner is the ribosome. An antibiotic targeting the protein itself or its interactions to the ribosome may disable only the bacteria, but have no effect on the eukaryotic host. A comprehensive model of BipA bound to the 70S ribosome will provide unparalleled insight into BipA's binding site and its mechanism.

Toward this goal, cryo-EM techniques were employed to visualize and characterize the binding site of BipA on the 70S ribosome. Over the last years, the introduction of new automated algorithms for particle selection (AutoPicker) and classification (RELION) for the cryo-EM technique has revolutionized the workflow of the entire imaging and reconstruction process. We have taken full advantage of these advancements to obtain the final resonstruction of the 70S–BipA complex.

An X-ray structure of isolated BipA–GMPPNP was elucidated, by collaborators, and used for further molecular modeling of the protein to reveal possible atomic interactions between BipA and 70S ribosome. Additional biochemical studies were performed to fully characterize the specific ribosomal complex that optimizes binding of the factor. Together, the cryo-EM reconstruction, the BipA X-ray structure, the subsequent molecular modeling, and the additional biochemical studies provide a comprehensive model for BipA binding.

References

CDC. 2013. Antimicrobial Resistance Threat Report 2013 cdc.gov.

Crump, J.A., Luby, S.P., Mintz, E.D. 2004. The global burden of typhoid fever, pp. 346 Bulletin of the World Health Organization, Vol. 82. World Health Organization.

TABLE OF CONTENTS

LIST OF GRAPHS, IMAGES, AND ILLUSTRATIONS

APPENDIX

ACKNOWLEDGEMENTS

The journey towards achieving to my Ph.D. would have been too arduous to endure without the support and love of many people in my life. Some provided me with emotional support, others educated and mentored me, and then there are those countless friends who kept me inebriated through the most trying times. It is difficult to reflect back on the exact contribution of each individual; however, I will attempt to thank a specific few special family members, friends, and colleagues.

I would like to thank my mother and father, who have always waited patiently for their son to achieve his dreams. It was a very difficult time in their lives when I left home to enter a graduate program 3000 miles away. However, they supported my decision, understanding that my dreams of achievement were, in fact, their own dreams when they left Vietnam over 20 years ago to live in a foreign, but better country. The achievements in my life would not have been possible without their sacrifices. They have never stopped providing me with endless support and love. To my sister, who has always been my guiding post and hero, I would like to thank you for holding me back whenever I was ready to give up and making me believe that I could finish this journey. It is without a doubt that I live my life by her example. She has been the perfect sister, mother, and friend.

I would like to thank my dissertation advisor, Dr. Joachim Frank. His mentoring and guidance over the past years have made me a better writer and a better scientist than I ever thought possible. Before entering the lab, I was content to simply watching the scientific community from afar, perhaps adding tidbits of insight to the vast base of knowledge already there. However, Joachim encouraged me to take on active role in shaping the future of the community and thus, I have been blessed to meet and discuss ideas with countless legends of our scientific community.

My dissertation work is a collaboration between me and every individual in the lab. Each colleague has helped me to develop new ideas, troubleshoot problematic experimental situations, and shape the scope of my project. I want to express my deepest gratitude to all of my labmates. I must take this time to single out Amy Jobe and Bingxin Shen. They are two beautiful and magnificent women that have become more than colleagues to me; they have become my family. They have been confidants when I have needed to release anger, they have been therapists when I have needed counsel, and they have been my sisters when I needed family.

My friends, both within and outside the Biological Sciences program, have kept me sane during the last six years. Finding my own path towards acceptance of who I was a scientist, an educator, and most importantly, a gay man, would have been impossible without the constant support of my friends. There have been many times that I have been disheartened by the tribulations in my life. My friends have always steered me back towards the path of achievement. I have developed deep and meaningful relationships with too many extraordinary people to single out any individual. Thus, I will collectively give thanks to everyone who has touched my life and shared invaluable wisdom on my journey to this point today.

DEDICATION

I dedicate this thesis to my sister, Chau. I do not know what type of man I would be without her steadfast encouragement, extraordinary wisdom, and unbounding love.

PREFACE

The following dissertation aims to trace the progress, discuss the methods, and present new discoveries of the 70S–BipA complex and the single-particle cryo-electron microscopy technique. Chapter 1 introduces the single-particle cryo-electron microscopy technique and new procedures for particle selection and classification. Chapter 2 will orient the reader in the translation process, particularly focusing on the elongation cycle. This chapter ends by introducing the translational GTPase BipA and its known physiological functions. The reconstruction of the 70S–BipA complex was obtained by using novel methods of image processing (AutoPicker) and unsupervised particle classification (RELION). Chapter 3 provides an in-depth comparison between the use of old and new image procesisng and classification techniques for the BipA dataset. Chapter 4 provides the structural and biochemical analysis of the 70S–BipA complex reconstruction. Chapter 5 introduces the burgeoning projects aimed at further characterizing the unique aspects of BipA, building on the unexpected and exciting discoveries presented in Chapter 4. Finally, the appendix provides copies of works in which I am co-author: 1) the published work on AutoPicker and 2) the latest draft of an article detailing the design of a new electron micrscopy image processing suite, Arachnid.

CHAPTER 1:

SINGLE-PARTICLE CRYO-ELECTRON MICROSCOPY

CHAPTER 1 ABBREVIATIONS

Abbreviation	Full Title
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
Å	angstrom
CCD	charge-coupled device
CMOS	complementary metal–oxide–semiconductor
cryo-EM	single-particle cryo-electron microscopy
CTF	contrast transfer function
DED	direct electron detector
DP	diffraction pattern
DQE	detective quantum efficiency
EF-G	elongation factor G
EM	electron microscopy
EMDB	Electron Microscopy Data Bank
FT	Fourier transform
KDa	kilodalton
MAP	maximum a posterior
MDa	megadalton
ml	milliliter
nM	nanomolar
pM	picomolar
PSF	point spread function
RNA	ribonucleic acid
S	svedberg
SNR	signal-to-noise ratio
SPR	single-particle reconstruction
TEM	transmission electron microscope
α	alpha
β	beta
µl	microliter
µM	micromolar
µm	micron or micrometer

1.1 INTRODUCTION

Biological pathways depend not on the work of a single protein or enzyme, but on multiple partners working in concert with one another. There is an increasing need to explore the functions and mechanisms of multicomponent complexes. Insight into a macromolecule's function may be gained by studying its physical structure. For structure determination of such molecules, two methodologies are regularly employed: nuclear magnetic resonance, X-ray crystallography, and transmission electron microscopy. The goal of these methods is to obtain a meaningful model of the macromolecule's structure.

Traditionally, X-ray crystallography has been the method used for obtaining atomic resolution structure of biological molecules such as isolated proteins and multicomponent complexes. X-ray crystallography has allowed unparalleled insights into the mechanisms of large macromolecular machines such as the 3.2 MDa prokaryotic 70S ribosome (Ban et al., 2000) and 2.6 MDa yeast fatty-acid synthase (Lomakin et al., 2007). However, as will be discussed in Sections 1.2.1 and 1.2.3, the X-ray crystallography technique becomes increasingly more difficult as the size of the investigated macromolecule increases.

Electron microscopy allows visualization of biological specimens spanning a huge range of physical dimensions, from single molecules to whole cells. On the one hand, imaging of whole cells using electron tomographic techniques has provided information on the organization and supra-structures of cellular organelles, albeit at resolutions far from providing atomic information. On the other hand, single-particle cryo-electron microscopy ("cryo-EM" for brevity) routinely provides 3D reconstructions of macromolecules above 300 kDa in weight. A major advantage of the cryo-EM method, as will be discussed in Sections 1.2.2 and 1.3.6, is the ability to capture a spectrum of different conformational and functional states of a dynamic macromolecule in a single experiment. These reconstructions are representations of free, isolated macromolecules unhindered by crystallographic packing constraints.

Unfortunately, the vast majority of cryo-EM reconstructions have resolutions worse than 3.5 Å, not allowing *de novo* atomic modeling. However, in cases of reconstructions resolved to resolutions between 3.5 and 10 Å, fitting of known X-ray structures into the map allows a pseudo-atomic model of the molecule to be produced. Thus, X-ray and cryo-EM information can be combined to give a meaningful interpretation of the lower-resolution reconstruction (Rossmann et al., 2005).

While this dissertation will focus primarily on reconstructions determined using cryo-EM techniques, X-ray crystallographic information is often leveraged to give insightful interpretation of the cryo-EM density maps, as will be seen in Chapters 2, 3, and 4. To provide full appreciation of the research presented, this chapter will give a brief review of X-ray crystallography and a more in-depth treatment of cryo-EM. The chapter is organized as follows: first, the X-ray crystallography method is presented in Section 1.2.1. Next, the cryo-EM technique is introduced in Section 1.2.2 to provide a foundation for the remainder of the chapter. As cryo-EM and X-ray crystallography are the most frequently used techniques for structure determination, a comparison of the two methods follows in Section 1.2.3. Finally, the cryo-EM technique is expanded upon in individual subsections of Section 1.3.

1.2 METHODS OF STRUCTURE DETERMINATION

1.2.1 X-ray Crystallography

The X-ray crystallography method is regularly employed to provide atomic resolution structures of proteins and biological macromolecules. In X-ray crystallography, structurally identical molecules are packed into an ordered three-dimensional crystal lattice. Contacts at the interface between the molecules allow the molecules to reach a low energy state and pack into a periodic arrangement, forming a three-dimensional crystal.

When the crystal is bombarded by X-rays, the resulting interactions of the X-ray beam with lattice molecules diffract the beam in a way that is characteristic for the atomic structure (Bragg, 1913). The diffracted beams are recorded as an array of spots called a diffraction pattern. Diffracting the X-ray beam may be thought of as performing a Fourier transform of the investigated molecule's atomic structure, and the diffraction spots are in fact the recorded intensities of each sinusoid component. The intensity of each diffraction spot is enhanced by the periodic repetition of the structural motifs in the crystal lattice. Thus, diffraction to high resolution requires well-formed crystals with structurally homogenous molecules. As will be seen in Section 1.2.2, the same fundamental laws govern the diffraction of X-ray beams and scattering of electrons in cryo-EM.

Each diffraction pattern essentially represents a 2D central slice in Fourier space of the 3D Fourier transform of the molecule. Recording the diffraction of the crystal as the latter is rotated on a goniometer at regular intervals yields different diffraction patterns, which can be used to fill the 3D Fourier space. Taking the inverse transform of the 3D Fourier map gives the electron density distribution of the molecule in real space. However, structure determination is not so straightforward.

X-ray structure determination requires both amplitude and phase data of the incident X-ray beam. Unfortunately, diffraction patterns only supply information about the amplitude, not phase of the diffracted X-rays that have interacted with the crystal. This is commonly referred to as the "phase problem" in X-ray crystallography. Additional crystallographic experiments, such as the use of molecular replacement, must be conducted to obtain the missing phase data (Wlodawer et al., 2008). The reward of such efforts is a high-resolution near-atomic structure, usually at 3.5 Å or higher, depending on the quality of the crystal.

Each X-ray structure is a representation of a molecule in a single state. Multiple X-ray structures must be obtained to visualize the ensemble of functional and conformational states for dynamic macromolecular machines. For example, the ribosome experiences intrinsic, large-scale rotations of the two subunits to each other and appears in different conformational states that cannot be observed with a single X-ray structure (Fischer et al., 2010). In most cases, multiple crystallographic experiments must be performed, which can prove to be difficult because crystallization is a laborious, multiparameter, optimization process.

Obtaining crystals is an essential step in X-ray crystallography, but continues to be a bottleneck for the application of the method. Regions of flexibility on the surfaces of molecules prevent the molecules from reaching a low energy state and thus, disfavor productive crystallization. Larger molecules and assemblies are structurally more complex, with increased probability of having such flexible regions. As a rule of thumb, the larger the investigated molecule or complex, the more difficult it may be to crystallize (Wery and Schevitz, 1997).

There are exceptional achievements, such as the X-ray structures of ribosomes, which ultimately reinforce the rule of thumb. X-ray ribosomal structures were obtained only after decades of research and effort. The first X-ray structures of the 30S small subunit at 3 Å (Wimberly et al., 2000), 50S large subunit at 2.4 Å (Ban et al., 2000), and the complete 70S ribosome, initially at 7.8 Å in 1999 (Cate et al., 1999) and then 5.5 Å in 2001 (Yusupov et al., 2001), were a boon for crystallographers. Few multicomponent macromolecule as complex as the ribosome has been deposited in the Protein Data Bank (PDB). Another notable achievement in X-ray crystallography was the struture of RNA Polymerase II, solved by Kornberg and coworkers in 2001 (Cramer et al., 2001; Gnatt et al., 2001). Extensive reviews on the principles and protocols for X-ray crystallography can be found in (Drenth, 2007; Ladd and Palmer, 2013).

1.2.2 Single Particle Cryo-Electron Microscopy (Cryo-EM)

Cryo-EM has evolved into a powerful method for structure determination of macromolecular complexes (Frank, 2013). Cryo-EM is an abbreviated term to describe an ensemble of techniques involving sample preparation, imaging in the transmission electron microscope, processing of the data, reconstruction of the 3D object using the 2D experimental projection data, and interpretation of the density map obtained. For recent examples of this workflow, see (Hashem et al., 2013; Li et al., 2013a; Allegretti et al., 2014). This section provides a foundation for understanding the importance and relevance of the method while Section 1.3 expands upon each step in the process.

In cryo-EM, the investigated macromolecule is assumed to exist in multiple, isolated, structurally identical copies ("particles") in the sample. After a small aliquot of the sample is applied to the EM grid, the plunge-freezing technique (McDowall et al., 1983) (Section 1.3.1) is used to immobilize the particles in a thin layer of vitreous ice. The frozen-hydrated grid is then transferred to the transmission electron microscope (TEM) for data collection (Sections 1.3.2-1.3.4). Two-dimensional (2D) projection images of the specimen field, called electron micrographs, are recorded, as shown in Figure 1.1. Each micrograph can be treated as a projection of the specimen field's 3D Coulomb potential distribution onto a plane that is normal to the electron beam. By extension, each 2D particle image captured in the micrograph can be treated as a 2D projection of the macromolecule's Coulomb potential distribution. Each micrograph captures hundreds of projections of particles lying in unknown orientations. In terms of the Fourier transform, the micrograph images implicitly contain both phase and amplitude data. Thus, the phase problem of X-ray crystallography does not exist in electron microscopy. In fact, phase data originating from cryo-EM reconstructions of the ribosome have been used to provide phase information for crystallographic studies (Ban et al., 1998; Stuart and Abrescia, 2013).

The projection theorem (see Frank, 2006) states that the Fourier transform of each 2D projection image is a central slice in the 3D transform of the macromolecule. Thus, the 3D Fourier space of the macromolecule can be "filled up" using the 2D Fourier transformations of the experimental data if the projection angle of each particle image can be determined. One can see immediate parallels between the scattering of electrons and the diffraction of X-rays discussed in the previous section. Indeed, the same physical laws of scattering govern both phenomena (Williams and Carter, 2009d). The main difference is that an image is formed in the case of EM and the Fourier information is derived from such images, whereas in X-ray crystallography the Fourier information is derived from diffraction. As discussed below, the EM projection images collected are distorted representations of the true object due to the aberrations of the TEM imaging system, and correction for these aberrations is required before one can obtain a faithful model of the real object.

The single-particle reconstruction (SPR) technique, see (Frank, 2009), schematically outlined in Figure 1.2 and elaborated below in Sections 1.3.5-1.3.7, aims to solve the reconstruction problem: how to obtain a faithful 3D representation of the investigated molecule's Coulomb distribution from the experimental 2D projection data. The technique takes advantage of a large collection of EM projections of structurally identical particles, laying in random orientations. The SPR technique involves two main stages: determination of the projection image parameters, and 3D reconstruction of the desired macromolecule. In the first stage, parameters such as the defocus value of the micrograph must be determined for later image correction. A faithful reconstruction of the original macromolecule requires image correction of the experimental data for the aberrations of the electron microscope imaging system, to be discussed below. Individual particle images are isolated and their respective projection angles are determined for alignment of the data and reconstruction of the final density map.

The SPR technique implicitly assumes that multiple projection images originate from structurally identical particles. In X-ray crystallography, structural homogeneity among molecules is selected for intrinsically during the crystallization process, as discussed in Section 1.2.1. However, in cryo-EM, it is difficult to find biochemical means of ensuring complete structural homogeneity among free, isolated particles. Heterogeneity exists in two main forms: compositional and conformational. Regarding the former, variation in ligand occupancies can result in complexes of differing composition (Gao et al., 2004). Regarding the latter, the particle in its native state can thermodynamically assume a large range of conformational states (Frank and Gonzalez, 2010). Separation of conformers is essential because structural details are diminished if data from structurally different particles are merged together. Several classification techniques, to be discussed in Section 1.3.6 and Chapter 3, have been developed to separate different conformers into smaller but more homogeneous subsets, thus allowing an inventory of molecules in different functional states to be reconstructed from a single dataset. Originally considered a major obstacle for cryo-EM, structural heterogeneity can now be seen as an advantage of the SPR method because an entire spectrum of functional states can be captured and inventoried in a single experiment (Frank, 2013). Impressive studies with large datasets have successfully produced inventories for studying the dynamics of the tRNA with the 70S ribosome (Fischer et al., 2010) or the biogenesis of the 30S small ribosomal subunit (Mulder et al., 2010).

In recent years, the cryo-EM methodology has benefitted from a surge of new technologies paired with novel automated algorithms for single-particle reconstruction (Veesler et al., 2013; Langlois et al., 2014a). Taking advantage of these advancements has resulted in impressive reconstructions of ever-increasing resolutions, shown in Table 1.1. Of note is the recent cryo-EM reconstruction of the yeast mitochondrial ribosomal large subunit, resolved to ~3.2 Å. The resolution was high enough to enable *ab initio* atomic modeling of the rRNA and proteins (Amunts et al., 2014). The cryo-EM method has quickly become a popular choice for structure determination of macromolecules, especially those that have resisted crystallization. As of December 31, 2013, single particle reconstructions make up 1702 (78.1%) of the 2179 depositions in the Electron Microscopy Data Bank (EMDB), shown in Figure 1.3. A deeper treatment of the EM field can be found in the texts of (Frank, 2006), (Reimer and Kohl, 2008), and (Williams and Carter, 2009c).

1.2.3 Cryo-EM Versus X-ray Crystallography

A comparison of the two methodologies, Cryo-EM and X-ray Crystallography, is important for understanding the advantages and limitations of both techniques. In X-ray crystallography, as discussed in Section 1.2.1, intensities of diffraction pattern spots are amplified by the redundancy of structural motifs in the crystal lattice. The ordered lattice essentially boosts the "signal" from each atom in the molecule. The recorded diffraction spots are easily measured and discernible from the surrounding background. In contrast, cryo-EM images of biological samples have a low signal-to-noise ratio (SNR) due to high intrinsic levels of background noise. Some of the background noise, as will be discussed in Section 1.3.2, originates from inelastically scattered electrons. Most of the noise, however, is a result of the necessity to use a low electron dose to prevent extensive radiation damage to biological macromolecules. To solve the problem of low SNR, multiple particle images of the same projection view are averaged together: the real signal is boosted while the stochastic noise is reduced.

Sample quantity requirements differ dramatically between cryo-EM and X-ray crystallographic experiments. The native cellular concentrations of different biological molecules vary widely depending on the identity of the molecule of interest. Genetic and biochemical means of overexpressing and purifying a molecule may be laborious, or in some cases, impossible to carry out. For the preparation of cryo-specimens, microliters (µl) of samples at nanomolar (nM) concentrations are needed. In contrast, crystallization of the investigated molecule usually requires micromolar (µM) concentrations to supersaturate the solution for productive crystal formation. Also, because multiple sample conditions and parameters must be screened for optimal crystallization, an even larger amount of sample is often needed. For example, the typical concentration of ribosomal components needed for a cryo-EM experiment is in the ~30-50 nM range (Amunts et al., 2014), while crystallographic experiments report sample concentrations in the µM ranges (Khatter et al., 2014). Thus, cryo-EM appears as an attractive method to study proteins or macromolecules that are difficult to express in sufficient quantities for crystallization attempts.

Save for density maps of viruses and for recent achievements in the EM field, most cryo-EM reconstructions have not been resolved to the atomic resolution typical for X-ray structures. *Ab initio* atomic modeling requires resolutions of 3.5 Å or better (Zhou, 2011). To date, of the 1702 single-particle reconstruction depositions in the EMDB, only 18 (1.1%) have resolved to resolutions of 3.5 Å or better. In contrast, 436 single-particle reconstructions (25.6%) have intermediate resolutions between 5-10 Å. In this intermediate range, secondary structures such as α-helices and β-sheets, and the spatial arrangement of domains in the macromolecule can be visualized. Fitting of a known X-ray structure into the intermediate-resolution density map, which is the subject of Section 1.3.7, allows meaningful interpretation of the reconstruction (Trabuco et al., 2008).

1.3 CRYO-EM

While Sections 1.2.2 and 1.2.3 provided a foundation for appreciating the cryo-EM method, this section gives an in-depth treatment of each individual step of the method. A review of the mathematical concepts underlying SPR methods in cryo-EM are beyond the scope of this dissertation, but treatments can be found in the texts of (Frank, 2006) and (Reimer and Kohl, 2008). Instead, a conceptual and practical review of each step is provided here.

1.3.1. Sample Preparation

The ideal cryo-EM sample preparation method must protect the biological samples against radiation damage by the electron beam in the TEM. The plunge-freezing technique, see (Grassucci et al., 2007), fulfills this stipulation and has become the standard in preparing cryo-EM specimens. A holey-carbon EM grid, usually made of copper or gold, acts as the scaffold on which the sample is applied. Before sample application, a thin layer of amorphous carbon is placed onto the EM grid (Grassucci et al., 2007). The amorphous carbon assists in the determination of the contrast transfer function, to be discussed later in Sections 1.3.3 and 1.3.6. A droplet of the sample is then applied to the grid. The particles, free-floating in the solvent, are able to sample the full spectrum of orientations and conformational states.

The specimen grid is held at the tip of tweezers in a plunge freezer, shown schematically in Figure 1.4. Here, excess liquid must be blotted off, leaving a thin layer of the aqueous sample suitable for rapid freezing. In practice, aqueous layers of up to .2 µm (200 nm) allow rapid vitrification. Specimen thicknesses greater than 1 µm do not vitrify rapidly, resulting in crystalline ice formation and damage to the native specimen structure (Dobro et al., 2010). The grid is then rapidly plunged into a pool of liquid ethane or propane, vitrifying the aqueous solution into a glass-like vitreous ice

(McDowall et al., 1983), and immobilizing the particles.

Vitreous ice provides a variety of benefits. In this amorphous ice, the water molecules have not crystallized into a lattice, and thus, the particles' structures remain intact. At cryogenic temperatures, the vitreous ice protects the sample and reduces the amount of radiation damage incurred by the electron beam in the microscope (Grassucci et al., 2008). The sample is also preserved in a hydrated environment resembling native cellular conditions. The frozen-hydrated specimen grid is then transferred to the TEM for imaging.

1.3.2. The Electron Microscope and Image Formation

A schematic of a typical transmission electron microscope is shown in Figure 1.5 and in-depth protocols are found in (Grassucci et al., 2008). In the TEM, electrons are initially generated from an electron source such as a tungsten crystal or wire filament (Rose, 2008). A potential difference set up between the tip and a grounded anode determines the acceleration voltage. The best acceleration voltage to be used depends on the nature of the specimen. Increasing voltages will imbue electrons with greater penetrating power but reduce the number of scattering events that will occur between the electron beam and the specimen (Kanaya and Okayama, 1972; Baker et al., 2010). Structural information about the specimen is encoded in the distribution and phases of the scattered electrons, and thus, a reduction in scattering events also means less useful information is obtained. Typical voltages used to image biological specimens range from 80 keV to 300 keV.

The accelerated electrons are steered with magnetic deflectors and focused into a coherent parallel beam through a series of electro-magnetic lenses, called condensors. While optical lenses in light microscopes are made of glass, a lens in the electron microscope is a magnetic field, produced by running current through coils of conducting wire.

At the subatomic level, matter is mostly empty, and thus the majority of the electrons in the focused beam pass through the specimen unscattered (Reimer and Kohl, 2008). Interactions that occur between the incident electrons with the specimen result in two principal forms of scattering: elastic and inelastic. In the former, the path of an incident electron passing close to an atomic nucleus is changed due to its attraction to the positive potential of the nucleus. In this interaction, shown in Figure 1.6, the elastically scattered electrons exit the specimen with a change in phase from the incident phase but without change in kinetic energy (Williams and Carter, 2009b). Structural information about the particles is encoded within these elastically scattered electrons.

Interaction between incident electrons and the electrons in atomic orbitals result in inelastic scattering. In this type of scattering, kinetic energy is transferred from the incident electron to the atomic orbital electrons of the specimen (Williams and Carter, 2009a). The resulting inelastically scattered electrons leave the specimen at random angles with an overall loss of energy. Inelastic scattering ionizes the electron cloud, acts as a source of incoherent signal, and induces free radicals that scatter and break chemical bonds in the specimen (Baker and Rubinstein, 2010). This is the major source of the radiation damage to the specimen as well as a source of background noise in the electron micrograph. Inelastic scattering is diffuse and provides little structural information about the specimen.

The electron beam leaving the specimen is focused by the objective lens to produce a first image, which is subsequently magnified by a set of projection lenses. Image contrast in the final micrograph originates from both amplitude and phase contrast. Amplitude contrast is generated when the objective aperture, placed at the objective lens, prevents high-angle scattered electrons from proceeding to the image detection system. Elastically scattered electrons contain meaningful structural information about the specimen (Wade, 1992). It is the interference of the phase-shifted elastically scattered electrons with the unscattered electrons that generates phase contrast, which is the main contribution to the image contrast in the final micrograph. The pattern of phase shifts in the back focal plane can be changed by defocusing the microscope with the objective lens (Frank, 2006).

1.3.3. The Aberrated Microscope

Electromagnetic lenses experience many of the same defects as optical lenses. Given a perfect point as the input object in an imaging system, the final image recorded will be an approximation of the original object: it will convoluted by a function, called the point spread function (PSF), whose shape depends on the aberrations of the imaging system. Among the defects in the TEM are spherical aberrations, chromatic aberrations, and axial astigmatism (Penczek, 2010). The Fourier transform of the PSF is the contrast transfer function (CTF), whose terms include the spherical aberration, axial astigmatism, and defocus value (Wade and Frank, 1977; Penczek et al., 1997; Zhu et al., 1997). The signature of the CTF can be seen in the absolute-squared Fourier transform of the micrograph, called the computed power spectrum (Frank, 2006). The white Thon rings, shown in Figure 1.7, in the power spectrum originate from the scattering of the amorphous carbon film that was laid onto the EM grid during sample preparation (Section 1.3.1).

In practice, computing the power spectrum from the entire micrograph in a single transform operation gives a poor estimate for the CTF. Instead, computing the power spectrum of smaller, partially overlapping subregions of the micrograph and averaging them produces an overall cleaner, less noisy power spectrum (Frank, 2006). Assuming there is no astigmatism in the image, one can take the 1D azimuthal average of the 2D power spectrum to obtain the 1D profile of the CTF. Accurately determining the CTF, discussed in Section 1.3.5, will allow correction of the data to obtain a faithful reconstruction of the original object. A mathematical treatment of the PSF and CTF are beyond the scope of this dissertation, though their introduction here is important to understand how the data will be processed, once digitized.

1.3.4. Image Detectors for the Electron Microscope

The choice of image detector is very important, as each recording device has its own advantages and disadvantages. The experimentalist must match the resolution goals of the experiment with the correct detector for maximum performance and efficiency. In evaluating the performance of a detector, one parameter often used is the detective quantum efficiency (DQE) (Vulovic et al., 2010), defined as the ratio of the SNR output squared to the SNR input squared. The parameter can be plotted as the percentage of the incident electrons counted as a function of different spatial frequencies and shows how well a given detector can detect an electron event (Booth et al., 2006). A perfect device would have a DQE of 1, meaning that every electron event is detected and recorded. Three major types of devices have been used for data collection: photographic emulsion film, charge-coupled devices (CCD), and direct electron detectors (DED). DQE curves comparing CCD and DD cameras are shown in Figure 1.8.

Photographic emulsion film consists of silver halide grains suspended in a gel matrix. Electrons striking the film cause ionization of the halide grains, transforming them into metallic silver. Photographic emulsion film, such as the Kodak SO-163 with a grain size of 5 µm, allows the collection of a large specimen field and has a higher DQE performance than CCDs at high spatial frequencies. Until recently, the use of film was preferred for high-resolution projects. However, the great disadvantage is the required dark room development and digitization before assessment of the data and image processing can take place (Cheng and Walz, 2009).

With scintillator-coupled charge-coupled devices (CCDs), an indirect method of electron detection is used. In this system, incident electrons first collide with a scintillator, which converts the electron signal into a photon signal. The photons emitted are transported by a dense bundle of fiber optics to a CCD array consisting of potential wells where the light signal is converted into an electrical signal. The electrical charge is accumulated in the wells before being read out line by line and recorded as a digital micrograph (Brink and Chiu, 1994). The signal conversion process results in a loss of resolution, and thus the CCD has a lower DQE performance than film (Booth et al., 2006). An additional problem is "blooming", which occurs when a charge fills up a CCD well and bleeds out into the surrounding wells. Thus, the readout from a single pixel may be due to the contributions from multiple surrounding CCD wells. The great advantage of the CCD camera, as a newly introduced completely digital medium, is the convenience of use (Sander et al., 2005). The digital data are immediately available for quality assessment of the specimen field, image processing, and reconstruction techniques.

Within the last two years, the introduction of the direct electron detection (DED) device, also a completely digital medium, has revolutionized electron microscopy imaging strategies (Veesler et al., 2013). With DED devices, a CMOS active-pixel sensor is exposed directly to the electron beam, removing the need for a scintillator fiber optics system. Thus, the resolution loss associated with the signal conversion process in CCD cameras is avoided. The DED sensor has the ability to detect and map each electron strike event, instead of integrating an accumulated charge (Milazzo et al., 2011). The DED device has an improved DQE performance over those of both film and CCD cameras (Li et al., 2013b; Li et al., 2013c). Overall, the DED device offers convenience of automation, immediate access to the data, and has a high DQE performance across all spatial frequencies. Additionally, because of the increased readout speed on the CMOS chip, the exposure of each image can be broken up into dose-fractionated frames, essentially allowing for the correction of beam-induced movement and drift upon alignment of the frames (Li et al., 2013b).

1.3.5. Image Processing

Over the last few years, the SPR workflow has undergone significant improvements. These improvements have aimed to remove user intervention in the workflow, reduce subjectivity in choosing image processing parameters, and promote a more automated performance. Much of the work presented in this dissertation serves to show the quality of the new algorithms and techniques. In certain cases, elaborated upon in Chapter 3, both old and new methods were utilized to process the same data, and the results were compared to benchmark the quality of the new methodologies. An introduction to both traditional and recent workflows is required to appreciate the scope of the work in this dissertation.

Powerful software platforms such as SPIDER (Frank et al., 1981), EMAN2 (Tang et al., 2007), and FREALIGN (Grigorieff, 2007) provide users with procedures for analysis and processing of electron microscopy data. Arachnid (Langlois, R. E., Ho D. N., Frank, J., 2014. *In Preparation*), introduced below, was developed to streamline the entire process and introduce a comprehensive pipeline, shown in Figure 1.9. Before image analysis can take place, electron microscopy data must be digitized. Film data must be scanned by a densitometer, while CCD and DED data are already digital formats. If dose-fractionated frames are taken, frame alignment to correct for beam-induced drift must take place before additional image processing steps can follow.

Traditionally, the first step of the workflow is to estimate the defocus and determine the CTF of each micrograph for later correction of the images. In Arachnid, the procedure *ara-defocus* estimates the 2D power spectra from the image using Welch's periodogram method (Welch, 1967), then takes the azimuthal average to produce a 1D CTF profile, introduced earlier in Section 1.3.3. The procedure fits a model 1D CTF curve to this 1D profile to estimate the defocus of the micrograph. A viewing program, such as *ara-screen* in the Arachnid suite, allows the experimenter to view the computed powers spectra and assess the quality of the micrographs. Within the 2D power spectra one can easily see the imaging aberrations in the image such as drift or astigmatism, which deform the Thon rings, as shown in Figure 1.7. In this figure, only the micrograph in (a) would be kept because it has circular, full Thon Rings indicating minimal drift, while (b) and (c) would be discarded due to extensive thermal drift and astigmatism, respectively.

After the dataset has been manually assessed for high-quality micrographs, individual particle windows must be selected and cropped from each micrograph. In the SPIDER environment, the established procedure uses the template-matching algorithm *LFC-Pick* (Rath and Frank, 2004) to find a collection of candidate particle windows. The results are presented to the user in a graphics user interface (GUI) ranked in descending cross-correlation value to the given template reference. Traditionally, the user had to manually inspect all candidate particles to verify the selection, being careful to keep real particles and discard any non-particles, such as dust or crystalline ice. The process is prone to user subjectivity: individual particles have very low contrast on the background noise and choosing real particles with the human eye is problematic.

AutoPicker, a program using a new particle algorithm in the Arachnid suite, has been developed to make the particle-picking procedure completely automated (Langlois et al., 2014b). As *LFC-Pick*, AutoPicker uses a Gaussian blurred disc of a user-defined diameter as a template for finding candidate particles. While the user can still manually reject particles, the verification process is extremely robust to dust, noise, and other features termed "non-particles." An important feature in AutoPicker is its ability to reject noise windows. Rather than utilizing a user-specified number of windows or cross-correlation threshold, AutoPicker utilizes a simple Bayesian classifier to automatically find a cross-correlation threshold for each micrograph (Otsu, 1979). This allows AutoPicker to effectively eliminate noise windows. Results from the use of AutoPicker show that this automated selection method produces reconstructions resolved to the same resolutions as the previous goldstandard, the manual verification method (Langlois et al., 2011). The work in this dissertation uses both manual verification and automated selection methods and shows that the particle dataset chosen by the new AutoPicker procedure is comparable to, or better than, the results from the SPIDER procedure.

1.3.6. Classification of the Heterogeneous Dataset

An assumption of the SPR technique is that the dataset is comprised of structurally identical particles. However, as introduced earlier, unless chemical interventions are taken to ensure homogeneity of the sample, this is rarely the case (Frank, 2013). The use of particle images originating from heterogeneous structures for a single reconstruction results in a 3D density map that is a blend multiple underlying structures. The combination of heterogeneous structures blurs high-resolution details. Meaningful interpretation of the reconstruction is obviously problematic as the density map obtained may not be representative of a physiologically relevant conformation of the investigated particle. Heterogeneity is especially common for the ribosome, where often a multitude of functional states naturally co-exist (Agirrezabala et al., 2008; Zhang et al., 2008). Thus, it is imperative to separate the different structural classes before refinement of any single density map of interest. Disentangling the different compositional and conformational subpopulations allows the researcher to take an inventory of all the functional states that may exist in the sample (Scheres, 2010).

There are several methods to classify particles and improve the homogeneity of the dataset. In the *supervised classification* method (Shaikh et al., 2008), implemented in SPIDER, separation of particles into subsets is dependent on each particle's resemblance to one of two given 3D references. Difference in the cross-correlation values is graphed as a histogram, with the zero value at the center of the histogram, designating particles that are equally correlated to both references (Valle, et al. 2002). Unfortunately, non-particles and low-contrast particles may not have high correlation to either of the two references. The user has to determine the cross-correlation cutoff for subclasses. The technique is prone to bias, as reference models used always rely on *a priori* information (Frank, 2009). Further discussion of this method is provided in chapter 3.

Unsupervised classification techniques cluster the particles into subsets based on their intrinsic similarities to one another. ML3D (Scheres et al., 2007) and RELION (Scheres, 2012) are recent programs developed for unsupervised classification of particles selected from each electron micrograph, utilizes a maximum *a posteriori* (MAP), a regularized likelihood optimization algorithm, to separate particles representing different underlying structures. RELION requires the user to define the number of *K* classes, and acts to find the most probable reconstructions based on the observed data and available prior information. This program is highly robust and, in the case of the ribosome -- as will be shown in chapter 3 -- is sufficiently sensitive to separate out multiple different classes of ribosomes with different ligand and tRNA occupancies, as well as free subunits. Each class can be refined using an iterative refinement procedure to yield a final reconstruction. In this way, a spectrum of conformations may be inventoried from a single sample.

1.3.7 Interpretation of the Reconstruction

As discussed in Section 1.2, most cryo-EM reconstructions are not of sufficient resolution to allow *ab initio* atomic modeling. The typical density map, with intermediate resolutions between 5 and 10 Å, cannot resolve the molecular interactions between the components of an assembly (Bai et al., 2013). However, single-particle reconstructions are now routinely produced at subnanometer (< 10 Å) resolutions, with maps of asymmetric molecules within the last year approaching atomic-level resolution, defined at ~3 Å or better (Bai et al., 2013; Hashem et al., 2013). As discussed in Section 1.2.3., the fitting of X-ray structures into an intermediate-resolution cryo-EM density map leverages the atomic accuracy of an X-ray structure to produce a pseudo-atomic model of the reconstruction.

Two major fitting approaches have been utilized: rigid-body fitting and flexible fitting. Rigid-body fitting keeps the X-ray structure static and unperturbed as an exhaustive search over translational and rotational angles seek to place the X-ray structure into the cryo-EM map in a way that results in highest correlation between the density map and the X-ray structure. This rough positioning of the molecule is implemented in a variety of visualization programs such as VMD (Humphrey et al., 1996) and UCSF Chimera (Pettersen et al., 2004). Additionally, one can "fragment" the X-ray structure into its various domains, rigid-fit each domain separately, and perform a real-space refinement while reinforcing the integrity of the chemical bonds at the boundaries, for maximum correlation (Chapman, 1995; Gao et al, 2003). However, the segmentation can be highly subjective and may lead to mis-fitting and misinterpretation (Rossmann et al., 2001). A comprehensive review of the various rigid-body fitting methods and implementations is found in (Wriggers and Chacón, 2001).

Flexible fitting is available in a variety of different program suites including SITUS (Wriggers et al., 1999), RSRef (Chapman, 1995), and Molecular Dynamics Flexible Fitting (MDFF) (Trabuco et al., 2009). A review of these approaches is beyond the scope of this dissertation, and treatments can be found in (Sanbonmatsu, 2012) and (Ahmed et al., 2012). The technique used in this dissertation is the MDFF method implemented via the program VMD (Trabuco et al., 2011). In this technique, molecular dynamic (MD) simulations are utilized to drive the X-ray structure locally into the cryo-EM density. The density gradient of the reconstruction is translated into a new potential energy term in the MD simulations. This energy function acts as a steering force, driving the X-ray structure into high-density areas of the reconstruction. Additional harmonic restraints are provided to preserve the correct stereochemistry of the atoms and the secondary structures of the particle, preventing overfitting of the atomic structure due to the drive to occupy the density map (Mackerell et al., 1998). The result of MDFF is representative for a whole set of structures, each a pseudo-atomic model that is a compatible with the EM data. This is because density maps with resolutions between 5-10 Å lack sufficient information for the positioning of amino acid side chains, though secondary structures can be positioned accurately. Thus, except for the precise positions of the side chains, the pseudo-atomic model, integrating X-ray and cryo-EM data, is a plausible representation of the native, physiologically significant state of the particle, free from the errors of crystallization experiments. The method works well for reconstructions with subnanometer resolution. Lower-resolution maps, with large high-density areas, allow for atoms to erroneously sample a spectrum of positions, making interpretations of the model unreliable.

MDFF allows the X-ray structure conformational flexibility during the fitting process in order to occupy the reconstruction. In the case of EF-G, the translocase, the isolated X-ray structure (PDB: 1FNM) (Laurberg et al., 2000), is significantly different from the bound state of the factor, as revealed by cryo-EM (Valle et al., 2003). It was not until almost a decade later that an X-ray structure of EF-G bound to a 70S was solved (Gao et al., 2009). As shown in Figure 1.10, domain IV shifts approximately 27 Å towards the A site of the 30S small subunit upon binding to the ribosome (Li et al., 2011). In cases where X-ray structures of single components and a cryo-EM reconstruction of the entire assembly is known, such as the CRISPR RNA surveillance complex (Wiedenheft et al., 2011), one can fit individual proteins to provide a meaningful model of the entire assembly.

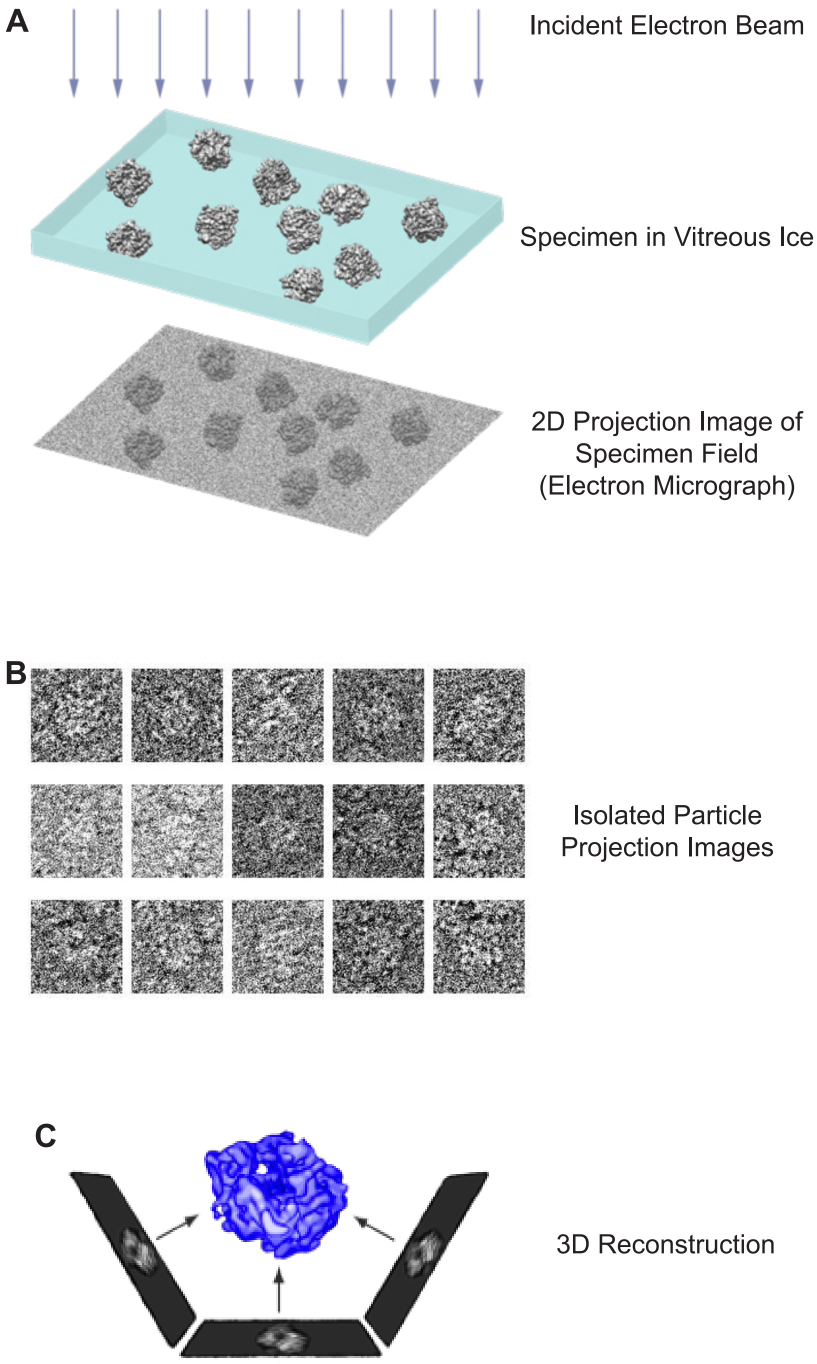

A
Incident Electron Beam
Specimen in Vitreous Ice
2D Projection Image of Specimen Field (Electron Micrograph)
B
Isolated Particle Projection Images
C
3D Reconstruction

Figure 1.1. Single particle reconstruction. (A) 2D projection images of the specimen field (electron micrographs) capture hundreds of particle projections, each assumed to be a structurally identical, differing only in their view orientation. (B) Each particle image is isolated from the micrograph, allowing a dataset of particles to be compiled. (C) Determination of each particle image's projection angle is necessary for 3D reconstruction. Using this information, methods such weighted back-projection or Fourier reconstruction techniques can be utilized to obtain a 3D reconstruction. Reproduced with permission from (Mitra and Frank, 2006)

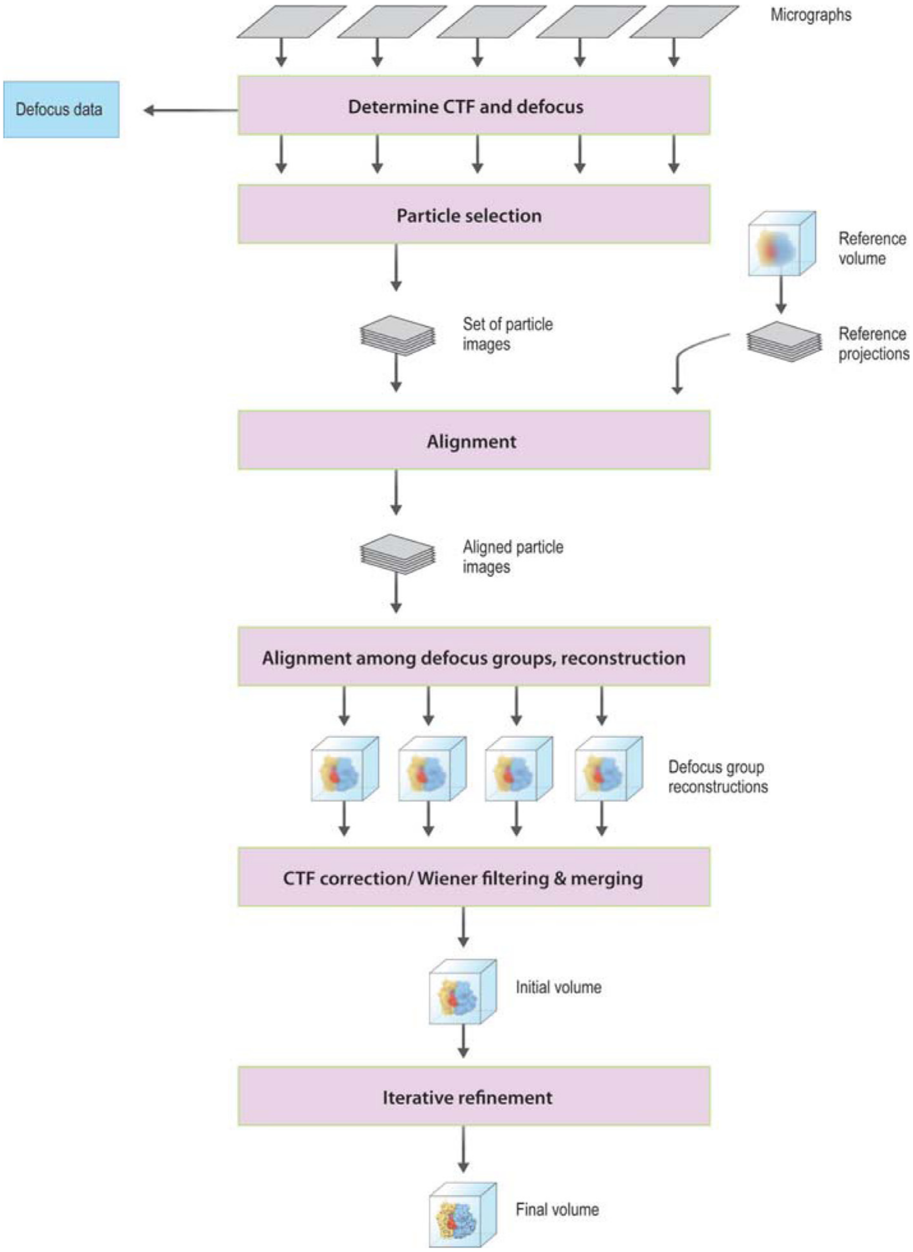


Figure 1.2. The traditional workflow of the single-particle reconstruction (SPR) technique. The workflow shown here assumes that micrographs have been 1) digitized and 2) reviewed for quality. Software suites such as SPIDER (Frank et al., 1981) are traditionally used to implement procedures such as automated particle selection. However, the candidate particles must be verified manually by the user to remove non-particles from the dataset before alignment of particle images can take place. Once aligned, each defocus group is reconstructed separately and corrected for the CTF (Wiener filtering). Following CTF correction, each defocus group reconstruction is merged to give an initial volume. Iterative refinement of the projection angles gives a final volume. Reproduced with permission from (Leith, 2012) (http://it.iucr.org/).

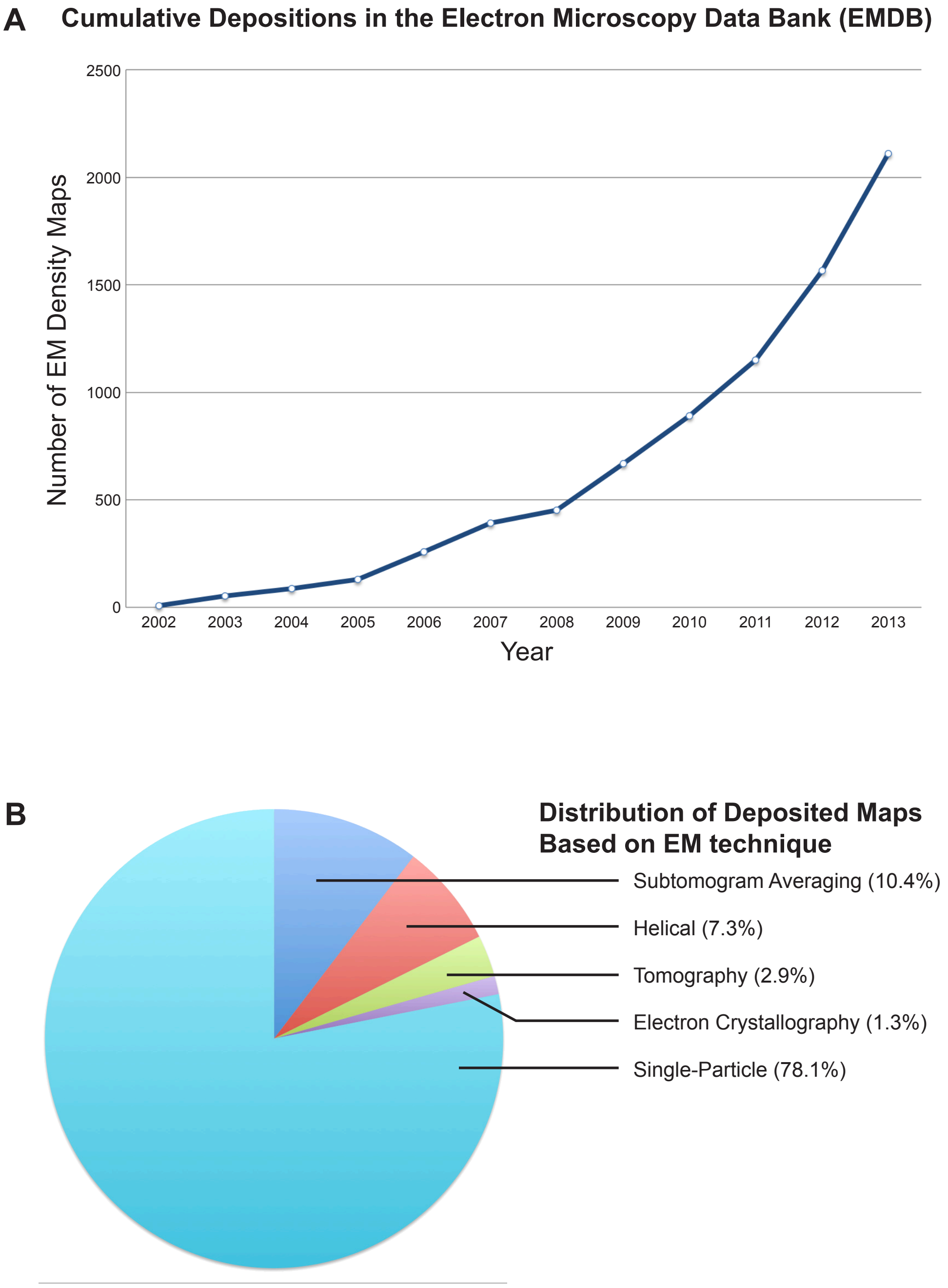


Figure 1.3. Current statistics for the Electron Microscopy Data Bank (EMDB). Over the last 10 years, the cumulative number of depositions in the EMDB had grown tremendously, as shown in (A). As of December 31, 2013, there are a total of 2111 map entries. Distribution of the released maps based on EM technique, in percentage of total depositions, is shown in (B), with each slice of the pie pertaining to a different technique. Statistics from EMDB (www.ebi.ac.uk/pdbe/emdb/statistics_main.html)

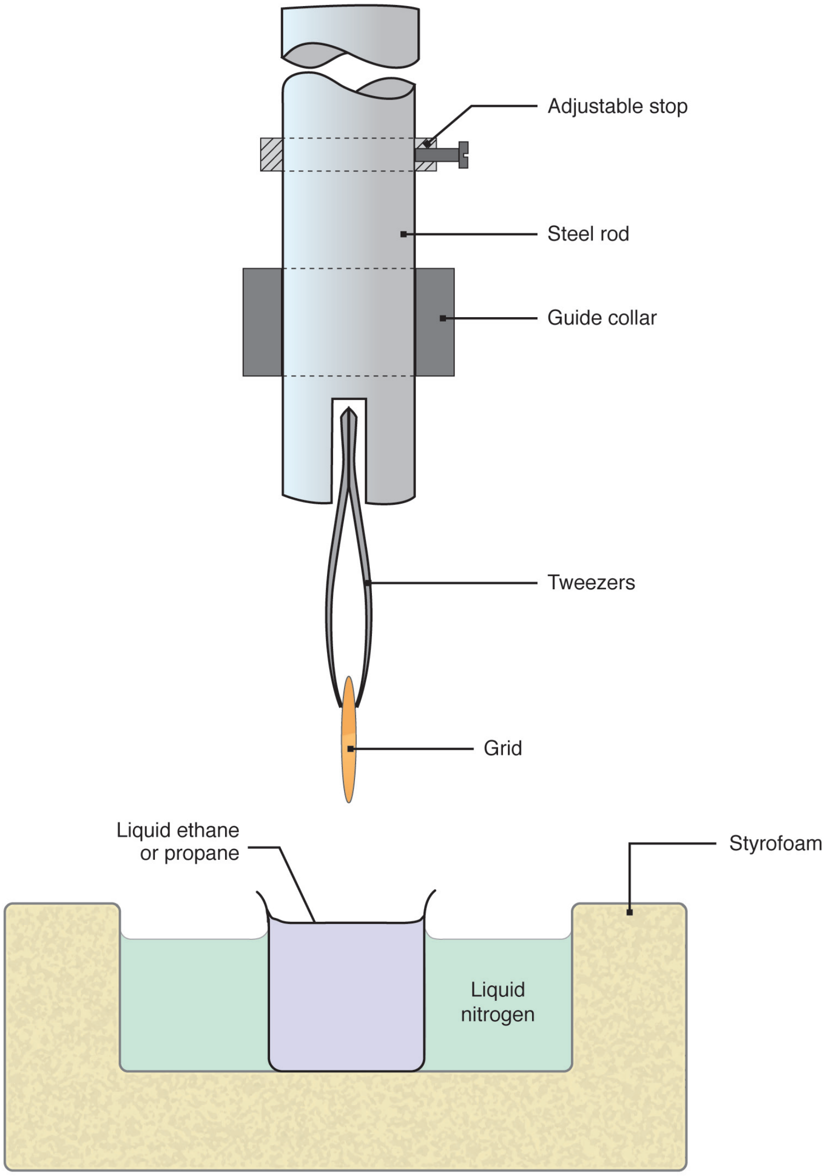


Figure 1.4. Schematic of a plunge-freezer. A small aqueous sample is applied to an EM grid, usually made of copper or gold. Excess liquid can be blotted off (not shown) to produce a thin layer of the sample. The grid, held by tweezers, is rapidly plunged into a pool of liquid ethane or propane, converting the aqueous solution into vitreous ice. The process immobilizes the biological sample in a frozen-hydrated environment that protects the sample from radiation as well as mimics natural cellular conditions. Reproduced with permission from (Frank, 2011).

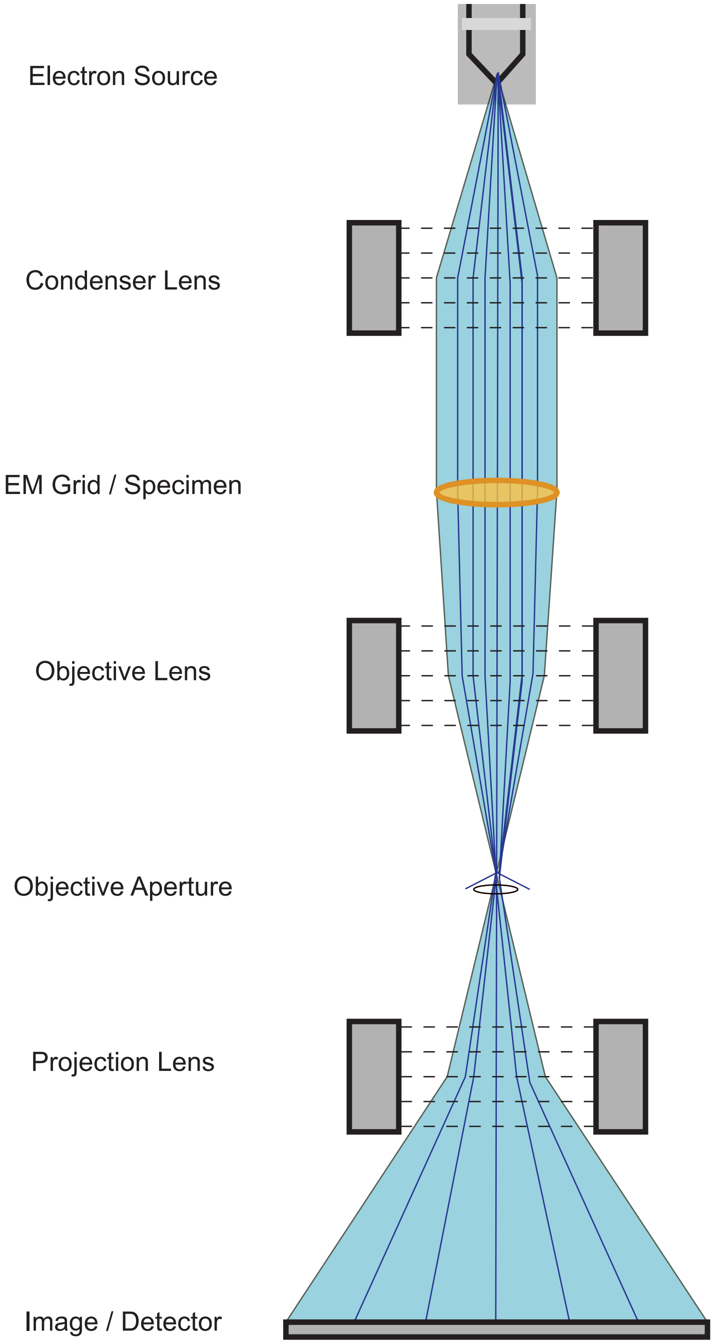


Figure 1.5. A schematic of a typical transmission electron microscope (TEM). An electron source at the top of the TEM generates a cloud of electrons, which are accelerated in the field between cathode and anode. Condenser lenses focus electrons into a coherent beam that travels down the microscope to interact with the specimen. The scattered electrons are focused by the objective lens into an initial image. Projection lenses magnify the image. The final image, called an electron micrograph, is recorded by a detector at the bottom of the TEM.

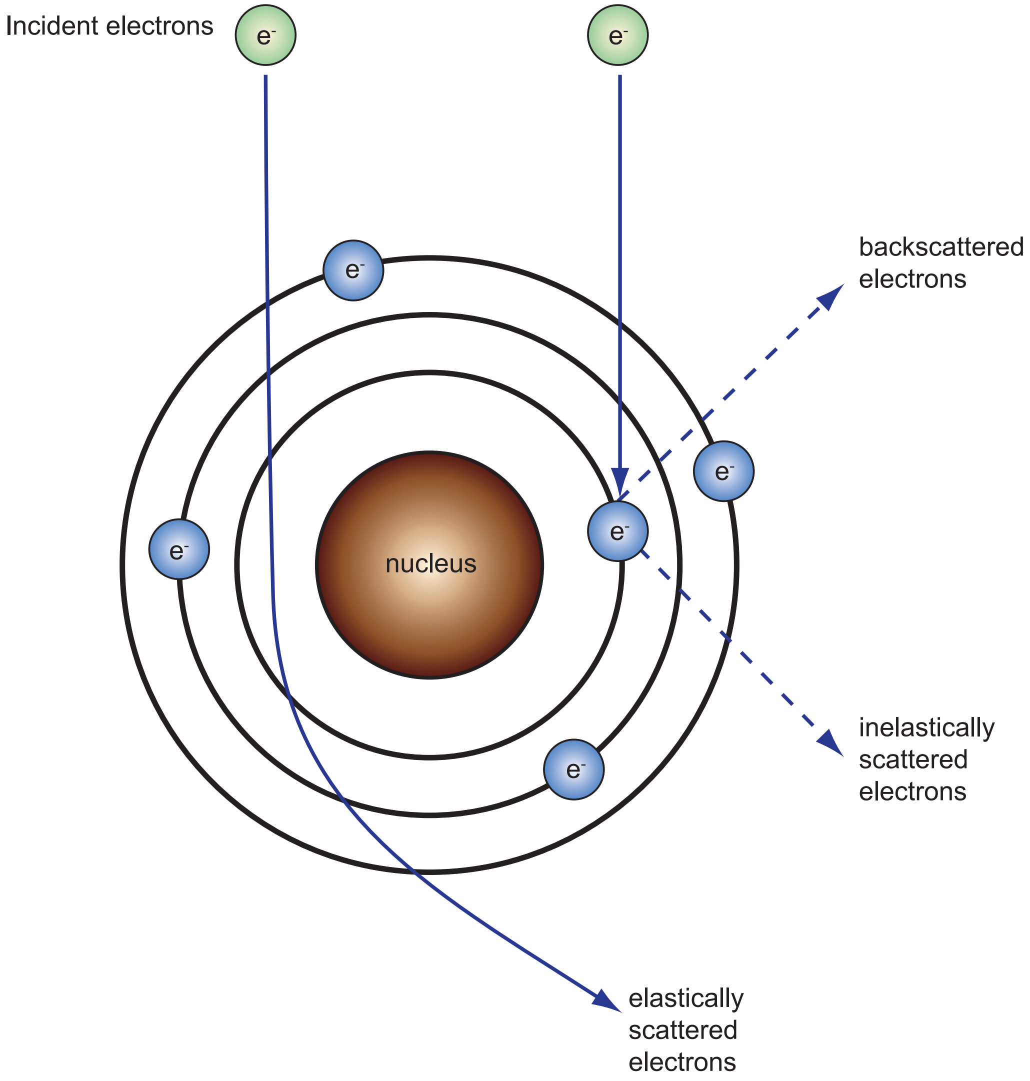

Incident electrons
e-
e-
backscattered electrons
e-
e-
e-
nucleus
e-
e-
inelastically scattered electrons
elastically scattered electrons

Figure 1.6. Elastic and inelastic scattering. The majority of incident electron pass through the specimen unscattered. Interactions between the incident electrons with a specimen's nucleus or orbital electrons results in two principal forms of scattering: elastic and inelastic, respectively. Elastically scattered electrons are deflected by the nucleus and leave the specimen without kinetic energy loss. Inelastically scattered electron interact with orbital electrons and ionize the electron cloud, causing breakage of chemical bonds. These electrons leave the specimen with a loss of energy.

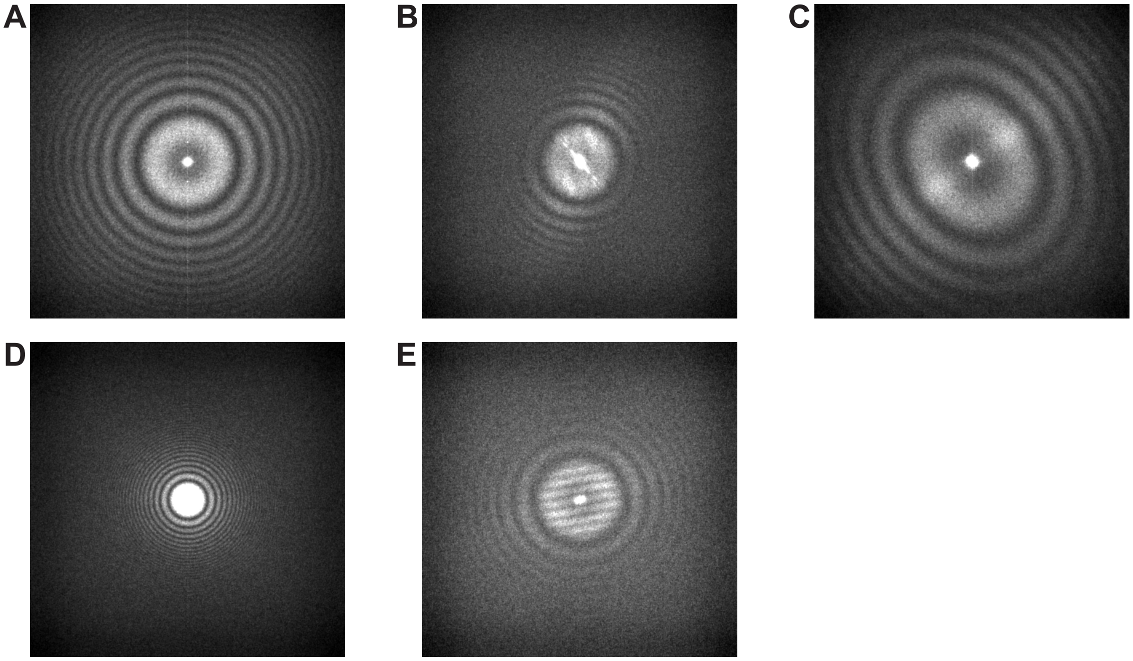


Figure 1.7. Computed power spectrum. The signature of the CTF can be seen in the Thon ring pattern of the computed power spectra of each micrograph image. Quality of the micrograph image can also be assessed at the same time. Good micrograph give rise to round, full Thon rings in (A). Patterns produced by continuous drift of the specimen stage and jump drift can be seen in (B) and (E), respectively. Aberrations of the microscope, such as axial astigmatism, can be seen in (C). Computed power spectra from severely underfocused micrographs have small, closely spaced Thon rings such as in (D).

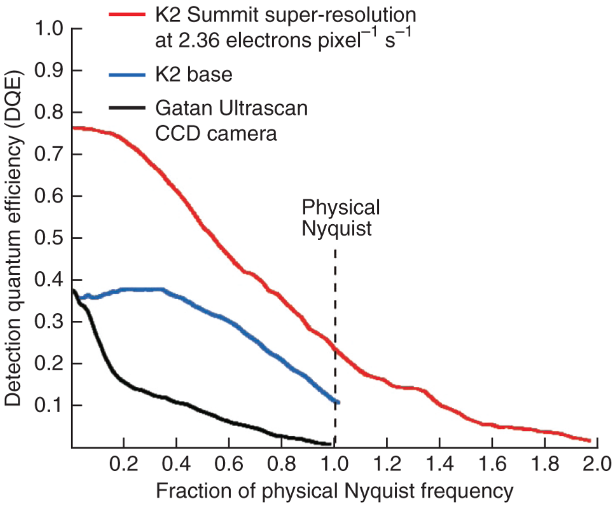


Figure 1.8. Comparison of the Detection Quantum Efficiency (DQE) of various digital detectors. The DQE is measured, for three different detectors, from 0 to 1.0 across fractions of the physical Nyquist frequency. The K2 base is an integrated camera, without the ability to count electrons and dose-fractionate the exposure. The K2 Summit, due to its ability to localize and count electron events, can produce images with subpixel accuracy in super-resolution mode. The Gatan Ultrascan CCD is a 4K x 4K integrated camera. Reproduced with permission from (Li et al, 2013).

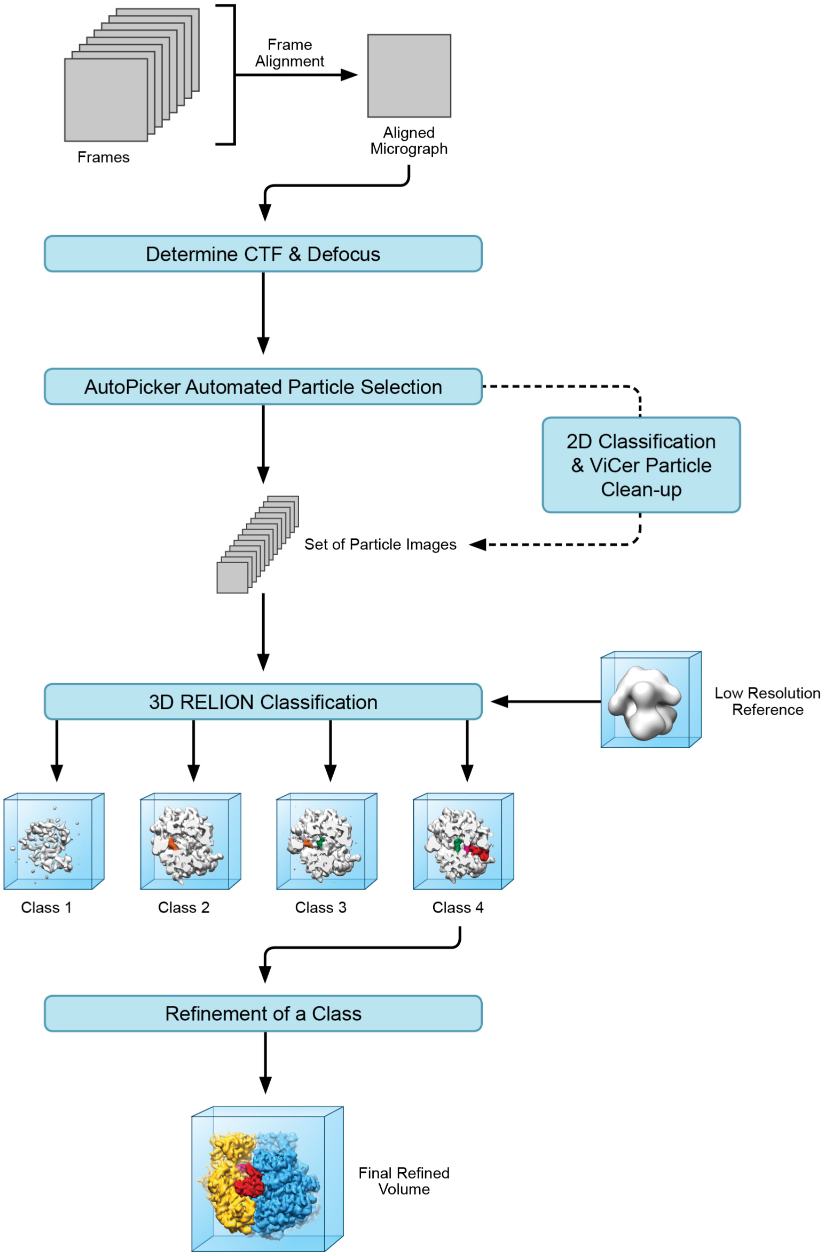


Figure 1.9. The Arachnid workflow. The Arachnid suite of procedures aims to streamline the image processing of electron micrograph data in preparation for classification techniques (Langlois, R. E., Ho D. N., Frank, J., 2014. *In Prep*). Like the traditional workflow in Figure 1.1, the first step is to determine the CTF and defocus. Quality assessment of the micrographs can be done concurrently with computing the power spectra. AutoPicker can be utilized to select particles, which can be cleaned via ViCer. The final set of particle images can be used for 3D unsupervised classification, here by RELION (Scheres, 2012). After classification, classes are visually inspected and refined.

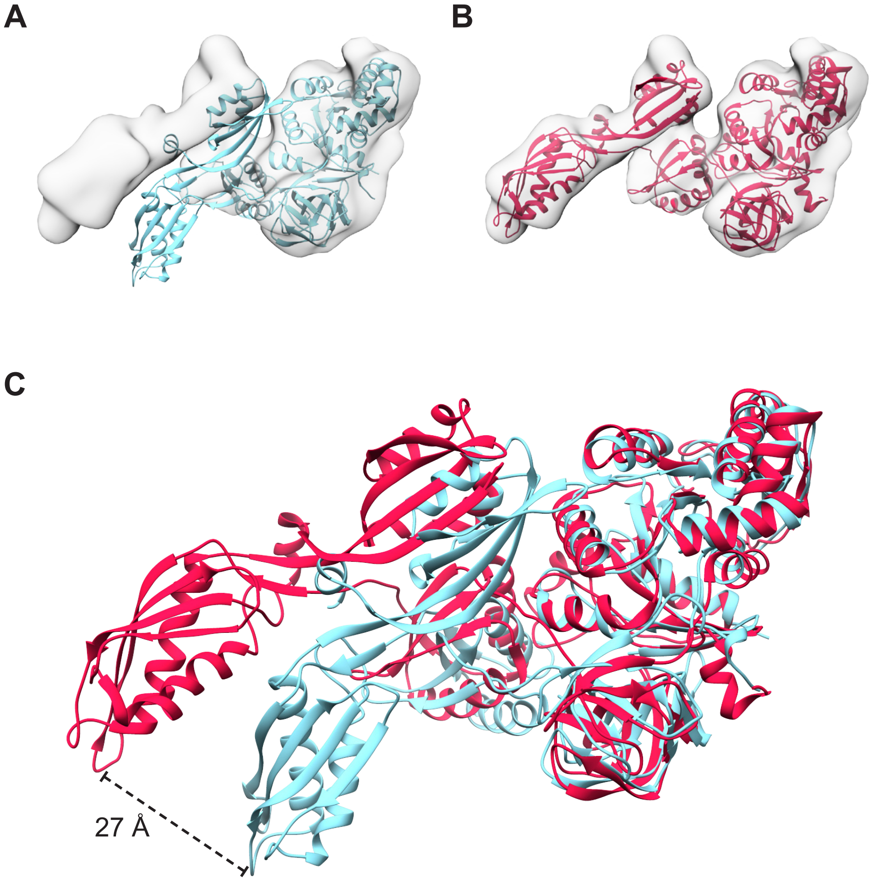


Figure 1.10. EF-G: isolated state versus the fitted state. In (A), the X-ray structure of EF-G in solution (PDB:1FNM) (Laurberg et al., 2000) is rigidly placed in a cryo-EM map of bound–EF-G resolved to 10.9 Å (Valle et al., 2003). After MDFF (Trabuco et al., 2008), domain IV displays a large 27 Å shift, shown in (C), to produce the fitted structure in (B) (Li et al., 2011).

REFERENCES

Agirrezabala, X., Lei, J., Brunelle, J.L., Ortiz-Meoz, R.F., Green, R., Frank, J., 2008. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. Mol. Cell 32, 190-197.

Ahmed, A., Whitford, P.C., Sanbonmatsu, K.Y., Tama, F., 2012. Consensus among flexible fitting approaches improves the interpretation of cryo-EM data. J. Struct. Biol. 177, 561-570.

Allegretti, M., Mills, D.J., McMullan, G., Kühlbrandt, W., Vonck, J., Sundquist, W. 2014. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector Elife, Vol. 3. eLife Sciences Publications Limited.

Amunts, A., Brown, A., Bai, X.-C., Llacer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H.W., Ramakrishnan, V., 2014. Structure of the yeast mitochondrial large ribosomal subunit. Science (New York, N.Y.) 343, 1485-1489.

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

Baker, L.A., Rubinstein, J.L., 2010. Radiation Damage in Electron Cryomicroscopy, p. 371-388, Cryo-EM, Part B: 3-D Reconstruction, Academic Press.

Baker, L.A., Smith, E.A., Bueler, S.A., Rubinstein, J.L., 2010. The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. J. Struct. Biol. 169, 431-437.

Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R.A., Sweet, R., Frank, J., Moore, P.B., Steitz, T.A., 1998. A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. Cell 93, 1105-1115.

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

Booth, C.R., Jakana, J., Chiu, W., 2006. Assessing the capabilities of a 4kx4k CCD camera for electron cryo-microscopy at 300kV. J. Struct. Biol. 156, 556-563.

Bragg, W.H., 1913. The Reflection of X-rays by Crystals. (II.). Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 89, 246-248.

Brink, J., Chiu, W., 1994. Applications of a slow-scan CCD camera in protein electron crystallography. J. Struct. Biol. 113, 23-34.

Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N., Noller, H.F., 1999. X-ray crystal structures of 70S ribosome functional complexes. Science (New York, N.Y.) 285, 2095-2104.

Chapman, M.S., 1995. Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function, p. 69-80, Acta Crystallogr A: Found Crystallogr.

Cheng, Y., Walz, T., 2009. The advent of near-atomic resolution in single-particle electron microscopy. Annu. Rev. Biochem. 78, 723-742.

Clemons, W.M., May, J.L.C., Wimberly, B.T., McCutcheon, J.P., Capel, M.S., Ramakrishnan, V., 1999. Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. Nature 400, 833-840.

Dobro, M.J., Melanson, L.A., Jensen, G.J., McDowall, A.W., 2010. Plunge Freezing for Electron Cryomicroscopy, p. 63-82, Cryo-EM, Part B: 3-D Reconstruction, Academic Press.

Drenth, J., 2007. Principles of Protein X-Ray Crystallography Springer.

Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., Stark, H., 2010. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature 466, 329-333.

Frank, J., Shimkin, B., Dowse, H., 1981. SPIDER—A modular software system for electron image processing. Ultramicroscopy 6, 343-357.

Frank, J., 2006. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State. Oxford University Press.

Frank, J., 2009. Single-particle reconstruction of biological macromolecules in electron microscopy – 30 years. Q. Rev. Biophys. 42, 139-158.

Frank, J., Gonzalez, R.L., 2010. Structure and dynamics of a processive Brownian motor: the translating ribosome. Annu. Rev. Biochem. 79, 381-412.

Frank, J., 2013. Story in a sample - the potential (and limitations) of cryo-electron microscopy applied to molecular machines. Biopolymers 99, 832-836.

Gao, H., Valle, M., Ehrenberg, M., Frank, J., 2004. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. J. Struct. Biol. 147, 283-290.

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

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., Kornberg, R.D., 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science 292, 1876-1882.

Grassucci, R.A., Taylor, D.J., Frank, J., 2007. Preparation of macromolecular complexes for cryo-electron microscopy. Nat Protoc 2, 3239-3246.

Grassucci, R.A., Taylor, D., Frank, J., 2008. Visualization of macromolecular complexes using cryo-electron microscopy with FEI Tecnai transmission electron microscopes. Nat Protoc 3, 330-339.

Grigorieff, N., 2007. FREALIGN: High-resolution refinement of single particle structures. J. Struct. Biol. 157, 117-125.

Hashem, Y., des Georges, A., Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C., Westhof, E., Madison-Antenucci, S., Frank, J., 2013. High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature 494, 385-389.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual Molecular Dynamics. J Mol Graph 14, 33-38- 27-38.

Kanaya, K., Okayama, S. 1972. Penetration and energy-loss theory of electrons in solid targets, pp. 43-58 J. Phys. D: Appl. Phys., Vol. 5. IOP Publishing.

Khatter, H., Myasnikov, A.G., Mastio, L., Billas, I.M.L., Birck, C., Stella, S., Klaholz, B.P. 2014. Purification, characterization and crystallization of the human 80S ribosome, pp. e49-e49 Nucleic Acids Res., Vol. 42.

Ladd, M., Palmer, R. 2013. Structure Determination by X-ray Crystallography. Springer US.

Kanaya, K., Okayama, S., 1972. Penetration and energy-loss theory of electrons in solid targets. J. Phys. D: Appl. Phys. 5, 43-58.

Khatter, H., Myasnikov, A.G., Mastio, L., Billas, I.M.L., Birck, C., Stella, S., Klaholz, B.P., 2014. Purification, characterization and crystallization of the human 80S ribosome. Nucleic Acids Res. 42, e49-e49.

Ladd, M., Palmer, R., 2013. Structure Determination by X-ray Crystallography Springer US.

Langlois, R., Pallesen, J., Frank, J., 2011. Reference-free particle selection enhanced with semi-supervised machine learning for cryo-electron microscopy. J. Struct. Biol. 175, 353-361.

Langlois, R., Ash, J.T., Pallesen, J., Frank, J., 2014a. Fully automated particle selection and verification in single-particle cryo-EM, p. 43-66-66, Applied and Numerical Harmonic Analysis, Springer New York.

Langlois, R., Pallesen, J., Ash, J.T., Nam Ho, D., Rubinstein, J.L., Frank, J., 2014b. Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. J. Struct. Biol. 186, 1-7.

Laurberg, M., Kristensen, O., Martemyanov, K., Gudkov, A.T., Nagaev, I., Hughes, D., Liljas, A., 2000. Structure of a mutant EF-G reveals domain III and possibly the fusidic acid binding site. J Mol Biol 303, 593-603.

Li, W., Trabuco, L.G., Schulten, K., Frank, J., 2011. Molecular dynamics of EF-G during translocation. Proteins 79, 1478-1486.

Li, W., Atkinson, G.C., Thakor, N.S., Allas, U., Lu, C.-c., Chan, K.-Y., Tenson, T., Schulten, K., Wilson, K.S., Hauryliuk, V., Frank, J., 2013a. Mechanism of tetracycline resistance by ribosomal protection protein Tet(O). Nature communications 4, 1477.

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

Li, X., Zheng, S.Q., Egami, K., Agard, D.A., Cheng, Y., 2013c. Influence of electron dose rate on electron counting images recorded with the K2 camera. J. Struct. Biol. 184, 251-260.

Lomakin, I.B., Xiong, Y., Steitz, T.A., 2007. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. Cell 129, 319-332.

Mackerell, J., Alexander D, Bashford, D., Bellott, M., Dunbrack, J., Roland L, Evanseck, J.D., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, I., W E, Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586-3616.

McDowall, A.W., Chang, J.J., Freeman, R., Lepault, J., Walter, C.A., Dubochet, J., 1983. Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. Journal of Microscopy 131, 1-9.

Milazzo, A.-C., Cheng, A., Moeller, A., Lyumkis, D., Jacovetty, E., Polukas, J., Ellisman, M.H., Xuong, N.-H., Carragher, B., Potter, C.S., 2011. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. J. Struct. Biol. 176, 404-408.

Mitra, K., Frank, J., 2006. Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. Annu Rev Biophys Biomol Struct 35, 299-317.

Mulder, A.M., Yoshioka, C., Beck, A.H., Bunner, A.E., Milligan, R.A., Potter, C.S., Carragher, B., Williamson, J.R., 2010. Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. Science (New York, N.Y.) 330, 673-677.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst., Man, Cybern. 9, 62-66.

Penczek, P.A., Zhu, J., Schröder, R., Frank, J., 1997. Three dimensional reconstruction with contrast transfer compensation from defocus series. Scanning Microsc 11, 147-154.

Penczek, P.A., 2010. Image Restoration in Cryo-Electron Microscopy, p. 35-72, Meth. Enzymol.

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

Rath, B.K., Frank, J., 2004. Fast automatic particle picking from cryo-electron micrographs using a locally normalized cross-correlation function: a case study. J. Struct. Biol. 145, 84-90.

Reimer, L., Kohl, H., 2008. Transmission Electron Microscopy: Physics of Image Formation Springer.

Rose, H.H., 2008. Optics of high-performance electron microscopes. Sci. Technol. Adv. Mater. 9, 014107.

Rossmann, M.G., Bernal, R., Pletnev, S.V., 2001. Combining electron microscopic with X-ray crystallographic structures. J. Struct. Biol. 136, 190-200.

Rossmann, M.G., Morais, M.C., Leiman, P.G., Zhang, W., 2005. Combining X-ray crystallography and electron microscopy. Structure 13, 355-362.

Sanbonmatsu, K.Y., 2012. Computational studies of molecular machines: the ribosome. Curr. Opin. Struct. Biol. 22, 168-174.

Sander, B., Golas, M.M., Stark, H., 2005. Advantages of CCD detectors for de novo three-dimensional structure determination in single-particle electron microscopy. J. Struct. Biol. 151, 92-105.

Scheres, S.H.W., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P.B., Frank, J., Carazo, J.-M., 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat Meth 4, 27-29.

Scheres, S.H.W., 2010. Visualizing molecular machines in action: Single-particle analysis with structural variability, p. 89-119, Recent Advances in Electron Cryomicroscopy, Part B, Academic Press.

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

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., Frank, J., 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat Protoc 3, 1941-1974.

Stuart, D.I., Abrescia, N.G.A., 2013. From lows to highs: using low-resolution models to phase X-ray data., p. 2257-2265, Acta Crystallogr. D Biol. Crystallogr.

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

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

Trabuco, L.G., Villa, E., Schreiner, E., Harrison, C.B., Schulten, K., 2009. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods 49, 174-180.

Trabuco, L.G., Schreiner, E., Gumbart, J., Hsin, J., Villa, E., Schulten, K., 2011. Applications of the molecular dynamics flexible fitting method. J. Struct. Biol. 173, 420-427.

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., Frank, J., 2003. Locking and unlocking of ribosomal motions. Cell 114, 123-134.

Veesler, D., Campbell, M.G., Cheng, A., Fu, C.-y., Murez, Z., Johnson, J.E., Potter, C.S., Carragher, B., 2013. Maximizing the potential of electron cryomicroscopy data collected using direct detectors. J. Struct. Biol. 184, 193-202.

Vulovic, M., Rieger, B., van Vliet, L.J., Koster, A.J., Ravelli, R.B.G., 2010. A toolkit for the characterization of CCD cameras for transmission electron microscopy, p. 97-109, Acta Crystallogr. D Biol. Crystallogr.

Wade, R.H., 1992. A brief look at imaging and contrast transfer. Ultramicroscopy 46, 145-156.

Wade, R.H., and Frank, J. 1977. Electron microscopic transfer functions for partially coherent axial illumination and chromatic defocus spread. Optik 49, 81-92.

Welch, P., 1967. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. 15, 70-73.

Wery, J.-P., Schevitz, R.W., 1997. New trends in macromolecular X-ray crystallography. Curr Opin Chem Biol 1, 365-369.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., van der Oost, J., Doudna, J.A., Nogales, E., 2011. Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477, 486-489.

Williams, D.B., Carter, C.B., 2009a. Inelastic Scattering and Beam Damage, p. 53-71-71, Transmission Electron Microscopy, Springer US.

Williams, D.B., Carter, C.B., 2009b. Elastic Scattering, p. 39-51-51, Transmission Electron Microscopy, Springer US.

Williams, D.B., Carter, C.B., 2009c. Transmission Electron Microscopy Springer US.

Williams, D.B., Carter, C.B., 2009d. Diffraction in TEM, p. 197-209-209, Transmission Electron Microscopy, Springer US.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M., 2008. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J. 275, 1-21.

Wriggers, W., Milligan, R.A., McCammon, J.A., 1999. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. J. Struct. Biol. 125, 185-195.

Wriggers, W., Chacón, P., 2001. Modeling tricks and fitting techniques for multiresolution structures. Structure 9, 779-788.

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

Zhang, W., Kimmel, M., Spahn, C.M.T., Penczek, P.A., 2008. Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis. Structure 16, 1770-1776.

Zhou, Z.H., 2011. Atomic Resolution Cryo Electron Microscopy of Macromolecular Complexes, p. 1-35, Recent Advances in Electron Cryomicroscopy, Part B, Academic Press.

Zhu, J., Penczek, P.A., Schröder, R., Frank, J., 1997. Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. J. Struct. Biol. 118, 197-219.

CHAPTER 2:

The Translational Elongation Cycle and the Role of BipA, a Novel Translational GTPase

CHAPTER 2 ABBREVIATIONS

Abbreviation	Full Title
(p)ppGpp	guanosine pentaphosphate (precursor of ppGpp)
30S IC	30S initiation complex
30S PIC	30S pre-initiation complex
3D	three-dimensional
70S IC	70S initiation complex
Å	angstrom
A site	aminoacyl site
aa-tRNA	aminoacyl tRNA
BipA	BPI-inducible protein A
BPI	bactericidal/permeability-increasing protein
CCA	cytosine-cytosine-adenosine
CTD	c-terminal domain
E site	exit site
EF-G	elongation factor G
EF-Tu	elongation factor Tu
EF4	elongation factor 4
EM	electron microscopy
GAP	GTPase-activating protein
GDP	guanosine diphosphate
GTP	guanosine triphosphate
IF	initiation factor
K_d	dissociation constant
Mb	megabases
mRNA	messenger RNA
nM	nanomolar
P site	peptidyl site
ppGpp	guanosine tetraphosphate
PTC	peptidyl transfer center
RF	release factor
RNA	ribonucleic acid
rRNA	ribosomal RNA
SD	shine-degarno
SHX	serine hydroxamate
SRL	sarcin ricin loop
trGTPase	translational GTPase
tRNA	transfer RNA
μM	micromolar

2.1 INTRODUCTION

While Chapter 1 aimed to provide a review of the methods used in this dissertation, this chapter focuses on the translational machinery and aims to introduce the biology behind the research later presented in Chapters 3 and 4. We begin with an introduction to ribosome structure and the mechanism of translation in Section 2.2. A deeper treatment of the bacterial elongation cycle is provided in Section 2.3. Finally, Section 2.4 introduces a recently discovered translational GTPase (trGTPase), BipA, whose structure bound to the ribosome and mechanism are the main foci of the rest of this dissertation.

2.2 TRANSLATION

The process of translation, whereby the genetic message encoded in messenger RNA (mRNA) is translated into a sequence of amino acids, is central to life in all organisms. At the heart of translation and protein synthesis is the ribosome, a large macromolecular complex composed of both protein and ribonucleic acid (RNA). Ribosomes across the different kingdoms of life differ in composition, as discussed below, but the overall structure is largely conserved. It should be noted that while the bacterial system will be used to illustrate the process of translation for the remainder of the chapter, the mechanism of translation is largely shared among all three kingdoms of life.

2.2.1 The Translational Machinery: The Ribosome

In all kingdoms of life, the ribosome is composed of a small and large subunit. In eubacteria and archaea, the small 30S subunit and the large 50S subunit form the complete 70S ribosome, while the eukaryotic counterparts are the 40S, 60S, and 80S, respectively. On each subunit lie three sites for the binding of transfer RNA (tRNAs), designated the aminoacyl (A), peptidyl (P), and exit (E) sites. Each subunit is a multicomponent, ribonucleoprotein complex. For example, in eubacteria, the small 30S subunit is composed of a 16S rRNA and 21 proteins while the large 50S subunit is composed of a 5S and an 18S rRNA and 34 proteins (Agrawal et al., 2011). Table 2.1 summarizes the ribosomal RNA (rRNA) and protein composition of ribosomes among various species. The complexity of this molecular machine reflects the monumental task that it must perform in translating mRNA nucleotide sequences with high fidelity into long amino acid polypeptides.

The first structures of the ribosome were described by Palade (Palade, 1955). Here, the ribosomes were seen as small particulates lining the endoplasmic reticulum. Studies in the *E. coli* system revealed that the complete 70S ribosome was composed of two subunits disparate in size (Huxley and Zubay, 1960). Using EM, Jim Lake was able to describe extensively the main topology of the ribosome and produce a 3D model of the macromolecular complex by inference from a few views (Lake, 1976). Unfortunately, ribosomal studies remained somewhat stagnant for the next decade. The first single particle structure of the 50S large subunit was done by Radermacher (Radermacher et al., 1987), albeit using the negative-stain technique. The first cryo-EM structure of the 70S was obtained at 40 Å in 1990 (Frank et al., 1991). It was not until 1995 that a cryo-EM reconstruction of the ribosome resolved to 25 Å (Frank et al., 1995), allowed characterization of topological features, intersubunit bridges, the binding sites of tRNAs, and the peptide exit channel.

The first atomic structures of the 30S subunit from *Thermus thermophilis* (Wimberly et al, 2000; Schluenzen et al., 2000) and 50S from *Haloarcula marismortui* in 2000 (Nissen et al., 2000a), followed by the complete 70S ribosome from *Thermus thermophilis* in 2001 (Yusupov et al., 2001) revolutionized the field of ribosomal research.

2.2.2 Translation at a Glance

The entire translation process, shown in Figure 2.1, is divided into four stages: initiation, elongation, termination, and recycling. During each stage, a number of translational factors modulate the ribosome to promote polypeptide production using amino acids delivered to the ribosome by tRNA. This section summarizes the entire process, but an extensive treatment of each stage can be found in (Rodnina et al., 2011).

During the *initiation* stage, a ribosomal complex is assembled in a step-wise manner to ultimately form a competent 70S initiation complex (the 70S IC), which is primed for the start of the first *elongation cycle*. Complementary binding between a portion of the 16S rRNA in the 30S small subunit with the Shine-Delgarno sequence of the mRNA, a conserved ribosomal binding sequence found upstream of the AUG start codon, positions the mRNA onto the platform of the 30S subunit and its start codon into the P site, leading to the formation of the 30S pre-initiation complex (30S PIC). Initiation factors (IFs) 1, 2, and 3 act in concert to position the first aminoacylated tRNA (aa-tRNA), bound to a formyl derivative of methionine (fMet), into the P site. The IFs facilitate productive interaction between the anticodon of the fMet-tRNA^{fMet} and the AUG start codon, forming the 30S initiation complex (30 IC). Dissociation of IF1 and IF3 and the joining of a 50S subunit with the 30S IC results in the 70S initiation complex (70S IC) primed for the first cycle of elongation.

The *elongation* cycle is divided into three principal steps: decoding, tRNA accommodation/peptide bond formation, and translocation. In the decoding step, incoming aa-tRNAs are delivered to the A site on the 30S subunit by elongation factor Tu (EF-Tu) where the anticodon stem loop of the aa-tRNA samples the codon of the mRNA in the 30S A site. Successful cognate-codon recognition stabilizes the aa-tRNA in the A site, releasing free energy that is used to induce conformational changes in the 30S head and EF-Tu, thereby promoting GTP hydrolysis on EF-Tu. Following GTP hydrolysis, EF-Tu dissociates from the ribosome, leaving the aa-tRNA in the A site to be accommodate dinto the peptidyl transfer center (PTC) located on the 50S subunit. Here, peptide bond formation occurs, transferring the nascent polypeptide chain bound to peptidyl tRNA in the P site to the aa-tRNA in the A site. In the subsequent translocation, the concerted movement of the tRNAs (from the A and P sites, to the P and the E sites, respectively) and the mRNA by one codon, catalyzed by elongation factor G (EF-G), results in an empty A site for the next incoming aa-tRNA. The process starts anew, with each elongation cycle incorporating an additional amino acid into the growing nascent polypeptide chain, which exits the ribosome via an exit tunnel of the 50S subunit.

Termination of translation occurs when a stop codon (UAA, UAG, or UGA) is reached in the mRNA template. Here, the codon is recognized by release factors (RF) 1 or 2 instead of an aa-tRNA. RF1 recognizes the stop codons UAA and UAG while RF2 recognizes UAA and UGA. The specific release factor binds to the A site of the 30S subunit, mimicking an A-site tRNA. The RF not only recognizes the stop codon but also breaks the ester linkage of the polypeptide chain to the peptidyl-tRNA, thereby releasing the nascent peptide into the exit tunnel. Subsequently, RF3 binds to facilitate the release of either RF1 or RF2, thus completing translation termination.

The research in this thesis focuses largely on the translational GTPase BipA and its homologues. Thus, a more in-depth treatment of the elongation cycle and the various canonical elongation factors is provided below in Section 2.3.

2.3 THE BACTERIAL ELONGATION CYCLE

Each round of the elongation cycle incorporates an amino acid to a growing nascent polypeptide chain. The process is divided into three main steps: decoding, accommodation/peptide bond formation, and translocation. The entire cycle is coordinated, regulated, and catalyzed by the ribosome and two enzymes belonging to the elongation factor family of translational GTPases (trGTPases): EF-Tu and EF-G. Here, the structural basis for the elongation cycle is described to provide an understanding for the high fidelity, rapidity, and mechanism of the process.

2.3.1 tRNA Structure and The Ternary Complex

tRNAs are the adaptor molecules that act as the translators between the nucleic acid and polypeptide codes. The tertiary structure of a phenyalanine tRNA molecule is shown in Figure 2.2. At the 3' CCA end of the tRNA, named for the typical cytosine-cytosine-adenosine sequence found, amino acids are attached via an ester linkage, "charging" the tRNA to form an aminoacyl-tRNA. At the other end of the aa-tRNA is the anticodon, three nucleotide bases capable of complementary binding to a specific codon of the mRNA blueprint. Other regions, such as the D-loop and the variable loop, are very flexible, allowing for distortion of the tRNA, an ability important for decoding and accommodation, discussed below.

This entire assembly of EF-Tu–GTP–aa-tRNA is termed the ternary complex. EF-Tu is a translational GTPase consisting of three domains, shown schematically in Figure 2.3. Domain I, termed the GTPase domain, binds to GTP and catalyzes its hydrolysis to GDP and an inorganic phosphate. This domain is conserved among all translational GTPases. Domains II and III both adopt a β-barrel motif and are responsible for binding the aa-tRNA. The L7/L12 protein of the 50S large subunit is thought to recruit the ternary complex (as well as a number of other factors) to the ribosome (Wahl and Moller, 2002; Diaconu et al., 2005). Binding of the ternary complex to the ribosome, shown in Figure 2.4, positions the aa-tRNA into the A site of the 30S small subunit and EF-Tu adjacent to the A site of the ribosome, with Domain I positioned at the GTPase-associated center (GAC) on the 50S large subunit and Domain II interacting with the 16S rRNA of the 30S small subunit. The GAC is characterized by the region of the 50S subunit encompassing protein L11 at the base of the L7/L12 stalk and the Sarcin-ricin loop (SRL), A2660-A2664 of the 23S rRNA (Valle et al., 2003a; Clementi et al., 2010; Voorhees et al., 2010; Shi et al., 2012).

2.3.2 Decoding and the GTPase-Activation Mechanism

Once the ternary complex has successfully bound, a distortion in the anticodon stem and a rotation of the D-loop in the aa-tRNA results in the tRNA adopting a conformation known as the A/T state (Valle et al, 2003a). In this state, discussed further in Section 2.3.4, the tRNA anticodon samples the codon on the mRNA template in a process called decoding. Noncognate tRNA cannot be stabilized on the ribosome, and such ternary complexes readily dissociate from the ribosome. Near-cognate tRNAs are not as stable in the decoding center as a cognate tRNA, especially after GTP hydrolysis on EF-Tu has occured, to be discussed below. Successful cognate codon-anticodon recognition results in the release of binding energy used to induce a domain closure of the 30S small subunit (Ogle et al., 2002).

Adoption of the tRNA into the A/T state, codon recognition, the subsequent binding energy release, and the 30S domain closure are all important signals for GTPase activation of EF-Tu. The closure of the 30S induces further conformation changes in EF-Tu Domain II resulting in the disruption of the tRNA's contact with the switch I loop (residues 40-62) in EF-Tu Domain I. The disruption of switch I is accompanied by a rearrangement of two amino acids, Val20 and Ile60, that make up a "hydrophobic gate" (Villa et al., 2009). This rearrangement allows His84 to be coordinated into the catalytic site of Domain I by A2662 of the SRL (Berchtold et al., 1993; Aleksandrov and Field, 2013). On EF-Tu, His84 is the catalytic amino acid responsible for hydrolysis, as mutation of this residue to an alanine reduces the rate of GTP hydrolysis by six orders of magnitude (Daviter et al., 2003). While various proposed mechanisms differ in the exact role of His84, there is agreement on the reordering of His84 into the active site inline with a water molecule and the γ-phosphate of the GTP as necessary for productive hydrolysis (Voorhees et al., 2010; Liljas et al., 2011). This GTPase-activation mechanism is universally shared by all elongation trGTPases. The stabilizing of the transition state for GTP hydrolysis by the ribosome is akin to the stabilization mechanism of other traditional GTPase-activating proteins (GAPs) such as Ran and Ras. Thus, the ribosome itself can be considered a GAP (Liljas et al., 2011).

2.3.3 Accommodation and Peptide Bond Formation

GTP hydrolysis results in conformational changes in EF-Tu Domain III that disrupt its binding to the tRNA and to the ribosome. tRNA detachment from EF-Tu further promotes EF-Tu dissociation from the ribosome (Valle et al., 2003a). The correct codon-anticodon recognition stabilizes the remaining interactions of the tRNA with the ribosome, allowing the aa-tRNA to swivel and accommodate its 3' CCA end into the peptidyl transfer center (PTC) on the 50S subunit (Valle et al., 2003a; Gromadski and Rodnina, 2004). Near-cognate tRNAs are much more unstable in comparison to cognate tRNAs once EF-Tu has dissociated from the ribosome. Biochemical studies suggested that nearly a 50-fold increase in GTP consumption was required for amino-acid incorporation of near-cognate codons as compared to cognate codons (Ruusala et al., 1982). Crystallographic and biochemical experiments show that the PTC is composed of 23S rRNA residues (Voorhees et al., 2009), promoting the idea that the ribosome acts as a ribozyme during peptide bond formation.

Structural and biochemical data suggest that the aa-tRNA is accommodated into the PTC following an induced-fit model. Once positioned, the α-amino group of the aa-tRNA performs a nucleophilic attack on the aminoacyl ester linkage on peptidyl-tRNA, resulting in the transfer of the peptide chain from the P-site tRNA onto the A-site tRNA (Bashan et al., 2003; Trobro and Åqvist, 2005).

2.3.4 tRNA Conformations & Hybrid States

During elongation, tRNAs exhibit a variety of different conformations as they move sequentially from the A to the P and then to the E site. This movement is incremental and a number of tRNA hybrid states, shown in Figure 2.5, can been observed (Agirrezabala et al., 2012). These states, such as the A/T state mentioned in Section 2.3.2, are canonically named 'X/Y' where 'X' is the position of the anticodon stem loop and 'Y' is the position of the 3' CCA end of any particular tRNA relative to the three tRNA binding sites on the 30S and the 50S (Moazed and Noller, 1989). Aminoacyl-tRNAs during the decoding process, bound to the ternary complex, are said to be in the A/T state. Accommodated A-site aa-tRNAs, with their anticodons in the A site of the 30S and their 3' aminoacylated CCA ends in the PTC, are said to be in the A/A state. Peptidyl tRNAs, with their anticodon in the P site of the 30S and their 3' peptidyl end in the PTC of the 50S are in the P/P state. After peptidyl transfer has occurred, the deacylated P-site tRNA and the new A-site peptidyl tRNA must now move to the E site and P site, respectively, to make room for the next incoming aa-tRNA. Hybrid state formation has been linked to an intersubunit rotation of the 30S relative to the 50S by 5-10 degrees (Agirrezabala et al., 2008). This so-called "rotated" or "ratcheted" state, shown in Figure 2.4, is also important for the translocation step of elongation.

2.3.5 Translocation

In translocation, the movement of the mRNA and the anticodon of the tRNAs by one codon is mediated by the trGTPase elongation factor G (EF-G). As in EF-Tu, the first domain of EF-G is a GTPase domain, though EF-G contains an additional domain, called G', that interacts with the C-terminal of the L7/L12 protein (Datta et al., 2005). EF-G is possibly the best example of molecular mimicry: its crystal structure (Czworkowski et al., 1994) strongly resembles the general shape of the entire ternary complex (Nissen et al., 2000b). As shown in Figure 2.3, Domains I and II resemble the structure of EF-Tu, while III, IV, and V collectively mimic the tRNA, with domain IV specifically resembling the anticodon stem arm of the tRNA. Indeed, the first visualization of EF-G bound to the ribosome, via cryo-EM, revealed that domain IV of EF-G docks into the A site just as A/T tRNA (Agrawal et al., 1998). Comparisons between cryo-EM densities of bound EF-G (Valle et al., 2003b) and X-ray solution structures of EF-G (Czworkowski et al., 1994; al-Karadaghi et al., 1996) revealed a 27 Å swiveling of domain IV that must occur for binding to the A site (Li et al., 2011). Several recent crystallographic structures reveal that bound EF-G orders the catalytic His87 in its GTPase domain in a similar position as His84 of EF-Tu, suggesting a shared mechanism of GTPase activation (Pulk and Cate, 2013; Tourigny et al., 2013).

Binding of domain IV of EF-G to the A site seemingly "pushes" the tRNAs into their subsequent binding sites. While spontaneous translocation (absent of any elongation factors) has been observed (Cornish et al., 2008), the EF-G-mediated translocation rate is orders of magnitude faster (Rodnina et al., 1997). GTP hydrolysis and the resulting conformational changes in EF-G cause the protein to dislodge from its binding site. The reverse rotation of the 30S subunit back into the unrotated state of the ribosome leaves the tRNAs in the P and E sites while the mRNA has moved by one codon. This state essentially resets the ribosome for the next cycle of elongation.

2.3.6 EF4: The Backtranslocase

Errors in the elongation process, such as the mis-incorporation of a noncognate tRNA, must be identified and rectified. Beyond the kinetic proofreading steps employed during the decoding and accommodation process, the ribosome may use another elongation factor: EF4 (initially known as LepA). The factor's gene was first identified as the part of the *lep* operon, an operon upstream of the signal peptidase I gene in *E. coli* (March and Inouye, 1985a). Characterization of the gene product of the *lep* operon revealed that the protein product, called LepA, has high sequence homology to elongation factor EF-G and EF-Tu (March and Inouye, 1985b).

Since its discovery, further structure (Connell et al., 2008; Evans et al., 2008) and functional studies (Qin et al., 2006; Liu et al., 2011) have established the protein as the backtranslocase. The protein catalyzes a backtranslocation of the ribosome, allowing tRNAs in the E and P site to shift back to the P and A site, respectively, as part of an error-correction mechanism (Qin et al., 2006; Yamamoto et al., 2014). Spontaneous backtranslocation of the ribosome has been previously observed (Konevega et al., 2007; Fischer et al., 2010). Cryo-EM maps of the 70S–LepA complex, after backtranslocation has occurred, revealed that LepA distorts the A-site tRNA into a new, previously uncharacterized, A/L state (Connell et al., 2008). LepA was subsequently renamed EF4 to reflect its inclusion in the elongation factor family of GTPases.

2.4 BipA

2.4.1 Discovery and Initial Characterization of the BipA Protein

Pathogenic bacteria encounter a variety of environments during their life cycle. In order to survive, these bacteria must adapt to environmental stresses such as abrupt changes in temperature, pH, nutritional resources, and amino acid pools. To cause disease, these bacteria invade host cells, where they are often met with a battery of innate antimicrobial agents. Host cells may release antimicrobial agents such as defensins, lysozymes, and bactericidal/permeability-increasing protein (BPI). Quick response and adaptation to the stress imposed by the host immune system is imperative for successful bacterial survival.

BipA, an elongation trGTPase, was first discovered in 1995 by two-dimensional electrophoresis. Qi and coworkers (Qi et al., 1995) exposed *Salmonella enterica typhimurium* cells to BPI in order to activate the stress response of these bacteria. Two-dimensional gel electrophoresis was performed before and after exposure to BPI in an effort to detect proteins whose expression levels change over the course of the stress response. Comparison of the gels revealed the presence of a protein induced by more than sevenfold over basal levels during the stress response. After extraction and sequencing of the various proteins, these authors found high sequence homology between one 67.4 kDa protein, named BipA (BPI-Inducible Protein A), and the canonical elongation factors, such as EF-Tu and EF-G.

2.4.2 The bipA Gene

A study of 191 fully sequenced prokaryotic genomes found one copy of the *bipA* sequence in all but 26 genomes smaller than 1.5Mb (Margus et al., 2007). Among different genomes, the *bipA* gene is highly conserved, especially in Domain I and the C-terminal domain, a domain of unknown function (Scott et al., 2003). Recently, *bipA* has been found in the genomes of various lower-order eukaryotes such as *S. salsa* (Wang et al., 2008) and trypanosomes (unpublished data). However, higher eukaryotic organisms, such as humans, do not contain a homologous sequence. While there may be a homolog of BipA in higher order eukaroytes, it may not have been detected as the sequence could be quite different from those of lower eukaryotes and prokaryotes.

2.4.3 BipA Shares Homology with the Elongation Factor Family of Translational GTPases

BipA has overall sequence and structural homology to EF-Tu, EF-G, and EF4, the canonical members of the elongation factor family of GTPases. As discussed in Sections 2.3, EF-Tu delivers aminoacyl-tRNAs to the A site of the 30S subunit for decoding and accommodation, EF-G is the ribosomal translocase, and EF4 is the ribosomal backtranslocase. Table 2.2 provides a summary of the sequence homology between BipA and the elongation factors. Biochemical studies have shown that BipA does not share the same functions as EF4 or EF-G. As yet, no biochemical study has been able to deduce a role for BipA during the normal course of elongation. Thus, while BipA has high sequence homology to EF-Tu, EF-G, and EF4, it cannot be considered an elongation factor.

BipA, EF-G, and EF4 each have five domains, shown schematically in Figure 2.6. Domains I, II, III, and V are common to all three proteins. Domain I, the GTPase domain, is the most conserved domain. Point mutations of conserved amino acids at the predicted BipA GTP hydrolysis site negates its GTPase activity, reinforcing the idea that BipA, EF-G, and EF4 share the same mechanism of GTPase hydrolysis. The G’ domain of EF-G is absent from BipA and EF4. Studies have shown that this G’ domain interacts with ribosomal protein L7/L12 (Nechifor et al., 2007). Its deletion or the deletion of L7/L12 results in a reduction of EF-G’s GTPase activity. As the other elongation trGTPases, BipA catalyzes the hydrolysis of GTP to GDP, a point that will be discussed below in Section 2.4.4.

While four out of the five domains are common to all three proteins, each has a unique domain specific to its function and mechanism. EF-G's Domain IV extensively interacts with the ribosome in the intersubunit space by mimicking the anticodon loop of the A site tRNA (Connell et al., 2007). BipA and EF4 both lack a domain homologous to EF-G Domain IV. Instead, the C-terminal domains of BipA and EF4 are unique domains specific to each protein. For comparison purposes, the domain nomenclature established for EF-G is used to number both BipA's and EF4's corresponding homologous domains. Thus, BipA and EF4's domains are named I, II, III, V, and CTD, even though Domain V is the fourth domain in the linear sequence.

Finally, while all three proteins have been shown to bind to the same general site on the 70S ribosome (Owens et al., 2004), the sequence of BipA's CTD has no homology to any other protein sequence. Superimposing the crystal structures of EF-G bound to the 70S ribosome in the pre-translocation state (PDB: 3J5X from (Brilot et al., 2013)), and EF4 bound to the 70S ribosome in the post-backtranslocated state (PDB: 3DEG from (Connell et al., 2008)), reveals extensive topological differences between the EF4 CTD and EF-G Domain IV. As shown in Figure 2.7, not only are the topologies of the two domains completely different, but they also occupy different spatial regions, the significance of which will be discussed in Chapter 4. BipA's CTD, with its non-homologous sequence, may also have structural features that are distinct from EF-G and EF4 (deLivron et al., 2009). Finding the specific ribosomal interactions will be crucial to understanding BipA's mechanism and function.

2.4.4 The Ribosome Enhances BipA's GTPase Activity

Due to its high sequence homology to other elongation factors, it was expected that BipA's binding partner would be the ribosome. Pull-down assays corroborated this hypothesis (deLivron and Robinson, 2008). Biochemical studies pinpointed BipA's binding to the general docking site as other elongation factors. BipA's CTD mediates its interaction with the 70S ribosome (deLivron et al., 2009). A deletion of the CTD eliminates binding completely, suggesting a crucial important of the unique CTD in binding and, by extension as discussed below, GTP hydrolysis.

Akin to EF-G, EF-Tu, and EF4, BipA also experiences 70S ribosome-induced increase in GTPase activity, suggesting a shared mechanism for GTPase activation (deLivron and Robinson, 2008). Isolated, purified BipA in vitro has a basal turnover rate of 19.8 h^{-1}. This activity is increased fourfold to ~80 h-1 in the presence of 70S ribosomes. Compared to the rate of EF-G GTPase activity in the presence of 70S ribosomes, ~6-7 s^{-1} (Rodnina et al., 1997), BipA is approximately 315X slower at catalyzing the hydrolysis of GTP to GDP.

2.4.5 BipA Exhibits Two Ribosomal Binding Modes

BipA exhibits two distinct ribosomal binding modes. In the first binding mode, under normal cellular conditions, BipA binds to the 70S ribosome. Biochemical titration assays (unpublished results from Victoria Robinson) show that BipA competes with EF-G for binding near the A site. However, BipA can only effectively compete with EF-G when the BipA quantity is over 3X the quantity of EF-G.

Several antibiotics known to bind to the A site were used to confirm the general binding site of BipA. Studies found that thiostrepton, which binds to the base of the L7/L12 stalk, was able to block BipA binding (Mikolajka et al., 2011). This same antibiotic also blocks the binding of EF-G and EF4. (Walter et al., 2012). Interestingly, fusidic acid was shown to inhibit the GTPase activity of EF-G, but not that of BipA (Mikolajka et al., 2011). Also, fusidic acid, which inhibits translation by binding to EF-G, has no effect on BipA GTPase activity. The differential effects of antibiotics on the elongation factors indicate that BipA has distinct structural features.

In the second binding mode, when a stringent response is induced by the addition of serine hydroxamate (SHX), an amino acid analog that inhibits correct charging of serine tRNAs, BipA changes its binding affinity. Under these conditions, pull-down assays show that BipA associates solely with free 30S subunits, with little affinity for either the 50S large subunit or the 70S ribosome. Studies determined that this differential binding mode occurs upon BipA binding to ppGpp instead of GTP (deLivron and Robinson, 2008). The accumulation of the guanosine nucleotide (p)ppGpp is the hallmark of the stringent response (Goldman and Jakubowski, 1990; Magnusson et al., 2005). ppGpp is the derivative of (p)ppGpp and a global stress alarmone (Magnusson et al., 2005). (p)ppGpp is largely produced by RelA, a ribosomal factor which binds to the ribosome A site in the presence of a deacylated A-site tRNA (Haseltine and Block, 1973; Payoe and Fahlman, 2011). Production of (p)pGpp quickly causes a cascade of different stress-response pathways leading to expression of specific stress response-related proteins (Kanjee et al., 2012).

Its overproduction during the stringent response initializes a cascade of stringent-response pathways (Kanjee et al., 2012). The structural and biochemical changes responsible for this differential binding are not known. Interestingly, BipA is only found in organisms that also have RelA, the protein responsible for the creation of ppGpp. This further suggests that BipA and ppGpp work in concert with the ribosome to control gene expression during stress response.

2.4.6 The Physiological Function of BipA

Very little is known about the physiological function of BipA. While the ribosome is a known binding partner of BipA, no other protein or macromolecular complex has been identified as a potential target. Thus, while BipA's role has been implicated in a variety of pathways, the specific physiological function has never been fully elucidated.

The discovery of BipA, in *S. typhimurium* cells experiencing sudden stress conditions, led Qi and coworkers to speculate BipA's role in the stringent response (Qi et al., 1995). The stringent response was first identified in bacterial cells under amino acid starvation conditions (Forro, 1965). However, the term has come to mean the induction of a general stress response to a variety of sudden stresses such as amino acid starvation, nutrient depletion, gene expression repression, pH changes, heat shock, and antimicrobial agents such as defensins, lysozymes, and bactericidal/permeability increasing protein (BPI) (Jain et al., 2006; Carneiro et al., 2011).

Studies have reinforced BipA's vital role in the stringent response to a variety of environmental stresses. BipA has been shown to confer resistance to a variety of antimicrobial agents including sodium dodecyl sulfate (Kiss et al., 2004), BPI (Qi et al., 1995; Grant et al., 2003), and BPI derivatives (Barker et al., 2000). These mutants show greatly reduced growth upon exposure to these agents. BipA was found necessary for the expression of the group 2 capsule gene clusters (Rowe et al., 2000) which encode the proteins required for mediating resistance to host immune system responses (Merino and Tomás, 2001). Deletion or mutation in the *bipA* gene leads to increased bacterial sensitivity to heat shock (Rowe et al., 2000) and cold shock (Pfennig and Flower, 2001; Kiss et al., 2004). Inactivation of the protein in E. coli cells results in the disappearance of the other stress proteins, leading to speculation that BipA has a possible role in gene regulation (Freestone et al., 1998b; Freestone et al., 1998a). Addition of *S. meliloti* BipA can reverse the phenotype of ΔbipA *E. coli* (Krishnan and Flower, 2008). Such an extensive conservation suggests that BipA plays a major functional and physiological role in bacteria. Lastly, as detailed in the previous section, under amino acid starvation conditions, BipA experiences differential binding modes, to either the 70S ribosome or the 30S small ribosomal subunit, depending on the nucleotide concentration present in the bacterial system (deLivron and Robinson, 2008). BipA's dissociation constant, K_d, for GTP, GDP, and ppGpp are 22 µM, 29 µM, 11 µM, respectively (unpublished data from Victoria Robinson). BipA's higher affinity for ppGpp implies that BipA is primed to respond quickly under sudden stress conditions.

Beyond involvement in the stringent response, BipA has been implicated in the regulation of virulence-related gene expression in a variety of species. In enteropathogenic *E. coli*, BipA mutants have dramatically reduced expression of the LEE pathogenicity island, a cluster of genes that encodes proteins necessary for colonization and invasion of host cells (Grant et al., 2003). Additionally, pathogenic bacteria cells lacking a functional BipA protein have reduced motility, resulting in inability of these cells to attach to host cells (Farris et al., 1998; Møller et al., 2003). In *S. meliloti*, which invades and forms symbiotic relationships with plants, BipA is required for symbiosis and for the response to sudden pH fluctuations upon bacterial invasion of the plant host (Kiss et al., 2004). Interestingly, (Kiss et al., 2004) found that E. coli BipA can rescue the phenotype of the S. meliloti BipA mutant, suggesting that the protein has shared functions across bacterial species.

Species		*E. Coli*	*S. Cerevisiae*	*H. Sapiens*	*B. Taurus* *Mitochrondria*
Molecular Mass (MDa)		2.3	3.3	4.3	2.7
Diameter (Å)		~260	~300	~320	~320
Sedimentation Coefficient		70S	80S	80S	55S
Small Subunit	*Name*	30S	40S	40S	28S
	rRNA	16S	18S	18S	12S
	# of Proteins	21	33	33	29
Large Subunit	*Name*	50S	60S	60S	39S
	rRNA	23S and 5S	5S, 5.8S, 25S	5S, 5.8S, 28S	16S
	# of Proteins	34	46	47	50

Table 2.1. Comparison of ribosomes across different species. Structural information on ribosomes in different species is summarized. While ribosomes across the different kingdoms of life differ in composition, the overall architecture is largely conserved: each complete ribosome is composed of a small and large subunit. Each subunit is a multicomponent complex of both proteins and ribonucleic acid (RNA). Structural information was obtained from (Agrawal et. al, 2011) and (Anger et al., 2013)

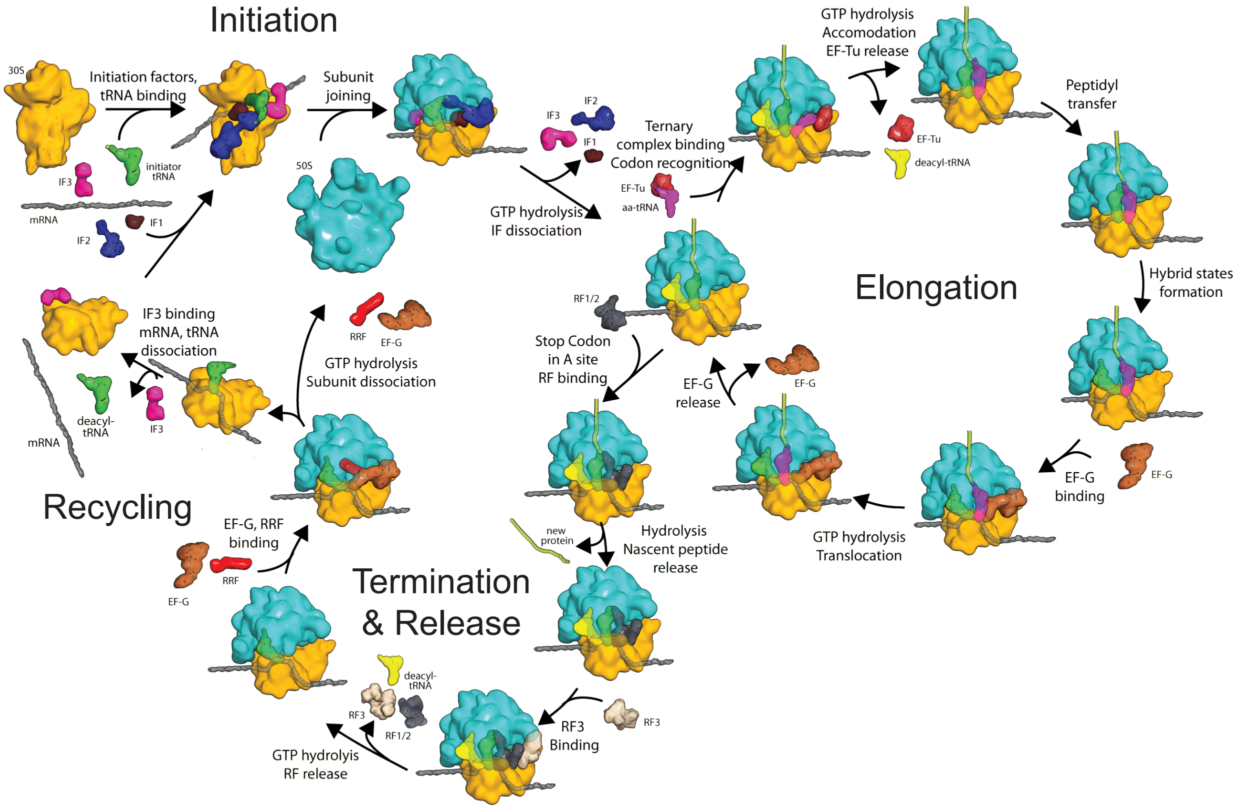


Figure 2.1. Translation in eubacteria. The entire process of translation is depicted here in four stages: initiation, elongation, termination/release, and recycling. During each stage, a number of translational factors modulate the ribosome to promote productive polypeptide synthesis. During initiation, a 70S complex is assembled with an fMet-tRNAfmet in the P site paired with an AUG start codon on the mRNA. In each round of elongation, a single amino acid is incorporated into the nascent polypeptide chain. During termination, release factors (RFs) recognize one of three stop codons and catalyze the release of the polypeptide chain from the ribosome. In recycling, the 70S is disassembled into its individual components to be reused in a new initiation stage. Reproduced with permission from (Schmeing and Ramakrishnan, 2009, with re-adjusted color scheme).

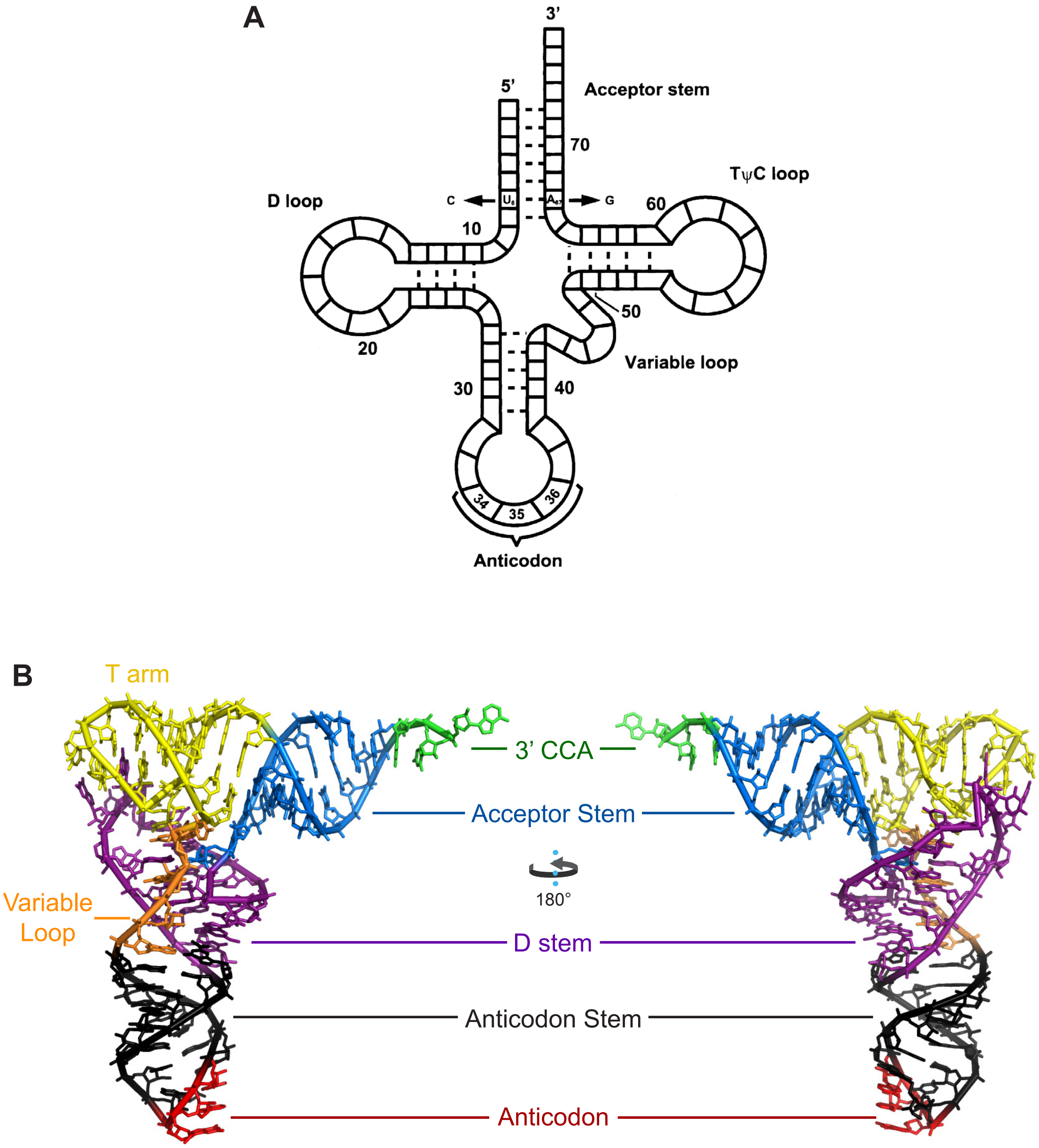


Figure 2.2. The secondary and tertiary structure of tRNA. The cloverleaf secondary structure of a phenylalanine tRNA is shown in (A). Its crystal structure (PDB:1EHZ) (Shi and Moore, 2000), resolved to 1.9 Å, is shown in (B) with various structural features color-coded and labeled accordingly. (A) is reproduced with permission from (Shi and Moore, 2000).

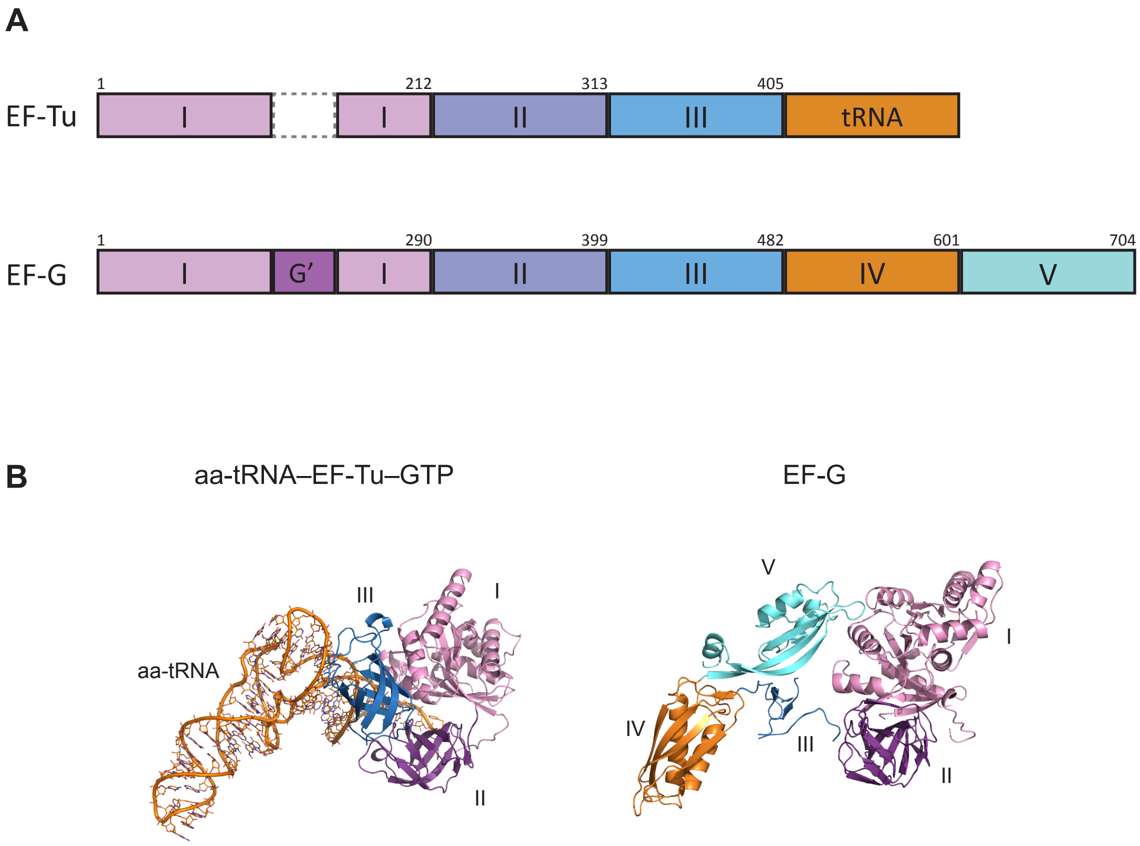


Figure 2.3. EF-Tu and EF-G. The domains of EF-Tu and EF-G are depicted schematically in (A). The crystal structures of the aa-tRNA–EF-Tu–GTP ternary complex (PDB: 1EHZ from Nissen et al., 1995) and EF-G (PDB: 1DAR from al-Karadaghi et al., 1996) are shown in (B), with each domain colored as the primary sequence in (A). The tertiary structure of the ternary complex and EF-G strongly resemble each other in overall shape. Domains I and II are homologous between EF-Tu and EF-G. EF-G domains III, IV, and V collectively resemble the shape of the aminoacyl-tRNA (aa-tRNA).

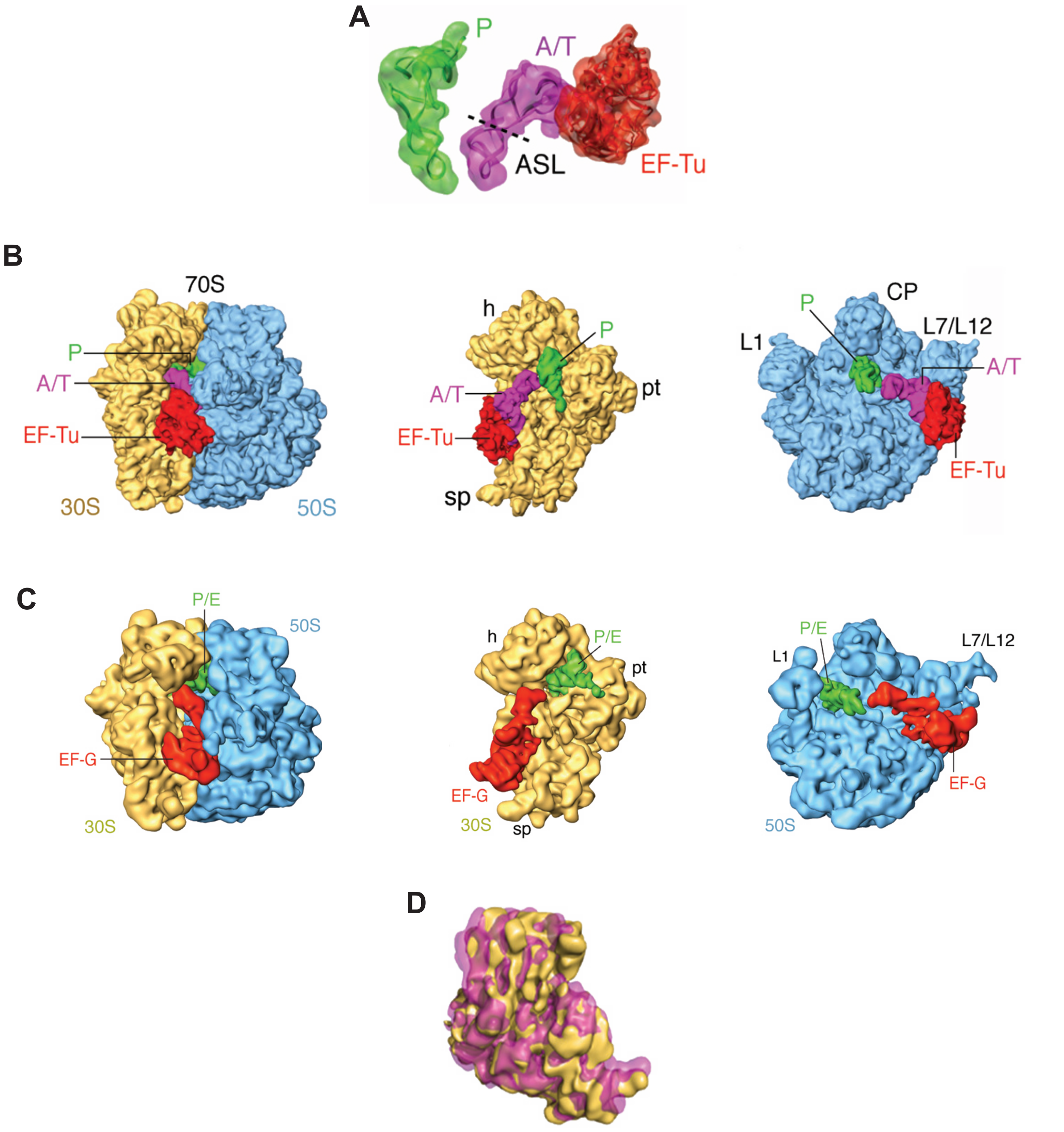


Figure 2.4. The binding of the ternary complex and EF-G. The cryo-EM map of the ternary complex bound to the 70S ribosome (Agirrezabala et al., 2011), shows EF-Tu in the GTPase-associated center (GAC) of the ribosome and the aa-tRNA anticodon stem loop (ASL) in the A site of the 30S. (A) Focused view of the architecture of the ternary complex. (B) The entire 70S–ternary complex. (C) Reconstruction of the 70S–EF-G complex (Valle et al., 2003b), showing EF-G occupying the same general binding site as the ternary complex. EF-G's domain IV extends into the A site of the 70S ribosome. EF-G binding stabilizes intersubunit rotation of the 30S with respect to the 50S in a "rotated" or "ratcheted" state (Frank et al., 2007). The unrotated and rotated states are shown superimposed in (C) in yellow and magenta, respectively. 30S small subunit landmarks are abbreviated as follows: h - head; pt - platform; sp - spur. Figures reproduced with permission from (Frank, 2009).

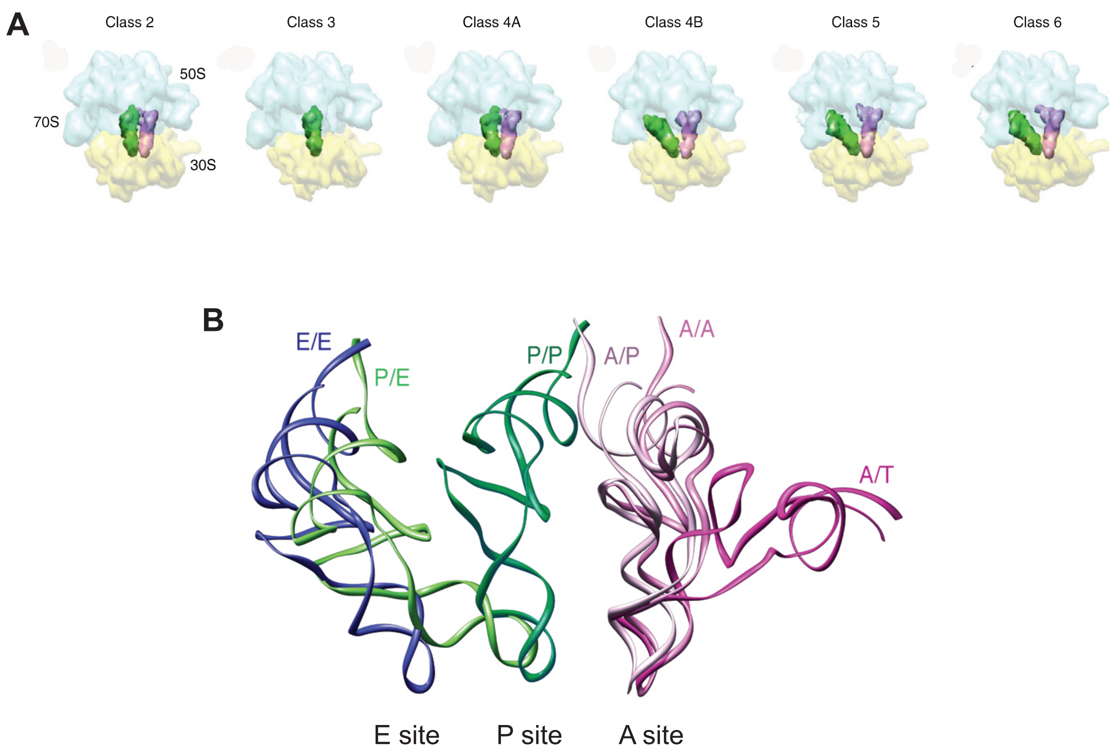


Figure 2.5. tRNA hybrid states. During elongation, tRNAs move sequentially with the mRNA, from the A (aminoacyl) to the P (peptidyl) and then to the E (exit) site. The movement is accompanied by the intersubunit rotation, which has been observed to occur spontaneously or with the binding of EF-G. In (A), different ribosomal classes, captured by cryo-EM (Agirrezabala et al., 2012), show distinct conformations of tRNAs in the classic A/A-P/P state (class 2), P/P state (class 3), P/P-P/E state (classes 5 and 6) and two intermediate classes (classes 4A and 4B). Fitting of tRNA X-ray structures into cryo-EM maps allow comparison of the different tRNA hybrid states in (B). A-site, P-site, and E-site tRNAs are shown in magenta, green, and blue, respectively. Figures (A) and (B) are reproduced with permission from (Frank, 2012) and (Agirrezabala and Frank, 2009), respectively.

A

Protein	BipA	EF-Tu	EF-G	EF4
Number of Amino Acids	607	394	704	598
Number of Domains	5	3	5	5
Sequence Homology	--	32% (48%)	27% (44%)	28% (47%)

B

Homologous Protein	EF-G	EF4
BipA Domain #		
I	39% (55%)	39% (58%)
II	28% (48%)	30% (52%)
III	35% (60%)	27% (46%)
V	27% (48%)	29% (50%)
CTD	--	--

Table 2.2. Sequence homology between *S. enterica* BipA and other *S. enterica* elongation factors. (A) Overall protein sequence homology between BipA and other canonical elongation factors (EF-Tu, EF-G, and EF4) with the BipA sequence used as the reference. Shown in parenthesis is the percentage of positives: amino acids who differ in identity but have the same chemical properties. (B) The BipA sequence, further divided into individual domains according to the numbering scheme in Figure 2.6. Each of BipA's domains were compared to the corresponding homologous domain in EF-G and EF4. The BipA CTD has no homology to EF-G domain IV or EF4 CTD. Because EF-Tu contains only three domains, its sequence was not used for individual domain comparisons. CTD = C-terminal domain.

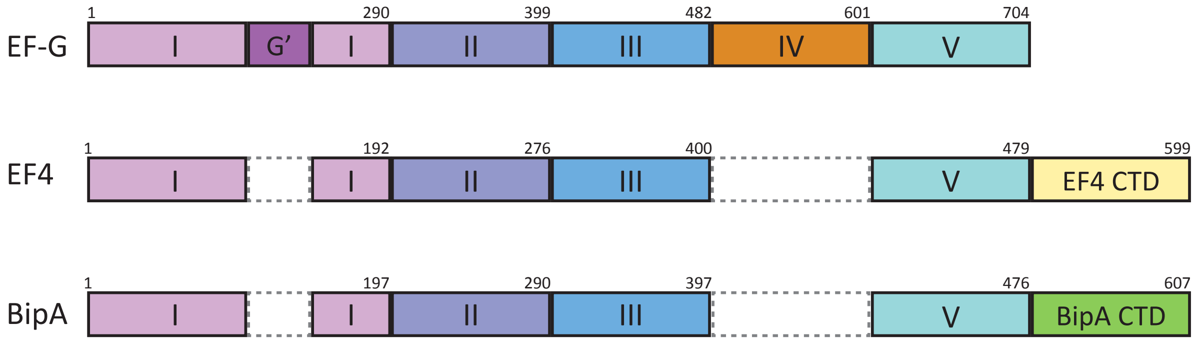


Figure 2.6. Domain comparisons between EF-G, EF4, and BipA. EF-G, BipA, and EF4 are composed of five individual domains. Domains I, II, III, and V are common to all three proteins. Domain IV of EF-G, which makes extensive contacts with the ribosome, is absent from both BipA and EF4. Instead, BipA and EF4 each have a unique CTD that is required for ribosomal binding.

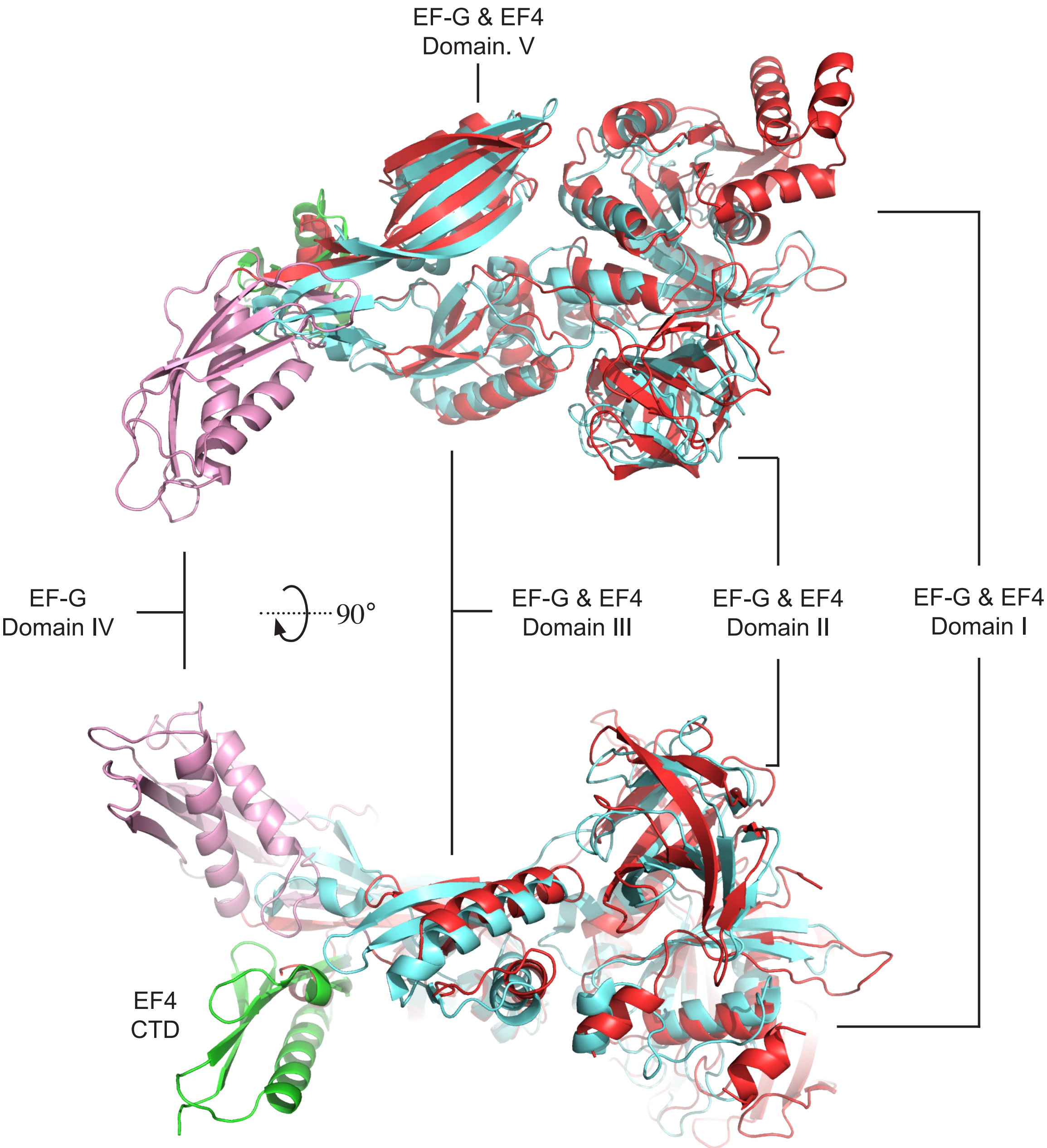


Figure 2.7. Superimposition of EF-G and EF4 crystal structures. The crystal structures of EF-G bound to the 70S in a pre-translocation state (PDB: 3J5X) and EF4 bound to the 70S ribosome in a post-backtranslocation state (PDB: 3DEG) are shown superimposed on one another. The superimposition reveals that four of the five domains (I,II, III, and V) overlay well, while EFG domain IV and EF4's CTD occupy different spatial regions. Domains I, II, III, and V in EF-G and EF4 are colored red and blue, respectively, per protein. The non-homologous domains, EF-G Domain IV and EF4 CTD, are colored in pink and green, respectively. CTD = C-terminal domain.

REFERENCES

Agirrezabala, X., Lei, J., Brunelle, J.L., Ortiz-Meoz, R.F., Green, R., Frank, J., 2008. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. Mol. Cell 32, 190-197.

Agirrezabala, X., Liao, H.Y., Schreiner, E., Fu, J., Ortiz-Meoz, R.F., Schulten, K., Green, R., Frank, J., 2012. Structural characterization of mRNA-tRNA translocation intermediates. Proc Natl Acad Sci USA 109, 6094-6099.

Agrawal, R., Sharma, M., Yassin, A., Lahiri, I., Spremulli, i., 2011. Structure and function of organellar ribosomes as revealed by cryo-EM. Ribosomes, 83-96-96.

Agrawal, R.K., Penczek, P.A., Grassucci, R.A., Frank, J., 1998. Visualization of elongation factor G on the *Escherichia coli* 70S ribosome: The mechanism of translocation. Proc Natl Acad Sci USA 95, 6134-6138.

al-Karadaghi, S., Aevarsson, A., Garber, M., Zheltonosova, J., Liljas, A., 1996. The structure of elongation factor G in complex with GDP: conformational flexibility and nucleotide exchange. Structure 4, 555-565.

Aleksandrov, A., Field, M., 2013. Mechanism of activation of elongation factor Tu by ribosome: catalytic histidine activates GTP by protonation. RNA 19, 1218-1225.

Barker, H.C., Kinsella, N., Jaspe, A., Friedrich, T., O'Connor, C.D., 2000. Formate protects stationary-phase *Escherichia coli* and *Salmonella* cells from killing by a cationic antimicrobial peptide. Mol Microbiol 35, 1518-1529.

Bashan, A., Agmon, I., Zarivach, R., Schluenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H.A.S., Kossoy, E., Kessler, M., Yonath, A., 2003. Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. Mol. Cell 11, 91-102.

Berchtold, H., Reshetnikova, L., Reiser, C.O.A., Schirmer, N.K., Sprinzl, M., Hilgenfeld, R., 1993. Crystal structure of active elongation factor Tu reveals major domain rearrangements. Nature 365, 126-132.

Brilot, A.F., Korostelev, A.A., Ermolenko, D.N., Grigorieff, N., 2013. Structure of the ribosome with elongation factor G trapped in the pretranslocation state. Proceedings of the National Academy of Sciences 110, 20994-20999.

Carneiro, S., Lourenço, A., Ferreira, E.C., Rocha, I., 2011. Stringent response of *Escherichia coli*: revisiting the bibliome using literature mining. Microb Inform Exp 1, 14-24.

Clementi, N., Chirkova, A., Puffer, B., Micura, R., Polacek, N., 2010. Atomic mutagenesis reveals A2660 of 23S ribosomal RNA as key to EF-G GTPase activation. Nat. Chem. Biol. 6, 344-351.

Connell, S.R., Takemoto, C., Wilson, D.N., Wang, H., Murayama, K., Terada, T., Shirouzu, M., Rost, M., Schuler, M., Giesebrecht, J., Dabrowski, M., Mielke, T., Fucini, P., Yokoyama, S., Spahn, C.M., 2007. Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. Mol Cell 25, 751-764.

Connell, S.R., Topf, M., Qin, Y., Wilson, D.N., Mielke, T., Fucini, P., Nierhaus, K.H., Spahn, C.M.T., 2008. A new tRNA intermediate revealed on the ribosome during EF4-mediated back-translocation. Nat Struct Mol Biol 15, 910-915.

Cornish, P.V., Ermolenko, D.N., Noller, H.F., Ha, T., 2008. Spontaneous intersubunit rotation in single ribosomes. Mol. Cell 30, 578-588.

Czworkowski, J., Wang, J., Steitz, T.A., Moore, P.B., 1994. The crystal structure of elongation factor G complexed with GDP, at 2.7 Å resolution. EMBO J. 13, 3661-3668.

Datta, P.P., Sharma, M.R., Qi, L., Frank, J., Agrawal, R.K., 2005. Interaction of the G′ Domain of elongation factor G and the c-terminal domain of ribosomal protein L7/L12 during translocation as revealed by cryo-EM. Mol. Cell 20, 723-731.

Daviter, T., Wieden, H.-J., Rodnina, M.V., 2003. Essential role of histidine 84 in elongation factor Tu for the chemical step of GTP hydrolysis on the ribosome. J Mol Biol 332, 689-699.

deLivron, M.A., Robinson, V.L. 2008. *Salmonella enterica serovar Typhimurium* BipA exhibits two distinct ribosome binding modes, pp. 5944-5952 J. Bacteriol., Vol. 190.

deLivron, M.A., Makanji, H.S., Lane, M.C., Robinson, V.L., 2009. A novel domain in translational GTPase BipA mediates interaction with the 70S ribosome and influences GTP hydrolysis. Biochemistry 48, 10533-10541.

Diaconu, M., Kothe, U., Schlünzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., Stark, H., Rodnina, M.V., Wahl, M.C., 2005. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. Cell 121, 991-1004.

Evans, R.N., Blaha, G., Bailey, S., Steitz, T.A., 2008. The structure of LepA, the ribosomal back translocase. Proc Natl Acad Sci USA 105, 4673-4678.

Farris, M., Grant, A., Richardson, T., O'Connor, C., 1998. BipA: a tyrosine-phosphorylated GTPase that mediates interactions between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells. Mol Microbiol 28, 265-279.

Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., Stark, H., 2010. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. Nature 466, 329-333.

Forro, J., Frederick, 1965. Autoradiographic studies of bacterial chromosome replication in amino-acid deficient *Escherichia coli 15T-*. Biophys. J. 5, 629-649.

Frank, J., Penczek, P., Grassucci, R., Srivastava, S., 1991. Three-dimensional reconstruction of the 70S *Escherichia coli* ribosome in ice: the distribution of ribosomal RNA. J. Cell Biol. 115, 597-605.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R.K., Agrawal, R.K. 1995. A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome, pp. 441-444 Nature, Vol. 376.

Freestone, P., Grant, S., Trinei, M., Onoda, T., Norris, V., 1998a. Protein phosphorylation in *Escherichia coli* L. form NC-7. Microbiology 144, 3289-3295.

Freestone, P., Trinei, M., Clarke, S.C., Nyström, T., Norris, V., 1998b. Tyrosine phosphorylation in *Escherichia coli*. J Mol Biol 279, 1045-1051.

Goldman, E., Jakubowski, H.Z., 1990. Uncharged tRNA, protein synthesis, and the bacterial stringent response. Mol Microbiol 4, 2035-2040.

Grant, A., Farris, M., Alefounder, P., Williams, P., Woodward, M., O'Connor, C. 2003. Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic *Escherichia coli* (EPEC). Mol Microbiol 48, 507-521.

Gromadski, K.B., Rodnina, M.V., 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. Mol. Cell 13, 191-200.

Haseltine, W.A., Block, R., 1973. Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes. Proc Natl Acad Sci USA 70, 1564-1568.

Huxley, H.E., Zubay, G., 1960. Electron microscope observations on the structure of microsomal particles from *Escherichia coli*. J Mol Biol 2, 10-IN18.

Jain, V., Kumar, M., Chatterji, D., 2006. ppGpp: stringent response and survival. J. Microbiol. 44, 1-10.

Kanjee, U., Ogata, K., Houry, W.A., 2012. Direct binding targets of the stringent response alarmone (p)ppGpp. Mol Microbiol 85, 1029-1043.

Kiss, E., Huguet, T., Poinsot, V., Batut, J., 2004. The typA gene is required for stress adaptation as well as for symbiosis of *Sinorhizobium meliloti* 1021 with certain *Medicago truncatula* lines. Mol. Plant Microbe Interact. 17, 235-244.

Konevega, A.L., Fischer, N., Semenkov, Y.P., Stark, H., Wintermeyer, W., Rodnina, M.V., 2007. Spontaneous reverse movement of mRNA-bound tRNA through the ribosome. Nature structural & molecular biology 14, 318-324.

Krishnan, K., Flower, A.M., 2008. Suppression of ΔbipA phenotypes in *Escherichia coli* by abolishment of pseudouridylation at specific sites on the 23S rRNA. J. Bacteriol. 190, 7675-7683.

Lake, J.A., 1976. Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. J Mol Biol 105, 131-159.

Li, W., Trabuco, L.G., Schulten, K., Frank, J., 2011. Molecular dynamics of EF-G during translocation. Proteins 79, 1478-1486.

Liljas, A., Ehrenberg, M., Åqvist, J., 2011. Comment on "The mechanism for activation of GTP hydrolysis on the ribosome". Science (New York, N.Y.) 333, 37-author reply 37.

Liu, H., Chen, C., Zhang, H., Kaur, J., Goldman, Y.E., Cooperman, B.S. 2011. The conserved protein EF4 (LepA) modulates the elongation cycle of protein synthesis. Proc Natl Acad Sci USA 108, 16223-16228.

Magnusson, L.U., Farewell, A., Nyström, T., 2005. ppGpp: a global regulator in *Escherichia coli*. Trends in Microbiology 13, 236-242.

March, P.E., Inouye, M., 1985a. Characterization of the *lep* operon of *Escherichia coli*. J. Biol. Chem. 260, 7206-7213.

March, P.E., Inouye, M., 1985b. GTP-binding membrane protein of *Escherichia coli* with sequence homology to initiation factor 2 and elongation factors Tu and G. Proc Natl Acad Sci USA 82, 7500-7504.

Margus, T., Remm, M., Tenson, T., 2007. Phylogenetic distribution of translational GTPases in bacteria. BMC Genomics 8, 15.

Merino, S., Tomás, J.M. 2001. Bacterial Capsules and Evasion of Immune Responses onlinelibrary.wiley.com. John Wiley & Sons, Ltd.

Mikolajka, A., Liu, H., Chen, Y., Starosta, A.L., Marquez, V., Ivanova, M., Cooperman, B.S., Wilson, D.N., 2011. Differential effects of thiopeptide and orthosomycin antibiotics on translational GTPases. Chem. Biol. 18, 589-600.

Moazed, D., Noller, H.F., 1989. Intermediate states in the movement of transfer RNA in the ribosome. Nature 342, 142-148.

Møller, A.K., Leatham, M.P., Conway, T., Nuijten, P.J.M., de Haan, L.A.M., Krogfelt, K.A., Cohen, P.S., 2003. An *Escherichia coli* MG1655 lipopolysaccharide deep-rough core mutant grows and survives in mouse cecal mucus but fails to colonize the mouse large intestine. Infect. Immun. 71, 2142-2152.

Nechifor, R., Murataliev, M., Wilson, K.S., 2007. Functional interactions between the G' subdomain of bacterial translation factor EF-G and ribosomal protein L7/L12. J. Biol. Chem. 282, 36998-37005.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A., 2000a. The structural basis of ribosome activity in peptide bond synthesis. Science (New York, N.Y.) 289, 920-930.

Nissen, P., Kjeldgaard, M., Nyborg, J., 2000b. Macromolecular mimicry. EMBO J. 19, 489-495.

Ogle, J.M., Murphy, F.V., Tarry, M.J., Ramakrishnan, V., 2002. Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell 111, 721-732.

Owens, R.M., Pritchard, G., Skipp, P., Hodey, M., Connell, S.R., Nierhaus, K.H., O'Connor, C.D., 2004. A dedicated translation factor controls the synthesis of the global regulator Fis. EMBO J 23, 3375-3385.

Palade, G.E., 1955. A small particulate component of the cytoplasm. The Journal of Biophysical and Biochemical Cytology 1, 59-68.

Payoe, R., Fahlman, R.P., 2011. Dependence of RelA-mediated (p)ppGpp formation on tRNA identity. Biochemistry 50, 3075-3083.

Pfennig, P., Flower, A., 2001. BipA is required for growth of *Escherichia coli* K12 at low temperature. Mol Genet Genomics 266, 313-317.

Pulk, A., Cate, J.H.D., 2013. Control of ribosomal subunit rotation by elongation factor G. Science (New York, N.Y.) 340, 12359-70.

Qi, S.-Y., Li, Y., Szyroki, A., Giles, I., Moir, A., O&apos;Connor, C.D. 1995. *Salmonella typhimurium* responses to a bactericidal protein from human neutrophils. Mol Microbiol 17, 523-531.

Qin, Y., Polacek, N., Vesper, O., Staub, E., Einfeldt, E., Wilson, D.N., Nierhaus, K.H., 2006. The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosome. Cell 127, 721-733.

Radermacher, M., Wagenknecht, T., Verschoor, A., Frank, J., 1987. Three-dimensional structure of the large ribosomal subunit from *Escherichia coli*. EMBO Journal 6, 1107-1114.

Rodnina, M.V., Savelsbergh, A., Katunin, V.I., Wintermeyer, W. 1997. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. Nature 385, 37-41.

Rodnina, M.V., Wintermeyer, W., Green, R. 2011. Ribosomes. Springer Vienna.

Rowe, S., Hodson, N., Griffiths, G., Roberts, I.S., 2000. Regulation of the *Escherichia coli* K5 capsule gene cluster: evidence for the roles of H-NS, BipA, and integration host factor in regulation of group 2 capsule gene clusters in pathogenic *E. coli*. J. Bacteriol. 182, 2741-2745.

Ruusala, T., Ehrenberg, M., Kurland, C.G., 1982. Is there proofreading during polypeptide synthesis? The EMBO journal 1, 741-745.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., Yonath, A., 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. Cell 102, 615-623.

Scott, K., Diggle, M., Clarke, SC., 2003. TypA is a virulence regulator and is present in many pathogenic bacteria. Br J Biomed Sci 60, 168-170.

Shi, X., Khade, P.K., Sanbonmatsu, K.Y., Joseph, S., 2012. Functional role of the sarcin-ricin loop of the 23S rRNA in the elongation cycle of protein synthesis. J Mol Biol 419, 125-138.

Tourigny, D.S., Fernandez, I.S., Kelley, A.C., Ramakrishnan, V., 2013. Elongation factor G bound to the ribosome in an intermediate state of translocation. Science (New York, N.Y.) 340, 1235490.

Trobro, S., Åqvist, J., 2005. Mechanism of peptide bond synthesis on the ribosome. Proc Natl Acad Sci USA 102, 12395-12400.

Valle, M., Zavialov, A., Li, W., Stagg, S.M., Sengupta, J., Nielsen, R.C., Nissen, P., Harvey, S.C., Ehrenberg, M., Frank, J., 2003a. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. Nat. Struct. Biol. 10, 899-906.

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., Frank, J., 2003b. Locking and unlocking of ribosomal motions. Cell 114, 123-134.

Villa, E., Sengupta, J., Trabuco, L.G., LeBarron, J., Baxter, W.T., Shaikh, T.R., Grassucci, R.A., Nissen, P., Ehrenberg, M., Schulten, K., Frank, J., 2009. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. Proceedings of the National Academy of Sciences of the United States of America 106, 1063-1068.

Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C., Ramakrishnan, V., 2009. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat Struct Mol Biol 16, 528-533.

Voorhees, R.M., Schmeing, T.M., Kelley, A.C., Ramakrishnan, V., 2010. The mechanism for activation of GTP hydrolysis on the ribosome. Science (New York, N.Y.) 330, 835-838.

Wahl, M.C., Moller, W., 2002. Structure and function of the acidic ribosomal stalk proteins. Curr. Protein Pept. Sci. 3, 93-106.

Walter, J.D., Hunter, M., Cobb, M., Traeger, G., Spiegel, P.C., 2012. Thiostrepton inhibits stable 70S ribosome binding and ribosome-dependent GTPase activation of elongation factor G and elongation factor 4. Nucleic Acids Res. 40, 360-370.

Wang, F., Zhong, N.-Q., Gao, P., Wang, G.-L., Wang, H.-Y., Xia, G.-X., 2008. SsTypA1, a chloroplast-specific TypA/BipA-type GTPase from the halophytic plant *Suaeda salsa*, plays a role in oxidative stress tolerance. Plant Cell Environ. 31, 982-994.

Yamamoto, H., Qin, Y., Achenbach, J., Li, C., Kijek, J., Spahn, C.M.T., Nierhaus, K.H., 2014. EF-G and EF4: translocation and back-translocation on the bacterial ribosome. Nat. Rev. Microbiol. 12, 89-100.

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

CHAPTER 3:

Image Processing of the BipA Dataset and Its Use as a Testbed for New Automated Particle Picking and Classification Techniques

CHAPTER 3 ABBREVIATIONS

Abbreviation	Full Title
2D	two-dimensional
3D	three-dimensional
ATP	adenosine triphosphate
CC	cross-correlation
CCF	cross-correlation function
EF-G	elongation factor G
EF-Tu	elongation factor Tu
GAC	GTP-associated center
MAP	*maximum a posterior*
ML	maximum likelihood
NP	new particle
RELION	REgularised LIkelihood OptimisatioN
SPR	single-particle reconstruction
tRNA	transfer RNA
ViCer	View Classifier
µm	micrometer or micron

3.1 INTRODUCTION

Following the protocol of single-particle reconstruction detailed in Chapter 1, a single 3D reconstruction requires thousands, if not hundreds of thousands, of particle projection images that must each be found and picked from the larger micrographs. The particle-picking problem can be divided into two steps: particle selection and particle verification. In the first step, candidate particles must be recognized and windowed from the larger micrograph. This task is made more difficult by inherent high levels of noise and the low contrast of the biological particles in the micrograph. Thus, the candidate particle dataset will comprise of true particles and contaminants (i.e. ice crystals, dust, and other features that by their size can be mistaken for particles – we call them "nonparticles"). At the same time, some true particle may have been missed by the particle selection procedure.

In the second step, the set of candidate particles must be verified to ensure that particle images entering the alignment and reconstruction phases are indeed true particles and not contaminants. This second step has posed a bottleneck for the single-particle reconstruction (SPR) workflow as researchers have had to visually inspect candidate particles and visually decide the veracity of each. Recently, the new algorithm AutoPicker (Langlois et al., 2014b), discussed in Chapter 1.3.5, was introduced as an unsupervised method for particle picking and verification.

After candidate particles have been verified, the dataset can be used for reconstruction. Initially, the underlying assumption of the reconstruction technique is that dataset is comprised of structurally identical particles. However, as discussed in Chapter 1.3.6, this assumption rarely holds true and classification techniques must be employed to separate out different conformations that may inhabit the sample. Traditionally, supervised classification or unsupervised classification techniques (i.e. RELION (Scheres, 2012a)) have been employed to separate the larger dataset into subsets of more homogenous particles.

Over the course of the last few years, the introduction of automated selection and verification programs such as AutoPicker and the classification program RELION has revolutionized the workflow of the SPR technique. Here in our lab, the BipA dataset was used as one of the first testbeds for the AutoPicker algorithm. I was fortunate to be a part of this project and helped optimize AutoPicker parameters and design the layout and ergonomics of the accompanying visualization program *Ara-Display* and the larger image processing suite Arachnid (Langlois, R., Ho, D., deGeorges, A., Frank, J. *in prep*). Thus, this chapter not only serves to sketch out the methods I used to obtain the final reconstruction of the 70S–BipA complex, but also provides an outline of the recent improvements to the SPR workflow. Characterization of the BipA data is given in Section 3.2. In Section 3.3, the candidate particle datasets used for manual verification and AutoPicker are characterized, respectively. In Section 3.4, the manually verified BipA dataset is compared with the AutoPicker BipA dataset. In Section 3.5, the supervised classification technique is introduced and its application to the BipA dataset is presented. In Section 3.6, the program RELION, a novel unsupervised classification method (Scheres, 2012a), and its application in classifying of the BipA dataset, are discussed. The chapter culminates in a description of the reconstruction of the 70S–BipA complex, resolved to a resolution of 8.5 Å, to be fully characterized in Chapter 4.

3.2 CHARACTERIZATION OF THE BIPA DATASET AND THE CHOICE OF VARIOUS PROCESSING PARAMETERS

Preparation of the 70S–BipA sample (*in vitro* reactions and cryo-EM grid preparation) is outside the scope of this chapter and will be detailed in Chapter 4. In all, 434 film micrographs were recorded on the FEI Tecnai F30 Polara electron microscope (FEI, Eindhoven) with Kodak Electron

SO-163 Image Film at a low dose (~18e$^-$/Å^2) and nominal 59,000x magnification setting. At the calibrated magnification of 58,269x (calibrated by Robert Grassucci on September 28, 2009; the data were collected on February 25, 2010), the pixel size is 1.2 Å at the specimen level when digitized using the 16-bit ZI Imagine Photoscan 2000 densitometer (Z/I Imagine, Aalen, Germany) at a sampling rate of 7 µm.

The data were decimated by a factor of two to increase the signal-to-noise ratio and improve the visibility of the power spectrum for CTF computation, giving a final pixel size of 2.4 Å. The power spectrum of each micrograph was computed using the program SPIDER (Frank et al., 1981; Frank et al., 1996). Visual inspection of the power spectra allowed quality assessment of the micrographs. Those micrographs with power spectra showing significant amounts of drift (see Figure 1.7B) were discarded, leading to a final film dataset of 312 micrographs.

The 70S ribosome is estimated to be approximately 250 Å in diameter. In the micrograph, at a pixel size of 2.4 Å, this corresponds to a ~105 pixel diameter for the 70S ribosome. Practical usage indicates that the particle of interest should inhabit about 65-70% of the final window size. Thus, the particle window size was designed as 160x160 pixels. The final three-dimensional map of the 70S–BipA complex, presented in Chapter 4, was reconstructed using undecimated data.

3.3 CHARACTERIZATION OF THE MANUALLY VERIFIED AND AUTOPICKER DATASET

3.3.1 The Particle Picking Algorithm, LFC-Pick

The program suite SPIDER (Frank et al., 1981) was used to process the micrographs. Specifically for particle picking and windowing, the procedure *LFC-Pick* was employed. This procedure utilizes a 2D template for finding candidate particles via the local normalized cross-correlation algorithm described by Alan Roseman (Roseman, 2003) and implemented in SPIDER by Rath and coworkers (Rath and Frank, 2004). In the procedure, a 2D projection is simulated from a given 3D reference. A circular mask with a user-defined diameter pertaining to the estimated particle diameter, is applied to the simulated image, producing the final 2D template. The user defines the window size of the isolated particle images.

In *LFC-Pick*, computation of the CCF between the template and the entire micrograph is followed by peak search to find candidate particles. As described in Chapter 1, a peak in the CCF indicates a positions where the micrograph resembles the template. The peak search finds all candidate particles that have the highest correlation, and consequently the highest CCF peaks, to the 2D template. To account for the possible fluctuation in beam illumination and consequently contrast in local areas of the micrograph, normalization is applied under the footprint of the 2D template mask. One can take the fast Fourier transform of the experimental micrograph area underneath the template mask to estimate the area's pixel density variance. *LFC-Pick* crops particle windows and places them in stacks, one per micrograph. Particles are stacked in descending order according to their cross-correlation value to the 2D template.

For the BipA dataset, a reconstruction of an empty 70S ribosome, provided in-house by Dr. Wen Li, was used as the 3D reference to generate the 2D template. Template matching and particle cropping with *LFC-Pick* yielded a complete candidate particle dataset of 503,592 images.

Each particle image stack had to be manually inspected and true particle selected from the gallery of candidate particles. I used the SPIDER-associated display program WEB (Frank et al., 1996) to visually inspect each of the 312 particle image stacks, which displayed the particles in descending order of cross-correlation values. Both highest- and lowest-ranked particles tended to be non-particles, as shown in Figure 3.1. Sharp, high-contrast edges are typical features of ice contaminants and other non-particles. Particle images showing these artifacts were immediately rejected from the dataset. Many non-particle images were found dispersed throughout the gallery of candidate particles, interspersed among good, true particles (Figure 3.1).

LFC-Pick found an overwhelming number of candidate particles, many of which were simply images containing pure noise. The transition from good particles to such noise images is readily visible in micrographs imaged at far-from-focus, defined here as a defocus > 3.0 µm, but more difficult to discern in close-to-focus micrographs (defocus < 2 µm), as low-defocus particles have, as a rule, low contrast. Thus, there were fewer particles verified in defocus groups at close-to-focus than defocus groups at far-from-focus. This distinction will become important in the comparison with the performance of the AutoPicker program, as discussed below. 133,782 particle images comprised the final manually verified dataset. The manual verification process requires considerable time investment by the researcher. In the case of the BipA manually verified dataset, the entire process was completed in approximately 4.5 weeks.

3.3.2 The AutoPicker-Verified Dataset

By design, the AutoPicker algorithm seeks to eliminate the time investment and user subjectivity of the manual verification process. Like *LFC-Pick*, the procedure also finds and crops candidate particles from the micrograph using a template-matching algorithm. However, instead of using a 3D reference to generate the 2D template, AutoPicker uses a Gaussian blurred disk as 2D template, an option also available in *LFC-Pick,* but employs additional tools, as detailed below. The radius of the disk is the estimated radius of the investigated particle. The algorithm is particularly powerful when paired with the visualization program *Ara-Display*, part of the Arachnid suite (Langlois, R., Ho, D., deGeorges, A., Frank, J. *in prep*), which I helped to design. The program allows the user to visualize the entire micrograph with the windowed candidate particles marked. In this way, the user can visually determine, for example, whether the selected particles may be contaminants or part of a cluster of aggregated particles.

Unlike *LFC-Pick*, the AutoPicker algorithm has additional tools which are able to 1) differentiate between true particles and non-particles and 2) find an appropriate threshold by which to reject the non-particles. After candidate particles have been found, principal component analysis (PCA) is employed on power spectra of particle windows. Assuming a Gaussian distribution, AutoPicker immediately rejects particle images that are extreme outliers (more than 4 standard deviations from the mean). This step removes the most obvious high-contrast non-particles (i.e., the first few particles shown in Figure 3.1C). Similarity between the remaining windows and the 2D template is measured via the CCF, and the results are graphed as a histogram of CC scores. The optimal threshold to separate particles from non-particles is found using a cutoff algorithm described in (Otsu, 1979). Here, the algorithm looks for an obvious indicator, in terms of the CC histogram, that marks the transition from true particles to non-particles.

For the BipA data set, the entire AutoPicker procedure was run on 312 micrographs in under four hours without user intervention, giving a final dataset of 293,036 particles. As shown in Figure 3.2, the AutoPicker-verified dataset shows no signs of the high contrast non-particles that plagued the datasets generated by *LFC-Pick*. The AutoPicker dataset was classified using RELION (Scheres, 2012a), to be discussed below in Section 3.6.

3.4 COMPARISON OF THE MANUALLY VERIFIED WITH THE AUTOPICKER DATASET

3.4.1 Scheme of Comparison

The AutoPicker algorithm has been previously benchmarked against two manually verified datasets, one of the 70S ribosome and the other of the ATP Synthase (Langlois et al., 2014b). For the 70S ribosome dataset, the particles were selected by *LFC-Pick* and then manually verified by the authors. Manual verification was considered the "gold standard" technique for particle verification. Thus, manually verified particles were considered true particles (true positives).

Particle coordinates after manual verification and AutoPicker were compared to determine the performance of AutoPicker against manual verification. Three criteria were used to gauge AutoPicker's performance: 1) the recall fraction, 2) the new particle (NP) fraction, and 3) the quality of the reconstructions from each dataset. The recall fraction is the fraction of manually verified particles that were also found by AutoPicker. The NP fraction is the fraction of total AutoPicker particles that were missed by manual verification. For calculating the recall and NP fractions, defocus groups were created to combine information from different micrographs of similar defocus.

In the benchmark study (Langlois et al., 2014b), of which I am a co-author, we observed that AutoPicker was able to consistently pick a vast majority (>85%) of the same particles found by manual verification, regardless of the defocus. However, the performance at low defocus and high defocus differed, a point to be discussed below. Additionally, AutoPicker picked many more particles than manual verification. Whether the additionally picked particles likely to be are true particles can be gauged by reconstructing a density map using the entire enlarged particle dataset. We showed that both the 70S ribosome and ATP Synthase reconstructions using particles selected by AutoPicker had resolutions equal or better than those of reconstructions using manually verified particles.

3.4.2 The Recall and New Particle Fractions of the AutoPicker Dataset

For the BipA dataset, the same statistical comparisons can be made between the manually verified dataset of 132,264 particles and the AutoPicker dataset of 293,036 particles. In terms of the quantity of particles picked, AutoPicker consistently picks approximately twice as many particles as manual verification across the entire defocus range. Also regardless of defocus, AutoPicker has a very high recall fraction (>.82 in all but one defocus group). This means that AutoPicker is able find the vast majority of the same particles that were manually verified.

The small fraction of manually verified particles that were not found by AutoPicker could either be non-particles erroneously chosen by me, or true particles that AutoPicker could not find. The former possibility is the likelier situation because the recall fraction rate increases with defocus. As discussed, particles at high defocus (>3 µm) have higher contrast than particles at low defocus. Thus, at high defocus, true particles are more easily visually discernable, resulting in a higher percentage of true particles in the manually verified high defocus groups. A very high recall fraction then means that AutoPicker has a strong ability to pick true particles.

All AutoPicker particles that were not found by manual verification are particles that may be new true particles missed by manual verification or non-particles erroneously chosen by AutoPicker. Both scenarios seem likely. Visual inspection of the AutoPicker dataset still shows signs of contaminants, although high-contrast non-particles have been rejected. Overall, the NP fraction decreases with increasing defocus. At low defocus, approximately 66% of the AutoPicker particles are new. At high defocus, this fraction falls to approximately 55%, indicating greater agreement between the two datasets as defocus increases. Thus, at low defocus, I may have missed selecting true particles and rejecting them as non-particles during the manual verification process. The benchmark study (Langlois et al., 2014b) showed the same trend for the 70S ribosome dataset, with the same interpretation. As the defocus increases, it is expected that the agreement between the AutoPicker dataset and the manually verified dataset will increase as particles become easier to discern by the human eye.

Overall, the recall fractions suggest that the AutoPicker dataset contains almost all the particles of the manually verified dataset, indicating AutoPicker's strong ability to find true particles. Combined with the NP fractions trend, this suggests that the AutoPicker dataset contains additional true particles missed by me. The ability to find true particles is important, especially at low defocus, where valuable high-resolution information about the investigated macromolecule is found. Additionally, the use of AutoPicker reduced the time investment from over four weeks to a few hours, and all but removes the bottleneck imposed by manual verification. This improvement moves the entire cryo-EM workflow closer to becoming a streamlined, automated high throughput process.

3.5 SUPERVISED CLASSIFICATION

As discussed in Chapter 1 (Sections 1.2.2 and 1.3.6), an assumption of the SPR technique is that the dataset is comprised of structurally identical particles. However, this is rarely the case in cryo-EM studies: heterogeneity, both compositional and conformational, often exists. Disentangling the different conformations that may exist in a sample is a necessity for meaningful interpretation of the data.

3.5.1 The Supervised Classification Technique

In the supervised classification technique (Valle et al., 2002; Gao et al., 2004), the particle dataset is divided into more homogenous subsets according to each particle's resemblance to two 3D references. For each 3D reference, a set of simulated projections of equally spaced views in angular space, are generated. Two rounds of reference-based projection alignment (Penczek et al., 1994) is applied to the particle dataset, once for each reference. Each particle image in the dataset is matched, using the cross-correlation function (CCF), with the highest-correlated simulated projection view image for each reference. Thus, each particle image is assigned two cross-correlation values, CC1 and CC2, one for each reference. The difference in CC value (ΔCC = CC2-CC1), allows the researcher to gauge how well the particle image is differentiated between the two references. A positive ΔCC value means that the particle has higher resemblance to a simulated projection from reference 2, while a negative value means that the particle has greater correlation with reference 1. ΔCC can be graphed as a histogram, with a zero center value indicating particle images that have equal correlation with each reference.

This classification technique has several problems. First, the choice of references requires *a priori* information about the structure of the investigated complex. Second, non-particles that have subsisted in the dataset even after verification will not align substantially to either reference, although a carefully chosen threshold range may exclude these images. Finally, while theoretically the classification should produce a bimodal distribution of resemblance, allowing for easy division of particles, the result is often unimodal (Gao et al., 2004; Scheres et al., 2007). The researcher must necessarily choose a ΔCC cutoff threshold range, defined as X ≤ ΔCC ≤ Y, where 'X' is the minimum allowed ΔCC value and 'Y' is the maximum allowed ΔCC value. All particles with a ΔCC value falling within this range are entered into a data subset and a density map is reconstructed from the subset. The choice of threshold range, as discussed below, is highly subjective and the optimal range is only found by visually inspecting multiple density maps reconstructed using different threshold ranges.

Model bias of the final reconstruction towards the reference can be assessed during the iterative 3D reconstruction process. Here, a low-quality 3D reference with no factor bound, such as an empty 70S ribosome, can be used for reference-based alignment and orientation determination in the first round of refinement. In the following rounds in this iterative process, if a density for the bound factor appears, one can be assured that additional density originates from information in the particle images and not from the reference used.

3.5.2 Supervised Classification of the BipA Manually Verified Dataset

For supervised classification, reconstructions of an empty 70S ribosome and a 70S–EF-G complex were chosen as reference 1 and reference 2, respectively, for reference-based alignment. The references are shown in Figure 3.4. The 70S–EF-G complex was chosen as a reference because of the similarity, both in structure and sequence, between BipA and EF-G. Also, as discussed in Section 2.4, biochemical studies pinpointed BipA's binding site near the GTPase-associated center (GAC) of the 70S ribosome, akin to the binding site of EF-G. Thus, the 70S–BipA complex was assumed to have overall structure resemblance to the 70S–EF-G complex. Both references were filtered to ~12.5 Å prior to the start of reference-based alignment. After alignment, every particle image was assigned two CC values, CC1 and CC2, corresponding to the particle image's resemblance to the empty 70S reference and the 70S–EF-G reference, respectively.

The normalized histogram of the distribution of resemblance, shown in Figure 3.4, is graphed with the number of particles on the y-axis and the difference of CC values (ΔCC = CC2-CC1) on the x-axis. The histogram shows a unimodal distribution of resemblance. Many particles had equal or near-equal resemblance to both references. There could be a variety of reasons for a unimodal distribution. Particle images with high amounts of drift, particle images of pure noise, and non-particles, will have a difficult time aligning to either reference, resulting in a very low ΔCC value. Most likely, due to resolution limitations, the difference in cross-correlation values might be in the order of the noise. In that case, the underlying bimodal distribution will blend into one unimodal one.

With this idea in mind, six cutoff thresholds, shown in Figure 3.4B, were chosen. The corresponding particle statistics and resolution of the subsequent respective reconstructions are also given. For cutoff #1, with a threshold of $-1 \leq \Delta CC \leq 1$, the corresponding reconstruction, resolved to 9.5 Å, used the entire dataset of 133,782 particles (Figure 3.5). While the cutoff #1 reconstruction had the highest resolution of any cutoff reconstruction, the map showed no density for BipA, expected to be bound proximal to the A site. Reconstructions using thresholds of only negative or positive ΔCC, (cutoffs #2 and #3) depicted a 70S ribosome with P-site tRNA occupancy and a 70S ribosome with marginal prospective density for BipA, respectively. Thus, the subsequent three choices of cutoff thresholds (#4 - #6) aimed to optimize the visibility of BipA in the 70S–BipA complex by incrementally limiting the acceptable ΔCC range.

The best reconstruction of the 70S–BipA complex was obtained from cutoff #6. While this reconstruction had the lowest resolution, ~13 Å, the BipA density is strong and fully realized. The reconstruction shows clear density of the protein proximal to the A site at the GAC, as expected from previous biochemical data. Unexpected was an additional density for an A-site tRNA and a slightly weaker density for a P-site tRNA. A potential, scattered mass of density for the E-site tRNA was also detected. Collectively, the cutoff reconstructions show that while imperfect, supervised classification is capable of separating the particles into more homogenous subsets for ligands of this size. Unfortunately, further analysis of the 70S–BipA complex was hindered by the low resolution of the reconstruction, which did not meet the criteria for MDFF fitting and modeling.

3.5.3 Moving Forward with the BipA AutoPicker Dataset

With AutoPicker's introduction, the BipA dataset was primed to be used as a testbed for the new algorithm. However, because the cryo-EM results for the 70S–BipA complex were unpublished, the data were not included in the benchmark study (Langlois et al., 2014b). Nevertheless, the testing I performed with the AutoPicker algorithm helped to optimize AutoPicker parameters and prompted the subsequent development of a new post-particle selection cleaning procedure, called ViCer (Langlois et al., 2014a).

Analysis of the AutoPicker dataset statistics compelled me to abandon the manually verified dataset. The suboptimal resolutions of the reconstructions from the supervised classification technique also compelled me to find new methods of classification. Thus, I used the newly introduced program RELION (Scheres, 2012a) for unsupervised classification of the AutoPicker dataset.

3.6 UNSUPERVISED CLASSIFICATION USING RELION

3.6.1 RELION: Implementation of the Maximum a Posterior (MAP) Approach to Cryo-EM Data

In cryo-EM we seek the solution of Θ, an unknown 3D reconstruction, based on χ, the collection of particle images (nomenclature of variables as introduced in Scheres, 2012b). By using the regularized likelihood optimization, also known as *maximum a posterior* (MAP) estimation, we seek to optimize the probability of observing the 3D model(s) (Θ) given the collected particle dataset (χ), which is defined by the term 'posterior.' The posterior is proportional to the maximum likelihood (ML) function (Sigworth et al., 2010) regularized by prior information (Scheres, 2012b). We often have additional prior information on the biological sample that can be given as constraints or regularization parameters on the estimator. A mathematical treatment of the MAP estimator is beyond the scope of this dissertation, but can be found in (Scheres, 2012b). The classification program RELION (Scheres, 2012a) implements the MAP for cryo-EM structure determination in the presence of heterogeneity.

In the initial setup of RELION, the user chooses the expected number of classes, designated as the parameter *K,* and provides a low-resolution (<60 Å) 3D reference, often a known reconstruction devoid of bound exogenous factors. RELION then uses an iterative process to optimize the posterior of the density maps of the K classes. In the first iteration of classification, the entire particle dataset is randomly divided evenly into the designated number of K classes. The low-resolution 3D reference is used for an initial alignment of the each subset of particles and then an initial set of K reconstructions are obtained. In the Fourier implementation of the algorithm (Scheres, 2012b), each particle in the entire dataset is then compared to Fourier central slices of each of the K reconstructions using a likelihood function, which gives each particle a likelihood probability distribution of belonging to each view of each class. This process is similar to reference-based orientation determination, which uses the cross-correlation function and not a likelihood function; however, an important difference is that an entire probability distribution is computed, taking into account the likelihood of the particle to belong to every simulated class view, rather than a single set of parameters assigned.

At the end of each iteration, new density maps are reconstructed for each class. All particle images contribute to the reconstruction for each class. However, each particle image's contribution is weighted proportionally based on their likelihood probability and prior information. The process starts anew in the next iteration, with the previous iteration's K reconstructions given as prior information. The classification, with finer angular sampling from iteration to iteration, seeks to find the most probable K reconstructions based on the observed data and available prior information.

3.6.1 Disentangling Conformations in the BipA Dataset Using RELION

The total Autopicker dataset consisted of 293,036 particles. RELION (Scheres, 2012a) was employed for three-dimensional particle classification and reconstruction, while the classification scheme, shown in Table 3.1 and Figure 3.6, was adapted from earlier work in this lab (Hashem et al., 2013). The initial round of 3D classification was performed to separate genuine particles from non-particles using a rough angular sampling of 15 degrees. Non-particle classes are recognizable by showing scattered, noisy density. Thus, RELION has the additional advantage of being able to distinguish non-particles from true particles, facilitating an additional particle clean-up step.

Iterative rounds of RELION, with gradually finer angular sampling and corresponding local search ranges, were employed to classify different ribosomal conformations and binding states. At the end of each round, classes deemed visually inconsistent with the appearance of a BipA-bound 70S complex were rejected and the corresponding particles were removed from the dataset in the next round of classification. Starting in Round 3, a class of an obvious, but low-quality 70S ribosome began to appear in the classification. These classes, seen also in later rounds, contain relatively few particle assignments. Particle images that make up these classes may have greater levels of noise or may contain higher levels of drift than other particle images, resulting in their assignment to a different low-quality 70S class.

Overall, five distinct ribosomal classes were observed: an empty 70S (21,540 particles); 70S with E-site tRNA occupancy (36,775 particles); 70S with P-site tRNA occupancy (57,295 particles); 70S with E-site and P-site tRNA occupancy (28,992 particles); and 70S with BipA as well as A- and P-site tRNAs (22,938 particles), designated as the 70S–BipA complex of interest to be discussed in this paper. While all five classes are shown in Figure 3.7, only the BipA-bound class was further refined, to give the final reconstruction. The cryo-EM reconstruction of the 70S–BipA complex, presented in Chapter 4, is resolved to a resolution of 8.5 Å as assessed using the gold standard protocol (Liao and Frank, 2010; Henderson et al., 2012).

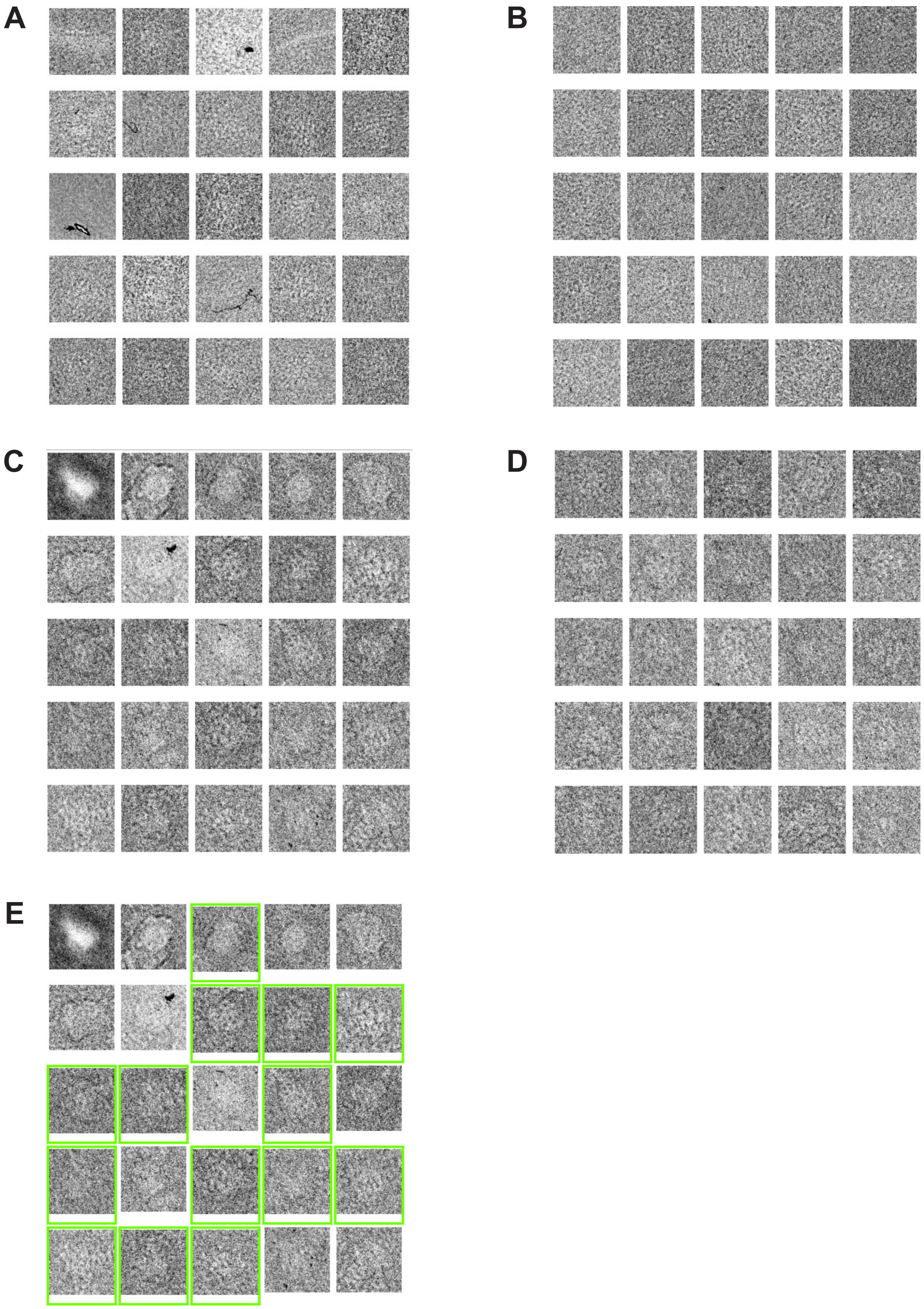


Figure 3.1. Galleries of *LFC-Pick* particles and manually verified particles. The *LFC-Pick* procedure was employed in SPIDER (Rath and Frank, 2004) to select and isolate candidate particle windows. (A) and (B) show candidate particles 1-25 and 400-425, respectively, from Micrograph 69 with a defocus of 1.5 µm. (C) and (D) show candidate particles 1-25 and 400-425, respectively, from Micrograph 188, with a defocus of 2.5 µm. (E) shows the manually verified particles from particles 1-25 of Micrograph 188, boxed in green.

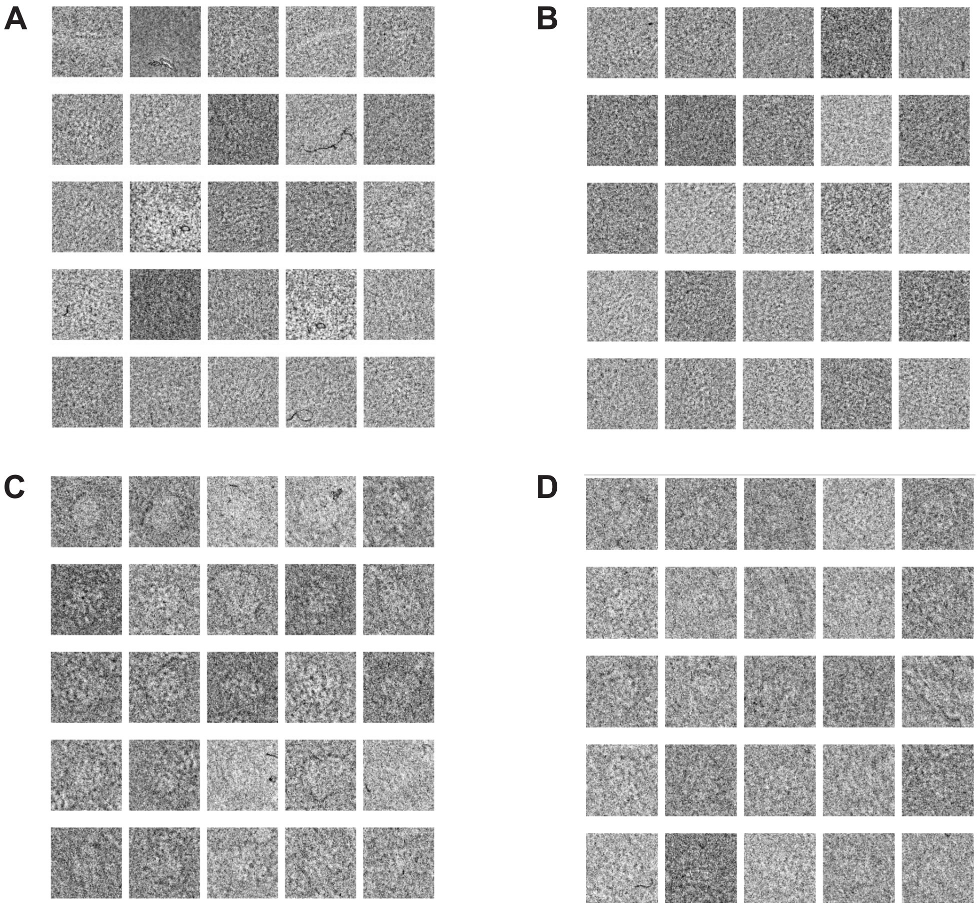


Figure 3.2. Galleries of AutoPicker candidate particles. The AutoPicker procedure (Langlois, R. et al, 2014a) was employed to select and isolate candidate particle windows. (A) and (B) show candidate particles 1-25 and 400-425, respectively, from Micrograph 69, with a defocus of 1.5 μm. (C) and (D) show candidate particles 1-25 and 400-425, respectively, from Micrograph 188, with a defocus of 2.5 μm.

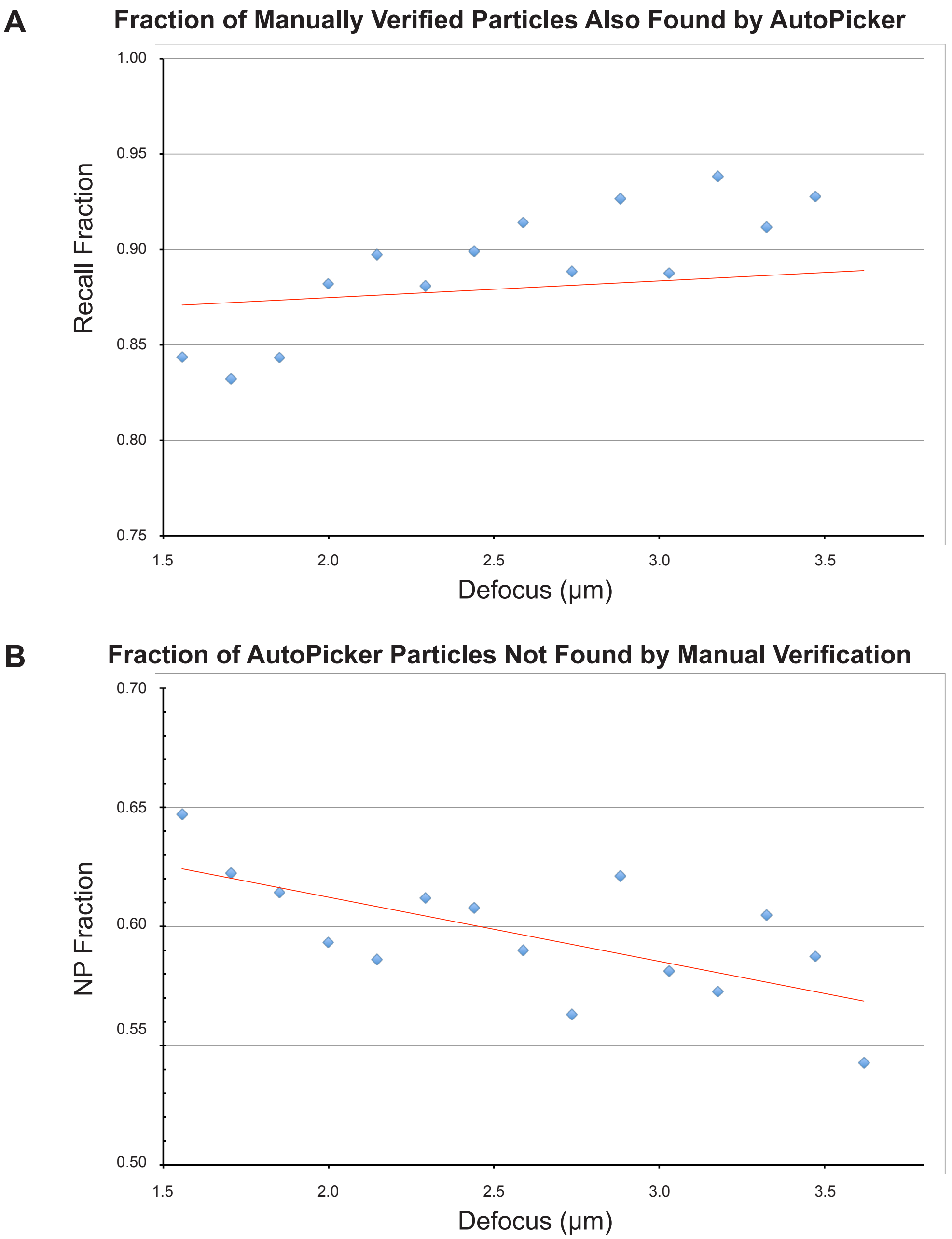
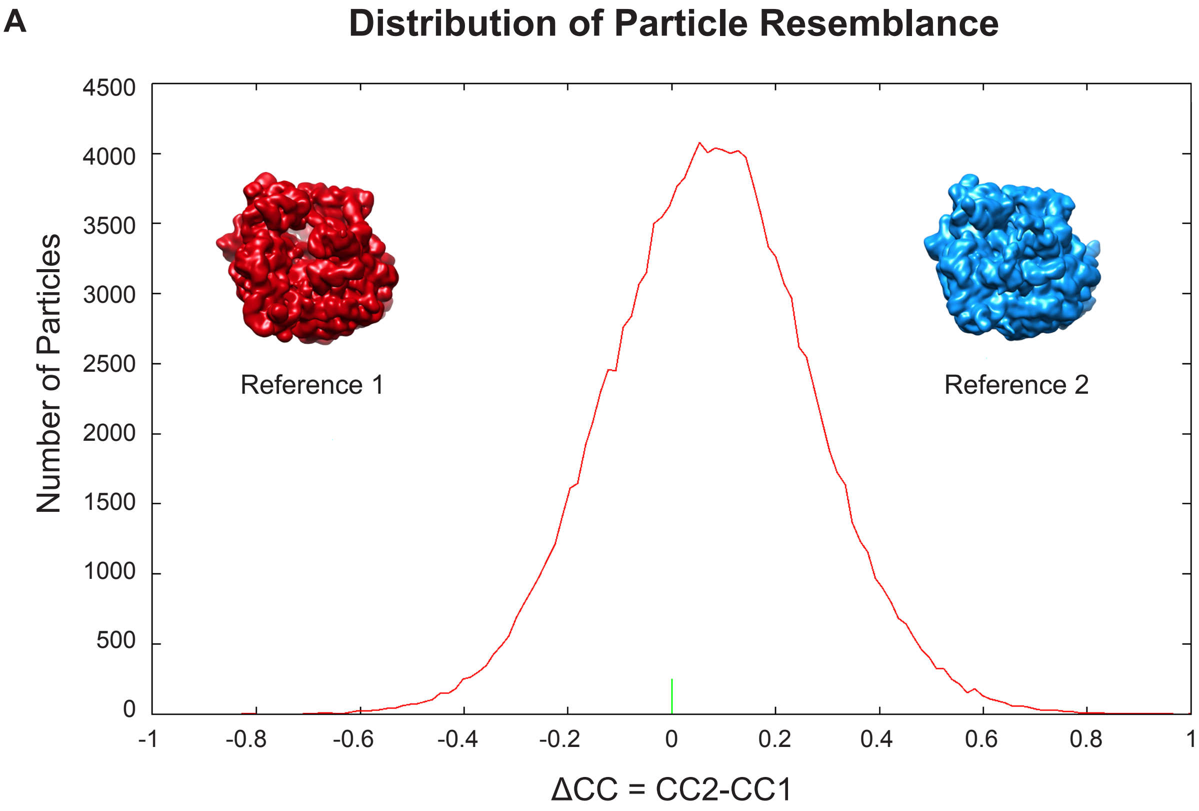


Figure 3.3. The recall and new particle (NP) fraction values from a comparison of the manually verified and AutoPicker datasets. 312 micrographs were divided into 15 defocus groups. (A) shows the fraction of manually verified particles, per defocus group, that were also found by AutoPicker, designated the recall fraction. The recall trend line, in red, displays an increasing slope, indicating that there is greater agreement between manually verification and AutoPicker as the defocus increases. (B) shows the fraction of AutoPicker particles, per defocus group, that were not found by the manual verification technique, designated the new particle (NP) fraction. The NP trend line, in red, displays a decreasing slope. This is also indicative of greater agreement between manually verification and AutoPicker as the defocus increases.



B

CUTOFF RECONSTRUCTION #	CUTOFF MIN	CUTOFF MAX	NO. OF PARTICLES	RESOLUTION (Å)
1	-1	1	133782	9.52
2	-1	0	45489	11.07
3	0	1	88331	9.91
4	0.1	1	62560	10.71
5	0.2	1	35673	11.39
6	0.26	1	23491	13.10

Figure 3.4. Results of supervised classification of the BipA manually verified dataset. (A) The normalized ΔCC = CC2-CC1, displaying the distribution of resemblance for the BipA manually verified particle dataset. The two density maps shown are the references used for supervised classification: am empty 70S ribosome (reference 1) and a 70S–EF-G complex (reference 2), in red and blue, respectively. Use of the technique did not generate a bimodal histogram, showing that the classification was imperfect. (B) A table displaying the statistics of the various ΔCC cutoff threshold ranges used to divide the particles into more homogeneous subsets. Reconstructions for each cutoff threshold range are shown in Figure 3.5.

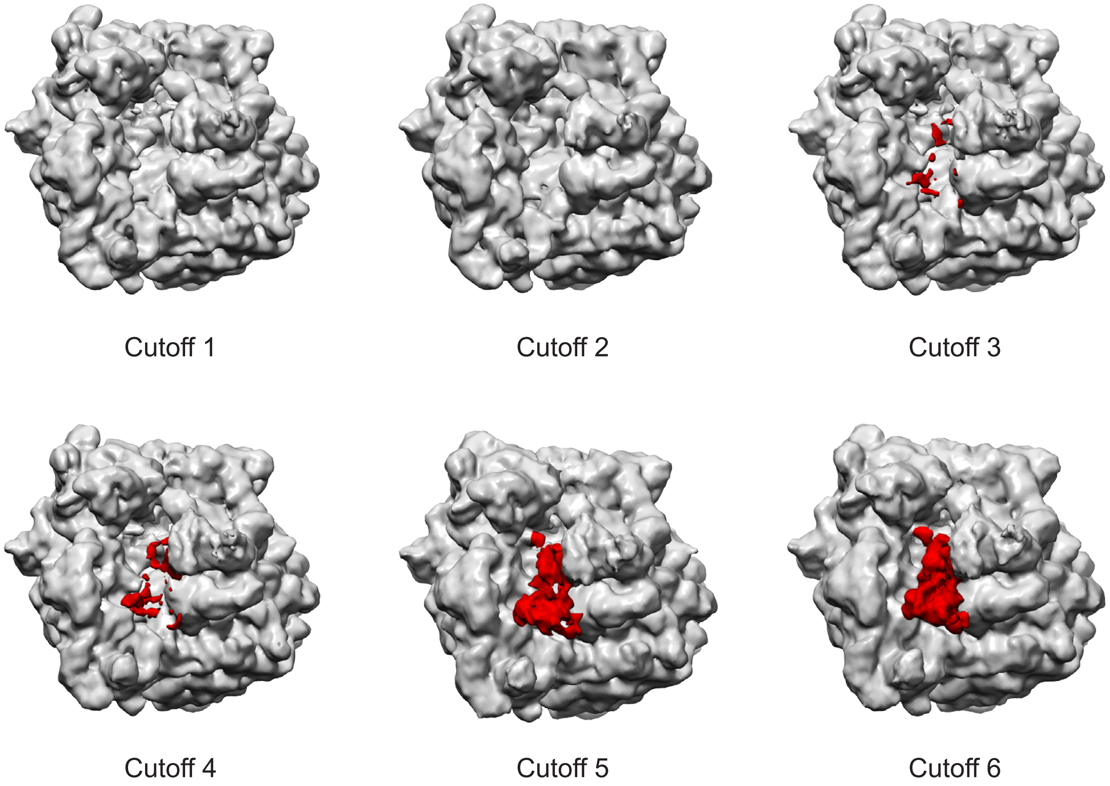


Figure 3.5. Reconstructions from supervised classification. A total of six density maps were reconstructed using particle subsets from various cutoff thresholds, shown in Figure 3.4. The cutoff #1 reconstruction, using all 133,782 particles, displays no density for the BipA protein, although its resolved resolution (~9.3 Å) is the best of any class. The cutoff #2 reconstruction (using only particles with negative ΔCC values) also display no density for the BipA protein, although there is a stronger density for a P-site tRNA. Starting with the cutoff #3 reconstruction, a density for BipA can be seen near the GTPase-associated center (GAC) of the 70S ribosome. The cutoff #6 reconstruction shows the strongest density for BipA, although its resolved resolution (~13 Å) is the worst of any of cutoff reconstruction. All prospective densities for BipA are colored in red.

RELION Round	Class	# Particles	Description	Rejected
Round 1	1	51185	70S–E-site tRNA	
Total # Particles:	2	62219	70S–All	
293036	3	45399	70S–All	
	4	42445	Non-particles	X
	5	44112	70S–All	
	6	47676	70S–All	
Round 2	1	15458	70S–E-site tRNA	
Total # Particles:	2	21896	Non-particles	X
250591	3	21553	70S–P-site tRNA–E-site tRNA	
	4	25145	70S–BipA	
	5	30975	70S–P-site tRNA–E-site tRNA	
	6	24050	70S–BipA	
	7	22497	70S	
	8	31241	70S–P-site tRNA–E-site tRNA	
	9	24684	Non-particles	X
	10	33092	Non-particles	X
Round 3	1	8602	Low-Quality 70S	X
Total # Particles:	2	3379	Non-particles	X
170919	3	22424	70S–BipA	
	4	28992	70S–P-site tRNA–E-site tRNA	X
	5	14065	70S–E-site tRNA	X
	6	22710	70S–E-site tRNA	X
	7	18888	70S–BipA	
	8	8800	70S	X
	9	20050	70S–BipA	
	10	23009	70S–P-site tRNA	X
Round 4	1	10386	70S–P-site tRNA	X
Total # Particles:	2	11920	70S–BipA	
61362	3	2175	Low-quality 70S	X
	4	12134	70S–BipA	
	5	14259	70S–BipA	
	6	10488	70S–BipA	
Round 5	1	8785	70S–BipA	
Total # Particles:	2	17099	70S–BipA	
48801	3	1963	Low-quality 70S	X
	4	4255	70S–P-site tRNA	X
	5	11101	70S–BipA	
	6	5598	70S–BipA	
Round 6	1	11840	70S–BipA	
Total # Particles:	2	9640	70S–P-site tRNA	
42583	3	11098	70S–BipA	
	4	10005	70S–P-site tRNA	

Table 3.1. The RELION classification scheme for the BipA AutoPicker dataset. The BipA AutoPicker dataset was classified in six rounds of RELION classification. In the initial three rounds, only non-particle classes were discarded. In later rounds of RELION, particles assigned to classes visually inconsistent with a 70S–complex were discarded. In Round 6, particles from class 2 and 4 were pooled together for the final reconstruction of the 70S–BipA complex. Abbreviations are as follows: ‘70S–All’ - 70S ribosome with scattered densities for BipA and three tRNAs; ‘70S–BipA’ - 70S ribosome complexed with BipA, a P-site tRNA, and an A-site tRNA.

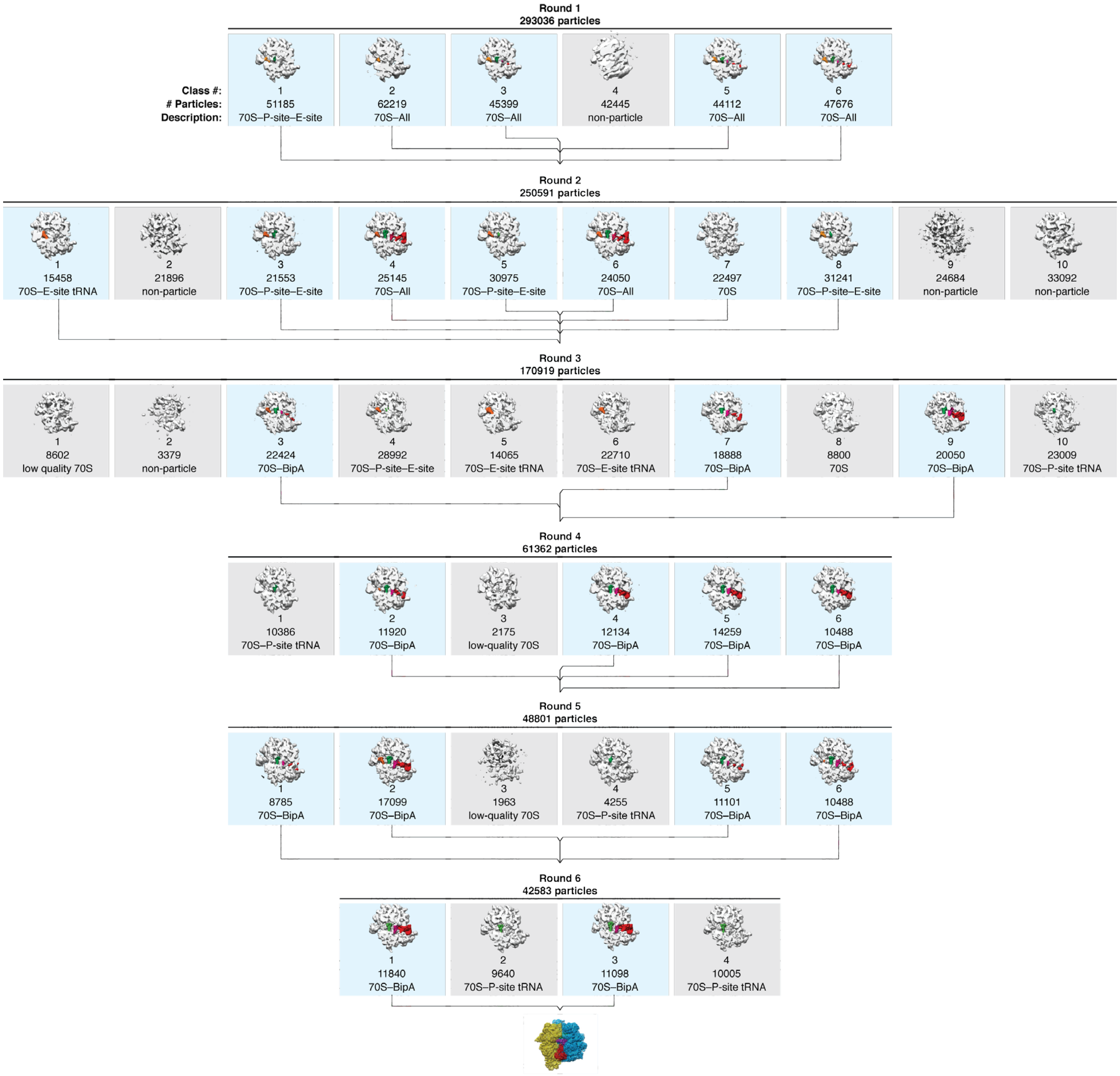


Figure 3.6. Density maps of the classes observed in the RELION classification scheme. Density maps of the ribosomal classes corresponding to those tabulated in Table 3.1 are presented here. Ribosomal components are highlighted in different colors: A-site tRNA - magenta; P-site tRNA - green; E-site tRNA - orange; BipA - red. Abbreviations are as follows: ‘70S–All’ - 70S ribosome with scattered densities for BipA and three tRNAs; ‘70S–P-site–E-site’ - 70S ribosome complexed with a P-site tRNA and E-site tRNA;‘70S–BipA’ - 70S ribosome complexed with BipA, a P-site tRNA, and an A-site tRNA.

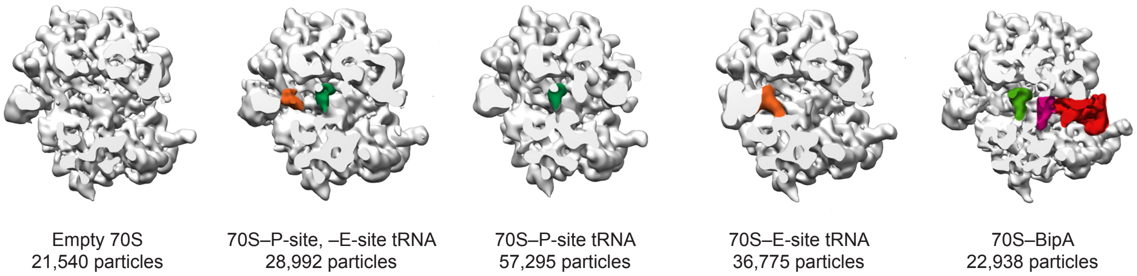


Figure 3.7. Observed ribosomal classes in the BipA AutoPicker dataset. RELION (Scheres, 2012a) was employed for 3D classification of the BipA AutoPicker dataset. Overall, five distinct ribosomal classes were observed: an empty 70S (21,540 particles); 70S with E-site tRNA occupancy (36,775 particles); 70S with P-site tRNA occupancy (57,295 particles); 70S with E-site and P-site tRNA occupancy (28,992 particles); and 70S with BipA as well as A- and P-site tRNAs (22,938 particles), designated as the 70S–BipA complex of interest. The 70S–BipA class was further refined to give the final reconstruction, presented in Chapter 4, resolved to 8.5 Å.

REFERENCES

Frank, J., Shimkin, B., Dowse, H., 1981. SPIDER—A modular software system for electron image processing. Ultramicroscopy 6, 343-357.

Frank, J., Radermacher, M., Penczek, P.A., Zhu, J., Li, Y., Ladjadj, M., Leith, A., 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. 116, 190-199.

Gao, H., Valle, M., Ehrenberg, M., Frank, J., 2004. Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. J. Struct. Biol. 147, 283-290.

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U.T., Pestova, T.V., Frank, J., 2013. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153, 1108-1119.

Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., Egelman, E.H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S.J., Medalia, O., Penczek, P.A., Rosenthal, P.B., Rossmann, M.G., Schmid, M.F., Schröder, G.F., Steven, A.C., Stokes, D.L., Westbrook, J.D., Wriggers, W., Yang, H., Young, J., Berman, H.M., Chiu, W., Kleywegt, G.J., Lawson, C.L., 2012. Outcome of the first electron microscopy validation task force meeting. Structure 20, 205-214.

Langlois, R., Ash, J.T., Pallesen, J., Frank, J., 2014a. Fully automated particle selection and verification in single-particle cryo-EM, p. 43-66-66, Applied and Numerical Harmonic Analysis, Springer New York.

Langlois, R., Pallesen, J., Ash, J.T., Nam Ho, D., Rubinstein, J.L., Frank, J., 2014b. Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. J. Struct. Biol. 186, 1-7.

Liao, H.Y., Frank, J., 2010. Definition and estimation of resolution in single-particle reconstructions. Structure 18, 768-775.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst., Man, Cybern. 9, 62-66.

Penczek, P.A., Grassucci, R.A., Frank, J., 1994. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53, 251-270.

Rath, B.K., Frank, J., 2004. Fast automatic particle picking from cryo-electron micrographs using a locally normalized cross-correlation function: a case study. J. Struct. Biol. 145, 84-90.

Roseman, A.M., 2003. Particle finding in electron micrographs using a fast local correlation algorithm. Ultramicroscopy 94, 225-236.

Scheres, S.H.W., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P.B., Frank, J., Carazo, J.-M., 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat Meth 4, 27-29.

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

Scheres, S.H.W., 2012b. A Bayesian view on cryo-EM structure determination. J Mol Biol 415, 406-418.

Sigworth, F.J., Doerschuk, P.C., Carazo, J.-M., Scheres, S.H.W., 2010. An introduction to maximum-likelihood methods in cryo-EM, p. 263-294, Meth. Enzymol.

Valle, M., Sengupta, J., Swami, N.K., Grassucci, R.A., Burkhardt, N., Nierhaus, K.H., Agrawal, R.K., Frank, J., 2002. Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. EMBO J. 21, 3557-3567.

CHAPTER 4:

The Cryo-EM Structure of BipA Bound to the Ribosome

CHAPTER 4 ABBREVIATIONS

Abbreviation	Full Title
(p)ppGpp	guanosine (penta)tetraphosphate
aa-tRNA	aminoacyl-tRNA
BipA	BPI-Inducible Protein A
CCA	cytosine-cytosine-adenosine
CTD	c-terminal domain
DTT	dithiothreitol
EF-G	elongation factor G
EF4	elongation factor 4
GAC	GTPase-associated center
GMPPNP	5'-Guanylyl imidodiphosphate
GTP	guanosine triphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
k_{cat}	turnover rate
k_{cat}/K_M	catalytic efficiency
M	molar
MDFF	molecular dynamics flexible fitting
mM	millimolar
mRNA	messenger RNA
NTD	n-terminal domain
OB-fold	oligonucleotide/oligosaccaride-binding fold
OD	optical density
pmol	picomoles
PTC	peptidyl transfer center
RMSD	root-mean-square deviation
rRNA	ribosomal RNA
SAD	single-wavelength anomalous diffraction
SDS	sodium dodecal sulfate
SeMet	selenomethionine
SRL	sarcin-ricin loop
trGTPase	translational GTPase
tRNA	transfer RNA
α	alpha
β	beta
βME	β-mercaptoethanol
μl	microliter
μM	micromolar

4.1 INTRODUCTION

The reconstruction of the *S. enterica* 70S–BipA complex, resolved to 8.5 Å, presents a previously uncharacterized structure. In order to elucidate the specific interactions between the BipA protein and the 70S ribosome, as well as the protein's interactions with the A-site tRNA, one can leverage the precision of known X-ray structures and employ fitting techniques to generate a quasi-atomic model of the entire complex. Ideally, we would fit an atomic structure of the 70S ribosome occupied with A-site and P-site tRNAs into the density map. The final necessary component for a complete model of the 70S–BipA complex would be a full X-ray structure of BipA, to be fitted into the portion of the density map corresponding to the protein.

To date, only a portion of the C-terminal domain of *Vibrio parahaemolyticus* BipA has been solved (PDB: 3E3X, unpublished structure by Nocek, B. et al). Sequence comparison between the *V. parahaemolyticus* BipA and the *S. enterica* BipA revealed 65% sequence homology. However, as only a portion of the protein was crystallized, the structure of *V. parahaemolyticus* BipA was unsuitable for use in further molecular modeling. Fortunately, our collaborators in the Robinson lab were able to solve the full X-ray structure of *S. enterica* BipA to 2.7 Å, allowing us to complete the molecular modeling of the entire 70S–BipA complex. Our model, along with additional biochemical experiments, provides exciting insights to the way BipA binds to the ribosome.

Much of the work presented in this chapter was performed in close collaboration with the lab of Dr. Victoria Robinson. For this reason, Ala Shaqra in the Robinson Lab is a co-first author on the work presented here. Demarcation of contributions by members of the Robinson lab is given where appropriate. The chapter is structured in the following way: Section 4.2 provides details on the elucidation of the BipA X-ray structure, solved by Dr. Victoria Robinson. The use of MDFF to fit the X-ray structures of the 70S ribosome, the two tRNAs, and the BipA X-ray structures into the reconstruction is provided in Section 4.3. The quasi-atomic model of the 70S–BipA complex is discussed in Section 4.4. Our model prompted the design of a new series of biochemical experiments, performed by Ala Shaqra, aimed at characterizing the specific ribosomal species needed for optimal BipA binding, to be discussed in Section 4.5. Section 4.6 concludes the chapter with an interpretation of the results and discusses the possible implications of the newfound insights.

4.2 THE BIPA X-RAY STRUCTURE

4.2.1 Expression and Purification of BipA

S. enterica BipA was purified as described previously (deLivron and Robinson, 2008). In brief, BipA was overproduced in *E. coli* BL21(DE3) (Novagen, Billerica, MA) and purified using HisTrap FF crude column (GE Biosciences, Piscataway, NJ) and gel filtration chromatography as previously described (deLivron and Robinson, 2008). Selenomethionine derivatized protein was obtained by transforming B834(DE3)pLysS (Novagen, Billerica, MA) cells with pWW3. Cells were grown according to a procedure adapted from a protocol supplied by J. Brannigan, R. Lewis and A. Wilkinson. In brief, a 1 ml overnight culture was used to inoculate 50 ml of LB. The cells were grown at 37 °C until reaching an OD_{600} ~ 1.0, harvested by centrifugation and washed twice with pre-warmed media containing 2X-M9 salts, 4% (w/v) glucose, 2 mM $MgSO_4$, 25 mg/ml $FeSO_4$–$7H_2O$, thiamine, riboflavin, pyridoxine monohydrate, niacinamide, 30 mg/ml kanamycin and 40 mg/ml of all the amino acids excluding Met which was substituted with SeMet. The cells were then resuspended in 1 L of the same media and grown at 37 °C to an OD_{600} ~ 0.6 before induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 18 °C. The protein was purified using the same procedure as described above for the native protein except that all buffers were supplemented with 1 mM DTT to prevent oxidation of the selenomethionine residues.

4.2.2 BipA Crystallization, X-ray Data Collection and Refinement

Crystals of *S. enterica* BipA, absent of any energy nucleotide, were grown by the hanging drop method of vapor diffusion from a PEG4000 solution. Selenomethionine-derivatized crystals were obtained under similar conditions. Crystals were flash-frozen after brief soaking in a reservoir solution containing 15% (v/v) glycerol. The crystals belong to space group $P2_1$ with unit cell parameters *a* = 89.5 Å, *b* = 84.1 Å, *c* = 95.61 Å, a = g = 90° and b = 106.2° corresponding to 2 molecules in the asymmetric unit. Diffraction data were collected at 100 °K at the National Synchrotron Light Source (Brookhaven, NY) on beamlines X6, X25 and X29. Processing, integration and scaling of data was done with DENZO and SCALEPACK (Otwinowski and Minor, 1997). Structure factors were rescaled for anisotropy and ellipsoidal truncation with the Diffraction Anisotropy web server (Strong et al., 2006).

The structure was solved using SAD phasing with data collected from a single selenomethionine-derivatized protein crystal. Positions of six pairs of Se sites related by a 2-fold NCS were determined from the peak wavelength utilizing the SAD phasing protocol implemented in SOLVE (Terwilliger and Berendzen, 1999). From these initial phases, an automated SOLVE search of the 15-2.7 Å SAD data located 21 additional Se-met residues present in the asymmetric unit which contains two BipA molecules. RESOLVE was used to generate a preliminary model of BipA consisting of approximately 2/3 of the polypeptide chain, as poly(alanine and glycine), in the asymmetric unit and assigned side chains to 156 amino acid residues (Terwilliger, 2003). The partial structure was rigid-body fitted into the asymmetric unit and refined with REFMAC5 (Murshudov et al., 1997).

A minimal starting model of BipA was made by removing all loops and extended polypeptide chain regions that did not correspond exactly to the SAD electron density. This model was then refined against the native data set at 2.7 Å and subjected to iterative cycles of model building using COOT (Emsley and Cowtan, 2004) and refinement with REFMAC5 (Murshudov et al., 1997) until convergence. Stereochemistry of the model was inspected with PROCHECK (Laskowski et al., 1993).

The final model of BipA contains 568 of the 607 residues in the protein and has a final R_{work} of 25.4 and an R_{free} of 29.6%. Undefined regions of the structure include the switch I (residues 34-55), a large flexible loop in the CTD (residues 541-552) and seven residues at the C-terminus of the protein (residues 600-607). Statistics for the refinement are given in Table 4.1. Figures were generated using PyMol (DeLano, 2002).

4.2.3 Overall Structure of S. enterica BipA

The BipA X-ray structure has five domains, as shown in Figure 4.1, which have been named according to the domain definitions of EF-G, as discussed in Chapter 2. Domain I (residues 1-197), a GTPase fold, is a six-stranded β-sheet surrounded by five α-helices. Domain II (residues 198-290) is a six-stranded Oligonucleotide/ Oligosaccaride-binding fold (OB-fold). All translational GTPases contain these two domains at their N-terminus (Margus et al., 2007). Domains III (residues 291-397) and V (residues 398-476) have similar split β-α-β folds. A ten-residue linker connects the fifth C-terminal domain (CTD) (residues 477-607) to the previous four domains. This domain has two centrally located β-sheets: one is a twisted extension to an adjoining β-sheet in domain III and the other surrounds a short helix, which in turn is bordered on the opposite side by a flexible loop. A small, basic helix (residues 593-601) is present at the C-terminus of the protein. This helix is crucial for both 70S and 30S ribosome binding (deLivron et al., 2009). The unique features of this domain prompted us to submit it to the DALI server in order to more closely examine its similarities and differences to other protein structures (Holm and Rosenstrom, 2010). DALI gave no significant hits, indicating that the CTD of BipA has a novel fold never observed before in any protein deposited in the PDB.

4.2.4 Comparison of the Unbound BipA, EF-G, and EF4 X-ray Structures

The overall structural features of isolated EF-G, EF4 and BipA resemble one another as would be expected from their high degree of sequence similarity, as discussed in Chapter 2 (Section 2.4.3) and in deLivron (2009) (deLivron et al., 2009). All three proteins have five domains, four of which are topologically equivalent (Aevarsson et al., 1994; al-Karadaghi et al., 1996; Evans et al., 2008). However, there are quite a number of distinguishing structural features in all three proteins (Figure 4.2). For example, both BipA and EF4 are missing the G' domain present in EF-G, and the switch regions in all three proteins are of different length. Domain II in BipA more closely resembles the corresponding domain in EF4 in that it is missing the first two strands of the β-barrel present in EF-G. Domains III and V in all three proteins superimpose well onto one another with RMSD of less than 3.0 Å for ~140 residues. As shown in Figure 4.2, the orientation of these domains relative to the GTPase and β-barrel domains are different in each protein.

Undoubtedly, the distinguishing structural component of each protein family is the domain that is positioned distal to the GTPase domain. For EF-G, this is domain IV whereas in EF4 and BipA it is the CTD. Interestingly, the CTDs of BipA and EF4 have novel folds but the argument can be made that the CTD of BipA is quite distinct, as shown in Figure 4.3. This is because domain IV in EF-G and the CTD of EF4 both have an antiparallel β-sheet with either one or two helices on one side. The CTD of BipA looks nothing like these domains, eluding to the idea that BipA may make very different contacts with the ribosome than either EF-G or EF4. In summary, even though the gross morphological features of the EF-G, BipA and EF4 families of proteins are similar, they each have distinctive structural attributes enabling them to interact selectively with a given biological state of the ribosome.

4.3 CHARACTERIZATION OF THE 70S–BIPA CRYO-EM RECONSTRUCTION

4.3.1 Sample Preparation and Electron Microscopy

Purified BipA and 70S ribosome samples were prepared as described in Sections 4.2.1 and 4.5.1, respectively. *In vitro* complexes were assembled by incubating 50 nM ribosomes, 500 nM BipA, and 1.25 µM GMPPNP together in a buffer of 10 mM Tris (pH 7.5), 5 mM $MgCl_2$, 30 mM NH_4Cl, 1 mM DTT for 20 minutes at 37 °C. A total of 4 µl of the *in vitro* reaction were applied to 200-mesh holey carbon grids (Quantifoil 2/4 grid, Quantifoil Micro Tools GmbH, Jena, Germany) as described previously (Grassucci et al., 2007). Grids were blotted and plunge-frozen in liquid ethan at with a Vitrobot (FEI, Portland, Oregon).

Data collection of the BipA dataset was described at length in Chapter 3 (Section 3.2). In brief, film micrographs were recorded on the FEI Tecnai F30 Polara electron microscope (FEI, Eindhoven) with Kodak Electron SO-163 Image Film at a low dose (~18e^-/A^2) and a calibrated magnification of 58,269x. The pixel size is 1.2 Å at the specimen level when digitized using the 16-bit ZI Imagine Photoscan 2000 densitometer (Zeiss, Aalen, Germany) at a sampling rate of 7 µm.

4.3.2 Reconstruction Procedures

Image processing and AutoPicker particle selection of the BipA film dataset were described in Chapter 3, Sections 3.2 and 3.3.2, respectively. RELION (version 1.2b7) (Scheres, 2012) was used for 3D classification and refinement of the BipA AutoPicker particle dataset. The classification scheme was discussed extensively in Chapter 3. In the final round (Round #6) of RELION classification, a total of 42,583 particles where classified into four classes, as described in Table 3.1 and shown in Figure 3.6. Classes 1 (11,840 particles) and 3 (11,098 particles) were visually consistent with a 70S ribosome bound with BipA as well as A- and P-site tRNAs, as shown in Figure 3.7. Superimposition of Class 1 and Class 3 reconstructions showed no differences in the 70S, A-site tRNA, P-site tRNA, or BipA conformations. Class 2 and Class 4 were reconstructions of a 70S bound with a P-site tRNA. Particles assigned to Class 1 and 3 were pooled together to give a final subset of 22,938 particles for 3D refinement 70S–BipA reconstruction.

RELION was employed for automated refinement of the reconstruction using a low-resolution empty 70S ribosome (EMD ID: 2277) filtered to 60 Å as an initial model. The final reconstruction of the 70S–BipA complex is resolved to a resolution of 8.53 Å as assessed by the gold standard protocol (Henderson et al., 2012). The FSC curve is shown in Figure 4.5. The estimated accuracy of assigned angles as reported by RELION was 1.826°. RELION post-processing with auto-mask and auto-bfactor determined the resolution of the map as 8.35 Å and the b-factor of the map as -718.818 Å^2. However, in each case, the reporting of the second and third decimal is questionable. The final cryo-EM reconstruction of the 70S–BipA complex, depicted in Figure 4.4, shows a 70S ribosome complexed with BipA as well as A- and P-site tRNAs. All figures of the reconstruction were generated using UCSF Chimera (Pettersen et al., 2004).

4.3.3 Segmentation and Display of Density Maps

The 70S–BipA cryo-EM reconstruction was segmented using several modules in UCSF Chimera (Pettersen et al., 2004). The SEGGER module was used for initial segmentation of the volume isolated (Pintilie et al., 2010; Baker and Rubinstein, 2011). Segments containing less than 10,000 voxels were discarded. Segments were refined manually using the Volume Eraser module implemented in UCSF Chimera. Overall, five segments were obtained, corresponding to the 50S large ribosomal subunit, the 30S ribosomal subunit, a P-site tRNA, an A-site tRNA, and BipA, as shown in Figure 4.4. The segments obtained were smoothed using a Gaussian filter in the volume filter module of Chimera.

4.3.4 The 70S–BipA Reconstruction

Overall, the ribosome is found in the classical, non-rotated conformation. As compared with the *E. coli* 70S ribosome, the *S. enterica* 70S ribosome here is equivalent in protein and rRNA composition. Thus, as discussed below in Section 4.4.1, the atomic structure of the *E. coli* ribosome will be used in the molecular modeling of the complex. There is only scattered density present for the highly flexible L9 and L7/L12 proteins, which have been difficult to visualize both in cryo-EM and X-ray structures. The P-site tRNA is in the canonical P/P conformation as seen by comparison with previous structures (Agirrezabala et al., 2012), while the A-site tRNA is in a deformed, previous uncharacterized conformation (Figure 4.7). While the density for the P-site tRNA is complete, the A-site tRNA evidently lacks density for its 3’ CCA end, a point we will elaborate on in a later section.

We see masses of density for the first four domains (I, II, III, and V) of BipA, and partial density for its CTD, as shown in Figures 4.6 and 4.9. When overlaid with the bound structures of the elongation factors (not shown), it is apparent that BipA's domains I and II bind with the 70S ribosome in a way that is similar to domains I and II of EF-G, EF-Tu, and EF4. These two domains contact the GTPase-associated center of the ribosome, at the Sarcin-Ricin Loop of the 23S rRNA. This result is congruent with previous biochemical studies that have pinpointed BipA binding to the same general -- though not equivalent – ribosomal binding site as the canonical elongation factors. Akin to these factors, BipA experiences an increase in ribosome-induced GTPase activity, suggesting a shared mechanism for GTPase activation. Density for contact between BipA's Domain III and the 30S small subunit is apparent only at lower thresholds, suggesting that the interaction between the protein and the subunit is either transient or weak. This finding agrees with previous sequence homology studies which predicted that few amino acids in Domain III are required for the binding of BipA (deLivron et al., 2009). In contrast, Domain V makes strong contacts with the L11 NTD lobe, similar to the binding between EF-G and this lobe. Although density for the CTD of BipA is only partially observed, the CTD is seen to make strong contacts with the A-site tRNA through multiple points. Moreover, the A-site tRNA is apparently deformed as a consequence of these interactions. This deformation will be further characterized below.

4.4 A QUASI-ATOMIC MODEL OF THE 70S–BIPA COMPLEX

4.4.1 Atomic structure modeling

To elucidate the interactions between the protein and the ribosome, the Molecular Dynamics

Flexible Fitting (MDFF) (Trabuco et al., 2008) technique was employed. Fitting of the atomic structures into the cryo-EM reconstruction was performed using MDFF (Trabuco et al., 2009) (assuming a generalized-Born implicit solvent) as implemented in program NAMD (Tanner et al., 2011). The cryo-EM map of the 70S–BipA complex was fitted with an atomic model of the 70S–P-tRNA–A-tRNA complex (PDBs: 3JOU and 3JOT) (Agirrezabala et al., 2012) and the X-ray structure of BipA. The X-ray structure of BipA lacks two fragments: One is the distal loop region (residues 542-552) in the CTD, where the structure is expected to be highly unstable in the ribosome-unbound form. The other is for the Switch I region (residues 35-52) immediately neighboring the GTP. The two missing regions were modeled as two loops to create a stereochemically complete structure for further fitting. This resultant model of BipA was first flexibly fitted into the segmented map for BipA, followed by a fitting of the entire model of the 70S–P-tRNA–A-tRNA–BipA into the complete unsegmented map.

4.4.2 A Model of the 70S–BipA Complex

The atomic model of BipA fitted into our density map is shown in Figure 4.6. The binding sites of Domains I and II proved to be quite similar as those of the homologous domains in EF-G, EF-Tu, and EF4, suggesting a similar mechanism of ribosome-induced GTPase activation. The map shows only partial density for the switch I (residues 34-55) region in Domain I, which indeed has been shown to be flexible in the crystal structure of BipA. For MDFF, as discussed above, this region was filled in as a flexible loop. Examination of this region in the 70S–BipA reconstruction, shown in Figure 4.8A, reveals that MDFF was unable to rebuild a model of the BipA switch I loop, likely due to poor starting model. To elucidate the amino acids of BipA that likely fits into this density, we superimposed the X-ray structures of *T. thermophilus* EF-G (PDD: 4JUW) and EF-Tu (PDB: 2XQD), both with structured switch I regions and in their GTPase-activated states, onto BipA's

Domain I. As shown in Figure 4.8A, residues 54-63 of EF-G and residues 52-61 of EF-Tu fits very well into this partial density in the 70S–BipA reconstruction. As discussed in Chapter 2, section 2.3.2, GTPase activation requires the correct coordination of three unverisally conserved amino acids in Domain I of translational GTPases. In *T. thermophilus* EF-Tu, these residues are: Val20 in the P-loop, Ile61 in the switch I loop, His85 in the switch II loop. In *T. thermophilus* EF-G, the homologous residues are Ile20 , Ile63, and His87, respectively. In *S. enterica* BipA, the homologous residues are Val14, Ile54, and His78, respectively. As shown in Figure 4.8B, there is high agreement in the positions of Val14 and His78 in the BipA model with the corresponding homologous residues of both EF-G and EF-Tu. Disagreement in the position of BipA Ile54 with EF-Tu Ile61 and EF-G Ile63 can be attributed to the ill-fitted BipA switch I loop, as discussed above. Thus, our model suggests that we have captured a GTPase-activated state of BipA bound to the 70S ribosome.

Within Domain III, amino acids 313-327 form a flexible loop, which is within the vicinity of the ribosomal protein S12 on the 30S small subunit. As previously noted, any interaction between Domain III and the small subunit may be transient and not sufficient to stabilize the binding of the protein. Additionally, we revisited a previous mutational study (deLivron et al., 2009) aimed at identifying amino acids necessary for binding. In that study, point mutations were performed on amino acids shared among elongation factors as well as uncharacterized peptides in the novel CTD. This study found that R375 in Domain III is not required for BipA binding. In our model, R375 is situated in the vicinity of a very large flexible region spanning residues 29-60 of Domain II, and thus, this residue may not be necessary for any kind of ribosomal binding or structure stabilization. Finally, the flexible residues in this section display a strong density in the reconstruction, suggesting that the bound form of BipA may be more rigid in this domain.

As compared with the X-ray structure of BipA, Domain V is shifted toward the L11 protein on the base of the L7/L12 stalk and makes strong contacts, especially along the α-helix and β-sheet formed by amino acids E409-G435. These interactions mirror the binding interactions of other canonical elongation factors to the ribosome. In a point mutation study (deLivron et al., 2009), R422 and K423 were found to be unnecessary for binding while K427, K434, and R436 were required. In our model, both R422 and K423 lie far from any 50S subunit proteins while K427, L434, and R436 are near the interface between Domain V and the L11 protein of the L7/L12 stalk base, and thus, may be required for stable binding of BipA to the ribosome (Figure 4.9).

The CTD is a novel non-homologous domain necessary for BipA binding and activity (deLivron et al., 2009). By means of MDFF fitting, we observe that various flexible regions of the CTD that lack density, such as the large distal loop, are in proximity of four 23S rRNA helices on the large 50S subunit, namely H89-H92 (Figure 4.9). These 23S helices may help coordinate the CTD into position for proper and efficient binding, as well as stabilize BipA's interaction with the A-site tRNA. Additionally, according to the aforementioned study by deLivron and coworkers (deLivron et al., 2009), H527 and R529 are both important for BipA binding. As shown by our model, these amino acids are in proximity to H89, perhaps aiding in the coordination of the CTD. Residue K562, also instrumental for the activity of BipA, is in proximity to the Sarcin-Ricin Loop (SRL) at H95, an important center for ribosome-induced GTPase activity. BipA's CTD extends toward the A-site tRNA and thereby appears to induce a new conformation of the tRNA previously unobserved in the literature.

While the position of the anticodon remains unchanged from the A/A state, the rest of the tRNA exhibits a distortion due to various interactions with regions of the BipA's CTD. The importance of BipA's distal loop (residues 535-556) was unexplained in the point mutation study by deLivron and coworkers (deLivron et al., 2009), which found that N536, K541, K542, and R547 are all necessary for effective BipA binding to the ribosome. This region was not ordered in the X-ray structure. Our model shows that all four of these amino acids lie in close proximity to the entire breadth of the A-site tRNA acceptor stem, and thus may be the reason why the acceptor stem itself is displaced by 6-8 Å, away from its position in the canonical A/A tRNA state, and toward BipA. The D-loop of the tRNA has strong interactions with the final basic α-helix of the BipA CTD (residues 593-601). In our model, G46 in the D-loop flips out of the tRNA ladder structure within the vicinity of R598 in the final α-helix (Figure 4.10). We suggest that the final α-helix is responsible for strong interactions with the tRNA, possibly by coordinating the nucleotides within the D-loop. These interactions apparently distort the D-loop of the tRNA, causing it to shift by 8 Å away from its position in the A/A state, toward BipA. Previous studies found that deletion or mutation of the final α-helix abrogates BipA's binding (deLivron and Robinson, 2008). Additionally, this helix is universally conserved in all BipA sequences across prokaryotes. Our model thus provides an explanation for the importance of the final α-helix in the binding of BipA.

The CCA end of the A-site tRNA could not be fitted using the atomic coordinates, as we lack sufficient density in the reconstruction. This observation led to questions about the tRNA's identity and acylation state, which are addressed below with biochemical assays. Additionally, the presence of the A-site tRNA in a BipA-bound complex was unexpected. Previous biochemical studies (Farris et al., 1998; deLivron and Robinson, 2008), aimed at characterizing the BipA-bound complex, did not suggest the presence or necessity of an A-site tRNA. Thus, the observation of an A-site tRNA in the reconstruction prompted a reexamination of previous biochemical assays (deLivron and Robinson, 2008; deLivron et al., 2009). Several studies have noted regular tRNA contamination in 70S samples and have stressed that extra precautions, such additional washing cycles or the use of puromycin must be taken to ensure a purified sample (Leshin et al., 2010; Meskauskas et al., 2011).

Additional complexes (not shown) were reconstructed using isolated BipA and purified 70S samples, with careful precautions taken to wash out tRNA contamination. In this case the reconstruction reflects a homogenous population of empty ribosomes, with no trace of a mass for BipA. These experiments suggest that BipA's interaction with the A-site tRNA is an integral requirement for its binding and activity.

SECTION 4.5 BIOCHEMICAL CHARACTERIZATION OF BIPA GTPASE ACTIVITY AND BINDING IN THE PRESENCE OF VARIOUS RIBOSOMAL COMPLEXES

4.5.1 Ribosome Isolation

S. enterica 70S ribosomes were obtained as described, with the following modifications (deLivron et al., 2009): Crude ribosome pellets were resuspended in 10 mM Tris (pH 7.5), 10 mM $MgCl_2$, 30 mM NH_4Cl, 1 mM DTT. 100 OD_{254} units were applied to a 7-47 % gradient and centrifuged at 96,000 *x g* for 7.2 hr at 4 °C in a Surespin 630 rotor (Thermo Scientific, Waltham, MA). Fractions containing 70S ribosomes were pooled, pelleted, resuspended in the same buffer, flash-frozen in liquid nitrogen and stored at -80 °C until use.

4.5.2 Steady State GTP Hydrolysis Assays

Guanine nucleotide hydrolysis activities of BipA were determined by measuring the release of free phosphate with the malachite green-ammonium molybdate assay (Lanzetta et al., 1979; deLivron et al., 2009). Assays were done in 96-well plate format. His-tagged BipA protein (1 μM) was incubated for 5 to 90 min at 37 °C in a 200 µl reaction mixture containing 20 mM Tris (pH 7.5), 200 mM NaCl, 10 mM $MgCl_2$, 2 mM β-mercaptoethanol (βME) and 50 – 2000 µM GTP. Where indicated, BipA was incubated with 25 nM of a given ribosomal species at 37 °C for 30 min. At the specified time points, 30 µl is removed and added to 200 µl of malachite green solution to quench the reaction. After 30 min, color formation was measured at 660 nm using a Synergy HT 96-well plate reader (Bio-Tek, Winooski, VT). Kinetic parameters were determined by a non-linear regression fit of the data to the Michaelis-Menten equation using GraphPad Prism (Version 5.0d). Kinetic values are reported as average values with standard deviations and correspond to a minimum of three independent experiments.

4.5.3 Ribosome Association Experiments

Examination of complex formation between BipA and the ribosome was done using co-sedimentation through a sucrose gradient using the same technique as in our previous studies (deLivron and Robinson, 2008). Specific ribosome complexes were prepared as described by Blaha and coworkers (Blaha et al., 2000). In brief, 70S ribosomes (100 pmol) are heat-activated in polymix buffer containing 20 mM $MgCl_2$ for 30 min at 37 °C. A custom-synthesized mRNA [5'-GGCAAGGAG-GUAAAAAUG-3'] was designed to accommodate an initiator tRNA at the P-site (FM-03, tRNA Probes, Inc) and either an aminoacylated (L-50, tRNA Probes, Inc, College Station, TX) or uncharged Lys–tRNA (L-01, tRNA Probes, Inc, College Station, TX) at the A-site. This mRNA was added to activated ribosomes at 3-fold molar excess and allowed to incubate for another 30 minutes at 37 °C. To assemble an initiation complex (70S IC), a P-site tRNA, at a 3-fold excess, was added to the ribosome and incubated for 3 min at 37 °C. Here, we use the notation of X-$tRNA^X$, where X designates the amino acid charged to the tRNA and the superscripted X denotes the anticodon.

To produce a 70S IC with an A-site tRNA, either Lys-tRNA^{Lys} or tRNA^{Lys} (designating charged or uncharged) was introduced to the 70S IC complex in 3-fold molar excess and incubated again for 3 min at 37 °C. The resulting complexes, presumably with either tRNA^{Lys} or Lys-tRNA^{Lys} in the A site, are designated 70S IC(A) and 70S IC(A)*, respectively. The complexes were then cooled to 30 °C and BipA:GMPPNP (pre-incubated on ice for 30 min) added to the mixture, followed by a final incubation for 30 minutes at 30 °C.

4.5.4 Steady-State Kinetics

Previous studies from the Robinson Lab demonstrated that BipA binds to the 70S ribosome with a 1:1 stoichiometry and that the GTPase activity of the protein is stimulated by this association (deLivron and Robinson, 2008). For the current studies, the malachite green-ammonium molybdate assay was adapted to a 96-well plate format by Ala Shaqra. In agreement with previous studies, we observed the same increase in the BipA turnover rate (k_{cat}) from 18.0 ± 0.6 hr^{-1} in its unbound, isolated form to 57.6 ± 5.2 hr^{-1} in the presence of purified 70S ribosome (Figure 4.11A). To determine if the presence of an A-site tRNA has any effect on the GTPase activity of BipA, three complexes were assembled as described above: a 70S IC, 70S IC(A) and 70S IC(A)*. GTP hydrolysis activity of BipA was measured in the presence of each of these complexes (Figure 4.11A). Interestingly, there was only a modest change in the ribosome-stimulated GTPase activity of BipA when an A-site tRNA was present on the 70S ribosome. Similar three- to four-fold increases in k_{cat} values were obtained for BipA–70S, BipA–70S IC and BipA 70S IC(A)* and a two-fold increase was measured for BipA–70S IC(A). There is, however, a difference in the catalytic efficiency of BipA in the presence of these various ribosomal species with the largest change observed between BipA in isolation, k_{cat}/K_M of 4.5 M^{-1}s^{-1}, and BipA–70S–IC(A)*, of 17.5 M^{-1}s^{-1}, as shown in Figure 4.11B. However, the fact remains that BipA is not a proficient enzyme, especially in comparison to EF-G, whose catalytic efficiency is 12 uM^{-1}s^{-1} in the presence of the 70S ribosome (Mohr et al., 2000). The increase observed in the presence of an A-site tRNA may indicate a quicker turnover of GTP by BipA in response to the presence of this moiety, however, steady state kinetic analysis done for this work is not suitable to address this question directly.

4.5.5 Association of BipA and the 70S Ribosome in the Presence of an A-site tRNA

To determine whether the association of BipA and the 70S ribosome is modified in the presence of an A-site tRNA, *in vitro* ribosome-binding assays were utilized. Purified His-tagged BipA was incubated with the various 70S ribosome complexes described above in the presence of excess GMPPNP. The samples were then applied to a 1.1 M sucrose cushion and centrifuged. Fractions were collected above the cushion, representing the free protein, and from the bottom of the tube that had passed through the cushion, representing ribosome bound samples, and analyzed on SDS–polyacrylamide gel electrophoresis for the presence of BipA. As shown in Figure 4.12, similar to our previous studies, in the absence of nucleotide, BipA was unable to bind to the 70S ribosome. Interestingly, no change in the relative binding of BipA to ribosomes programmed with occupancy of either an acylated or deacylated tRNAs was observed. This was a surprising finding, but it corroborates our kinetic data where the presence of an A-site tRNA did not substantially alter the GTPase properties of the protein. However, it should be noted that the assembled ribosome complexes are known to be unstable. The A site is shallower and wider than the P or E sites on the ribosome and has a lower affinity for tRNA. As such, a fraction of BipA may not be able to bind if these complexes fall apart *in vitro*. This may be the why we observe partial binding of BipA to the 70S ribosomal complexes and why there are very few known structural models with a stably bound A-site tRNA.

Competition binding studies were done with the antibiotic puromycin to corroborate the presence of the BipA CTD in the A site of the 70S ribosome and these studies brought some unexpected results. The binding assays were done as described above, except puromycin (2 mM) was incubated with heat-activated 70S ribosomes for 3 minutes at 37°C. These complexes were then cooled to 30°C upon which BipA was added in the presence or absence of GMPPNP (2 mM), and allowed to bind the ribosome for 30 min. Puromycin, an aminonucleoside antibiotic, binds near the peptidyl transfer center (PTC) of the ribosome, overlapping with the A-site tRNA (Hansen et al., 2002). In fact, part of the molecule resembles the 3' CCA end of the aminoacylated tRNA. It is thought to act by inhibiting peptide-bond formation by perturbing or preventing the correct positioning of the aminoacylated ends of tRNAs in the PTC. The use of puromycin surprisingly resulted in complete BipA binding to the ribosome. As shown in Figure 4.12, there is no detectable unbound fraction of BipA. Intriguingly, the association of BipA with the 70S ribosome in the presence of puromycin is guanine nucleotide-independent.

SECTION 4.6 DISCUSSION

We have captured and structurally characterized a novel ribosomal complex: the 70S–BipA complex. Our results corroborates years of biochemical data and provides answers to previously unsolved questions (deLivron and Robinson, 2008; deLivron et al., 2009) on the importance of a variety of BipA amino acids. The binding site of BipA's Domain I and Domain II in our map overlap well with the G domains of elongation factors such as EF-G (Agrawal et al., 1998), and the amino acids conserved between BipA and the canonical elongation factors have proven to be essential for GTPase hydrolysis activity (deLivron et al., 2009). Thus, BipA's GTPase mechanism is likely to be similar to that EF-G, EF-Tu, and EF4.

The results of the puromycin binding assays are intriguing for a number of reasons. First of all, this is one of the few examples of an antibiotic inducing such a binding state of an elongation factor. Fusidic acid is known to bind to the ribosome-bound state of EF-G, preventing its dissociation from the 70S ribosome after GTP hydrolysis (Bodley et al., 1969). While the mechanism by which puromycin stabilizes BipA binding is unclear, a situation analogous to that between fusidic acid and EF-G seems unlikely as BipA is able to bind, in the presence of puromycin, independent of the guanine nucleotide species.

Secondly, these experiments support the requirement of an A-site tRNA for BipA's association with the ribosome. Puromycin, with high affinity to the A site of the ribosome, may interact productively with the BipA CTD. The majority of the interactions between the BipA CTD and the A-site tRNA in the cryo-EM reconstruction are between the large distal loop region and the acceptor stem arm of the tRNA. It is conceivable that the high amount of flexibility in the CTD allows BipA to sample the A site of the ribosome. Productive interaction with a moiety in the A site, such as an A-site tRNA or puromycin, may stabilize BipA binding. By negating the requirement for the A-site tRNA, puromycin is changing the structure and the dynamics of the ribosome into a state that supports BipA association.

Our biochemical GTPase assays reveal that BipA's catalytic efficiency is highest in the presence of an A-site tRNA. Particularly, the three- to four-fold increases in BipA catalytic efficiency in both the presence of acylated and deacylated A-site tRNA are novel, unexpected findings. We believe the presence of an A-site tRNA may help to additionally stabilize the binding, allowing BipA to have longer contact with the GTPase-associated center on the 50S subunit, resulting in greater sampling of the GAC. The CTD is especially important for binding because of its interactions with the A-site tRNA and the surrounding helices of the 23S rRNA. Point mutations and deletions that have resulted in abrogation of binding of the protein (deLivron et al., 2009) are shown in our model at important regions of the CTD for A-site tRNA interaction. As we lack density for the 3' CCA end of the A-site tRNA in our reconstruction, there may be a mixture of acylated and deacylated tRNAs in our complex. Studies have suggested that because a deacylated A-site tRNA cannot be accommodated within the PTC, the tRNA could sample a large conformational space (Whitford et al., 2010) before dissociating from the 70S ribosome. We suggest that BipA may be able to bind to a deacylated A-site tRNA due to this unique attribute, which allows BipA's flexible distal loop to establish optimal interaction with the tRNA.

RelA is well-known enzyme capable of sensing a deacylated A-site tRNA. RelA is responsible for the production of guanosine tetraphosphate, (p)ppGpp, the universal alarmone in the bacterial stringent response. The production of this nucleotide triggers cascades of stress response pathways. A recent cryo-EM reconstruction of the 70S–RelA complex revealed that RelA binds proximal to the A-site of the 70S and strongly interacts with a highly distorted A-site tRNA which resembles the one observed in the A/T state (Agirrezabala et al., 2013). Thus, the conformation of the A-site tRNA resulting from interaction with BipA on the ribosome is different from that observed upon interaction with RelA. Previous studies show that RelA binding is optimal when there is an uncharged tRNA in the A site (Payoe and Fahlman, 2011; Agirrezabala et al., 2013). Here in our study, we suggest that BipA can also sense an A-site tRNA, aminoacylated or deacylated.

The mechanism of BipA binding and the structure of the 70S–BipA complex have been elusive. While studies have characterized BipA's participation in various stress (Wang et al., 2008; Neidig et al., 2013), stringent (Pfennig and Flower, 2001), and pathogenicity pathways (Grant et al., 2003), the exact mechanism by which BipA modulates gene expression has not been elucidated. BipA seemingly modulates gene expression levels (Krishnan and Flower, 2008), but contains no

DNA- or RNA-binding motifs.

Our model of the 70S–BipA complex brings newfound insights to the binding of BipA to the 70S ribosome. As BipA has been known to participate in stringent response pathways, BipA's ability to bind the 70S ribosome in the presence of an A-site tRNA may be crucial for its physiological function. It is tempting to speculate that BipA could be modulating the stringent response, not by recognizing specific genes or mRNA sequences, but through its interaction with the A-site tRNA. If BipA's stable binding also helps to stabilize a deacylated A-site tRNA, then the uncharged tRNA may be held in the A site long enough for other proteins, such as RelA, to bind and recognize the deacylated tRNA. In this scenario, BipA, by providing a scaffold for RelA, would indrectly cause a cascade of gene expression specific to stress response and adaptation pathways. As yet, no biochemical or structural studies have attempted to find a link between BipA and RelA. While there is still much to be elucidated, our reconstruction and model of the 70S–BipA complex provides answers to previous unexplained results as well as present clues to BipA's mechanism.

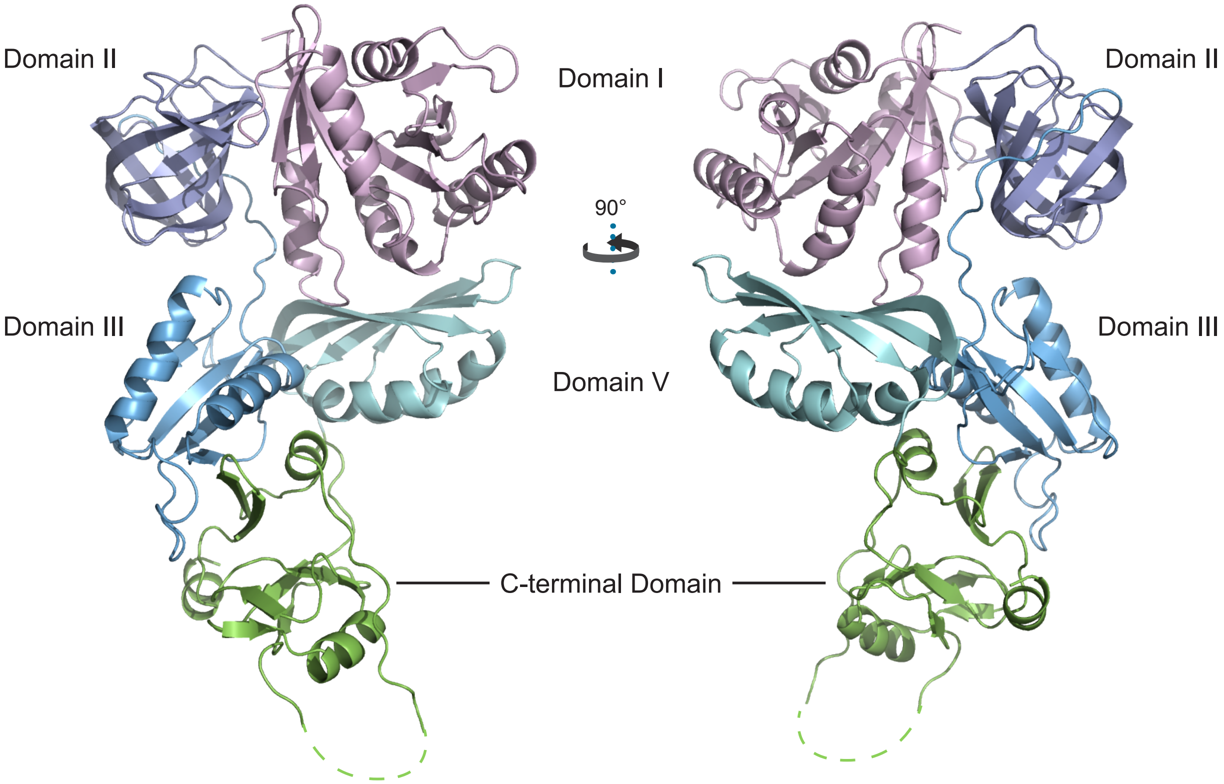


Figure 4.1. The X-ray Structure of BipA. The X-ray structure of *S. enterica* BipA, absent of any nucleotide, was solved to 2.7 Å. Overall, Domains I and II exhibit the canonical structure of the GTPase and β-barrel domain, respectively, that are shared by all elongation translational GTPases (trGTPases). Domains III and V resemble the tertiary architectures of the corresponding homologous domains in EF-G and EF4. The C-terminal domain topology is novel and previously uncharacterized. Several undefined regions of the structure could not be visualized: the switch I (residues 34-55), a large flexible loop in the C-terminal domain (CTD) (residues 541-552), shown as a dashed line above, and seven residues at the C-terminus of the protein (residues 600-607). Domains are colored according to the scheme presented in Figure 2.6.

	Native BipA	Se-met BipA
Data Collection and Phasing		
Space Group	$P2_1$	$P2_1$
Molecules per asymmetric unit	2	2
Unit Cell Parameters	a=89.4 Å, b=84.0 Å, 95.6 Å, b = 106.2 °	a=90.0 Å, b=83.3 Å, 96.2 Å, b = 106.2 °
Resolution (Å)	15-2.7 (2.8-2.7)	30-2.4 (2.47-2.4)
Wavelength (Å)	0.97934	0.9783
Unique reflections (N, F > 0)	34,022	46,063
Completeness, %	96.4 (87.8)	82.3 (35.2)
I/σI	15.1 (4.2)	13.7 (2.7)
R_{merge}[a]	0.089 (0.207)	0.096 (0.277)
Multiplicity	5.0 (2.8)	5.7 (1.9)
Figure-of-merit	0.88 (0.94)	0.31 (0.11)

Refinement	
Reflections, work/free	32,246 / 1,776
Protein Atoms	8,842
Water Atoms	44
R_{work}[b]	25.40%
R_{free}[c]	29.60%
RMSD bond lengths (°)	0.0078
RMSD bond angles (°)	1.54

Ramachandran Plot	
Most favored (%)	92.9
Allowed (%)	6.3
Disallowed (%)	0.8

Table 4.1. Crystallographic statistics of data collection and refinement. Highest shell values are in parenthesis. Completeness and R_{merge} are given for all data and for data in the highest resolution shell.

[a] $R_{merge} = \sum_{hkl}\sum_{i}|I_{hkl},i - I_{hkl}|/\sum_{hkl}\sum i_{Ihkl},i$, where I_{hkl},i is the i^{th} observed intensity and I_{hkl} is the average intensity over symmetry equivalent measurements

[b] $R_{work} = \Sigma_{hkl} |F_o\text{-}F_c| / \Sigma_{hkl}F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively for all reflections *hkl* used in refinement.

[c] R_{free} is calculated for 5% of the data that were not used in refinement.

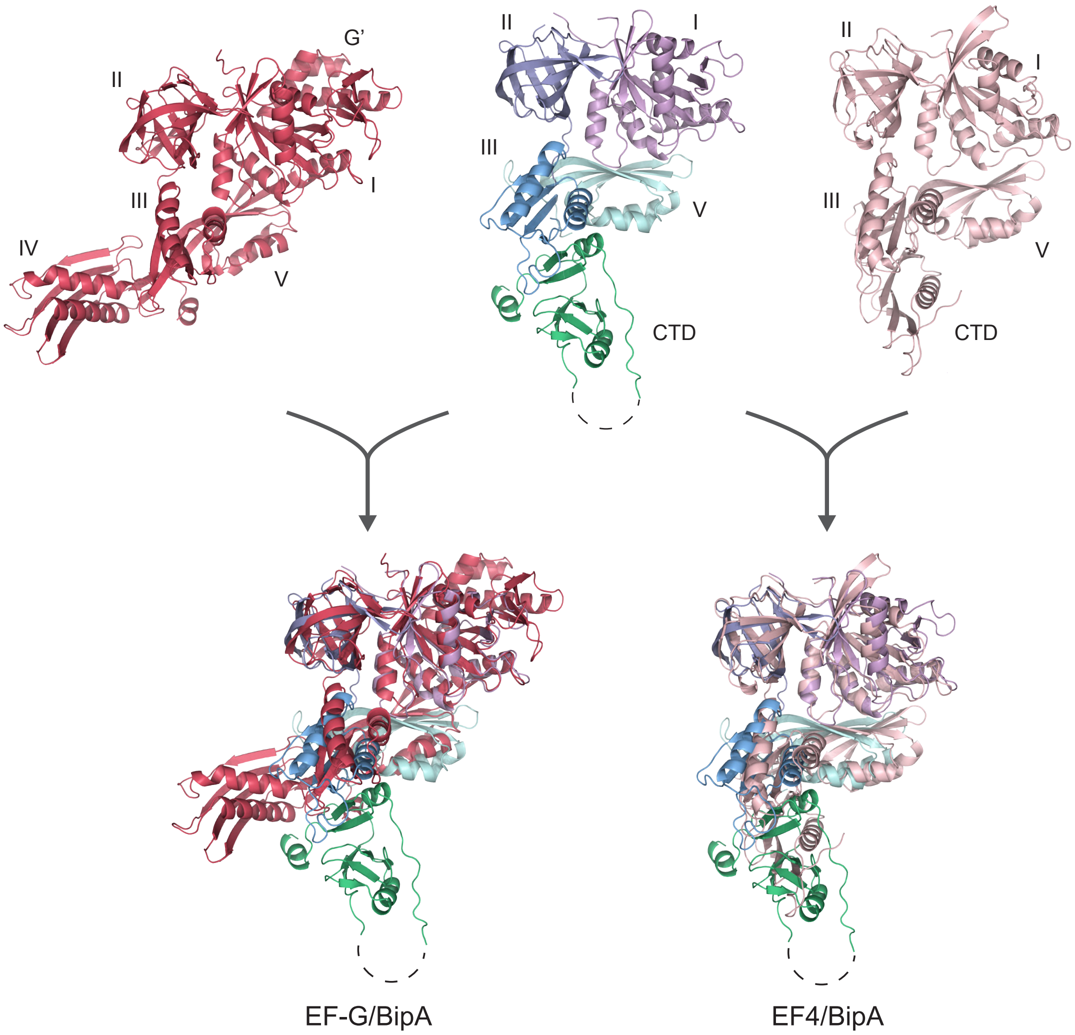


Figure 4.2. Superimposition of the X-ray structures of unbound EF-G, BipA, and EF4. (A) The crystal structures of EF-G (PDB: 1FNM), BipA (PDB: 3BV5), and EF4 (PDB: 3CB4) are shown individually. (B) The crystal structure of EF-G superimposed on BipA shows that BipA lacks a G' domain and a homologous EF-G domain IV. In addition, the CTD of BipA and EF-G domain IV are positioned in opposite directions in relation to the rest of the protein. The BipA CTD may provide new contacts to the ribosome. (C) The CTD of BipA and LepA have different topologies. Thus, BipA may have distinct contacts to the ribosome from either LepA or EF-G (*11, 12*). The LepA structure, however, lacks the final 50 amino acids at the c-terminal. EF-G and LepA are colored differently for comparison purposes, the domain colors of BipA follow the color scheme presented in Figure 4.1.

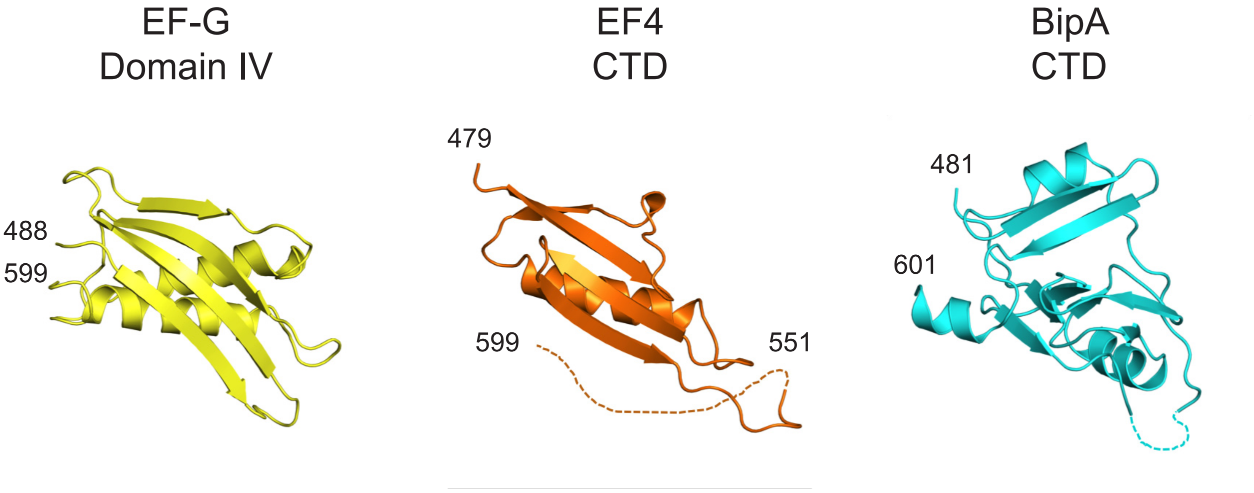


Figure 4.3. Comparison of the structural topologies of EF-G Domain IV, EF4 CTD, and BipA CTD. Domain IV in EF-G and the CTD of EF4 both have an antiparallel β-sheet with either one or two helices on one side. However, the two domains do not share sequence homology. The CTD of BipA looks nothing like these domains, suggesting that BipA may have specific uncharacterized interactions with the ribosomes.

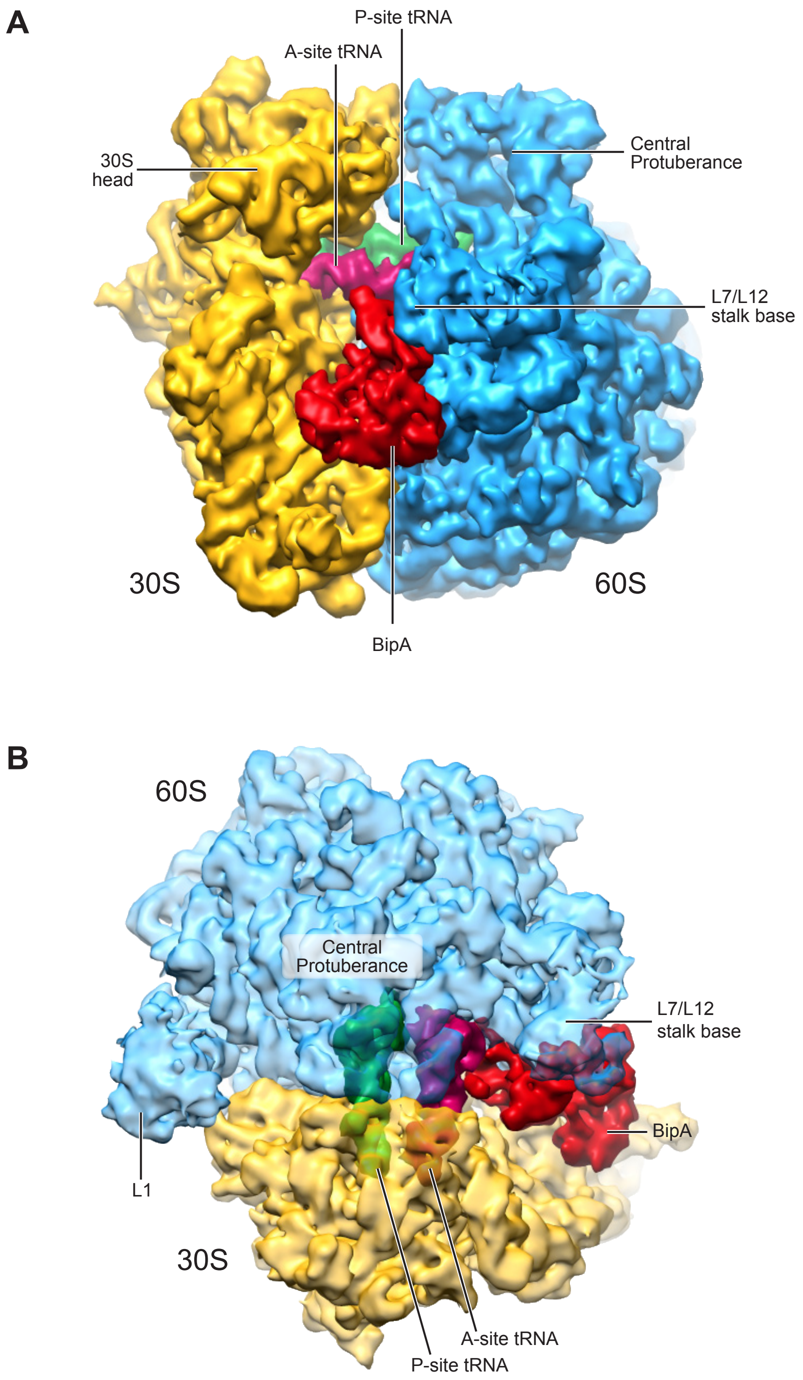


Figure 4.4. Cryo-EM reconstruction of the 70S–BipA complex. The segmented map of the 70S–BipA complex is shown here in two views: (A) side view and (B) front view with transparent subunits. Visualized in the reconstruction is a complete 70S ribosome, segmented into the 50S subunit (blue), and the 30S subunit (yellow), BipA (red), A-site tRNA (magenta), and P-site tRNA (green). Various ribosomal landmarks are labeled for orientation.

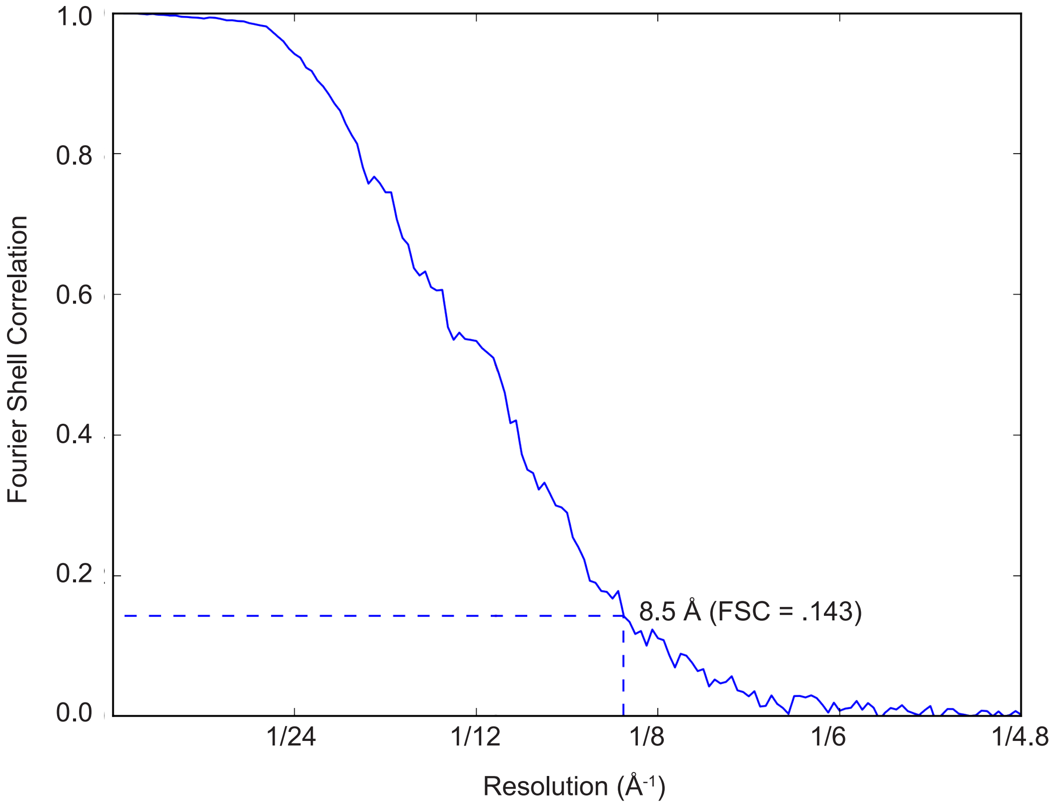


Figure 4.5. The Fourier Shell Correlation (FSC) curve for the 70S–BipA reconstruction. The ~23,000 particles assigned to the 70S–BipA class were randomly divided into half-sets using RELION. Separate 3D volumes were reconstructed from each half-set and the cross-correlation values between the half-volumes across Fourier shells were calculated. The resolution of the 70S–BipA reconstruction was determined using the FSC = .143 cutoff, according to the gold standard protocol (Henderson et al, 2012).

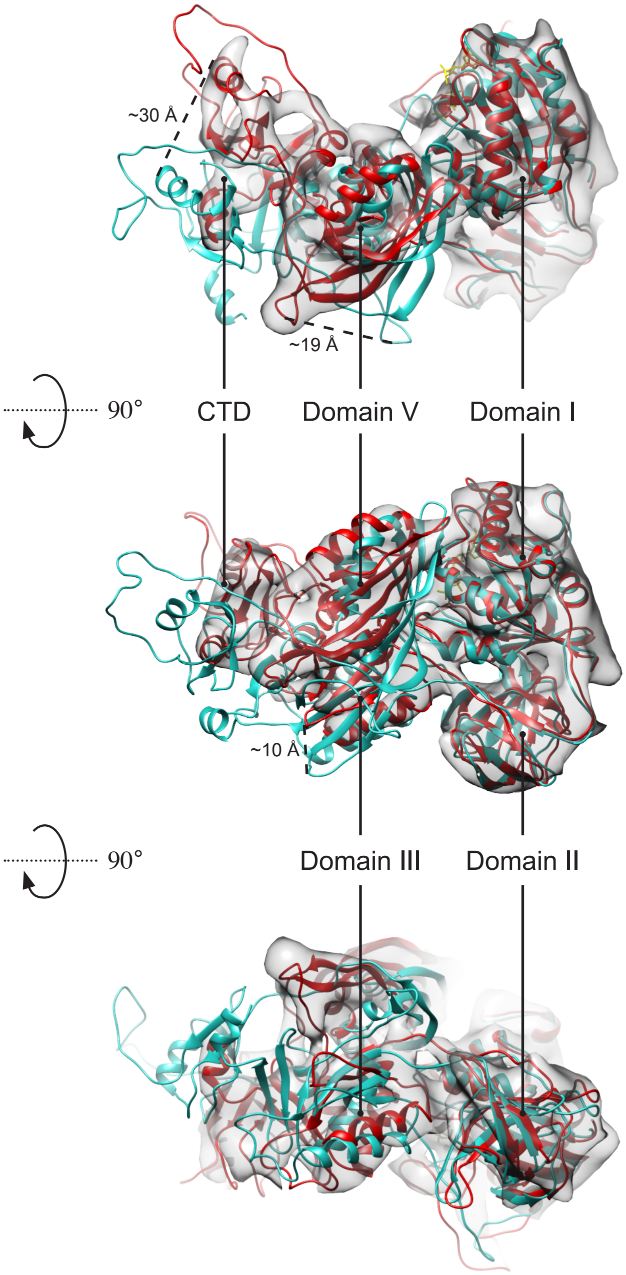


Figure 4.6.Fitting of the BipA X-ray structure into the BipA cryo-EM density. MDFF was employed to fit the BipA X-ray structure (blue) into the protein's cryo-EM density (grey), to produce the final quasi-atomic model of BipA in a ribosome-bound state (red). There is little change in the position of Domains I and II before and after fitting. Domains III and V shift by ~10 Å and 19 Å, respectively, towards the base of the L7/L12 stalk to fit their corresponding density. The most dramatic difference between the x-ray structure and the fitted structure occurs in the c-terminal domain, which shifts by ~30 Å to fit into the corresponding cryo-EM density.

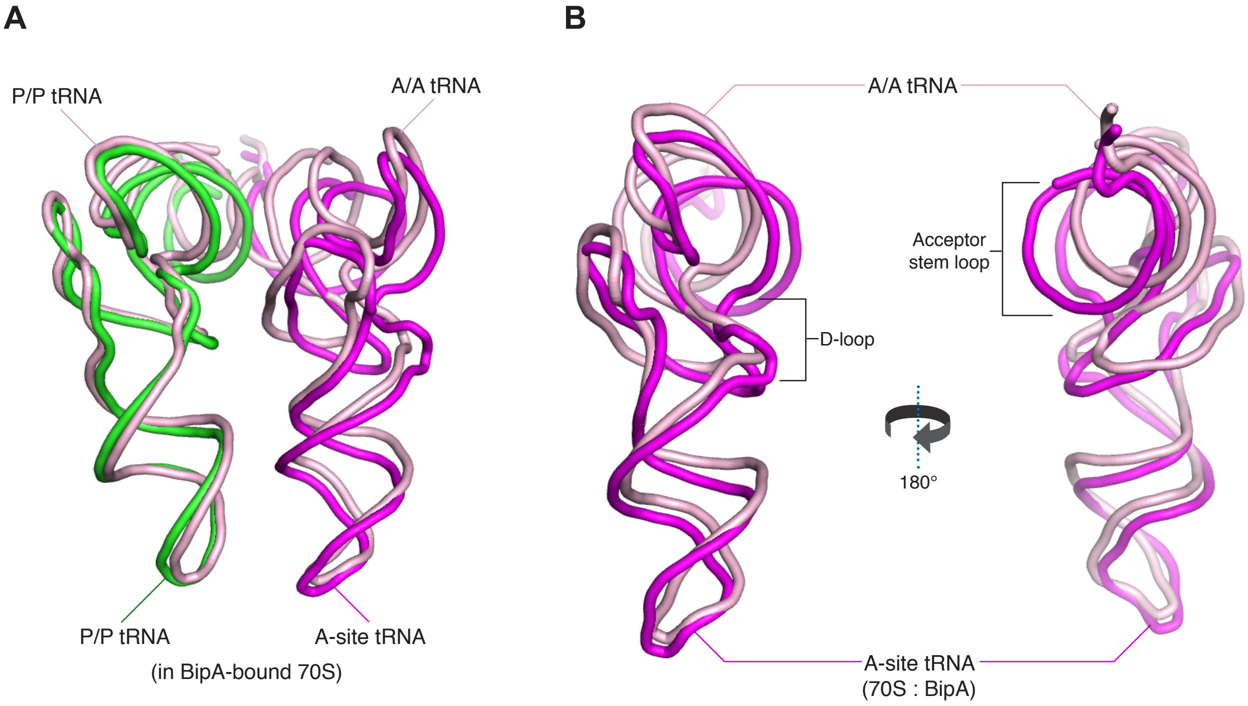


Figure 4.7. Conformations of the A-site and P-site tRNAs in the 70S–BipA complex. Ribbon diagrams of A-site and P-site tRNA obtained by flexible fitting, are shown in magenta and green, respectively. In peach are the corresponding A/A and P/P tRNA configurations determined in a previous study 2. In the 70S–BipA complex, while the P-site tRNA remains in the P/P state (A), the A-site tRNA is subjected to an 8-Å deformation throughout the D-loop and the acceptor stem loop (B), away from the A/A configuration, towards BipA.

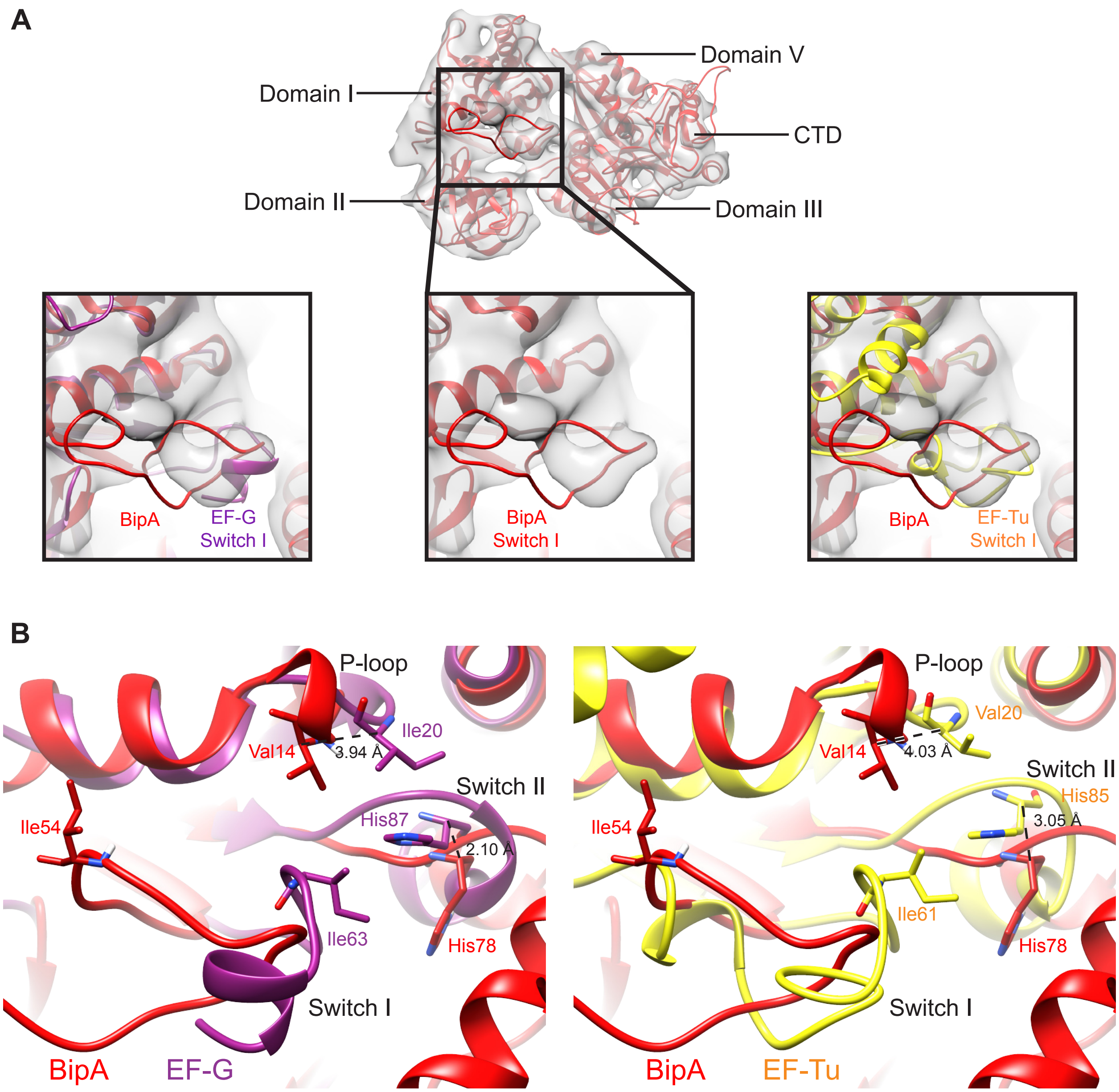


Figure 4.8. The switch I and switch II regions of BipA, EF-G, and EF-Tu. Residues 34-55 of switch I region are undefined in the BipA X-ray structure. Thus, a flexible loop was filled in to create a stereochemically complete X-ray structure for MDFF. (A) The BipA switch I region has clear density within the reconstruction. However, MDFF was unable to fit the flexible loop into the corresponding density due to the poor initial model. The X-ray structures of EF-G (PDB: 4JUW) and EF-Tu (PDB: 2XQD), both in a GTPase activated state with an ordered switch I loop, were superimposed onto the fitted model of BipA. EF-G and EF-Tu show good agreement in the position of their switch I loops with the BipA switch I density. (B) Comparison of the amino acids necessary for BipA GTPase activity (Val14, Ile54, and His78) with the homologous residues in EF-G and EF-Tu show high agreement in the positions of the P-loop amino acid (BipA: Val14, EF-G: Ile20, EF-Tu: Val20) and Switch II catalytic histidine (BipA: His78, EF-G: His87, EF-Tu: His85).

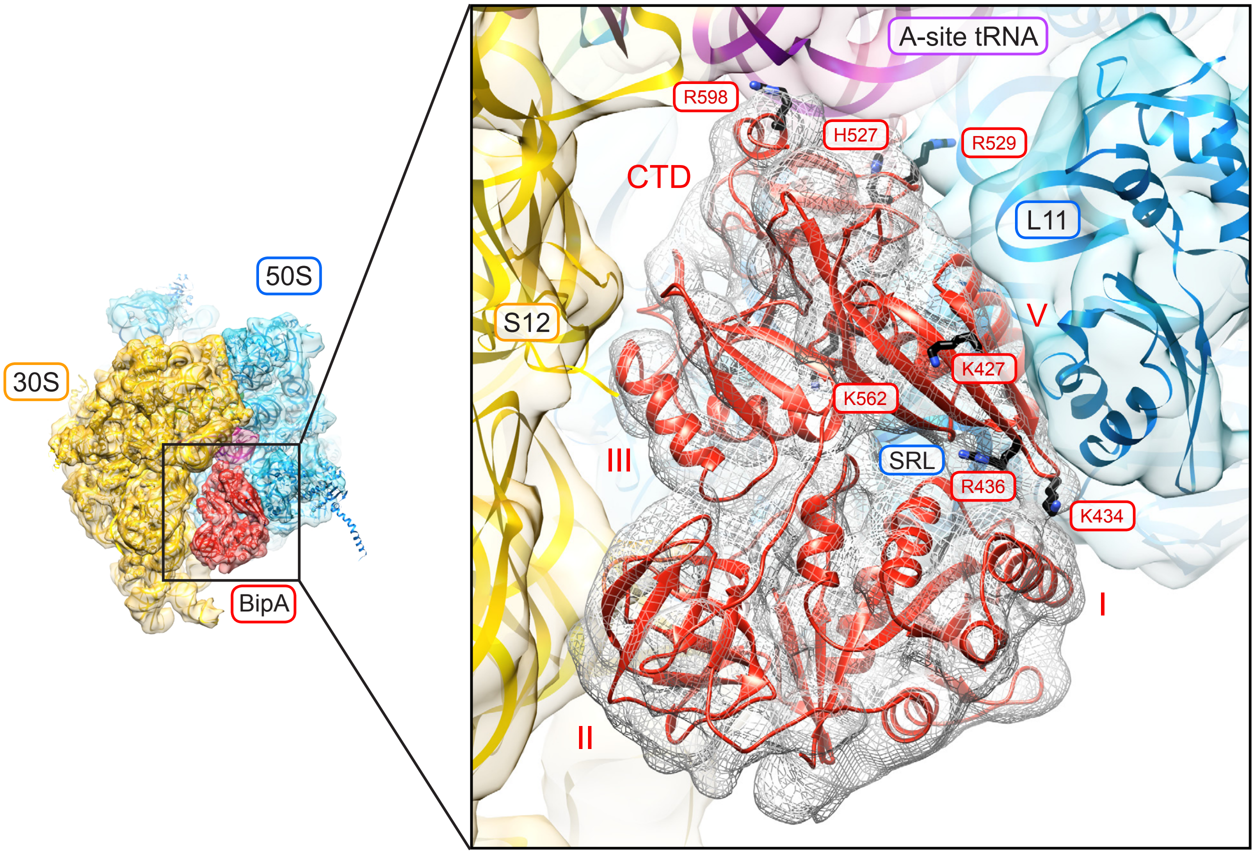


Figure 4.9. The binding site of BipA on the ribosome. The X-ray structure of BipA was flexibly fitted into its corresponding cryo-EM density to elucidate the binding contacts of BipA. BipA is shown in red, 50S subunit proteins and 23S rRNA in blue, and 30S subunit proteins and 16S rRNA in yellow. Domains I and II, highlighted in (A), show that BipA exhibits similar binding contacts as several other elongation factors. Amino acids found to be important for BipA binding in point mutational studies (deLivron et al, 2009) are shown as stick representations.

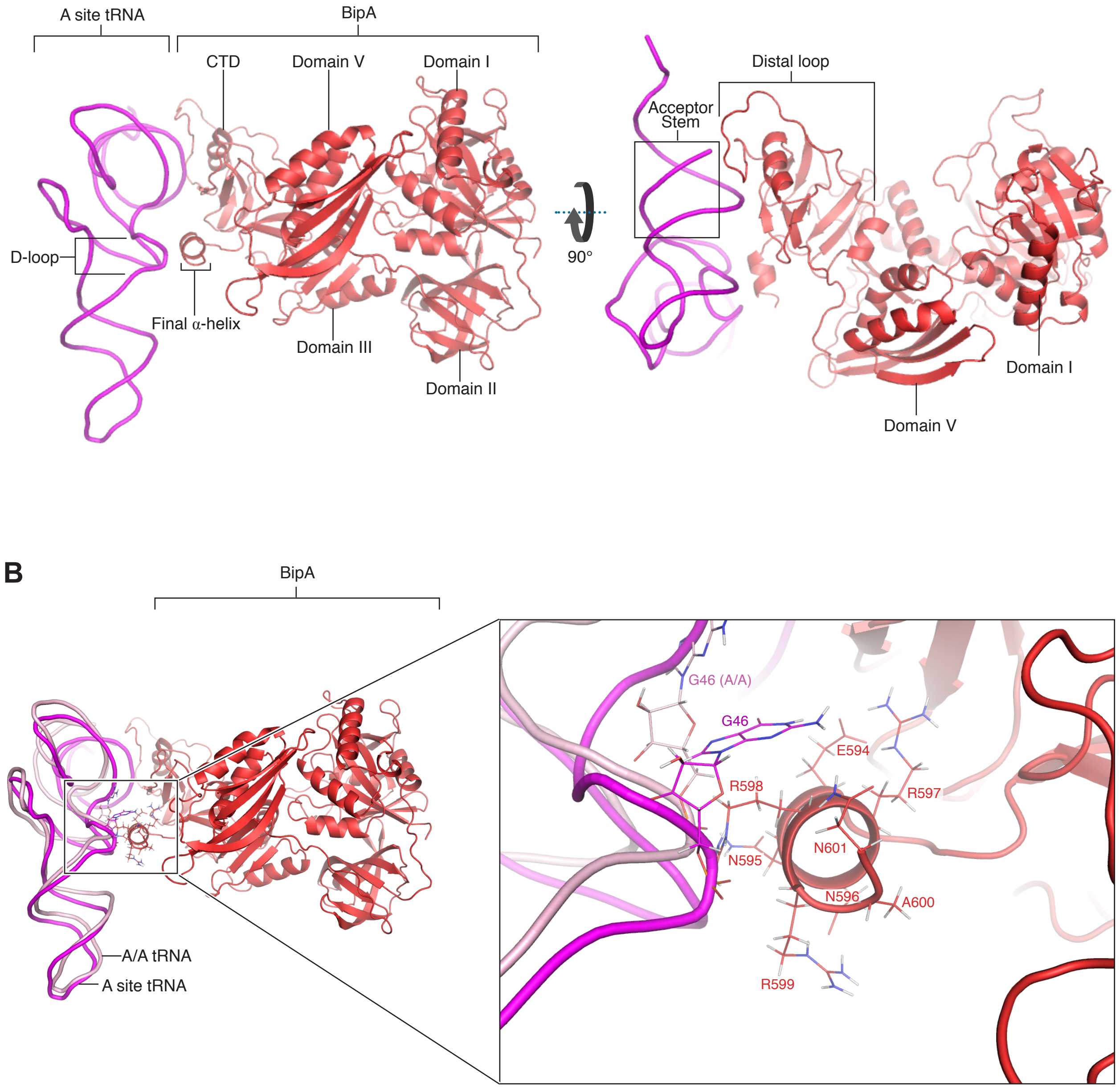


Figure 4.10. Interaction between A-site tRNA and BipA's CTD. The BipA CTD has extensive interactions with the A-site tRNA, namely in two regions: (A) the distal loop with acceptor stem loop, and (B) the final α-helix with the D-loop. In (B), a close-up of the various possible interactions between the final α-helix with the D-loop. A canonical A/A tRNA (in peach) is shown for comparison. Mutations in the final α-helix results in significant reduction of BipA binding (deLivron 2009).

A

Ribosomal Species	Vmax	K_M	k_{cat}	k_{cat}/K_M
	(pmol of PO_4/pmol of BipA/min)	(mM)	(h^{-1})	($M^{-1}s^{-1}$)
BipA	0.30 ± 0.01	1.11 ± 0.12	18.00 ± 0.60	4.55 ± 0.50
Bip + 70S	0.96 ± 0.09	1.67 ± 0.30	57.60 ± 5.10	9.66 ± 1.93
BipA–70S IC	1.33 ± 0.19	2.50 ± 0.62	79.80 ± 11.40	8.94 ± 2.56
BipA–70S IC(A)	0.61 ± 0.03	0.84 ± 0.12	36.60 ± 1.80	12.20 ± 1.87
BipA–70S IC(A)*	1.04 ± 0.09	1.00 ± 0.23	62.40 ± 5.40	17.47 ± 4.48

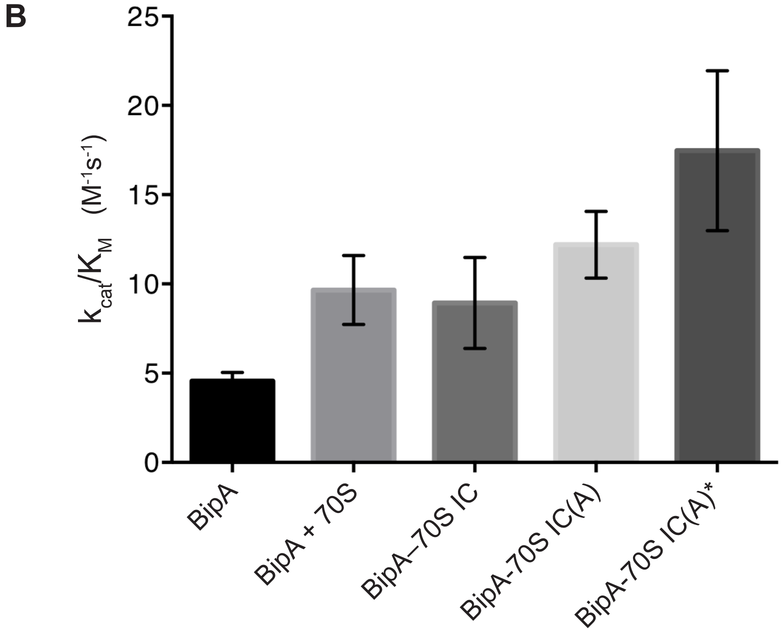
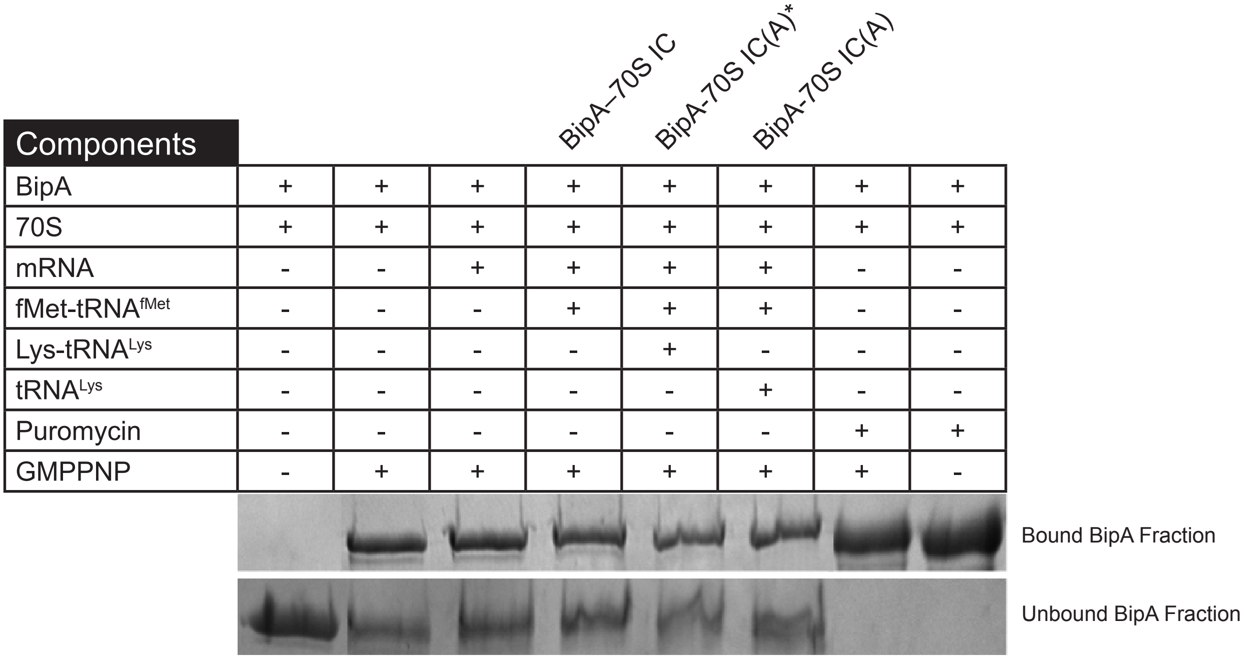


Figure 4.11. The GTPase activity of BipA in the presence of various ribosomal species. (A) tabulates the steady state kinetic parameters of BipA in the presence of various 70S species. As shown, the presence of an A-site tRNA, aminoacylated or deacylated, increases BipA's catalytic efficiency (k_{cat}/K_M) , graphed in (B). BipA exhibits the highest catalytic efficiency in the presence of an A-site tRNA. Ribosomal species are named as follows: BipA - the isolated protein; BipA + 70S - BipA reacted with purified 70S ribosomes; BipA–70S IC - BipA reacted with the an assembly of 70S–P-site fMet-$tRNA^{fMet}$; BipA–70S IC(A) - BipA reacted with the a pre-assembled 70S–P-site fMet-$tRNA^{fMet}$–A-site $tRNA^{Lys}$; BipA–70S IC(A)* - BipA reacted with the a pre-assembled 70S–P-site fMet-$tRNA^{fMet}$–A-site Lys-$tRNA^{Lys}$.



Components				BipA–70S IC	BipA-70S IC(A)*	BipA-70S IC(A)		
BipA	+	+	+	+	+	+	+	+
70S	+	+	+	+	+	+	+	+
mRNA	-	-	+	+	+	+	-	-
fMet-tRNA^{fMet}	-	-	-	+	+	+	-	-
Lys-tRNA^{Lys}	-	-	-	-	+	-	-	-
tRNA^{Lys}	-	-	-	-	-	+	-	-
Puromycin	-	-	-	-	-	-	+	+
GMPPNP	-	+	+	+	+	+	+	-

Figure 4.12. Binding of BipA in the presence of various ribosomal species. *In vitro* ribosome-binding assays were utilized to determine whether various ribosomal complexes had an effect on the the association of BipA with the 70S ribosome. Purified His-tagged BipA was incubated with the various 70S ribosome complexes in the presence of excess GMPPNP. Three specific complexes, are demarcated above the chart: BipA–70S IC, BipA–70S IC(A), and BipA–70S IC(A)*. These three complexes correspond to the same complexes introduced in Figure 4.9. In the presence of puromycin, BipA completely binds to the 70S ribosome in a guanine nucleotide independent manner.

REFERENCES

Aevarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., al-Karadaghi, S., Svensson, L., Liljas, A., 1994. Three-dimensional structure of the ribosomal translocase: elongation factor G from *Thermus thermophilus*. EMBO J. 13, 3669-3677.

Agirrezabala, X., Liao, H.Y., Schreiner, E., Fu, J., Ortiz-Meoz, R.F., Schulten, K., Green, R., Frank, J., 2012. Structural characterization of mRNA-tRNA translocation intermediates. Proc Natl Acad Sci USA 109, 6094-6099.

Agirrezabala, X., Fernandez, I.S., Kelley, A.C., Cartón, D.G., Ramakrishnan, V., Valle, M., 2013. The ribosome triggers the stringent response by RelA via a highly distorted tRNA. EMBO Rep. 14, 811-816.

Agrawal, R.K., Penczek, P.A., Grassucci, R.A., Frank, J., 1998. Visualization of elongation factor G on the *Escherichia coli* 70S ribosome: The mechanism of translocation. Proc Natl Acad Sci USA 95, 6134-6138.

al-Karadaghi, S., Aevarsson, A., Garber, M., Zheltonosova, J., Liljas, A., 1996. The structure of elongation factor G in complex with GDP: conformational flexibility and nucleotide exchange. Structure 4, 555-565.

Baker, L.A., Rubinstein, J.L., 2011. Edged watershed segmentation: a semi-interactive algorithm for segmentation of low-resolution maps from electron cryomicroscopy. J. Struct. Biol. 176, 127-132.

Blaha, G., Stelzl, U., Spahn, C.M.T., Agrawal, R.K., Frank, J., Nierhaus, K.H., 2000. Preparation of functional ribosomal complexes and effect of buffer conditions on tRNA positions observed by cryoelectron microscopy. Meth. Enzymol. 317, 292-309.

Bodley, J.W., Zieve, F.J., Lin, L., Zieve, S.T., 1969. Formation of the ribosome-G factor-GDP complex in the presence of fusidic acid. Biochemical and Biophysical Research Communications 37, 437-443.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. .

deLivron, M.A., Robinson, V.L., 2008. *Salmonella enterica serovar Typhimurium* BipA exhibits two distinct ribosome binding modes. J. Bacteriol. 190, 5944-5952.

deLivron, M.A., Makanji, H.S., Lane, M.C., Robinson, V.L., 2009. A novel domain in translational GTPase BipA mediates interaction with the 70S ribosome and influences GTP hydrolysis. Biochemistry 48, 10533-10541.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica D 60, 2126-2132.

Evans, R.N., Blaha, G., Bailey, S., Steitz, T.A., 2008. The structure of LepA, the ribosomal back translocase. Proc Natl Acad Sci USA 105, 4673-4678.

Farris, M., Grant, A., Richardson, T., O'Connor, C., 1998. BipA: a tyrosine-phosphorylated GTPase that mediates interactions between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells. Mol Microbiol 28, 265-279.

Grant, A., Farris, M., Alefounder, P., Williams, P., Woodward, M., O'Connor, C., 2003. Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic *Escherichia coli* (EPEC). Mol Microbiol 48, 507-521.

Grassucci, R.A., Taylor, D.J., Frank, J., 2007. Preparation of macromolecular complexes for cryo-electron microscopy. Nat Protoc 2, 3239-3246.

Hansen, J.L., Schmeing, T.M., Moore, P.B., Steitz, T.A., 2002. Structural insights into peptide bond formation. Proceedings of the National Academy of Sciences of the United States of America 99, 11670-11675.

Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., Egelman, E.H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S.J., Medalia, O., Penczek, P.A., Rosenthal, P.B., Rossmann, M.G., Schmid, M.F., Schröder, G.F., Steven, A.C., Stokes, D.L., Westbrook, J.D., Wriggers, W., Yang, H., Young, J., Berman, H.M., Chiu, W., Kleywegt, G.J., Lawson, C.L., 2012. Outcome of the first electron microscopy validation task force meeting. Structure 20, 205-214.

Holm, L., Rosenstrom, P., 2010. Dali server: conservation mapping in 3D. Nucleic acids research 38, W545-549.

Krishnan, K., Flower, A.M., 2008. Suppression of ΔbipA phenotypes in *Escherichia coli* by abolishment of pseudouridylation at specific sites on the 23S rRNA. J. Bacteriol. 190, 7675-7683.

Kucukelbir, A., Sigworth, F.J., Tagare, H.D., 2013. Quantifying the local resolution of cryo-EM density maps. Nat Meth 11, 63-65.

Lanzetta, P.A., Alvarez, L.J., Reinach, P.S., Candia, O.A., 1979. An improved assay for nanomole amounts of inorganic phosphate. Anal. Biochem. 100, 95-97.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26, 283-291.

Leshin, J.A., Rakauskaitė, R., Dinman, J.D., Meskauskas, A., 2010. Enhanced purity, activity and structural integrity of yeast ribosomes purified using a general chromatographic method. RNA Biol 7, 354-360.

Margus, T., Remm, M., Tenson, T., 2007. Phylogenetic distribution of translational GTPases in bacteria. BMC Genomics 8, 15.

Meskauskas, A., Leshin, J.A., Dinman, J.D., 2011. Chromatographic purification of highly active yeast ribosomes. J Vis Exp.

Mohr, D., Wintermeyer, W., Rodnina, M.V., 2000. Arginines 29 and 59 of elongation factor G are important for GTP hydrolysis or translocation on the ribosome. EMBO J. 19, 3458-3464.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. Acta Crystallographica Section D D53, 240-255.

Neidig, A., Yeung, A.T., Rosay, T., Tettmann, B., Strempel, N., Rueger, M., Lesouhaitier, O., Overhage, J., 2013. TypA is involved in virulence, antimicrobial resistance and biofilm formation in *Pseudomonas aeruginosa*. BMC Microbiol 13, 77.

Otwinowski, Z., Minor, W., 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology 276, 307-326.

Payoe, R., Fahlman, R.P., 2011. Dependence of RelA-mediated (p)ppGpp formation on tRNA identity. Biochemistry 50, 3075-3083.

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

Pfennig, P., Flower, A., 2001. BipA is required for growth of *Escherichia coli* K12 at low temperature. Mol Genet Genomics 266, 313-317.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W., 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.

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

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., Eisenberg, D., 2006. Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proceeding of the National Academy of Sciences 103, 8060-8065.

Tanner, D.E., Ma, W., Chen, Z., Schulten, K., 2011. Theoretical and computational investigation of flagellin translocation and bacterial flagellum growth. Biophysical journal 100, 2548-2556.

Terwilliger, T.C., Berendzen, J., 1999. Automated MAD and MIR structure solution. Acta Crystallographica D55, 849-861.

Terwilliger, T.C., 2003. Automated main-chain model-building by template-matching and iterative fragment extension. Acta crystallographica D59, 45-49.

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

Trabuco, L.G., Villa, E., Schreiner, E., Harrison, C.B., Schulten, K., 2009. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods 49, 174-180.

Wang, F., Zhong, N.-Q., Gao, P., Wang, G.-L., Wang, H.-Y., Xia, G.-X., 2008. SsTypA1, a chloroplast-specific TypA/BipA-type GTPase from the halophytic plant *Suaeda salsa*, plays a role in oxidative stress tolerance. Plant Cell Environ. 31, 982-994.

Whitford, P.C., Geggier, P., Altman, R.B., Blanchard, S.C., Onuchic, J.N., Sanbonmatsu, K.Y., 2010. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. Rna 16, 1196-1204.

CHAPTER 5:

Conclusion

I began the 70S–BipA project aiming to find new insights into the binding of the ubiquitously conserved BipA protein. As BipA shares high structural and sequence homology to the canonical elongation factors, I expected a reconstruction of the 70S–BipA complex to resemble the structure of 70S–EF-Tu and 70S–EF-G complexes. However, previous studies (deLivron and Robinson, 2008) predicted that BipA has unique interactions and binding modes to the ribosome that is not displayed by any other elongation factor. This is especially true for BipA's CTD, which has no homology to any known protein. Indeed, our model of the 70S–BipA complex provides structurally characterization of a novel ribosomal complex. Our map helps to corroborate years of biochemical studies and suggests the importance of the BipA CTD in not only BipA binding, but also in the promotion of BipA's GTPase activity. The surprising structural findings of the 70S–BipA reconstruction, such as the simultaneous presence of the A-site tRNA in our complex, led us to design a series of biochemical assays that characterized the ribosomal species optimal for BipA binding to the 70S ribosome.

As BipA is a ubiquitous protein in bacteria and lower eukaryotes, and has an essential role in the bacterial stringent response, it presents an attractive target for drug research. BipA's novelty as a translational factor and stringent response regulator was suggested time and time again in previous studies (Qi et al., 1995; Farris et al., 1998a; Farris et al., 1998b; Grant et al., 2003; Vogt et al., 2011; Neidig et al., 2013). Elucidation of the 70S–BipA complex structure, BipA's X-ray structure, and characterization of the ribosomal complexes that optimizes its binding may help in the further development of antibiotics exploiting the protein's specific interactions with the 70S ribosome.

The idea that the 70S–BipA complex could indirectly induce adaptation pathways by acting as a scaffold for other proteins, such as RelA, is an exciting prospect. This scenario would neatly position BipA at the beginning of the stringent response and help to explain how BipA participates in such a variety of stress adaptation pathways. Biochemical studies are currently underway to find a link between RelA and the 70S–BipA complex. In these studies, the *S. enterica* RelA protein will be overexpressed and purified. Pull-down assays, performed as previously described (deLivron and Robinson, 2008), will be used to detect whether the purified RelA protein can bind to the 70S–BipA complex. Crystallization attempts of unbound RelA or a 70S–BipA complex have been unsuccessful. The only cryo-EM reconstruction of the 70S–RelA complex (Agirrezabala et al., 2013) visualizes only 80% of the protein's mass, prompting speculation as to whether the RelA ribosome-bound state captured by cryo-EM is truly the functional state in which the protein binds to the ribosome (Starosta et al., 2014).

However, the 70S–BipA complex is only one of two ribosome–BipA complexes. Recall that BipA has two binding modes, one to the complete 70S ribosome in the presence of GTP and the other with the isolated 30S in the presence of ppGpp. The second binding mode was discovered by deLivron and coworkers (deLivron and Robinson, 2008), but characterization of the 30S–BipA complex remains unstudied. Thus, there is yet no knowledge as to how and in what conformation BipA binds to the free, isolated 30S subunit.

My initial attempts, at the beginning of my dissertation research, to image an *in vitro* sample of purified 30S subunits complexed with BipA–ppGpp were substantially hindered by sample aggregation and contamination. Biochemical interventions, such as varying the magnesium or salt concentration, were unsuccessful in alleviating such problems. This meant that per micrograph, relatively few particles were free and isolated. An extraordinary amount of data would have been required to collect enough non-aggregated particle images. Particle verification, at that time, was still done by manual visual inspection of the candidate particle images and an unacceptable amount of effort and time would have been required. However, with the development of AutoPicker to which I contributed, processing the large amounts of required data is no longer a daunting task. Also, use of RELION has been successful in classifying recent complexes of the small eukaryotic subunit (Hashem et al., 2013). Thus, a project to visualize the 30S–BipA complex would be much easier to attempt now.

A project focused on elucidating the structure of the 30S–BipA complex is exciting as it will provide answers to another piece of the BipA puzzle. Biochemical studies have been unable to elucidate the conformational changes that must occur within BipA in order to induce its binding to the isolated 30S subunit. It is suggested that 30S–bound state of BipA may exhibit a completely novel conformation than the 70S–bound state (deLivron and Robinson, 2008). This idea is further reinforced by our 70S–BipA reconstruction, which shows few interactions between BipA and the 30S subunit. In fact, the research represent here suggests that BipA requires susbtantial amounts of interactions between the 70S ribosome and the A-site tRNA in order to stabilize its binding. Absent of a 50S subunit or A-site tRNA, additional interactions between the 30S subunit and BipA would be necessary to stabilize BipA binding to the subunit. Thus, elucidation of the 30S–BipA structure will provide insights as to how BipA conformationally changes when bound to ppGpp. This structure will subsequently provide insights as to how BipA responds to increasing levels of ppGpp during stress.

The mechanism by which puromycin stabilizes BipA binding remains elusive. As puromycin binds at the PTC of the 50S subunit, the CTD of BipA is the only domain of the protein that can interact with the antibiotic. We can only speculate that perhaps the interactions between BipA and puromycin stabilizes a conformation of the CTD that has not been visualized in both the X-ray structure and our cryo-EM reconstruction. A new extension to the 70S–BipA cryo-EM project has begun, aimed at reconstructing the 70S–BipA–puromycin complex. In addition to puromycin, biochemical and structural studies into the effects of other antibiotics on the binding of BipA, may further insights to how the CTD of BipA interacts with the ribosome.

Structural characterizations of ribosome–BipA complexes provide answers to one part of a larger BipA story. There is still much to learn about this mysterious protein. Studies have been unable to determine how the binding of BipA affects cellular viability and expression of stress-related proteins. Studies have clearly shown that BipA is essential for bacterial stress adaptation to adverse cellular conditions, but no knowledge has been garnered about BipA's specific position in any stress pathway. Ongoing structural and biochemical studies will be needed to elucidate BipA's specific physiological function.

Finally, the work presented here not only served to characterize the structure of the 70S–BipA complex, but also showcased several exceptional advancements in the field of cryo-EM. I have shown that AutoPicker can remove the substantial cost of time investment and the user subjectivity of manual particle verification, with no detriment to the quality of the particle dataset used for reconstruction. In addition, RELION has replaced the imperfect supervised classification technique to allow an inventory of conformations to be classified and visualized from a single sample. Finally, as introduced in Chapter 1, the new direct electron detectors are enabling the collection of data with unprecedented quality. Collectively, these advancements harken in a new age in the cryo-EM field, where a high-throughput single-particle cryo-EM workflow is now readily accessible and producing near-atomic resolution reconstructions.

REFERENCES

Agirrezabala, X., Fernandez, I.S., Kelley, A.C., Cartón, D.G., Ramakrishnan, V., Valle, M., 2013. The ribosome triggers the stringent response by RelA via a highly distorted tRNA. EMBO Rep. 14, 811-816.

deLivron, M.A., Robinson, V.L., 2008. *Salmonella enterica serovar Typhimurium* BipA exhibits two distinct ribosome binding modes. J. Bacteriol. 190, 5944-5952.

deLivron, M.A., Makanji, H.S., Lane, M.C., Robinson, V.L., 2009. A novel domain in translational GTPase BipA mediates interaction with the 70S ribosome and influences GTP hydrolysis. Biochemistry 48, 10533-10541.

Farris, M., Grant, A., Jane, S., Chad, J., O'Connor, C.D., 1998a. BipA affects Ca^{++} fluxes and phosphorylation of the translocated intimin receptor (Tir/Hp90) in host epithelial cells infected with enteropathogenic *E. coli*. Biochem. Soc. Trans. 26, S225.

Farris, M., Grant, A., Richardson, T., O'Connor, C., 1998b. BipA: a tyrosine-phosphorylated GT-Pase that mediates interactions between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells. Mol Microbiol 28, 265-279.

Grant, A., Farris, M., Alefounder, P., Williams, P., Woodward, M., O'Connor, C., 2003. Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic *Escherichia coli* (EPEC). Mol Microbiol 48, 507-521.

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U.T., Pestova, T.V., Frank, J., 2013. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153, 1108-1119.

Neidig, A., Yeung, A.T., Rosay, T., Tettmann, B., Strempel, N., Rueger, M., Lesouhaitier, O., Overhage, J., 2013. TypA is involved in virulence, antimicrobial resistance and biofilm formation in *Pseudomonas aeruginosa*. BMC Microbiol 13, 77.

Qi, S.-Y., Li, Y., Szyroki, A., Giles, I., Moir, A., O&apos;Connor, C.D., 1995. *Salmonella typhimurium* responses to a bactericidal protein from human neutrophils. Mol Microbiol 17, 523-531.

Starosta, A.L., Lassak, J., Jung, K., Wilson, D.N., 2014. The bacterial translation stress response. FEMS microbiology reviews. *In press.*

Vogt, S.L., Green, C., Stevens, K.M., Day, B., Erickson, D.L., Woods, D.E., Storey, D.G., 2011. The stringent response is essential for *Pseudomonas aeruginosa* virulence in the rat lung agar bead and *Drosophila melanogaster* feeding models of infection. Infect. Immun. 79, 4094-4104.

Appendix

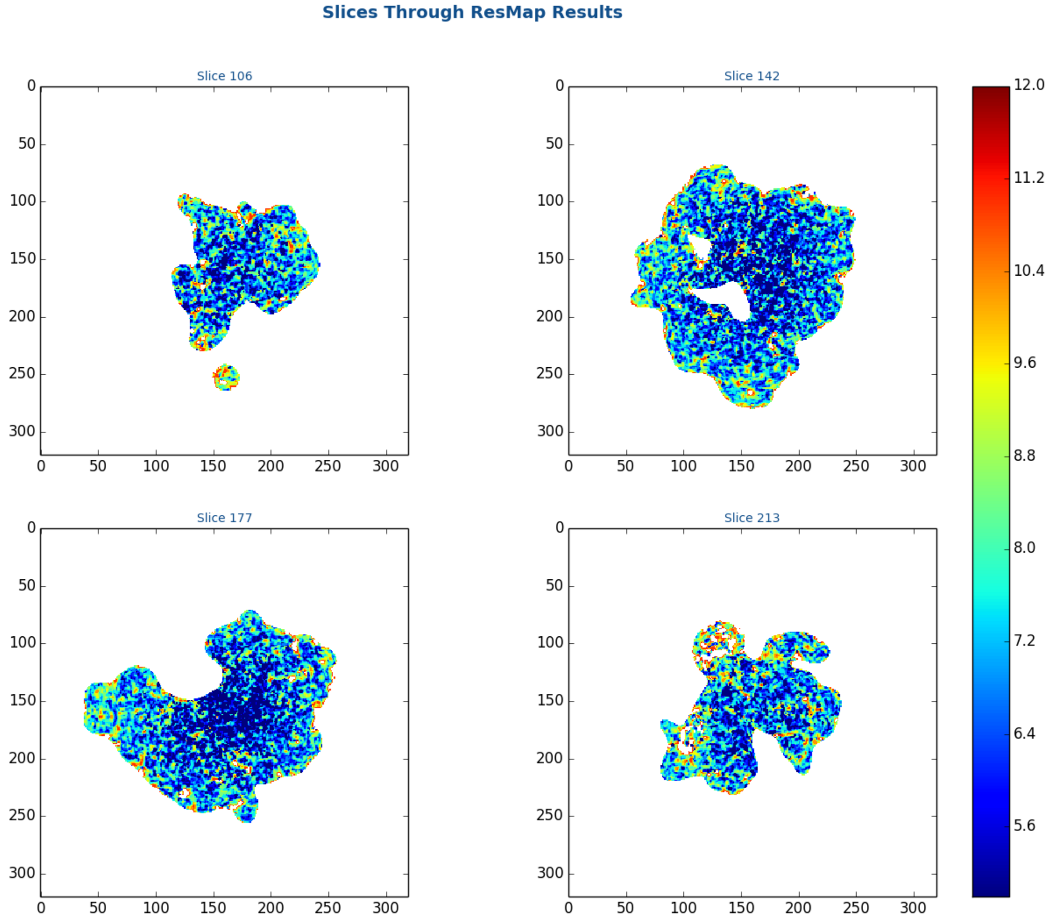


Figure A1. Local resolution assessment of the 70S–BipA reconstruction. The program ResMap (Kucukelbir et al., 2013) was employed to assess the local resolution of the 70S–BipA reconstruction, resolved to an overall resolution of 8.5 Å. The core of the density map, corresponding to the core of the ribosome, exhibits the highest local resolution, generally 7.5 Å or better. The regions on the periphery show lower resolution. This is expected, as periphery regions of the 70S ribosome have more flexibility. One such example is the L1 protein, which has the lowest resolution (10 Å or lower) of any part of the map. The BipA protein, overall, has intermediate resolutions between 7.5 and 9.0 Å.

Journal of Structural Biology 186 (2014) 1–7

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Automated particle picking for low-contrast macromolecules in cryo-electron microscopy

Robert Langlois [a], Jesper Pallesen [a,b], Jordan T. Ash [a,c,1], Danny Nam Ho [d], John L. Rubinstein [e,f], Joachim Frank [a,b,d,*]

[a] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, United States
[b] Howard Hughes Medical Institute, Columbia University, New York, NY 10032, United States
[c] Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, United States
[d] Department of Biological Sciences, Columbia University, New York, NY 10027, United States
[e] The Hospital for Sick Children Research Institute, Toronto M5G 0A4, Canada
[f] Departments of Biochemistry and Medical Biophysics, University of Toronto, Toronto M5S 1A8, Canada

ARTICLE INFO

Article history:
Received 2 January 2014
Received in revised form 11 February 2014
Accepted 2 March 2014
Available online 6 March 2014

Keywords:
Particle selection
Machine-learning
Automation
High-resolution
Cryo-EM

ABSTRACT

Cryo-electron microscopy is an increasingly popular tool for studying the structure and dynamics of biological macromolecules at high resolution. A crucial step in automating single-particle reconstruction of a biological sample is the selection of particle images from a micrograph. We present a novel algorithm for selecting particle images in low-contrast conditions; it proves more effective than the human eye on close-to-focus micrographs, yielding improved or comparable resolution in reconstructions of two macromolecular complexes.

1. Introduction

The term single-particle reconstruction refers to the reconstruction of a macromolecule from multiple projections, each presenting a single, freestanding copy of the macromolecule. These projections are obtained by cryo-electron microscopy (cryo-EM). The plunge-freeze procedure traps the molecules in a thin layer of vitreous ice. A low-dose electron beam captures a low-contrast, two-dimensional projection image (referred to as a micrograph) containing a collection of the molecules trapped in random orientations. The images of the molecules are then subjected to a computational workflow commonly referred to as single-particle analysis, which results in a 3D density map of the macromolecule.

A single high-resolution reconstruction of a 3D macromolecular complex requires the collection of thousands of micrographs, which typically yield hundreds of thousands of particle images. In cases where contrast is extremely low (e.g. with low electron exposures and low defocus settings), a researcher currently spends a substantial amount of time picking particle images from the micrographs. From an image-processing standpoint, the particle-picking problem can be broken down into two steps. First, candidate particle images must be selected from the micrograph; this step historically has been referred to as *particle selection*. Second, the “true” particles (i.e. those representing biological molecules) must be identified among those candidates that may contain falsely discovered non-particles such as contaminants or noise; this step is commonly referred to as *particle verification*. This effort is often compounded by specimen heterogeneity, i.e. multiple conformational states coexisting within the same sample. This problem makes it necessary to collect a larger dataset to ensure there is sufficient relevant data left, after classification, to build a high-resolution map of the structure of interest. Hence, particle picking, especially the second step of particle verification, represents a significant barrier to a completely automated, reproducible single-particle analysis workflow.

Considerable effort has been made to develop algorithms that aid the human eye in selecting good particle images in these extremely low-contrast micrographs (Glaeser, 2004; Langlois et al., 2011; Zhu et al., 2004). A strategy often used is to employ a

[*] Corresponding author at: Howard Hughes Medical Institute, Columbia University, Department of Biochemistry and Molecular Biophysics, 650 West 168th Street, Black Building 2-221, New York, NY 10032, United States. Fax: +1 212 305 9500.
E-mail address: jf2192@columbia.edu (J. Frank).

[1] present address: Department of Computer Science, Princeton University, Princeton, NJ 08540, United States.

http://dx.doi.org/10.1016/j.jsb.2014.03.001
1047-8477/© 2014 Elsevier Inc. All rights reserved.

cross-correlation search over the micrograph in identifying data windows containing candidate particles and then manually verify each window (Rath and Frank, 2004; Roseman, 2003). Another approach to limit the false discovery rate (Langlois and Frank, 2011) is to use hand-tuned thresholds, which can be applied on a micrograph-by-micrograph basis (Chen and Grigorieff, 2007; Tang et al., 2007) or over the entire set (Voss et al., 2009). The elements of subjectivity can be reduced using a machine-learning algorithm referred to as a *classifier*, a supervised learning tool which requires the user to define an initial selection comprising several hundred examples of "good" and "bad" windows (Arbeláez et al., 2011; Langlois et al., 2011; Zhao et al., 2013). Alternatively, candidate particle images identified can be aligned in 2D and then clustered into classes based on intrinsic information; this enables the user to look at the average of each class and either verify or reject the entire class, or further inspect individual particles within that class (Arbeláez et al., 2011; Shaikh et al., 2008). Nevertheless, current methods still require significant effort by the user to verify particles.

We envision a new type of tool that uses unsupervised learning to select particles from the micrograph with minimal user intervention. The user is only required to provide the approximate size of the macromolecule. Unsupervised learning leverages the observation that images of physical objects have limited complexity, and thus, can be described by a compact representation. We seek to further reduce this compact representation by exploiting the fact that the views of the macromolecules are linked by rigid-body transformations: azimuthal rotation and translation.

In the present study, we introduce a two-step automated particle-picking procedure. The first step is a modified template-matching procedure, termed AutoPicker, which identifies a set of candidate particle images from a collection of micrographs and rejects high-contrast contamination and noise using an unsupervised learning procedure. The second step employs an unsupervised one-class classifier, termed View Classifier or ViCer, which exploits the similarity among *aligned* true particles to reject outliers. To assess the quality of the final particle selection, we have applied the algorithm to identify and verify particles from two independent datasets recorded under low-contrast conditions: one of micrographs containing 70S ribosomes from *Escherichia coli* and the second containing molecules of the V/A-ATPase from *Thermus thermophilus*. The density maps obtained using the automatically selected particle images were compared to maps derived from manually selected particle images, which led to high-quality structures. We demonstrate that the particle images selected from of the AutoPicker/ViCer workflow lead to density maps with comparable, if not better, resolved features, and find that this outcome is in part a consequence of AutoPicker/ViCer's ability to identify additional true particles in close-to-focus micrographs.

2. Methods

2.1. Proposed particle-picking algorithm

The proposed automated particle-picking algorithm naturally reduces to two steps: (1) identification as well as an initial verification of potential particles with AutoPicker and (2) further verification using outlier rejection with ViCer.

2.1.1. AutoPicker

The AutoPicker algorithm, as outlined in Supplemental Fig. 1a, uses template matching to identify windows that contain candidate particle images in a micrograph and classification by unsupervised learning to reject both high contrast contaminants and noise windows. Template matching alone provides an excellent ranking of low-contrast, noisy particle (SNR ~0.06) windows over noise, yet provides no means for selecting the optimal threshold to distinguish these two groups. In addition, a micrograph may contain high-contrast contaminants such as ice crystals and bubbles in the ice after radiation damage of the specimen; depending on their size, windows containing contaminants are ranked, according to the cross-correlation score between each window and a template, higher than, or at the same level as, those containing particles. The unsupervised learning algorithm introduced by AutoPicker handles both of these limitations.

First, AutoPicker employs principal component analysis (PCA) over the power spectra of the extracted image windows, reducing each image to a single principal component. Then, assuming a Gaussian distribution, it rejects windows that fall in the tail, i.e. more than 4 standard deviations from the mean. While this cutoff might seem extreme, in practice only the noise windows follow a Gaussian distribution, whereas contaminants tend to follow a more skewed distribution on the tail. This cutoff targets only a specific type contaminant that proves deleterious to the next step. AutoPicker then repeats this procedure over the background surrounding the particle as defined by a ring around the particle; the size of the ring is defined as the particle radius multiplied by the exclusion multiplier and the width is the exclusion distance. Large contaminants and aggregation violate this ring of exclusion, and consequently, become outliers. This step eliminates the most obvious high-contrast contaminants.

Second, AutoPicker applies Otsu's algorithm (Otsu, 1979) on the cross-correlation scores of the remaining windows with the template in order to determine the optimal threshold that separates candidate particles from noise. Note that the order of these two steps is important because high-contrast contaminants tend to skew the cross-correlation histogram, causing Otsu's method to find a suboptimal threshold. In this work, the template was chosen as a disk with a radius corresponding to the particle size and its edges softened by application of a kernel with a Gaussian falloff.

2.1.2. ViCer 2.0

For relatively clean micrographs lacking ice crystals and other artifacts, the AutoPicker algorithm is sufficient to ensure good particle selection. However, many contingencies can contrive to produce less than ideal micrographs and in such cases additional contaminant removal proves necessary. The View Classifier (ViCer) can then be used to further clean the candidate particles of contaminants.

The original ViCer outlier rejection algorithm (Langlois et al., 2012), as outlined in Supplemental Fig. 1b, works by maximizing the similarity between true particles and, as a byproduct, is able to recognize contaminants as outliers. ViCer requires that the particle images have been aligned and grouped into views; it then uses the translation-invariant bispectral transforms of the particle images to further increase the similarity among true particles. Next, PCA is used to represent the bispectral transforms in a two-dimensional feature space. Visual inspection of this space revealed that the true projections tend to form a single cluster, surrounded by outlier contaminants.

The new ViCer algorithm includes two substantial improvements over the original algorithm. First, the PCA is replaced with an outlier-robust version of PCA called DHR-PCA (Feng et al., 2012). This robust PCA prevents corruption of the covariance matrix by contaminants, and as a consequence, yields principal components that better separate contaminants from true particles. Second, the Mahalanobis distance score (a multivariate *z*-score) replaces the ad hoc multivariate extension of the median absolute deviation (MAD) score (Hoaglin et al., 1983) to define the decision boundary between true particles and outlier contaminants. The Mahalanobis distance is defined as follows:

$$D(x) = \sqrt{(x-\mu)^T S^{-1}(x-\mu)}$$

where x is a vector comprising the first two components from PCA, μ is a vector representing the mean for each coordinate in x and S is the covariance matrix estimated from the data in factor space.

In addition, the covariance matrix used to estimate the Mahalanobis distance is estimated using the robust minimum covariance determinant algorithm (Rousseeuw, 1984). The decision boundary cutoff is then determined by a chi-squared distribution (Pearson, 1900), with two degrees of freedom (corresponding to the two principal components) and a probability of 0.97. This probability value is a meaningful parameter, as it defines an outlier as any image with a probability less than 0.03. Note that while this parameter value can be adjusted based on the needs of the experimentalist, the current value performs consistently well in empirical trials.

2.2. Characterization of the datasets

2.2.1. 70S ribosome from E. coli

Micrographs of 70S *E. coli* ribosomes were imaged with an FEI (Portland, OR, USA) Tecnai F30 Polara electron microscope equipped with a field emission gun operating at 300 kV. The data were recorded under low-dose conditions (~20 e$^-$/Å^2) at a temperature of −180 °C and captured on Kodak (Rochester, NY) SO-163 film. The objective aperture in the microscope was 100 µm, and the magnification was set to 59,000. Defocus ranged from 1.5 to 5 µm. A Zeiss-Imaging scanner (Intergraph, Huntsville, Alabama, USA) was used to digitize the micrographs with a step size of 7 µm; for more details see Pallesen et al. (2013).

2.2.2. V/A-ATPase from T. thermophilus

Micrographs of V/A-ATPase from *T. thermophilus* HB8 were obtained in a previously described study (Lau and Rubinstein, 2012). Specimens were imaged with an FEI Tecnai F20 electron microscope equipped with a field emission gun and operated at 200 kV. The data were recorded under low-dose conditions (~18–20 e$^-$/Å^2) at a temperature of −180 °C and captured on Kodak SO-163 film at a magnification of 50,000 with a defocus range of 2.5–4.5 µm. The micrographs were digitized with a Zeiss-Imaging scanner with a 7 µm step size.

2.3. Manually verified benchmarks

This work applies the AutoPicker/ViCer particle-picking software to two independent datasets that had already led to high-quality structures (Lau and Rubinstein, 2012; Pallesen et al., 2013) by employing manual particle verification. The manual picking for each dataset was performed using different protocols that are standard to each lab. Note that the selections obtained by manual picking represent a substantial investment of time and effort by each lab and that it would be arduous to create a more consistent benchmark. Each protocol is subsequently described in detail.

2.3.1. SPIDER LFCPick (70S ribosome)

This work includes two manually verified subsets of particle windows for the ribosome dataset, which are used as benchmarks for the described algorithm. The particle windows for the first benchmark were selected using SPIDER's LFCPick (Adiga et al., 2005) and then manually verified by one of the authors (J.P.). The particle windows for the second benchmark were selected using DoGLFC with AffinityRank (Langlois et al., 2011) and then manually verified by another coworker (Gyanesh Sharma, G.S.).

2.3.2. MRC Ximdisp (ATP synthase)

This work also includes a benchmark derived from a single set of manually verified particle windows from a *T. thermophilus* V/A-ATPase dataset. The particles images were interactively selected with Ximdisp (Crowther et al., 1996).

2.4. Single-particle reconstruction

The orientation parameters are derived from references as described in the previous subsections. Both complexes were subjected to RELION "gold standard" refinement for spatial frequencies higher than 1/(40 Å) and the angular search was oversampled by two orders (Scheres, 2012a). The final density maps were filtered to the target resolution and amplitude enhanced using the program bfactor (http://emlab.rose2.brandeis.edu/grigorieff/download_b.html). CTF correction was performed using the defocus values as estimated by standard SPIDER procedures for the 70S ribosome and CTFFIND (Mindell and Grigorieff, 2003) for the V/A-ATPase.

3. Results and discussion

The primary concern when replacing manual with automated particle picking is whether the procedure employing the automated algorithm can perform comparably to the manual particle picking. Such a procedure will have several significant advantages including speed, consistency and reproducibility. Therefore, we chose two different asymmetric complexes as test cases, ATP synthase (a rod-like protein complex embedded in an amorphous detergent micelle) and the 70S ribosome (a globular protein/RNA complex), in order to determine whether the AutoPicker/ViCer workflow could reliably detect particle images in two widely different cases of sample type and experimental conditions. The structures of both complexes were previously determined using manual particle picking leading to high-quality structures (Lau and Rubinstein, 2012; Pallesen et al., 2013). Thus these two datasets represent interesting, real-life cases where particle picking presents a challenge due to low contrast in the image and imperfect experimental conditions that give rise to contaminants or aggregated particles.

Earlier studies have measured the performance of an automated particle-picking algorithm against a "gold standard" benchmark data set created by an expert manually picking the particles within the micrograph (Langlois and Frank, 2011; Zhu et al., 2004). The often-used dataset in the field (Zhu et al., 2004) is derived from a high-contrast data collection of keyhole limpet hemocyanin (KLH) molecules where the micrographs contain minimal contamination. This is, however, not a dataset typical for current high-resolution studies where micrographs are usually collected at higher voltage and have less contrast. Under these conditions, it is hard to define a "good" particle or "true positive" with a sufficient level of confidence or agreement among experts (Zhao et al., 2013). In addition, this task comes with the superfluous and onerous stipulation that one of the two prominent views should be discarded, requiring the algorithm to perform view classification in addition to particle selection. We conclude that, for low-contrast datasets, benchmarking the automated picking directly against manual picking will not result in a meaningful comparison given both the ambiguity in assigning *true positives* and in the lack of a meaningful benchmark.

The lack of a meaningful comparison is a known issue in the field and efforts are currently underway to remedy this situation. One notable effort is the 3DEM benchmark,[2] which ambitiously

[2] http://i2pc.cnb.csic.es/3dembenchmark/LoadHome.htm

aims to provide a high-quality benchmark for every step in single-particle reconstruction. This project is bound to have a large impact on methods development for the field, but it is currently still a work in progress.

Instead, we compare the performance of automated to manual selection/verification using the end product of the single-particle analysis workflow: the density map. As outlined in Methods, we apply the same workflow to process both datasets using RELION (Scheres, 2012b) for automated angular refinement, which measures the resolution using the Fourier Shell Correlation (FSC) at 0.143 (Rosenthal and Henderson, 2003). During map refinement with RELION, the dataset is divided into two halves and each half is refined independently in order to prevent an overly optimistic resolution estimate from the FSC resulting from overfitted noise

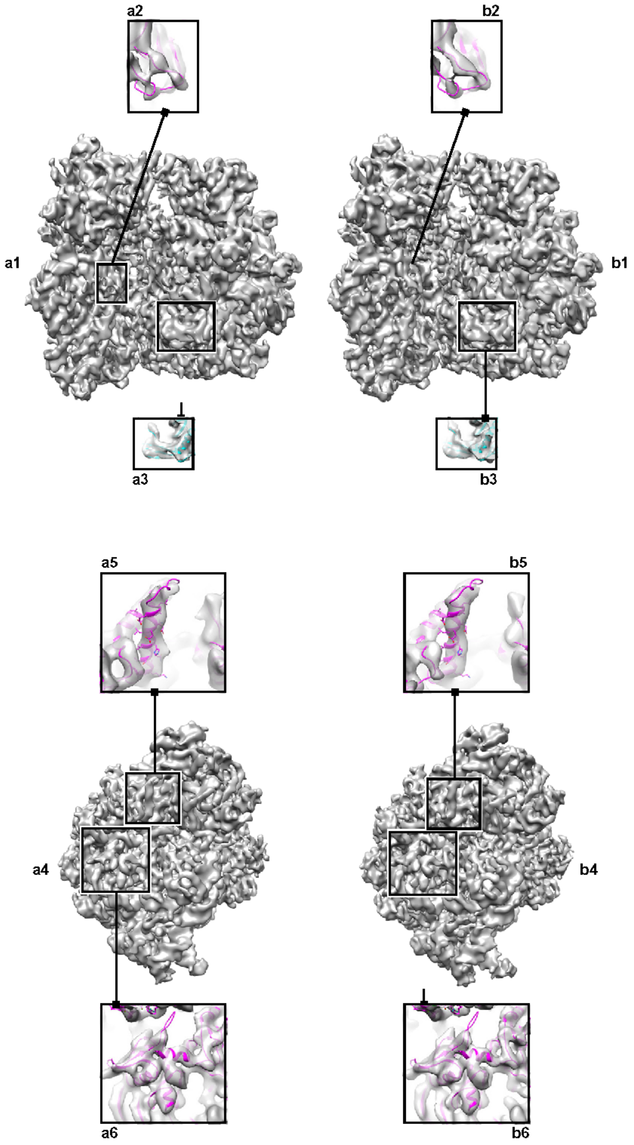


Fig.1. 70S ribosome reconstructed from (a) selections made by AutoPicker/ViCer (b) manual verification performed by J.P. The top panel (a1,b1) shows the ribosome in the classic view with the 30S subunit on the left and the 50S on the right. Within this view, specific features highlight differences in resolvability of a loop (a2,b2) and a β-sheet (a3,b3). The bottom panel (a4,b4) shows an end on view of the 30S subunit. Within this view, specific features highlight differences in resolvability of an α-helix (a5,b5) and the separation of secondary structure elements (a6,b6).

within the data. This approach provides an unbiased means to compare automated to manual selection/verification while preventing the influence of most subjective decisions by the user, which could lead to a misleading comparison of the maps. Subjectivity could be introduced through either a custom mask or tailored parameter settings; we avoid this problem by not including a mask and by using the same parameter settings for both angular refinements.

A concern with the application of the FSC to judge the quality of candidate particles selected either manually or automatically is that the quantity and not the quality of the particles might inflate the resolution estimate. For example, in the case were one processes all the data together and then divides the data into half-sets for resolution estimation. However, the use of "gold-standard" refinement measures the resolution from two independent datasets, avoiding inflation of the resolution due to overfitting of the data.

In the case of the 70S ribosome dataset, angular refinement of the automated AutoPicker/ViCer selection/verification produced a higher-resolution map, at 7.1 Å, compared to two manual verifications of the particle images found by template matching, at 7.4 Å (J.P) and 8.5 Å (G.S.). A visual inspection of the density maps obtained by AutoPicker/ViCer versus the manual verification (as performed by J.P.) reveals that the automated method produces a better-quality map than the better of the two manual verifications (as performed by J.P.). That is, the map derived from the automated method (Fig. 1a) contains numerous features that are better resolved when compared to the map derived from manual verification (Fig. 1b). Consider as an example the loop region (residues 79–81) between strands of a β-sheet in protein S12 of the 30S subunit where the map derived from our automated particle-picking workflow contains sufficient density to accommodate the entire loop (Fig. 1a2), whereas the density mass in the same region of the map derived from manual particle verification is fragmented (Fig. 1b2). The density mass encompassing a β-sheet in the protein L6 of the 50S subunit (residues 88–121) exhibits some separation between the strands in the map from AutoPicker/ViCer selection/verification (Fig. 1a3), whereas the corresponding density in the map from manual verification exhibits no such separation (Fig. 1b3). Further inspection of an α-helix in the protein S8 of the 30S subunit (residues 29–51) reveals density in the AutoPicker/ViCer map corresponding to a large aromatic side chain (Fig. 1a5), which is entirely missing in the manual map (Fig. 1b5). Finally, the density enveloping two neighboring secondary structure elements in protein S2 of the 30S subunit (residues 154–163) resolves the separation between the two elements in the automated map (Fig. 1a6) but not in the manually verified map (Fig. 1b6).

Similarly, a map calculated by RELION for the V/A-ATPase from the automatically picked particle images was at least as good as a map calculated from manually selected particle images. One of the most striking features in the published map of the *T. thermophilus* V/A-ATPase (Lau and Rubinstein, 2012) was the ability to resolve some of the α-helices in soluble and membrane regions of the complex. The membrane region of the complex contains a ring of helical-hairpin subunits known as subunit L, which is equivalent to subunit c in eukaryotic V- and F-type ATPases. The L-ring is comprised of two concentric rings of α-helices, with each subunit contributes one α-helix to the inner ring and one α-helix to the outer ring. The maps produced by RELION from both the manually and automatically selected particle images resolve the α-helices of the outer ring. Furthermore, the two extended peripheral stalk structures in the soluble region of the complex consist of a right-handed coiled coil of elongated α-helices from the E and G subunits. The α-helices of the E and G subunits were also partially resolved in both of the maps calculated here (Fig. 2). Other short α-helices in the structure could also be resolved (Fig. 2, inset).

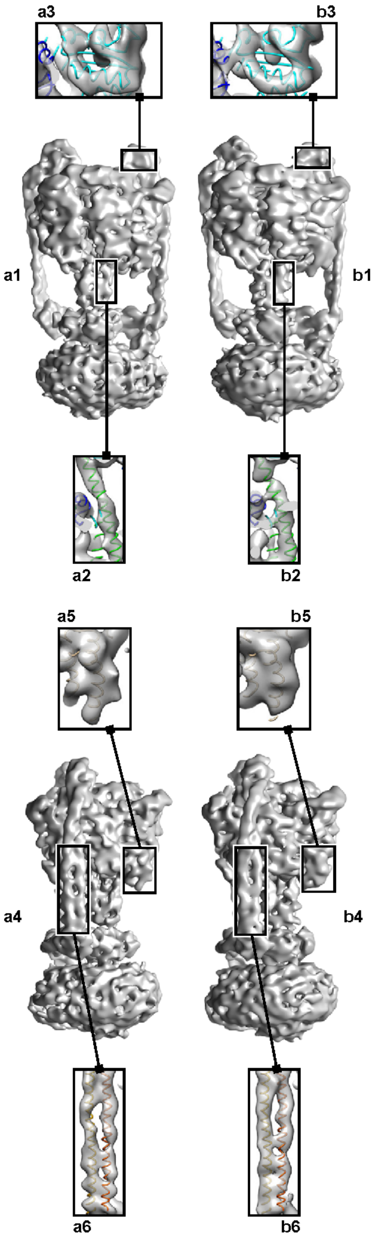


Fig.2. 3D map of the V/A-ATPase from *T. thermophilus*. A map reconstructed from (a) automatically and (b) manually picked particle images. The top panel (a1,b1) shows the V/A-ATPase from the front. Within this view, specific features highlight the resolvability of an α-helix from the central core, residues 1–37 of the D subunit, (a2,b2) versus a peripheral α-helix, residues 195–214 of the A subunit, (a3,b3). The bottom panel (a4,b4) shows the V/A-ATPase from the side. Within this view, specific features that highlight resolvability of a helix-turn-helix motif, residues 538–576 of chain A, (a5,b5) as compared to an α-helix in the stalk, residues 3–57 of the B subunit and residues 25–79 of the E subunit, (a6,b6).

We observed that the number of particles found by the AutoPicker/ViCer workflow was the same irrespective of changes in defocus, whereas for manual picking this number decreased on micrographs captured closer to focus, e.g. fewer particles are typically selected manually from the micrographs shown in Fig. 3a and e compared to those shown in Fig. 3b and f. This trend is quantified in Fig. 3c and g, which plot, for each defocus group (*x*-axis), the fraction of particles (*y*-axis) found by AutoPicker/ViCer but missed by manual selection/verification (V/A-ATPase) or only verification (70S ribosome) with respect to the total number of particles found by AutoPicker/ViCer; this fraction is denoted as disagreement with respect to AutoPicker. From these values we note that the level of disagreement increases as the micrograph is captured closer to focus. As a control, we also plot the level of disagreement with respect to manual verification (Fig. 3d) and manual selection/verification (Fig. 3h), which measures, for each defocus group (*x*-axis), the fraction of particles (*y*-axis) found by manual verification but missed by AutoPicker/ViCer with respect to the total number of particles found by manual verification. In this case, the level of disagreement is virtually the same irrespective of changes to defocus, demonstrating that the fraction of manually verified particles missed by AutoPicker/ViCer does not change with defocus.

This observed trend, where the proficiency of manual verification decreased with decreasing defocus, proved stronger for the 70S ribosome dataset (Fig. 3c) than the V/A-ATPase dataset (Fig. 3g). This difference can be attributed to the difference in contrast of the particles arising from differences in the composition of the molecule and the conditions unique to each data collection. That is, the 70S ribosome dataset was imaged at 300 kV with the defocus ranging from 1.5 to 5 μm whereas the V/A-ATPase dataset at 200 kV, with the defocus ranging from 2.5 to 5 μm.

The automated particle-picking algorithm has already been employed to calculate density maps for several other biomolecular complexes. For example, two publication-quality density maps of the 40S subunit in complex with the protein DHX29 (Hashem et al., 2013b) and the HCV IRES (Hashem et al., 2013a) have been obtained recently in a short period of time, thanks in part to the proficiency of the new automated particle-picking algorithm. This algorithm fills a gap in automated data processing making it possible to perform a fully automated single-particle reconstruction while simultaneously collecting data. In other words, the experimentalist can now view a preliminary structure before the data collection has even finished and, as a consequence, judge whether the quality of the data is sufficient for further image processing.

While the AutoPicker/ViCer workflow will significantly advance high-throughput processing of images captured by cryo-electron microscopy, it has certain inherent restraints. First, it relies on template matching to perform a fast initial search of the micrograph so that when there is a significant amount of contamination or aggregation the current peak exclusion misses good particles, as seen with the V/A-ATPase dataset. This is still an open problem and will be the focus of future work. Second, as a prerequisite for being completely unsupervised in the steps following template matching, this approach makes assumptions about the distribution of data within the micrograph, i.e. that particles representing biomolecules will constitute the largest fraction of objects in the micrograph. This assumption does not present a major problem, however, because in almost all cases, the user screens both grids and micrographs to ensure a minimum level of sample quality.

The rapid advances in the technology underlying data collection in cryo-EM necessitate on-going development of the image-processing workflow. To this end, the AutoPicker/ViCer algorithms have been released as part of Arachnid, a new open-source image-processing package aimed toward images collected by cryo-EM (http://www.arachnid.us).

In sum, we demonstrate that our automated particle-picking algorithm, AutoPicker/ViCer, can accurately identify true particles for very different macromolecular complexes imaged under imperfect experimental conditions: low contrast with significant levels of contamination. The two samples differ in two important ways. First, the samples are composed of different types of macromolecular assembly, i.e. the 70S ribosome is composed of both RNA

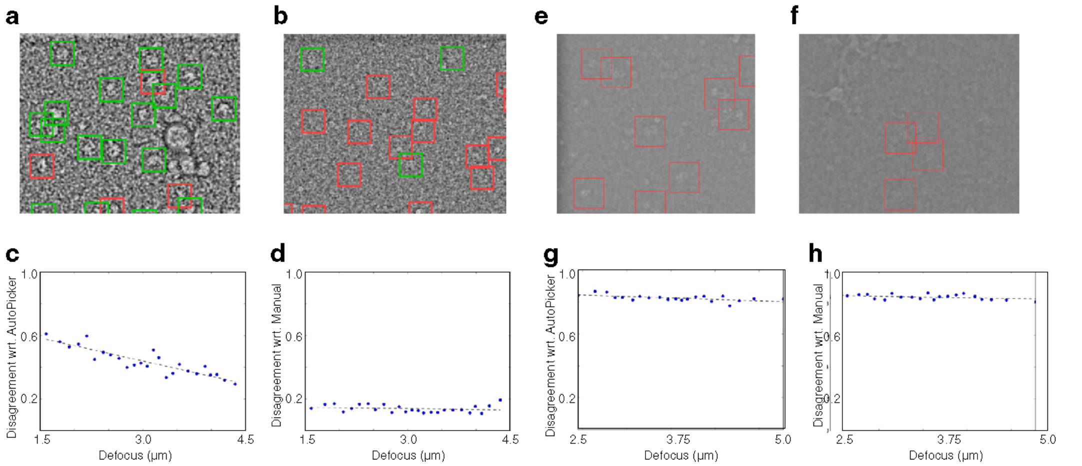


Fig.3. A comparison of manual selection versus AutoPicker/ViCer selection at close and far from focus. Two example micrographs from the 70S ribosome dataset (a) far from focus (4.1 μm) and (b) close to focus (1.8 μm) followed by two example micrographs from the V/A-ATPase dataset (e) far from focus (4.1 μm) and (f) close to focus (2.6 μm). The green windows in panels a and b coincide with those manually verified by J.P. The windows marked in red contain unverified particles. Note that due to the low amount of overlap in the V/A-ATPase dataset, no green windows appear in right most micrographs (e,f). All micrographs have been preprocessed in the same manner: Gaussian band-pass filter, outlier pixel removal and down-sampled by a factor of four. Plots comparing AutoPicker/ViCer to manual selection over the 70S dataset (c,d) and ATP synthase dataset (g, h) showing disagreement with respect to (wrt) AutoPicker (c,g) and wrt manual verification (d,h) on the *y*-axis and defocus on the *x*-axis; the data was partitioned into 25 defocus groups (blue dots). The black dashed line is a linear trend line fit to the data.

and protein while the V/A-ATPase is composed of protein with a detergent micelle. Second, the samples present very different shapes, with the 70S ribosome being globular and the V/A-ATPase rod-like in shape. Despite these differences, the selections made by the AutoPicker/ViCer workflow yielded high-quality density maps without manual intervention, saving substantial costs in time and labor and preventing the results from being influenced by subjective decisions.

Acknowledgments

This work was supported by the Howard Hughes Medical Institute and National Institutes of Health Grants R37 GM 29169 and R01 GM 55440 (to J. Frank). J.T.A. acknowledges support by an AMGEN scholarship. We would like the thank Melissa Thomas for assistance with the preparation of the illustrations and Gyanesh Sharma for performing a second manual verification. We would also like to thank early adopters of AutoPicker in our group, namely Bo Chen, Bingxin Shen, Ming Sun, Amedee des Georges, Yaser Hashem and Sanchaita Das, for their feedback.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2014.03.001.

References

Adiga, U., Baxter, W.T., Hall, R.J., Rockel, B., Rath, B.K., et al., 2005. Particle picking by segmentation: a comparative study with SPIDER-based manual particle picking. J. Struct. Biol. 152, 211–220.

Arbeláez, P., Han, B.-G., Typke, D., Lim, J., Glaeser, R.M., et al., 2011. Experimental evaluation of support vector machine-based and correlation-based approaches to automatic particle selection. J. Struct. Biol. 175, 319–328.

Chen, J.Z., Grigorieff, N., 2007. SIGNATURE: a single-particle selection system for molecular electron microscopy. J. Struct. Biol. 157, 168–173.

Crowther, R.A., Henderson, R., Smith, J.M., 1996. MRC image processing programs. J. Struct. Biol. 116, 9–16.

Feng, J., Xu, H., Yan, S., 2012. Robust PCA in high-dimension: a deterministic approach. In: International Conference on Machine Learning, Edinburgh, Scotland.

Glaeser, R.M., 2004. Historical background: why is it important to improve automated particle selection methods? J. Struct. Biol. 145, 15–18.

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., et al., 2013a. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. Nature 503, 539–543.

Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., et al., 2013b. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153, 1108–1119.

Hoaglin, D.C., Mosteller, F., Tukey, J.W., 1983. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons.

Langlois, R., Frank, J., 2011. A clarification of the terms used in comparing semi-automated particle selection algorithms in cryo-EM. Struct. Biol. 175, 348–352.

Langlois, R., Pallesen, J., Frank, J., 2011. Reference-free particle selection enhanced with semi-supervised machine learning for cryo-electron microscopy. J. Struct. Biol. 175, 353–361.

Langlois, R., Ash, J.T., Pallesen, J., Frank, J., 2012. Fully automated particle selection and verification in single-particle cryo-EM. In: Frank, J., Herman, G., (Eds.), Minisymposium on Computational Methods in Three-Dimensional Microscopy Reconstruction, New York, NY.

Lau, W.C.Y., Rubinstein, J.L., 2012. Subnanometre-resolution structure of the intact *Thermus thermophilus* H^+-driven ATP synthase. Nature 481, 214–218.

Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.

Pallesen, J., Hashem, Y., Korkmaz, G.r., Koripella, R., Huang, C., 2013. Cryo-EM visualization of the ribosome in termination complex with apo-RF3 and RF1. eLife 2.

Pearson, K., 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. In: Philosophical Magazine Series 5, vol. 50, pp. 157–175.

Rath, B.K., Frank, J., 2004. Fast automatic particle picking from cryo-electron micrographs using a locally normalized cross-correlation function: a case study. J. Struct. Biol. 145, 84–90.

Roseman, A.M., 2003. Particle finding in electron micrographs using a fast local correlation algorithm. Ultramicroscopy 94, 225–236.

Rosenthal, P.B., Henderson, R., 2003. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. 333, 721–745.

Rousseeuw, P.J., 1984. Least median of squares regression. J. Am. Stat. Assoc. 79, 871–880.

Scheres, S.H.W., 2012a. A Bayesian view on cryo-EM structure determination. J. Mol. Biol. 415, 406–418.

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

Shaikh, T.R., Trujillo, R., LeBarron, J.S., Baxter, W.T., Frank, J., 2008. Particle-verification for single-particle, reference-based reconstruction using multivariate data analysis and classification. J. Struct. Biol. 164, 41–48.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., et al., 2007. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46.

Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., Carragher, B., 2009. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. J. Struct. Biol. 166, 205–213.

Zhao, J., Brubaker, M.A., Rubinstein, J.L., 2013. TMaCS: a hybrid template matching and classification system for partially-automated particle selection. J. Struct. Biol. 181, 234–242.

Zhu, Y., Carragher, B., Glaeser, R.M., Fellmann, D., Bajaj, C., et al., 2004. Automatic particle selection: results of a comparative study. J. Struct. Biol. 145, 3–14.

Arachnid: A Scientific Python Toolkit for Electron Microscopy

Robert E. Langlois, Danny N. Ho, Ming Sun, Amedee des Georges, others, and Joachim Frank

Abstract

Arachnid is a new scientific, open-source computing platform to process images captured by electron microscopy. This platform provides a rich set of tools for rapid code development while, at the same time, providing a simple and easy-to-use interface for users. Arachnid focuses on automated preprocessing of the images for later orientation determination and classification.

Introduction

Single-particle reconstruction (SPR) of macromolecules imaged by cryo-electron microscopy is a rapidly evolving field with many challenging computational problems. In this approach, an electron microscope captures a 2D projection image containing multiple copies of a single, freestanding macromolecule frozen in a random orientation within a thin layer of vitreous ice. A computational workflow called single-particle reconstruction, then, is used to isolate the individual particle images, orient each image in 3D space, sort the images into homogenous classes related to conformation states and reconstructs a 3D density map from the individual molecular images for each state (Figure 1).

Numerous open source packages have been developed to tackle the SPR workflow. One of the original software packages, SPIDER (Frank et al., 1981; Shaikh et al., 2008), provided a set of modular commands along with a custom scripting language providing the user with the ability to design custom batch procedures or workflows. The SPIDER package was released open source in 2006. Another popular approach was to provide a set of command line tools where the user used a shell scripting language to build a workflow (Crowther et al., 1996; Elmlund and Elmlund, 2012; Grigorieff, 2007; Heymann and Belnap, 2007; Scheres, 2012; Sorzano et al., 2004). More recent packages have adopted scientific scripting languages such as Matlab (Aspire, http://spr.math.princeton.edu/) and Python (de la Rosa-Trevin et al., 2013; Hohn et al., 2007; Lander et al., 2009; Ludtke et al., 1999; Tang et al., 2007).

Scientific scripting languages offer many advantages over traditional languages such as C/C++ and Fortran, such as providing an environment that facilitates faster code development, easier maintenance and less debugging while sacrificing little efficiency in terms of speed and memory. Scientific scripting draws from extensive, well-documented libraries that further facilitate code development rather than requiring the developer to "reinvent the wheel".

We introduce a new open source software platform that aims to simplify both the use and development of single-particle reconstruction procedures. This platform utilizes the

Python language as a high-level framework that glues together optimized C/C++ and Fortran routines.

Scientific Python

Python is an established, all-purpose scripting language whose design philosophy puts stress on the readability of code. Unlike C or Fortran, Python allows the programmer to express concepts in fewer lines of code and promotes readability by enforcing indentation rather than brackets to demarcate programing constructions such as if-statements or for-loops.

The scientific scripting paradigm best exemplified by Matlab and Python helps the programmer avoid many pitfalls in code development such as premature optimization, unreadable and overly complicated code. The programmer can focus on development of the algorithm while utilizing vectorized libraries to ensure computing efficiency for these procedures. This encourages fast prototyping of new concepts, allowing the programmer to refine an idea before optimization.

The foundation of scientific Python is formed by the NumPy (Dubois et al., 1996) and SciPy (Jones et al., 2001) packages. These packages are further expanded by a growing number of scientific libraries such as IPython (Perez and Granger, 2007), Matplotlib (Hunter, 2007) and Spyder (https://code.google.com/p/spyderlib/). Combined, these libraries form an environment that, in many ways, exceeds its commercial counterparts like Matlab, which lacks certain features in plotting and interaction with the interpreter.

SciPy also provides a mechanism for developers to provide specialized extensions called SciKits (SciPy Toolkits). For example, scikit-learn (Pedregosa et al., 2011) focuses on building simple, efficient tools for machine learning in Python. Likewise, scikit-image (van der Walt et al., 2011) encompasses tools for image processing in Python. These extensions present a mechanism to customize scientific computing in Python for specific areas of science.

Arachnid Design

The design of the Arachnid platform follows that of the aforementioned SciKits, extending Scientific Python to image processing for electron microscopy. Arachnid can be broken down into three main components:

1. Command-line applications: All applications are written in Python to ensure a consistency of interface;
2. Graphical User Interface: Simplified user interface to command line programs;
3. Core library: While primarily written in Python, this library also contains optimized procedures written in C/C++ and Fortran.

Command-line applications

The primary interface to a specific Arachnid application is through the command line. A user can change options either using the command line arguments or using a configuration file. The application will then run on the terminal and provide the user with feedback regarding the state of the job. Note that the configuration file and command line arguments can be used in conjunction with the understand that arguments specified on the command line override those in the configuration file.

The configuration file is a powerful tool for interacting with an application as well as providing the user with substantial documentation describing the features of the application. As shown in Figure 2, the configuration file starts with a header giving a short description of the application along with an example of how to run the application. To ensure reproducibility, the header also includes an exhaustive version of the application, indicating the specific *commit* (a logged message referring to a specific change in the source code) in the source repository. The configuration file then provides a list of every supported option, organized into meaningful groups. Each option has associated with it, a short comment documenting its use. The configuration file also provides a link to the full online-documentation for the application listed in the header.

The configuration file can serve as a script which, by default, invokes the appropriate application using the current configuration file as input. This simplifies running an application while, at the same time, providing the user with the ability to combine scripting and options for an Arachnid application in the same file. In addition, all scripts have default settings that have been carefully tuned over many experimental trials with datasets from different samples including ribosomes, ion channels, and ATPases.

When an Arachnid application is run, it gives feedback utilizing a user-friendly logging system. By default, only notes on the progress of the application, warnings or errors are printed to the terminal. However, the user can customize the level of verbosity. Also, errors and warnings are highlighted in red and blue, respectively, to enhance their visibility (for terminal windows supporting color highlighting). When an error occurs in an Arachnid application, a concise error message is reported to the user along with a pointer to the location of an additional crash report, which contains the full exception trace that is, generally, only useful to the developer in debugging.

Arachnid also supports automated restart. In the event of a crash, a restarted application will continue from where it left off. Changes to key parameters, all input files, partial output files or specifying the `--force` parameter will cause an application to restart from the beginning. Changes to a single input file will cause an application to reprocess only that specific input file.

Most of the scripts in Arachnid process individual files, independently. For example, AutoPicker (Langlois et al., 2014b) or frame-alignment work independently on micrographs. For this type of script, Arachnid supports two levels of parallelism: (1) cluster computing, by specifying the `--use_MPI` option and (2) multi-processing, by specifying the `--worker-count` option. These two levels can even be used concurrently for great efficiency.

The command line scripts in Arachnid can be grouped into three main categories: (1) Native Applications, (2) Native Utilities and (3) pySPIDER Applications. The application scripts encompass non-trivial steps in the single-particle reconstruction workflow comprising frame alignment for direct electron detector movies supporting both L2-optimization (Li et al., 2013) and sequential - *ara-alignmovie*, automated particle picking - *ara-autopick* (Langlois et al., 2014b) and *ara-vicer* (Langlois et al., 2014a; Langlois et al., 2014c), defocus estimation - *ara-fastctf* and reconstruction - *ara-reconstruct*.

The utility scripts encompass utility preprocessing, optional post-processing, debugging and organizational steps in the single-particle reconstruction workflow. Such post-processing scripts include particle orientation histograms - *ara-coverage*, particle selection analysis - *ara-bench* (Langlois and Frank, 2011) and volume processing - *ara-prepvol*. Debugging scripts include image file information - *ara-info* and internal Arachnid testing - *ara-sanitycheck*. Finally, the organizational scripts include project workflow creation (Figure 3) - *ara-project*, file enumeration - *ara-enumfiles*, isolating windows from micrographs - *ara-crop*, Relion (Scheres, 2012) selection file generation - *ara-selrelion* and preprocessing micrographs for screening - *ara-screenmics*.

The pySPIDER scripts utilize a Python wrapper for SPIDER commands. They can be used to create customized scripts with SPIDER while replacing the SPIDER scripting language with Python. In Arachnid, the pySPIDER scripts fill gaps in Arachnid application functionality. Those gaps currently include angular refinement - *sp-autorefine*, resolution estimation - *sp-resolution* and amplitude enhancement - *sp-enhancevol*.

Note that pySPIDER scripts are differentiated from Arachnid scripts with the *sp-* prefix whereas Arachnid scripts use *ara-*. These prefixes not only differentiate Arachnid and pySPIDER scripts, but uniquely identify these scripts as belonging to the Arachnid package.

Graphical User Interface

A graphical user interface (GUI) allows the user to interact with a program through visual widgets that help guide the user's decisions. Arachnid provides two basic types of GUIs, those that either (1) set options in the scripts or (2) interact with the data. For the first type of GUI, *option setters*, Arachnid includes two subtypes. The first subtype of GUI is called the AutoGUI, which automatically generates a graphical user interface from the available options for that script. An AutoGUI can be displayed for any script by specifying the '-X' option on the command line when invoking the script. The second subtype of GUI is the *project manager interface*, which opens a wizard that takes the user step-by-step through setting up a workflow. This interface can, optionally, connect to an existing Leginon (Carragher et al., 2000) database and extract the location of the image files as well as the important parameters that describe the data collection such as microscope acceleration voltage, magnification, spherical aberration and pixel size.

Arachnid also provides several GUIs that allow the user to interact with the data. This includes image visualization with *ara-display*, micrograph and power spectra screening with *ara-screen* and particle and power spectra distribution analysis with *ara-plot* (Figure 4).

The ara-screen GUI supports many features for micrograph and power spectra screening. First, it allows the user to switch between micrographs and power spectra so the user has to only go through the data once. Second, for the power spectra, it provides the ability to view rings indicating resolution and/or a 1D representation of the power spectra. Third, for the micrographs, it allows the user to views the particle selection in order to verify that no parameters were incorrectly set. Finally, the ara-screen GUI is integrated with AutoPicker providing simple controls so the user can test various settings in order to optimize particle selection for a new molecule and see the result immediately.

Another Arachnid GUI program, called *ara-plot*, allows the user to analyze data in a scatter plot. For instance, ara-plot can be used to visualize the decision boundary used by *ara-vicer* to reject contaminants by displaying a scatter plot of each image in 2D Eigenspace and clicking on any point (representing that image) displays the corresponding particle image. Similarly, *ara-plot* can be used to analyze the contrast transfer function (CTF) parameters determined by *ara-fastctf*. That is, the defocus, astigmatism or error can be displayed on a scatter plot, and by clicking on the individual points representing a single micrograph, displays the corresponding power spectra and model.

Core Library

The core library of Arachnid is intended to facilitate the development of new applications and utilities. The following list summarizes the organization of the core library into 9 packages, which encompass:

1. app - Application framework that provides a unified interface to every Arachnid script. It also provides many of the features previously discussed in the "Command-line Scripts" section.
2. gui - Graphical User Interface (GUI) framework that uses PySide, a Python-binding to the Qt4 libraries, to provide the GUI elements previously described in the "Graphical User Interface" section.
3. image - Image processing framework, which includes routines optimized in both Fortran and C/C++.
4. learn - Machine learning framework, which also includes routines optimized in C/C++ and bindings to Blas (Dongarra et al., 1990) and Lapack (Anderson et al., 1990) optimized libraries.
5. metadata - Metadata reading/writing framework. This provides support reading and writing in SPIDER, Relion, CSV and FREALIGN document formats. It also provides the interface to the Leginon database using SQLAlchemy.
6. orient - Orientation framework that includes conversions between Euler angle conventions, quaternions and rotation matrices. It also provides support for the HEALPix library (Gorski et al., 2004), which can be used to tessellate a sphere in 3D.
7. parallel - Parallel programming framework, which includes Python multi-processing routines, message passing interface (MPI) routines and routines to control the

number of threads for OpenMP and MKL OpenMP threading in the C/C++, Fortran optimized code as well as the blas/lapack optimized libraries.

8. spider - pySPIDER framework that defines a Python interface to many SPIDER commands.
9. util - Additional modules to standardized plotting and image drawing.

Each package, module, class, method and function is documented in the Arachnid package. Both user and developer information is documented at the module level and the user manual is built by stripping out only the user documentation from the source code. This ensures that the developer has easy access to the full documentation of a module. Arachnid uses the Sphinx reStructuredText processer to extract documentation and build the Arachnid website dynamically from the source code (http://www.arachnid.us). As shown in Figure 5, this website includes a user manual, reconstruction protocols, developer's manual and the application programmer's interface (API). The website also links to a blog where the latest information on Arachnid can be found.

Arachnid Distribution

Source Code

Arachnid uses the Git (http://git-scm.com/) version control system to maintain its source code and distributes the code from GitHub (http://www.github.com). The basic function of a version control system, e.g. Git, is that it provides tracking for changes in the source code. However, the real power is in that it provides a framework under which a community of developers can work on a single project.

Using GitHub, developers can easily create a *fork* (or copy at the current time of forking) of the Arachnid source code. Then, they can *clone* a local working copy of the code in which they can add their contribution. When the code is ready for release, they can immediately *push* it to their GitHub fork of Arachnid, and, if they choose, distribute it from there. Finally, they can contribute their addition to the original Arachnid GitHub repository by making a *pull* request. The code is then evaluated by an overseer (a person who safeguards Arachnid) and *merge*d into the main repository when it meets certain standards. These standards include:

1. Does the code work?
2. Has the developer provided sufficient documentation?
3. Is the code readable?

Installation via Binary Packages

Arachnid uses the Anaconda (Continuum Analytics, Austin, TX USA) package to simplify installation. Anaconda is a system to simplify packaging for Python and includes all the standard Scientific Python packages. Indeed, it provides binaries for all the Arachnid prerequisites, properly compiled to run efficiently on most machines. Binary Arachnid packages are distributed on binstar.org (Continuum Analytics, Austin, TX USA), a repository for publicly and privately available Anaconda packages.

The Arachnid Binstar repository provides 4 types of binary packages. These include a stable build and a daily build and an accelerated version of the stable and daily build. The accelerated version requires a premium package available for a modest price from continumm.io; note, however, that these premium packages are free for academic use. The accelerated version provides optimized NumPy, SciPy and Arachnid routines using Intel's Math Kernel Library (MKL).

Arachnid includes conda recipes, which are scripts that allow a developer to easily build a binary version of Arachnid for the Binstar repository. These recipes define not only how to build Arachnid binaries, but also define all the dependencies for Arachnid, namely Python source, executable and shared library dependencies. Thus, installing the Arachnid package in Anaconda also installs all its dependencies in one easy step.

Forums

Keeping an open line of communication between developers and users is crucial to a healthy open source community. Arachnid provides several mechanisms to facilitate that communication. These tools include a mailing list where users can ask general questions about Arachnid, which other users or developers can answer in a timely fashion. They also include an *Issues Manager* where users can request new features or report bugs. Another important tool is a blog where quick announcements can be made and simple examples for new programs can be outlined.

Conclusion

The Arachnid platform has been used to solve several structures including 70s ribosomes and ATPases (Langlois et al., 2014b) . It has also been used to solve the structure of two 40S ribosomal subunits in complex with DHX29 (Hashem et al., 2013b)and the Hepatitis-C-virus entry sites (Hashem et al., 2013a).

The aim of Arachnid is to provide both the user and developer a solid scientific computing platform for analysis of images captured by electron microscopy. We regard it, starting from its general release, as a communal project lives not at any one institution, but in the "Cloud" inviting everyone to contribute, either directly to the Arachnid platform, or by forking the Arachnid source code for his or her own project. Arachnid's current focus is on automating the preparation of data for classification and orientation assignment. However, in future versions, we plan to add new approaches to classification and orientation assignment. The Arachnid platform and its documentation can be found at http://www.arachnid.us.

Acknowledgements

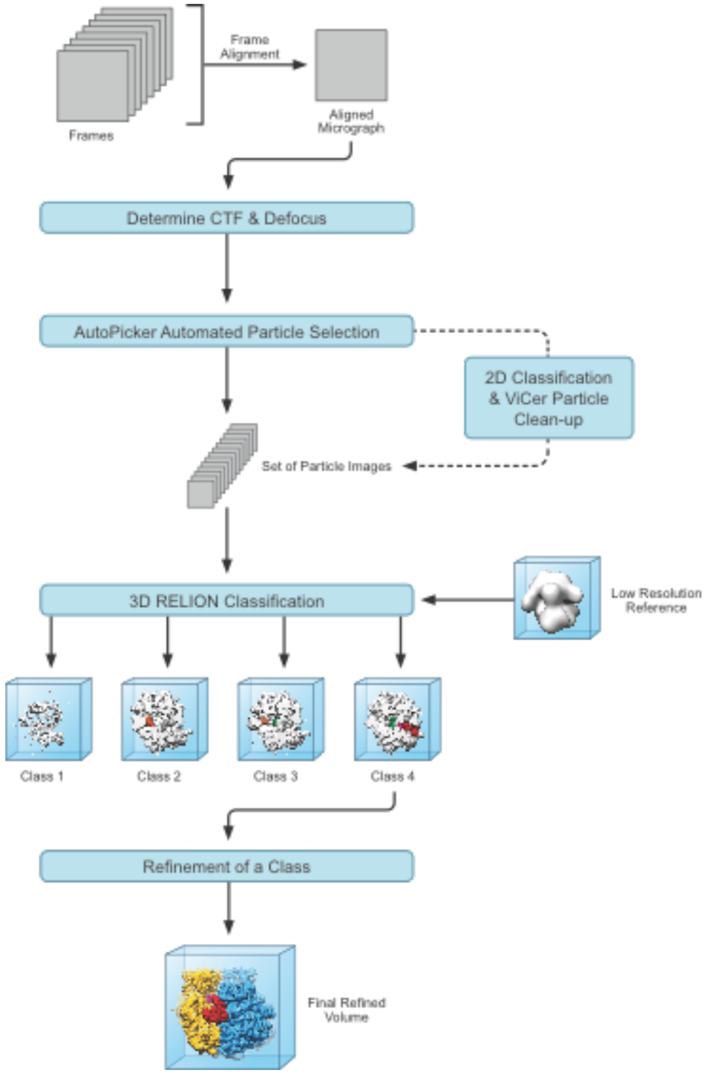
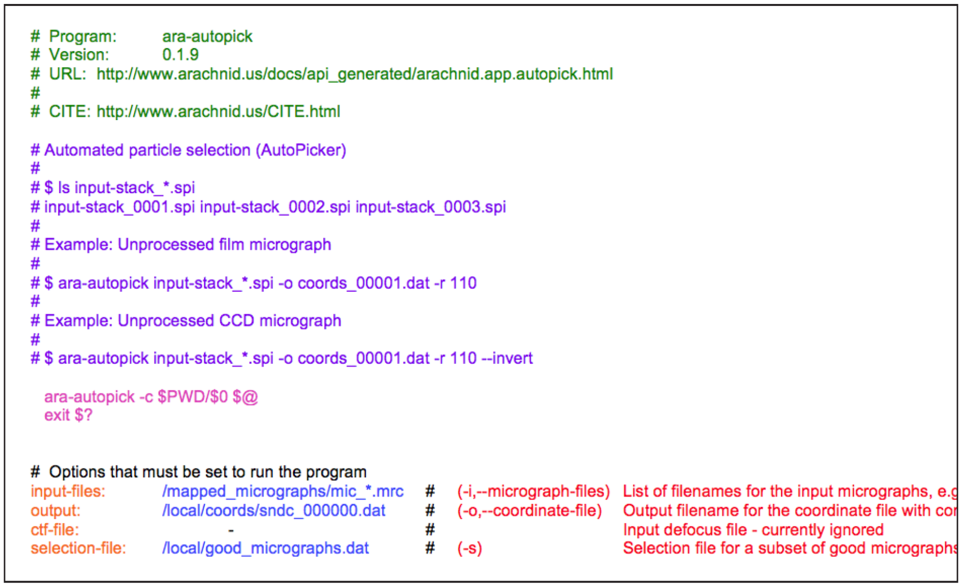


Figure 1 – Single particle reconstruction workflow. This workflow incorporates several recent advancements in the field of cryo-EM. When images are collected in movie-mode on a direct electron camera, dose-fractionated frames will be aligned in order to correct drift. The contrast transfer function (CTF) is then characterized for each averaged image or micrograph. Next, AutoPicker will select potential particle windows from each image. These windows can then be subjected to 2D-classification at which point further cleanup can be performed with ViCer; the dotted lines indicate that this is an optional step. Finally, RELION (Scheres, 2012) can be used for 3D classification of the heterogeneous dataset into homogenous subsets of particles. The class of particles that compose a density map that visually shows the complex of interest can then be subsequently be refined to give the final reconstruction.



```
# Program:        ara-autopick
# Version:        0.1.9
# URL:  http://www.arachnid.us/docs/api_generated/arachnid.app.autopick.html
#
# CITE: http://www.arachnid.us/CITE.html

# Automated particle selection (AutoPicker)
#
# $ ls input-stack_*.spi
# input-stack_0001.spi input-stack_0002.spi input-stack_0003.spi
#
# Example: Unprocessed film micrograph
#
# $ ara-autopick input-stack_*.spi -o coords_00001.dat -r 110
#
# Example: Unprocessed CCD micrograph
#
# $ ara-autopick input-stack_*.spi -o coords_00001.dat -r 110 --invert

  ara-autopick -c $PWD/$0 $@
  exit $?


# Options that must be set to run the program
input-files:      /mapped_micrographs/mic_*.mrc    #   (-i,--micrograph-files)  List of filenames for the input micrographs, e.g
output:           /local/coords/sndc_000000.dat    #   (-o,--coordinate-file)   Output filename for the coordinate file with cor
ctf-file:                   -                      #                            Input defocus file - currently ignored
selection-file:   /local/good_micrographs.dat      #   (-s)                     Selection file for a subset of good micrographs
```

Figure 2 - A typical configuration file supported by Arachnid. The configuration file is rendered in color to highlight specific features encompassing (1) header, (green), which includes the application name, version and documentation URL, (2) a short description showing examples of how to run the application (purple), (3) brief shell script to launch the application using the current configuration file (magenta), (4) a short description of a group of options (black), (5) option name (orange) followed by its value (blue) and then by a comment describing the option (red).

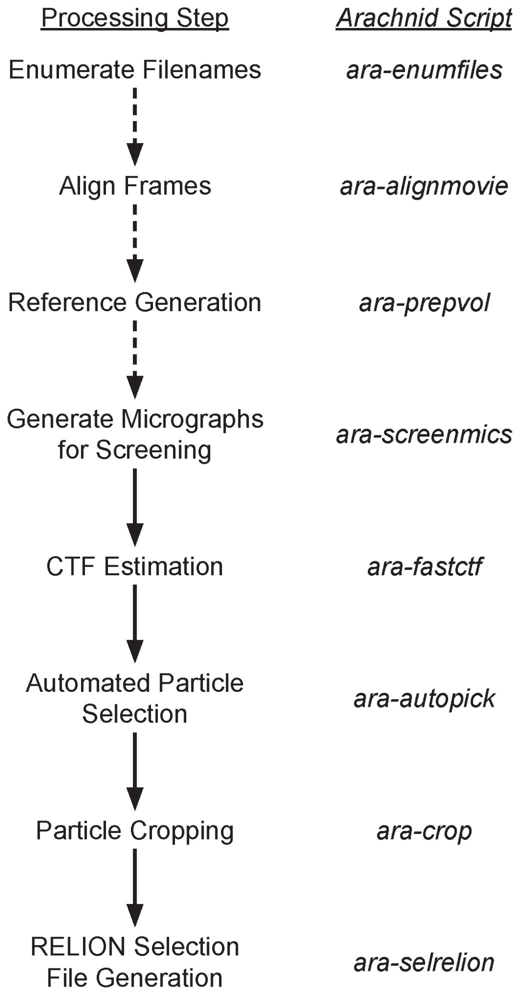


Figure 3 - Arachnid workflow and the corresponding command-line applications. Optional steps, depending on the instrumentation available, naming convention for input images and the needs of the experimentalist, are shown as dashed lines.

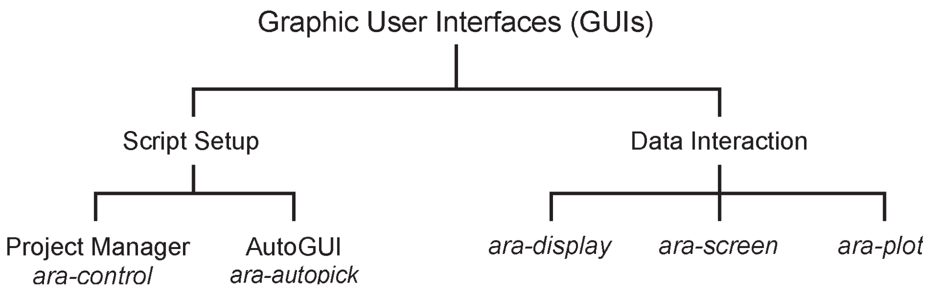
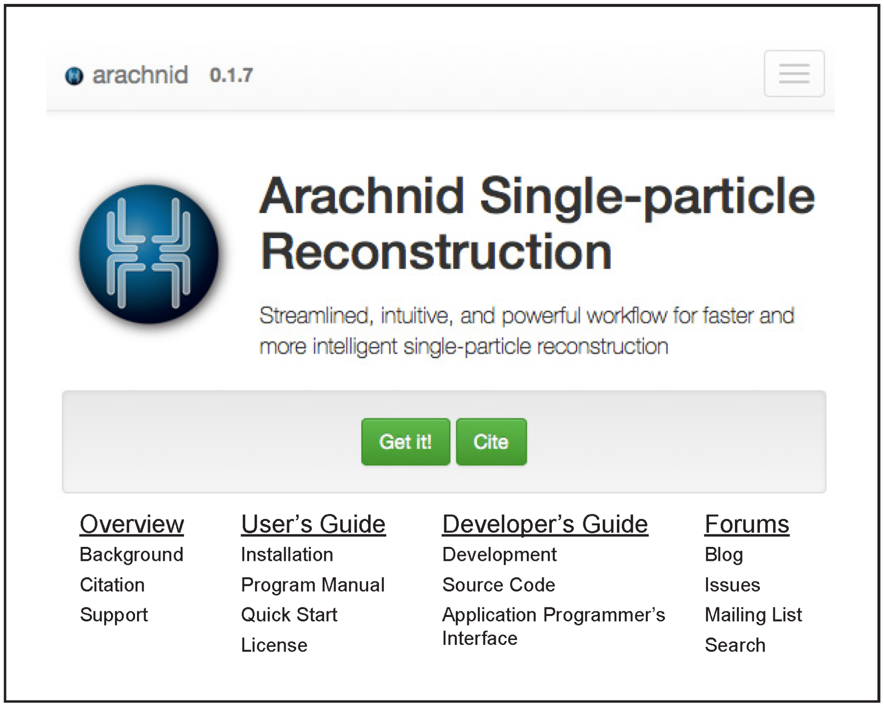


Figure 4 - Organization of the various Graphic User Interfaces (GUIs). The five available GUIs are divided, based upon their purpose of use, into two principal categories: those that set options in applications, *e.g.* Workflow Manager (ara-control) and AutoGUI (available for every non-GUI, *i.e.* command-line, script) and those that interact with the data for screening (i.e. ara-screen) or display (i.e. ara-display and ara-plot) purposes.



Arachnid Single-particle Reconstruction

Streamlined, intuitive, and powerful workflow for faster and more intelligent single-particle reconstruction

Overview
- Background
- Citation
- Support

User's Guide
- Installation
- Program Manual
- Quick Start
- License

Developer's Guide
- Development
- Source Code
- Application Programmer's Interface

Forums
- Blog
- Issues
- Mailing List
- Search

Figure 5 - Organization of website. Arachnid is made available online at www.arachnid.us. At the top of the website, the latest stable version of the platform is shown. A brief description of the platform is provided underneath the title. Downloads and citations are immediately available via green buttons prominently placed below the title and description. For expediency, the navigation on the website is not shown, instead, a brief summary outlining the site organization is given below.

References

Anderson, E., Z. Bai, J. Dongarra, A. Greenbaum, A. McKenney*, et al.*, 1990. LAPACK: A portable linear algebra library for high-performance computers, Supercomputing '90. Proceedings of, pp. 2-11.

Beazley, D., 2003. Automated scientific software scripting with SWIG. Future Generation Computer Systems 19, 599-609.

Behnel, S., R. Bradshaw, C. Citro, L. Dalcin, D.S. Seljebotn*, et al.*, 2011. Cython: The Best of Both Worlds. Computing in Science &amp; Engineering 13, 31-39.

Carragher, B., N. Kisseberth, D. Kriegman, R.A. Milligan, C.S. Potter*, et al.*, 2000. Leginon: An Automated System for Acquisition of Images from Vitreous Ice Specimens. Journal of Structural Biology 132, 33-45.

Chekanov, S., 2010. Scientific Data Analysis using Jython Scripting and Java. Springer London.

Crowther, R.A., R. Henderson, J.M. Smith, 1996. MRC Image Processing Programs. Journal of Structural Biology 116, 9-16.

Dalcin, L., R. Paz, M. Storti, J. Delia, 2008. MPI for Python: Performance improvements and MPI-2 extensions. Journal of Parallel and Distributed Computing 68, 655-662.

de la Rosa-Trevin, J.M., J. Oton, R. Marabini, A. Zaldivar, J. Vargas*, et al.*, 2013. Xmipp 3.0: An improved software suite for image processing in electron microscopy. Journal of Structural Biology 184, 321-328.

Dongarra, J.J., J. Du Croz, S. Hammarling, I.S. Duff, 1990. A set of level 3 basic linear algebra subprograms. ACM Trans. Math. Softw. 16, 1-17.

Dubois, P.F., K. Hinsen, J. Hugunin, 1996. Numerical Python. Computers in Physics 10.

Elmlund, D., H. Elmlund, 2012. SIMPLE: Software for ab initio reconstruction of heterogeneous single-particles. Journal of Structural Biology 180, 420-427.

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

Gorski, K.M., E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen*, et al.*, 2004. HEALPix &#45;&#45; a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere. The Astrophysical Journal 622, 759-771.

Grigorieff, N., 2007. FREALIGN: High-resolution refinement of single particle structures. Journal of Structural Biology 157, 117-125.

Hashem, Y., A. des Georges, V. Dhote, R. Langlois, H.Y. Liao*, et al.*, 2013a. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. Nature 503, 539-543.

Hashem, Y., A. des Georges, V. Dhote, R. Langlois, H.Y. Liao*, et al.*, 2013b. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. Cell 153, 1108-19.

Heymann, J.B., D.M. Belnap, 2007. Bsoft: Image processing and molecular modeling for electron microscopy. Journal of Structural Biology 157, 3-18.

Hohn, M., G. Tang, G. Goodyear, P.R. Baldwin, Z. Huang*, et al.*, 2007. SPARX, a new environment for Cryo-EM image processing. Journal of Structural Biology 157, 47-55.

Hunter, J.D., 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9, 90-95.

Jones, E., T. Oliphant, P. Peterson, e. al., 2001. SciPy: Open source scientific tools for Python.

Klockner, A., N. Pinto, Y. Lee, B. Catanzaro, P. Ivanov*, et al.*, 2012. PyCUDA and PyOpenCL: A scripting-based approach to GPU run-time code generation. Parallel Computing 38, 157-174.

Lander, G.C., S.M. Stagg, N.R. Voss, A. Cheng, D. Fellmann*, et al.*, 2009. Appion: An integrated, database-driven pipeline to facilitate EM image processing. Journal of Structural Biology 166, 95-102.

Langlois, R., J. Frank, 2011. A Clarification of the Terms Used in Comparing Semi-automated Particle Selection Algorithms in Cryo-EM. Structural Biology 175, 348-52.

Langlois, R., J. Ash, J. Pallesen, J. Frank, 2014a. Fully Automated Particle Selection and Verification in Single-Particle Cryo-EM, in: Herman, G, Frank, J, Eds.), Computational Methods for Three-Dimensional Microscopy Reconstruction, Springer, pp. 97-132.

Langlois, R., J. Pallesen, J.T. Ash, D. Nam Ho, J.L. Rubinstein*, et al.*, 2014b. Automated particle picking for low-contrast macromolecules in cryo-electron microscopy. Journal of Structural Biology 186, 1-7.

Langlois, R., J. Pallesen, J. Ash, D. Ho, J. Rubinstein*, et al.*, 2014c. {Automated particle picking for low-contrast macromolecules in cryo-electron microscopy}. Journal of Structural Biology In Press.

Li, X., P. Mooney, S. Zheng, C.R. Booth, M.B. Braunfeld*, et al.*, 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat Meth 10, 584-590.

Ludtke, S.J., P.R. Baldwin, W. Chiu, 1999. EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. Journal of Structural Biology 128, 82-97.

Pedregosa, F., G.l. Varoquaux, A. Gramfort, V. Michel, B. Thirion*, et al.*, 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825-2830.

Perez, F., B. Granger, 2007. IPython: A System for Interactive Scientific Computing. Computing in Science & Engineering 9, 21-29.

Peterson, P., 2009. F2PY: a tool for connecting Fortran and Python programs. International Journal of Computational Science and Engineering 4, 296-305.

Scheres, S.H.W., 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology 180, 519-530.

Shaikh, T.R., H. Gao, W.T. Baxter, F.J. Asturias, N. Boisset*, et al.*, 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protocols 3, 1941-1974.

Sorzano, C.O.S., R. Marabini, J. Vel√° zquez-Muriel, J.R. Bilbao-Castro, S.H.W. Scheres*, et al.*, 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. Journal of Structural Biology 148, 194-204.

Tang, G., L. Peng, P.R. Baldwin, D.S. Mann, W. Jiang*, et al.*, 2007. EMAN2: An extensible image processing suite for electron microscopy. Journal of Structural Biology 157, 38-46.

van der Walt, S., E. Gouillart, T. Yu, J. Schönberger, 2011. Scikit-Image: Image Processing in Python.