STRUCTURE DETERMINATION OF HETEROGENEOUS BIOLOGICAL SPECIMENS IN CROWDED ENVIRONMENTS

by

Benjamin Andrew Himes

BS, The Pennsylvania State University, 2011

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Benjamin Andrew Himes

It was defended on

April 25, 2018

and approved by

James Conway, Professor, Department of Structural Biology

Rieko Ishima, Associate Professor, Department of Structural Biology

Joachim Frank, Professor, Columbia University, Distinguished Professor, SUNY

Dissertation Directory: Peijun Zhang, Professor, University of Oxford,
Associate Professor, University of Pittsburgh

STRUCTURE DETERMINATION OF HETEROGENEOUS BIOLOGICAL SPECIMENS IN CROWDED ENVIRONMENTS.

Benjamin Andrew Himes, PhD

University of Pittsburgh, 2018

The central dogma of molecular biology describes a strictly linear flow of genetic information stored in DNA transferred through RNA and translated into protein products. In the "post-genomic era" however, it is evident that abundant information flows from protein to protein and even protein back to DNA. The field of Structural Biology seeks to understand how the spatial and temporal organization of that information is stored and transmitted via the three-dimensional structure and dynamics of biological macromolecules. X-ray crystallography, nuclear magnetic resonance, and single particle cryo-electron microscopy (cryo-EM) are the primary techniques available to the structural biologist to deduce structure and dynamics at or near atomic resolutions. These tools are generally limited to the study of stable molecules that can be purified biochemically. Other approaches, like super-resolution light microscopy and cryo-electron tomography (cryo-ET), are amenable to the study of more labile macromolecular complexes or those found in situ; however, they are limited to resolutions of tens of nanometers. Improving the resolving capability of cryo-ET with sub-tomogram averaging to routinely reach beyond 10 Å is the primary goal of this work. My unique contribution to the field of structural biology is a suite of software tools called emClarity (enhanced macromolecular classification and alignment for high-resolution in situ tomography) which allows scientists with minimal computational background to probe the structural states of conformationally variable molecules present in complex and crowded environments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

2D/3D	two/three-dimensional
cryo-EM	cryo-electron microscopy
cryo-ET	cryo-electron tomography
cryoSTAC	cryo-sub tomogram alignment and classification
CTF	contrast transfer function
FSC	Fourier shell correlation
HIV	human immunodeficiency virus
HIV-RT	human immunodeficiency virus reverse transcriptase
HK	histidine kinase
MWE	missing-wedge effect
MTF	modulation transfer function
NMR	nuclear magnetic resonance
PSF	point spread function
SPA	single particle analysis
SNR	signal to noise ratio
SSNR	spectral signal to noise ratio
RR	response regulator

PREFACE

In memory of Dr. Klaus Schulten and Dr. Roger Hendrix who offered inspirational advice and thoughtful encouragement toward the completion of this work.

I owe a great debt to my brother David, for support at the beginning of my time in Pittsburgh as well as for challenging me to continue exploring academic thought in fields unrelated to my own. I also thank my parents Rick and Deb, for their unwavering support and love which among other things continues to embolden me to take risks and pursue the path less traveled. And to my sister Kellie, thank you for demonstrating what it looks like to return to school, kick ass and grow a relationship all with a smile. I also owe special gratitude to Brenda, from whom I continue to draw a great deal of inspiration.

I would also like to thank my adviser Dr. Peijun Zhang for providing me with ample advice and freedom throughout my studies to pursue my ambitions. And finally, I appreciate the opportunity to study at both the University of Pittsburgh and Carnegie Mellon University as part of the joint Molecular Biophysics and Structural Biology Graduate Program, headed up by Dr. James Conway and Dr. Gordon Rule.

** **Figures** or [brackets] in the following pages indicate intra-document links.
To return to where you left off, please press simultaneously the alt and left arrow key.

1.0 INTRODUCTION

Structural Biology is a discipline that seeks to find a functional and mechanistic understanding of the molecular interactions governing the life of the cell, and ultimately the organism, by determining the three-dimensional (3D) distribution of atoms in macromolecular complexes. These macromolecules are commonly referred to as “particles,” a convention I adopt here. Structural “snapshots” of particles in a given conformation, which may or may not correspond to a functional state, can be obtained from purified specimens *in vitro.* These snapshots sample a conformational landscape that is often dynamic and thereby require additional information from biochemical and biophysical experiments to build a functional model.

The techniques traditionally used to probe biomolecules on the length scale of hydrogen bonds, Nuclear Magnetic Resonance (NMR) and Macromolecular X-ray crystallography (MX), measure the characteristics of an *ensemble* of molecules. Cryo-electron microscopy (cryo-EM) on the other hand, can measure information from individual particles, and so is particularly well suited to the study of heterogeneous samples of large biological macromolecules and their complexes.

Cryo-EM records images formed in a transmission electron microscope (TEM) that are, to close approximation, projections of the 3D-Coulomb potential of the specimen as

described further in section 1.1. A simple mathematic relationship referred to as the central section theorem[1] shows that the Fourier transform of a 2D projection is equal to a central section of the 3D Fourier transform of the object. Strictly speaking, this theorem only applies where the curvature of the Ewald Sphere is negligible [1,2], i.e., where the "projection approximation" holds. In this regime, the Fourier transforms of many projections from different angles are resampled onto their respective central sections in 3D, and upon filling up Fourier space, the inverse Fourier transform returns the 3D-Coulomb potential of the object (a so-called EM map[2]). This relationship is shown schematically in **Figure 1.1**.

There are two primary varieties of cryo-EM used to obtain the angular distribution of projections needed to perform this 3D reconstruction, single particle analysis (SPA) and cryo-electron tomography (cryo-ET).

[1] Sometimes the Fourier slice theorem
[2] This is not quite the same as the result from X-ray crystallography, which is a 3D electron density map.

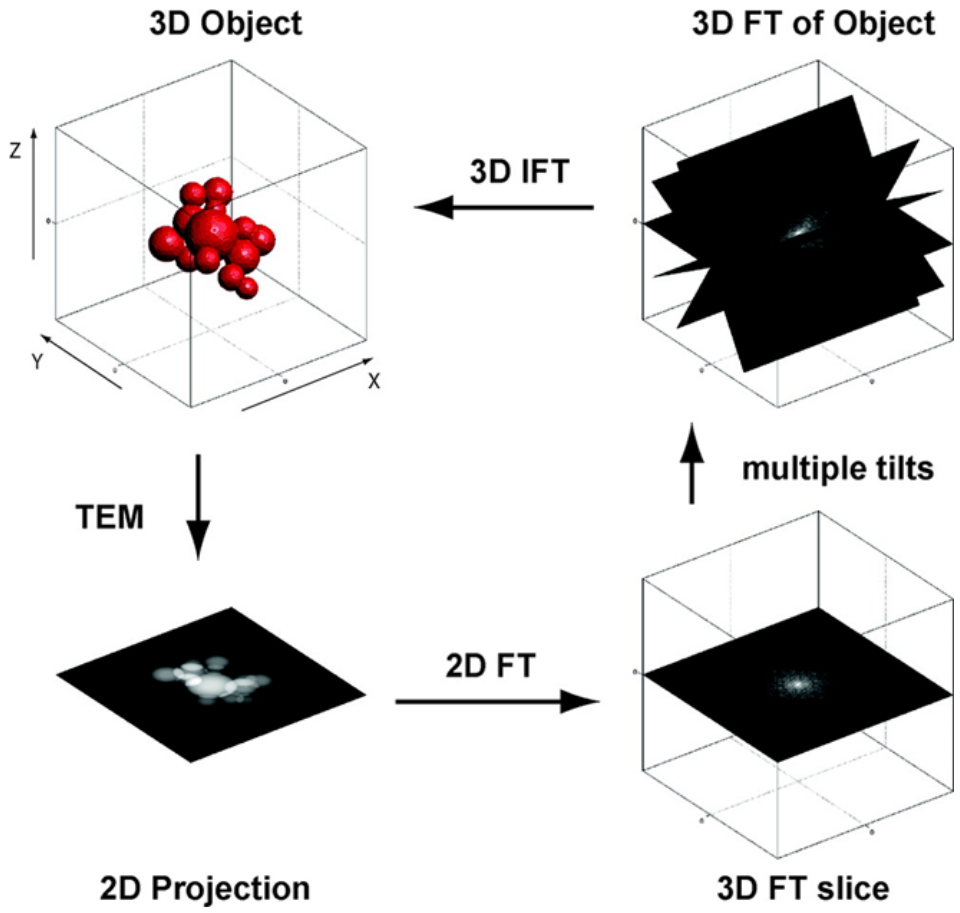

Figure 1.1 Schematic illustrating Fourier Reconstruction

The Fourier slice theorem is invoked to reconstruct the 3D electrostatic potential of the specimen by filling in central sections of its 3D Fourier transform with the Fourier transform of 2D projections from many angles. Reproduced from [3] with permission.

Cryo-ET is an alternative technique that provides 3D reconstructions for individual instances of macromolecules, even in a crowded environment where particles overlap in projection [7]. The additional information about each particle is made available by collecting many projections of the same sample as it is rotated relative to the optical axis in the TEM [8]. These tilted images provide a modest resolution along the optical-axis at the expense of the attainable global resolution, which is usually in the tens of nanometers [9].

This resolution may be improved using a hybrid of SPA and cryo-ET called cryo-electron tomography with sub-tomogram averaging and classification (cryoSTAC[3]) [10]. The input for cryoSTAC is many 3D sub-tomograms that are cut out *in silico* from the full tomogram, although in principle, nothing is preventing the use of projection data directly as in SPA. The main distinction between SPA and cryoSTAC is that the latter produces a 3D reconstruction for every individual particle.

Another more practical difference is that cryoSTAC lags significantly behind SPA in the resolutions routinely obtained [11], owing to a number of obstacles to reaching a sub-nanometer resolution *in situ* using cryoSTAC. In this dissertation, I focus on new methods in image processing address these challenges, in particular, correction for Ewald sphere curvature, spatially variable sample distortions and displacements, and reduction of the computational cost of working with the thousands of 3D volumes required. All of these

[3] For an in-depth treatment of cryoSTAC and background on cryo-ET the reader is referred to [41].

are shown to contribute to making resolutions better than one nanometer routinely accessible in cryoSTAC. In the remainder of this chapter, I introduce the essential background needed to understand the results of this dissertation. References to comprehensive treatments of specific topics are made where needed; for a history of the development of the field of cryo-EM with an emphasis on cryo-ET the reader is referred to the introduction in [12].

1.1 IMAGE FORMATION IN THE TRANSMISSION ELECTRON MICROSCOPE

The transmission electron microscope (TEM) has the same essential components as an optical light microscope operating in transmission mode: a radiation source, a condenser lens system that focuses the radiation onto the sample, a transparent sample, an objective lens that focuses the scattered radiation, a magnifying lens system, and a detector. Compared to light, electrons interact very strongly with matter, and many specialized forms of electron microscopy have been developed to take advantage of this fact (chapters 2-9, Hawkes and Spence) [14]. This strong interaction requires the TEM column to be maintained at a high-vacuum to prevent spurious scattering of electrons outside of the specimen. This high-vacuum in turn causes a liquid sample to rapidly evaporate; one solution to this problem that also preserves high-resolution details, is to

freeze the specimen and maintain it at temperatures around 80 K using liquid nitrogen[4]. To avoid the formation of crystalline ice and instead produce a vitrified ice, preserving a nearly native hydrated state, the freezing may be accelerating by using a cryogen with a larger heat capacity than liquid nitrogen, such as liquid ethane or liquid propane [13].

For our purposes, we are concerned with the high-resolution TEM also known as phase-contrast TEM. The three primary events are shown schematically in **Figure 1.2**: the electrons scatter from the specimen, they are focused into a diffraction pattern in the back focal plane of the objective lens, and finally, they form an image in the conjugate image plane.

"Real" and "reciprocal" spaces referred to in **Figure 1.2** are synonymous with "position" and "momentum" space[5]. These are dual spaces mathematically related to one another by the Fourier transform, a fact that is used extensively in image formation theory and image processing in TEM [14].

---

[4] It is thought that using even lower temperatures via liquid helium would help to provide further protection against radiation damage, however, there is a poorly understood loss of contrast and apparently increased specimen motion [149,150].

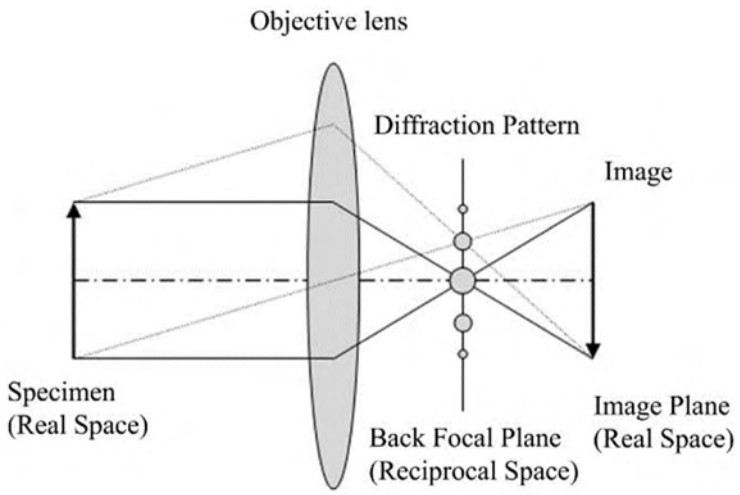

Figure 1.2 three central aspects of image formation

1.1.1 Imaging aberrations

Due to the imperfect nature of magnetic lenses, several wave aberrations distort the exit wave. The strongest of these are the third-order[6] aberrations, also called the primary Seidel aberrations, of which the spherical aberration, coma, and astigmatism are most pertinent to cryo-EM of biological macromolecules [15].

The coefficient of spherical aberration is fixed for a given objective lens [20]. The other aberrations may be corrected through microscope alignment [21,22], and residual errors in microscope alignment may be corrected computationally in later computer

[6] The ideal gaussian wave front is quadratic.

processing of the images. Correction of residual astigmatism becomes increasingly relevant beyond 8 Å resolution and so has been routinely done in SPA [23]. Correcting residual axial coma (beam tilt) was discussed early on [24], and first demonstrated in 1986 by Henderson and colleagues [25]. These image processing corrections are rooted in contrast transfer theory.

1.1.2 Electron scattering

When speaking of scattering, it is perhaps intuitive to envision electrons as individual particles, however, considering the wave nature of electrons makes many aspects of the theory of image formation easier to discuss. The incoming high-energy electrons may be represented by a wave function, and if the object is thin enough[7] to be treated as if it were at a single distance from the focal plane of the objective lens, then the "projection approximation" simplifies the expression for the exit wave function to:

$$\psi_{exit}(x,y) \cong \psi_{in}(x,y)e^{i\sigma \int V(x,y,z)dz} \qquad eq\ 1.1$$

Where $V(x,y,z)$ is the object's Coulomb potential and $\sigma = 2\pi\lambda m_e e h^{-2}$, with the relativistic electron wavelength (λ) mass (m_e) and charge (e) and the Planck constant h.

[7] "Thin enough" depends on the atoms present, the imaging parameters, and the resolution of a particular experiment as discussed in detail in [16].

For a derivation of the transmission function (the exponential on the right in eq 1.1) the interested reader is referred to the treatment by De Graef [15].

If, in addition to being thin, the specimen is composed primarily of light atoms, as is the case in biological specimens, the resulting projected potential will be << 1. In this case, equation 1.1 may be expanded into a Taylor series and truncated after the first two terms, giving the approximate exit wave function under the weak phase object approximation (WPOA) in equation 1.2 [17]:

$$\psi_{exit}(x,y) \approx \left(1 + i\sigma V_z(x,y)\right) \qquad eq\ 1.2$$

The exit wave function then propagates through the vacuum in the TEM column and is focused using electromagnetic lenses, as first practically demonstrated in 1931 through the efforts of Knoll and Ruska [18]. These lenses are quite imperfect and modify the exit wave as it is focused. The resulting wave aberrations (section 1.1.2) were quantified by Scherzer in the late 1940s [19], and their impact on image formation is described by contrast transfer theory (section 1.1.3).

1.1.3 Contrast transfer theory

The collective effect of spherical aberration, imaging out of focus, and axial astigmatism is a spatial frequency dependent phase shift [76]

$$\chi_q = \pi\lambda q^2 \left(-\frac{1}{2}C_s\lambda^2 q^2 + \left(\frac{\Delta f_1 + \Delta f_2}{2} + \frac{\Delta f_1 - \Delta f_2}{2} cos\ 2(\varphi_0 - \varphi)\right)\right) \qquad eq\ 1.3$$

The relativistic wavelength (λ) is calculated from the accelerating voltage[8], and the spherical aberration (C_s) is constant for a given microscope[9]. Spatial frequency (q) is the independent variable leaving the astigmatism angle (φ) which is defined as the angle between the major axis and the image X-axis, and under focus[10] (Δf).

How this phase shift affects the transfer of information from object to image is the subject of contrast transfer theory, and the reader is referred to chapter 3.3 in [27] for an accessible treatment. A more detailed discussion relating contrast transfer theory to image reconstruction can be found in [151]. Mathematically speaking, for a weakly scattering object, the Fourier transform of the bright field image is obtained by multiplying the Fourier transform of the projected object with the contrast transfer function (CTF).

$$\boldsymbol{CTF} = -\sqrt{1-A^2}sin(\chi_q) - Acos(\chi_q) \qquad eq\ 1.4$$

In equation 1.4 “A” refers to the percentage of amplitude contrast, which can be approximated as constant with respect to spatial frequency and is typically assumed to lie between 0.07-0.14 [26]. As the sinusoidal form of equation 1.4 suggests, the CTF

[8] 2.51 pm for 200 KeV and 1.97 pm for 300 KeV radiation.
[9] For the instruments currently used for biological specimens this is usually around 2 mm, while for those used in material science the value is often closer to 0.5 mm. The worse aberration in biological TEM is a compromise on the size of the lens pole piece gap [20].

[10] Note that in biological TEM, the convention is to define under focus > 0, such that the origin of the microscope is set to the *back*-focal plane of the objective lens.

produces characteristic oscillations in the power spectrum of the object, which may be estimated from the periodogram[11] as described in more detail in chapter 2.2.4.

A qualitative understanding of the effect of the CTF is useful and is illustrated in **Figure 1.3** Using the atomic coordinates for the enzyme Beta-galactosidase (PDB 6CVM), the projected object potential[12] as calculated using the multi-slice approach [152] is shown in **Figure 1.3** A. The CTFs for three under-focus values are plotted in **Figure 1.3** B. Including these in the wave propagation shows that the signal is increasingly delocalized as the lens is set farther from focus (**Figure 1.3** C-E). Also, the spread is frequency dependent; the low-resolution contrast is enhanced while the high-resolution details are now primarily outside the particle envelope. This increase in low-resolution contrast far from focus is commonly exploited to facilitate later image processing steps in both SPA and cryoSTAC. **Figure 1.3** F-H shows images simulated with the same defocus values but with explicit water molecules, 1.5x the particle thickness, and shot noise following a Poisson distribution.

[11] The Fourier transform of the image auto-correlation function.

[12] the integral inside the exponential term in equation 1.1.

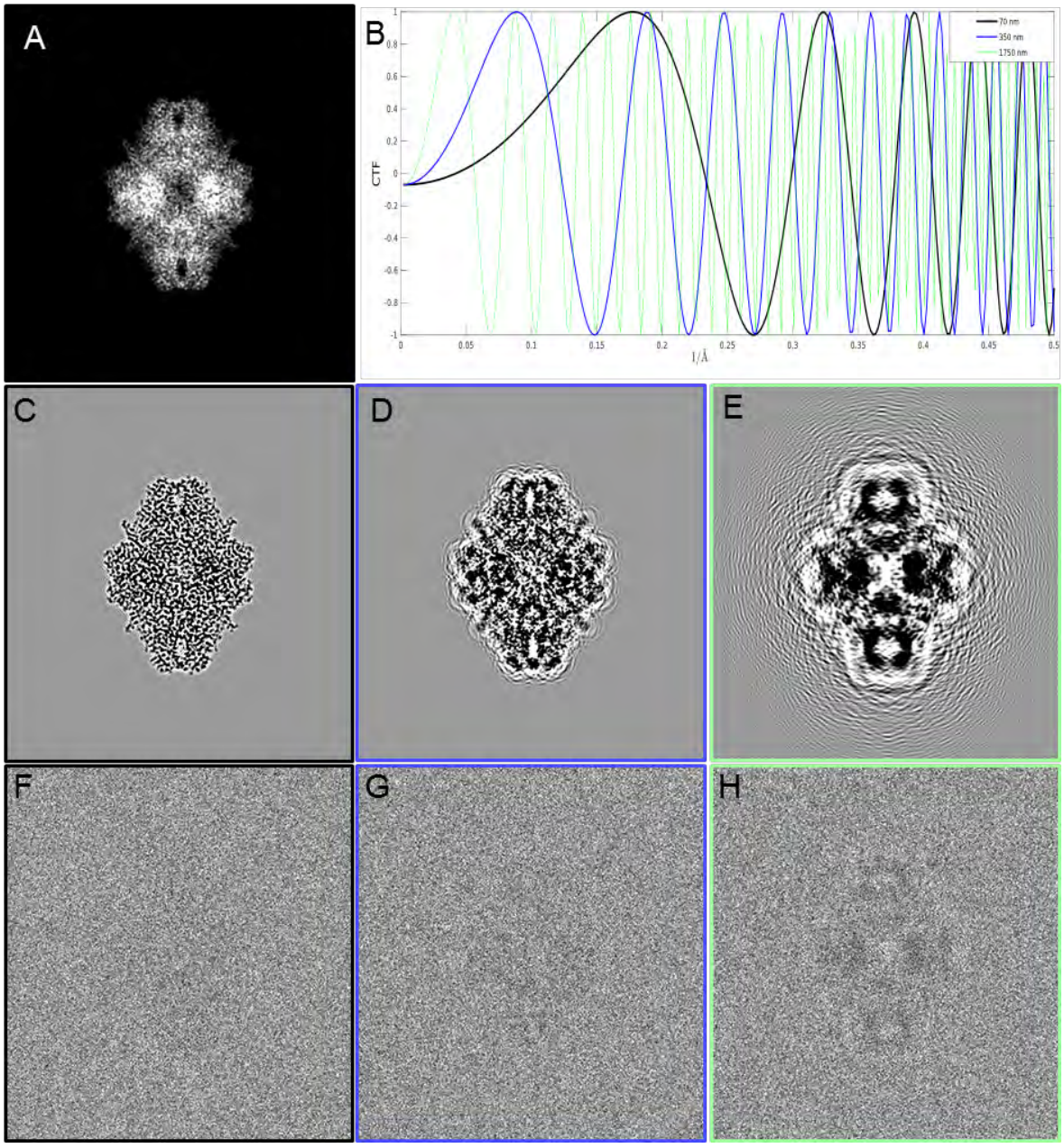

Figure 1.3 Visualizing the CTF as a function of under-focus

(A) Projected electrostatic potential using atomic coordinates from PDB-6CVM. (B) The plot of the CTF for 70 nm (black), 350 nm (blue), and 1750 nm (green) under-focus. (C-E) Multi-slice propagation of assuming the increasing defocus plotted in (C). (F-H) the same simulation now including explicit water molecules to give a realistic image.

1.2 IMAGE RESTORATION

Fortunately, the effects of the CTF, within the approximations of linear imaging theory, are relatively straightforward to correct. Because there are regions where the CTF does not transmit any information to be recorded, a range of imaging parameters must be used on different images, which are then combined into the final result. The modulation by the CTF is commonly restored by a Wiener filter, which has been described succinctly as a "careful division" by the CTF [27]. The care that is referred to here is the inclusion of an estimate of the spectral signal to noise ratio in the denominator, which prevents overamplification of noise.

$$F^{Wiener} = \frac{CTF_i}{\sum_i^N \|CTF_i\|^2 + {}^1/_{SSNR}} \qquad eq\ 1.5$$

The estimate of the SSNR is strongly impacted by the solvent surrounding the particles in the sample [28], and in Chapter 2 I present a solution to this problem tailored to cryoSTAC. In this chapter, I also discuss solutions that help to alleviate deviations from linear imaging theory that are particularly relevant to thicker samples often used in cryoSTAC, in particular, the breakdown of the projection approximation.

One additional assumption in image restoration is that the 3D Fourier transform of the object is adequately sampled to permit interpretation up to some resolution. A simple condition was given by Crowther that reveals a linear relationship between the resolution attainable and the diameter of the specimen in real space divided by the angular sampling [29].

1.3 THE MISSING WEDGE-EFFECT

While I have stated that there is a 3D reconstruction for each particle in a tomogram, it is vital to note that this reconstruction is distorted due to an incomplete angular sampling of Fourier space. The angular range the specimen is tilted to is limited in practice to ~+/- 60˚ although for samples that are very thin, especially 2D crystals, closer to 70˚ may be reached [30]. This sampling is limited primarily by the increased probability of scattering due to the additional thickness of the specimen, inversely proportional to the cosine of the tilt angle, which results in fewer electrons reaching the detector in total, while a higher percentage of those that do are due to inelastic scattering such that the higher tilts have lower SNR. **Figure 1.4** A, B show the 3D sampling function which represents the achievable sampling in Fourier space. The oscillations are due to the CTF, while the color runs from low-resolution (blue) to high-resolution (orange).

Not only does the high-resolution information fade at high tilts, but it also progressively degrades as the specimen accumulates radiation damage. The region of lowest SSNR is the wedge-shaped region perpendicular to the tilt axis in **Figure 1.4** A.

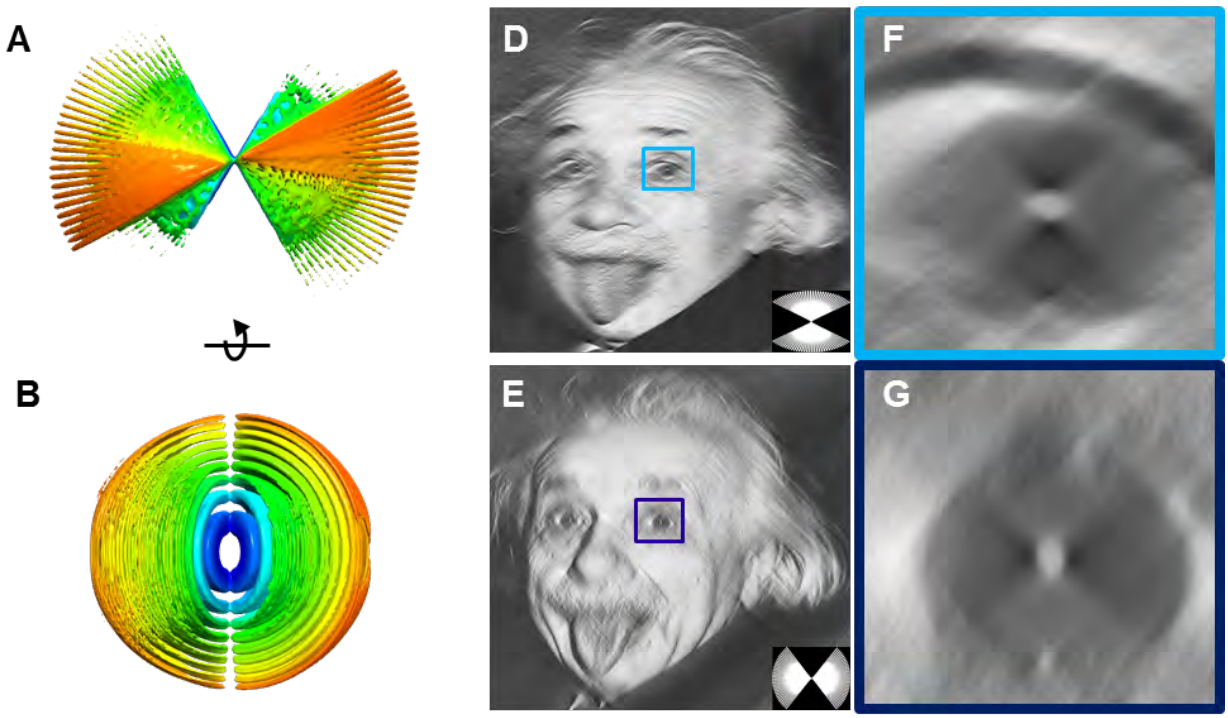

Figure 1.4 The missing-wedge effect

A) 3D sampling function in Fourier space for a typical tomogram, acquired with a bi-directional tilt-scheme starting from 30˚. The colors indicate spatial frequency (1/resolution) running from low (blue) to high (red). (B) looking down along the beam direction, the oscillations of the CTF are apparent. (D-E) applying a 2D missing-wedge to the same image illustrate the two primary effects: elongation along the missing-wedge direction, and removal of linear/planar features perpendicular to the wedge. (F-G) magnified view of Einstein's eyes from D and E.

The “missing-wedge effect” refers to the distortions in the 3D reconstruction due to this under-sampling. In **Figure 1.5** D an image of Einstein with a horizontal missing-wedge (inset) and zoomed in portion of his eye in **Figure 1.5** F shows the characteristic drop in resolution and elongation of density parallel to the missing-wedge, especially evident in the normally circular pupil. A second significant effect is that linear or planar features perpendicular to the wedge are not well represented, as illustrated in **Figure 1.5** F, G where Einstein’s eyelid is no longer visible.

1.4 CHALLENGES IN STRUCTURAL STUDY OF HETEROGENEOUS SAMPLES

Arguably, the greatest strength of cryo-EM as a technique, compared with NMR and MX, is the ability to record and analyze measurements from individual instances of biological macromolecules. Although averaging is necessary to reach SNR levels high enough to reconstruct interpretable density maps, the ability to use pattern recognition and machine learning algorithms permits the analysis of individual particles. In the case of cryoSTAC where each particle is represented by a unique 3D-density map, the analysis of individual particles is particularly promising granted the image features due to the missing-wedge can be accounted for.

The missing-wedge effect distorts the sub-tomograms strongly and impacts all aspects of image processing in cryoSTAC. Current state-of-the-art software makes compromises in accurately compensating this effect in exchange for computational

tractability. In addition to the well-known missing-wedge effect, local specimen distortions and the accuracy of CTF correction considering the breakdown of the projection approximation in thick specimens also need to be addressed if cryoSTAC is to reach a higher resolution routinely. The primary focus of my research has been developing methods for the determination of these maps from variable and heterogeneous specimens, at resolutions better than 10 Å as demonstrated for the icosahedral hepatitis B virus capsid using SPA some 20 years ago [31]. These methods have been developed and tested on three primary biological samples: ribosomes, bacterial chemotaxis receptor signaling arrays, and purified HIV-1 virus-like particles.

2.0 HIGH-RESOLUTION STRUCTURAL DETERMINATION OF HETEROGENEOUS SPECIMENS USING CRYOSTAC

Macromolecular complexes are intrinsically flexible and often challenging to purify for structure determination by single particle cryo-EM. Such complexes may be studied using cryoSTAC, which in exceptional cases reaches a sub-nanometer resolution, yielding insight into structure-function relationships. All maps from cryoSTAC currently deposited in the EMDB with resolution < 9 Å are from macromolecules that form ordered structural arrays, like viral capsids, which dramatically simplifies structural determination. Extending this approach to more common specimens that exhibit conformational or compositional heterogeneity and may be available in limited numbers remains challenging. To this end, I developed **emClarity,** a GPU-accelerated image processing package, and demonstrate significant improvements in the resolution of maps compared to those generated using the current state-of-the-art software. Furthermore, I devised a novel approach to sub-tomogram classification that reveals conformational states not previously observed with the same data.

2.1 INTRODUCTION

Recent advances in the capabilities of cryo-EM are changing how we think about the structural determination of complex biological assemblies. In addition to reaching a near-atomic resolution, the development of advanced classification techniques, such as maximum-likelihood classification as implemented in FREALIGN [32] or RELION [33], has enabled possibilities for probing macromolecular functional dynamics using an SPA approach. For a sample to be suitable for SPA, it must yield tens to hundreds of thousands of individual instances of biological macromolecules [34]. Commonly called “particles,” they must first be purified to relatively high compositional and conformational homogeneity [35] and subsequently imaged in many different orientations. These two conditions are often difficult to achieve, especially as the size and number of components increases, and are features common to the assemblies of biological complexes often found at the heart of cellular activities [9]. When this is the case, cryo-ET is capable of generating 3D reconstructions of pleomorphic samples *in situ* is the preferred approach

These reconstructions (tomograms) are generally limited to 3-4 nm resolution. This limit on the resolution of a tomogram is largely due to the extremely limited electron dose (1-2 electrons/Å^2 per projection, ~60-120 electrons/Å^2 in total) chosen to minimize radiation damage to samples during the collection of many projection images. Additionally, the signal in these noisy images is not distributed evenly in the tomogram, a consequence of primarily the increasing specimen thickness at high tilt angles which limit the angular

sampling to about +/- 60° in practice. This uneven sampling of the 3D specimen in 2D projections results in anisotropic resolution, and the resulting distortion is colloquially referred to as the “missing-wedge effect” (MWE) named for the shape of the sub-sampled region in Fourier space of a single-tilt-axis tomogram.

These resolution-limiting issues can be overcome when many copies of a macromolecule are present in a tomogram, extracted *in silico*, aligned to a common reference frame, and averaged using cryoSTAC procedures that share many similarities to SPA. Such averaging can increase the signal to noise ratio (SNR) in the final map and also complete the angular sampling in the average, alleviating the MWE. Classification refers broadly to the use of statistical analysis to sort out heterogeneity as described later.

For cryoSTAC to work, the bias due to the MWE during image processing must be mitigated. This is accomplished by explicitly considering the contribution of each sub-tomogram to the final average as a function of spatial frequency in Fourier space. In practice, this means the (dis)similarity metric used for alignment and classification must be modified to consider only the sampled regions of each volume. The two most common metrics are the constrained cross-correlation [36–38] and the constrained-Euclidean distance [39,40])

The function used to constrain these distance metrics depends on the sampling of the specimen and describes the extent of information transfer during the imaging process. This 3D-sampling function (3D-SF) defines the information transfer based on all contributing projection’s CTFs, the exposure in each projection, the sample thickness as a function of the tilt angle, and the weights applied in the reconstruction of the tomogram

(**Figure 1.4**). The 3D-SF has also been referred to as the 3D Contrast Transfer Function (3D-CTF) [40]. However, this creates some confusion with an unrelated process called 3D-CTF correction. To avoid ambiguity, I avoid the term "3D-CTF" altogether as it is a bit of a misnomer given the CTF is inherently 2D. Instead, I refer to "3D-CTF correction" as Ewald sphere correction since the variation of focus with depth is equivalent to the curvature of the Ewald sphere [1]. The reader is referred to recent reviews which cover both the general principles [41] and computational approaches [42,43] in greater depth.

Even a rough estimate of the 3D-SF using a binary missing-wedge mask makes it possible to obtain sub-tomograms averages at low resolutions, 15-20 Å. With the particles aligned to a common reference frame, any conformational differences that exist should be identified and the resulting subgroups separated into different averages. This characterization is commonly referred to as "classification" and allows for separation of multiple biological states from the same sample, in turn permitting *in situ* structural determination of functional conformations [30,44].

The ability to perform 3D-classification of unknown states semi-independently from the alignment procedure is arguably the greatest strength of cryoSTAC relative to SPA because each particle is reconstructed as a unique, albeit distorted, 3D volume. With these per-particle 3D reconstructions, it is possible to directly analyze the 3D-variance without any bootstrapping techniques, an approach useful for focusing in on dynamic portions of the specimen, the value of which has been discussed extensively [45–47].

To date, few structures have been solved at resolutions better than 10 Å using cryoSTAC [40,48–53]. This resolution is a critical threshold beyond which flexible molecular fitting approaches are more reliable [54]; an avenue for integrating high-resolution data into medium resolution maps and investigating dynamic or transient complexes [30,49,55].

I present here a complete set of GPU-accelerated programs called **emClarity** for **e**nhanced **m**acromolecular **cl**assification and **a**lignment for high-**r**esolution ***i****n situ* **t**omograph**y**, designed to make the sub-nanometer resolution the norm rather than the exception for cryoSTAC. I have focused my efforts on the steps in cryoSTAC image processing that are likely to yield the greatest improvements, as suggested by empirical observation and theoretical calculations [50,56–58]: accuracy of tilt-series alignment, improved defocus determination and CTF correction, explicit treatment of anisotropic resolution, and more robust classification. A typical cryoSTAC workflow is illustrated in **Figure 2.1** with areas of significant improvement in emClarity highlighted in red, while novel additions to the pipeline are in orange boxes.

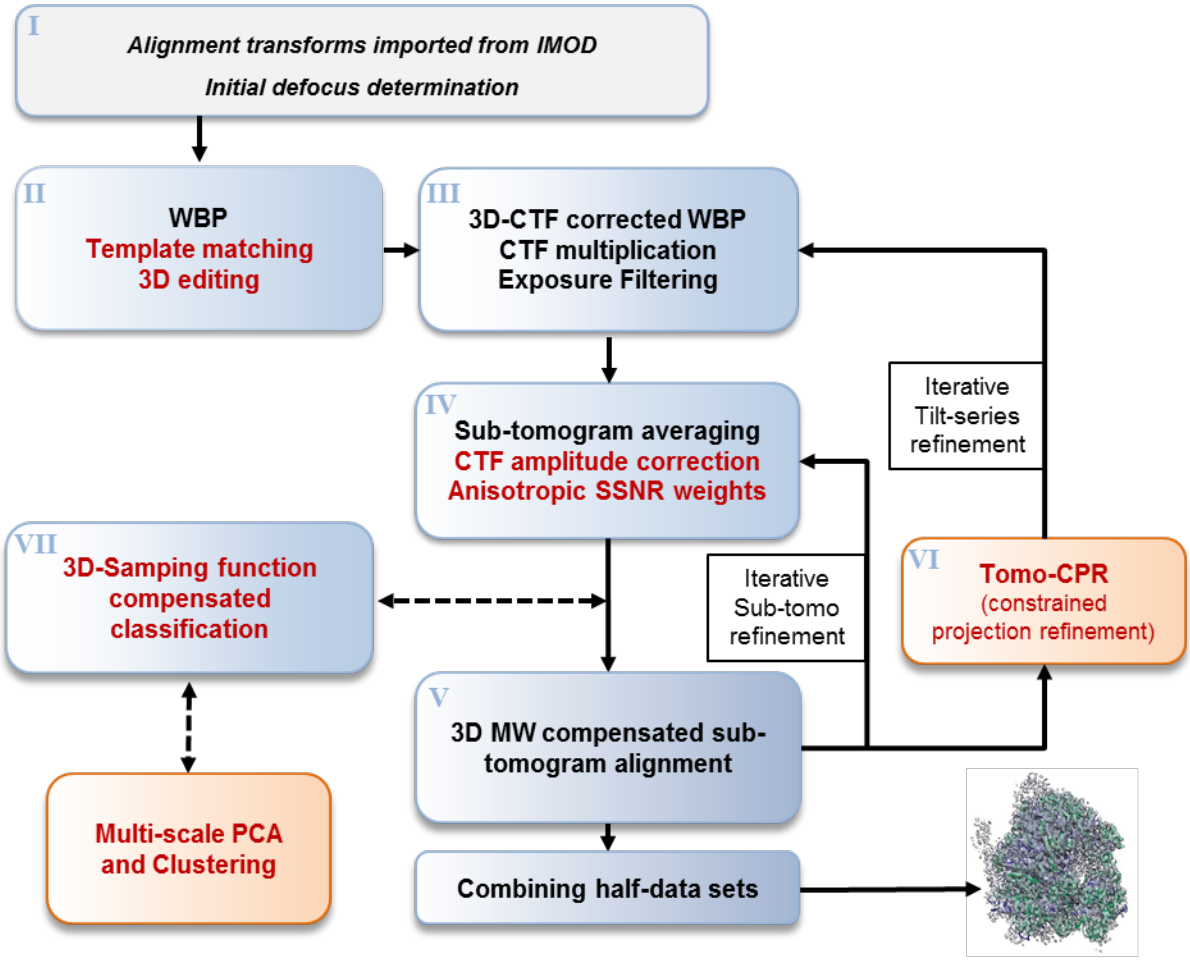

Figure 2.1 Typical cryoSTAC workflow

Blue background indicates steps in a typical cryoSTAC workflow, where red text denotes routines substantially improved in emClarity. Orange background indicates novel algorithms not available in other published software. Double-headed, dashed arrows indicate optional branch points that may be included during any given cycle of refinement.

2.2 METHODS AND ALGORITHMIC DEVELOPMENTS

All cryo-EM methods seek to refine the orientation of each particle with respect to the microscope reference frame. In SPA this is a 5-parameter search space, while in cryoSTAC it is a 6-parameter search space, three Euler angles, and three spatial coordinates. This search is carried out by comparing a reference structure to each particle at different orientations and scoring each using constrained-cross-correlation (least squares minimization) or Euclidean l2-norm (likelihood maximization.)

Caution must be taken to avoid fitting to the noise in cryo-EM data [59]. A modification to the iterative methods used, deemed the gold-standard approach [60], robust to this problem and independent of the alignment algorithm, has been demonstrated to produce satisfying results [61]. Care must be taken in cryoSTAC where a reference-based search (template matching) is often needed to pick out the particles of interest from the tomograms before entering the sub-tomogram workflow. The information used in the template matching program is generally restricted to 40Å as is common in the field [36,53]. This resolution has been sufficient to locate and roughly orient particles of various shape and dimension while also limiting any correlation between half-sets to at most ~ 26Å[13], a resolution beyond which the data are considered independent [62].

[13] The real-space masking of the particle introduces an extra correlation in Fourier space equal to $2/D$, where *D* is the diameter of the mask [144]. For a globular protein of molecular weight 250 KDa, the diameter is ~ 75 Å, and so an upper bound on the correlation introduced is $(40\ \text{Å})^{-1} - (75/2\ \text{Å})^{-1} \cong (26\ \text{Å})^{-1}$.

During alignment, a low-pass filter shaped according to a modified version of the figure-of-merit [63] which is based on the gold-standard Fourier Shell Correlation (FSC) is applied to the data. The modification I make forces the Fourier amplitudes to zero after the point where the 1-bit criterion [64] interests the FSC curve. This point is often near an FSC of 0.33 and seems to provide a reasonable balance between reducing over-fitting while also permitting convergence, which may be very slow if too little detail is available in the reference used for refinement.

2.2.1 Refinement of tilt-series alignment

A prerequisite for 3D reconstruction of a tomogram is the refinement of the projection geometry, including tilt-axis angle, in-plane shifts and rotations, magnification, tilt angle, and possibly other distortions like non-perpendicularity of the electron beam or skew between the x & y-axes [65]. This process (tilt-series alignment) is most commonly accomplished by using gold beads as high-contrast fiducial markers. Additional approaches, based on locating and tracking image features [66–68] or on projection matching using an intermediate tomogram as a 3D model [38] are available, but their success is generally sample-dependent [41], and these require significant user input, which has been somewhat ameliorated by automation with the recently released Appion package [69].

I have integrated into emClarity a novel algorithm called tomo-CPR (tomogram constrained particle refinement) for the iterative refinement of the tilt-series alignment using an approach that shares some similarity with the “particle polishing [70]”

implemented for SPA in RELION. The most important difference is that the reference projections used for orientation determination include information from neighboring particles as well as non-particle information from the tomogram. Adding this signal to the reference accomplishes essentially the same thing as subtracting it from the data. Another smaller difference is in how tomo-CPR constrains neighboring particles to behave similarly. As in SPA they are constrained within a given projection but are additionally required to vary smoothly as a group from projection to projection through the tilt-series.

Tomo-CPR is illustrated in **Figure 2.2.** First, I replace the density corresponding to our particles of interest in the original tomogram with a copy of the high SNR reconstruction and then re-project that synthetic tomogram along with a 3d model of the sub-tomogram origins using the IMOD program *tilt*. This also includes any local alignments previously determined and allows us to create a reference tilt-series along with a model for each sub-tomograms position in the 2d-projections. Tiles around each projected sub-tomogram origin are masked out and convolved with the CTF of the data projection at that point, using a defocus calculated from geometric considerations of the offset to the tilt-axis and the tilt-angle. The sub-tomogram fiducial positions in the data projections are then refined via cross-correlation, and these refined positions are then fed into IMOD's tiltalign as if they were derived from gold-fiducial, allowing us to take advantage of local refinements and robust fitting as described previously [71].

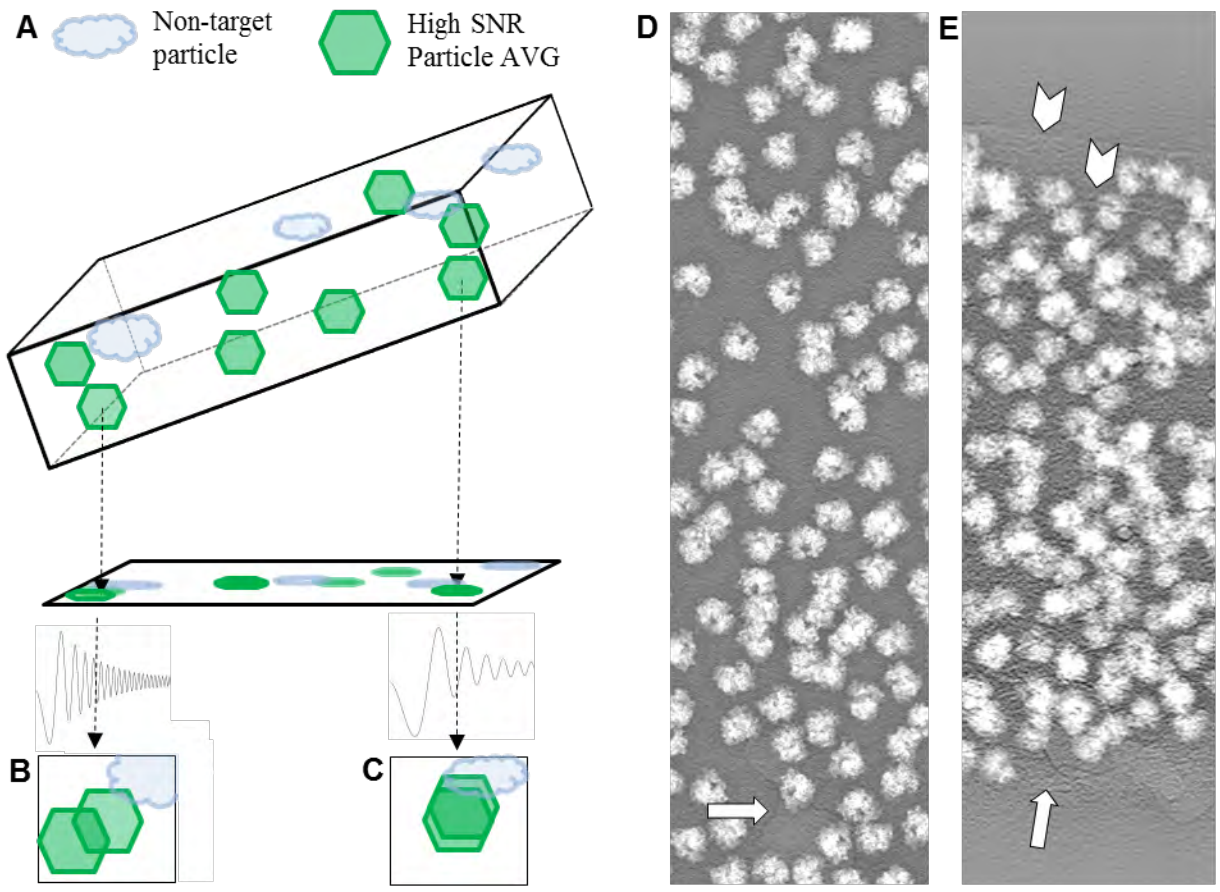

Figure 2.2 Tomogram constrained projection refinement (tomo-CPR)

(A) Schematic overview for reference generation in the Tomo-CPR. The instances of structural noise resulting from the complex 3D environment of the sample are captured by projecting the full tomogram with the particle of interest replaced by the high SNR average. (B-C) Cartoon illustrating overlapping information in the projections arising from other particles, other components in the specimen, and variable defocus as a function of tilt. Examples of non-tilted (D) and tilted (E) projections used to generate references for the yeast 80S tomo-CPR. Tiles centered on the projection origin of each sub-tomogram are convolved with the CTF considering the individual sub-tomograms defocus, which varies with tilt angle and location in the tomogram. Particles are observed to overlap, even without tilting, while features due to contaminants (white arrows) and the carbon edge (white chevron) would affect the accurate determination of local particle drift using a simple projection of the average for cross-correlation.

2.2.2 Maximizing weak signal in the reconstructions

2.2.2.1 Real space masking

Real space masks that follow the particle envelope are useful in the maximization of the SSNR as they exclude solvent. However, care must be taken to avoid inadvertently introducing spurious correlations by application of the same real space masks to data from each half-set [72]. For each half set, I apply a simple but effective iterative dilation algorithm to derive a soft-edged mask that loosely follows the particle's envelope. I start from the highest intensity pixels found within an envelope defined by a smoothed version of the current map, and gradually relax the threshold for neighboring pixels of those included in the previous iteration, thereby enforcing connectedness while allowing weak density to contribute **Figure 3.3**. Including weak density based on spatial constraints means that flexible regions of macromolecules or complexes that have variable occupancy are retained by using such a masking approach that is based on connectivity.

← Median filter a copy of input map

← Select pixels > 4σ above the mean as seeds to exclude dust

← Dilate the binary mask progressively adding connected pixels above a gradually relaxed threshold

← Calculate the Euclidean distance from all zero pixels to the nearest non-zero, adding those that are < 10 Å.

← Smooth the binary mask three times and normalize to a max value of one.

Figure 2.3 Iterative real-space masking algorithm

2.2.2.2 Fourier space masking

Image features of a given size distributed across the map in real space are co-localized in reciprocal space by spatial frequency, such that masks can be applied to specific resolution bands. High-resolution features degrade at a faster rate as a function of electron exposure, and so I have adapted for projections of tilted specimens the optimal exposure filter [73] described initially for SPA. Briefly, I apply the exposure-based filter to each projection during correction of the CTF phase inversions before reconstruction. Based on some anecdotal observations, this filter may be too severe, and future investigation into sample dependent filtering may prove fruitful in further extending the resolution.

2.2.3 3D-SF Calculation

$$3D\ Sampling\ Function\ \equiv SF^{3D} = \sum_{j=1}^{S}\sum_{i=1}^{Z} T^{i,j}\left|CTF_i^{2d}\right|^2 R^{2d} ExpFilter_i^{2d} \qquad eq\ 2.1$$

The first term in the summation is the combined transformation of projection (i) into the tomogram and sub-tomogram (j) into the final average. The second term is the standard expression for the CTF limited to third-order Seidel aberrations [74], the third is the Radial weighting used for single-axis tilt geometry, and the fourth is the optimal-exposure filter as defined [73], Z is the number of projections in each tilt-series, and S the number of sub-tomograms.

2.2.4 Improved defocus determination

Even with the advent of direct electron detectors, estimating the defocus in an image with a dose of only 1-2 $e^-/Å^2$ requires precise microscope alignment, optimized data collection schemes and a sample that provides abundant signal [48]. For cases where these conditions are not met, I have devised a new algorithm to maximize the information available for estimating the defocus value. For the initial defocus determination, I apply rotational averaging to the power spectrum as estimated from periodogram averaging, as described previously [75]. After obtaining this primary estimate of the mean defocus at the height of the tilt-axis, I then fit a 2D astigmatic function to the average power spectrum without rotational averaging, as is routine in SPA [76] using a low pass version of the power spectrum for background subtraction [77].

The number of periodograms available for coherent averaging is substantially fewer in tilted images due to the defocus gradient perpendicular to the tilt-axis. To increase the number of periodograms available for coherent averaging, the known defocus gradient may be used to resample the portion of the power spectrum between the first two zeros of the CTF [78], an approach currently used in IMOD to enhance the accuracy of the global defocus determination. This factor is determined to be the ratio of the given tile to the defocus at the tilt-axis (personal communication, David Mastronarde). I derived a more accurate scaling factor which extends this concept to resample *all* the CTF information.

The approach takes advantage of the discrete Fourier transform's implicit dependence on the sampling rate (S) and image size (N) in a way that effectively maps an image with defocus $\Delta f_0 + \Delta\Delta f \rightarrow \Delta f_0$ by changing where the spatial frequency is sampled in the image of the transform.

First, I note that the phase aberration of the CTF is dominated by the defocus term:

$$\chi_q \cong -\pi\lambda\Delta f q^2 \quad \because \quad \frac{\lambda\Delta f q^2}{\frac{1}{2}C_s\lambda^3 q^4} = \frac{\left(2\cdot10^{-12}\right)^1\cdot\left(4\cdot10^{-6}\right)^1\cdot(10^9)^2}{\frac{1}{2}\cdot(2\cdot10^{-3})^1\cdot(2\cdot10^{-12})^3\cdot(10^9)^4} = \frac{8}{8\cdot10^{-3}} = 10^3 \qquad \text{eq 2.2}$$

From there it is simple to show that:

$$-\pi\lambda\Delta f_0 {q_0}^2 \;=\; -\pi\lambda(\Delta f_0 + \Delta\Delta f){q_1}^2$$

$$\Delta f_0({q_0}^2 - {q_1}^2) = \;\Delta\Delta f {q_1}^2$$

$$\frac{\Delta f_0}{\Delta\Delta f}\;({q_0}^2 {q_1}^{-2} - 1) \;= 1$$

$$q_1(n) = \frac{n - N/2}{N \cdot S}\left(1 + \frac{\Delta\Delta f}{\Delta f_o}\right)^{\frac{1}{2}} \qquad eq\ 2.3$$

Any error due to the approximations made in Eq. 2.2 will become increasingly significant as $\Delta\Delta f \rightarrow \Delta f_o$, i.e., for images recorded closer to focus. To address this shortcoming, I use eq. 2.3 only as an initial estimate for the scaling factor, which is then refined for each row of tiles by maximizing the correlation measured between calculated 1D CTFs with defocus $\Delta f_0 + \Delta\Delta f \;\rightarrow\; \Delta f_0$ at frequencies beyond the first zero crossing.

2.2.5 Improved CTF correction

The two predominant software packages for tilt-series CTF correction are CTFPLOTTER and CTFPHASEFLIP [78] included in the IMOD package [79], and TOMOPS and TOMOCTF [75]. They differ mainly in their approach to background subtraction and amplitude restoration, the former relying on an estimate of the detector's MTF, and the latter implementing a novel filter that acts similarly to a Wiener filter while tuning the strength of signal dampening in three different regions. Both approaches are sub-optimal as they rely on amplification of intensity in individual projections which are particularly noisy, and thereby reduce the fidelity with which the CTF amplitude modulations may be restored [27].

A more attractive approach is to correct the phases on the projections, and then to address the amplitudes after reconstructing the 3D map. One such approach uses a Wiener-like filter which takes an estimate of the SSNR(q) as an integral part of the statistical model underlying the adaptation of RELION for sub-tomogram averaging [40]. The 3D sampling model used in this case is also responsible for the phase correction of the individual particles, which I suggest is a poor choice given the high defocus (2-4μm) used in tomography as it inevitably leads to significant aliasing of the CTF unless images too large to be practical are used [80].

2.2.5.1 CTF phase correction

Rather than extracting and correcting the phases on tiles, I compute the forward Fourier transform once for each projection and then multiply by the CTF for a given defocus, inverse transform and cut out the valid region. This helps to prevent aliasing as the projection images are typically 3-5K pixels square

2.2.5.2 CTF amplitude correction

I have adapted a version of the “volume normalized Single-particle Wiener Filter,” and as our filter is necessarily a post-reconstruction filter, I start with equation eight from the original paper [81]. Note that equation 2.4 assumes the ad-hoc Wiener constant in equation 2.5 to be negligible.

$$F^{SPW_t}(\boldsymbol{q}_{hkl}) = \frac{SF^{3D}}{SF^{3D} + \frac{f_{particle}}{f_{mask}}\left(\frac{1 - FSC_{mask}(\boldsymbol{q}_{hkl})}{2FSC_{mask}(\boldsymbol{q}_{hkl})}\right)\left(\frac{1}{n_q}\sum_{q \in \boldsymbol{q}_{h'k'l'}} SF^{3D}\right)} F^{LSQ}(\boldsymbol{q}_{hkl}) \qquad eq\ 2.4$$

The least squares estimate, which is a Wiener filtered reconstruction with an ad-hoc Wiener constant is defined below in Eq 2.5. I have made three major changes to the filter:

$$F^{LSQ}(\boldsymbol{q}_{hkl}) = \frac{\sum_{j=1}^{S}\sum_{i=1}^{Z} T^{i,j} CTF_i^{2d} R^{2d} ExpFilter_i^{2d} F_i^{2d}}{\sum_{j=1}^{S}\sum_{i=1}^{Z} T^{i,j} \left|CTF_i^{2d}\right|^2 R^{2d} ExpFilter_i^{2d} F_i^{2d} \quad + \quad 1} \qquad eq\ 2.5$$

First, the 3D-SF is weighted for critically under-sampled regions, where the SSNR estimated by the FSC is less reliable. This is done by choosing a minimum acceptable sampling threshold, 0.2 * median(SF^{3D} > 0), and then scaling SF^{3D} to replace less sampled regions by smoothly transition from this value to some new larger number chosen by the maximum in the original SF^{3D}.

Second, the spherical shells normally used in the FSC calculation are further subdivided into conical sections, which captures directionally anisotropic SSNR [82]. I currently use cones incremented over 30˚ with a half-angle of 36˚, such that they overlap substantially, producing a smooth estimate of the directional resolution. The impact of the amount of overlap and the sampling increment is not well characterized and should be investigated in future work.

Third, the average sampling over spherical shells (final term in the denominator of eq 2.4) that is used to scale the SSNR estimate to represent the average SSNR in a single

sub-tomogram is replaced with a Gaussian smoothed version of the 3D-sampling function. Again, to account for anisotropy in the sampling.

$$F^{SPW_t}(\boldsymbol{q}_{hkl}) = \frac{1}{|SF^{3D}|^2 + \frac{f_{particle}}{f_{mask}}\left(\frac{1 - FSC_{aniso,mask}(\boldsymbol{q}_{hkl})}{2FSC_{aniso,mask}(\boldsymbol{q}_{hkl})}\right)(G \otimes |SF^{3D}|^2)} F^{LSQ}(\boldsymbol{q}_{hkl}) \qquad eq\ 2.6$$

2.2.6 3D-SF-compensated classification

I use multivariate statistical analysis (MSA) for classification. This involves first defining a set of descriptors or *features* which are then searched for patterns common to a sub-group called a pattern-class, or simply a *class*. Grouping inputs by these patterns is accomplished with a clustering algorithm; commonly k-means, hierarchical ascendant classification (HAC) or a neural networks approach via a self-organizing map (nn-SOM), all three are implemented in emClarity via MATLAB's statistics and machine learning toolbox, incorporated in the (free) MATLAB compiled runtime. Since the "missing wedge" produces significant artifacts that are specific to the orientation of each particle in the sample, but not necessarily to its identity or conformation, it has been challenging to resolve meaningful patterns in cryoSTAC data. Estimating the effect of the missing-wedge by using a binary mask has shown a good first order correction called wedge masked differences (WMDs) [83]. I replace this mask with our 3D-SF, which results in a more accurate estimate of the artifacts introduced by the MWE and allows higher resolution information to be considered in the clustering. The creation of these feature vectors is outlined in **Figure 2.4.**

Initialize D_i matrices in main memory, with M X N pixels, where d is the number of resolution bands, M is the number of pixels in the real space mask, N is the number of sub-tomograms in the data-set (or random-subset)
FOR each resolution band
FOR each sub-tomogram

← Rotate sub-tomogram into microscope reference frame, for even half-set include the transformation found in FSC calculation at the outset of the cycle.
← Rotate the 3D-sampling function by the particles orientation matrix and apply to the Fourier transform of the global average.
← Form the difference of the masked average and the rotated particle
← Place the values under the mask into the storage matrix

End Loop
End Loop

← Center each row, by subtracting the mean. (This is the mean of all sub-Tomograms at this particular voxel)
← Determine the partial Singular value decomposition for the selected number of coefficients (usually ~ 30) $U_i\, S_i\, V^T{}_i = \; D_i$
← IF running on a partial data set, save each matrix U_i
- Load full data set,
  - FOR each resolution band
  - FOR each sub-tomogram
    - Rotate sub-tomogram into microscope reference frame, for even half-set include the transformation found in FSC calculation at the outset of the cycle.
    - Project the data onto the highest variance directions $U_i{}^T D_i$ = P_i
    - Concatenate each of these P_i matrices along the rows

← ELSE
← Concatenate each of the $S_i\, V^T{}_i$ matrices along the rows, such that each column is a feature vector with projections from each resolution band along its rows

Figure 2.4 Formation of feature vectors

2.2.7 Multi-scale clustering

In a naïve approach, a clustering algorithm will interpret every individual voxel as an independent measurement along one dimension of an N-dimensional space, where N is the number of voxels in each sub-tomogram at the time of analysis. To reduce the impact of the "missing wedge" and to introduce a correlation between pixels, a smoothing filter is typically applied to the data before clustering. In effect, this tells the clustering algorithm that neighboring pixels are measurements of related features in the full sample space. Because our 3D-SF WMD feature vectors are robust to the "missing wedge" at higher resolutions, this idea may be used to introduce inter-voxel correlations at biologically relevant length scales. I do this by forming multiple feature vectors for each sub-tomogram, each being differentially bandpass-filtered, for example, ~ 10 Å to emphasize alpha-helical density, 18-20 Å for RNA helices or small protein domains, and ~ 40 Å for larger protein domains. This approach is similar to existing ideas that use the discrete wavelet transform with a limited set of coefficients, followed by clustering of the data reconstructed ***independently*** using a limited subset of wavelet basis [84]. In our case, each Gaussian kernel can be viewed as a simple wavelet, localized at the origin and defined in frequency by the biological length-scales mentioned above. The primary difference is that the coefficients are concatenated into a single matrix, allowing consideration of each length-scale ***simultaneously***, providing a more comprehensive description of the feature space. In effect, this teaches the clustering algorithm to "see the forest for the trees."

2.3 RESULTS

2.3.1 emClarity improves resolution in sub-tomogram averaging

Given the inherent difficulty in working with extremely low SNR cryo-EM data, and the sensitivity of the results to optimal selection of parameters[14], I have elected to test and demonstrate our software using two publicly available data sets from the Electron Microscopy Pilot Image Archive [85] (EMPIAR). I show these published/ deposited maps, juxtaposed with the maps obtained with emClarity in **Figure 2.5**. A total improvement in the yeast 80S ribosome from EMPIAR-10045 using RELION version 1.4 (EMD-3228 [43]) from 12.9Å to 7Å is achieved using emClarity (**Figure 2.5** A). For the mammalian 80S ribosome from EMPIAR-10064 using pyTOM (EMD-3420 [51]), I obtained an improvement from 11.2Å to 8.6Å (**Figure 2.5** B)[15].

To evaluate the relative impact of each of the individual features implemented in emClarity, I incrementally included them into several reconstructions of the yeast 80S ribosome. To control for errors in alignment and to have a one-to-one comparison with EMD-3228, I used precisely the same particles and orientation parameters from the star files that accompany the raw data EMPIAR-10045. I compare each map to an external reference map derived from SPA (EMD-2275 [86]), via a cross-Fourier Shell Correlation

[14] It is worth noting that the authors for the maps we use for comparison are also authors on the primary publications for their respective software packages, which helps to ensure the resolutions reported are likely optimal for the given data.
[15] This resolution is likely an underestimate. During revision, bugs were found that improved the yeast 80S resolution from 8.2 Å to 7Å, but due to time restraints, the mammalian 80S data could not be re-processed.

(cross-FSC), starting from the RELION reconstruction as a control (**Figure 2.5** C). The accuracy of our combined CTF correction approach, phases on oversampled 2D-tiles combined with optimal-exposure filtering and 3D-CTF based Wiener filtering is reflected in the magenta curve in **Figure 2.5** C, which shows a significant improvement over the cross-FSC of the control, even though they are reconstructions using the same particles and orientations. The most substantial improvement comes from the tomo-CPR which is shown in green (and obviously includes the features in the magenta curve as well.) A more modest improvement is measured when I add in a per-tilt defocus estimation using our novel approach to resample periodograms from tilted images, as reflected in the cyan cross-FSC.

In addition to improved resolution, as noted in **Figure 2.5** B, there is a density (possibly sec-61) outside the peptide-exit tunnel of the ribosome (white arrow) that is present in the map derived with emClarity, but not in the map derived with pyTOM. Finally, in **Figure 2.5** D, I show the density from a peripheral region with a rigidly docked model of the yeast 80S ribosome (PDB-4V7R) that underscores the difference in interpretability between the maps derived from the current state-of-the-art and emClarity. There is a definite improvement in both RNA and protein structures.

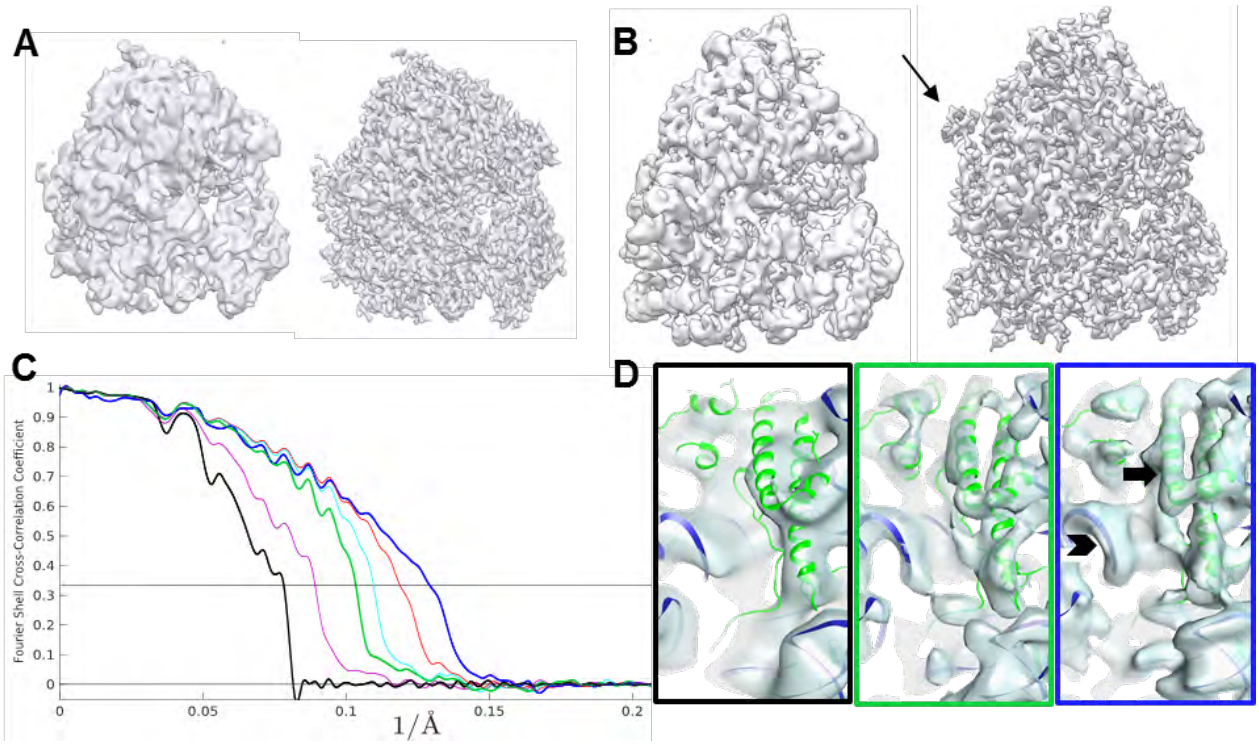

Figure 2.5 Feature-wise improvement in resolution

A) Comparison of the sub-tomogram average of yeast 80S ribosome by RELION (EMD-3228) (left, at 12.9 Å resolution) and by emClarity (right, at 7.0 Å resolution), using the same raw data from the electron microscopy public image archive (3,233 Yeast 80S ribosomes, EMPIAR-10045). (B) Comparison of sub-tomogram averages of rabbit 80S ribosome by pyTOM (EMD-3420) (left, at 11.2 Å resolution) and emClarity (right, at 8.2 Å resolution), using the same raw data (1,400 Rabbit 80S ribosomes, EMPIAR-10064). Arrow points to an additional feature outside peptide exit tunnel only revealed with the more conservative masking procedure in emClarity. (c) Cross-FSC between the sub-tomogram averages by emClarity and the SPR cryo-EM map (EMD-2275) of yeast 80S ribosome. The first five curves use orientation parameters from the Relion 1.4 alignment, cumulatively including additional features from emClarity: magenta, improved CTF estimation and correction with the optimal exposure filter; green, one round of tomo-CPR; cyan, adding in the per-tilt defocus estimation; red, adding in explicit consideration of resolution anisotropy in the adapted single particle wiener filter. The final dark blue curve incorporates all these features plus the alignment parameters determined from scratch in emClarity. Representative views of sub-tomogram averages, with a rigid body docking of the yeast 80S atomic model (PDB-47VR) for visualization. Arrow and chevron highlight the resolved alpha helices and RNA structures, respectively.

The yeast 80S sample that was used has a strong preferential orientation which is reflected in the plot of the FSC as calculated over conical sections **Figure 2.6 A,** and a plot of the angular distribution in **Figure 2.6** B, C**.** When the anisotropy in the SSNR that results from this preferred orientation is included in our adaptation of the single particle Wiener filter, another substantial improvement in the cross-FSC is made, shown in the red curve. The final and highest resolution curve represents an alignment carried out in emClarity with all features added, illustrating the additional impact these advances have on the accuracy in the orientation determination.

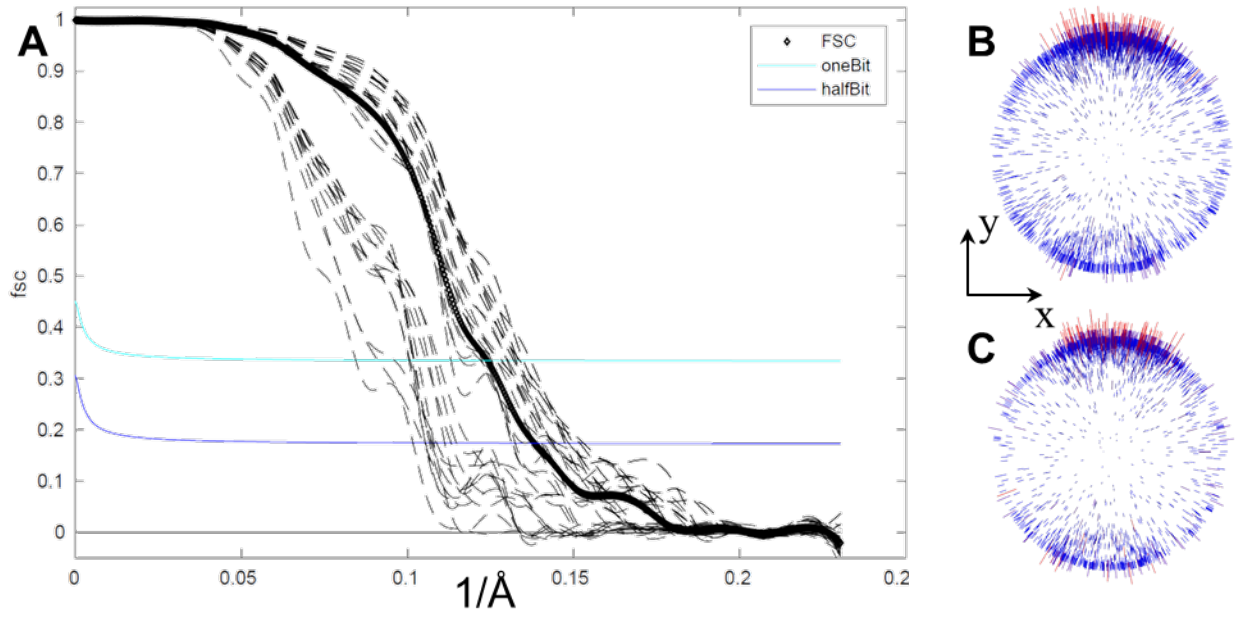

Figure 2.6 Measuring anisotropic resolution

(A) the plot of the FSC calculated over 38 conical shells (dashed lines) which range from 6.6 Å - 9.6 Å at the 0.143 cutoff, while the normal spherical shells (bold line) show the average directional resolution 7.0 Å. (B) Angular distribution from the alignment in emClarity shows a very similar distribution to that published for the yeast 80S EMD-3228.

2.3.2 Classification in emClarity reveals multiple functional states

Using multi-scale clustering combined with 3D-SF compensated Principal Component Analysis (PCA), emClarity reveals subtle conformational differences and distinguishes minor populations from noisy and distorted images, as demonstrated with yeast 80S ribosome data from EMPIAR-10045, and mammalian 80S ribosome data from EMPIAR-10064. Such results were not previously obtainable using existing software [43,51].

2.3.2.1 Classification of non-translating Yeast 80S ribosomes

The ribosome is a complex molecular machine composed of RNA and protein which exists in many functional states and interacts with an array of co-factors. The major domains are named by their sedimentation coefficients (S, Svedberg) where the eukaryotic ribosome is composed of two major domains dubbed the large subunit (60S) and small subunit (40S). While the ribosome has a well-conserved catalytic core which mediates the peptidyl transferase reaction [87], it is increasingly subject to more complex regulation in higher organisms resulting in an expanded set of both RNA and protein components. RNA expansion segments are found primarily at the periphery of the ribosome and are typically highly dynamic and challenging to resolve in structural analysis. One particularly good example is es27, an approximately 150 Å RNA helix which predominantly adopts one of two conformations separated by about 90^0, shown in orange in **Figure 2.7**. The first situates the end of the RNA helix just outside the peptide exit tunnel on the 60S subunit ($es27_{pet}$, **Figure 2.7** A, B, D, E) and the second points toward the tRNA exit site ($es27_{L1}$, **Figure 2.7** C). This dynamic domain is generally observed in

cryo-EM maps as a superposition of these two states, as is the case with the currently published results by Maximum likelihood (ML) classification in RELION [43]. A notable exception being ribosomes with accessory complexes bound at the peptide exit tunnel, e.g., Sec61, are known to bias the conformation to the $es27_{L1}$ [88].

Another example of a highly dynamic ribosome domain is the L1 stalk – comprised of protein L1, and RNA helices h75, h76 and h79 from the 25s portion of the 60S subunit [89]. The motions of L1 are well correlated with several defined functional translocational states as observed using single-molecule FRET and SPA [90]. Using emClarity, three oL1 conformational states are discerned as isolated from the thermal (stochastic) fluctuations of the non-translating yeast 80S ribosome: $L1_{open,}$ $L1_{int,}$ and $L1_{closed}$ shown in green with variable occupancy in the five classes in **Figure 2.7**. In addition to isolating dynamic states, identifying very sparsely populated classes is a particularly important and challenging task for classification of cryo-EM data. In **Figure 2.7** E the dissociated 60S subunit occupying a minor class, only ~4% of the data set or roughly ~140 sub-tomograms. In contrast, the ML approach implemented in RELION found three classes, one designated as a junk class and two relatively indistinguishable classes [43]. This minor population could only be isolated in the case where feature vectors built from the projection on the principal components from at least three length-scales were simultaneously clustered.

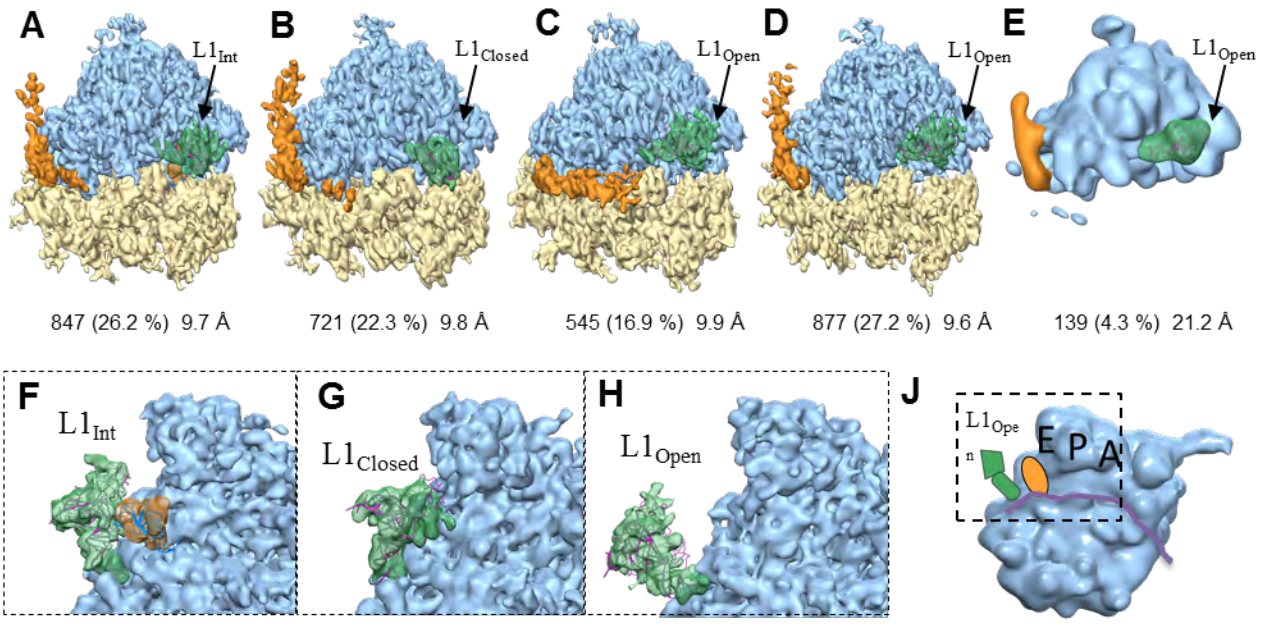

Figure 2.7 Classification of stochastically fluctuating yeast 80S domains

Classification of yeast 80S ribosome (EMPIAR-10045) with full 3dCTF compensated missing-Wedge and multi-scale PCA in emClarity. (A-E) Four major classes and a minor class contributing 96.9% of sub-tomograms are shown with number and percentage of contributing units and resolution indicated below. The remaining 3.1% comprised a 6th minor class with no significant structural features which were removed from the analysis. The highly dynamic L1 protuberance (green) and RNA expansion-segment 27 (orange) are captured in distinct conformations in these classes. Lower row, enlarged views of the L1 protuberance in an (F) intermediate position bound to P/E tRNA shown in orange (class a), (G) fully closed interacting with the 60S central protuberance (class b), and (H) fully open (class c,d,e) respectively. L1 protuberance from PDB 3J78 colored magenta (rpL1,h76,h79) docked for visual aid.

2.3.3 Improved estimation of MWE shown in 3D variance maps

Regions of significant variance across a data set may be visualized by overlaying a 3D "variance map" with the average structure. The "missing wedge" produces significant artifacts that are specific to the orientation of each particle in the sample, but not necessarily its identity or conformation. Left uncorrected these artifacts obscure meaningful differences among particles, reflected in a diffuse variance across the dataset which can be seen in **Figure 2.8** H-J. A previously demonstrated technique for estimating the effect of the "missing wedge" by using a binary mask, called "wedge masked differences (WMDs)", was shown to be a good first-order correction [83]; however, the accuracy of this model breaks down when higher-resolution features are considered (**Figure 2.8** D-F). To allow higher-resolution information in the classification, I replace this binary wedge mask with our 3D-SF, resulting in a more accurate estimate of the artifacts introduced by the "missing wedge" as shown in **Figure 2.8** A-C. It is worth noting that this does not "fill in" any missing data. Instead it estimates what a given particle should look like by distorting the current sub-tomogram average by that particle's 3D-SF, and clusters based on the difference between this expected value and the observed particle.

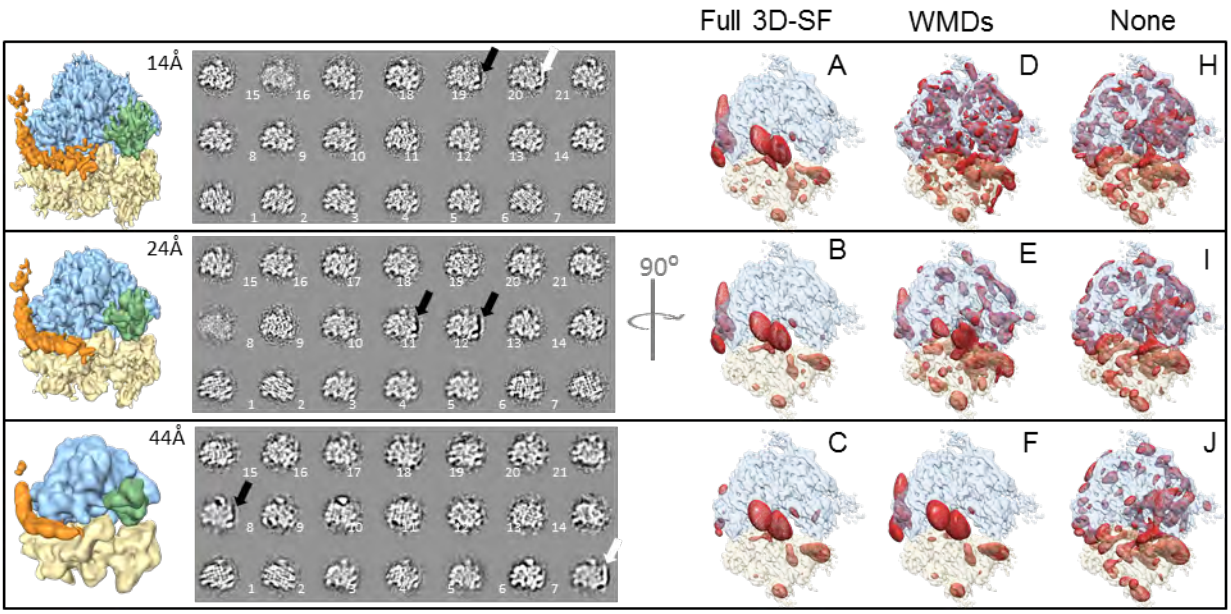

Figure 2.8 missing wedge compensated 3D variance maps improve significantly with accurate Fourier sampling consideration via 3D-SF

Illustration of the impact and compensation of the missing-wedge at multiple length scales. Far left column, the total average filtered by Gaussian kernels of variable width to correlate voxels over the given length scales. Center column, eigen-images composed of the average plus the eigenvector, sorted from the most variance explained (1) to least (21). Black and white arrows show an example where es27 density is absent or present in the L1 position. Right three columns show 3D-variance maps in red over-laid with the average as a visual guide. (A-C) Variance is concentrated on L1 (green arrow), es27 (orange arrow), E-site tRNA, and the mRNA channel entrance (blue arrow) at all resolutions. (D-F) Regular WMDs are dominated by noise at all but the lowest resolution. (H-J) Negative control shows no meaningful concentration of the variance due to severe missing wedge bias. Black scale bars represent 100 Å. White scale bars represent 300 Å.

2.3.3.1 Mammalian 80S ribosome

In contrast to the non-translating yeast specimen, the mammalian ribosomes imaged in EMPIAR-10064 were prepared from clarified rabbit reticulocyte lysate using a buffer low in Mg^{2+} but lacking polyamines, such that cofactors should co-purify excepting perhaps some loss of E-site tRNA [91]. I extracted 3,090 ribosomes from the four tilt-series deposited as the "mixed-CTEM" data set on EMPIAR. emClarity identified five predominant classes as shown in **Figure 2.9** A-E. Three of these classes show ribosomes adopting a non-rotated 40S conformation with variable tRNA, eEF1A occupancy (class I-III), while two very similar classes adopted a mid-rotated (~5-6^0) 40S conformation with eEF2 present (class IV-V).

A rigid body docking of the full 80S mammalian ribosome in the non-rotated POST state from PDB-4UJE [92] showed very clear agreement with the conformation of the 40S subunit, which combined with the co-factors observed suggest classes II and III are POST trans-locational ribosomes differing in retention of E site tRNA while class I is most similar to the "sampling" state. Classes IV and V both have eEF2 bound and differ in rotation of the 40S subunit of 5.9^0 and 5.0^0, respectively. A rigid body docking of the yeast 80S structure of eEF2 from PDB-4UJO [93] into classes IV and V show overall good agreement with their eEF2•sordarin•GDP position and our density. There are, however, noticeable differences, particularly in domain IV of eEF2 which is known to be dynamic and plays a key role in translocation [90,94]. I analyzed these differences qualitatively by comparing the rigidly docked model solved with Sordarin present (**Figure 2.9** J-K) with the same model after running a short (1ns) Molecular Dynamics Flexible Fitting (MDFF)

(**Figure 2.9** H-I). The antibiotic Sordarin is highly specific for binding to fungal eEF2 and permits GTP hydrolysis, yet prevents conformational changes that result in the subsequent release of eEF2 after translocation [95]. Although Sordarin is not present in the sample under study here, there is a pronounced difference in electron density between domains III-V of eEF2 in class V (**Figure 2.9** I black arrow) that coincides with the Sordarin binding pocket. This density is not present in class IV which also exhibits a rotation of eEF2-domain IV (**Figure 2.9** J).

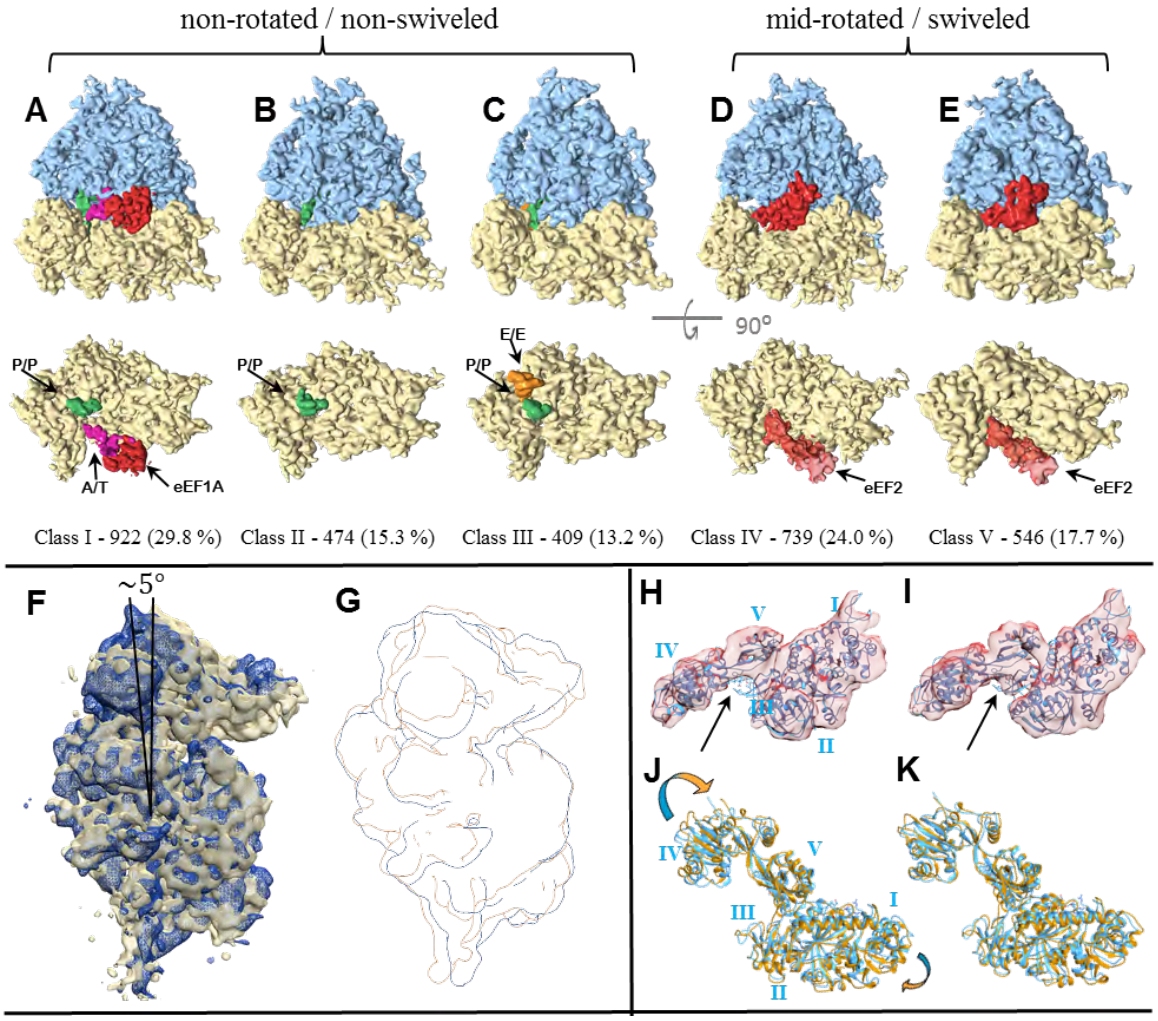

Figure 2.9 Classification using multi-scale PCA with fully compensated CTF estimators reveal five distinct translocational species from only 3,090 particles.

Multi-scale Clustering with 3D-SF compensated feature vectors reveals five distinct translocational classes from 3,090 particles. (A-C) Classes I-III represent a post-translocational state with the co-factors shown in the lower row from the inter-subunit surface with the 60S subunit removed for clarity. (D-E) Classes IV-V have a mid-rotated 40S and a swiveled head corresponding to a pre-translocational intermediate. (F) Class IV is shown in the dark blue overlaid with class III in gold, showing the mid-rotated 40S state. (G) Outline from a low-pass filtered overlay as in (F). (H) MDFF of eEF2 (orange) with the density from class-IV starting from PDB-4ujo (cyan ribbon) shows similar conformation in eEF2 domains II, III, and V, while eEF2

domain IV deviates the most. (I) Same as (H), but with class V showing smaller deviations. (J-K) The rigid body docking of PDB-4ujo into the density from class IV/V respectively shows overall close agreement, except the stronger density between eEF2 domains III & V in (K). Arrows point to this density which is occupied by the antibiotic Sordarin in PDB-4ujo but is not present in the sample used in this study.

2.4 DISCUSSION

With the rapid expansion of cryo-EM resources available at major universities and with the development of shared use models like eBIC at Diamond Light Source, the ability to collect high-quality cryo-EM and cryo-ET data is now arguably less limiting than the ability to effectively process the data. I have created a set of image processing routines incorporated into the program emClarity, which have demonstrated much greater accuracy in alignment and image restoration compared to current state-of-the-art approaches as demonstrated by using the same raw data sets which are publicly available.

While the improvement in resolution can be analyzed feature by feature, showing the largest impact is from the refinement of the tilt-series alignment using tomo-CPR, taken together they are greater than the sum of their parts. The per-tilt defocus determination and adaptation of the single particle wiener filter are robust and need little adjustment from default parameters.

The parameter space for tomo-CPR is less well explored: the weighting of the background tomogram relative to the mapped back sub-tomogram is one such parameter which the user may need to adjust if unsuitably large image shifts are found in the solution from *tiltalign*.

Our approach for image classification in the presence of the “missing-wedge” effect by combining the correction for wedge differences with multi-scale clustering which helps to encode biologically relevant information for the clustering algorithms gives promising results, and some clear improvements would further advance its utility. While the goal is to anticipate and remove MWE bias in the analysis, the approach does so by concentrating the variance due to the MWE in a few eigenvectors. Currently, the user needs to recognize these and remove them from consideration as feature vectors used in clustering. While this requires some prior knowledge, the “streaky” appearance has been consistent across a wide variety of specimens analyzed. Future work may include training a simple convolutional neural network to recognize these streaky features and then distributing this pre-trained CNN with emClarity to suggest to the user which eigenvectors to consider for subsequent processing.

The application of these advances to study samples in an environment with relaxed biochemical restraints *ex vivo* shows promise, having isolated functional intermediates of translocation from a cell lysate. Looking at classes IV and V of the mammalian ribosome suggests that the binding of the antifungal Sordarin, which stabilizes an interaction between eEF2-domain III/V, is re-enforcing an on-pathway interaction that exists in functional ribosomes. This also hints that nearby intermediates on the energy landscape

may be found by improving the statistics available via a larger sample. In addition to isolating well-resolved class averages with minor populations, and finding nearby minima in the energy landscape, our approach also results in the production of accurate 3D-variance maps which will be beneficial to exploring macromolecular dynamics.

By highlighting key regions of dynamic behavior, our approach should be useful for direct analysis and the design of complementary biophysical experiments. While these advances in classification are in the pre-processing and dimensionality reduction stage, future work to explore modern approaches in pattern recognition and machine learning will likely establish another substantial improvement in the technique.

Ultimately, I hope that emClarity will advance the study of structural biology *in situ*, as methods used to thin cellular samples, particularly cryo-FIB milling, also remove gold-fiducial markers, making the alignment of the tilt-series a major limiting factor.

2.5 CONCLUSION

I have developed emClarity, an image processing package for GPU-accelerated high resolution cryoSTAC. The programs run on Nvidia graphics cards with > 11Gb memory and benefit from fast disk storage as do most other cryo-EM software. To demonstrate the improvements possible with emClarity, I have shown maps at substantially improved resolutions compared to those obtained in the original studies, where the authors are experts in their respective programs. I also reveal previously obscured conformational

sub-populations from two publicly available data sets. With the release of the HIV-1 data from Schur *et al.* I have demonstrated that emClarity can achieve the highest-resolution maps from cryoSTAC to date, reaching 3.1 Å.

2.6 SOFTWARE AND INSTRUCTIONAL MATERIAL

Detailed methods aimed at reproducibility for the results in this chapter are included in the Appendix.

The software is freely available from https://www.github.com/bHimes/emClarity

Tutorial documentation and videos at https://www.github.com/bHimes/emClarity/wiki

2.7 ACKNOWLEDGMENTS

I thank Dr. Joachim Frank and Dr. Wen Li for very helpful discussions. Mr. Doug Bevan for technical assistance with computer clusters, and Drs. Teresa Brosenitsch, Frances Joan-Alvarez, and J. Peter Rickgauer for reading the manuscript. I thank Dr. Xiaofeng Fu for testing the emClarity software. I also thank Dr. Niko Grigorieff for material support and patience while this work was finalized, and Dr. Srini Turaga for kindly sharing his GPU cluster, allowing for a broader search of parameter space. This work was supported by the National Institutes of Health (GM085043, GM082251) and the UK Wellcome Trust Investigator Award (206422/Z/17/Z) to Dr. Peijun Zhang.

3.0 FUNCTIONALLY DYNAMIC QUATERNARY STRUCTURE OF BACTERIAL CHEMOTAXIS SIGNALING ARRAYS DETERMINED BY CRYOSTAC

Bacteria move in response to their local environment by switching their flagellar motors between clockwise and counterclockwise rotation, producing a biased random walk. This switching is controlled by a network of transmembrane signaling receptors, enzymes, and adaptor proteins, which form extended and dynamic arrays in the inner membrane of *E. coli*. How the organization of these arrays enables the cell to simultaneously integrate signals from multiple ligands, while also maintaining sensitivity over large ligand concentration gradients is not well understood. We determined the structure of the core-signaling unit [96] (CSU) of the *E. coli* chemotaxis receptor signaling array at 11 Å resolution using cryo-Electron Tomography with sub-tomogram averaging and classification. We built a pseudo-atomic model of the core-signaling unit using the high-resolution crystal structures. Extended molecular dynamics simulations of the 64 million atom unit-cell, revealed a conformational switch in the histidine kinase CheA related to signaling. My contribution to this work was the development of a new computational approach to address specimen heterogeneity and preferred orientation and deriving a map of a trimer of CSU at 11 Å resolution.

3.1 INTRODUCTION

Bacteria have evolved to inhabit practically every environmental extreme to be found, from the acidic lumen of the human stomach to the deepest ocean trenches [97]. Changes in nutrient availability, pH, temperature, osmolarity, and many other conditions require bacteria to constantly monitor external conditions and adjust their structure, physiology, and behavior [98]. Chemotaxis is the movement of an organism toward or away from a chemical signal [99]. In bacteria, the locomotive force necessary for chemotaxis is generated by the rotation of one or more filamentous protein structures called flagella [100]. The flagella are under the control of a two-component signaling regulatory system; a transmembrane receptor/ histidine kinase complex serves as the sensor which phosphorylates the diffusible response regulator CheY (**Figure 3.1** A). The balance of phosphorylated and apo-CheY determines either a clockwise or counter-clockwise rotation that creates a pattern of tumbling or smooth swimming respectively (**Figure 3.1** B). The balance of these two states results in a biased random walk [101] along with a concentration gradient (**Figure 3.1** C).

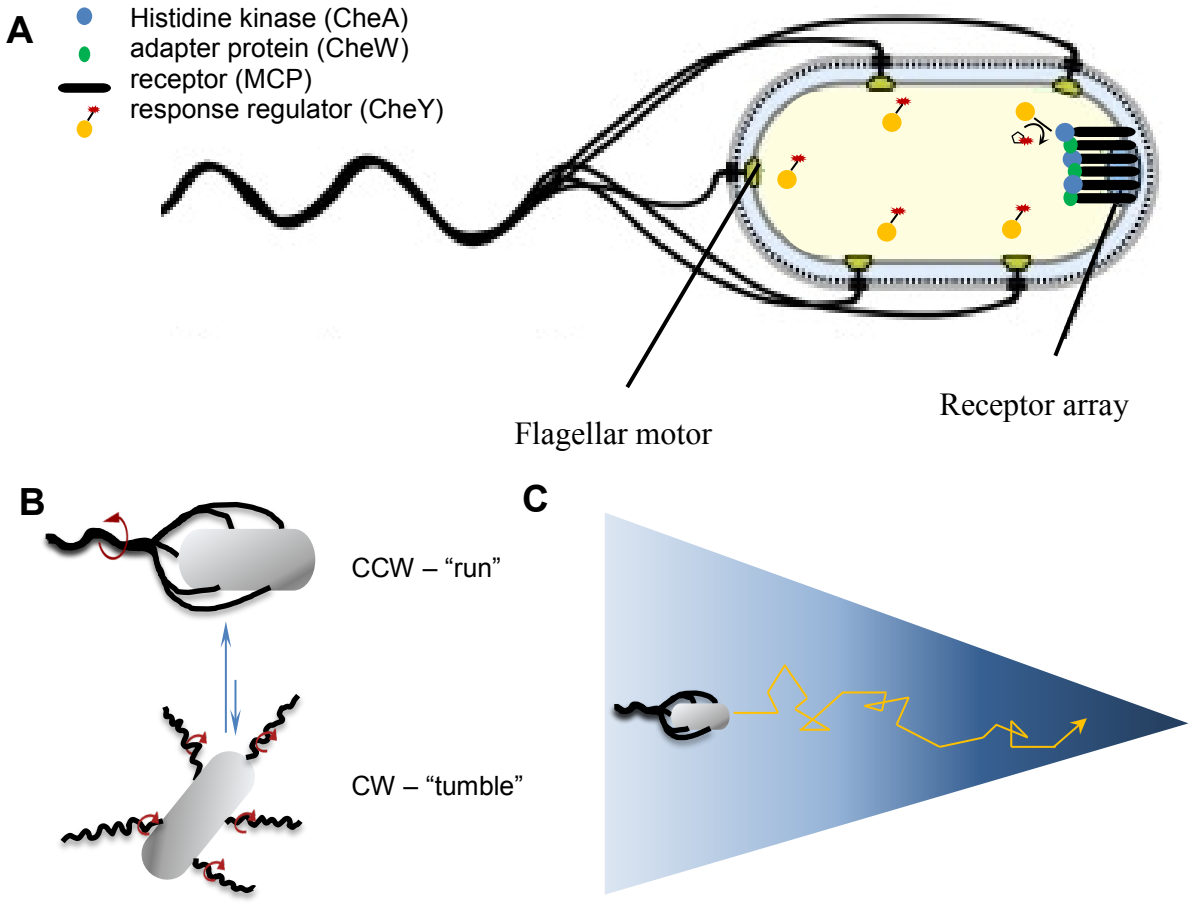

Figure 3.1 Essential components of motility

(A) The essential components needed to generate a CCW signal overcoming the motors "default" CW bias. (B) The direction of flagellar rotation determines whether a coherent bundle is formed causing the cell to either run forward or not resulting in "tumbling" which serves to randomize the orientation of the cell and the direction of the subsequent run. (C) Chemotaxis results from a controlled switching between these two states culminating in a biased random walk up an attractant gradient or down a repellant gradient.

Early experiments demonstrated the robust nature of the chemotactic behavior in *E. coli*, maintaining responsiveness to changes in ligand concentration over a range of 6 orders of magnitude. For example, the detection threshold and saturating concentrations for Aspartate are 30 x 10^{-9} M and 100 x 10^{-3} M respectively [102,103].

The cell stores a record of recent environmental conditions in a pattern of methylated glutamate residues on the transmembrane receptors name methyl-accepting chemotaxis proteins (MCPs) (black and white dots, **Figure 3.2**). When a positive stimulus is detected (binding attractant) the receptor bound kinase is turned OFF, biasing the cell toward a run (**Figure 3.2** A) by reducing the pool of phospho-CheY. Reducing the kinase activity of CheA also reduces the activity of the methylesterase CheB allowing the constitutively active methyltransferase CheR to "catch-up" which slowly returns the receptor to the ON state [104,105]. It is the balance of receptors biased ON or OFF that determine the actual output signal, but how this balance is affected by the quaternary organization of the signaling arrays is not understood.

A major limitation in interpreting the biochemical and genetic information available is the lack of a detailed structural model of the assembled array. The primary goal of this study is to derive the cryoEM structure of the signaling array, upon which assemble a pseudo-atomic model of the array system using known high-resolution crystal structures of the array components (**Figure 3.2** B).

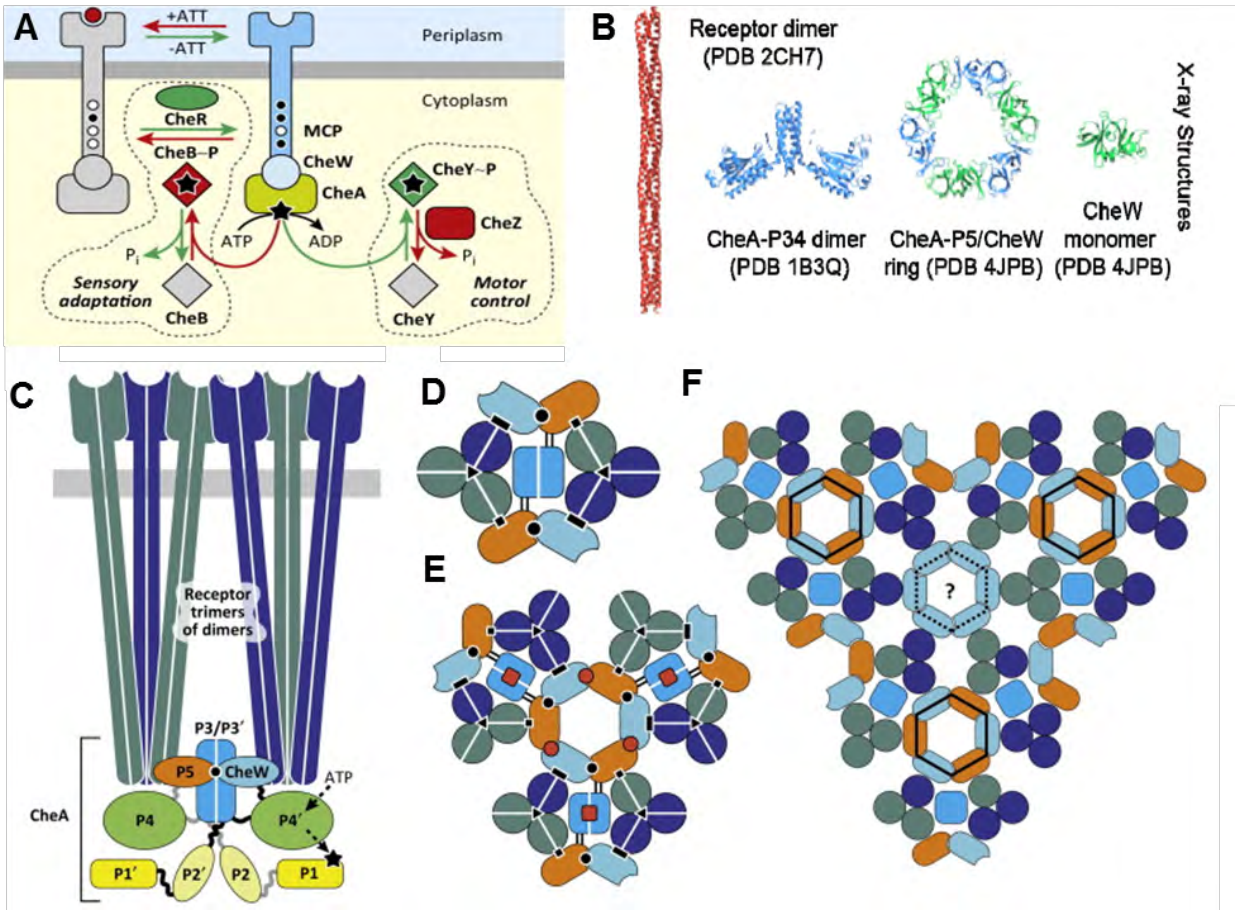

Figure 3.2 Overview of chemotaxis signaling and core-unit structure

(A) Cartoon representing the simplest complete signaling pathway from a positive or negative stimulus. (B) Component crystal structures are available for modeling the full chemotaxis receptor core signaling unit. (C) Side view of the CSU which is a dimer of trimers of receptor homo-dimers, the homodimers within each trimer may respond to different ligands. Each domain of the five CheA domains is labeled and placed in the approximate location expected in the CSU. (D) Looking down through the membrane, view of the CSU. (E) A trimer of CSUs and (F) along with hexamer of CSUs with putative 6-fold CheW ring shown with the question mark. Except for labels, panel A, B-F reproduced with permission from [106], and B adapted from [30].

The minimal assembly needed for signaling is a dimer of trimers of MCP dimers, along with a dimer of CheA and two copies of the adaptor protein CheW [107,108], referred to as the core-signaling-unit or CSU (**Figure 3.2** C, D). The CSUs are thought to assemble into trimers which fill out into a pseudo p6 lattice [109,110] **Figure 3.2** E-F.

The organization of the structural domains of CheA varies in the published crystal structures and is at odds with the available crosslinking, biochemical, and ESR data [111–113]. Defining the positions of these domains, particularly the CheA-P3 dimerization domain, CheA-P5/Receptor interaction (CheA-P5 is structurally homologous to CheW) and the CheA-P4 Kinase domain is of particular interest. To better understand how the activity of CheA is regulated by the MCPs we have developed an *in vitro* reconstitution system that recapitulates the signaling arrays found in cells, with a minimal functional set of well-defined components.

The reconstitution of heterologously expressed proteins also permits us to manipulate the signaling state represented by the *in vitro* arrays. By knocking out the CheR and CheB genes in the expression host, the methylation dependent function of the receptors may be controlled and studied *in vivo* and *in vitro* by using genetic mutation of glutamate to glutamine at these key residues to produce an analog for the methylated glutamyl residue [114]. We can then look to the WT pattern as QEQE, a mix of ON/OFF signaling, EEEE biased strongly OFF, and QQQQ biased strongly ON [115].

While this *in vitro* system provides a well-defined specimen with abundant signal for structure determination by cryoSTAC, it also presents several challenges for existing image processing algorithms, including severely anisotropic resolution resulting from the

preferred orientation imposed by the lipid monolayer support, as well as substantial heterogeneity due to the conformational plasticity of the core signaling complex. Once these hurdles have been addressed, it would be desirable to then look at the complexes in native cell membranes.

3.2 EXPERIMENTAL PROCEDURES

A broad overview of experimental procedures supporting the structural determination of the CSUs is provided below. For details on the lipid-monolayer array reconstitution, carried out by Dr. Jun Ma; tomographic data collection, carried out by Dr. Peijun Zhang and Dr. Gongpu Zhao; biochemical analysis of the key amino acid residues suggested by the model, designed and carried out by Dr. Frances Alvarez; and aspects of the computational modelling carried out by Dr. C. Keith Cassidy the reader is referred to our publication in eLife [30].

3.2.1 Protein expression and purification

Plasmids and cell strains used in the study were gifts from Dr. Sandy Parkinson, University of Utah, and Dr. Bob Weis, University of Massachusetts Amherst. Component proteins were expressed in *E. coli* strain RP3098 which is a null mutant for all Che proteins as well as the chemoreceptors. Tar, CheA, and CheW were expressed from plasmids induced with IPTG. Expression and purification conditions varied slightly, but all proteins produced good (~6mg/Liter culture) yield and purified to high homogeneity.

3.2.2 Signaling array reconstitution

A lipid monolayer system containing Ni^{2+}-NTA lipids was used to reconstitute the chemotaxis arrays. A mixture of 9:18:18 mM of TarCF:CheA:CheW in a buffer containing 75 mM Tris- HCl, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$ was applied to a Teflon well, over which a lipid monolayer containing 2:1 DOPC:DOGS-NTA-Ni^{2+} lipid mixture, at 2 mg/ml concentration was gently overlaid and incubated overnight to allow the signaling arrays to assemble. Following assembly, a holey carbon coated electron microscope specimen grid was laid over the monolayer and gently peeled back, transferring the monolayer supported array to the specimen grid.

3.2.3 Cryo-Electron Tomography

Because the monolayer is delicate, it required special care in blotting from just the non-sample side of the grid before being plunge-frozen in liquid ethane. The grids were pre-coated with gold beads before picking up the monolayer sample. Data were imaged at 200 KeV in a Tecnai Polara microscope (FEI, OR) at a nominal magnification of 39,000x giving a nominal pixel size of 3 Å/pixel as recorded on a Gatan 4k x 4k CCD detector. The total accumulated dose was 60 $e^-/Å^2$ collected using a bi-directional tilt-scheme starting at 0˚ proceeding in 3˚ increments to +70˚, returned to 0˚ in -3˚ increments to -70˚ using a target under-focus value of 5-8 μm.

3.2.4 Tilt-series alignment and tomogram reconstruction

Twenty tilt-series with negligible mechanical or physical artifacts were selected for image processing and tomographic volume reconstruction. The tilt-series were roughly aligned using cross-correlation in IMOD with default parameters [79]. The tilt-series alignments were further refined using fiducial-free “area matching with geometry refinement” as implemented in the Protomo [38] software package. The resolution of the reference was used in template matching was bandpass filtered to 40 Å to suppress noise and include information only before the first zero-crossing of the contrast transfer function. Alignments were initialized at a binning of 4, with an ~12 Å nominal pixel-size, which was incrementally decreased each time the image shifts dropped below 1 Å. Using the refined geometry parameters, the raw projections were centered and rotated so that the tilt-axis coincided with the Y-axis of the micrographs. The CTF was corrected on the 2D projections with phase flipping using tomoCTF [78]. A strip-width of 180 Å was chosen based on the thickness of the sample at high tilt. Tomograms were reconstructed using simultaneous-iterative refinement (SIRT) as implemented in IMOD. I observed that at least 30 iterations were needed to avoid strong low-pass filtering of the algorithm when it is not run until convergence. These were calculated using a GPU, thereby removing an additional interpolation in the reconstruction step, by avoiding the use of cosine stretching of the input projections. Using the observed threshold of 30 iterations for approximate convergence, volumes calculated from 20 SIRT iterations, providing higher contrast, were used for the initial cycles of sub-tomogram extraction and alignment, while those from 60 SIRT iterations were used for the final cycles.

3.2.5 Template matching

To extract sub-tomograms, initial positions of the receptor complexes, respective to a Cartesian grid defined by each tomogram, were approximated by using a template matching algorithm implemented in MATLAB with a reference that emphasized the receptor dimers with little influence from the histidine kinase CheA. Both the template and tomograms were low-pass filtered to 40 Å and binned to a 9 Å pixel size. This resolution, as well as a coarse angular search, were chosen to limit the potential for model bias at the target resolution of 10 Å. Following template matching, the data were randomly split into two halves, which were processed independently for all subsequent steps.

3.2.6 Sub-tomogram alignment and classification

Sub-tomogram alignment and classification were carried out using custom scripting of the image processing utilities from the Protomo i3 package [38]. Alignment and classification were carried out simultaneously, where multiple references representing a trimer of CSUs and a hexamer of CSUs were selected from class averages. The classification was performed using Multivariate Statistical Analysis and Hierarchical Ascendant Classification.

In each cycle, eight class averages were produced from each half data set by focusing the analysis on the CheA portion of the complex using a cylindrical mask, offset from the center of the volume in Z. Initial references for each half set were selected from these class averages by choosing the “best” (visually) trimers and hexamer of CSUs.

These references were then used to align class averages chosen to each have ~50 contributing sub-volumes.

In the following cycle, the raw sub-tomograms were subject to multi-reference alignment, but only a small in-plane and translational adjustment were allowed. This alignment by classification was repeated five times while allowing the automatic exclusion of high variance outliers after the second cycle. After the final cycle, class averages containing either the trimer of CSUs or hexamer of CSUs were manually selected and averaged together for each half data set.

The corresponding gold-standard FSC was calculated to evaluate the reliability of the data. Soft cylindrical masks were used, rather than spherical masks, given the extended slab like nature of the specimen. The final averages of the trimer of CSUs or hexamer of CSUs contained 3,000 sub-tomograms or 300 sub-tomograms respectively, and an empirical correction for the CTF envelope was applied for sharpening

To assess the degree of resolution anisotropy, conical Fourier shell correlations from the two independent half data sets were calculated along each of the principal axes, as well as the ten axes bisecting them [82]. The averaged density map of a trimer of CSUs was then low-pass filtered according to the conical FSCs along three principle axes by using cones with a half angle of 30˚.

3.3 RESULTS

Our analysis of the *in vitro* reconstituted arrays, revealed the domain organization of CheA in the CSU, as well as the quaternary structure of the hexamer of CSUs to ~ 18 Å and the trimer of CSUs to ~ 11 Å. I also show that CheW does form 6-membered rings at the center of the hexamer of CSUs, and by mapping the sub-tomogram positions back to the tomograms confirmed the array organization *in vitro* matched the previously observed *in vivo* pseudo-p6 lattice.

3.3.1 In Vitro reconstituted array re-capitulates observed in vivo array

Incubating CheW and CheA with TarCF tethered to a lipid monolayer is sufficient to produce extended arrays of signaling complex (**Figure 3.3** A), which recapitulate the pseudo-p6 order previously observed *in vivo* [110] (**Figure 3.3** B). At the center of each 6-fold symmetry center, there is a clearly defined ring of protein density at the same height in the protein interaction region as the CheA-p5/ CheW ring at the 3-fold symmetry center. This confirms the presence of a CheW only ring which has been speculated (as alluded to in the introductory figure) to add stability to the array (**Figure 3.3** C).

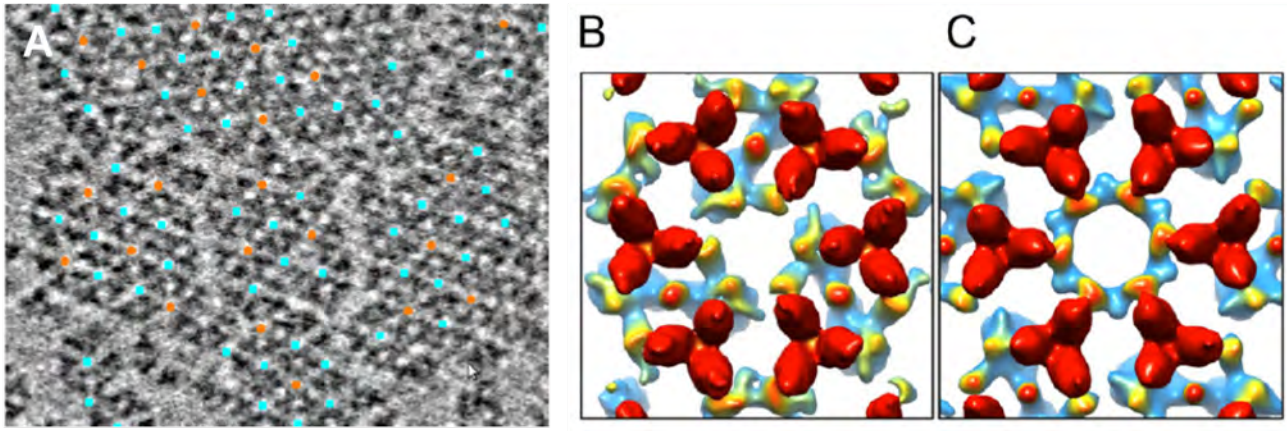

Figure 3.3 Trimer and Hexamer of CSUs

A) Mapped back locations of the trimers and hexamers as identified by Multivariate Statistical Analysis reveal the pseudo p6 lattice, 3-fold centers marked by blue dots, and 6-fold centers marked by orange dots. Sub-tomogram averages of B) the timer of core-signaling units and C) the hexamer of core signaling units, which displays a ring of density that is CheW [30].

This approach has also revealed that the degree of methylation correlates with how ordered the arrays are, where the un-methylated OFF state produces extended but disordered arrays visible at moderate resolution (~3 nm) in cryo-electron tomograms (**Figure 3.4** A), while the methylated ON state produces extended planar arrays with apparent long-range order (**Figure 3.4** B). Note that the contrast appears higher due to the extended low-resolution (1 – 5 nm) order in the QQQQ arrays. It is an open question in the field to what degree the array disassembles and what role this may play in signaling regulation [116].

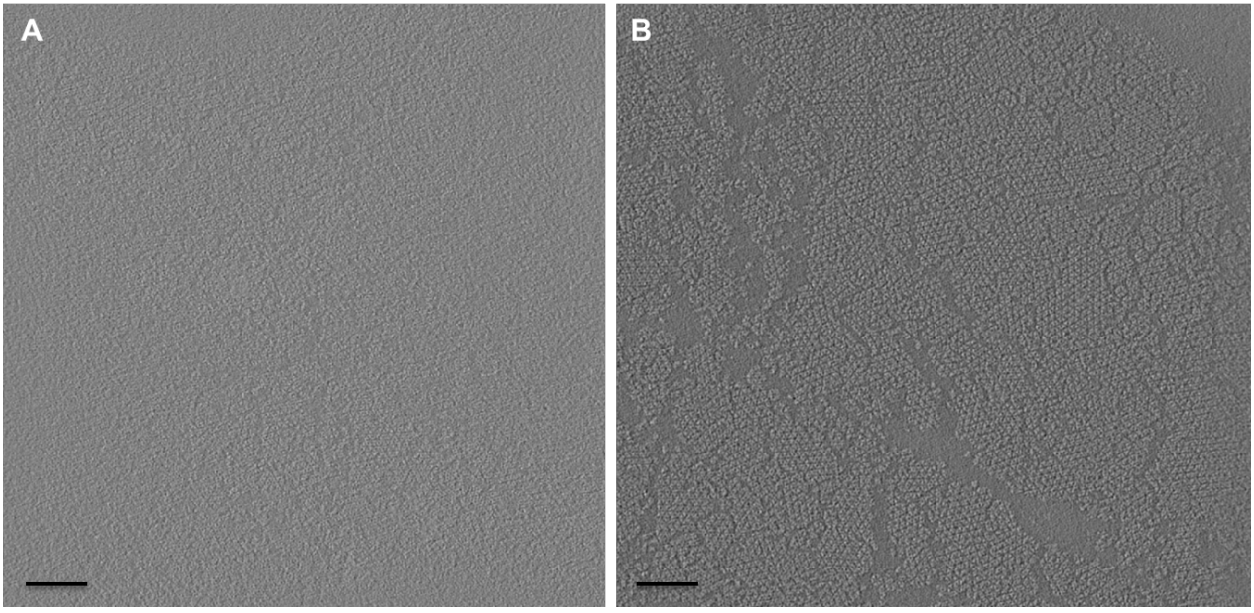

Figure 3.4 Receptor signaling state determines long range array order

(A*) Tar-CF CheA/CheW ternary complex in 4E (OFF) state form extended but loosely packed arrays. (B) Tar-CF CheA/CheW ternary complex in 4Q (ON) state form extended and highly ordered, planar arrays. Scale bars 100 nm.*

3.3.2 Domain architecture revealed

The arrays in this study should be biased slightly toward the ON state as they have the "neutral" QEQE receptor modification. In agreement with cross-linking data from a number of studies the CheA-p5/CheW interface I is clearly resolved and the density at interface II is weaker as emphasized with the black arrows in **Figure 3.5** A. This is in keeping with the hypothesis that as arrays are biased to the kinase ON state, the arrays become more ordered by strengthening interface II. By extension likely also forming

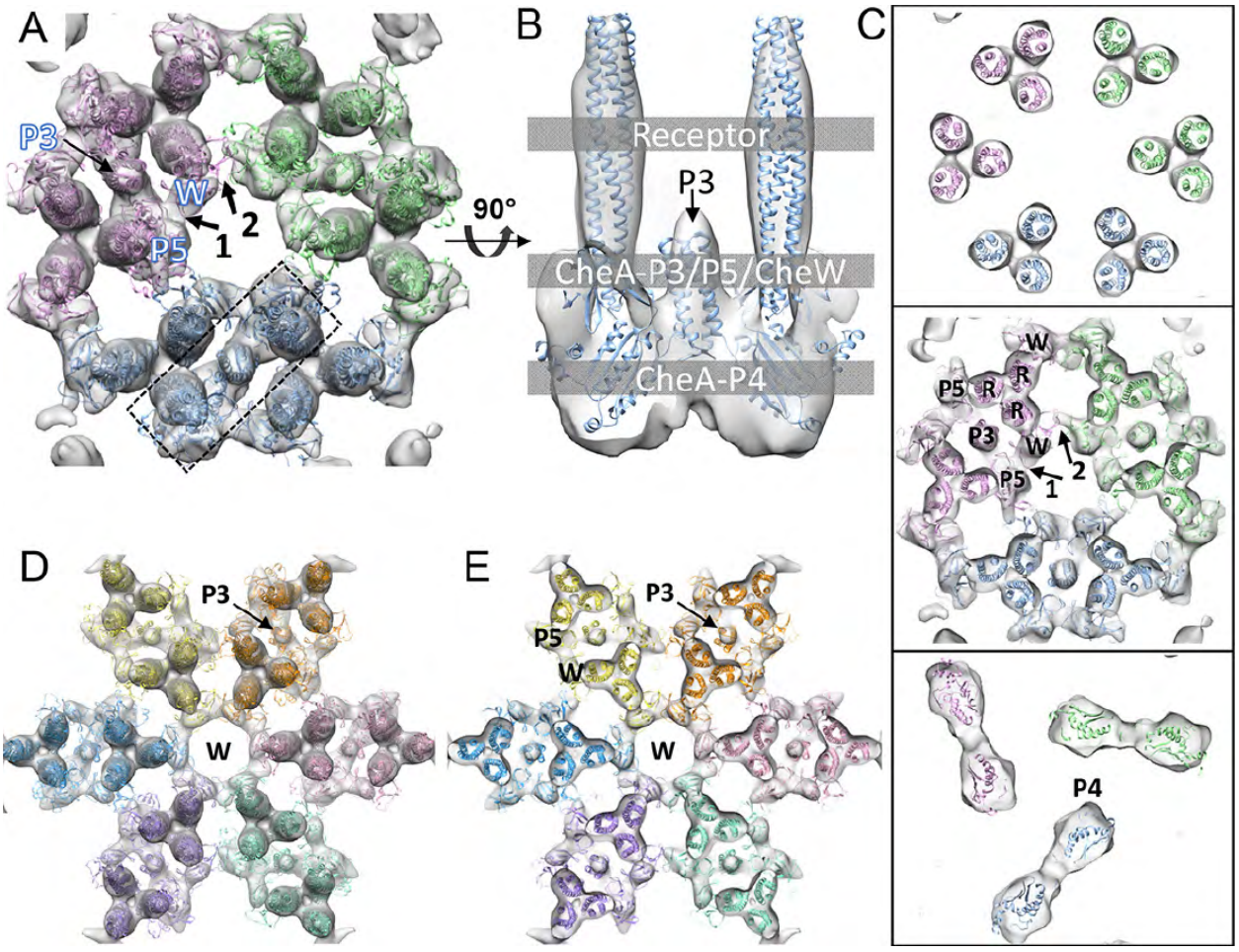

Figure 3.5 Domain architecture of the core-signaling unit and its higher order assemblies

A) A trimer of CSUs with the backbone atomic model docked in viewed from the membrane looking into the cytoplasm. The density between the receptor trimer of dimers matches the expected 4-helix bundle for the CheA-p3 dimerization domain. In this density from WT, I observe a strong density at CheA-p5/CheW interface 1, while interface 2 is much weaker. B) CSU rotated 90˚ showing the fit of CheA. C) Slices through three different positions of the trimer of CSUs shown in A, highlighting the 20A receptor TODs, the interfaces in the protein interaction region, and the fitting of the CheA-p4 kinase domain at the most membrane distal position. D/E highlight the hexamer of CSUs with the CheW only ring marked as W while the CheA-p5/CheW ring is marked as well [30].

stronger CheW only rings. We also resolve for the first time the location of the dimerization 4-helix bundle formed by CheA-P3 domain (**Figure 3.5** B) allowing us to establish the organization of the CSU to unprecedented resolution.

3.4 EXAMINING RECEPTOR ARRAYS *IN SITU*

To build on these findings further, the arrays need to be studied in a lipid bilayer. We previously showed that the membranes left after *E. coli* cells that are gently lysed by expressing the phage "E" gene were thin enough for TEM [117]. While the inner membrane does retract from the cell wall (**Figure 3.6** A), we show that the chemotaxis receptor arrays remain intact (**Figure 3.6** B, C). In addition to the array ultrastructure in (B), we show the trimer of CSUs averaged in panel C.

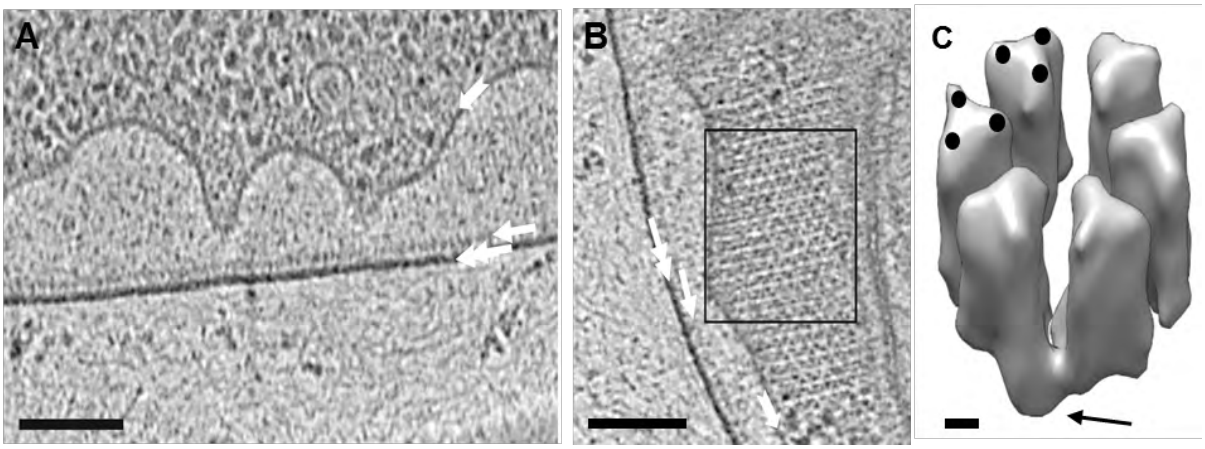

Figure 3.6 Integrity of the inner-membrane following lysis

A) The change in osmotic pressure upon cell lysis causes the inner membrane (chevron) to retract from the peptidoglycan layer (single arrow) and the outer-membrane (double arrow). B) A large patch of chemotaxis receptor signaling complexes is shown intact within the retracted inner-membrane. Arrows denote the same features as in A. C) Sub-tomogram average of the trimer of receptor CSUs from the patch in B at ~ 20A resolution. Each trimer of receptor dimers (black circles) is bridged by the soluble kinase CheA (black arrow). Scale bars 100 nm. Figure adapted from [117]

3.4.1 Reducing false positives *in situ* template matching

Compared to the *in vitro* reconstituted arrays, these lysed cellular specimens provide several new challenges. The identity of the of macromolecules present is not well-defined, the contrast in the images is substantially worse (**Figure 3.7** A), and the membranes themselves present an additional obstacle to particle detection as they produce strong non-specific correlation peaks with target proteins when searching tomograms of these cell ghosts using template matching.

I show promising initial results that improve the signal to noise ratio of the cross-correlogram (output of the matched filter) by up to seven-fold. Here the SNR is defined as the peak height of the true cross-correlation peak compared to the mean intensity of cross-correlation values in a local neighborhood. In **Figure 3.7** B**.** the ratio of the SNR_{decoy}/SNR_{normal} is plotted as a function of the relative resampling rate. As expected a small change in the pixel size ~ 95% sampling, produces the strongest effect.

I have also tested this approach using purified ribosomes which provide a visual conformation of the location of the real peak, and obtain similar results; however, the optimal change seems to be larger at ~ 85% sampling rate. It is currently unclear if the sample dependence may be calculated, or if a simple optimization routine might be included either at the beginning of template matching or as a stand-alone program. Further testing is needed, with realistic ground-truth simulated data.

.

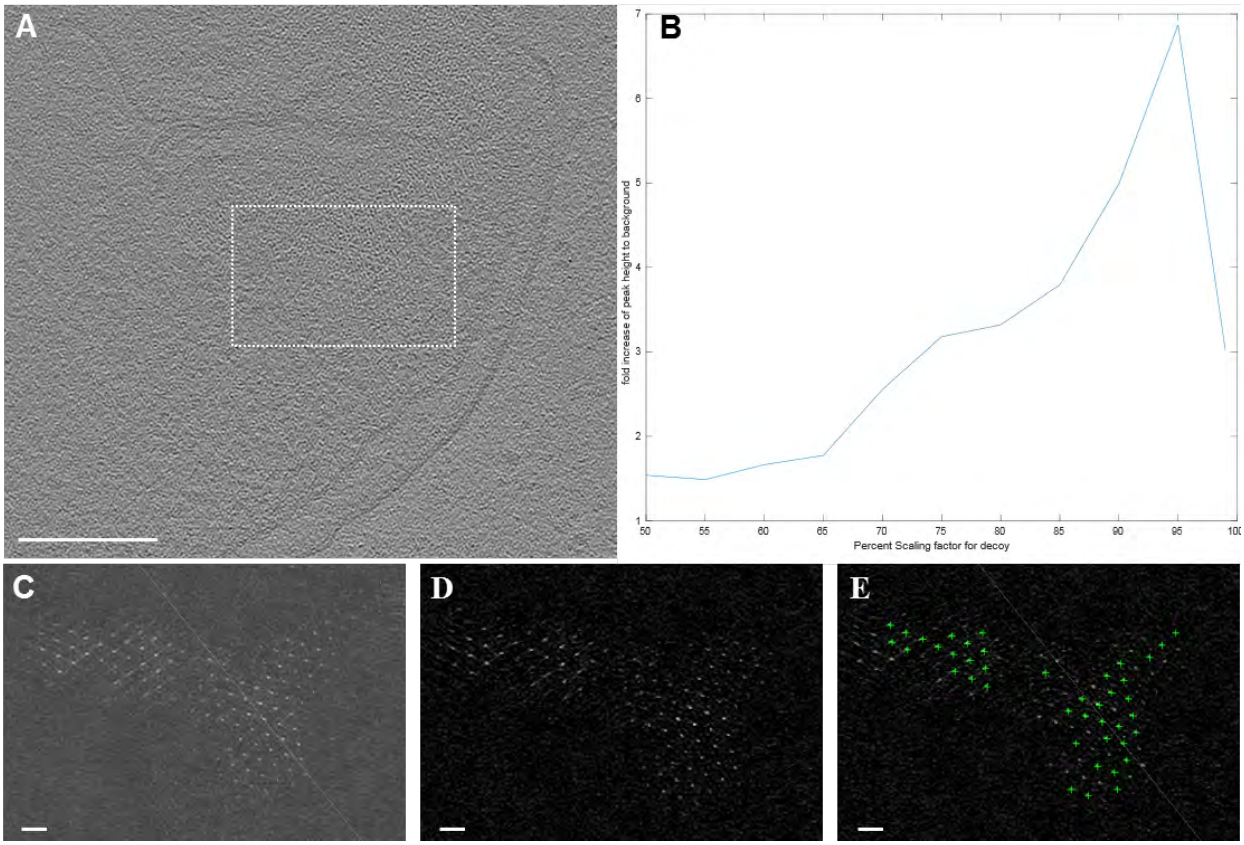

Figure 3.7 Improving SNR in template matching using a decoy for noise floor subtraction

A) central 20 nm from a cellular tomogram of an e-gene lysed E. coli cell, where the dashed box surrounding a patch of chemotaxis receptors shown in the cross-correlation maps in C-E. Scale bar 100 nm. B) Plot illustrating the ratio SNR_{decoy}/SNR_{normal} *as a function of the percent scaling of the decoy. C) CCC map of the patch in A, D) ccc-decoy map of the patch in A, with 95% scaled decoy, E) same as D with positions used to calculate B marked with green. Scale bars C-E 10 nm.*

3.5 DISCUSSION

The *in vitro* reconstituted bacterial chemotaxis signaling arrays we developed provide images of well-defined components with several advantages for cryoSTAC analysis, including large numbers of molecules in a very thin layer of ice. This enabled us to solve the structure of the assembled components of the CSU at unprecedented resolution, permitting the construction of a pseudo-atomic model for the extended array as described in our eLife paper.

This was possible since we could for the first-time be confident in our placement of the kinase CheA-P3 dimerization 4-helix bundle, as well as proving the presence of the CheW only ring which we believe serves to stabilize the active form of the receptor array. The CheA-P4 domain of the kinase proved to be highly dynamic as suspected from EPR and Fluorescence studies; this is consistent with our density map, where CheA-P4 has a poorer resolution. Looking forward, we should be able to apply the classification approaches we have developed for the *in vitro* arrays, to these dynamic complexes *in situ*, given that we may achieve an improved SNR with data collected using new direct electron detectors as well as imaging conditions more favorable to high-resolution structure determination, particularly avoiding very far from focus.

Despite these successes, the monolayer also creates substantial difficulty due to the strong preferred orientation the specimen assumes. We were able to implement a successful alignment strategy to both prevent this from biasing the alignment, as well as accounting for it in the final map reconstruction.

The ultimate limitation on the resolution, however, was the noise from the CCD based detector, combined with the very high defocus used which coupled together create a strong envelope function attenuating the structure factor amplitudes.

3.6 CONCLUSION

The study of the *E. coli* chemotaxis receptor signaling arrays using *in vitro* reconstituted lipid monolayers has provided valuable insight into the organization of the bacterial chemotaxis core signaling unit, and its higher-order assemblies. These findings also suggest how increased large-scale order may play a role in signaling regulation. By imaging a restricted set of components *in vitro,* we can assign their location with a high degree of certainty. This should result in the ability to carry out more informed biophysical studies of these dynamic assemblies, leading to a clearer understanding of cooperativity in array function.

3.7 ACKNOWLEDGMENTS

I thank Dr. J. Sandy Parkinson for bacterial strains and plasmids. I also thank Doug Bevan for assistance with computing infrastructure. This work was supported by the National Institutes of Health NIGMS Grant R01GM085043, P50GM082251-7518, 9P41GM10460, and 5R01GM098243 to Dr. Peijun Zhang.

4.0 HIGH-RESOLUTION STUDIES OF HIV-1 VIRUS LIKE PARTICLES

HIV/AIDS is a significant global health concern, and development of new therapeutics is central to its effective treatment, as the genetically hypervariable virus rapidly evolves resistance. A relatively new class of antiretrovirals termed “Maturation inhibitors” have shown initial promise, yet so far have failed to deliver a clinically successful drug. Structural study of the viral maturation process is complicated by the irregular nature of the immature HIV capsid. Cryo-Electron Tomography with sub-tomogram averaging and classification (cryoSTAC) is well suited to this problem but is limited in the resolution attainable. Using our newly developed software “emClarity,” I have pushed the envelope of this technique to 3.1 Å, producing the highest resolution map published to date by cryoSTAC, using a dataset of bevirimat stabilized immature HIV Gag virus-like particles from John Briggs group [118]. I also use emClarity to investigate the structural consequences of a single point mutation in the spacer peptide-1 Threonine 8 to Isoleucine in HIV-1 Gag polyprotein.

4.1 INTRODUCTION

HIV/AIDS remains a significant global health concern has resulted in over 35 million deaths so far, with roughly 2 million new infections each year [119]. Treatment of HIV is complicated by the virus integrating its genome into the host-cell genome which can result in a dormant phase of the viral life-cycle. Whether or not the virus lies dormant can be cell specific, and depends on whether pro-viral integration occurs in host genes related to cell growth [120]. If so, the viral life cycle may exist in a steady state for much of the infection; where infection, cell death, and subsequent turnover are in balance, resulting in a large number of replication cycles [121]. Each cycle allows for mutations to the viral genome, a fact which is exacerbated by the low fidelity of the HIV reverse transcriptase (HIV-RT). Taken together, this results in a rapid accumulation of genetic diversity creating a range of quasi-species within an infected individual [122], making the development of consistently effective therapeutics very challenging.

While the first anti-retroviral drugs for HIV were approved as early as 1987, it was not until the development of protease inhibitors in 1996 that HIV could be effectively treated [123]. Not that PR inhibitors were some “magic bullet” by themselves, rather they could be combined with existing RT inhibitors to form a potent triple-drug cocktail called combination therapy, also known as highly-active antiretroviral therapy (HAART) [124]. Since that time additional combination therapies have been formulated based on the development of new fusion inhibitors, nucleoside-analog reverse transcriptase inhibitors (NRTIs), non–nucleoside reverse transcriptase inhibitors (NNRTIs), integrase inhibitors, and protease inhibitors [125]. The efficacy of these cocktails comes at some expense as

long-term treatment has revealed several toxicities: *myopathy; sensory neuropathies,* including distal symmetric poly-neuropathy, inflammatory demyelinating polyneuropathy, mononeuritis multiplex, progressive polyradiculopathy, and autonomic neuropathy*; lipoatrophy; mitochondrial toxicity,* including hyperlactatemia lactic acidosis*,* and hepatic steatosis [126].

The primary viral structural protein to initiate its assembly is the gag polyprotein which encodes six covalently linked proteins: matrix (MA), capsid (CA), nucleocapsid (NC), spacer-peptide 1 (SP1), spacer-peptide 2 and p6 (**Figure 4.1** A) [128]**.** The structures of the isolated MA, CA, and NC proteins are known from a combination of X-ray crystallographic and NMR studies, shown in connected by flexible linkers in (**Figure 4.1** A).

Gag polyproteins assemble at the host membrane in a structure having the appearance of radial spokes. This initial structure is transformed by well-coordinated proteolysis of gag starting with the SP1-NC juncture, followed by MA-CA, and culminating the cutting of the CA-SP1 juncture (arrows **Figure 4.1** A, B) which is required for the virus to mature [129].

A new class of anti-viral compounds has emerged that prevent new viral particles from maturing by blocking this final cleavage by stabilizing the immature capsid lattice [127]. I focused my efforts on determining the structural mechanism underlying the maturation process in the late stages of the viral life cycle when the virus is preparing to infect new cells.

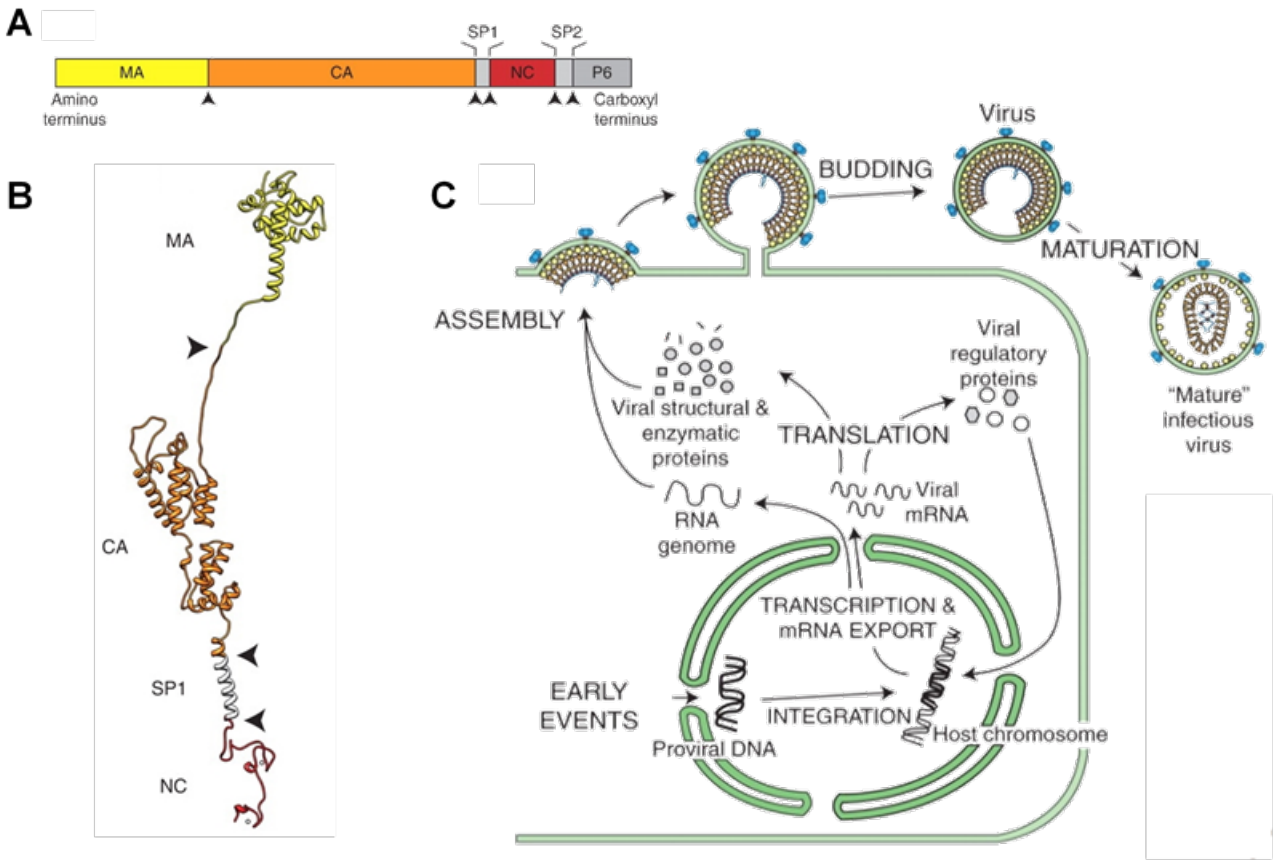

Figure 4.1 HIV-I life cycle.

Schematic showing the primary structure and domain organization of the HIV-I gag polyprotein. (A) Atomic models from x-ray crystallographic and NMR studies arrange in radial spokes from the membrane proximal MA inward. The relative orientations depicted are only approximate given the large degree of flexibility in the linkers. (A, B) arrows indicate cleavage sites for PR. (C) Schematic illustrating the infection cycle of HIV. Figure adapted from [130]

A maturation inhibitor (MI), 3-O-(3',3'-dimethylsuccinyl) betulinic acid, a derivative of betulinic acid, was isolated from the leaves of *Syzygium clavifloru* [131]. This compound, also called PA-457 or bevirimat (BVM) was subsequently shown to act a late stage in gag processing, preventing the proteolytic cleavage from p25→p24 (CA) [132]. Another MI PF-46396 has been isolated through drug screens of replicating *in vitro* viruses [133]. Initial phase I and II clinical trials of BVM showed that the drug was generally well tolerated and produced significantly reduced viral loads in a single-dose response [134]. Ultimately the sensitivity of the drug is extensively modulated by high baseline polymorphisms in the CA-SP1 region, particularly in clade B genotypes [135,136].

Keller et al. used cryo-ET to compare the ultra-structure of virions where the final gag cleavage at the CA-SP1 juncture was blocked by either sterically or allosterically by BVM or by genetic changes to the CA-SP1 cleavage site (L364I and M368I, a.k.a. "CA5" mutant [137].) They observed a layer of density, attributed to CA-SP1, which remained organized in a hexagonal lattice in contrast to the CA5 mutant which did not [138]. This finding suggests BVM inhibits HIV-protease by blocking its action allosterically through a stabilization of the immature Gag lattice. It then follows that a high-resolution structure of the CA-SP1 helical bundle may permit rational design of other stabilizing interactions that act similarly.

The recently reported cryoSTAC structure of CA-SP1 + BVM at 3.9 Å by Schur et al. revealed for the first time a six-helix bundle [48] formed by SP1 and the C-terminal residues of CA. The concurrently published crystal structure of isolated CA-CTD/SP1 by

Wagner et al. further supports the conclusion [139]. Interestingly, a density appearing stronger in the center of that 6-helix bundle was attributed to BVM, though the C6 symmetry applied precludes any conclusive identification.

An alternative stabilization of the immature lattice has been identified which is due to a single point mutation in Gag, threonine-371 to isoleucine [140]. This is residue 8 in SP1, and following the literature, I will refer to it as gag-T8I. Sequence numbers reference gag polyprotein (a.k.a. pr55-gag) UniProtKB/Swiss-Prot: P03347.3.

4.2 EXPERIMENTAL PROCEDURES

Our lab has expressed the Gag protein containing the T8I point mutation in *E. coli* cells, leading to self-assembly in a crowded cellular environment. Assembled Gag virus-like particles (VLPs) were purified from the cell lysate via ultra-centrifugation through a sucrose gradient.

4.2.1 Cryo-ET and Image processing of HIV-1 gag + BVM

Schur et al collected these data. under very similar conditions [48], including the use of a newly developed tilt-scheme, the “Hagen” or “Dose-symmetric” scheme, designed to concentrate the electron dose in the low angle tilt [141]. I used the same movie alignment strategy as with the Gag T8I data. However, the aligned sums were Fourier cropped to 1.0 Å rather than the physical Nyquist of 1.35 Å. Cropping to physical Nyquist is typically

done to remove aliasing of high-resolution imaging artifacts. However there is useful information beyond physical Nyquist [142], and I reasoned that the higher sampling rate might help to suppress signal degradation at high resolution from the multiple interpolations in the cryoSTAC pipeline.

Rather than selecting individual VLPs to reconstruct (as a matter of convenience) each tilt-series was divided into four quadrants, and the template matching results were cleaned automatically using a new feature in emClarity that uses constraints based on neighboring peaks to decide if a hit is likely a false positive or not. For this, I enforced any retained peaks to have at least 5 (of the expected 6) neighboring peaks within 100 Å and ± 20˚.

4.2.2 Cryo-ET and Image processing of Gag T8I

Dr. Jiying Ning prepared gag T8I particles, Dr. Xiaofeng Fu froze cryoEM grids with 10 nm gold beads, and Dr. Alistair Seibert collected tilt-series at the electron bioimaging center (eBIC) at the Diamond Light Source, UK. Eight dose fractionated frames were recorded for each tilt angle, in super-resolution mode with a Gatan K2 direct electron detector on a bioquantum energy filter, with a physical Nyquist sampling rate of 1.35 Å per pixel. Data were collected using a bi-directional tilt scheme starting from 0˚, proceeding to -51˚ then 0 to +51˚ in 3˚ increments accumulating a total electron dose of 120 electrons/$Å^2$.

I aligned dose-fractionated movie frames using the full field of view with the program *unblur* [73], without applying an exposure filter, which was done later in emClarity. The summed movie frames were concatenated into an image stack and aligned using the 10 nm gold-fiducial markers in IMOD [79]. I note that at this magnification, 5 nm beads would have provided sufficient signal for tracking while obscuring less of the sample.

These tilt series were transformed and aligned according to their tilt geometry (rotation, mag, shift) and then cropped in Fourier space to the physical Nyquist of 1.35 Å. Individual virus-like particles were selected for further analysis from bin ten tomograms using scripts provided with the emClarity software. Estimation of the defocus including astigmatism was carried out for each tilted projection, and this information was used for 3D-CTF correction in emClarity.

4.3 RESULTS

Using emClarity and the excellent data of Schur et al. from the EMPIAR database (EMPIAR-10164) I have reached the highest resolution using cryoSTAC to date, at 3.1 Å. I additionally present an initial map for gag-T8I at 5 Å that suggest the density attributed to the maturation inhibitor BVM may indeed be correct. These preliminary findings also suggest that on improvements in data collection, the gag-T8I sample should reveal new insights into any structured interactions between CA-SP1-NC/RNA. Specific changes to imaging include selecting areas with a higher particle density to improve defocus

estimation and a total number of asymmetric units available for averaging, as well as collecting data closer to focus to improve the coherence of high-frequency information by reducing the impact of small errors in CTF correction.

4.3.1 Near-atomic resolution using cryoSTAC and emClarity

The HIV-1+BVM tilt-series that resulted in the publication of the 3.9 Å map from cryoSTAC [48] have been released during the final stages of the writing of this thesis, EMPIAR-10164. A second publication by the same group showed they could reach a nominally identical resolution (but noisier map) with only 5-tilt series (~10% of the total) by additionally correcting for the curvature of the Ewald sphere ("3D-CTF"). The full data set then reached to 3.4 Å and was the highest resolution map from cryoSTAC published before this work [143]. On a first pass at the data using emClarity, I reached a resolution of 3.6 Å using the same 5-tilt series subset of the data.

In **Figure 4.2** A I compare the density for a single CA monomer from 10% of the data using emClarity (left) to that from the full data from EMD-3728. The backbone from model PDB-5l93 is overlaid. Our density on the left compares favorably with the published map from the full data-set on the right, clearly showing the pitch of the backbone helices and several side chain densities. The densities themselves are somewhat "clunky" which is consistent with the resolution measured.

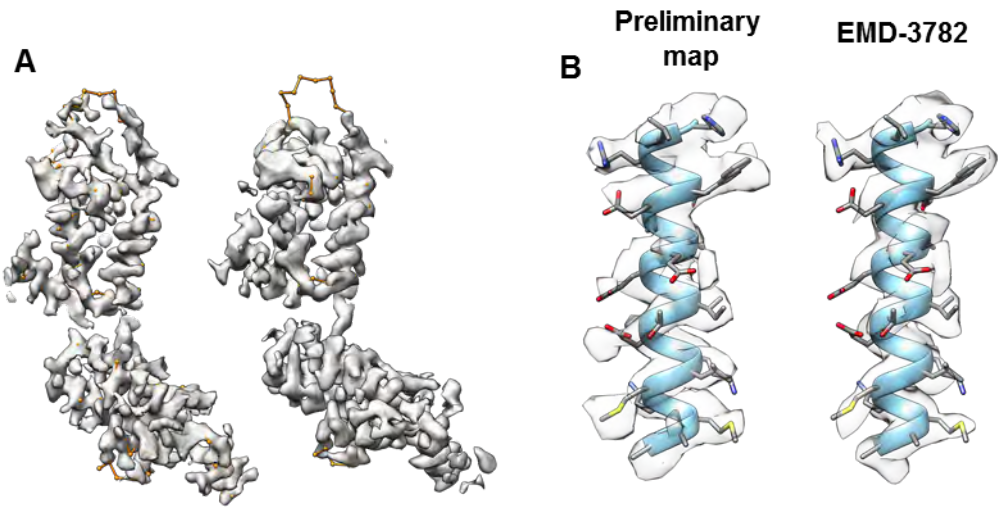

Figure 4.2 HIV-1 + BVM at 3.6 Å from 10% of EMPIAR-10164

A) Partial dataset map of CA monomer using emClarity (left) compared to that from the published 3.4 Å structure EMD-3728. B) Representative alpha helix from CA-NTD showing side-chain densities from the respective maps in (A)

The full data set reaches 3.1 Å in emClarity and is clearly better resolved than the published 3.4 Å map juxtaposed in **Figure 4.3** A. The gag hexamer has large pockets of solvent that would be included in a simple geometric mask (a cylinder for example) and failure to consider this would substantially underestimate the resolution [28]. One popular method to reduce the impact of the solvent on the resolution estimation is to calculate a very tight mask and then to correct the FSC for any effects that mask may have had on the resolution estimation [72].

Briefly, this is accomplished by calculating the uncorrected FSC of the data under the mask (FSC_U), as well as the data under the mask with high resolution noise replacing the signal (FSC_n), and removing any spurious correlation found only in FSC_n, as in equation 4.1 .

$$FSC_{true} = \frac{(FSC_U - FSC_n)}{1 - FSC_n} \qquad eq\ 4.1$$

This approach is susceptible to masking out real density, particularly for flexible specimen, which would lead to an overestimation of the resolution if care is not taken to prevent this. An alternative is approach is to compensate the FSC curve by estimating the fractional volume of the sample that may be contributed to the solvent.

$$FSC_{true} = \frac{f_{part} FSC_U}{1 + (1 - f_{part}) FSC_U} \qquad eq\ 4.2$$

Here f_{part} is estimated by using the ratio of the volume under the mask to the volume occupied by protein, which is in turn estimated from the molecular weight [144]. This density estimation approach is convenient for purified and well isolated samples, however it i,s not, suitable for heterogeneous assemblies where the mass may not be accurately known. As an alternative, I use our iterative mask dilation to estimate the non-solvent fraction as all positive density in the connected region determined in the masking. Given that the procedure is designed to err on the side of including more features, any systematic error in this approach would tend to over-estimate the solvent content and thereby err on the side of under-estimating the resolution. In this case, I show that the results from both approaches give the same resolution at the 0.143 cutoff, and very similar

overall FSC curves, with the tight mask compensated FSC shown in red and the solvent compensated FSC shown in dark blue in **Figure 4.3** B. Further support for the measured resolution is shown in the clear improvement amino acids with small side chains like alanine (**Figure 4.3** C). One region of particular interest is highlighted with the orange callout box in panel A which is expanded in Figure 4.3 D. Schur et al., interpreted this density as a stacking interaction between Y277-P279, which they hypothesized may help to stabilize the interdomain linker. I show that this density was likely an artifact, and that Y277 appears to form a hydrogen bond with neighboring H194. The SP1 helix, which was the primary focus of the original paper is shown side by side in **Figure 4.3** E (callout from blue oval the in A).

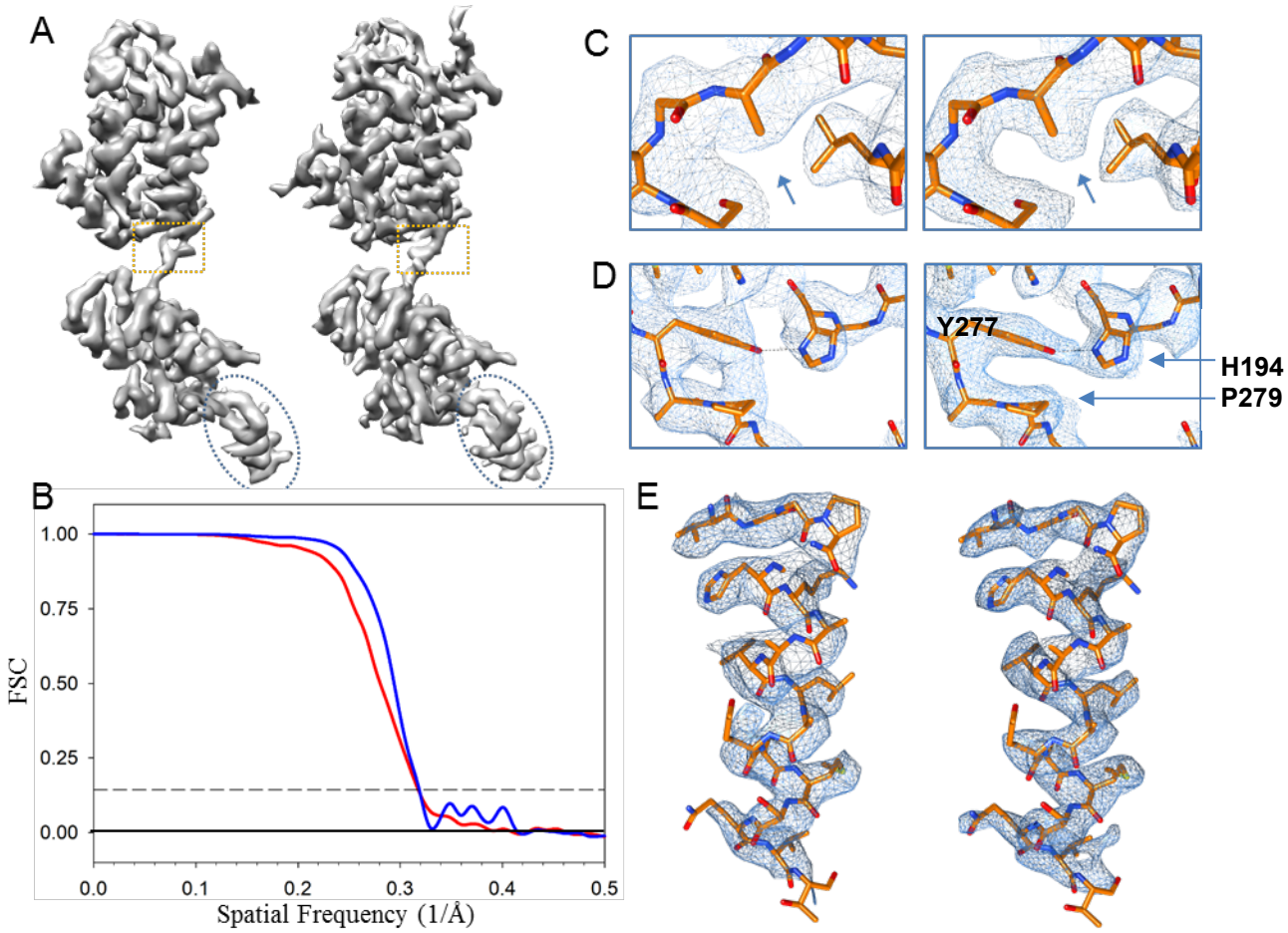

Figure 4.3 emClarity achieves the highest resolution sub-tomogram average to date

(A) EMD-3782 CA-monomer at 3.4 Å resolution (left) and emClarity at 3.1 Å (right). (B) "gold-standard" FSC between half sets of HIV-1 Gag (EMPIAR-10164). Red, fsc "true" tight mask with correction via phase randomization. Blue, soft mask with solvent corrected fsc. Both approaches indicated a resolution of 3.1 Å at the 0.143 cutoff (dashed line). (C) Iso-surface view scaled to match the leucine residue shows cleaner backbone, andresolution of even small side-chains like alanine (blue arrow). (D) Expanded view the CA-NTD/CTD linker (orange box in panel a) rotated by 180˚. The putative Y277-P279 stacking interaction does not appear in our map, however, emClarity shows a strong density bridging Y227-H194 which are ~ 3 Å apart indicating a hydrogen bond. (E) the CA-SP1 helix (blue dashed oval in panel a) as another example of the improved resolution of important regions in the map.

4.4 GAG-T8I MUTATION STABILIZES THE IMMATURE LATTICE

I show that the T8I mutation stabilizes the immature lattice particularly well in the very C-terminus of CA where continuous density with CA-NC and RNA density are shown in **Figure 4.4** A. I additionally find that the density that is coordinated by two rows of lysines in the Schur et al. structure is also present in our map, indicating it is likely the common small molecule inositol-hexaphosphate (IP6) as previously hypothesized, and recently confirmed [145]. Despite a well resolved 6-helix bundle in the CA-SP1 region, our map lacks any density in the region that Schur et al. hypothesized to be BVM (black arrow **Figure 4.4**) This seems to confirm that the density they attribute to BVM is correct. I also observe an additional coordinated density (this time by methionines) at the very end of the CA-SP1 region in the HIV-1 + BVM data set.

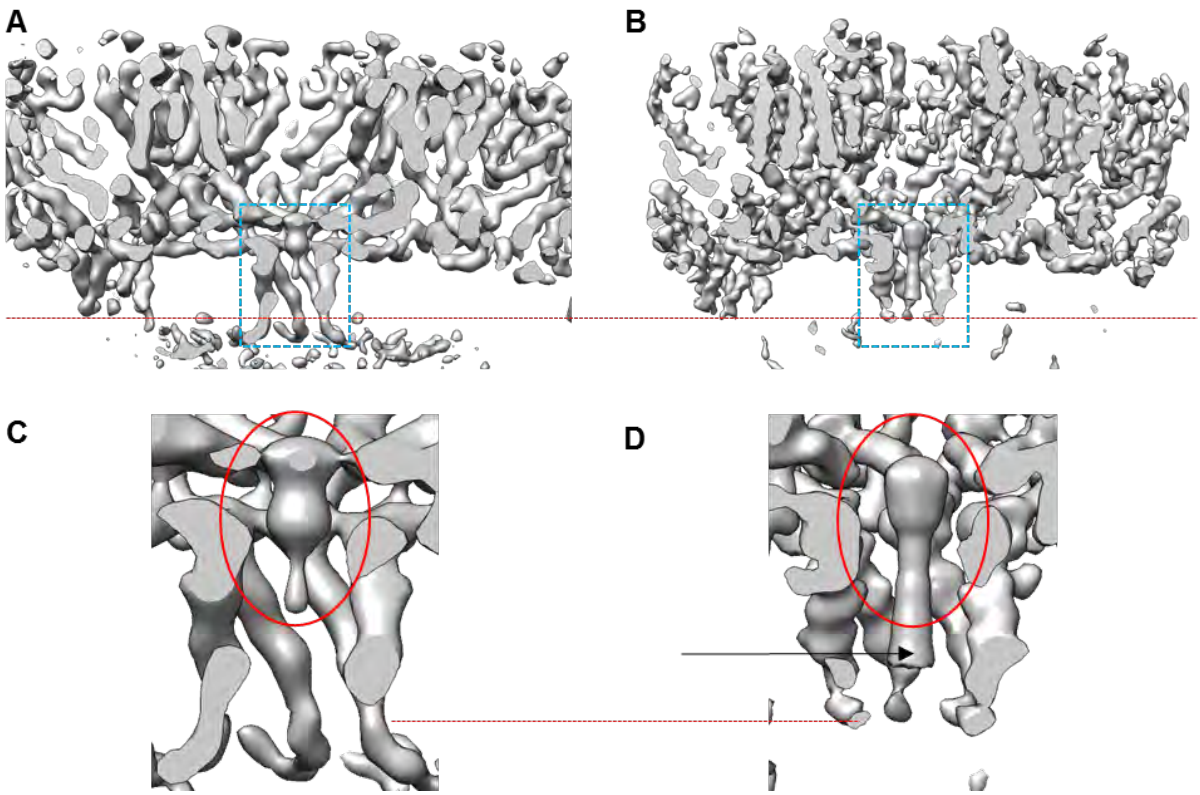

Figure 4.4 Comparison of T8I vs. BVM stabilized VLPs

(A) Cross-section through the CA-lattice as stabilized by the T8I mutation. The mutation is at the very c-terminus of the CA-SP1 helix, marked by the red-dashed line in all panels. (B) Cross-section through the CA-lattice as stabilized by the MI BVM. Blue box highlights the CA-SP1 6-helix bundle as magnified in C and D. The red oval in C/D highlights the region coordinated by two rows of lysines, which are seen more clearly in the T8I map. The red dashed line marks the location of T8(I) and the end of the CA-SP1 helix. Clear continuous density is seen in the T8I stabilized lattice in panel C, which should be CA-NC. D) The black arrow indicates additional density in the center of the 6-helix bundle that is attributed to BVM. The absence of this density in the T8I map further supports this hypothesis.

4.5 DISCUSSION

Using emClarity, I have extended the resolution achievable by cryoSTAC to 3.1 Å. This required a very large data set, of ~ 960,000 asymmetric units. The B-factor applied was 150 $Å^2$ which is substantially worse than that obtained using SPA. I suspect the following issues: First, every time the electron beam applied to the specimen (each tilt acquisition) there is a sizeable unresolved motion in the early milliseconds of exposure. This means data collection with multiple exposures will necessarily have more blurring per accumulated dose than a single exposure. Second, the exposure per movie frame is between 0.3 and 0.5 electrons/ $Å^2$, which leads to an additional b-factor due to further motional blurring by limiting the accuracy of the frame alignment. Third, multiple interpolations are currently required to align the tilt-series, reconstruct the tomograms and finally average the sub-tomograms.

The latter could be rather easily addressed by writing a program to reconstruct the sub-tomogram average directly from the tilt-series. The first two might be addressed by increasing the tilt-increment thereby reducing the number of “initial” exposures, while also increasing the total dose in each frame. Not only should this improve the movie frame alignment, but it should also improve the accuracy of the tomogram-constrained projection refinement in emClarity.

4.6 CONCLUSION

The HIV Gag T8I mutation provides a valuable specimen for studying the stabilization of the immature HIV gag lattice, which is a promising target for anti-retroviral maturation inhibitors. This mutant produces a more stable CA-SP1 bundle as compared to HIV gag+BVM, as evidenced by the extended region resolved in the gag-T8I map. Using the advances in emClarity, the highest resolution achievable with cryoSTAC has been pushed to 3.1 Å revealing noticeably cleaner side-chain densities. This is an important step for the field as well as the advance in our understanding of HIV biology.

4.7 ACKNOWLEDGMENTS

Dr. Jiying Ning and Dr. Xiaofeng Fu for their work with expression, purification, cryoEM sample preparation and imaging the Gag T8I particles. Dr. Alistair Siebert for cryoET data collection of Gag T8I. Dr. John Briggs and the members of his group for sharing their HIV + BVM data with me before its full public release on EMPIAR.

5.0 SUMMARY OF PROJECTS AND FUTURE PROSPECTS

Cryo-ET can resolve a broad range of length scales critical for connecting the structural details of molecules and their assemblies, to their broader cellular function. I have created a library of software called “emClarity” which extends the resolution attainable by cryoSTAC to 3.1 Å. While spatial resolution in cryo-EM/ET is assesed by the FSC which measures the self-consistency of the data, it is also relevant to consider the conformational resolution of different sub-populations of the data into homogeneous groups or classes. I demonstrate considerable improvements in the classification of compositional and conformational heterogeneity as compared to current state-of-the-art software.

Classification in cryoSTAC is complicated by the missing-wedge effect, which distorts each particle according to its orientation in the microscope. This creates strong features in the resulting image that are unrelated to the specimen. The missing-wedge effect is usually modeled by a simple binary wedge mask, which I have demonstrated is inadequate at spatial resolutions better than 2 nm. Accordingly, I have developed methods for a more accurate estimation of what information is present in a given sub-tomogram via the “3D-sampling function,” integrating established ideas from SPA.

In chapter 2 I detail an approach accounting for spatially anisotropic resolution and its impact on both the alignment process and the interpretation of the final structure. This sort of anisotropic resolution is the result of preferred specimen orientation, which is a common problem in cryo-EM. I also discuss extensions to improve the correction for the aberrations of the contrast transfer function. I have tried to make the tools as easy to use as possible while also remaining flexible enough to be useful for the informed computational experimentalist. I expect that it would be of great utility to have a simple graphical user interface to simplify further the use of emClarity such that it may enjoy a broader impact on the community.

The first routine developed in the library that became emClarity was the template matching program, aimed at one of the most significant hurdles in cryoSTAC – locating particles of interest in the 3D tomogram. This development of a fast template matching algorithm using GPUs, able to locate molecules in crowded environments was brought about in our studies regarding novel sample preparation of intact cell membranes, discussed in chapter 3**.** It has proven to be very useful in other areas, like isolate HIV-gag particles which provide too large some sub-tomograms to pick by eye. Key ideas advanced include efficient means of calculating local image statistics, challenges regarding computational expediency, and the means to reduce the impact of challenging sources of structural noise that result in many false positives, e.g., carbon film and biological membranes. This is of critical importance if the technique is to be useful in samples obtained from *in situ* specimens, particularly of FIB milled lamella.

Finally, in chapter 4 I have presented exciting results demonstrating cryoSTAC at resolutions where even small side chains are discerned. I also present preliminary findings that suggest the single-point mutation at CA-SP1, T8I may be very useful for high-resolution studies of immature HIV capsid stability. These results also seem to confirm the binding of the maturation inhibitor Bevirimat at the center of the CA-SP1 6-helix bundle.

In addition to new algorithmic approaches, new data-collection schemes could also be investigated. In particular, it would seem that collecting fewer exposures, each with more total electron dose, would be beneficial for a couple of reasons. First, the large beam-induced motion that occurs each time the specimen is exposed to the electron beam occurs for each tilted image. This means the first few frames of each movie contain little high-resolution information and ultimately that a tilt-series contains less useful high-resolution information than a single exposure micrograph. Second, all of the tools developed here, particularly the CTF refinement and tomoCPR would benefit from the stronger signal in the individual projections.

Looking forward, the newest generation of GPUs to be released are supposed to have 32 Gb of memory compared to the 12 Gb emClarity has been designed to run with. This, combined with the wealth of resources being invested in machine learning approaches promise to provide a fertile ground for the future development of high resolution *in situ* cryoSTAC.

5.1 LIST OF PUBLICATIONS

B.A. Himes, P. Zhang, emClarity: Software for High Resolution Cryo-electron Tomography and Sub-tomogram Averaging, Nat. Methods. (2018). doi:10.1101/231605.

C.K. Cassidy, **B.A. Himes**, Z. Luthey-Schulten, P. Zhang, CryoEM-based hybrid modeling approaches for structure determination, Curr. Opin. Microbiol. 43 (2018) 14–23. doi:10.1016/j.mib.2017.10.002.

J. Ning, G. Erdemci-Tandogan, E.L. Yufenyuy, J. Wagner, **B.A. Himes**, G. Zhao, C. Aiken, R. Zandi, P. Zhang, In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway, Nat. Commun. 7 (2016) 13689. doi:10.1038/ncomms13689

C. Liu, J.R. Perilla, J. Ning, M. Lu, G. Hou, R. Ramalho, **B.A. Himes**, G. Zhao, G.J. Bedwell, I.-J. Byeon, J. Ahn, A.M. Gronenborn, P.E. Prevelige, I. Rousso, C. Aiken, T. Polenova, K. Schulten, P. Zhang, Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site, Nat. Commun. 7 (2016) 10714. doi:10.1038/ncomms10714

C.K. Cassidy[‡], **B.A. Himes[‡]**, F.J. Alvarez[‡], J. Ma, G. Zhao, J.R. Perilla, K. Schulten, P. Zhang, CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling., Elife. 4 (2015). doi:10.7554/eLife.08419

X. Fu, **B.A. Himes**, D. Ke, W.J. Rice, J. Ning, P. Zhang, Controlled Bacterial Lysis for Electron Tomography of Native Cell Membranes, Structure. 22 (2014) 1875–1882. doi:10.1016/j.str.2014.09.017

[‡]These authors contributed equally.

APPENDIX

Datasets

The datasets used in emClarity processing are from Electron Microscopy Public Image Archive (EMPIAR), including the yeast 80S ribosome (EMPIAR-10045), the mammalian 80S ribosome (EMPIAR-10064), and the HIV-1 immature Gag (EMPIAR-10164).

emClarity programs

emClarity is run from the command line and is easily scripted to run in a manner most suited to a user's particular project. A text parameter file is used to input project specific details, like microscope parameters, mask dimensions, and angular search ranges. I typically make a copy of the parameter file for each cycle of averaging and alignment, which I refer to as paramC.m, where C refers to the cycle number. The meta-data of each project is tracked in a binary database which is named using the "subTomoMeta" parameter. Each tilt-series may have multiple areas slated for reconstruction, tiltN M refers to tilt-series "N" and reconstruction area "M." A brief description of the major functions (*in italic*) in emClarity is below:

emClarity init paramC.m

Read in the desired dimensions for each sub-region of each tilt-series to reconstruct, initialize the subTomoMeta (metadata binary.)

emClarity ctf estimate paramC.m tiltN

estimate the ctf for a given tilt-series N

emClarity reScale mapNameIN mapNameOUT AngPixIN AngPixOUT CPU/GPU

resample a map to a new pixel size, particularly for template matching.

emClarity templateSearch paramC.m tiltN M reference.mrc symmetry gpuIDX

Reconstruct tomogram M from tilt-series N without ctf correction, run template matching on GPU # gpuIDX, randomize results at symmetry-related positions.

emClarity ctf 3d paramC.m

Run 3d-CTF corrected weighted-back projection.

emClarity avg paramC.m N RawAlignment

Every cycle begins by creating a sub-tomogram average, calculating the “gold-standard” FSC, and weighting the average accordingly while compensating for amplitude attenuation by the CTF to produce references for the alignment.

emClarity alignRaw paramC.m N

Run the alignment

emClarity removeDuplicates paramC.m N

cleans out any sub-tomograms that have drifted to the same position. Not needed every cycle.

emClarity tomo-CPR paramC.m N

Run tomogram constrained particle refinement; this is generally done before a step down in binning, e.g., bin4 → bin3.

emClarity ctf update paramC.m N

Only needed after a run of tomo-CPR, this updates the tilt-series geometry in the subTomoMeta, and also resamples the raw tilt-series applying rotation, shift, and magnification scaling all in Fourier space to reduce interpolation losses of high-resolution information.

Since the tilt-series alignments are updated, and usually also the binning is reduced, a new round of 3D-CTF reconstructions need to be made.

If the classification is to be run, the cycle starts the same, but with the “flgClassification” parameter enabled.

emClarity avg paramC.m N

emClarity pca paramC.m N previousPCA

Run 3D-sampling function compensated PCA at each length scale specified in the "pcaScaleSpace" parameter. The command line argument previousPCA is always zero in the first run. If a random subset (25% or ~3000, whichever is larger) is to be analyzed by setting "Pca_randSubset", then a subsequent round of pca must be run with "previousPCA" set to one to project all of the sub-tomograms along the principal component axes.

emClarity cluster paramC.m N

Cluster the data based on selected eigenvectors from the pca step.

emClarity avg paramC.m N Cluster_cls

Notice the last argument (a string) which creates a montage of the class averages selected in the parameter file. Classes with different memberships may be selected.

At the end of processing, the half-sets may be aligned and combined by running:

emClarity avg paramC.m N FinalAlignment

Align and combine half-sets, optionally creating multiple, differently sharpened maps.

Image processing

For each specimen, I make a project directory, which I will refer to generically asThe alignment and classification procedures are generally identical for all the samples, except for the HIV-1 Gag data, which were not classified and had C6 symmetry applied. All parameters are unique to each dataset, including the angular search range and iterations used. emClarity is only tested on Linux operating systems, and all references to command line operations are to be understood in that manner.

Project set-up and coarse tilt-series alignment

For each specimen, I make a project directory, referred to generically as "projectDir." The HIV-1 Gag data consist of dose fractionated frames, which I aligned using the program "unblur" version 1.0 included in the cisTEM package. The aligned frames were summed and saved *without* any exposure-based filtering because this is handled later inside emClarity. All tilt-series were aligned using the default parameters in IMOD version 4.10.12 using the eTomo interface, with the available gold-fiducial markers. For the ribosome datasets, all gold fiducials were selected, while for the HIV-1 Gag data ~20-30 closest to the protein and distributed on both surfaces of the ice were selected. Local alignments with fixed XYZ global coordinates were run for the HIV-1 Gag data only. After generating the final aligned stack, the gold beads were located using *find beads3d*. Only the fiducial model describing the location of the beads is needed, so they were not erased.

Note that for EMPIAR 10045 the pixel size in the header must be corrected to 2.17 Å before beginning. This may be done with the IMOD program *alterheader* from the command line.

The files describing the projection transformations, any local alignments, and fitted tilt-angles are copied to the fixedStacks directory and renamed.

```
>$ mv specimen_name_1_fid.xf projectDir/fixedStacks/tilt1.xf

>$ mv specimen_name_1_fid.tlt projectDir/fixedStacks/tilt1.tlt

>$ mv specimen_name_1_local.xf projectDir/fixedStacks/tilt1.local

>$ mv specimen_name_1_erase.fid projectDir/fixedStacks/tilt1.erase
```

If outlier pixels are removed in IMOD, this “fixed” stack may be moved to projectDir/fixedStacks/tilt1.fixed, otherwise you may just link the raw data.

```
>$ cd projectDir/fixedStacks

>$ ln -s ../rawData/specimen_name_1.st tilt1.fixed
```

This is repeated for all tilts-series, of which there are 7, 4, and 41 in the yeast, mammalian, and HIV-1 Gag data sets respectively.

Ctf estimation

The mean defocus at the tilt-axis was then estimated in emClarity for each tilt-series using a 3.5 ± 2.5 μm window covering the range of expected defocus values for all three data sets. For the HIV-1 Gag data, the per-tilt defocus was determined using "emClarity ctf refine" to produce the power-spectra, which were subsequently fit using ctffind4 with the –amplitude-spectrum input flag and default parameters. For the yeast ribosome data which have a thin layer of carbon providing an extra signal in the power spectrum, the per-tilt defocus values were refined during tomo-CPR. To do so, the height of the cross-correlation peak is maximized by scanning through a small range of defocus values as applied to each reference tile [146].

Selecting sub-regions for further analysis

The selection of sub-regions of each tilt-series for reconstruction is defined by a text file with the minimum and maximum values in x, y, z for each region. The script "recScript2.sh" provided with emClarity was used to first create reconstructions of each tilt-series at a binning of 10 and thickness of 300 covering the full X, Y dimension of the images. Each region is then defined in IMOD by making an IMOD model with six points per region, xmin, xmax, ymin, ymax, zmin, zmax, in that order. A second run of "recScript2.sh" creates a projectDir/recon directory and converts these model files into the text files read in by emClarity to be used for the rest of the procedure. These are called tilt1_recon.coords and list the tilt-series base name, number of regions

to reconstruct, and for each region the width, first and the last slice in y, thickness, x-origin offset, and z-origin offset.

The ribosome data were divided on the x-axis into two regions per tilt-series. The HIV-1 Gag data were divided into quadrants. Additionally, the flag “fscGoldSplitOnTomos=1” is set in the parameter file for the HIV-1 Gag data, so that the even/odd half sets are divided based on tomogram, not randomly on sub-tomograms. This is necessary to avoid mixing neighboring particles which would violate the gold-standard hypothesis.

Template matching

References were derived from SPA EMD-3228 [48] (yeast 80S ribosome), EMD-5592 [147] (human 80S ribosome) and EMD-8403 (HIV-1 Gag) [148] and rescaled to the full pixel size of each data set using “emClarity reScale.” These references were then passed to “emClarity templateSearch” binned to achieve a nominal pixel size ~ 8-12 Å depending on the size of the specimen. All maps and tomograms are automatically low-pass filtered to 40 Å resolution by default in emClarity. Non-CTF corrected tomograms are reconstructed by the templateSearch program as needed for template matching.

The results for the ribosome dataset were cleaned manually by comparing the maximum intensity projection maps and the binned tomograms overlaid with an IMOD model showing the x,y,z coordinates of each peak detected.

For the HIV-1 Gag data, emClarity removeNeighbors was used to automatically clean the results based on geometrical restraints. Only peaks that had five neighbors within 100 Å

and also oriented within 20˚ were retained, resulting in 179,168 sub-tomograms to start. (This number dropped to 162,213 in the first round of averaging as particles too close to the edge to allow padding by 1.5 x particleRadius were excluded.

Particles with symmetry pose a special challenge to all missing-wedge compensation approaches as any error in the compensation will result in the particle looking different at its symmetry-related orientations. To help with this, I set the orientation found in template matching to any of the equivalent symmetry-related positions, and subsequently only search an angular range small enough to not reach the neighboring positions.

Iterative alignment

Each cycle of alignment is initiated by calculating averages of the two half sets, calculating the gold-standard FSC, and then applying re-weighting each average to generate a FOM weighted reference.

I alternate searching over just the azimuthal and polar angles, and an in-plane search. For each specimen, I started at a binning of chosen to produce a pixel size of ~7-8 Å. I then go through three rounds of averaging and alignment, followed by removing any positions that may have drifted to overlap using “emClarity removeDuplicates.” I then run a round of tomo-CPR, which requires updating the aligned tilt-series and the 3d CTF corrected tomograms.

```
>$ emClarity ctf update paramX.m
```

```
>$ emClarity ctf 3d paramX.m
```

This reconstruction is generated at a binning one finer than the previous, and the same pattern was repeated until reaching full sampling.

Classification

The ribosome data for the yeast 80S were classified in a single pass, using three resolution bands 10,18, and 28Å, 36 of the top eigenvalues were saved, and five from each band were selected (parameter Pca_coefficients=[7:11;7:11;1:11])] for clustering via kmeans the class averages were then generated by running

```
>$  emClarity avg paramX.m X cluster_cls
```

The ribosome data for the mammalian 80S were classified in two passes. First, they were split into groups displaying either a rotated or un-rotated 40S small subunit. To do this, the subTomoMeta file (projectName.mat) was copied to two new files: project_smallSU.mat and project_largeSU.mat. The classes are selected for removal by viewing the class average montage in IMOD and selecting any point in the region of a given class. These models are then used to remove their contributing members in the subTomoMeta.

```
>$  emClarity geometry paramX.m X RemoveClasses [X,0,0] STD.
```

Since both branches of the project access the same raw data, it is convenient to remain in the same project directory, and all subsequent output will be identified by the new subTomoMeta base name.

A subsequent round of classification was run using 12,22,32 Å resolutions. Unlike the yeast 80S which had some Eigen images with clear missing wedge bias, revealed as “streakiness” in the density, the mammalian displayed sufficient true variability to overpower the noise from the missing-wedge bias, and all 36 eigenvectors from each resolution band were used in clustering.

Analysis

Models PDB-3J78 for yeast were rigid body docked in using Chimera.

Models PDB-4UJO for mammalian were docked in using Chimera, in combination with the “Segger” plugin.

Models PDB-5l93 were docked in using Chimera, refined in real-space using Phenix version 1.13-2998-000, and manually edited in COOT version 0.8.9.

BIBLIOGRAPHY

[1] D.J. DeRosier, Correction of high-resolution data for curvature of the Ewald sphere, Ultramicroscopy. 81 (2000) 83–98. doi:10.1016/S0304-3991(99)00120-5.

[2] X. Zhang, Z. Hong Zhou, Limiting factors in atomic resolution cryo electron microscopy: No simple tricks, J. Struct. Biol. 175 (2011) 253–263. doi:10.1016/j.jsb.2011.05.004.

[3] H. Friedrich, P.E. De Jongh, A.J. Verkleij, K.P. De Jong, Electron tomography for heterogeneous catalysts and related nanostructured materials, Chem. Rev. 109 (2009) 1613–1629. doi:10.1021/cr800434t.

[4] N. Grigorieff, Direct detection pays off for electron cryo-microscopy, Elife. 2 (2013) 2–4. doi:10.7554/eLife.00573.

[5] A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M.I. Davis, R. Pragani, M.B. Boxer, L.A. Earl, J.L.S. Milne, S. Subramaniam, Breaking Cryo-EM Resolution Barriers to Facilitate Article Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery, Cell. 165 (2016) 1–10. doi:10.1016/j.cell.2016.05.040.

[6] Y.Z. Tan, S. Aiyer, M. Mietzsch, J.A. Hull, R. McKenna, J. Grieger, R.J. Samulski, T.S. Baker, M. Agbandje-McKenna, D. Lyumkis, Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant, Nat. Commun. 9 (2018) 1–11. doi:10.1038/s41467-018-06076-6.

[7] S. Subramaniam, Bridging the imaging gap: Visualizing subcellular architecture with electron tomography, Curr. Opin. Microbiol. 8 (2005) 316–322. doi:10.1016/j.mib.2005.04.012.

[8] T.S. Baker, R. Henderson, Electron cryomicroscopy of biological macromolecules, in: Int. Tables Crystallogr., 2012: pp. 593–614.

[9] C.M. Oikonomou, G.J. Jensen, Cellular Electron Cryotomography : Toward Structural Biology In Situ, Annu Rev Biochem. (2017) 1–24. doi:10.1146/annurev-biochem-061516-044741.

[10] J. Walz, D. Typke, M. Nitsch, A.J. Koster, R. Hegerl, W. Baumeister, Electron Tomography of Single Ice-Embedded Macromolecules: Three-Dimensional Alignment and Classification, 395 (1997) 387–395.

[11] B.A. Himes, P. Zhang, emClarity: Software for High Resolution Cryo-electron Tomography and Sub-tomogram Averaging, Nat. Methods. (2018). doi:10.1101/231605.

[12] J. Frank, The electron microscope as a structure projector, 2006. doi:10.1007/978-0-387-69008-7_4.

[13] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A.W. McDowall, P. Schultz, Cryo-electron microscopy of vitrified specimens, Q. Rev. Biophys. 21 (1988) 129. doi:10.1017/S0033583500004297.

[14] P.W. Hawkes, J.C.H.H. Spence, Science of Microscopy, in: Intergovernmental Panel on Climate Change (Ed.), Clim. Chang. 2013 - Phys. Sci. Basis, Cambridge University Press, Cambridge, 2007: pp. 1–30. doi:10.1017/CBO9781107415324.004.

[15] M. De Graef, Introduction to Conventional Transmission Electron Microscopy, Cambridge University Press, 2003.

[16] M. Vulović, L.M. Voortman, L.J. Van Vliet, B. Rieger, When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-EM, Ultramicroscopy. 136 (2014) 61–66. doi:10.1016/j.ultramic.2013.08.002.

[17] P.J.B. Koeck, A. Karshikoff, Limitations of the linear and the projection approximations in three-dimensional transmission electron microscopy of fully hydrated proteins, J. Microsc. 259 (2015) 197–209. doi:10.1111/jmi.12253.

[18] T. Mulvey, Origins Rand Historical Development of the Electron Microscope, Brittish J. Appl. Phys. 13 (1962) 197. doi:10.1088/0508-3443/13/5/303.

[19] O. Scherzer, The Theoretical Resolution Limit of the Electron Microscope, J. Appl. Phys. 20 (1949) 20–29. doi:10.1063/1.1698233.

[20] C. Hetherington, Aberration correction for TEM, Mater. Today. 7 (2004) 50–55. doi:10.1016/S1369-7021(04)00571-1.

[21] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, K.H. Herrmann, Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms, Ultramicroscopy. 3 (1978) 49–60. doi:10.1016/S0304-3991(78)80006-0.

[22] F. Zemlin, A practical procedure for alignment of a high resolution electron microscope, Ultramicroscopy. 4 (1979) 241–245. doi:10.1016/S0304-3991(79)90301-2.

[23] R. Yan, K. Li, W. Jiang, Defocus and magnification dependent variation of TEM image astigmatism, Sci. Rep. 8 (2018). doi:10.1038/s41598-017-18820-x.

[24] D.J. Smith, W.O.O. Saxton, M.A.A. O'Keefe, G.J.J. Wood, W.M.M. Stobbs, The importance of beam alignment and crystal tilt in high resolution electron microscopy, Ultramicroscopy. 11 (1983) 263–281. doi:10.1016/0304-3991(83)90006-2.

[25] R. Henderson, J.M. Baldwin, K.H. Downing, J. Lepault, F. Zemlin, Structure of purple membrane from halobacterium halobium: recording, measurement and evaluation of electron micrographs at 3.5 Åresolution, Ultramicroscopy. 19 (1986) 147–178. doi:10.1016/0304-3991(86)90203-2.

[26] C. Toyoshima, N. Unwin, Contrast transfer for frozen-hydrated specimens: determination from pairs of defocused images., Ultramicroscopy. 25 (1988) 279–291.

[27] J. Frank, Electron Microscopy of Macromolecular Assemblies, in: Three-Dimensional Electron Microsc. Macromol. Assem., Oxford University Press, 2006: pp. 15–69.

[28] C. V. Sindelar, N. Grigorieff, An adaptation of the Wiener filter suitable for analyzing images of isolated single particles, J. Struct. Biol. 176 (2011) 60–74. doi:10.1016/j.jsb.2011.06.010.

[29] R.. A. Crowther, D. DeRosier, A. Klug, The reconstruction of a three-dimensional structure from projections and its application to electron microscopy, Proc. R. Soc. London A Math. Phys. Eng. Sci. 317 (1970) 319–340. doi:10.1098/rspa.1970.0119.

[30] C.K. Cassidy, B.A. Himes, F.J. Alvarez, J. Ma, G. Zhao, J.R. Perilla, K. Schulten, P. Zhang, CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling., Elife. 4 (2015). doi:10.7554/eLife.08419.

[31] J.F. Conway, N. Cheng, A. Zlotnick, P.T. Wingfield, S.J. Stahl, A.C. Steven, Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy, Nature. 386 (1997) 91–94. doi:10.1038/386091a0.

[32] D. Lyumkis, A.F. Brilot, D.L. Theobald, N. Grigorieff, Likelihood-based classification of cryo-EM images using FREALIGN, J. Struct. Biol. 183 (2013) 377–388. doi:10.1016/j.jsb.2013.07.005.

[33] S.H.W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination, J. Struct. Biol. 180 (2012) 519–530. doi:10.1016/j.jsb.2012.09.006.

[34] R.M. Glaeser, R.J. Hall, Reaching the information limit in cryo-EM of biological macromolecules: experimental aspects., Biophys. J. 100 (2011) 2331–7. doi:10.1016/j.bpj.2011.04.018.

[35] Y. Cheng, N. Grigorieff, P.A. Penczek, T. Walz, A primer to single-particle cryo-electron microscopy, Cell. 161 (2015) 439–449. doi:10.1016/j.cell.2015.03.050.

[36] A.S. Frangakis, J. Böhm, F. Förster, S. Nickell, D. Nicastro, D. Typke, R. Hegerl, W. Baumeister, Identification of macromolecular complexes in cryoelectron tomograms of phantom cells, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 14153–14158. doi:10.1073/pnas.172520299.

[37] A. Bartesaghi, P. Sprechmann, J. Liu, G. Randall, G. Sapiro, S. Subramaniam, Classification and 3D averaging with missing wedge correction in biological electron tomography, J. Struct. Biol. 162 (2008) 436–450. doi:10.1016/j.jsb.2008.02.008.

[38] H. Winkler, 3D reconstruction and processing of volumetric data in cryo-electron tomography, J. Struct. Biol. 157 (2007) 126–137. doi:10.1016/j.jsb.2006.07.014.

[39] M. Stölken, F. Beck, T. Haller, R. Hegerl, I. Gutsche, J.-M. Carazo, W. Baumeister, S.H.W.W. Scheres, S. Nickell, Maximum likelihood based classification of electron tomographic data, J. Struct. Biol. 173 (2011) 77–85. doi:10.1016/j.jsb.2010.08.005.

[40] T.A.M. Bharat, C.J. Russo, J. Löwe, L.A. Passmore, S.H.W. Scheres, Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging, Structure. 23 (2015) 1743–1753. doi:10.1016/j.str.2015.06.026.

[41] W. Wan, J.A.G. Briggs, Cryo-Electron Tomography and Subtomogram Averaging, 1st ed., Elsevier Inc., 2016. doi:10.1016/bs.mie.2016.04.014.

[42] J.-J. Fernandez, Computational methods for electron tomography., Micron. 43 (2012) 1010–30. doi:10.1016/j.micron.2012.05.003.

[43] T.A.M. Bharat, S.H.W. Scheres, Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION, Nat. Protoc. 11 (2016) 2054–2065. doi:10.1038/nprot.2016.124.

[44] T. Zeev-Ben-Mordehai, D. Vasishtan, A. Hernández Durán, B. Vollmer, P. White, A. Prasad Pandurangan, C.A. Siebert, M. Topf, K. Grünewald, Two distinct trimeric conformations of natively membrane-anchored full-length herpes simplex virus 1 glycoprotein B., Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 4176–4181. doi:10.1073/pnas.1523234113.

[45] P.A. Penczek, J. Frank, C.M.T. Spahn, A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation, J. Struct. Biol. 154 (2006) 184–194. doi:10.1016/j.jsb.2005.12.013.

[46] P.A. Penczek, C. Yang, J. Frank, C.M.T. Spahn, Estimation of variance in single-particle reconstruction using the bootstrap technique, J. Struct. Biol. 154 (2006) 168–183. doi:10.1016/j.jsb.2006.01.003.

[47] H.Y. Liao, Y. Hashem, J. Frank, Efficient Estimation of Three-Dimensional Covariance and its Application in the Analysis of Heterogeneous Samples in Cryo-Electron Microscopy, Structure. 23 (2015) 1129–1137. doi:10.1016/j.str.2015.04.004.

[48] F.K.M. Schur, M. Obr, W.J.H. Hagen, W. Wan, A.J. Jakobi, J.M. Kirkpatrick, C. Sachse, H.-G. Kräusslich, J.A.G. Briggs, H.-G. Krausslich, J.A.G. Briggs, H.-G. Kräusslich, J.A.G. Briggs, H.-G. Krausslich, J.A.G. Briggs, An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation, Science (80-.). 353 (2016) 506–508. doi:10.1126/science.aaf9620.

[49] F.K.M. Schur, W.J.H. Hagen, M. Rumlová, T. Ruml, B. Müller, H.-G. Kräusslich, J. a G. Briggs, Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution., Nature. 517 (2015) 505–508. doi:10.1038/nature13838.

[50] F.K.M. Schur, W.J.H. Hagen, A. de Marco, J. a G. Briggs, Determination of protein structure at 8.5Åresolution using cryo-electron tomography and sub-tomogram averaging., J. Struct. Biol. 184 (2013) 394–400. doi:10.1016/j.jsb.2013.10.015.

[51] M. Khoshouei, S. Pfeffer, W. Baumeister, F. Förster, R. Danev, Subtomogram analysis using the Volta phase plate, J. Struct. Biol. 197 (2017) 94–101. doi:10.1016/j.jsb.2016.05.009.

[52] S. Mattei, B. Glass, W.J.H. Hagen, H.-G. Kräusslich, J.A.G. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions, Science (80-.). 354 (2016) 1434–1437. doi:10.1126/science.aah4972.

[53] S. Pfeffer, L. Burbaum, P. Unverdorben, M. Pech, Y. Chen, R. Zimmermann, R. Beckmann, F. Förster, Structure of the native Sec61 protein-conducting channel, Nat. Commun. 6 (2015) 8403. doi:10.1038/ncomms9403.

[54] L.G. Trabuco, E. Villa, E. Schreiner, C.B. Harrison, K. Schulten, Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography, Methods. 49 (2009) 174–180. doi:10.1016/j.ymeth.2009.04.005.

[55] S. Maji, R. Shahoei, K. Schulten, J. Frank, Quantitative Characterization of Domain Motions in Molecular Machines, J. Phys. Chem. B. (2017) acs.jpcb.6b10732. doi:10.1021/acs.jpcb.6b10732.

[56] L.M. Voortman, M. Vulovi??, M. Maletta, A. Voigt, E.M. Franken, A. Simonetti, P.J. Peters, L.J. van Vliet, B. Rieger, Quantifying resolution limiting factors in subtomogram averaged cryo-electron tomography using simulations, J. Struct. Biol. 187 (2014) 103–111. doi:10.1016/j.jsb.2014.06.007.

[57] M. Kudryashev, D. Castaño-Díez, H. Stahlberg, Limiting Factors in Single Particle Cryo Electron Tomography, Comput. Struct. Biotechnol. J. 1 (2012) 1–6. doi:10.5936/csbj.201207002.

[58] D. Vanhecke, S. Asano, Z. Kochovski, R. Fernandez-Busnadiego, N. Schrod, W. Baumeister, V. Lučić, Cryo-electron tomography: Methodology, developments and biological applications, J. Microsc. 242 (2011) 221–227. doi:10.1111/j.1365-2818.2010.03478.x.

[59] A. Stewart, N. Grigorieff, Noise bias in the refinement of structures derived from single particles, Ultramicroscopy. 102 (2004) 67–84. doi:10.1016/j.ultramic.2004.08.008.

[60] R. Henderson, A. Sali, M.L. Baker, B. Carragher, B. Devkota, K.H. Downing, E.H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S.J. Ludtke, O. Medalia, P.A. Penczek, P.B. Rosenthal, M.G. Rossmann, M.F. Schmid, G.F. Schröder, A.C. Steven, D.L. Stokes, J.D. Westbrook, W. Wriggers, H. Yang, J. Young, H.M. Berman, W. Chiu, G.J. Kleywegt, C.L. Lawson, Outcome of the first electron microscopy validation task force meeting, Structure. 20 (2012) 205–214. doi:10.1016/j.str.2011.12.014.

[61] S.H.W. Scheres, S. Chen, Prevention of overfitting in cryo-EM structure determination, Nat. Methods. 9 (2012) 853–854. doi:10.1038/nmeth.2115.

[62] S.C. Murray, J. Flanagan, O.B. Popova, W. Chiu, S.J. Ludtke, I.I. Serysheva, Ways & Means Validation of Cryo-EM Structure of IP 3 R1 Channel, Struct. Des. 21 (2013) 900–909. doi:10.1016/j.str.2013.04.016.

[63] P.B. Rosenthal, R. Henderson, Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy, J. Mol. Biol. 333 (2003) 721–745. doi:10.1016/j.jmb.2003.07.013.

[64] M. Van Heel, M. Schatz, Fourier shell correlation threshold criteria, J. Struct. Biol. 151 (2005) 250–262. doi:10.1016/j.jsb.2005.05.009.

[65] F. Amat, D. Castaño-Diez, A. Lawrence, F. Moussavi, H. Winkler, M. Horowitz, Alignment of cryo-electron tomography datasets, Methods Enzymol. 482 (2010) 343–367. doi:10.1016/S0076-6879(10)82014-2.

[66] D. Castaño-Díez, A. Al-Amoudi, A.M. Glynn, A. Seybert, A.S. Frangakis, Fiducial-less alignment of cryo-sections, J. Struct. Biol. 159 (2007) 413–423. doi:10.1016/j.jsb.2007.04.014.

[67] D. Castaño-Díez, M. Scheffer, A. Al-Amoudi, A.S. Frangakis, Alignator: A GPU powered software package for robust fiducial-less alignment of cryo tilt-series, J. Struct. Biol. 170 (2010) 117–126. doi:10.1016/j.jsb.2010.01.014.

[68] S.S. Brandt, Markerless Alignment in Electron Tomography, in: J. Frank (Ed.), Electron Tomogr. Methods Three-Dimensional Vis. Struct. Cell, Springer New York, New York, NY, 2006: pp. 187–215. doi:10.1007/978-0-387-69008-7_7.

[69] A.J. Noble, S.M. Stagg, Automated batch fiducial-less tilt-series alignment in Appion using Protomo, J. Struct. Biol. 192 (2015) 270–278. doi:10.1016/j.jsb.2015.10.003.

[70] S.H. w Scheres, Beam-induced motion correction for sub-megadalton cryo-EM particles, Elife. 3 (2014) e03665. doi:10.7554/eLife.03665.

[71] D.N. Mastronarde, Fiducial Marker and Hybrid Alignment Methods for Single- and Double-axis Tomography, in: J. Frank (Ed.), Electron Tomogr. Methods Three-Dimensional Vis. Struct. Cell, Springer New York, New York, NY, 2006: pp. 163–185. doi:10.1007/978-0-387-69008-7_6.

[72] S. Chen, G. McMullan, A.R. Faruqi, G.N. Murshudov, J.M. Short, S.H.W. Scheres, R. Henderson, High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy., Ultramicroscopy. 135 (2013) 24–35. doi:10.1016/j.ultramic.2013.06.004.

[73] T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6., Elife. 4 (2015) e06980. doi:10.7554/eLife.06980.

[74] K.V. Fernando, S.D. Fuller, Determination of astigmatism in TEM images, J. Struct. Biol. 157 (2007) 189–200. doi:10.1016/j.jsb.2006.07.021.

[75] J.J. Fernández, S. Li, R. a Crowther, CTF determination and correction in electron cryotomography., Ultramicroscopy. 106 (2006) 587–96. doi:10.1016/j.ultramic.2006.02.004.

[76] J.A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in electron microscopy, J. Struct. Biol. 142 (2003) 334–347. doi:10.1016/S1047-8477(03)00069-8.

[77] A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs, J. Struct. Biol. 192 (2015) 216–221. doi:10.1016/j.jsb.2015.08.008.

[78] Q. Xiong, M.K. Morphew, C.L. Schwartz, A.H. Hoenger, N. David, CTF Determination and Correction for Low Dose Tomographic Tilt Series, J. Struct. Biol. 168 (2010) 378–387. doi:10.1016/j.jsb.2009.08.016.CTF.

[79] J.R. Kremer, D.N. Mastronarde, J.R. McIntosh, Computer visualization of three-dimensional image data using IMOD., J. Struct. Biol. 116 (1996) 71–76. doi:10.1006/jsbi.1996.0013.

[80] P.A. Penczek, J. Fang, X. Li, Y. Cheng, J. Loerke, C.M.T. Spahn, CTER-Rapid estimation of CTF parameters with error assessment, Ultramicroscopy. 140 (2014) 9–19. doi:10.1016/j.ultramic.2014.01.009.

[81] C. V. Sindelar, N. Grigorieff, Optimal noise reduction in 3D reconstructions of single particles using a volume-normalized filter, J. Struct. Biol. 180 (2012) 26–38. doi:10.1016/j.jsb.2012.05.005.

[82] C.A. Diebolder, F.G.A. Faas, A.J. Koster, R.I. Koning, Conical fourier shell correlation applied to electron tomograms, J. Struct. Biol. 190 (2015) 215–223. doi:10.1016/j.jsb.2015.03.010.

[83] J.M. Heumann, A. Hoenger, D.N. Mastronarde, Clustering and variance maps for cryo-electron tomography using wedge-masked differences, J. Struct. Biol. 175 (2011) 288–299. doi:10.1016/j.jsb.2011.05.011.

[84] B.K. Alsberg, Multiscale cluster analysis, Anal. Chem. 71 (1999) 3092–3100. doi:10.1021/ac9811672.

[85] R. Marabini, S.J. Ludtke, S.C. Murray, W. Chiu, J.M. de la Rosa-Trev??n, A. Patwardhan, J.B. Heymann, J.M. Carazo, The Electron Microscopy eXchange (EMX) initiative, J. Struct. Biol. 194 (2016) 156–163. doi:10.1016/j.jsb.2016.02.008.

[86] X.C. Bai, I.S. Fernandez, G. McMullan, S.H.W. Scheres, Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles, Elife. 2013 (2013) 2–13. doi:10.7554/eLife.00461.

[87] R.R. Gutell, B. Weiser, C.R. Woese, H.F. Noller, Comparative anatomy of 16-S-like ribosomal RNA., Prog. Nucleic Acid Res. Mol. Biol. 32 (1985) 155–216.

[88] R. Beckmann, C.M.T. Spahn, N. Eswar, P.A. Penczek, A. Sali, J. Frank, Architecture of the Protein-Conducting Channel Associated with the Translating 80S Ribosome, Cell. 107 (2001) 361–372.

[89] S. Mohan, H.F. Noller, Recurring RNA structural motifs underlie the mechanics of L1 stalk movement, Nat. Commun. 8 (2017) 14285. doi:10.1038/ncomms14285.

[90] C.M.T. Spahn, M.G. Gomez-Lorenzo, R.A. Grassucci, R. Jørgensen, G.R. Andersen, R. Beckmann, P.A. Penczek, J.P.G. Ballesta, J. Frank, Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation, EMBO J. 23 (2004) 1008–1019. doi:10.1038/sj.emboj.7600102.

[91] D.N. Wilson, K.H. Nierhaus, The E-site story: the importance of maintaining two tRNAs on the ribosome during protein synthesis, Cell. Mol. Life Sci. 63 (2006) 2725–2737. doi:10.1007/s00018-006-6125-4.

[92] T. V. Budkevich, J. Giesebrecht, E. Behrmann, J. Loerke, D.J.F.F. Ramrath, T. Mielke, J. Ismer, P.W. Hildebrand, C.S. Tung, K.H. Nierhaus, K.Y. Sanbonmatsu, C.M.T.T. Spahn, Regulation of the Mammalian Elongation Cycle by Subunit Rolling: A Eukaryotic-Specific Ribosome Rearrangement, Cell. 158 (2014) 121–131. doi:10.1016/j.cell.2014.04.044.

[93] P.D. Abeyrathne, C.S. Koh, T. Grant, N. Grigorieff, A.A. Korostelev, Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome, Elife. 5 (2016) 1–31. doi:10.7554/eLife.14874.

[94] M.G. Gomez-Lorenzo, C.M. Spahn, R.K. Agrawal, R.A. Grassucci, P. Penczek, K. Chakraburtty, J.P. Ballesta, J.L. Lavandera, J.F. Garcia-Bustos, J. Frank, Three-dimensional cryo-electron microscopy localization of EF2 in the Saccharomyces cerevisiae 80S ribosome at 17.5 A resolution., EMBO J. 19 (2000) 2710–2718. doi:10.1093/emboj/19.11.2710.

[95] B. Chakraborty, R. Mukherjee, J. Sengupta, Structural insights into the mechanism of translational inhibition by the fungicide sordarin, J. Comput. Aided. Mol. Des. 27 (2013) 173–184. doi:10.1007/s10822-013-9636-8.

[96] M. Li, G.L. Hazelbauer, Core unit of chemotaxis signaling complexes., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 9390–9395. doi:10.1073/pnas.1104824108.

[97] P. Rampelotto, Extremophiles and Extreme Environments, Life. 3 (2013) 482–485. doi:10.3390/life3030482.

[98] J.S. Parkinson, in Bacterial Signaling, Prospects. (1992) 71–112.

[99] J. Adler, Chemotaxis in Bacteria Motile Escherichia coli migrate in bands that are, Adv. Sci. 153 (2008) 708–716. doi:10.1126/science.153.3737.708.

[100] R.M. Harshey, A. Toguchi, Spinning tails: Homologies among bacterial flagellar systems, Trends Microbiol. 4 (1996) 226–231. doi:10.1016/0966-842X(96)10037-8.

[101] H.C. Berg, D.A. Brown, Chemotaxis in Escherichia coli analysed by three-dimensional tracking, Nature. 239 (1972) 500–504. doi:10.1038/239500a0.

[102] J. Adler, A method for measuring chemotaxis and use of the method to determine optimum conditions for chemotaxis by Escherichia coli., J. Gen. Microbiol. 74 (1973) 77–91. doi:10.1099/00221287-74-1-77.

[103] R. Mesibov, G.W. Ordal, J. Adler, The range of attractant concentrations for bacterial chemotaxis and the threshold and size of response over this range. Weber law and related phenomena., J. Gen. Physiol. 62 (1973) 203–223. doi:10.1085/jgp.62.2.203.

[104] J.B. Stock, D.E. Koshland, Changing reactivity of receptor carboxyl groups during bacterial sensing, J. Biol. Chem. 256 (1981) 10826–10833.

[105] D. Sherris, J.S. Parkinson, Posttranslational processing of methyl-accepting chemotaxis proteins in Escherichia coli., Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 6051–5. doi:10.1073/pnas.78.10.6051.

[106] J.S. Parkinson, G.L. Hazelbauer, J.J. Falke, Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update, Trends Microbiol. 23 (2015) 257–66. doi:10.1016/j.tim.2015.03.003.

[107] P. Ames, C. a Studdert, R.H. Reiser, J.S. Parkinson, Collaborative signaling by mixed chemoreceptor teams in Escherichia coli., Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 7060–7065. doi:10.1073/pnas.092071899.

[108] C. a Studdert, J.S. Parkinson, Crosslinking snapshots of bacterial chemoreceptor squads., Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 2117–2122. doi:10.1073/pnas.0308622100.

[109] A. Briegel, X. Li, a. M. Bilwes, K.T. Hughes, G.J. Jensen, B.R. Crane, Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins, Proc. Natl. Acad. Sci. 109 (2012) 3766–3771. doi:10.1073/pnas.1115719109.

[110] A. Briegel, D.R. Ortega, E.I. Tocheva, K. Wuichet, Z. Li, S. Chen, A. Müller, C. V Iancu, G.E. Murphy, M.J. Dobro, I.B. Zhulin, G.J. Jensen, Universal architecture of bacterial chemoreceptor arrays., Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 17181–17186. doi:10.1073/pnas.0905181106.

[111] A.M. Bilwes, C.M. Quezada, L.R. Croal, B.R. Crane, M.I. Simon, Nucleotide binding by the histidine kinase CheA, Nat Struct Biol. 8 (2001) 353–360. doi:10.1038/86243.

[112] J. Bhatnagar, P.P. Borbat, A.M. Pollard, A.M. Bilwes, J.H. Freed, B.R. Crane, Structure of the ternary complex formed by a chemotaxis receptor signaling domain, the CheA histidine kinase, and the coupling protein CheW As determined by pulsed dipolar ESR spectroscopy, Biochemistry. 49 (2010) 3824–3841. doi:10.1021/bi100055m.

[113] S.L. Gloor, J.J. Falke, Thermal domain motions of CheA kinase in solution: Disulfide trapping reveals the motional constraints leading to trans-autophosphorylation, Biochemistry. 48 (2009) 3631–3644. doi:10.1021/bi900033r.

[114] R.M. Weis, D.E. Koshland, Reversible receptor methylation is essential for normal chemotaxis of Escherichia coli in gradients of aspartic acid., Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 83–7. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=279486&tool=pmcentrez&rendertype=abstract.

[115] R.C. Stewart, A.F. Roth, F.W. Dahlquist, Mutations that affect control of the methylesterase activity of CheB, a component of the chemotaxis adaptation system in Escherichia coli, J. Bacteriol. 172 (1990) 3388–3399.

[116] S. Schulmeister, K. Grosse, V. Sourjik, Effects of receptor modification and temperature on dynamics of sensory complexes in Escherichia coli chemotaxis, BMC Microbiol. 11 (2011) 222. doi:10.1186/1471-2180-11-222.

[117] X. Fu, B. a Himes, D. Ke, W.J. Rice, J. Ning, P. Zhang, Controlled Bacterial Lysis for Electron Tomography of Native Cell Membranes, Structure. 22 (2014) 1875–1882. doi:10.1016/j.str.2014.09.017.

[118] D.E. Martin, K. Salzwedel, G.P. Allaway, Bevirimat: A novel maturation inhibitor for the treatment of HIV-1 infection, Antivir. Chem. Chemother. (2008). doi:10.1177/095632020801900301.

[119] W.H. Organization, HIV / AIDS factsheet, (2017) 1–6. http://www.who.int/mediacentre/factsheets/fs360/en/ (accessed June 26, 2017).

[120] A.J. Murray, K.J. Kwon, D.L. Farber, R.F. Siliciano, The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence, J. Immunol. 197 (2016) 407–417. doi:10.4049/jimmunol.1600343.

[121] J.M. Coffin, HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy, Science (80-.). 267 (1995) 483 LP-489. http://science.sciencemag.org/content/267/5197/483.abstract.

[122] R.P. Smyth, M.P. Davenport, J. Mak, The origin of genetic diversity in HIV-1, Virus Res. 169 (2012) 415–429. doi:10.1016/j.virusres.2012.06.015.

[123] E. De Clercq, Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV, Int. J. Antimicrob. Agents. 33 (2009) 307–320. doi:10.1016/j.ijantimicag.2008.10.010.

[124] A.C. Collier, R.W. Coombs, D.A. Schoenfeld, R.L. Bassett, J. Timpone, A. Baruch, M. Jones, K. Facey, C. Whitacre, V.J. McAuliffe, H.M. Friedman, T.C. Merigan, R.C. Reichman, C. Hooper, L. Corey, Treatment of Human Immunodeficiency Virus Infection with Saquinavir, Zidovudine, and Zalcitabine, N. Engl. J. Med. 334 (1996) 1011–1018. doi:10.1056/NEJM199604183341602.

[125] E.J. Arts, D.J. Hazuda, HIV-1 Antiretroviral Drug Therapy BASIC PRINCIPLES OF ANTIRETROVIRAL ir, Cold Spring Harb. Perspect. Med. 2 (2012) 1–24. doi:10.1101/cshperspect.a007161.

[126] A.M. Margolis, H. Heverling, P.A. Pham, A. Stolbach, A Review of the Toxicity of HIV Medications, J. Med. Toxicol. 10 (2014) 26–39. doi:10.1007/s13181-013-0325-8.

[127] D. Wang, W. Lu, F. Li, Pharmacological intervention of HIV-1 maturation, Acta Pharm. Sin. B. 5 (2015) 493–499. doi:10.1016/j.apsb.2015.05.004.

[128] J.A.G. Briggs, H.G. Kräusslich, The molecular architecture of HIV, J. Mol. Biol. 410 (2011) 491–500. doi:10.1016/j.jmb.2011.04.021.

[129] K. Wiegers, G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, H.G. Kräusslich, Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites., J. Virol. 72 (1998) 2846–2854. doi:10.1016/S0969-2126(02)00720-7.

[130] G. Received, E.S. August, H. Resolution, E. Microscope, C. Cb, ' I, 13 (1984) 57–70.

[131] T. Kanamoto, Y. Kashiwada, K. Kanbara, K. Gotoh, M. Yoshimori, T. Goto, K. Sano, H. Nakashima, Anti-human immunodeficiency virus activity of YK-FH312 (a betulinic acid derivative), a novel compound blocking viral maturation, Antimicrob. Agents Chemother. 45 (2001) 1225–1230. doi:10.1128/AAC.45.4.1225-1230.2001.

[132] F. Li, R. Goila-Gaur, K. Salzwedel, N.R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D.E. Martin, J.M. Orenstein, G.P. Allaway, E.O. Freed, C.T. Wild, PA-457: A potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing, Proc. Natl. Acad. Sci. 100 (2003) 13555–13560. doi:10.1073/pnas.2234683100.

[133] W.S. Blair, J. Cao, J. Fok-Seang, P. Griffin, J. Isaacson, R.L. Jackson, E. Murray, A.K. Patick, Q. Peng, M. Perros, C. Pickford, H. Wu, S.L. Butler, New small-molecule inhibitor class targeting human immunodeficiency virus type 1 virion maturation, Antimicrob. Agents Chemother. 53 (2009) 5080–5087. doi:10.1128/AAC.00759-09.

[134] P.F. Smith, A. Ogundele, A. Forrest, J. Wilton, K. Salzwedel, J. Doto, G.P. Allaway, D.E. Martin, Phase I and II study of the safety, virologic effect, and pharmacokinetics/pharmacodynamics of single-dose 3-O-(3???3???-dimethylsuccinyl)betulinic acid (bevirimat) against human immunodeficiency virus Infection, Antimicrob. Agents Chemother. 51 (2007) 3574–3581. doi:10.1128/AAC.00152-07.

[135] K. Van Baelen, K. Salzwedel, E. Rondelez, V. Van Eygen, S. De Vos, A. Verheyen, K. Steegen, Y. Verlinden, G.P. Allaway, L.J. Stuyver, Susceptibility of human immunodeficiency virus type 1 to the maturation inhibitor bevirimat is modulated by baseline polymorphisms in Gag spacer peptide, Antimicrob. Agents Chemother. 53 (2009) 2185–2188. doi:10.1128/AAC.01650-08.

[136] O.N. Prophylaxis, Research letters, 128 (2012) 136–137.

[137] K. Wiegers, G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, H.G. Kräusslich, Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites., J. Virol. 72 (1998) 2846–2854. doi:10.1016/S0969-2126(02)00720-7.

[138] P.W. Keller, C.S. Adamson, J.B. Heymann, E.O. Freed, A.C. Steven, HIV-1 Maturation Inhibitor Bevirimat Stabilizes the Immature Gag Lattice, J. Virol. 85 (2011) 1420–1428. doi:10.1128/JVI.01926-10.

[139] J.M. Wagner, K.K. Zadrozny, J. Chrustowicz, M.D. Purdy, M. Yeager, B.K. Ganser-Pornillos, O. Pornillos, Crystal structure of an HIV assembly and maturation switch, Elife. 5 (2016) 1–18. doi:10.7554/eLife.17063.

[140] J. Fontana, P.W. Keller, E. Urano, S.D. Ablan, A.C. Steven, E.O. Freed, Identification of an HIV-1 Mutation in Spacer Peptide 1 That Stabilizes the Immature CA-SP1 Lattice, J. Virol. 90 (2016) 972–978. doi:10.1128/JVI.02204-15.

[141] W.J.H. Hagen, W. Wan, J.A.G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging, J. Struct. Biol. 197 (2017) 191–198. doi:10.1016/j.jsb.2016.06.007.

[142] P.L. Chiu, X. Li, Z. Li, B. Beckett, A.F. Brilot, N. Grigorieff, D.A. Agard, Y. Cheng, T. Walz, Evaluation of super-resolution performance of the K2 electron-counting camera using 2D crystals of aquaporin-0, J. Struct. Biol. 192 (2015) 163–173. doi:10.1016/j.jsb.2015.08.015.

[143] B. Turoňová, F.K.M.M.M. Schur, W. Wan, J.A.G.G.G. Briggs, Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4Å, J. Struct. Biol. 199 (2017) 187–195. doi:10.1016/j.jsb.2017.07.007.

[144] T. Grant, A. Rohou, N. Grigorieff, cis TEM , user-friendly software for single- particle image processing, (2018) 1–24.

[145] R.A. Dick, K.K. Zadrozny, C. Xu, F.K.M. Schur, T.D. Lyddon, C.L. Ricana, J.M. Wagner, J.R. Perilla, B.K. Ganser-Pornillos, M.C. Johnson, O. Pornillos, V.M. Vogt, Inositol phosphates are assembly co-factors for HIV-1., Nature. (2018). doi:10.1038/s41586-018-0396-4.

[146] R.R. Meyer, A.I. Kirkland, W.O. Saxton, A new method for the determination of the wave aberration function for high-resolution TEM. 2. Measurement of the antisymmetric aberrations, Ultramicroscopy. 99 (2004) 115–123. doi:10.1016/j.ultramic.2003.11.001.

[147] A.M. Anger, J.P. Armache, O. Berninghausen, M. Habeck, M. Subklewe, D.N. Wilson, R. Beckmann, Structures of the human and Drosophila 80S ribosome, Nature. 497 (2013) 80–85. doi:10.1038/nature12104.

[148] J. Ning, G. Erdemci-Tandogan, E.L. Yufenyuy, J. Wagner, B.A. Himes, G. Zhao, C. Aiken, R. Zandi, P. Zhang, In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway, Nat. Commun. 7 (2016) 13689. doi:10.1038/ncomms13689.

[149] B.E. Bammes, J. Jakana, M.F. Schmid, W. Chiu, Radiation damage effects at four specimen temperatures from 4 to 100 K, J. Struct. Biol. 169 (2010) 331–341. doi:10.1016/j.jsb.2009.11.001.

[150] E.R. Wright, C. V. Iancu, W.F. Tivol, G.J. Jensen, Observations on the behavior of vitreous ice at ???82 and???12 K, J. Struct. Biol. 153 (2006) 241–252. doi:10.1016/j.jsb.2005.12.003.

[151] J. Zhu, P.A. Penczek, R. Schröder, J. Frank, Three-Dimensional Reconstruction with Contrast Transfer Function Correction from Energy-Filtered Cryoelectron Micrographs: Procedure and Application to the 70SEscherichia coliRibosome, J. Struct. Biol. 118 (1997) 197–219. doi:10.1006/jsbi.1997.3845.

[152] K. Ishizuka, N. Uyeda, A New Theoretical and Practical Approach to the Multislice Method, Acta Crystallogr. 2 (1977) 740–749.