2010-12-28

Strategies to Resolve the Three-Dimensional Structure of the Genome of Small Single-Stranded Icosahedral Viruses

Eduardo Sanz Garcia
Brigham Young University - Provo

Strategies to Resolve the Three-Dimensional Structure of the Genome of Small Single-Stranded Icosahedral Viruses

by

Eduardo Sanz García

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

David M. Belnap, Chair
Gregory F. Burton
Allen R. Buskirk
Steven R. Herron
Richard K. Watt

Department of Chemistry and Biochemistry

Brigham Young University

April 2011

ABSTRACT

Strategies to Resolve the Three-Dimensional Structure of the

Genome of Small Single-Stranded Icosahedral Viruses

Eduardo Sanz García

Department of Chemistry and Biochemistry

Doctor of Philosophy

The aim of this study is the three-dimensional structural characterization of the genome packaging inside viral capsids via cryo-electron microscopy and three-dimensional reconstruction. The genome of some single-stranded viruses can be densely packaged within their capsid shells. Several stretches of the genome are known to adopt stable secondary structures, however, to date, little is known about the three-dimensional organization of the genome inside their capsid shells. Two techniques have been developed to facilitate the structural elucidation of genome packaging: the asymmetric random-model method, and the symmetry-mismatch, random model method. Both techniques were successfully tested with model and experimental data. The new algorithms were applied to study the genome structure of poliovirus and satellite tobacco mosaic virus. We have not yet found a consistent structure for the two genomes. Nevertheless, we have found that the genome of satellite tobacco mosaic genome is very stable, supporting a model where the RNA acts as a scaffold, with potential implications in capsid stability and assembly.

Keywords: Asymmetric reconstruction, Cryo-electron microscopy, Poliovirus, Satellite tobacco mosaic virus, Single-particle reconstruction

ACKNOWLEDGMENTS

I would like to acknowledge David Belnap for his advise and support on this project. I also wish to thank Aaron Stewart, and Peter Shen who are coauthors on the asymmetric random-model method (Chapters 2, 3) and the icosahedral reconstruction of STMV (Chapter 6), respectively. I thank Bernard Heymann, Milagros Rodríguez, Jun Lin, Robert Swenson, David Eng, Chuan-Hsi Hung, Kristin Pande, Benjamin Leishman, Roane Noel, Jennifer Nelson, and Rebecca Baum for their help and support. I want to thank the members of my committee, Allen Buskirk, Gregory Burton, Steven Herron, and Richard Watt, for their suggestions, guidance and insightful comments on this manuscript.

I thank the following individual for sharing their images or samples:

- J. Sengupta and J. Frank for the *E. coli* 70S ribosome images
- P. Shen, D. Enderlein, and R. Johne for the avian polyomavirus images
- T. Baker, G. Wang, M. Cahoon, and R. Swenson for the octahedral mouse polyomavirus VP1 images
- The GroEL data set was provided by the National Resource for Automated Molecular Microscopy (NIH P41 RR-17573).
- J. Hogle laboratory for providing the poliovirus sample and the 160S poliovirus images
- A. McPherson for providing the STMV purified virus

I thank J. Farrer, and M. Standing for support in microscopy.

I relied heavily on the Fulton Supercomputing Laboratory (Brigham Young University) for computational work. The project was supported by Brigham Young University institutional funds and by grant R21AI081221 from the National Institute of Allergy and Infectious Diseases. The content of this thesis is solely the responsibility of the author and does not necessarily represent official views of NIAID or the National Institutes of Health.

"It is a truism that nucleic acids are conformationally more complex than proteins."

S. Neidle, Oxford Handbook of Nucleic Acid Structure. Oxford University Press, ed. 1998.

Contents

List of Figures

Nomenclature

The page number refers to the first time the term is used.

(+) Nucleic acid with positive sense, 1

(-) Nucleic acid with negative sense, 1

2D Two-dimensional, 8

3D Three-dimensional, 8

ex virio Genome in solution, 1

in virio Genome fully encapsidated, 1

AFM Atomic force microscopy, 5

ARMM Asymmetric random-model method, 17

CC Correlation coefficient, 18

cryoEM Cryo-electron microscopy, 1

CTF Contrast transfer function, 29

DNA Deoxyribonucleic acid, 1

DPR Differential phase-residual, 30

ds Double-stranded nucleic acid, 1

EMDB Electron Microscopy Data Bank, 28

FSC Fourier shell correlation, 30

HIV Human Immunodeficiency Virus, 4

IRES Internal ribosome entry site, 72

PCA Principal component analysis, 29

PDB Protein Data Bank, 89

PV Poliovirus, 86

$R_{(A+B)}$ Reliability index, 30

RMM Random-model method, 17

RNA Ribonucleic acid, 1

SD Standard deviation, 24

SMRMM Symmetry-mismatch, random-model method, 59

SNR Signal-to-noise ratio, 20

ss Single-stranded nucleic acid, 1

SSNR Spectral signal-to-noise ratio, 30

STMV Satellite tobacco mosaic virus, 14

VPg 5' genome linked protein, 14

Chapter 1

Introduction

1.1 Overview

Viruses encode their genetic information in the form of deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), which can be arranged as double-stranded (ds) or single-stranded (ss). Single-stranded nucleic acids, independent of their sense (+ or -), have the potential ability to produce compact and ordered structures [1–3]. The focus of this study is the structural determination of the genome of small icosahedral viruses.

Currently, our knowledge about the spatial organization of viral genomes, both in solution (*ex virio*) or fully encapsidated (*in virio*), is very limited. Some stretches of nucleic acid that follow the same symmetry of the viral capsid have been resolved at higher resolution by X-ray crystallography and at moderate resolution by cryo-electron microscopy (cryoEM, Table 1.1). These nucleic acid structures range from a few nucleotides to extensive networks (see satellite tobacco, Pariocoto and Flock viruses in Table 1.1). A common motif that is frequently observed in cryoEM structures of tailed bacteriophages and other ds DNA and RNA viruses is the arrangement of the genome in concentric "shells".

Other portions of the viral genomes, essential for the virus "life" cycle, are known to assume characteristic secondary structures. Among these sequences are the internal ribosome entry sites [4], replication signals [5] and encapsidation panhandles [6]. Recently, Watts *et al.* [7] reported, for the first time, the secondary structure determination of an entire human immunodeficiency virus (HIV) RNA genome. But to date, no tertiary structure of viral genome has been entirely described at moderate or high resolution (Table. 1.1).

Table 1.1 Viral genomes which three-dimensional structures have been partially characterized.[a]

Family	Nucleic Acid	Virus	Method[b]	Symmetry Applied	Observation[c]	Reference
Bromoviridae	(+) ssRNA	cowpea chlorotic mottle virus	X-ray	icosahedral	base "stacking" arrangement of a few nucleotides at protein subunit interfaces; ∼20% of genome	[8]
Comoviridae	ssRNA	bean pod mottle virus	X-ray	icosahedral	trefoil arrangement of nearly 20% of genome	[9]
Cystoviridae	dsRNA	bacteriophage $^{\phi}$6	cryoEM	icosahedral	concentric layers, outer layer perforated	[10]
Hepadnaviridae	ssRNA, dsDNA	hepatitis B virus	3D cryoEM	icosahedral	one shell near capsid, dodecahedral	[11–13]
Herpesviridae	dsDNA	herpes simplex virus, type 1	2D cryoEM	none	DNA "fingerprints" with 26 Å spacing in images	[14]
Leviviridae	ssRNA	bacteriophage AP205	3D cryoEM	icosahedral	network of RNA around 5-fold and 3-fold axes	[15]
		bacteriophage MS2	3D cryoEM, X-ray	icosahedral, asymmetry	Interaction between RNA operator and capsid protein across 2-fold axes	[15–18]
		bacteriophage PP7	3D cryoEM	icosahedral	Network of RNA around 5-fold and 3-fold axes	[15]
		bacteriophage Qβ	3D cryoEM	icosahedral	nearly identical to MS2	[15]
Microviridae	ssDNA	bacteriophage $_{\Phi}$X174	X-ray	icosahedral	11 nucleotides, only 1 out of 5 positions occupied	[19]
Myoviridae	dsDNA	bacteriophage T4	3D cryoEM	icosahedral, 5-fold rotational	concentric layers	[20, 21]
		bacteriophage $^{\phi}$KZ	3D cryoEM	icosahedral	concentric layers	[22]
Nodaviridae	ssRNA	black beetle virus	X-ray	icosahedral	10 nucleotides of duplex RNA	[23]

Continue on next page...

Table 1.1–Continued[a]

Family	Nucleic Acid	Virus	Method[b]	Symmetry Applied	Observation	Reference
		Flock House virus	X-ray, 3D cryoEM	icosahedral	20 nucleotides of duplex RNA, dodecahedral arrangement	[24, 25]
		Pariacoto virus	X-ray, 3D cryoEM	icosahedral	duplex RNA in dodecahedral arrangement	[26]
Papovaviridae	dsDNA	Murine polyomavirus	X-ray	icosahedral	no regular order	[27]
Parvoviridae	ssDNA	canine parvovirus	X-ray	icosahedral	11 nucleotides	[28]
		minute virus of mice	X-ray	icosahedral	29 nucleotides	[29]
Picornaviridae	(+) ssRNA	coxsackieviruses A9, A21, B3	X-ray	icosahedral	nucleotide stacked against conserved Trp	[30–32]
		human rhinovirus 16	X-ray	icosahedral	4 nucleotides stacked against 2 aromatic residues, including conserved Trp	[33]
		human rhinovirus 3, 14	X-ray	icosahedral	nucleotide stacked against conserved Trp	[34, 35]
		human rhinovirus 2	X-ray	icosahedral	nucleotide stacked against conserved Trp	[36]
		Mengo encephalomyelitis virus	X-ray	icosahedral	diffuse density against capsid, likely RNA	[37]
		poliovirus 2	X-ray	icosahedral	1 nucleotide stacked against conserved Trp, 2 nucleotides stacked against 2 other aromatic residues	[38]
		poliovirus 1, 3	X-ray	icosahedral	1 nucleotide stacked against conserved Trp, another nucleotide stacked against another aromatic residue	[39]
		swine vesicular disease virus	X-ray	icosahedral	nucleotide stacked against conserved Trp	[40]
Podoviridae	dsDNA	bacteriophage P22	3D cryoEM	icosahedral, asymmetry	concentric shells, coaxial rings, tight ring near connector	[41–43]
		bacteriophage T7	2D and 3D cryoEM	circular (2D), icosahedral and 5-fold (3D)	concentric rings (2D) and shells (3D) of DNA	[44–46]
		bacteriophage ϕ29	cryoEM, modeling	asymmetry, C5	concentric shells, folded toroidal conformation	[47–49]
Siphoviridae	dsDNA	bacteriophage T5	3D cryoEM	icosahedral	concentric shells	[50]
Tectiviridae	dsDNA	bacteriophage PRD1	3D cryoEM	icosahedral	concentric shells	[51]
Tombusviridae	(+) ssRNA	hibiscus chlorotic ringspot virus	3D cryoEM	icosahedral	shell inside capsid	[52]

Continue on next page...

Table 1.1–Continued[a]

Family	Nucleic Acid	Virus	Method[b]	Symmetry Applied	Observation	Reference
		red clover necrotic mottle virus	3D cryoEM	icosahedral	"cage"-like structure near capsid	[53]
		tomato bushy stunt virus	3D cryoEM	icosahedral	change in RNA with swelling	[54]
Tymoviridae	ssRNA	desmodium yellow mottle virus	X-ray	icosahedral	two ssRNA fragments of 7 and 2 nucleotides	[55]
		turnip yellow mosaic virus	3D cryoEM, X-ray	icosahedral	duplex RNA on 2-fold, single-stranded loops in capsomeres	[56, 57]
Reoviridae	dsRNA	bluetongue virus	X-ray	icosahedral	concentric shells	[58]
		reovirus	X-ray	icosahedral	concentric layers	[59]
		rotavirus	3D cryoEM	icosahedral	non-transcribing is dodecahedral, transcribing is on 5-fold axes	[60–62]
Satellite	ssRNA	satellite panicum mosaic virus	X-ray	icosahedral	dsRNA along 5-fold axis, ssRNA near N-terminus of capsid protein	[63, 64]
		satellite tobacco mosaic virus	X-ray, modeling, molecular dynamics	icosahedral	portion of genome ordered in X-ray density maps, modeled genome	[57, 65–67]
unassigned	dsDNA	bacteriophage epsilon15	3D cryoEM, modeling	asymmetry	coaxial coils, ring of DNA near portal, ∼90 base pairs in portal-core complex; models predict both coaxial spool and folded toroid structures are present	[48, 68]
	ssRNA	sobemoviruses	cryoEM	icosahedral	concentric layers	[69]
	ssRNA	cocksfoot mottle virus	X-ray	icosahedral	at least two nucleotides stacked against Trp residue of capsid	[70]

[a] List compiled by D. Belnap. Revised by E. Sanz García.
[b] "X-ray" refers to X-ray crystallography. "2D cryoEM" is two-dimensional cryo-EM. 2D cryo-EM involves analysis of 2D images, *e.g.* averaging of several images or observations on individual particle images. "3D cryoEM" is three-dimensional cryoEM. It involves computing a 3D reconstruction from 2D images.
[c] If a number of nucleotides is given, the number refers to the number of nucleotides per asymmetric unit. There are 60 asymmetric units per icosahedron

Several of bioinformatic, biochemical and structural studies support the hypothesis that the genome of certain viruses could have a single predominant structure:

1. The first 900 bases of the 5' end of the Human Immunodeficiency Virus (HIV) genome was shown to have almost identical secondary structures *in virio*, *ex virio*, or when the genome was *in vitro* transcribed and folded [71].

2. Bioinformatic studies have suggested that a significant population of viral genomes adopt large-scale ordered foldings [72, 73].

3. Evolutionary conservation of a compact structure appears to impose a considerable constraint to neutral mutation drift [72, 74, 75].

4. Atomic force microscopy (AFM) experiments have confirmed that *ex virio* genomes spontaneously assume spheroid shapes in the absence of the capsids [76]. The diameter of these spheroids correspond to the inner radius of the capsid shell [76].

5. Hybridization accessibility assays showed that, for certain viruses, nucleic-acid probes have limited access to portions of extracted viral genome, thus, indicating that such genomes fold into compact structure preventing the binding to the complementary probes [73].

Two additional constraints support the idea that *in virio* genomes adopt a single conformation. First, the genome has to fit within the limited volume inside the capsid. Second, the anchoring effect of the capsid protein has an important effect upon the genome structure [9, 17, 66, 77, 78].

1.2 Biological significance

Relevant functions have been attributed to the *structure* of viral genomes, including capsid assembly [78, 79], capsid stability [65] and infection persistence [72]. Because regulatory *structures* within viral genomes are responsible to activate viral replication, transcription, encapsidation, and immune responses, the secondary and tertiary structural characterization of viral genomes is an area of promising growth [7, 79].

The genome of certain viruses acts not only as a mere carrier of the genetic information but as an active player on the structural assembly of the virion. Interaction

between the capsid protein and the viral genome has been observed in numerous cases (Table 1.1). Anchoring of the genome to the viral capsids often involves electrostatic interactions between positively charged residues in the N or C termini of the coat proteins and negatively charge nucleic acids [57, 80–82]. Specific interactions have been reported between packaging signals present in the viral genome and capsid proteins [83–85]. In some instances, capsid proteins direct the folding of the genome, independently of the sequence and size of the nucleic acid [25, 86, 87].[1] Other times, the RNA component modulates the size and architecture of the capsid shells [24, 85, 88, 89].

The viral genome also plays a critical role in maintaining capsid strength and integrity. Molecular dynamic studies have suggested that the viral capsids might not be stable in the absence of genome [65]. Heterogeneous collection of particles have been observed when capsid proteins were expressed in the absence of wild-type viral genome [79, 90, 91]. Many investigators, including the author, have observed that empty viruses (without their genomes) are more prone to disruption and deformation [92]. In some particular instances viruses without genome have never or have rarely been observed [65].

The ability of viruses to persist in their natural hosts strongly correlates with the presence of ordered genome structures [72, 73]. Viruses that have been predicted to adopt large globular structure have been found to be able to escape intracellular defense mechanisms. However, at present, it is difficult to conceptualize what mechanism underlies the association between large-scale genome structure and host persistence.

The vast majority of the structures of nucleic acids are limited to short sequences (Table 1.1). The determination of additional number of large DNA and RNA 3D

[1] An analogous process to the condensation of the chromatin upon the interaction with histones.

structures would be of great interest. Furthermore, a precise knowledge of the 3D genome structure could be used to validate secondary structure models.

The study of the genome structure is not only theoretically relevant, but also medically relevant. Small ssDNA and ssRNA viruses include a number of important human, animal and crop pathogens; including picornaviruses (poliovirus, enteroviruses, foot-and-mouth disease virus), flaviviruses (hepatitis C virus, dengue and yellow fever virus), togaviruses (rubella virus), caliciviruses (winter vomiting disease), bromoviruses (alfalfa mosaic virus), comoviruses (tobacco ringspot virus), and so forth. A more precise knowledge about how the structure of the genome affects critical viral life cycle events, such as cell entry, replication or assembly, could suggest therapies for the treatment of these infections.

1.3 Introduction to cryo-electron microscopy

Cryo-electron microscopy is able to resolve the structure of macromolecular complexes in an unstained, natural, fully-hydrated conformation. CryoEM consists of embedding biological materials in a vitreous layer of ice and imaging under a low electron dose in a transmission electron microscope (Fig. 1.1).

Two major obstacles of cryoEM are preventing (a) crystalline ice formation during sample preparation and (b) radiation damage under the electron beam. In the early 1980s, Jacques Dubochet and coworkers realized that cooling rates were too slow to allow vitrification of the water [93–95]. They developed a procedure in which a thin layer of water was flash-frozen in cryogenic coolant. This, prevented the water from crystallizing, keeping the solution in a vitreous state and preserving embedded particles in a hydrated-like state. The amorphous water is a meta-stable state as long as the temperature is kept below -137° C (under typical EM vacuums, 10^{-7}–10^{-9} Pa),

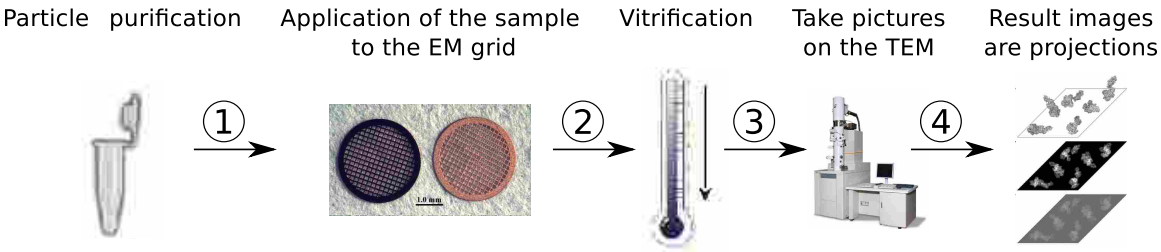


Figure 1.1 Schematic representation of a normal cryoEM experiment. (1) The purified virus is applied to an electron microscope grid. (2) The sample is blotted to make a thin specimen and flash-frozen. (3) Viruses are trapped in their natural conformation in an amorphous layer of vitrified water. (4) Micrographs are taken at low electron doses to avoid radiation damage. The resulting images represent projections of the structures at random views.

at which the water devitrifies irreversibly [96].

Organic specimens are damaged by the electron beam in the microscope [97]. In order to minimize radiation damage a low electron dose (on the order of 6-10 electrons per $Å^2$) is routinely used to image biological samples. Because of the low number of electrons the resulting images are too noisy to be interpreted directly. Hundreds or thousands of images are averaged to obtain interpretable results.

The capacity to produce three-dimensional (3D) reconstructions from EM images. is of paramount importance. This principle was demonstrated even before frozen-hydrated specimens were available. In 1968, De Rosier and Klug [98] produced a reconstruction of tobacco mosaic virus from pictures of negatively stained particles. Electron micrographs are two-dimensional (2D) projections in the view direction of the electron beam.[2] The central-slice, or projection theorem states that the Fourier transform of a 2D projection is a central section of the 3D transform of the object (Fig. 1.2). This mathematical relationship is the basis for 3D reconstruction tech-

[2]A 2D projection is the summation of the densities of a 3D object along certain view direction. Medical X-rays are good examples of 2D projections.

niques.[3] By applying an inverse Fourier transform, the 3D object, in real space, can be synthesized from the 3D transform. Therefore if many views from different angles are combined, the 3D object can be reconstructed.

To accomplish a 3D reconstruction the orientation (view angle) of the 2D projection, as well as the origin (defined center), has to be carefully determined (Fig. 1.3). As previously mentioned, because of the noisy nature of cryoEM images this determination is challenging. A number of methods have been devised to accurately find the origin and orientation of each image particle image (for a detailed discussion see Section 2.1).

1.4 Technical challenges for the structural determination of viral genomes

The determination of 3D structure is *per se* a challenging undertaking even more so, if the subject of the study has been historically elusive [79]. The tertiary structure determination of viral genomes is a tractable problem only if the structure is ordered or assumes a small number of conformations.

Methodological limitations are likely the major reason for the limited knowledge that we have about the structure of viral genomes. Several techniques have been used for the 3D structural characterization of viral genomes. Small-angle X-ray, neutron scattering, AFM and Raman spectroscopy have provided general information about the shape and size of the genomes, however, they lack the necessary resolving power.

In contrast to the aforementioned techniques, X-ray crystallography and cryoEM, have the potential to uncover the 3D structure of entire genomes at high or interme-

[3]Reconstruction algorithm used in medical imaging, such as, X-ray computed tomography or magnetic resonance imaging (MRI), are also based in this principle.

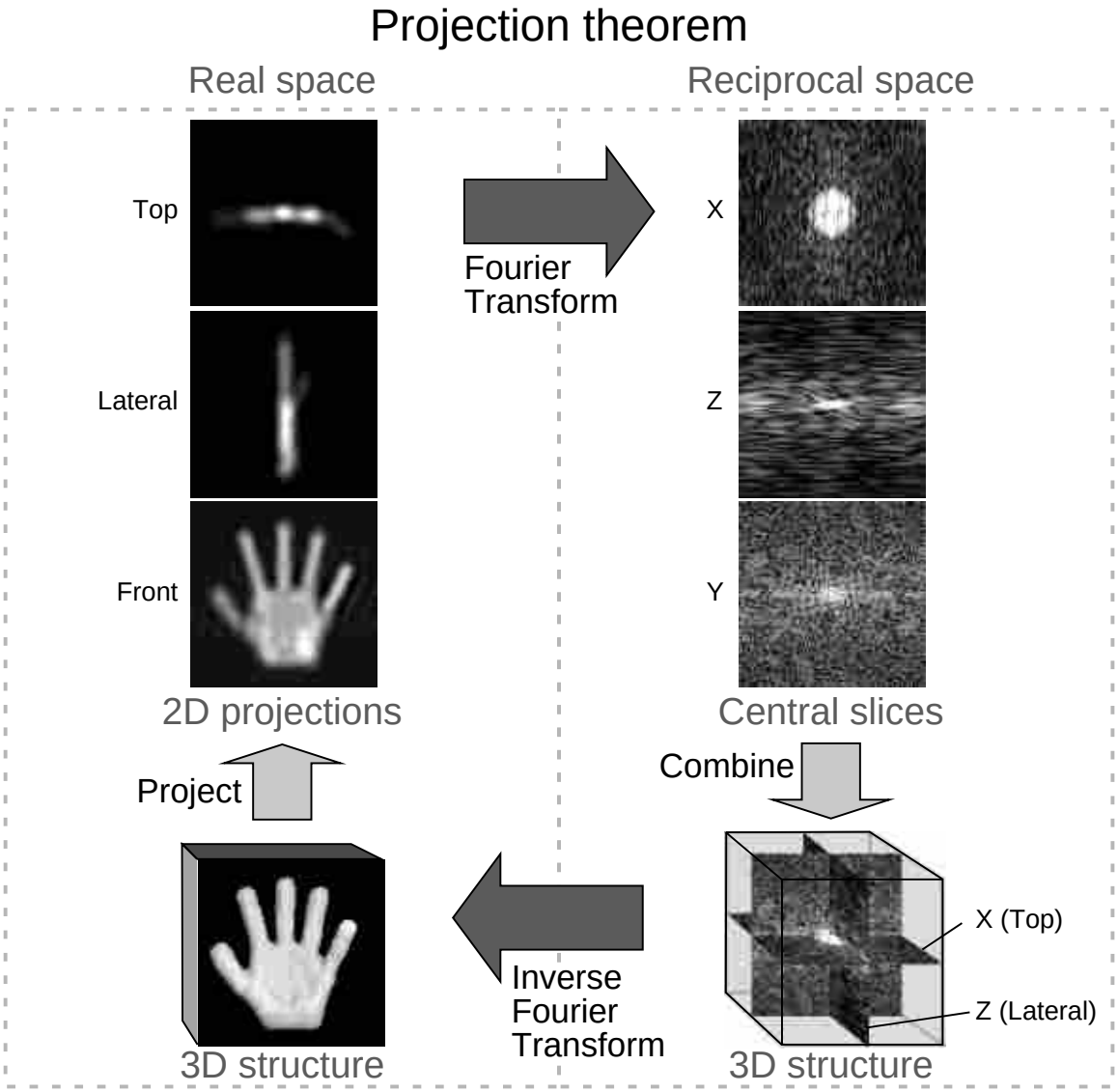


Figure 1.2 The central-slice theorem states that there is a reciprocal relationship between 2D projections in real space and central slices of a 3D structure in Fourier space. This relationship allows the 3D reconstructions from 2D projections [adapted from 99].

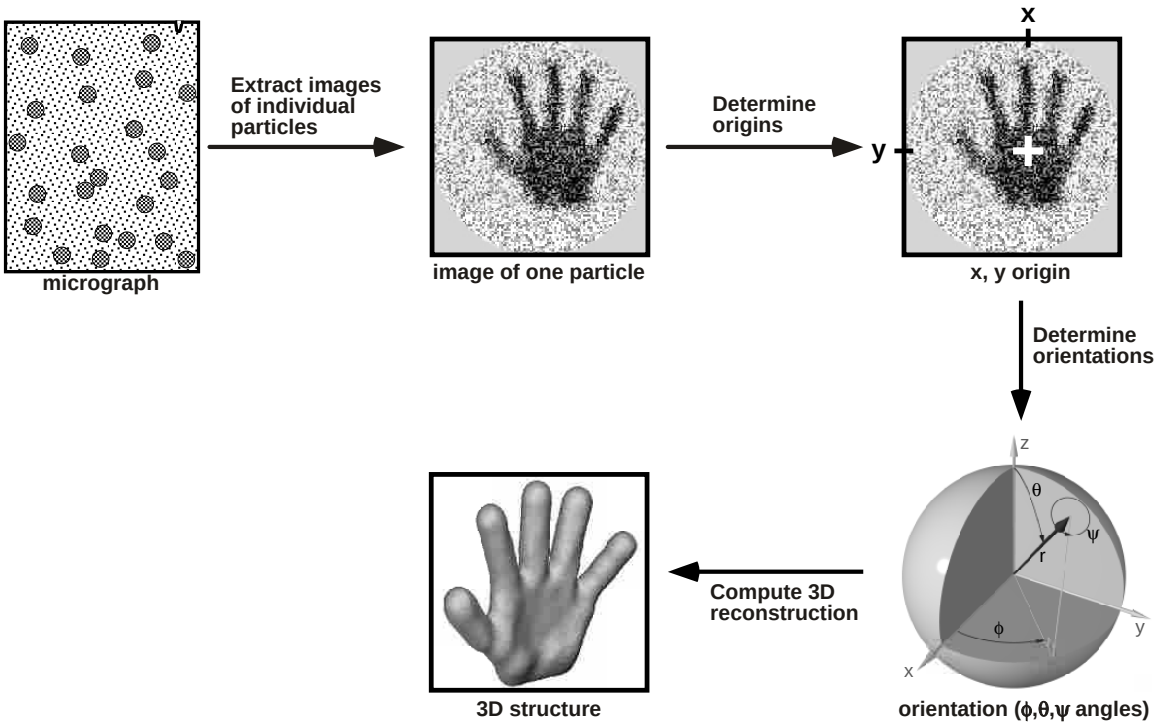


Figure 1.3 To compute a 3D reconstruction the origin and orientation of each particle image has to be carefully determined. Origins correspond to a defined centers (*e.g.* center of mass), and orientations represent the view angles of the projections. Figure adapted from David Belnap.

diate resolution, respectively.

Virus crystallography has been limited to the structural determination of viral capsids. Virus crystal growth is directed by interactions of the outer elements of the viral capsid. Therefore, the crystal has no chance to orient with respect to the genome (compare Fig. 1.4D,E). The lack of synchronization between the symmetric capsid and the internal genome results in a smearing of the structural information of the genome (Fig. 1.4C,D). Attempts to direct crystal growth using gravitational or electrostatic forces have failed [100]. In several cases, small portions of the viral genome that become associated with the internal surface of the capsid, have been resolved (Table 1.1).

CryoEM can allow the structural determination of ordered viral genomes. Unlike crystallography, cryoEM doesn't require the challenging process of growing crystals. Nevertheless, the resolution power of cryoEM is typically limited to one or two nanometers due to poor alignment precision of noisy images taken at low electron doses. Because every image has to be analyzed individually, specific components within each individual image can be extracted or aligned [79]. This methodological flexibility has allowed the resolution of structures containing symmetry mismatches [41, 101–103] and can be advantageously used to orient the genomic component independently of the capsid (Fig. 1.4C-D).

Nevertheless, model-based cryoEM reconstructions algorithms depend critically on a starting model. This model, usually a low- or moderate-resolution structure, is typically derived from a pre-existing related structure. When no prior structural information is available, the starting model must be derived directly (*ab initio*) from the imaged particles. This represents a common bottleneck for model-based cryoEM reconstructions. In this study, we present new strategies to address *de novo* determination of starting models and the structural determination of genomes inside

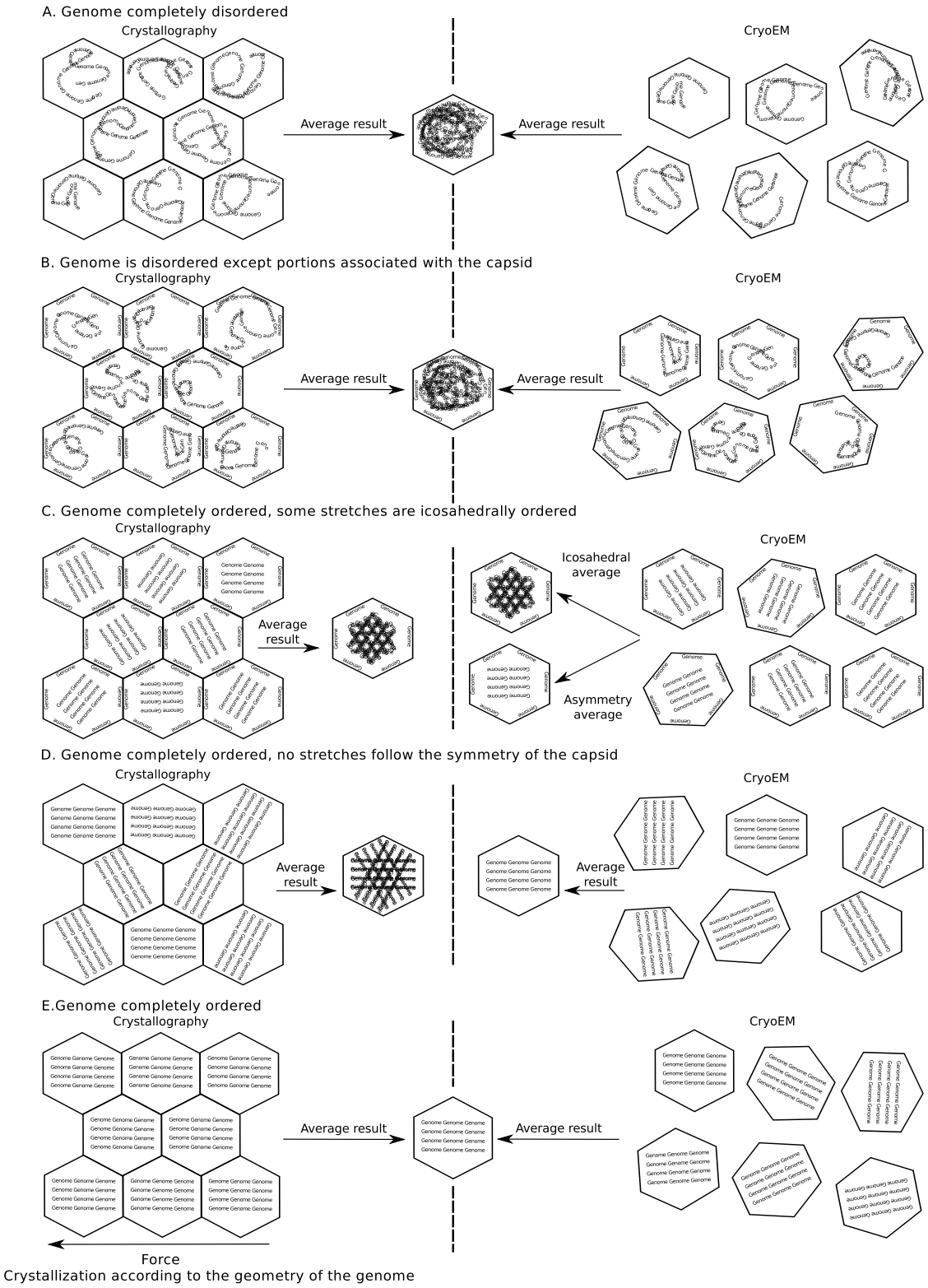


Figure 1.4 Possible arrangements of the viral genome and structural determination results, using X-ray crystallography and cryoEM. A-B, portions of the genome assume total or partial disorder, respectively. B-C, some sequences interact with the protein shell acquiring the same symmetry as the viral capsid. C-E, the viral genome is completely ordered.

icosahedral capsids (Chapters 2 and 4).

1.5 Subjects of the study

1.5.1 Poliovirus

Poliovirus is the causative agent of poliomyelitis, and the archetype of the family *Picornaviridae* [104]. Poliovirus is regarded as a useful model system because of its simple composition and well-characterized biology. Poliovirus is a 33-nm, non-enveloped, icosahedral virus with a (+) ssRNA genome of 7,500 nucleotides [105]. The 5' end is linked to a small protein (VPg) and the 3' end is polyadenylated. The RNA core of poliovirus is densely packed [92] with portions of known secondary structures [4, 5]. Although recent bioinformatic studies have suggested that the RNA is not ordered [72, 73], AFM studies suggest otherwise [76].

The virion capsid is composed of 60 copies of the VP1, VP2, VP3 and VP4 coat proteins arranged in a pseudo T=3 lattice [105]. VP4, as opposed to the other capsid proteins, is not exposed to the outer surface in the native conformation of poliovirus (160S). In addition to the native state, two other conformations (135S and 80S) have been resolved [106]. These conformations likely represent cell-entry transition states (see more detail in Chapter 5).

1.5.2 Satellite tobacco mosaic virus

Satellite tobacco mosaic virus (STMV) is one of the smallest icosahedral viruses with a diameter of 17 nm [57]. Its T=1 capsid is formed by 60 copies of one protein that encloses a (+) ssRNA genome of 1,059 nucleotides. The icosahedrally symmetric crystal structure revealed that at every dyad axis the RNA is arranged as a

9 base double-helical segment. These 30 helices account for about 50% of the whole genome. The viral genome likely plays an important role in capsid assembly [67] and capsid stability [65]. Extracted RNA cores seem to adopt a globular shape with a diameter corresponding to the interior of the capsid and are able to unfold and fold reversibly [76] (see more detail in Chapter 6).

Chapter 2

The asymmetric random-model method

2.1 Background

In reconstructing 3D structure from 2D electron images of freestanding particles, model-based or projection-matching algorithms can effectively determine image origins and orientations [107–113]. In these techniques, a 3D structure ("model") is projected into 2D images (projections) at orientations throughout the asymmetric unit.[1] Each model projection is compared to each image in the experimental data set. The orientation assigned to the most similar projection becomes the particle orientation, and translational alignment of this projection gives the particle origin. A 3D reconstruction is then computed from the experimental images. This new structure is used as the model for another round of origin and orientation refinement. The cycle is repeated until no further improvement in resolution of the 3D reconstruction

[1] The smallest part of the structure from which the complete structure can be built by applying the corresponding symmetry operators.

is observed.

Model-based reconstruction techniques rely on a suitable starting model to initiate convergence to the correct 3D structure. Commonly, the model is a pre-existing, related 3D structure, resolved by cryoEM or X-ray crystallography [*e.g.* 114, 115]. The absence of a related structure is a serious obstacle and usually requires that the initial model be computed *ab initio*—directly from the experimental images.

Several *ab initio* methods exist. Random conical tilt [116], orthogonal tilt [117], and tomography [118] all require two or more views of the specimen taken at different tilt angles. Common lines [119, 120], angular reconstitution [121], and simulated annealing [122–124] do not require tilting of the specimen. Computer-generated shapes [111, 125], or reconstructions from one image of a particle [120, 126] have been used as starting models. Despite these advances, in 2008, Taylor and Glaeser [97] suggested that new tools are still needed to obtain starting models for unknown particles.

In the mid-1980s, van Heel [127] and Harauz and van Heel [128] proposed a "random model" strategy for *ab initio* 3D reconstruction. They suggested that an initial "random model" be created by first assigning random orientations to aligned, averaged 2D images (class averages) and then computing a 3D reconstruction. This model would be used to start iterative model-based, projection-matching searches for orientation and origin. After several refinement cycles, the correct structure could eventually emerge. The random method was proposed to work for non-averaged, experimental projection-images only if (1) the random orientation assignments were close to the correct values or (2) the structure exhibited high symmetry [113]. Recently, Yan *et al.* [129] and Liu *et al.* [122] confirmed the prediction that the method would work for high symmetry objects by generating icosahedral starting models. Both groups suggested that the so-called "random-model method" (RMM) might work for particles with lower symmetries.

2.2 Material and methods

2.2.1 Random-model algorithm

We implemented the RMM as a six-step process (Fig. 2.1):

1. A random orientation was assigned to each imaged particle. The center of each image (*i.e.* the center of the image "box") was used as the initial origin [122, 129]. A 3D "random model" was computed from the images.

2. The random model was projected at regular angular intervals throughout the asymmetric unit. The number and interval of the projections was determined by the angular step-size. Standard Euler angles (Φ, Θ, Ψ) [130] relate with the angular step-size by

$$\Delta\Theta = \Delta\Psi = \text{angular step-size} \tag{2.1}$$

   and

$$\Delta\Phi = \text{angular step-size}/\sin(\Theta) \tag{2.2}$$

3. Each particle image was compared to the projections of the model for orientation and origin determination. We used a version of the model-based *PFT* algorithm [111] modified to (1) use both phase and amplitude information in selecting the best match of projections to each image, (2) add a Cartesian real-space correlation coefficient (CC) and (3) analyze particles that were asymmetric or had non-icosahedral point-group symmetry [131] (D. Belnap, B. Heymann, J. Conway, and T. Baker, unpublished).

4. Particle images with the highest CCs were selected until 80% of the particles were included. These images were used to compute the new reconstruction. A

low-pass Fourier filter of 1/20 $Å^{-1}$ was applied when computing the reconstruction.

5. The new 3D model was projected to begin a new cycle of origin and orientation determination.

6. We repeated steps 1-5 ten or more times. In each independent test, the same particle images were analyzed, but different random orientations were assigned and a new random model was computed for each test. The resulting starting models were compared visually and quantitatively.

2.2.2 Software and computer usage

In the *PFT2* algorithm [111], model projections (each with corresponding Φ, Θ values) were first ranked by similarity to the particle image based on correlation of polar Fourier transforms (amplitude-only). Because the projection set only contained one-half of the asymmetric unit (the other half is related by mirror symmetry), Φ and Θ values have two possible solutions. Next, the highest correlating 30 projections were tested with phase and amplitude information, via correlation in real space of polar images, to determine the correct Φ, Θ and Ψ values for each projection. The CC is defined in the following way:

$$\mathrm{CC} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i y_i}{\sqrt{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}\sqrt{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2}} \tag{2.3}$$

where x and y are images of a model projection and a particle image and $i = 1, 2, \ldots, n$ are pixels. The origin for each projection was determined via cross-correlation. Finally, the set of 30 projections was compared in Cartesian real-space coordinates to the particle image by a CC. The projection-image match with the highest CC was taken to be the new orientation and origin. Comparisons that used both phase and

amplitude information were limited by high- and low-pass Fourier filtering, 1/90 $Å^{-1}$ and 1/25 $Å^{-1}$, respectively. Only low-pass filtering was applied for the amplitude-only comparison and for origin determination. We compared data within a radius that encompassed the whole particle (same radius for all orientations).

Tests ran on a Linux cluster composed of 64-bit processors at 2.6 or 2.8 GHz. Typically 16 processors were used per job, requiring 15-59 hours to complete 40-100 iterations.

2.2.3 Synthetic particle images

Synthetic image particles were created by projecting an electron density model of an *E. coli* 70S ribosome (EMDB 1006 [132]) using random orientations and origins. The program *emgrand* (Bsoft [131]) was used to compute random orientation and origins. The actual projections were performed with *bint* (Bsoft). Gaussian noise was added (*brandom*, Bsoft) to simulate conventional cryoEM images. Different signal-to-noise ratios (SNR) were tested ranging from "1/0" (absence of noise) to 1/100. We introduced four different levels of randomness on the origins, such as the standard deviation of the particle origins were: 0.0 (particles centered on the image boundaries), 2.5, 5.0, 10.0 pixels. The standard deviation of the origin is defined in the following manner:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=0}^{n} \Delta x^2 + \Delta y^2} \tag{2.4}$$

where x and y are the origin coordinates and $i = 1, 2, \ldots, n$ are the images.

2.3 Results and discussion

As a preliminary step, before applying the asymmetric RMM to experimental data (see Chapter 3), we decided to test the asymmetric RMM on synthetic images. We

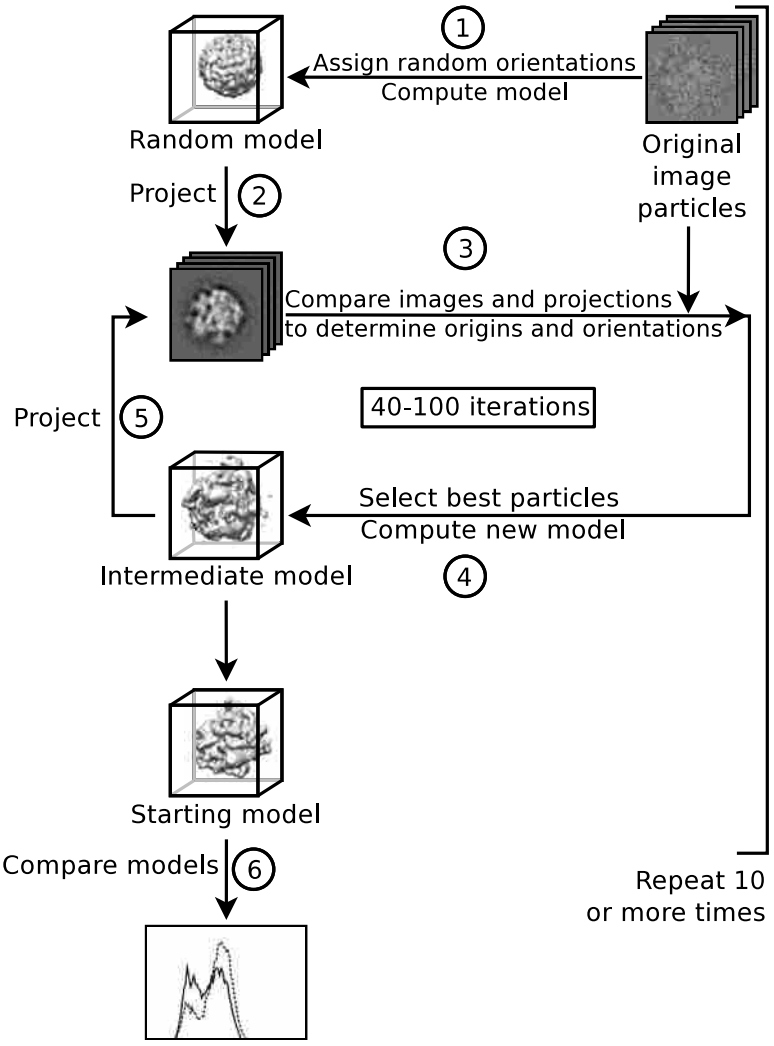


Figure 2.1 Random-model method. (1) A random orientation is assigned to every individual image. The center point of the image "box" is used as the initial origin for each image. A 3D reconstruction is then computed. (2) The 3D "random model" is projected at even angular steps throughout the asymmetric unit. (3) These projections are used to determine orientations and origins. (4) A figure-of-merit (*e.g.* correlation coefficient) measures the similarity of each matched projection and image. The value is used to determine the "best" rotational (orientation) and translational (origin) alignments in the set. Aligned particles are used to compute a new (intermediate) model. (5) The intermediate model serves as the input model for refinement of orientations and origins (step 3). Typically, steps 3-5 are repeated 40-100 times. The whole scheme (steps 1-5) is repeated independently 10 or more times. In each repetition, the same images are used, but a different random model is generated. (6) Resulting (starting) models are analyzed and compared both visually and quantitatively to determine suitability for further use.

successfully reconstructed the *E. coli* 70S ribosome using synthetic particle images.

2.3.1 Effect of image noise on the success rate of the asymmetric RMM

We studied the effect of noise on the success rate of the asymmetric RMM. We introduced different levels of noise on the data sets. As a way to compensate for increased amounts of noise, we employed proportionally higher numbers of particle images to compute the reconstruction. We found that the asymmetric RMM was relatively insensitive to typical cryoEM SNRs (Table 2.1). We were unable to obtain suitable starting models when the SNR was 1/100, even though a very large number of particles were used (Table 2.1). In one case, using 6,000 particle images, the absence of noise reduced the success rate. In this instance, the lack of image noise caused the random models to be very smooth and spherical. This lack of structural features in the random models appears to disable successful orientation refinements (see also Section 3.3.4).

2.3.2 Effect of the randomness of particle origins on the success rate of the asymmetric RMM

We studied the effect of the randomness of the particle origins on the asymmetric RMM by creating four different particle data sets. On each data set, particle images were translated randomly from the center of the image "box". Standard deviation of the data sets ranged from 0 to 10 pixels (Table 2.2). We detected a negative effect on the success rate with increased deviation from the center of the image "box" (Table 2.2).

Table 2.1 Effect of the image noise on the success of the asymmetry RMM using synthetic ribosome particle images.[a]

Number of particles	SNR	Success rate (%)[b]
6,000	1/0	20
6,000	1/10	100
24,000	1/20	100
24,000	1/40	100
37,520	1/50	100
66,680	1/66	80
6,000	1/100	0
12,000	1/100	0
24,000	1/100	0
50,000	1/100	0
100,000	1/100	0
150,000	1/100	0
300,000	1/100	0

[a] Standard deviation of the particle origins was zero.
[b] Ten independent runs

Table 2.2 Effect of the particle origins on the asymmetry RMM.[a]

Number of particles	SD of particle origins[b]	Success rate (%)[c]
500	0	70
500	2.5	50
500	5.0	30
500	10.0	30

[a] SNR of particle images was "1/0".
[b] SD refers to standard deviation.
[c] Ten independent runs

2.3.3 Combined effect of noise and randomness of the particle origins on the success rate of the asymmetric RMM

We studied the combined effect of image noise and randomness of particle origins on the asymmetric RMM. We chose a standard deviation of the particle origins and ranges of SNRs that matched typical experimental cryoEM data. We observed an additive negative effect of both factors on the success rate of the asymmetric RMM. (Table 2.3).

Table 2.3 Cumulative effect of image noise and randomness of the particle origins on the asymmetric RMM.[a]

Number of particles	SNR	Success rate (%)[b]
24,000	1/40	50
37,520	1/50	20
66,680	1/66	10

[a] For a meaningful comparison with previous results, we used the same number of particles as was used in the corresponding test reported in Table 2.1. Standard deviation of the particle origins was 10.

[b] Ten independent runs

Chapter 3

Applications of the asymmetric random-model method to experimental data

3.1 Background

Extending 3D-reconstruction algorithms from high symmetry to no symmetry usually only requires redefinition of the asymmetric unit and approximately n-fold or more particle images. However, adapting the icosahedral RMM [122, 129] to work with asymmetric particles proved to be non-trivial. Our attempts to simply use asymmetry, 60-fold more particle images, and methodology given by Yan *et al.* [129] failed to compute suitable starting models with experimental cryoEM images, though tests with model images (with and without noise) were successful (see Chapter 2). One of the most important factors for the success of the asymmetric RMM was dynamically varying the angular step size.

Using cryoEM images of ribosomes, we confirmed that the RMM can produce

suitable starting models of experimental asymmetric particles. We also produced suitable, symmetric starting models for dihedral and icosahedral data sets without imposing symmetry. Therefore, the RMM is also a tool for symmetry determination. Because the symmetry of the particles can, in some instance, be challenging to determine, the RMM is a valuable tool to overcome this obstacle.

3.2 Material and methods

3.2.1 Particle images

We analyzed images of ribosome, GroEL, and other particles that were prepared via standard cryoEM procedures for other studies (Table 3.1). We applied a radial mask to particle images and set the background to the mean of intensities between the mask and the edge of the image (Bsoft [131]) or the mask and the edge of a user-defined square that was several pixels larger than the mask (*X3DPREPROCESS* [133]).

Table 3.1 Tests of the asymmetric random-model method: data sets, conditions, and results.

Specimen	Native symmetry[a]	Components/Notes	Number of particles[b]	CTF Correction[c]	Electron source	Accelerating Voltage (kV)	Nominal magnification	Recording medium[d]	Underfocus range (μm)	Data Reference	Success rate (%)[e]	Highest resolution map (Å)[f]	EMDB ID[g]
Ribosome	C1	*E. coli* 70S ribosome complex, 70S-fMet-tRNA^{fMet}-Phe-tRNA^{Phe}-EF-Tu-GDP-kirromycin	10,000	Phases only	Field Emission Gun (FEG)	200	50k	film	1.6-4.0	[134]	40	23	5141
GroEL	D7	Publicly available data set[h]	6,613	Phases only	FEG	120	100k	CCD camera	2.0-9.9	[135]	70	18	5143
Avian polyomavirus[j]	I	Triangulation number of 7*dextro*	2,099[i]	None	FEG	200	39k	film	0.7-4.9	j	55[k]	35	5142
Poliovirus-receptor complex	I	Soluble poliovirus receptor complexed with 160S poliovirus.	2,813[i]	Phases only	FEG	120	38k	film	0.6-2.6	[106]	90	30	5144
Poliovirus	I	Native (160S) conformation	5,016	Phases only	FEG	300	39k	film	2.6-5.7	l	0	—	—
Mouse polyomavirus	O	Major capsid VP1 assembled with octahedral symmetry	2,143[i]	Phases only	Tungsten filament	100	50k	film	0.8-2.1	m	0	—	—

[a] Point-group symmetry of native particle given in Schoenflies notation.
[b] For the ribosome data, 10,000 particle images were randomly selected from a set of 75,996 images to encompass the full range of defocus levels in the set. For the GroEL data, we simply used the first 6,613 images that we downloaded (from a total set of 55,251 particle images). From a larger set of 44,218 images, 5,016 of the most highly defocused poliovirus images were picked. Other data sets included all available images.
[c] "Phases only", correction for phase-flipping effects of the contrast transfer function (CTF) [131, 136].
[d] Film digitized on a Nikon Super Coolscan 9000 ED or Z/I Imaging SCAI scanner.
[e] for 10 repetitions, unless specified otherwise.
[f] Where FSC curve crosses 0.5.
[g] One representative structure was deposited in the Electron Microscopy Data Bank (EMDB, http://emdatabank.org).
[h] http://ami.scripps.edu/experiment
[i] Number of focal pairs.
[j] P. Shen, D. Enderlein, W. Carter, R. Swenson, R. Johne and D. Belnap, unpublished [see also 137, 138].
[k] Percentage for 9 independent repetitions.
[l] E. Sanz-García, B. Leishman, D. Bubeck, R. Noel, J. Hogle, and D. Belnap, unpublished; a second test was performed with 44,000 images (see Section 3.3.1 and Fig. 3.6).
[m] R. Swenson, G. Wang, P. Shen, M. Cahoon, T. Baker, D. Belnap, unpublished, see also[139].

3.2.2 Random model algorithm

We modified the previously described algorithm (see Section 2.2.1) to suite experimental particle images. For each successive iteration, we used an angular step-size different from the previous iteration (Table 3.3).

3.2.3 Image processing

We used the Bsoft [131] and PFT3DR [111, 131, 140] packages and *X3DPREPROCESS* [133] for image processing and reconstruction. Bsoft programs or *X3DPREPROCESS* were used for particle extraction, background subtraction, masking, and linear-gradient correction. Contrast transfer function (CTF)[1] parameters were determined via *bshow* and images were corrected via *bctf* (Bsoft). *PFT2* (PFT3DR) was used for orientation and origin determination. *breconstruct* (Bsoft) was used to compute 3D reconstructions. We developed a new Bsoft program, *emgrand*, to assign random orientations (Fig. 2.1, step 1) and two scripts to use *PFT2* (*pft2p*) and *breconstruct* (*breconstructp*) on clustered computers. (Scripts and source code are available at http://bsoft.ws and http://people.chem.byu.edu/belnap/pft3dr.)

3.2.4 Validation of the starting models

We validated our starting models via visual inspection; principal component analysis (PCA); and assessments of resolution, CCs, and map variance. Maps generated from independent runs were compared to each other and to reference structures. A *t*-test, with unequal variance assumed, was applied to resolution, CC, and map-

[1]The CTF is the reciprocal space counterpart of the point spread function, and describes the response of an imaging system to a point source. The CTF models the decay of the signal and inversion of the contrast relative to the spatial frequencies and focus levels. Usually, the astigmatism (anisotropic differences in focus levels) is detected and corrected simultaneously with the CTF.

variance assessments. Fourier shell correlation (FSC) [121, 141], differential phase-residual (DPR) [142], spectral signal-to-noise ratio (SSNR) [143] and reliability index $R_{(A+B)}$ [144], were calculated in the last iteration by splitting selected particles into two subsets from which two maps were computed. Curves were calculated via *bresolve* (Bsoft). Resolution was estimated to be where the curve crossed the 0.5, 45°, 1 and 0.5 values, for FSC, DPR, SSNR and $R_{(A+B)}$, respectively. CCs were obtained from the last *PFT2* orientation and origin search (Fig. 2.1, step 3). For selected particles, we analyzed the mean and histogram of the CCs. The standard deviation of the densities within each 3D map was calculated. We used a radius that encompassed the whole particle. For PCA [113], we first aligned starting models from independent tests to a common origin, orientation, and handedness. (PCA was very sensitive to these three alignments.) The model with the highest resolution was the template. Models were aligned via the Bsoft program *bfind* or the alignment utility in *UCSF Chimera* [145]. We computed dissimilarity matrices by comparing aligned maps to each other via *bmapdist* (Bsoft). For the similarity functions, we tested two real-space criteria, -log CC and R-factor (least-squares measure), and two reciprocal-space criteria, FSC and DPR (0.5 and 45°cut-offs respectively). We used classical (metric) multi-dimensional scaling to display the similarities graphically via the cmdscale function in the R package [146] (http://www.R-project.org).

3.3 Results and discussion

We found that the model-based, projection-matching RMM, previously demonstrated only for asymmetric model data and for icosahedral symmetric particles [122, 129], can produce suitable starting models from images of experimental asymmetric particles (Fig. 3.1, Table 3.2). The RMM is another tool in the *ab initio* starting-model

repertoire and is relatively easy to implement and use. The technique also can be used to determine symmetry.

3.3.1 Ribosome data set

Changes in angular step-size

After many tries using constant angular step-sizes—including adjustments to resolution range, radial range, and number of particle images we discovered that changes in angular step-size allowed the asymmetric RMM to work for ribosome cryoEM images (Table 3.3; Figs. 3.1-3.2). In only one case, discovered later, did a constant step-size result in suitable models (Table 3.3, row 10; Fig. 3.2C). Because the angular step-size was critical for success with the asymmetric RMM, we tested several different changes in angular step-size among successive iterations (Table 3.3, Fig. 3.2). Highest success rates occurred with non-uniform changes in step-size (Table 3.3, rows 13-14; Fig. 3.2D-F). Moderate success rates were observed for slow-uniform decreases and a constant step-size of 10° (Table 3.3, rows 1-4, 10; Fig. 3.2A,C). Low rates were observed for fast-uniform decreases (Table 3.3, rows 5-8; Fig. 3.2B). Use of a constant step-size of 1°and 20° gave no starting models matching the reference structure (Table 3.3, rows 9, 11; Fig. 3.2C). Forty to 100 iterations were usually necessary to get suitable starting models. In general, more rapid declines in angular step-size gave lower success rates but did give suitable maps after fewer iterations.

Variations in angular step-size may allow particles to more easily escape local minima (wrong orientations or origins), especially if consecutive iteration cycles do not contain a subset of the same projections. For example, if an iteration with a 2° step-size is followed by an iteration at 1° steps, all projections used in the 2° set are also present in the 1° set. Few projections in the subsequent set will be

identical if the second iteration has a 1.7° step-size. During RMM tests, the 3D structure evolves from a spherical-like model to correct or incorrect structures through a series of intermediate maps (*e.g.* Fig. 3.1A,B) that induce adjustment of origins and orientations to correct or incorrect values. Although random models look very similar (Fig. 3.1A,B), the stochastic orientation assignments (Fig. 2.1, step 1) produce small density differences that determine the subsequent evolution of the model. Despite repeated refinement cycles, unsuitable maps changed very little and showed at most only partial similarity to the correct structure. In our experience, these "local-minima structures" often corresponded to rotationally symmetric structures (*e.g.* Fig. 3.6C1). These models generated nearly identical projections over the asymmetric unit, making escape from local minima difficult.

Image defocus

Image defocus affects success rates of the asymmetric RMM. We grouped ribosome images by average defocus. Three sets of 8,000 ribosome images had average underfocus values of 1.7, 2.5, and 3.8 μm. The highest defocus set gave the highest rate of suitable starting models (70%). The two lower-defocus sets had success rates of 40%. A similar trend was observed previously [129].

Number of particles

For the ribosome data set and the asymmetric RMM, we obtained suitable starting models with as few as 500 particle images and as many as 76,000 (Table 3.4). To avoid bias because of defocus level, the average defocus of each set of images was the same. Each test used the same pattern of changes in angular step-size. We observed that the success rate increased as the number of images increased, from 20% with 500 images to 90% with 76,000 images.

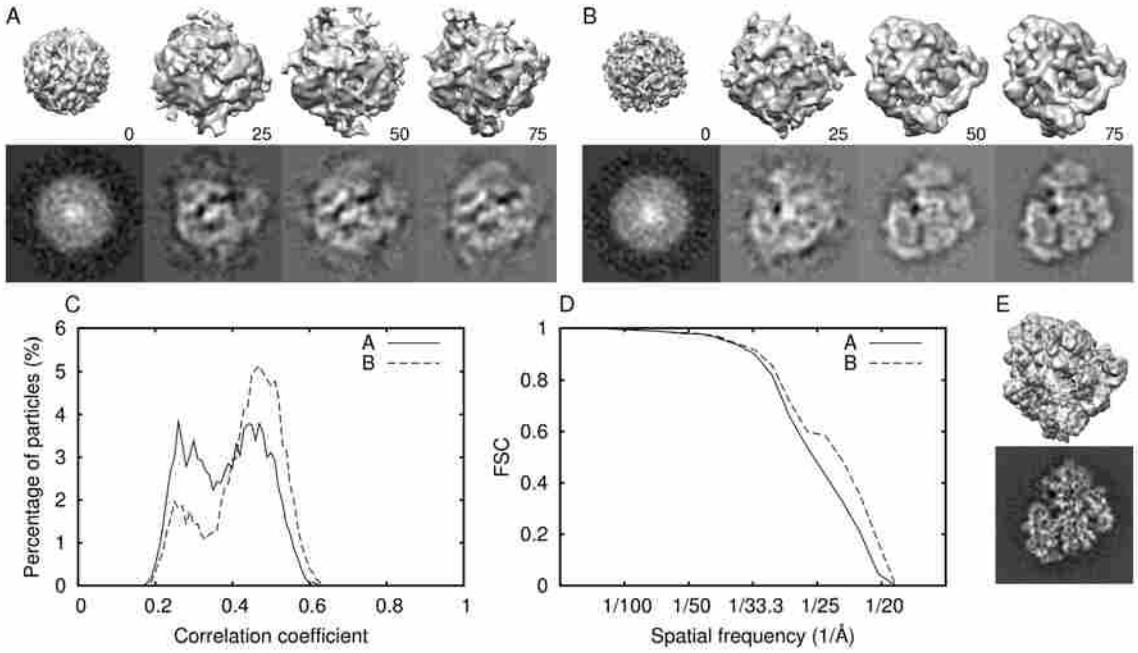


Figure 3.1 Ribosome reconstruction by the asymmetric RMM. Panels A and B show, respectively, unsuitable and suitable starting models: top row, surface representation; bottom row, a central section. Numbers indicate the iteration number. Panel E shows the original published map [134, 147] (EM Data Bank code 1055) at sub-nanometer resolution using the same data set: surface rendering, top; central section, bottom. Panels C and D display the CC histograms and the Fourier shell correlation plots (FSC), respectively, for the unsuitable (A) and the suitable maps (B).

Table 3.2 Validation of the starting models obtained from the ribosome data set.[a]

Model ID[b]	Qualitative classification by visual inspection[c]	Resolution estimation against standard (Å)[d]	Resolution (Å)[e,f]	Mean CC[e,g]	SD densities ($\times 10^{-3}$)[h]
A[i]	Incorrect	115.0	25.4	0.419	1.14
B[j]	**Correct**	**21.9**	**23.0**	**0.471**	**1.27**
C	Incorrect	107.0	24.8	0.422	1.13
D	Incorrect	117.0	24.2	0.407	1.12
E	Incorrect	103.0	25.4	0.419	1.12
F	**Correct**	**22.3**	**23.3**	**0.467**	**1.26**
G	**Correct**	**22.3**	**23.2**	**0.472**	**1.27**
H	Incorrect	132.0	24.3	0.421	1.15
I	Incorrect	117.0	25.0	0.418	1.14
J	**Correct**	**22.2**	**23.2**	**0.468**	**1.26**

[a] Columns in grey represent comparison of the starting models against the reference structure. In bold, values obtained by suitable starting models. Angular step-size: first iteration 20° step-size decreased by 0.19° each successive iteration.
[b] Each random model (Fig. 2.1, step 1) is derived from an independent asymmetric RMM run.
[c] Visual inspection comparing to reference structure [134, 147]
[d] Resolution at which the FSC curve intersects 0.5. The random model is compared by against the published structure [134, 147]
[e] Welch's t-test (Student's t-test but assumes unequal variance of the two samples) performed to assess the likelihood that unsuitable and suitable populations differ.
[f] Same as c, but here the particle images are split in two subsets and reconstructions computed. The two maps are compared by FSC. Resolution differences between suitable and unsuitable maps were highly significant (p=3.2×10^{-4}).
[g] Mean of the correlation coefficient of the particles in the last iteration.CC differences between suitable and unsuitable maps were highly significant (p=9.9×10^{-8}).
[h] Standard deviation of the density values. Map densities varied between -1 and 1. Deviation differences between suitable and unsuitable were highly significant (p=3.0×10^{-8}).
[i] Model corresponds with Fig. 3.1A.
[j] Model corresponds with Fig. 3.1B.

Table 3.3 Effect of angular step-size on the success rate of the asymmetric random-model method (with ribosome data set).[a]

Angular step-size (°)				
Initial[b]	Final[c]	Change[d]	Success rate (%)[e]	Convergence[f]
30	1	-0.3	40	—
20	1	-0.2	55	—
20	1	-0.19[g]	40[h]	58 (±15)
10	1	-0.1	30	30 (±9)
30	1	-0.6	10	—
20	1	-0.4	5	20 (±0)
10	1	-0.2	25	28 (±8)
20	< 1	$\times 2/3$	0[h]	—
20	20	0	0	—
10	10	0	40	43 (±18)
1	1	0	0	—
20,10[i]	1	-0.2	40	—
20,15[i]	1	-0.2	75	49 (±18)
20,18[i]	1	-0.2	70	—

[a] Classification of changes in angular step-size: slow-uniform decrease (rows 1-4), fast-uniform decrease (rows 5-8), constant (rows 9-11), and non-uniform (rows 12-14).
[b] Angular step-size used in the first iteration.
[c] Angular step-size lower limit.
[d] For each successive iteration.
[e] Percent success for 20 independent jobs, except as noted.
[f] Iteration at which suitable models appeared (average ± standard deviation), made by visual assessment of maps at iterations 10, 20, 30,. . . , 100.
[g] The reason for this odd step-size is that in our initial RMM tests with model particle images we originally tried the RMM on model data (Chapter 2) and found that 100 iterations were sufficient to see if a suitable starting model would appear. After we decided to try varying the angular step-size, for our first attempts we decided to start at 20° and end at 1° with 100 0.19° steps. This starting angle and step-size gave suitable starting models for the ribosome, GroEL, avian polyomavirus, and poliovirus-receptor data sets (Tables 3.5-3.7).
[h] Ten independent jobs
[i] The first number is the initial angular step-size for odd iterations; the second number is the initial angular step-size for even iterations.

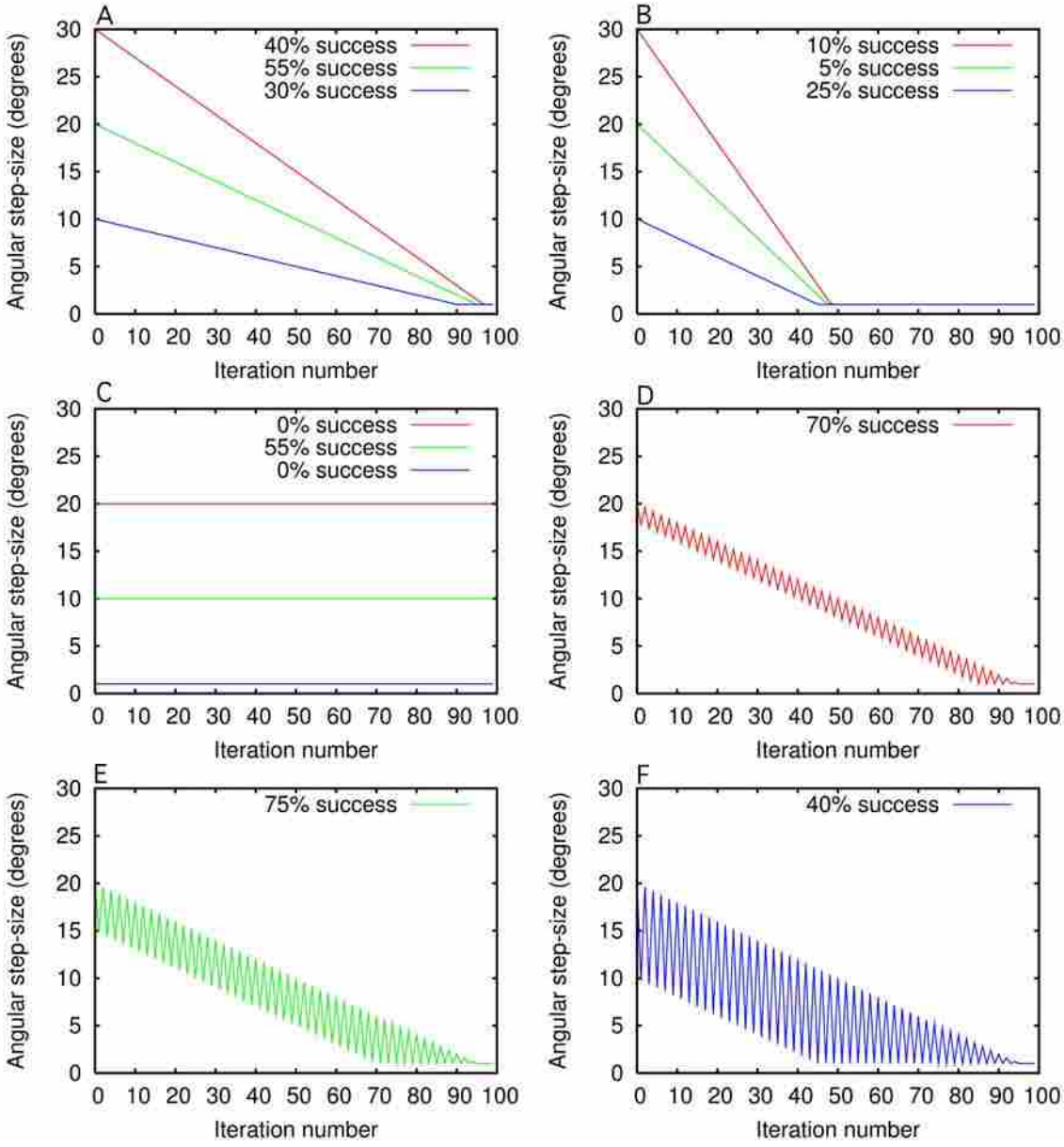


Figure 3.2 Effect of angular step-size on the success rate of the asymmetric RMM. Four basic changes in angular step-size among successive iterations were tested with the same 10,000 experimental ribosome particles (Table 3.3). The highest success rates were found when non-uniform changes in step-size were used (panels D and E).

Table 3.4 Relationship of success rate to the number of particle images.[a]

Number of particle images	Success rate (%)	Iteration at convergence[b]
500	20	—
1,000	20	—
3,000	40	80 (±23)
10,000	40	58 (±15)
76,000	90	57 (±23)

[a] Ribosome data set with 10 independent tests per entry. Angular step-size: initial 20˚ each successive iteration changed the step-size by -0.19˚. Average defocus, 2.9 μm underfocus.

[b] Iteration at which suitable models appeared (average ± standard deviation), made by visual assessment of maps at iterations 10, 20, 30,. . . , 100. The average number of iterations required for convergence to a suitable model declined substantially from 3,000 to 10,000 particle images, but declined insignificantly between 10,000 and 76,000 particles. Only a representative number of cases were assessed.

3.3.2 GroEL data set

Lower symmetry determination via asymmetric RMM

The asymmetric RMM can determine the symmetry of a particle. We obtained suitable starting models of known dihedral particles via the asymmetric RMM (Fig. 3.3; Table 3.5). GroEL particles have dihedral 7-fold symmetry [148]. The asymmetric RMM showed a high success rate (70%) of GroEL models with correct shape and symmetry, even though the GroEL particles showed preferential orientations in the ice, as top and side views [135].

3.3.3 Avian polyomavirus data set

Higher symmetry determination via asymmetric RMM

Icosahedral symmetry, with correct triangulation number and pentameric subunits (capsomeres), was observed in 55% of the asymmetric RMM tests with images of avian polyomavirus (Fig. 3.4; Table 3.6).

3.3.4 Poliovirus and poliovirus-receptor complex data sets

Prominent structural features

Prominent structural features facilitate production of suitable models via the asymmetric RMM. Poliovirus is a relatively smooth, icosahedral virus [106]. Our asymmetric RMM tests with 5,016 poliovirus images were unsuccessful (Fig. 3.6; Tables 3.1 and 3.8). A number of modifications were tried without success, including changes in high- and low-pass Fourier filters, inner and outer radial boundaries, and angular step-size. When a similar number of images of a poliovirus-receptor complex were used, the asymmetric RMM generated suitable starting models (Fig. 3.5; Tables 3.1 and 3.7). Yan *et al.* [129] also reported lower success rates of the icosahedral RMM if particles lacked prominent features. Random models are round and lack prominent features. Without constraints imposed by symmetry, random models appear unable to initiate successful orientation and origin searches if the particles are also round and relatively featureless, like some spherical viruses. The success rate of the poliovirus-receptor tests—compared to the poliovirus tests—indicates that addition of antibodies or ligands facilitates reconstruction of refractory "featureless" particles, as was shown previously [111]. Additional particles may also facilitate reconstruction of such particles. When we used 44,000 poliovirus images, two of our 10 tests produced maps that appear suitable (Fig. 3.6).

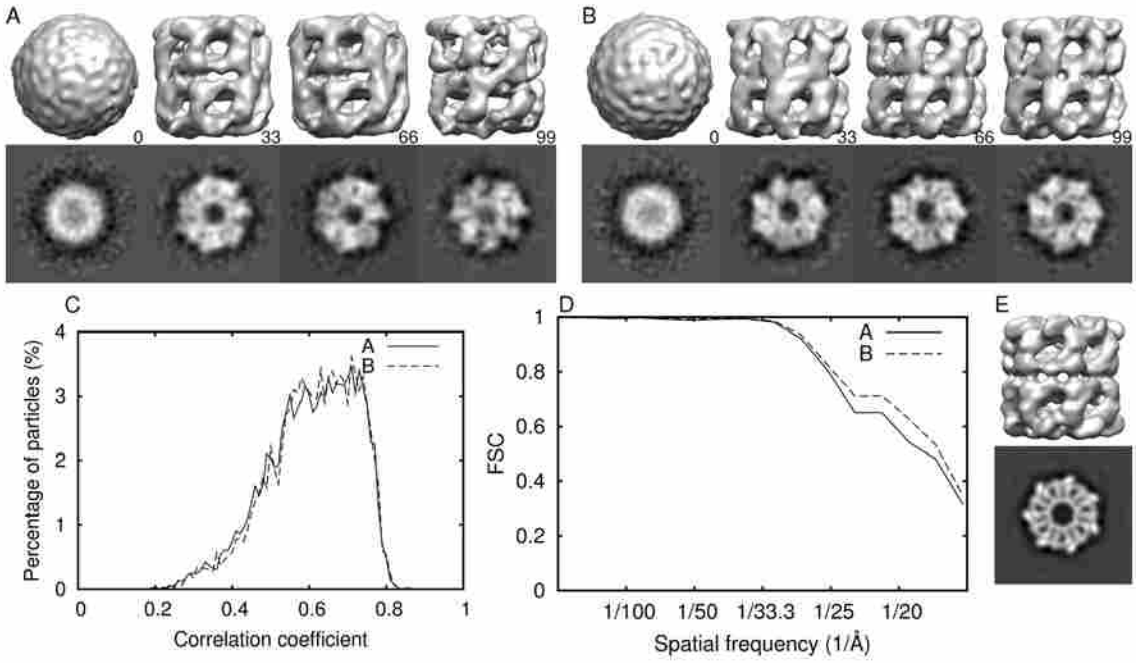


Figure 3.3 GroEL reconstruction by the asymmetric RMM. Panels A and B show, respectively, an unsuccessful and successful reconstruction: top row, surface representation; bottom row, a section through a 7-fold ring domain. Maps were computed from a set of 6,613 particle images [135]. Numbers in these panels indicate the iteration number. Panel E shows the originally published map [135] (EMDB 1458) filtered at 20 Å using the same data set: surface rendering, top; a section through a 7-fold ring domain, bottom. Panels C and D display the CC histograms and the Fourier shell correlation plots, respectively, for the unsuccessful (A) and the successful reconstruction (B). Map A resembles map B in the surface rendered view, but the map sections show more differences. Indeed, map A correlated significantly worse with the originally published map than other starting models (Table 3.5).

Table 3.5 Validation of the starting models obtained from the GroEL data set.[a]

Model ID[b]	Qualitative classification by visual inspection[c]	Resolution estimation against standard (Å)[d]	Resolution (Å)[e,f]	Mean CC[e,g]	SD densities ($\times 10^{-2}$)[h]
A[i]	Incorrect	45.1	18.5	0.650	2.29
B[j]	**Correct**	**20.9**	**17.8**	**0.653**	**2.07**
C	**Correct**	**25.7**	**18.4**	**0.652**	**2.10**
D	**Correct**	**17.8**	**17.5**	**0.659**	**2.07**
E	**Correct**	**20.9**	**18.1**	**0.652**	**2.07**
F	**Correct**	**25.3**	**18.0**	**0.651**	**2.10**
G	Incorrect	28.2	17.8	0.647	2.17
H	**Correct**	**25.3**	**18.3**	**0.652**	**2.07**
I	Incorrect	29.0	17.9	0.648	2.20
J	**Correct**	**24.7**	**18.0**	**0.653**	**2.09**

[a] Columns in grey represent comparison of the starting models against the reference structure. In bold, values obtained by suitable starting models. Angular step-size: first iteration 20° step-size decreased by 0.19° each successive iteration.

[b] Each random model (Fig. 2.1, step 1) is derived from an independent asymmetric RMM run.

[c] Visual inspection comparing to reference structure [135]

[d] Resolution at which the FSC curve intersects 0.5. The random model is compared by against the published structure [135]. Note that models "G" and "I" correlate only slightly worse than the "correct" models. Map "A" correlated notably worse, though it too resembled the correct models in surface renderings (Fig. 3.3).

[e] Welch's t-test (Student's t-test but assumes unequal variance of the two samples) performed to assess the likelihood that unsuitable and suitable populations differ.

[f] Same as c, but here the particle images are split in two subsets and reconstructions computed. The two maps are compared by FSC. Resolution differences between suitable and unsuitable maps were not significant (p=0.84).

[g] Mean of the correlation coefficient of the particles in the last iteration.CC differences between suitable and unsuitable maps were moderately significant (p=0.0094).

[h] Standard deviation of the density values between radii 0-175 pixel. Map densities varied between -1 and 1. Deviation differences between suitable and unsuitable were not significant (p=0.058).

[i] Model corresponds with Fig. 3.3A.

[j] Model corresponds with Fig. 3.3B.

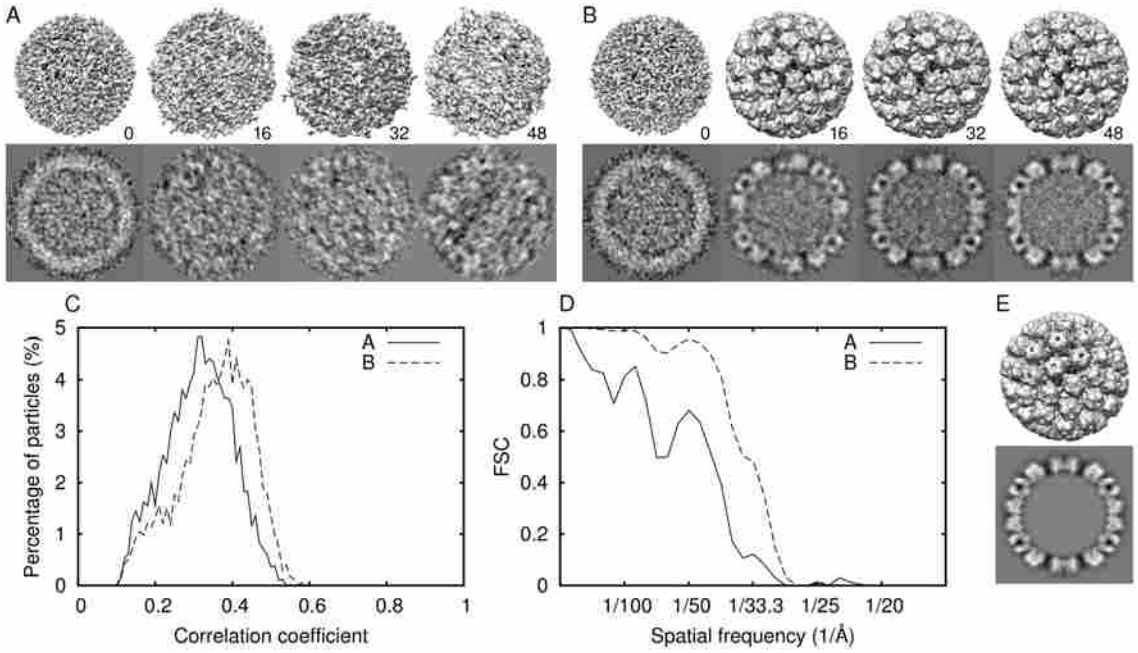


Figure 3.4 Avian polyomavirus reconstruction by the asymmetric RMM. Panels A and B show, respectively, an unsuccessful and successful reconstruction: top row, surface representation; bottom row, a central section. Maps were computed from a set of 2,099 focal pairs. Images were not CTF corrected. Numbers in these panels indicate the iteration number. Panel E shows the reference model computed at ∼12 Å using the same data set (surface rendering, top; central section, bottom) (P. Shen, D. Enderlein, W. Carter, R. Swenson, R. Johne and D. Belnap, unpublished). Panels C and D display the CC histograms and the Fourier shell correlation plots, respectively, for the unsuccessful (A) and the successful reconstruction (B).

Table 3.6 Validation of the starting models obtained from the avian polyomavirus data set.[a]

Model ID[b]	Qualitative classification by visual inspection[c]	Resolution estimation against standard (Å)[d]	Resolution (Å)[e,f]	Mean CC[e,g]	SD densities[h]
A[i]	Incorrect	131.0	42.3	0.351	0.790
B[j]	**Correct**	**38.2**	**34.9**	**0.397**	**0.721**
C	**Correct**	**35.6**	**34.9**	**0.398**	**0.720**
D	Incorrect	111.0	42.5	0.338	0.911
E	Incorrect	112.0	40.8	0.345	0.818
F	**Correct**	**35.7**	**35.2**	**0.398**	**0.720**
G	**Correct**	**37.9**	**35.2**	**0.396**	**0.717**
H	Incorrect	110.0	44.7	0.334	0.991
I	**Correct**	**35.5**	**34.8**	**0.397**	**0.718**

[a] Columns in grey represent comparison of the starting models against the reference structure. In bold, values obtained by suitable starting models. Angular step-size: first iteration 20˚ step-size decreased by 0.19˚ each successive iteration.
[b] Each random model (Fig. 2.1, step 1) is derived from an independent asymmetric RMM run.
[c] Visual inspection comparing to reference structure (P. Shen, D. Enderlein, W. Carter, R. Swenson, R. Johne and D. Belnap, unpublished).
[d] Resolution at which the FSC curve intersects 0.5. The random model is compared by against the published structure.
[e] Welch's t-test (Student's t-test but assumes unequal variance of the two samples) performed to assess the likelihood that unsuitable and suitable populations differ.
[f] Same as c, but here the particle images are split in two subsets and reconstructions computed. The two maps are compared by FSC.
[g] Mean of the correlation coefficient of the particles in the last iteration.CC differences between suitable and unsuitable maps were highly significant (p=6.3×10^{-4}).
[h] Standard deviation of the density values between radii 0-175 pixel. Map densities varied between -1 and 1. Deviation differences between suitable and unsuitable maps had low significance (p=0.041).
[i] Model corresponds with Fig. 3.4A.
[j] Model corresponds with Fig. 3.4B.

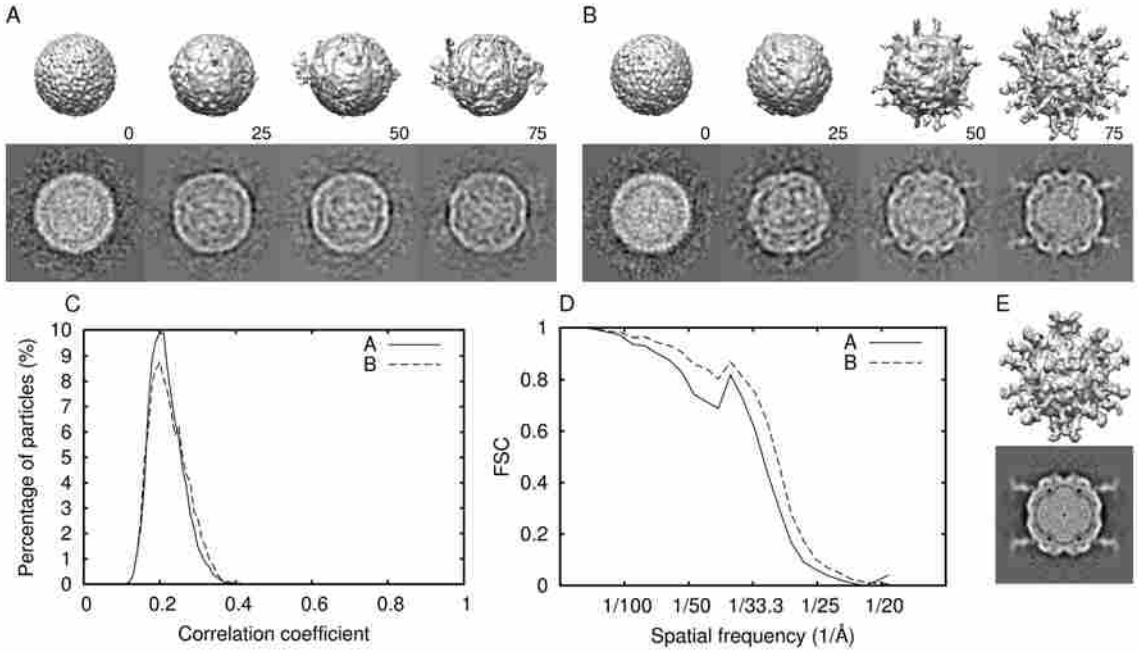


Figure 3.5 Poliovirus-receptor reconstruction by the asymmetric RMM. Panels A and B show, respectively, an unsuccessful and successful reconstruction: top row, surface representation; bottom row, a central section. Maps were computed from a set of 2,813 focal pairs [106]. Numbers in these panels indicate the iteration number. Panel E shows the published model at 21 Å resolution using the same data set: surface rendering, top; central section, bottom. Successful reconstruction (B) shows a level of detail comparable to the published structure. Panels C and D display the CC histograms and the Fourier shell correlation plots, respectively, for the unsuccessful (A) and the successful reconstruction (B).

Table 3.7 Validation of the starting models obtained from the poliovirus-receptor complex data set.[a]

Model ID[b]	Qualitative classification by visual inspection[c]	Resolution estimation against standard (Å)[d]	Resolution (Å)[e,f]	Mean CC[e,g]	SD densities[h]
A[i]	Incorrect	96.0	31.1	0.226	0.529
B[j]	**Correct**	**30.0**	**30.0**	**0.238**	**0.537**
C	**Correct**	**30.0**	**30.0**	**0.238**	**0.535**
D	**Correct**	**29.7**	**30.0**	**0.238**	**0.538**
E	**Correct**	**29.9**	**30.2**	**0.238**	**0.536**
F	**Correct**	**29.9**	**29.9**	**0.239**	**0.535**
G	**Correct**	**29.7**	**29.9**	**0.238**	**0.533**
H	**Correct**	**30.0**	**29.9**	**0.237**	**0.535**
I	**Correct**	**30.0**	**30.1**	**0.238**	**0.534**
J	**Correct**	**29.9**	**30.1**	**0.239**	**0.540**

[a] Columns in grey represent comparison of the starting models against the reference structure. In bold, values obtained by suitable starting models. Angular step-size: first iteration 20° step-size decreased by 0.19° each successive iteration.
[b] Each random model (Fig. 2.1, step 1) is derived from an independent asymmetric RMM run.
[c] Visual inspection comparing to reference structure [106].
[d] Resolution at which the FSC curve intersects 0.5. The random model is compared by against the published structure [106].
[e] Welch's *t*-test not done because unsuitable population was too small.
[f] Same as c, but here the particle images are split in two subsets and reconstructions computed. The two maps are compared by FSC.
[g] Mean of the correlation coefficient of the particles in the last iteration.
[h] Standard deviation of the density values between radii 0-140 pixel. Map densities varied between -1 and 1.
[i] Model corresponds with Fig. 3.5A.
[j] Model corresponds with Fig. 3.5B.

3.3.5 Octahedral mouse polyomavirus data set

Mixed handedness

The asymmetric RMM failed to produce suitable starting models with octahedral polyomavirus images because assigned orientations likely contained a mixture of handedness (Fig. 3.7). We flipped the handedness of our reference structure, aligned four adjacent pentameric subunits (capsomeres), and summed the two oppositely handed structures. The resulting map resembles models obtained via the asymmetric RMM. Therefore, if the RMM is unable to converge to a unique handedness, the resulting model will likely be unsuitable. (Similarly, failure of the RMM is more probable if particle conformations vary significantly.)

3.3.6 *Ab initio* reconstruction of refractory symmetric particles

Symmetric particles that were refractory to the asymmetric RMM were amenable to symmetric RMMs (Figs. 3.6 and 3.7). Using the same images and experimental parameters employed in the asymmetric RMM experiment, we computed maps via symmetric RMMs imposing symmetries both compatible and incompatible with the native symmetry. Suitable starting models were computed via some, but not all, symmetries compatible with the native particle symmetry. Native particle symmetry gave the highest success rates. As expected, incompatible symmetries produced unsuitable maps. The fact that symmetric RMMs worked for refractory icosahedral poliovirus and octahedral polyomavirus images indicates that symmetry constraints can overcome, at least some, deficiencies in the RMM.

Table 3.8 *Ab initio* reconstruction of refractory particles using the RMM.[a]

Specimen	Native point-group symmetry	Point-group symmetry and success rate (%)[b,c]												
		C1	C2	C3	C4	C5	C6	D2	D3	D4	D5	T	O	I
Octahedral polyomavirus	O	0	**0**	**0**	**0**	0	0	**0**	**0**	**60**	0	**60**	90	0
Poliovirus	I	0	**0**	**70**	0	**70**	0	**10**	**30**	0	**30**	**60**	0	80

[a] Angular step-size: first iteration 20° step-size decreased by 0.19° each successive iteration.
[b] Schoenflies notation.
[c] Percentage of success, out of 10 independent trials. Numbers in bold indicate percentage of suitable starting models found by use of the RMM with lower symmetries compatible with the particle's native symmetry.

3.3.7 Validation of starting models

Our experience is that any single validation criterion can be misleading, and the most reliable way to validate RMM results is to use multiple criteria. We validated starting models by comparing them visually and by FSC to reference structures. Our visual judgments were confirmed by FSC (Tables 3.2-3.7). Next, without reference maps, we used FSC, map variance, CC, and principal component analysis (PCA) to corroborate the previous validations.

Visualization is a reliable way to identify well-defined, consistent structures that may be suitable and smeared structures likely to be unsuitable. In general, if the maps represent the correct structure, particle features should have the highest definition and be more consistent (from map to map) than incorrect structures. Smeared features likely indicate unsuitable starting models. For example, in the octahedral polyomavirus tests (Fig. 3.7), we saw smeared or distorted capsomeres for all symmetries except dihedral 4-fold, tetrahedral, octahedral, and icosahedral. In our tests, some unsuitable starting models resembled the correct structure in surface renderings and by quantitative assessment, but not in density section (*e.g.* Fig. 3.1, Table 3.2). Therefore, visual inspection of surface renderings and density sections is prudent.

Consistent, well-defined features do not necessarily imply correct structures. When the icosahedral RMM was used on the octahedral polyomavirus data set, the resulting maps were consistent with each other, but incorrect (Fig. 3.7). The incorrect "icosahedral" capsomeres were also well resolved. Highly detailed structures, correct or incorrect, typically correlate with high variance values [126]. In our variance tests, one data set showed higher variance of suitable maps than unsuitable maps (Table 3.5); two data sets showed the opposite (Tables 3.6 and 3.7). Although p-values of the latter t-tests were greater than 0.01 and the p-value of the former test was significantly less than 0.01, map variance seems unreliable for validation.

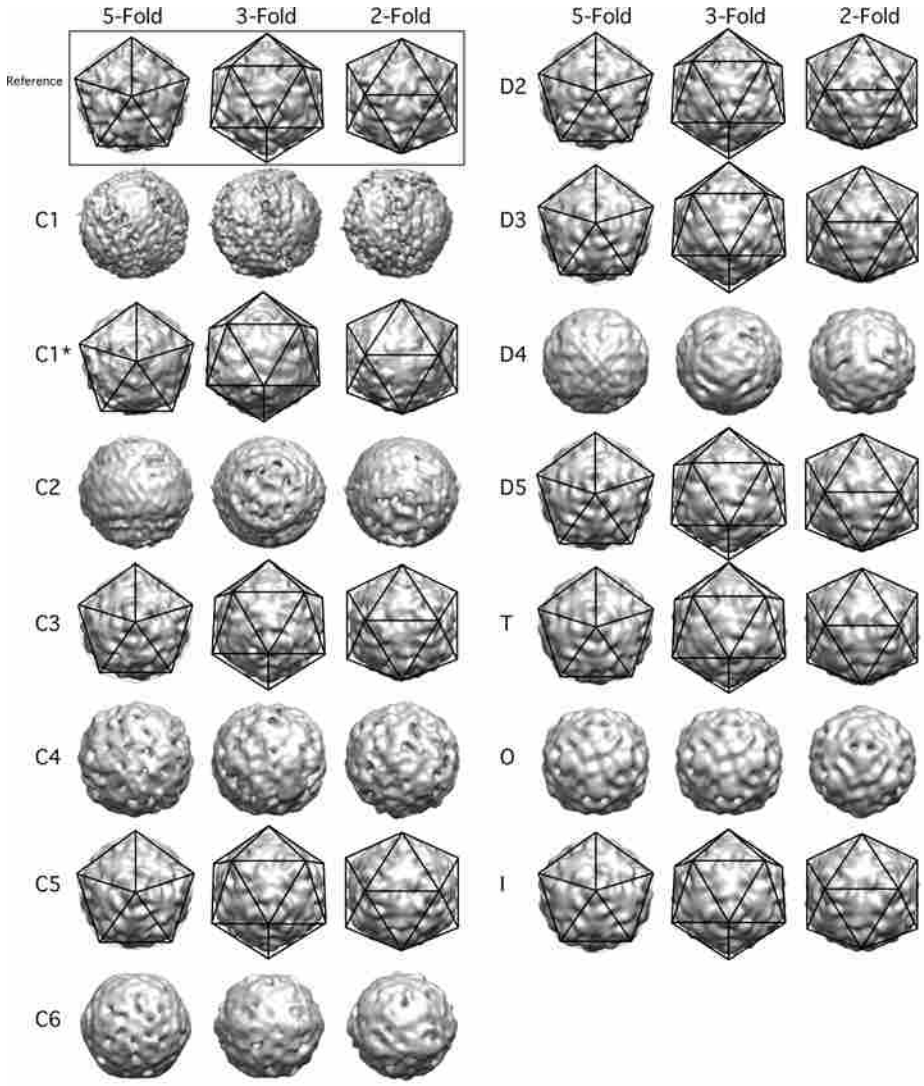


Figure 3.6 *Ab initio* reconstruction of the poliovirus data set by the random-model method. Thirteen different symmetries were tested (C1 to I). Three views, from 5-, 3-, and 2-fold axes (suitable maps) or from three different view directions, are presented for the "most suitable" reconstruction from a set of 10 independent runs with 5,016 particle images (same images were used in all tests). A published structure [106] is shown as reference (boxed, top-left). The RMM failed with C1 and C2 symmetries. Other icosahedral-compatible symmetries (C3, C5, D2, D3, D5, T, I) satisfactorily reconstructed starting models via symmetric RMMs. An icosahedral cage surrounds suitable models. The map labeled "C1∗" was computed via the asymmetric RMM with 44,000 particle images, *cf.* suitable models from symmetric RMMs (C3, C5, D2, D3, D5, T, I).

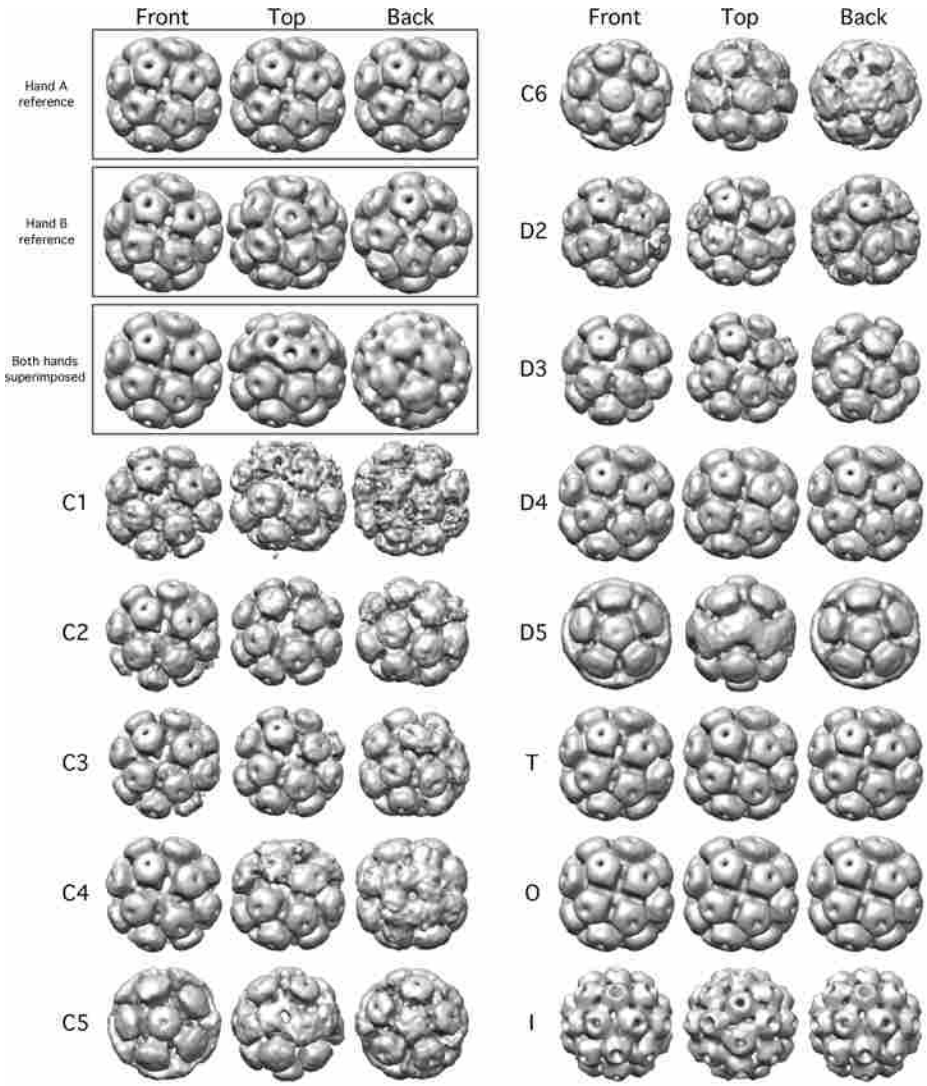


Figure 3.7 *Ab initio* reconstruction of the octahedral polyomavirus data set by the random-model method. Three orthogonal views, front, top and back, are presented for each map. Two top-left boxes, reference structures of opposite handedness with the "front" four capsomeres aligned. Bottom box, summation of hand "A" and "B" maps. Although the front view shows well defined and octahedrally arranged capsomeres, the top and back views show smeared capsomeres. Thirteen different symmetries (C1 to I) were tested (Schoenflies notation). The same image set, 4,286 particles, was used in all tests. RMM tests using asymmetry (C1) or cyclic symmetries compatible with octahedral symmetry (C2, C3, C4) produced maps similar to the map that is a sum of hands "A" and "B" (bottom box). Octahedral structures with correctly formed pentameric capsomeres were obtained only when D4, T and O symmetries were used. C1, C2, C3, C4, D2, D3, D4, and T symmetries are compatible with octahedral symmetry. (Reference structure by R. Swenson, G. Wang, P. Shen, T. Baker, D. Belnap, unpublished.)

PCA could be a useful quantitative validation tool. We used four similarity functions combined with PCA to quantitatively compare all resulting maps to each other (Section 3.3.7). The four criteria usually clustered all or most suitable maps together and apart from unsuitable maps (Fig. 3.8). In most cases, unsuitable maps did not cluster, except in the poliovirus and octahedral polyomavirus asymmetric RMM tests (Fig. 3.8E,F). However, if starting models show smeared features, as we observed in poliovirus and octahedral polyomavirus, then clustering could be disregarded.

Our FSC and CC tests were reliable and consistent with our visual and FSC assessments if the t-test (between the suitable- and unsuitable-map populations) gave a p-value < 0.01 (Tables 3.2-3.7). Otherwise, differences between the two populations were insignificant.

Plots of CCs and resolution curves could be helpful for validation. CC values displayed as histograms showed patterns consistent with average CC values (Figs. 3.1-3.5C, Tables 3.2-3.7). Starting models that deviated significantly from the correct structure had poorer FSC-resolution curves (Fig. 3.4D). If incorrect models resembled the correct structure, the FSC curves were similar (Figs. 3.1D and 3.3D). DPR, SSNR, and reliability index $R_{(A+B)}$ gave similar assessments to the FSC results (data not shown). Although a higher-resolution 3D map is more likely to be correct than a lower-resolution one, acceptable resolution does not guarantee map correctness. Maps judged unsuitable often had resolution values similar to those of suitable maps (Tables 3.2-3.7). Erratic fluctuations of resolution curves likely signify unsuitable maps [129].

For unknown structures analyzed via the asymmetric RMM, the following steps could be helpful in identifying a suitable starting model:

1. Compute multiple starting models.

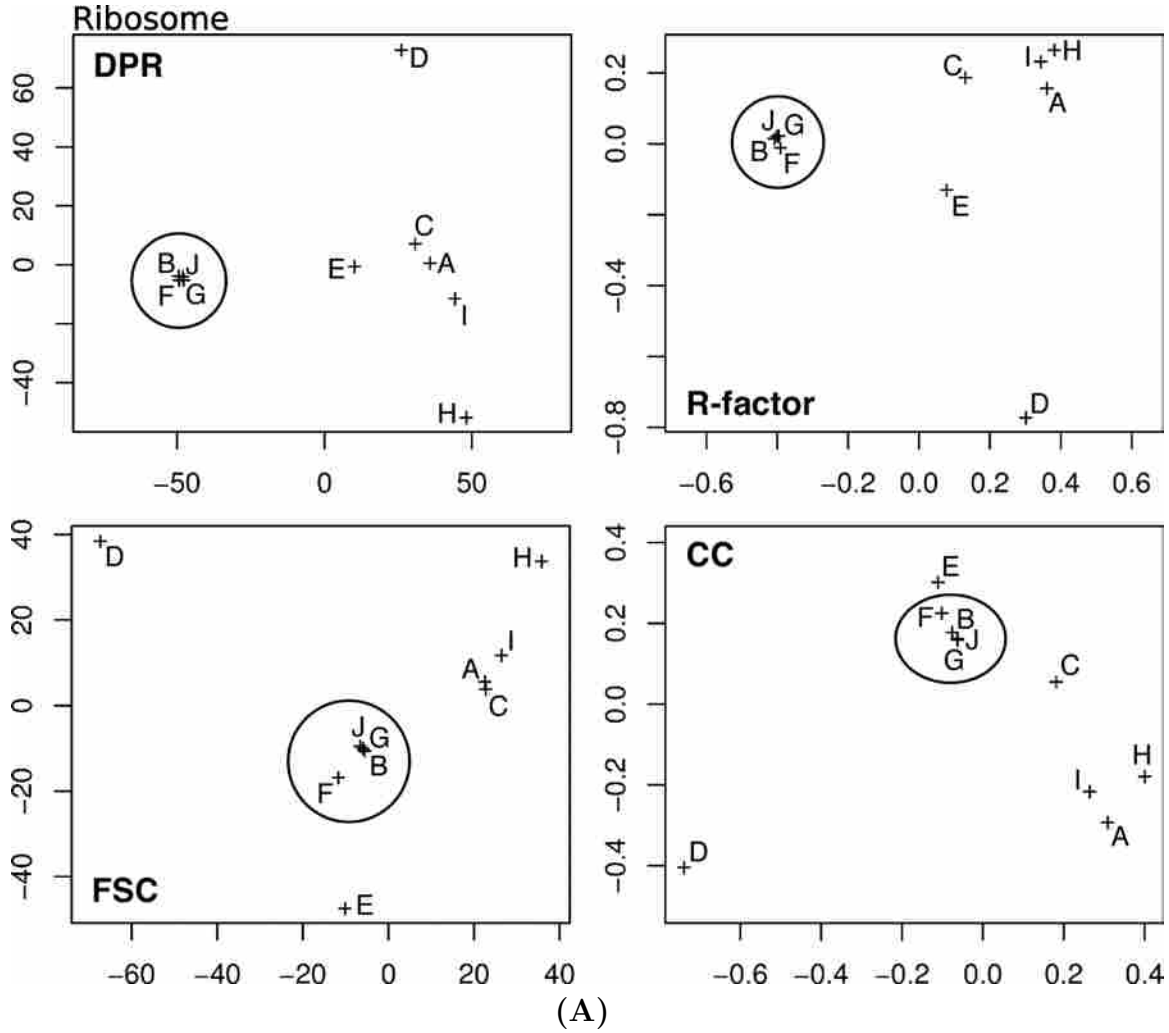
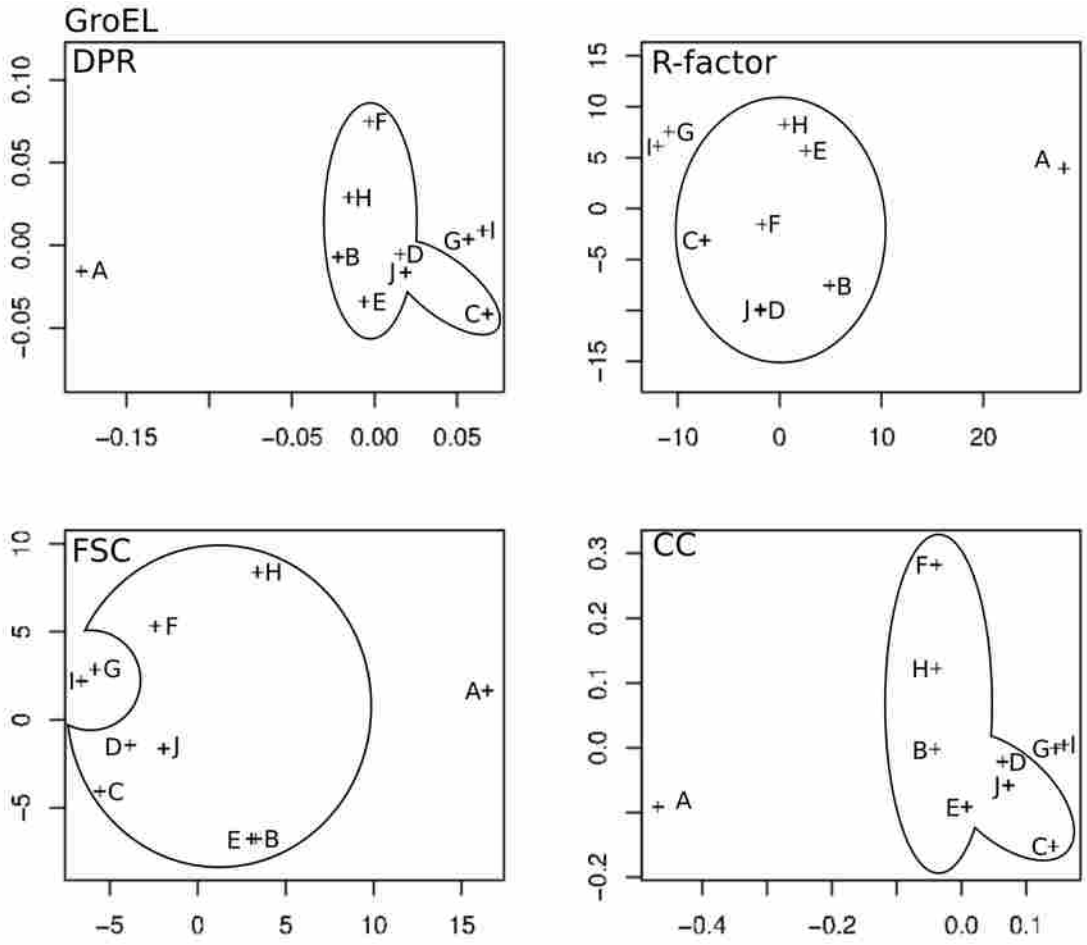
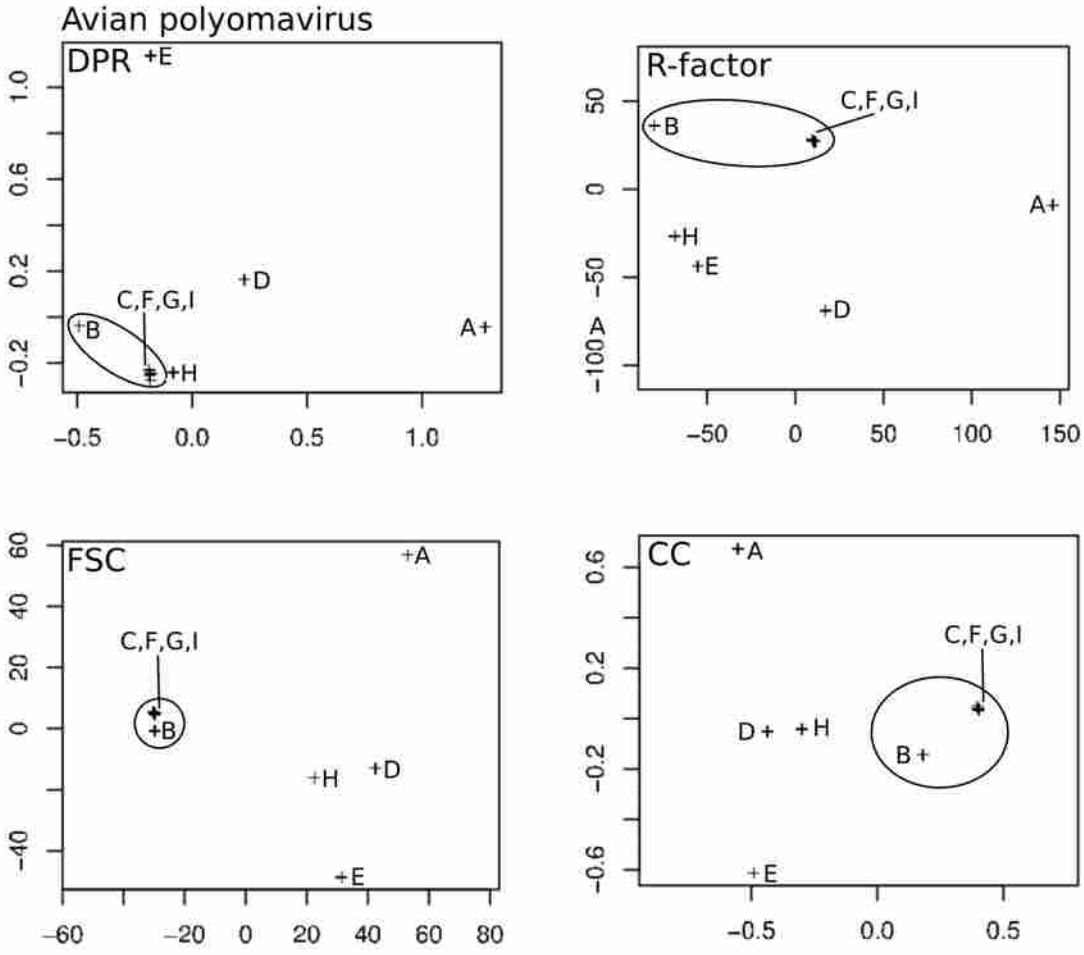


Figure 3.8 Principal component of analysis starting models obtained by the asymmetric RMM. Each plot graphically displays similarities among the resulting starting models (scales are unit-less). To get the data for each plot, each starting model was compared to every other model in the set using DPR, FSC, R-factor, or CC, as labeled. Marks indicate relative relationships among all maps. The closer the crosses, the more similar the models as assessed by the respective function and vice versa. Suitable models are encircled. (A) ribosome models, (B) GroEL models, (C) avian polyomavirus models, (D) poliovirus-receptor complex models, (E) poliovirus models, and (F) octahedral polyomavirus.



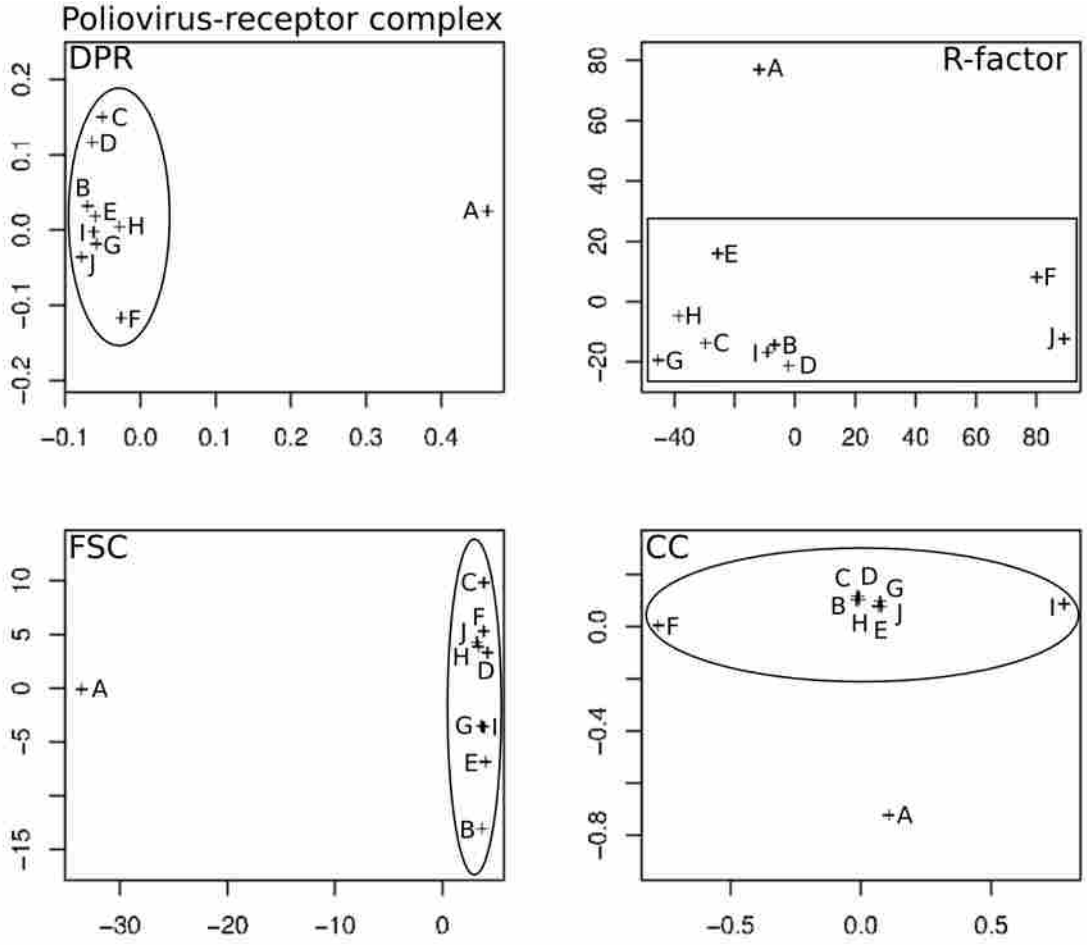
(B)

Figure 3.8 –Continued



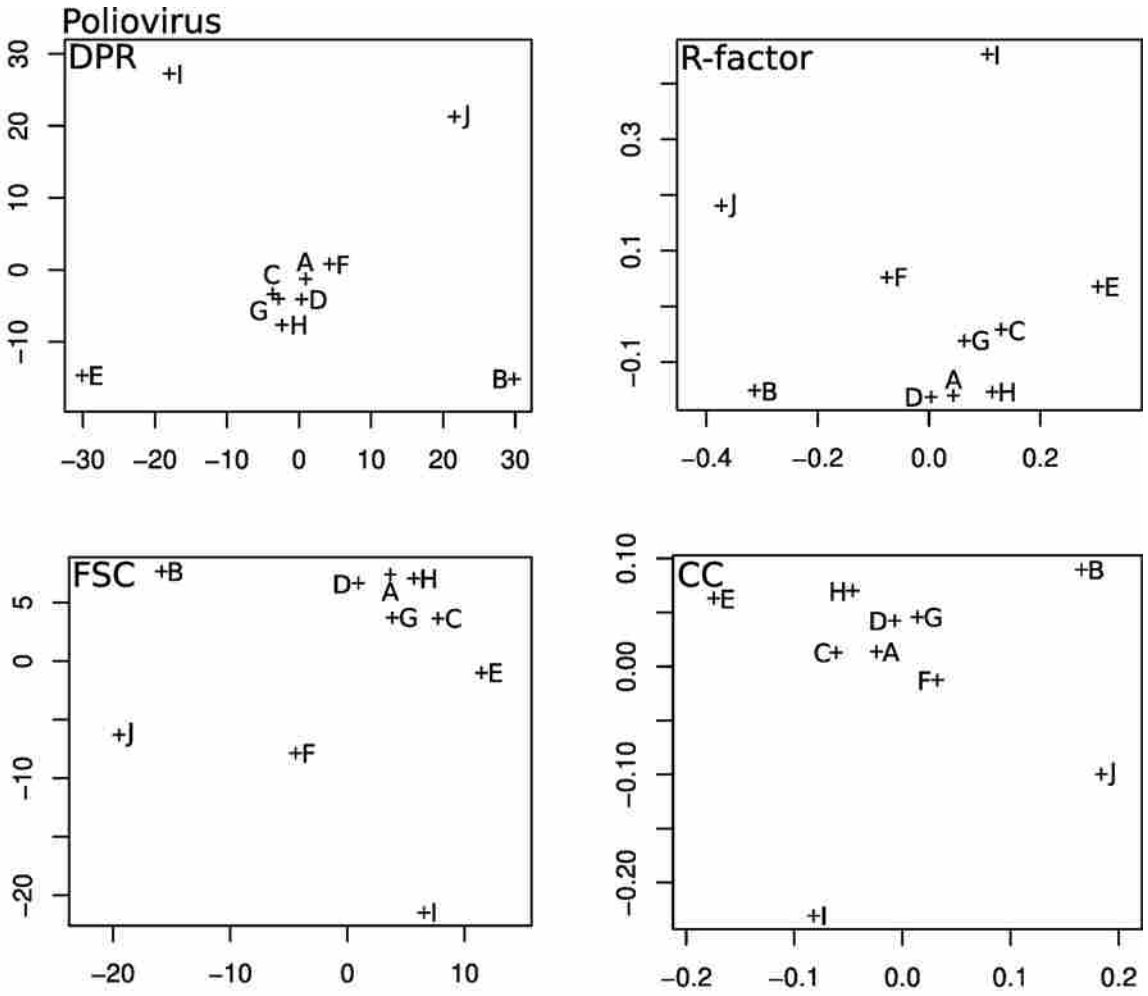
(C)

Figure 3.8 –Continued



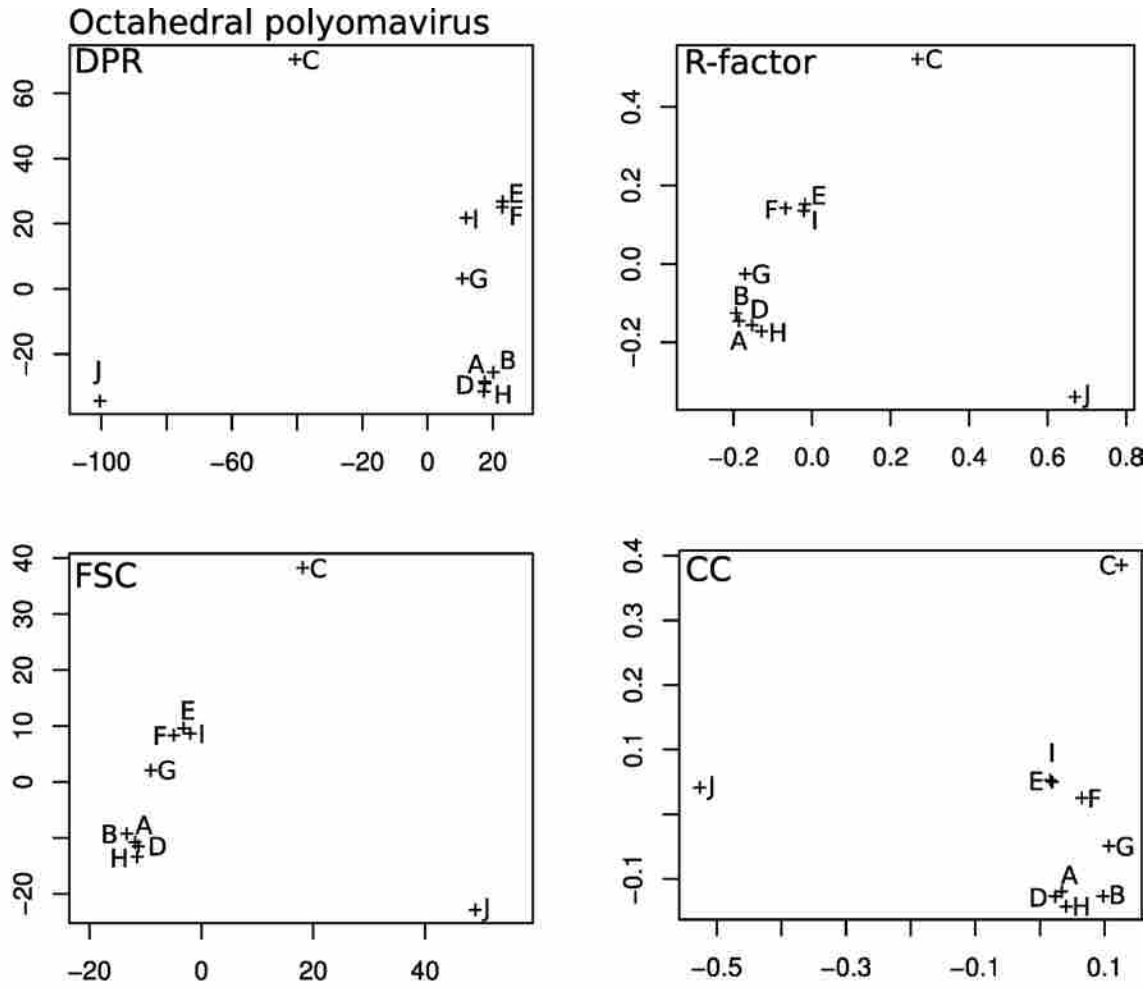
(D)

Figure 3.8 –Continued



(E)

Figure 3.8 –Continued



(F)

Figure 3.8 –Continued

2. Visualize maps. Look for consistent, well-defined structures and smeared features.

3. Use PCA to compare each map to every other map. Inspect the dissimilarity matrix or display distances graphically (via a distance-function algorithm, *e.g.* Fig. 3.8).

4. Check resolution and CC (from orientation, origin finding) results to see if they are consistent with visual inspection and PCA. Apply a t-test to compare populations of "suitable" and "unsuitable" maps.

5. Start model-based searches with the starting model. As suggested [129], if a starting model only yields a low-resolution structure despite intensive computations, then the starting model is likely unsuitable.

6. If the particle is thought to have symmetry, try the RMM with the suspected symmetry or a compatible symmetry. If correct, definition of structural features should improve.

Our experiments with known structures suggest RMM results are authentic if multiple maps a) correlate well visually and quantitatively, b) do not have smeared features, and c) differ statistically from unsuitable maps (*e.g.* PCA or p-value $<$ 0.01). Also, a suitable RMM starting model should produce an improved structure in subsequent refinement. As recently noted [97], validation of 3D reconstructions is an important problem in cryoEM studies. Current methods rely on comparisons that can be misleading—comparison of map projections to particle images or comparison of two separately computed 3D maps to each other—because incorrect structures may also correlate well. In the absence of atomic-resolution information (which is generally thought to validate structures) or complementary data, 3D maps of unknown

structures are usually accepted if multiple tests or techniques produce the same result. Therefore, validation of RMM starting models must also rely on convergence of multiple tests or subsequent production of a near atomic-resolution structure. For an unknown structure, the RMM could be used with other *ab initio* methods to test convergence of multiple *de novo* experiments. In addition, if images were taken of the same field-of-view at two tilt angles, the tilt-verification method [149] could be used.

Chapter 4

The symmetry-mismatch, random-model method

4.1 Introduction

Several cryoEM methods have been proposed to resolve asymmetric structures present in icosahedral viruses [100, 101, 150, 151]. We present a method that uses subtraction of the capsid from particle images [*e.g.* 100] but that does not employ multivariate statistical analysis, utilize common lines approaches, nor depend on prominent asymmetric features like phage tails and connectors [44, 151]. The symmetry-mismatch, random-model method (SMRMM) is a model-based three-dimensional reconstruction procedure that is an extension of the previously proposed random-model methods (Chapters 2, 3) [129, 152]. The SMRMM first resolves the higher symmetry component and relies on a fixed symmetry mismatch to then limit the random-model search for the orientation of the lower symmetry or asymmetric component [see 101, 150].

4.2 Method

4.2.1 Algorithm

We implemented the SMRMM in the following manner (Fig. 4.1):

1. An icosahedral reconstruction was computed from the virus-particle images using a model-based algorithm [111]. The origins and orientations are determined within the icosahedral asymmetric unit.

2. Once an acceptable icosahedral reconstruction was obtained, the genome density was removed computationally.

3. The 3D structure was re-projected for each icosahedral orientation (determined in step 1).

4. Densities in re-projected images were scaled to match the capsid intensities of the original particle images. Re-projections were subtracted from the original particle images resulting in "capsidless" images. This subtraction prevented the assignment of the asymmetric orientation based on icosahedral densities.

5. An asymmetric random model was generated from the subtracted particles. For each particle, the 59 other icosahedrally symmetric orientations were computed from the orientation determined in step 1, and one of the 60 orientations was chosen randomly. The set of random orientations was used to compute a 3D model without imposing symmetry. As in other random-model tests [129, 152], the model had the appearance of a nearly spherical mass.

6. The 3D model computed in step 5 served as a starting model for the asymmetric round of orientation determination. During each iterative cycle, the previously determined (step 1) icosahedral orientation again was used to compute the 60

icosahedrally related projections of the asymmetric model for each particle. A fixed symmetry mismatch means that the asymmetric alignment must be one of the 60 icosahedrally related orientations. The new orientation corresponds to the projection with the best correlation to the particle image. A final 3D model is the product of the iterative cycle.

Steps 5-6 were repeated ten or more times. In each independent test, the same images were analyzed, but different random orientations were assigned. Consequently, different random starting models were generated.

4.2.2 Image subtraction

Density scaling was achieved by computing the circularly averaged density plots for both original particle images and model re-projections [153]. We used the Bsoft [131] program *bradial* to fit the densities to a linear function ($y = ax + b$). Only densities inside an annulus that included the capsid shell were used. An identical function was used to normalize all the particles in the same micrograph, via the program *bop* (Bsoft). A sample script to achieve image subtraction is available at http://people.chem.byu.edu/belnap/pft3dr/utilities.

4.2.3 Software

We modified *PFT2* from the package PFT3DR [111, 152] (http://people.chem.byu.edu/belnap/pft3dr) to calculate the symmetry-related projections for each image particle (based on the icosahedral orientation determined in step 1). The orientation of the projection image that best matched the original image particle was assigned as the new orientation. Projection and particle images were compared following a procedure described elsewhere [see 152, Supplementary Section 2.2].

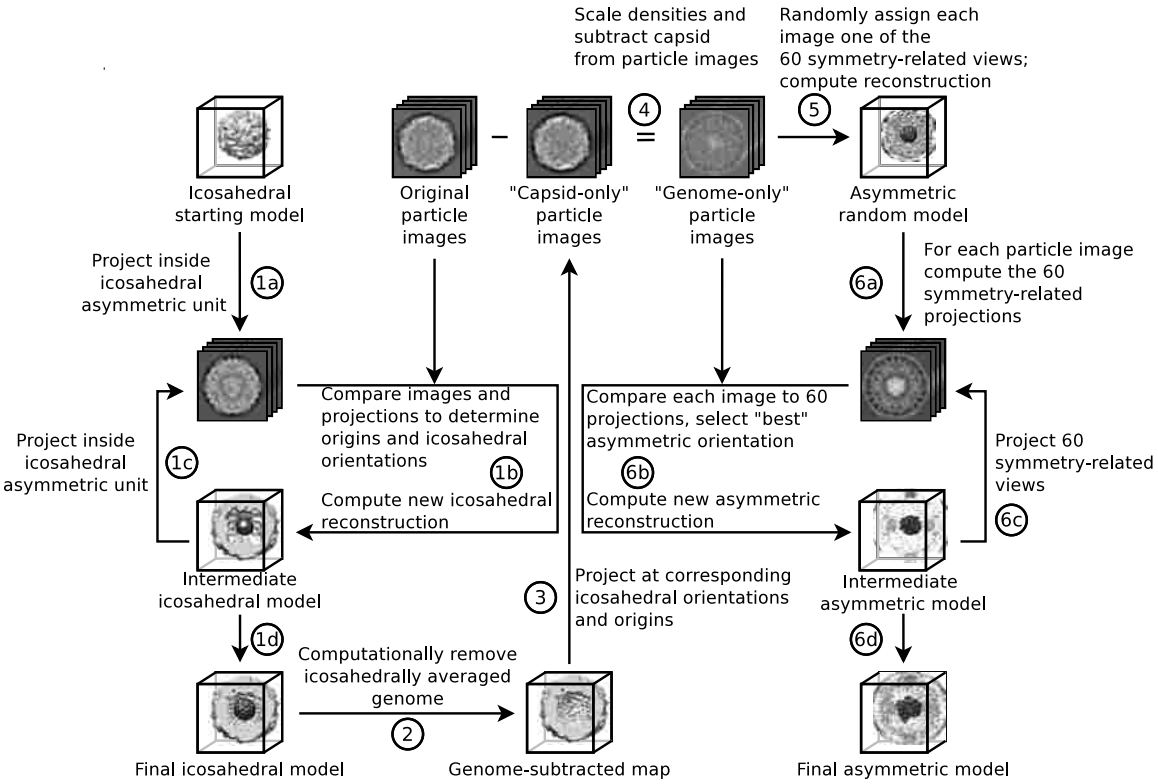


Figure 4.1 Overview of the SMRMM. The SMRMM is a pipeline of an icosahedral reconstruction (left, 1a-d) followed by an asymmetric reconstruction (right, 6a-d). (1a-d) An icosahedral model is obtained via model-based alignment procedures. The asymmetric "genome" structure (in this case a ribosome) is smeared by icosahedral averaging. (2) Genome density is removed from the final icosahedral reconstruction. (3) The empty capsid is re-projected using the determined orientations. (4) Re-projections and original image particles are density scaled and subtracted. The capsid densities of the resulting particles are greatly reduced (represented by weak density capsid). (5) One of the 60-related orientations is randomly assigned to each "genome-only" particle image and a 3D model is computed without imposing symmetry. (6a-d) For every particle, 60-related projections are computed. From this subset of projections, a new orientation is found. In our tests, only 8-10 iterations were necessary to render a final asymmetric reconstruction. Steps 5-6 are repeated ten or more times. In each independent test, the same images are analyzed, but different random orientations are assigned.

The Bsoft programs *breconstruct* [131] and *emgrand* [152], respectively, were used to 1) generate the asymmetric reconstructions and 2) randomly select orientations, related by icosahedral point-group symmetry, to create the initial (random) asymmetric model. A script to perform steps 5 and 6 is available at http://people.chem.byu.edu/belnap/pft3dr/utilities.

4.2.4 Origin determination

Because the origin determined during the icosahedral orientation search should be the same for the asymmetric structure, origin refinement during the asymmetric orientation search may not be necessary. Reuse of the icosahedrally determined origin may also prevent spurious origin changes from an origin determination based on only image intensities from asymmmetric densities. We successfully determined asymmetric orientations without modifying the icosahedrally determined origin.

4.3 Results and discussion

We tested if an asymmetric object could be reconstructed from icosahedrally smeared random models. In a first attempt we avoid the capsid subtraction. 5,000 synthetic images were generated by projecting a left hand model and an *E. coli* 70S ribosome (EMDB 1006 [132]). We successfully reconstructed the left hand and the *E. coli* 70S ribosome using synthetic "capsid-less" images (Fig. 4.2).

Mimicking a more realistic scenario, the subtraction of the capsid component was tested. 8,000 synthetic image particles were created by inserting an *E. coli* 70S ribosome (EMDB 1006 [132]) inside an icosahedral poliovirus capsid [106]. This artificial model was projected in random orientations using *emgrand* (Bsoft) and *bint* (Bsoft). Gaussian noise was added with *brandom* (Bsoft) to simulate conventional cryoEM

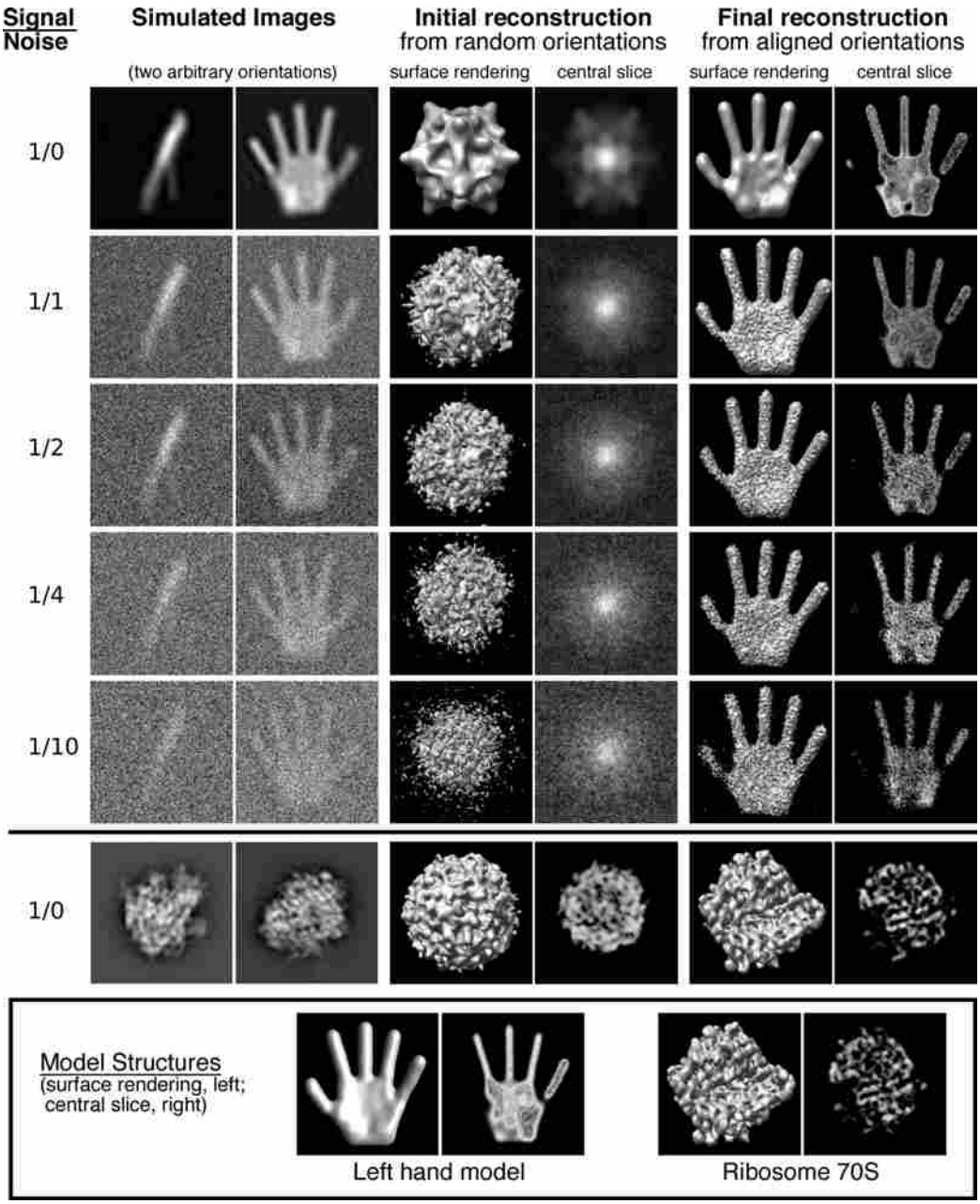


Figure 4.2 Validation of the SMRMM using model data. Model data was used to test if an asymmetric object could be reconstructed from an icosahedrally smeared reference model. Model structures (bottom row) were projected to make simulated projections (2nd and 3rd columns from left). Except for "noiseless" ("1/0") data, noise was added to give the specified signal-to-noise level (left column). Random models were generated by assigning random icosahedral orientations to each simulated image (4th and 5th columns from left). The final reconstruction is shown in the two rightmost columns.

images. Three signal-to-noise ratios (SNR) were tested including absence of noise ("1/0"), 1/10 and 1/100. We obtained successful asymmetric reconstructions for all three SNRs (Fig. 4.3) with success rates (percentage of tests giving the correct model) of 100%. The starting model for the SMRMM showed smeared densities. In these control tests, the correct ribosome structure emerged in less than 10 iterations. Spurious densities resulting from imperfect image subtractions (Fig. 4.1, step 4) seemed not to interfere in the final result.

Additionally, we assessed the SMRMM with 1,000, 10,000 and 75,000 experimental cryoEM images of *E. coli* 70S ribosome [132] in a semi-artificial scenario, as if the ribosome images were derived from an icosahedral shell. The intensity value of the background was set to match the mean of intensities between a radial mask and the edge of the image. The images were corrected for phase-flipping effects of the contrast transfer function [152]. First, we determined the asymmetric orientations and origins using a ribosome model (EMDB 1055 [132]). Second, we randomly assigned, to each imaged particle, one of its 60 symmetry-related orientations and computed a 3D reconstruction. This step produced a smeared 3D model. Third, we tested the SMRMM to see if the correct ribosome model could be derived from the smeared structure. Successful reconstructions were observed in 10 iterations or less, as with the synthetic images (Fig. 4.3). Success rates varied from 70% (1,000 particles) to 100% (10,000 and 75,000 particles). The percentage of particles being assigned a correct orientations also improved with higher number of particles. When 1,000 particles were used, 80% of the particles had orientations that were less 0.2° apart (on a sphere) from the expected orientations, improving to 95% when 75,000 particles were used.

Because the SMRMM uses random models as starting models we refer the reader to [129] and Chapter 3 [152] for a discussion about validation of 3D maps produced by the random-model method. Final asymmetric SMRMM models could be used as

starting maps for additional model-based refinements.

Our observation with experimental ribosome images suggests that if the asymmetric component is uniformly ordered with respect to the icosahedral shell, then the asymmetric structure should be resolvable by the SMRMM. Work to test the SMRMM with experimental virus images is ongoing. In this effort, success of the SMRMM could be hindered if the viral genome is structurally heterogeneous, where, in the best scenario, the SMRMM would result in an average of all the sub-structures present in the sample. Additionally, the viral genome might not always assume a fixed orientation relative to the symmetric shell, an absolute requirement for the success of the SMRMM. Provided the components are fixed relative to each other, the SMRMM is, theoretically, generalizable to all symmetry mismatches.

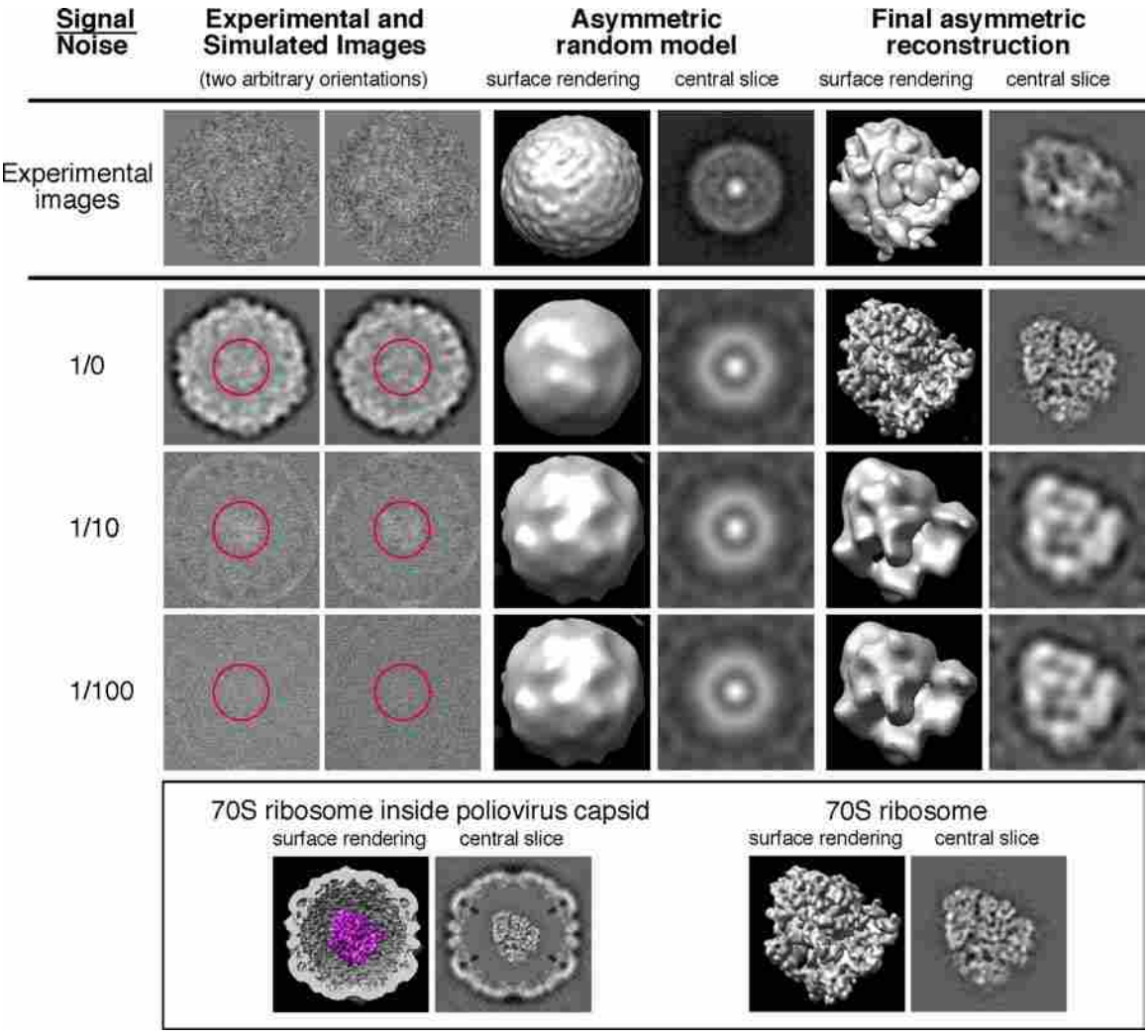


Figure 4.3 Validation of the SMRMM using experimental and model data. The data set is designated in left column. Columns two and three (from left, "Experimental and Simulated Images") are two representative images from the data set. Columns four and five ("Asymmetric random model") show the asymmetric random model (from step 5, see Fig. 4.1). Columns six and seven ("Final asymmetric reconstruction") show the final result (step 5, Fig. 4.1). Columns 4-7 show either a surface rendering or the central density slice of the 3D reconstructions. First row, the SMRMM was tested using experimental ribosome images [132]. Second to fourth rows, the SMRMM was assessed using the synthetic model data. Red circles outline the ribosome in non-closeup views. For those rows with a number in the left-most column, noise was added to give the specified signal-to-noise ratios ("1/0" indicates the lack of applied noise). An asymmetric random model was produced by assigning randomly one of the 60-related orientations (from the icosahedral reconstruction) to each image particle. Boxed inset (bottom row) shows model data used with ribosome [132] placed inside poliovirus capsid (left) and the 70S ribosome reconstruction alone (right).

Chapter 5

Poliovirus

5.1 Background

Poliovirus is a small nonenveloped virus from the *Enterovirus* genus, *Picornaviridae* family [104]. Poliovirus occurs in three serotypes, termed types 1, 2, and 3. Polioviorus, together with other members of this family, such as, rhinovirus, foot-and-mouth disease virus, and hepatitis A, are important human pathogens. Because of its implications to human health, poliovirus has been the focus of extensive research. As a result, a number of groundbreaking achievements have been associated with poliovirus: (i) first virus to be grown in cell cultures [105]; (ii) first modern mass vaccination campaign, that has lead to nearly complete eradication; (iii) one of the first animal viruses to be resolved by X-ray crystallography; (iv) first virus to be artificially synthesized and assembled from its genetic code [154]. Despite this wealth of knowledge, several areas of the biology of poliovirus are not yet well understood, such as cell entry [106, 140, 155], genome release [156], tropism to the central nervous system [157] (where it causes acute paralysis of the limbs), and 3D genome organization, the topic of this study.

The capsid of poliovirus is a 33-nm diameter, pseudo T=3 icosahedral shell composed of 60 copies of the VP1 (33 kDa), VP2 (30 kDa), VP3 (26 kDa), and VP4 (7.5 kDa) [105]. The total molecular mass of poliovirus, including its genome, is 8.5×10^6 Da. Autocatalytic cleavage of VP0, also referred to as "maturation cleavage", leads to VP2 and VP4. The cleavage of VP0, during the late stage of virus assembly, is essential for the stability of the viral capsid [158]. VP1, VP2, and VP3 are genetically unrelated (Fig. 5.1), but they share a common structure termed β-barrel jelly roll (Fig. 5.2). This motif is formed by two antiparallel β-sheets packed against other and twisted around a barrel. This topology is the most common structure among viruses, and is found in animal and plant viruses [159], as well as RNA and DNA viruses. VP4 is a myristolated protein with an extended structure [105]. VP4 has been predicted to form amphitathic helices and to be involved in the formation of membrane pores [140].

Three different conformations of poliovirus have been described [106], each of them characterized by their sedimentation coefficient (Svedberg units). The 160S corresponds to the native structure, 135S to a putative cell-entry intermediate, and 80S is the empty virion (Fig. 5.3). This conformation involves an expansion and thinning of the capsid shell. VP4 is exposed to the outer surface in the 135S conformation [140].

5.1.1 Pseudo-crystalline structure of poliovirus RNA

Poliovirus genome consist of a single (+) ssRNA molecule of ∼7,500 nucleotides. The genome, which has been completely sequenced [161], acts both as a messenger RNA and as template for replication. Translation of the poliovirus genome produces a single polyprotein precursor which is cleaved to produce all the structural and non-structural proteins necessary for replication and assembly. The poliovirus genome also acts as a template for the synthesis of the negative-strand RNA.

```
VP1       GLGQMLESMIDNTVRETVGAATSRDALPNTEASGPTHSKEIPALTAVETGATNPLVPSDT 60
VP2       --SPNIEACGYSDRVLQLTLGNSTITTQEAANSVVAYGRWPEYLRDSEANPVDQPTEPDV 58
VP3       --GLPVMNTPGSNQYLTADNFQSPCALPEFDVT-PPIDIPGEVKNMMELAEIDTMIPFDL 57
            .  :    .          *  :  :   :  . .         *    :    *

VP1       VQTRHVVQHRSRSESSIESFFAR-GACVTIMTVDNPASTTNKDKLFAVWKITYKDTVQLR 119
VP2       AACRFYTLDTVSWTKESRGWWWKLPDALRDMGLFGQNMYYH---------------YLG 102
VP3       SATKKNTMEMYRVRLSDKPHTDDPILCLSLSPASDPRLSHT---------------MLG 101
             :  . .     . .       .:     .                       *

VP1       RKLEFFTYSRFDMELTFVVTANFTETNNGHALNQVYQIMYVPPGAPVPEKWDDYTWQTSS 179
VP2       RSGYTVHVQCNASKFHQGALGVFAVPEMCLAGDSNTTTMHTSYQNANPGEKGGTFTGT-- 160
VP3       EILNYYTHWAGSLKFTFLFCGSMMATG-------KLLVSYAPPGADPPKKRKEAMLGT-- 152
          .           ::     . :  .            :..     * :      *

VP1       NPSIFYTYGTAPARISVPYVGISNAYSHFYDGFSKVPLKDQSAALGDSLYGAASLNDFGI 239
VP2       --FTPDNNQTSPARRFCPVDYLLGNGTLLGNAFVFPHQIINLRTN-----NCATLVLPYV 213
VP3       -HVIWDIGLQSSCTMVVPWISNTTYRQTIDDSFTEG-------------GYISV----- 192
                    :..    *          : :.*                .  ::

VP1       LAVRVVNDHNPTKVTSKIRVYLKPKHIRVWCPRPPRAVAYYGPGVDYKDGTLTPLSTKDL 299
VP2       NSLSIDSMVKHNNWGIAILPLAPLNFASESSPEIPITLTIAPMCCEFNG--LRNITLPRL 271
VP3       -------------------FYQTRIVVPLSTPREMDILGFVSACNDFSVRLLRDTTHIEQ 233
                                      *.    :       ::.   *   :

VP1       TTY-- 302
VP2       Q---- 272
VP3       KALAQ 238
```

Figure 5.1 Sequence alignment of poliovirus VP1, VP2, and VP3 using *Clustal W* [160]. Degree of conservation is denoted by the following symbols: "*", identical residues; ":", conserved substitutions; ".", semi-conserved substitutions; blank, no chemical similarity.

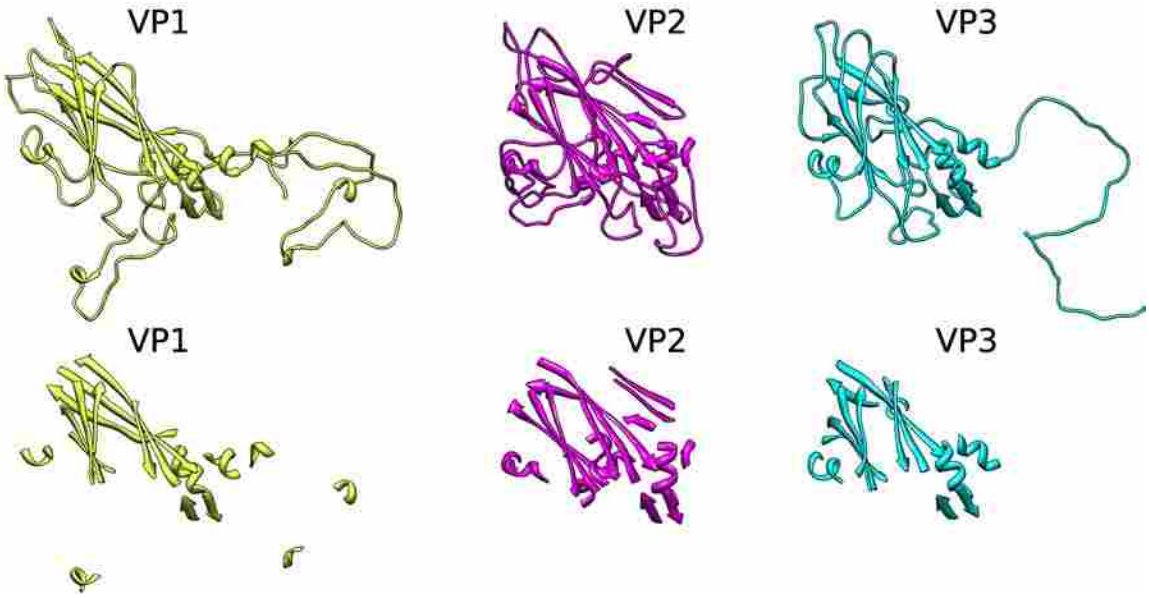


Figure 5.2 Structural homology among poliovirus VP1, VP2, and VP3 capsid proteins. Structural alignment of VP1 (yellow), VP2 (magenta) and VP3 (cyan) using *UCSF Chimera* [145]. Top panel, loops show a high degree of variability. Bottom panel, only β-sheets and α-helices are displayed. Note the striking similarity of the β-barrel jelly rolls.

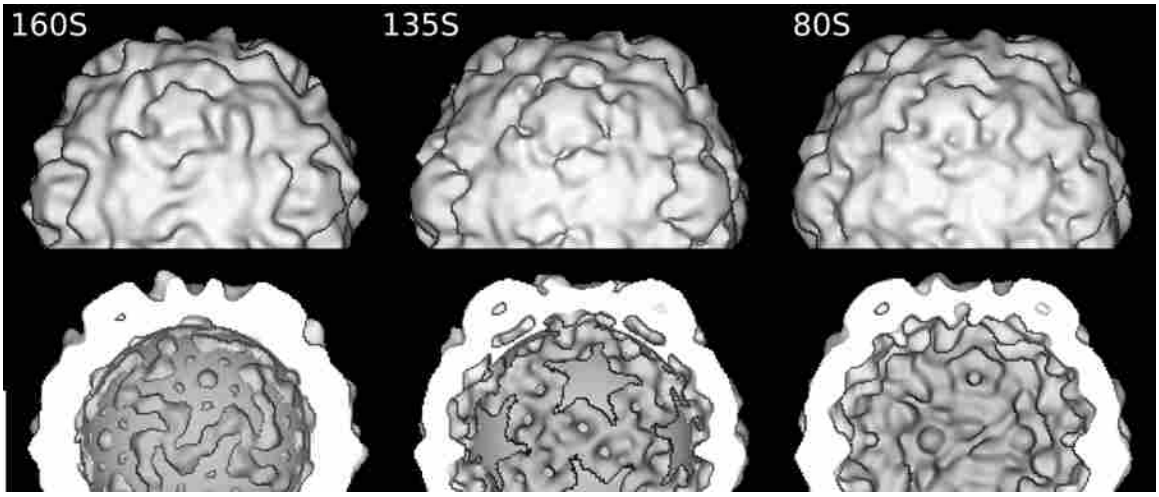


Figure 5.3 CryoEM reconstructions of poliovirus 160S, 135S and 80S conformations [106]. Top panel, surface rendering of outer views. Bottom panel, surface rendering showing the inner capsid surface (RNA density removed in 160S and 135S structures). Poliovirus capsid experiences an expansion and thinning from the native structure (160S) to the empty virion (80S)

Some portions of the genome are known to adopt stable secondary structures [4, 5, 162]. The genome can be subdivided into three main regions (Fig. 5.4):

A. The 5' untranslated region harbors a significant degree of secondary structure.

B. The middle portion encloses the open reading frame. This region connects both untranslated regions and accommodates a cis-acting element, formed by a 61-nucleotide stem-loop, essential to the virus replication.

C. The 3' end contains a polyadenylated tract of variable length. In contrast to the 5' untranslated region, the 3' end is comparatively less structured.

Poliovirus 5' end is linked to VPg, a 22-amino acid peptide. This region also contains a cloverleaf motif and a potent internal ribosome entry site (IRES). Poliovirus IRES mediates cap-independent translation of the viral genome by direct interaction with the 40S ribosome subunit [104]. Mutations in the IRES attenuate the virulence of poliovirus infection by reducing the level of translation [162]. Mutations in

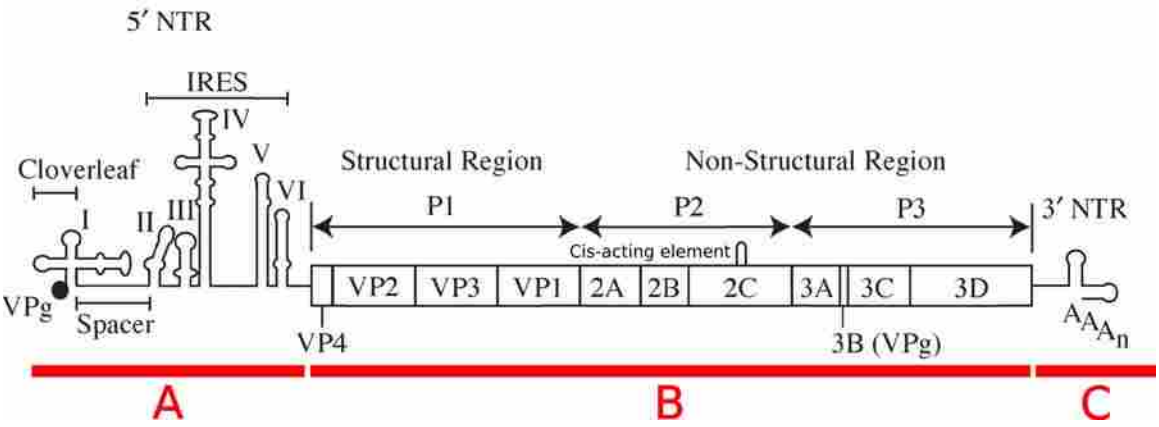


Figure 5.4 Genome organization of poliovirus [162]. The genome consists of (+) ssRNA, which encodes a single polyprotein. (A) The 5' end is covalently linked to VPg peptide and contains two essential structures: the cloverleaf and the internal ribosome entry site (IRES). (B) The open reading frame connect to the 3' untranslated region (C) which is poly-adenylated.

the IRES are the primary attenuating factor in oral poliovirus vaccines causing low neuropathogenicity, while inducing immunity. IRES structures have been directly visualized by EM and are able to fold *in vitro* [163].

Poliovirus RNA is densely packed inside the capsid shell [92]. Table 5.1 shows representative V_m values for several virus genomes. V_m is a crystallographic parameter that indicates the solvent content, and is defined as the volume occupied per dalton of biological material (inverse of density). As a reference, dehydrated RNA has a V_m value of 0.91 Å^3/Da, the minimum possible, and indicates the maximum degree of packaging (0% solvent content). Typical RNA crystals have slightly more than 50% solvent content. Note how human rhinovirus, structurally similar to poliovirus, packages its RNA genome more tightly than RNA crystals.

Bioinformatics studies have suggested that poliovirus genome lacks genome-scale ordered RNA structure [72]. Hybridization experiments have also indicated that *in vitro* transcripted products from full-length poliovirus clones are largely accessible

Table 5.1 Experimentally determined packing volumes for RNA in representative virus particles and a protein crystal [92].

Virus	RNA M_r ($\times 10^{-6}$)	Volume ($\times 10^{-6}$Å^3)	V_m (Å^3/Da)
Cowpea mosaic virus			
RNA 1	2.02	4.32	2.14
RNA 2	1.22	4.32	3.54
Southern bean mosaic virus	1.39	4.32	3.11
Cowpea chlorotic mottle virus	1.20	3.59	3.00
Satellite tobacco necrosis virus	0.34	0.97	2.87
Crystalline $U(UA)_6A$	0.0336	0.0733	2.18
Human rhinovirus 14	2.55	4.32	1.69
Dehydrated RNA			0.91

to hybridization probes in solution, indicating that poliovirus genome assumes an extended structure [73].

Two AFM studies addressed the packaging nature of poliovirus genome. Davis *et al.* [73] reported a pleomorphic and unfolded conformation of the RNA, and Kuznetsov *et al.* [76] visualized tightly compacted spheroids with a diameter corresponding to the inner diameter of poliovirus caspid (30 nm). Moreover, Kuznetsov *et al.* showed that the unfolding (by high temperature) and folding (by cooling) events are reversible. Other virus genomes observed to experience reversible contractions/expansion processes upon shifts in pH and ionic strength [61]. Although not fully explicable, discrepancies between Davis *et al.* and Kuznetsov *et al.* could be due to the difference in RNA isolation: Davis *et al. in vitro* transcribed poliovirus RNA, while Kuznetsov *et al.* directly extracted the RNA from intact viruses.

5.2 Material and methods

5.2.1 CryoEM and image processing

Poliovirus was purified following a protocol described elsewhere [164]. For this study two different data sets were used. A first data set was recorded at Harvard Medical School (Boston, Massachusetts, USA). CryoEM images were recorded in a Philips CM200 (Hillsboro, Oregon, USA) equipped with Gatan 626 cryoholder (Pleasanton, California, USA), at 200 keV, 50,000× magnification, at low electron dose. Micrographs were recorded on Kodak S0-163 and digitized on a SCAI scanner (Z/I Imaging, Huntsville, Alabama, USA). 4,500 defocal pairs were extracted using *X3DPREPROCESS* [133].

Because 60× more images were needed for the asymmetric reconstruction of poliovirus RNA, a second data set was obtained at Brigham Young University (Fig. 5.5). 3.5 μL from a $\sim$0.1 mg/mL suspension was applied to a homemade holey carbon grid.

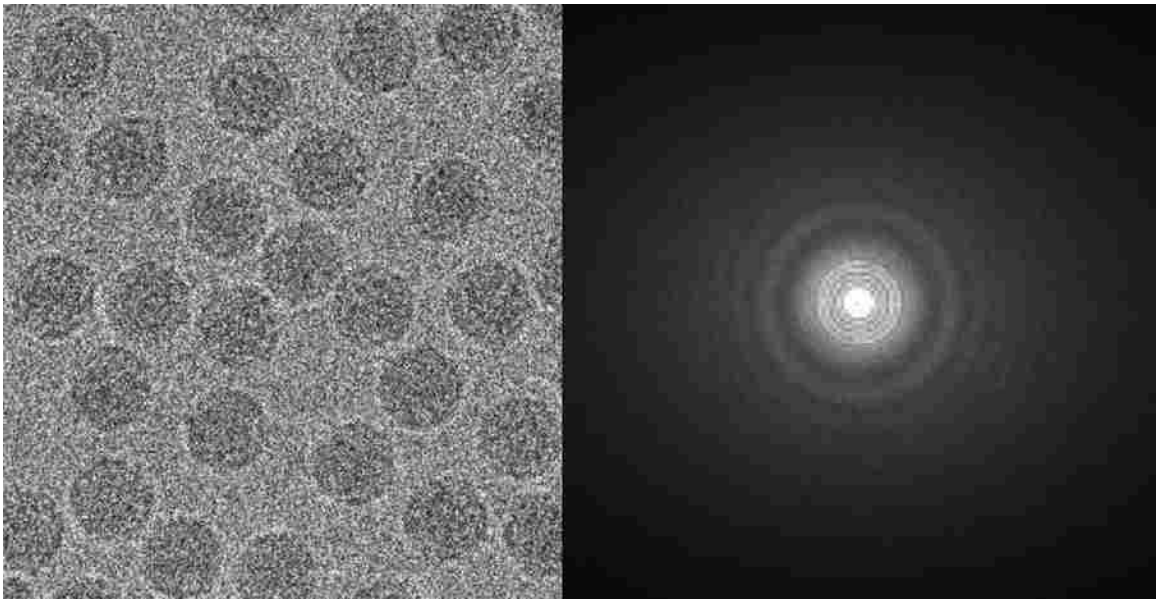


Figure 5.5 Frozen-hydrated poliovirus particles imaged by cryoEM obtained at Brigham Young University. Left, micrograph showing field of concentrate particles. Right, power spectrum of micrograph indicating that the micrograph is of good quality and can be used for three-dimensional reconstruction. The particles are in the native (160S) state. The FEI Tecnai F30 instrument was operated at 300 keV and at a magnification of 39,000×. The defocus is approximately 3 μm underfocus. The particles are approximately 30 nm in diameter.

The grids were vitrified in liquid ethane using a Vitrobot (FEI) vitrification device. Specimens were transferred to a Gatan 626 cryoholder (Pleasanton, California, USA) cooled with liquid nitrogen. CryoEM images were recorded in an FEI Technai F30 transmission electron microscope (Hillsboro, Oregon, USA) at 300 keV, 39,000× magnification, using a low electron dose procedure on Kodak SO-163 film. Underfocus settings ranged from 1.4–5.7 μm. Electron micrographs were digitized on a Nikon Super Coolscan 9000 ED scanner in 16-bit, negative mode. 22,000 defocal pairs were extracted from electron micrographs using *X3DPREPROCESS* [133]

For both data sets, we determined and corrected for contrast transfer function

and signal-decay parameters using *bshow* and *bctf* (Bsoft [131]).

Origins and orientations of the extracted particles were determined using the model-based technique of *PFT2*, which was adapted to use phase and amplitude information in orientation selection [see 152, Supplementary Section 2.2].

5.3 Icosahedral reconstruction

Using the images recorded at Brigham Young University, we resolved the capsid structure of poliovirus at 11 Å (FSC 0.5 criteria; Fig. 5.6). The final 3D reconstruction was calculated using *EM3DR2* [165] imposing icosahedral symmetry. This resolution gave us confidence in the quality of the images and in the potential to resolve the genome structure.

5.4 Asymmetric reconstructions

We tried to resolve the structure of poliovirus RNA using both the asymmetric RMM (Chapters 2, 3) and the SMRMM (Chapter 4). Table 5.2 summaries our tests and includes reconstruction parameters used. Additionally, we experimented with other approaches, such as, common lines (from raw images or class averages), and starting 3D reconstructions from random blobs. We found that all our reconstructions of the poliovirus RNA were inconsistent between each other. This could indicate that the viral capsid dominates the orientation refinement [17], or that the genome is disordered. Currently, we are investigating other subtraction methodologies and the feasibility of imaging and resolving the structure of extracted RNA cores (Section 7.1).

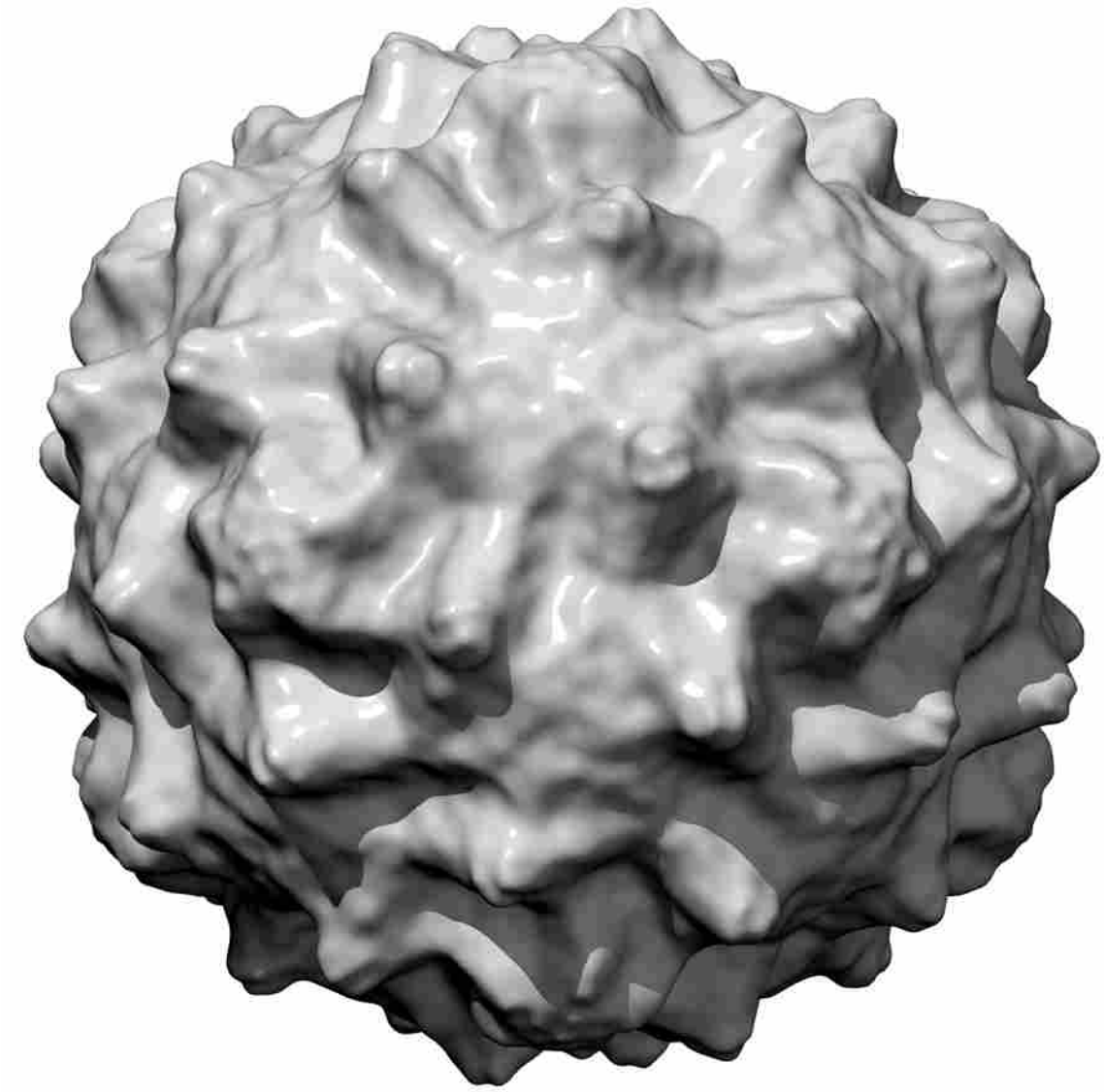


Figure 5.6 Poliovirus icosahedral reconstruction.

Table 5.2 Asymmetric reconstructions of poliovirus[a]

Date	Capsid-subtracted images	Method	Number of particles[b]	Resolution cutoffs (Å)[c]	Defocal pairs merged	Accelerating voltage (keV)	Nominal magnification	Number of repetitions
Feb/2006	No	SMRMM	4,500	25, 90	No	200	50,000	4[d]
Feb/2006	No	SMRMM	4,500	20, 90	No	200	50,000	2[d]
Mar/2006	No	SMRMM	4,500	20, 90	Yes	200	50,000	2[d]
Jul/2007	No	SMRMM	4,500	17, 90	Yes	200	50,000	21
Mar/2006	No	SMRMM	4,500	25, 90	Yes	200	50,000	3[d]
Jul/2006	Yes	SMRMM	4,500	20, 90	No	200	50,000	3
Sep/2007	Yes	SMRMM	4,500	20, 90	No	200	50,000	7[e]
Oct/2007	No	SMRMM	22,000	40, 90	Yes	300	39,000	10
Apr/2008	Yes	SMRMM	22,000	40, 90	No	300	39,000	10
Apr/2008	No	ARMM	22,000	40, 90	Yes	300	39,000	10
Nov/2009	Yes	ARMM	22,000	25, 90	No	300	39,000	20[f]
Oct/2010	Yes	ARMM	22,000	25, 90	No	300	39,000	114

[a] All data sets were CTF corrected for the phase component only: inversion of the phases caused by aberrations in the imaging system [131, 136].
[b] Refers to number of defocal pairs
[c] Low and high resolution cut-offs for orientation and origin determination. 3D reconstructions were typically computed at a maximum resolution equal to the low resolution cuf-off - 2 Å.
[d] Different particle selection schemes, reconstruction programs and origin refinement methodologies were tested.
[e] Random models were computed using half of the particle images.
[f] gradual and non-uniform decrease of the angular step-size

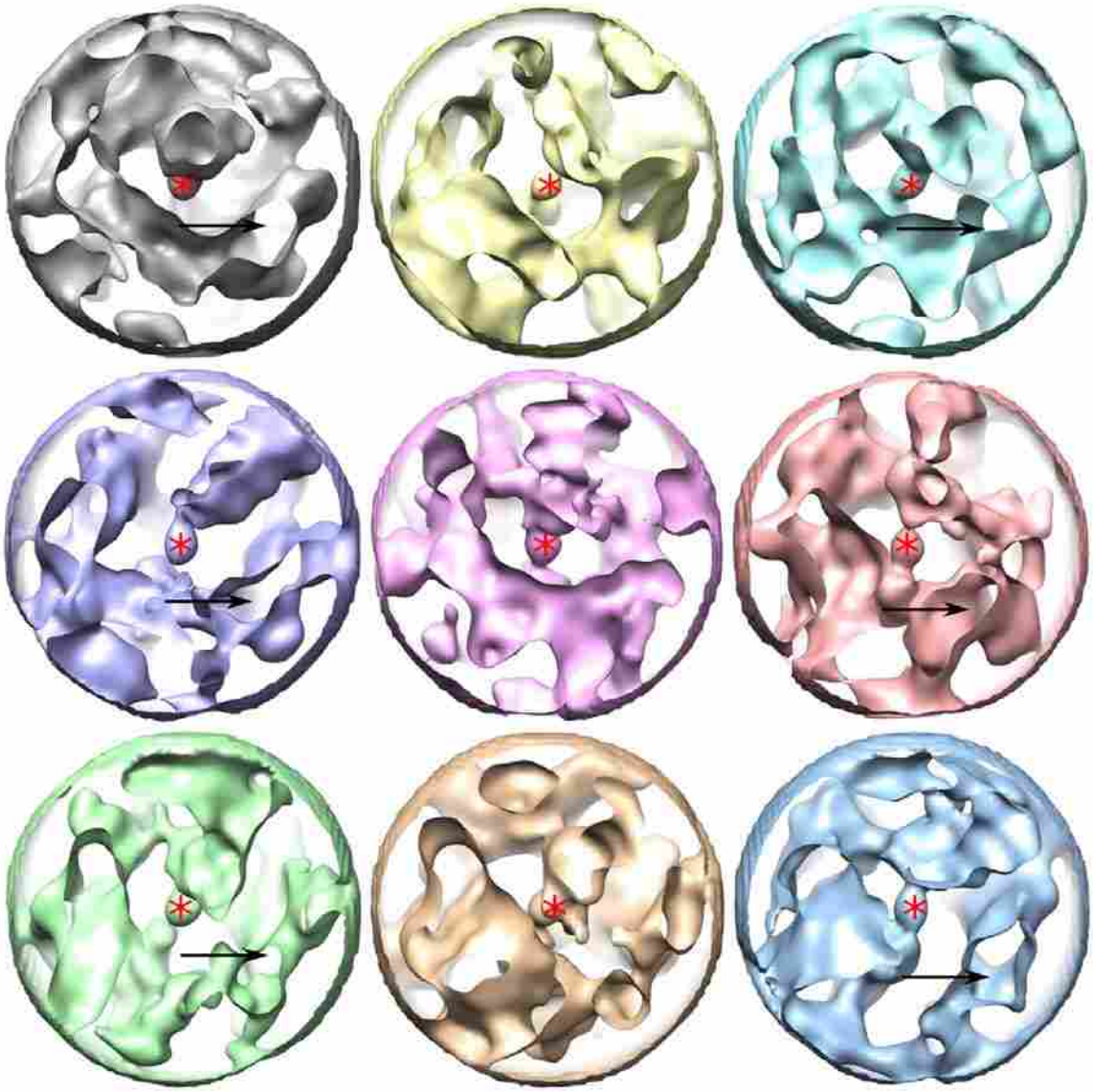
5.4.1 Asymmetric random-model method

Our results with the asymmetric RMM indicate some degree of homology among the reconstructions (Fig. 5.7). The adoption of common motifs could indicate the presence of several populations of alternative structures that share common foldings. We are currently investigating this hypothesis by applying classification algorithms to unravel the heterogeneity among the data set.

Previously, we observed the failure of the asymmetric RMM to resolve the capsid structure of poliovirus (Section 3.3.4). The reason for the failure was the lack of structural features which could help in the alignment process. We were able to partially overcome this deficiency with increased number of particle images. It is likely, that the outer radii of the RNA also adopts an spheroid structure, hence, it is expected to observe similar behavior when the asymmetric RMM is applied to poliovirus RNA.

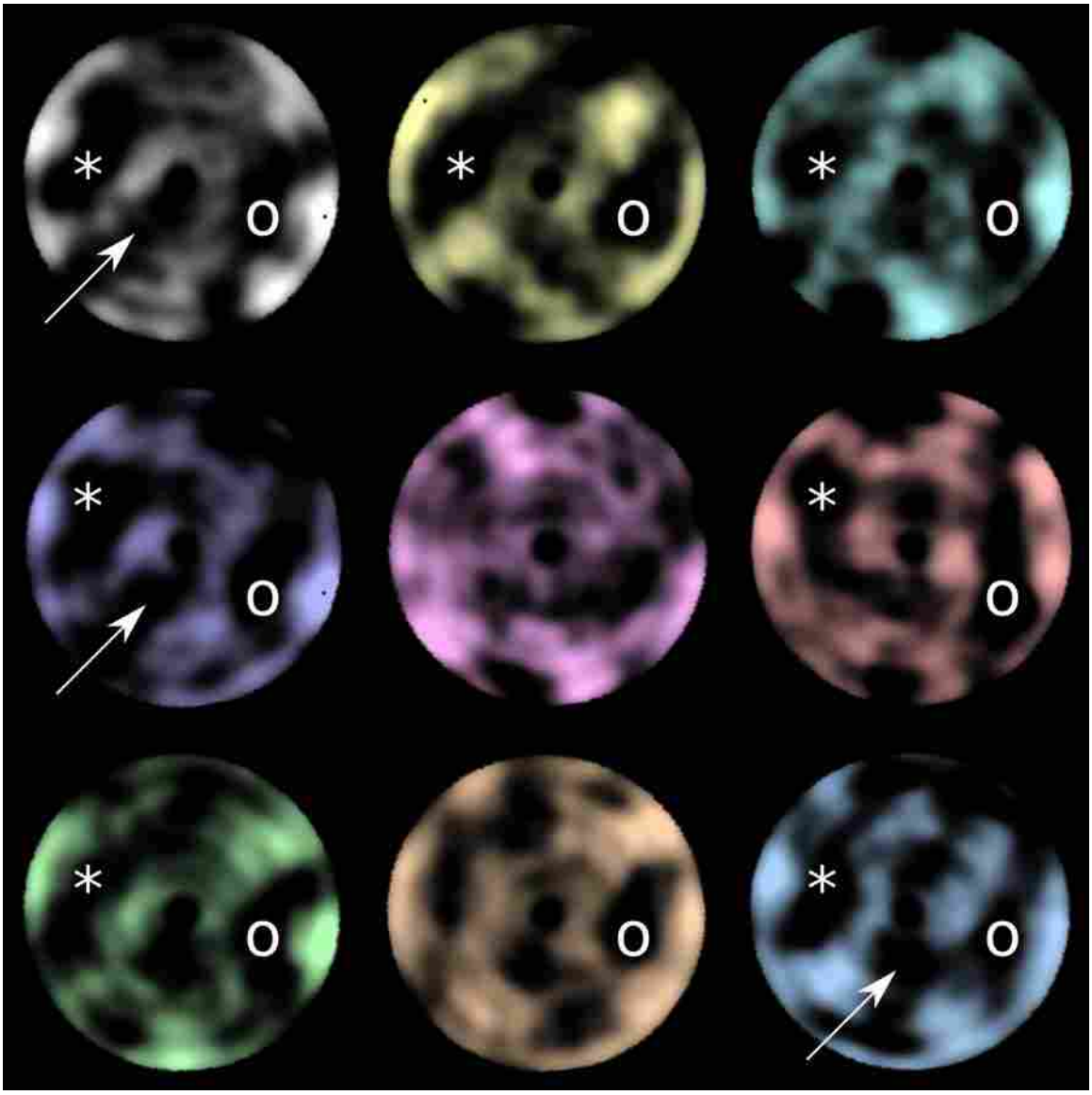
5.4.2 Symmetry-mismatch, random-model method

All SMRMM reconstructions were equally unsuccessful. Figure 5.8 shows the result of one of such reconstructions. We believe one reason for the failure of the method is that subtraction of the capsid affected negatively the RNA component of the images. Subtraction of the capsid effectively removed the capsid portion but could distort the signal of the RNA component. We are currently testing other subtracting algorithms that take into account the effects of imperfect CTF correction. One of this subtraction procedures produces reconstructions from the particles of each micrograph (one-micrograph reconstructions). These reconstructions are uncorrected for the CTF and they are used to subtract the capsid of the particles from the same micrograph. Because the CTF pattern is the same in one-micrograph reconstructions and corresponding particles no additional artifacts should be introduced. Preliminary



(A)

Figure 5.7 Asymmetric reconstruction of poliovirus RNA by the asymmetric RMM. (A) 10 sections of different reconstructions of poliovirus RNA. Stars and arrows mark common motifs among the different structures. (B) 10 central slices of the same reconstructions. Some common motifs, highlighted by white stars, circles and arrows, could be seen among the asymmetric reconstructions, suggesting partially consistent conformations of the RNA.



(B)

Figure 5.7 –Continued

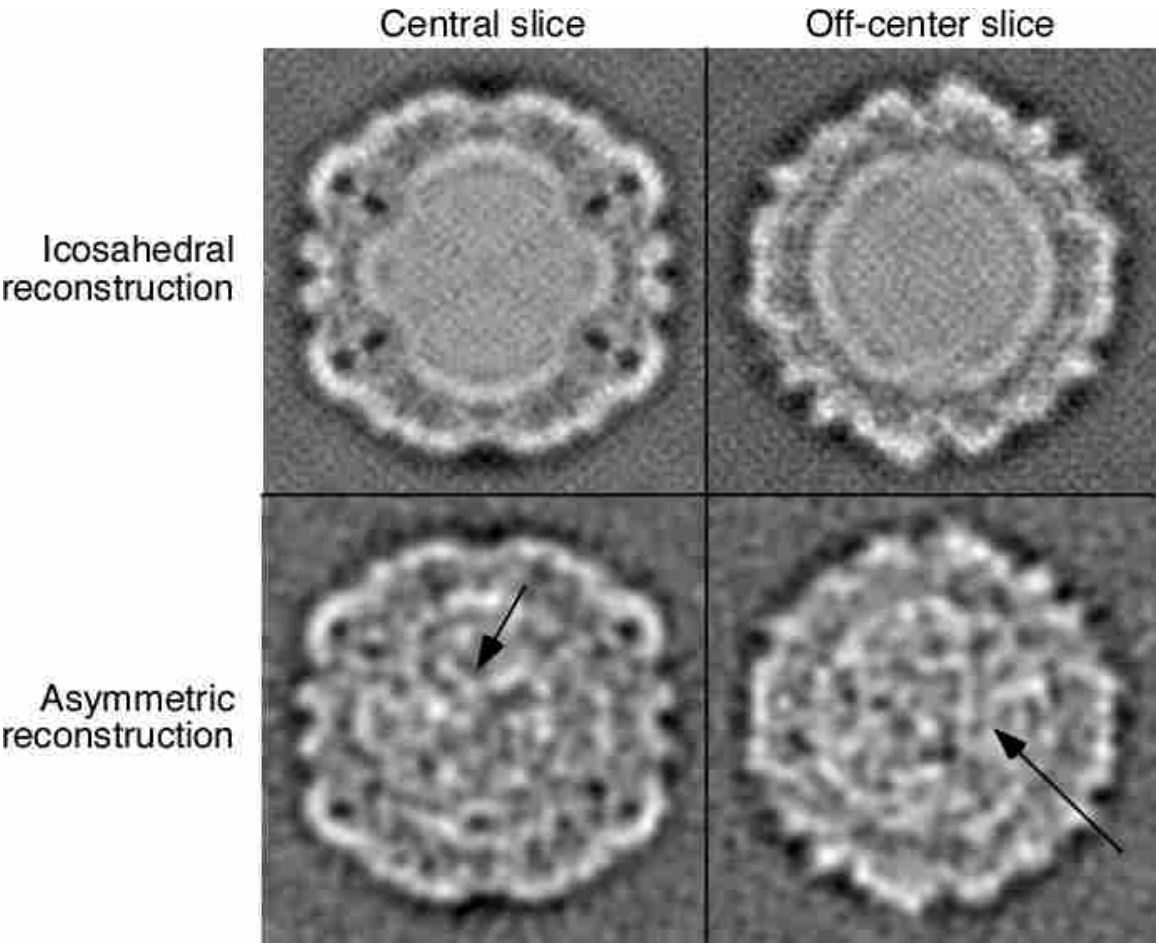


Figure 5.8 Asymmetric reconstruction of poliovirus RNA by the SMRMM. Differences between an icosahedral reconstruction compared to an asymmetric reconstruction. The arrows indicate putative RNA strands.

result indicate that this procedure could be advantageous. An alternative approach was to compute one-particle reconstructions (non-CTF-corrected), and use these to perform the actual subtraction. This approach seem to perform poorly because the noisy nature of the one-particle reconstructions introduced new artifacts.

Chapter 6

Conformational variance of satellite tobacco mosaic virus capsid around a stable RNA scaffold.

6.1 Background

Satellite tobacco mosaic virus (STMV) is a small, T=1 icosahedral particle approximately 17 nm in diameter [57, 66]. The capsid is composed of 60 identical copies of the major capsid protein (17 kDa) and encases a 1,059 nucleotide (+) ssRNA genome. The total molecular mass of STMV, including its genome, is 1.3×10^6 Da. Infection by STMV is only possible through co-infection with a tobamovirus [166], the result of which may heighten symptoms compared to infection by tobamovirus alone [167].

The molecular structure of STMV has been resolved by X-ray crystallography [57, 66]. The capsid protein arranges in a β-barrel jelly roll with extended amino-terminal arms that interact with the RNA. Capsid protein dimers were presummed to be the assembly unit of the capsid due to the large amount of hydrophobic and hydrogen-

bond interactions at the dyad axes. Conversely, capsid protein interactions are fewer at the 3- and 5-fold axes.

The RNA core of STMV is inherently stable in the absence of capsid proteins [76, 168]. The crystal structure of STMV revealed partially-ordered RNA throughout the capsid interior [57, 66]. The RNA arranges as 9-base double-helical segments inside the capsid at every dyad axis. These 30 helices account for about 50% of the whole genome. The capsid-RNA interface is lined with amino acids that interact with the nucleic acid backbone via hydrogen bonding and salt bridges. Extracted RNA cores adopt a globular shape with a diameter corresponding to the interior of the capsid and are able to unfold and fold reversibly [76]. Molecular dynamic simulations suggest that the viral genome plays an important role in capsid assembly [67] and stability [65]. These observations led to the assembly model where RNA must be completely folded before the binding of capsid proteins [65]. Therefore, STMV assembly seems to be driven by the RNA core rather than by the tendency of capsid proteins to adopt an icosahedral shape. In fact, capsids without RNA have rarely been observed [65].

Here, we present two solution structures of STMV, solved by cryoEM, which revealed different capsid arrangements than the X-ray crystal structure of the virion [57, 66]. Rigid-body fitting of the crystal coordinates into the cryoEM density maps accounted for the conformational differences between the crystal and cryoEM structures. Yet, the reconstructions contained conserved densities that matched well to the icosahedrally ordered portion of RNA found in the crystal structure. The observed capsid variation supports the notion of flexibility in capsid protein arrangement around a stable RNA scaffold.

6.2 Material and methods

6.2.1 CryoEM and image processing

Purified STMV was either kept in pure water or diluted in Buffer A (10 mM Tris-Cl, 50 mM NaCl, 0.01 mM $CaCl_2$, pH 7.4; Table 6.1). For each sample, 3.5 μL was applied to a holey carbon copper grid with a thin layer of carbon applied over C-flat holey grids (Protochips, Raleigh, North Carolina, USA). The grids were vitrified in liquid ethane using a Vitrobot (FEI) vitrification device (5 second blot) at 4.4° C and 100% relative humidity. Specimens were transferred to a Gatan 626 cryoholder (Pleasanton, California, USA) cooled with liquid nitrogen. CryoEM images were recorded at low electron dose at underfocus levels ranging from 0.5–4.0 μm on Kodak SO-163 film in an FEI Tecnai F30 transmission electron microscope (Hillsboro, Oregon, USA).

Micrographs were digitized on a Nikon Super Coolscan 9000 ED scanner in 16-bit, positive mode. Particle images were extracted from electron micrographs using *X3DPREPROCESS* [133]. We determined and corrected for the contrast transfer function signal and decay parameters using *bshow* and *bctf* (Bsoft [131]). Origins and orientations of the extracted particles were determined using the model-based technique of *PFT2* [111], which was adapted to use phase and amplitude information in orientation selection [see 152, Supplementary Section 2.2]. A 15 Å electron density map of STMV was generated from the atomic coordinates to initiate model-based reconstruction. Final reconstructions were calculated using *EM3DR2* (http://people.chem.byu.edu/belnap/pft3dr). *blocres* (Bsoft [131]; Cardone, in preparation) was used to estimate quasi-local-resolution by comparing 60 voxel (2.13×10^5 Å^3) sub-volumes centered at each voxel in the capsid portion of the final STMV reconstruction [169, 170].

For calibration purposes, we used poliovirus (PV; 160S state [106]) as an inter-

nal standard. Poliovirus particles were mixed with STMV in Buffer A and vitrified together (Table 6.1). 294 poliovirus particles were extracted from the micrographs, which produced a 20 Å reconstruction. A previously calibrated map of 160S poliovirus was used to correct the nominal pixel size of the poliovirus reconstructions. The same correction factor was used to amend the sampling size of the STMV reconstruction. The sampling sizes of all other STMV reconstructions, from images taken at same buffering conditions and voltage levels, were adjusted by comparing the reconstructions to the STMV calibrated map.

Table 6.1 Conditions of data sets acquired to reconstruction STMV

Specimen	Buffer[a]	pH	Accelerating voltage (keV)	Nominal magnification	Number of particles[b]	Figure
STMV	water	7.4	200	59,000	620	6.1A, col. 1
STMV	water	7.4	200	59,000	1,051	6.1A, col. 2
STMV	water	7.4	200	59,000	25,434	6.1B, col. 3
STMV	Buffer A	7.4	200	59,000	47,425	6.1B, cols. 4-5
STMV, polio	Buffer A	7.4	200	59,000	498, STMV; 147, PV	6.1B, col. 6
STMV, polio	Buffer A	10	200	59,000	974, SMTV; 134, PV	6.8
STMV	water	<6.5	200	59,000	4,400	6.1C, col. 7
STMV	water	<6.5	300	59,000	4,130[c]	6.1C, col. 8
STMV	water	<6.5	300	79,000	2,115[c]	6.1C, col. 9

[a] Buffer A: 10 mM Tris-Cl, 50 mM NaCl, 0.1 mM $CaCl_2$
[b] Refers to number of defocal pairs, except when the opposite is noted
[c] Total number of particles, no defocal pairs

6.2.2 Structure modeling and analysis

Rigid body fitting of STMV coordinates (Protein Data Bank (PDB) accession code: 1A34 [57]) into electron density maps was accomplished using the Situs package [171]. This tool performs a full exhaustive search in 3D search space (3 translational and 3 rotational degrees of freedom) for the most favorable docking of the crystal coordinates in the cryoEM map. Crystal contacts and atomic clashing were determined with *UCSF Chimera* [145]. Flexible loops were modeled using *MODELLER* [172].

6.3 Icosahedral reconstruction

Despite using the STMV crystal structure as the reference model, the iterative refinement process of computing the 3D reconstruction resulted in a divergent, but fairly consistent, structure for all data sets (Table 6.1; Fig. 6.1). The atomic coordinates of the STMV capsid were fitted into the final 14 Å STMV cryoEM reconstruction. All main-chain atoms of the β-barrel jelly roll were fitted into the density map rendered at 1.0 σ density level. The most prominent feature to protrude from the density map was the loop formed between strands B and C (BC) and a portion of strand A. Differences between the two maps were prominent at the 5-, 3-, and 2-fold axes of the icosahedral shell (Fig. 6.2).

6.3.1 Conformational changes of STMV capsid

Although the radius of the STMV reconstruction was identical to the radius of the STMV crystal structure (Fig. 6.3), the capsid structure is clearly dissimilar between the two structures (Fig. 6.2C, and bottom row). At the 5-fold axes, the cavity formed by the convergence of five protein subunits is present in either map, but the star shape formed at each vertex is rotated counter-clockwise in the cryoEM reconstruction.

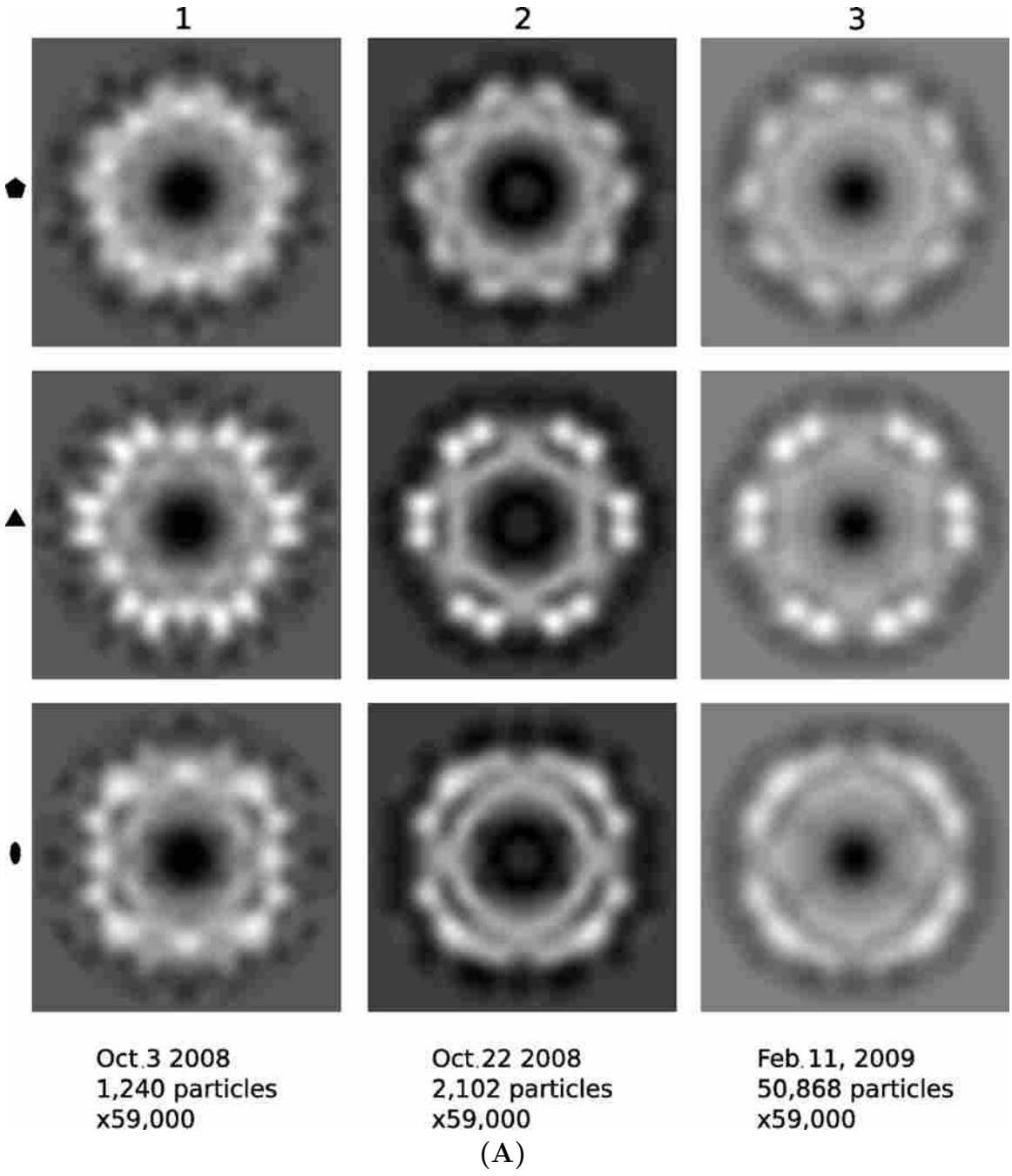
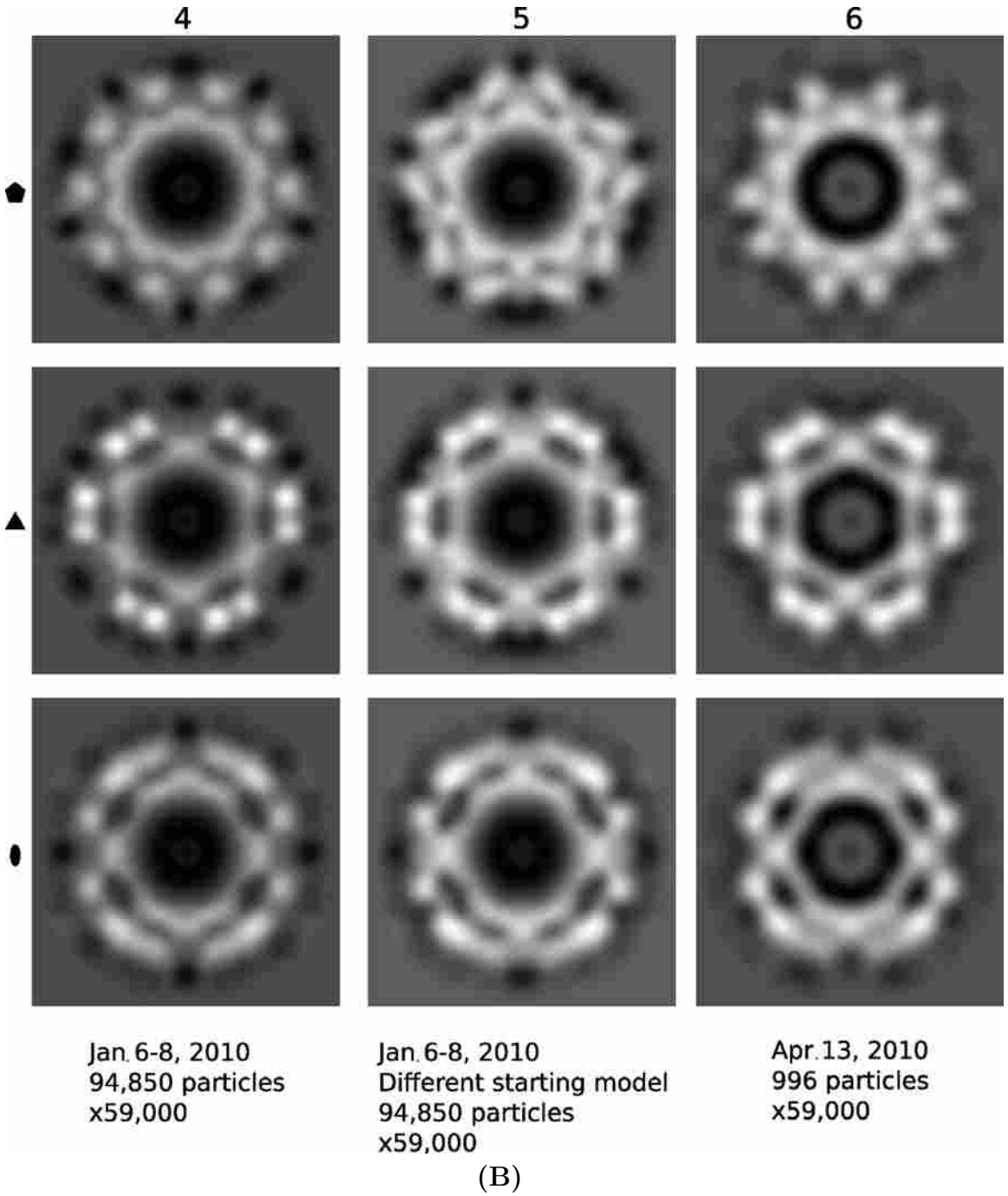
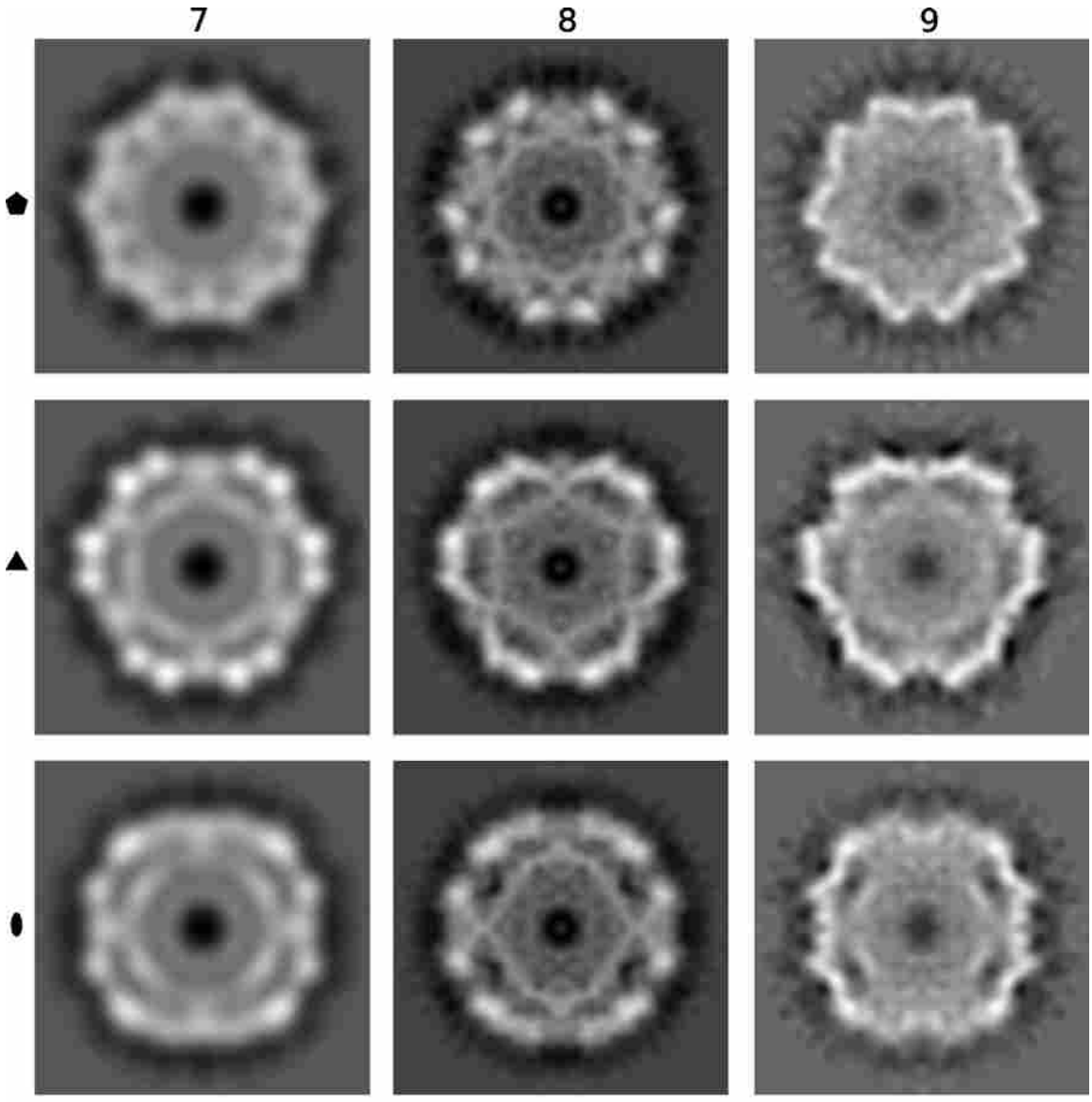


Figure 6.1 Icosahedral reconstructions of SMTV capsid from different data set (Table 6.1). Reconstructions are displayed in the same order as Table 6.1, except for STMV reconstruction at pH 10 which is depicted in Figure 6.8. Three central sections from different views are displayed. Top panel, capsid is oriented with a 5-fold axis toward the reader. The 3-, and 2-fold axes are oriented toward the reader in the middle and bottom panels, respectively.



(B)

Figure 6.1 –Continued



(C)

Figure 6.1 –Continued

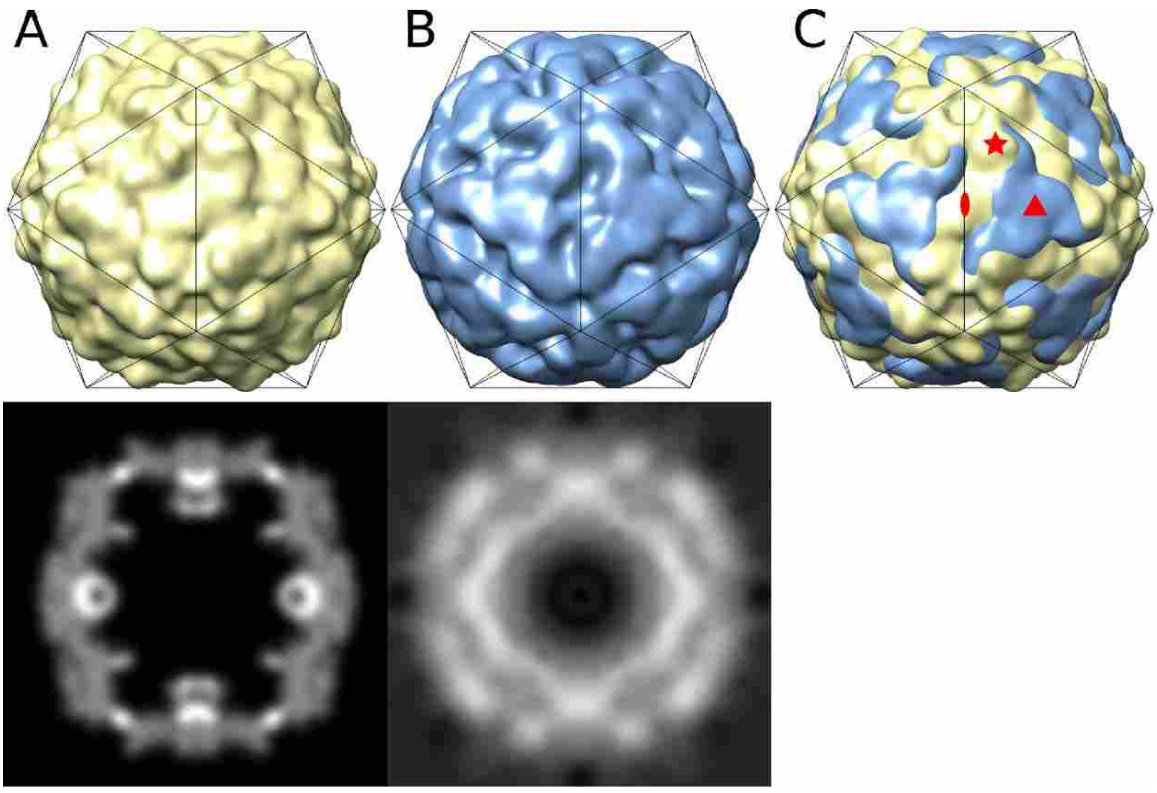


Figure 6.2 Differences between STMV crystal and solution structure. Panel A, electron density map computed from the crystal coordinates (PDB 1A34 [57]) used as reference model. Panel B, 14-Å cryoEM reconstruction using 79,000 image particles. Panel C, superimposition of the crystal and cryoEM structures. Major differences are around the 2-fold (oval), 3-fold axes (triangle) and the direction of the star like shape (star). Top row, surface rendering at 1.0 σ contour level. Bottom row, central slices.

This rotation results in a redistribution of density throughout the capsid without affecting the overall radius. The 3-fold axis from the crystal coordinates consists of shallow grooves surrounded by bumps formed around the 5-fold axis. In the cryoEM reconstruction, the 3-fold axes are filled with densities that converge at each axis such that the grooves are lost. Instead, the 3-fold axis of the cryoEM reconstruction is characterized by smoothness, compared to the varying contours seen in the crystal structure map at the same location. Quite the opposite was observed at the 2-fold axis, where high levels of densities formed in the crystal structure map are in contrast to the relatively deep grooves formed in the cryoEM reconstruction.

To characterize the apparent structural differences between STMV from X-ray crystallography and cryoEM, we fitted the atomic coordinates of the STMV major fold into the cryoEM reconstruction using rigid-body fitting techniques. We achieved a favorable fitting by shifting and rotating the capsid protein about the 5-fold axis (Fig. 6.4). The fitting matched very well strong densities that corresponded to the core of the β-barrel jelly roll and resulted in only 3.5% (70 of 1970) of the atoms protruded out of the density contours at 1.0 σ density level. The fitting did not produce any steric clashing among protein subunits. The new fitting accounted for the rotated star formation about the 5-fold axis, increased densities at the 3-fold axis, and showed absence of densities at the 2-fold axis.

Rigid-body fitting of the capsid protein resulted in the formation of deep grooves at the 2-fold axes, as seen in the cryoEM reconstruction (Fig. 6.5). These conspicuous features indicate that capsid protein interactions are minimal at the dyad axes. However, the axial point is marked by a noticeable bulge that stems out of the groove. Our fitting did not reveal any noticeable interactions at the 2-fold axes, including the bulge. However, the large, flexible loop formed between β-strands D and E (DE) could presumably be modeled into the bulge. We used *MODELLER* to determine the

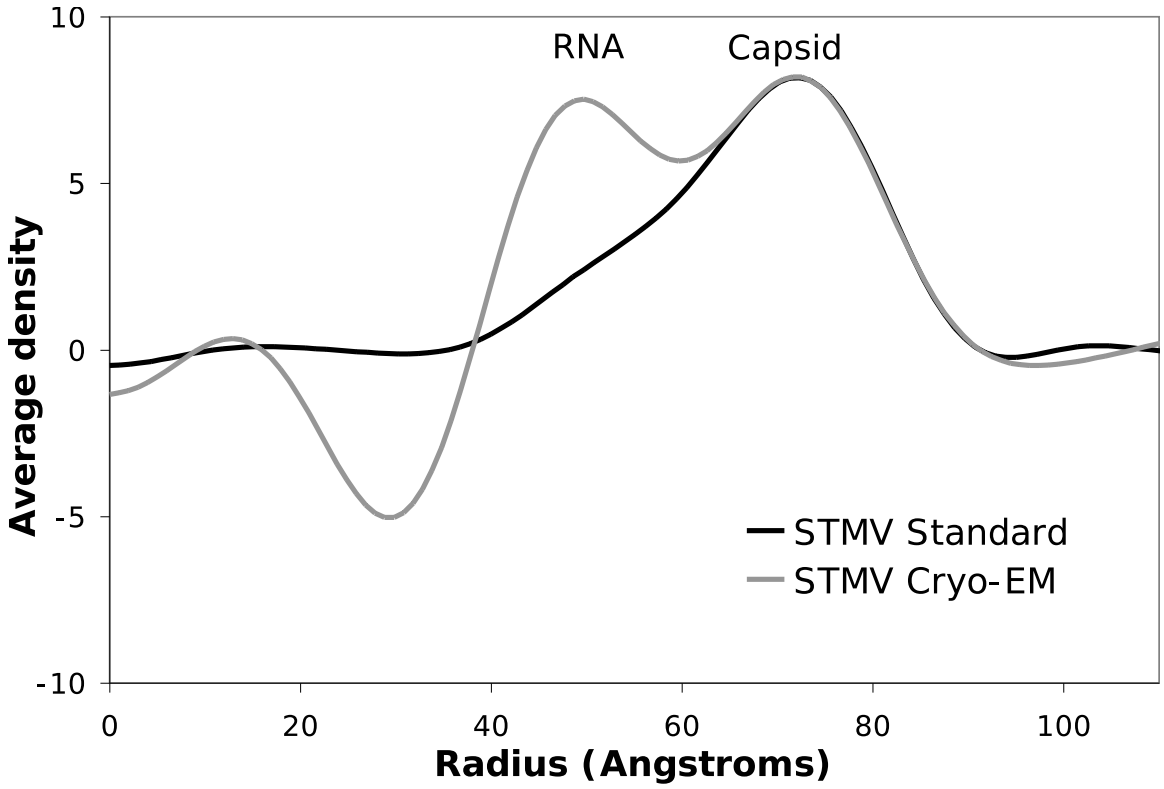


Figure 6.3 Spherically-averaged density profiles for the cryoEM structure and the electron densities derived from the crystal structure (PDB 1A34 [57]). After cryoEM models were calibrated using PV as internal standard, the radial profile was compared to the crystal structure. The densities corresponding to the capsid (first peak from the right) fit very well. The second peak from the right corresponds to the extra densities of RNA that are not visible in the crystal structure.

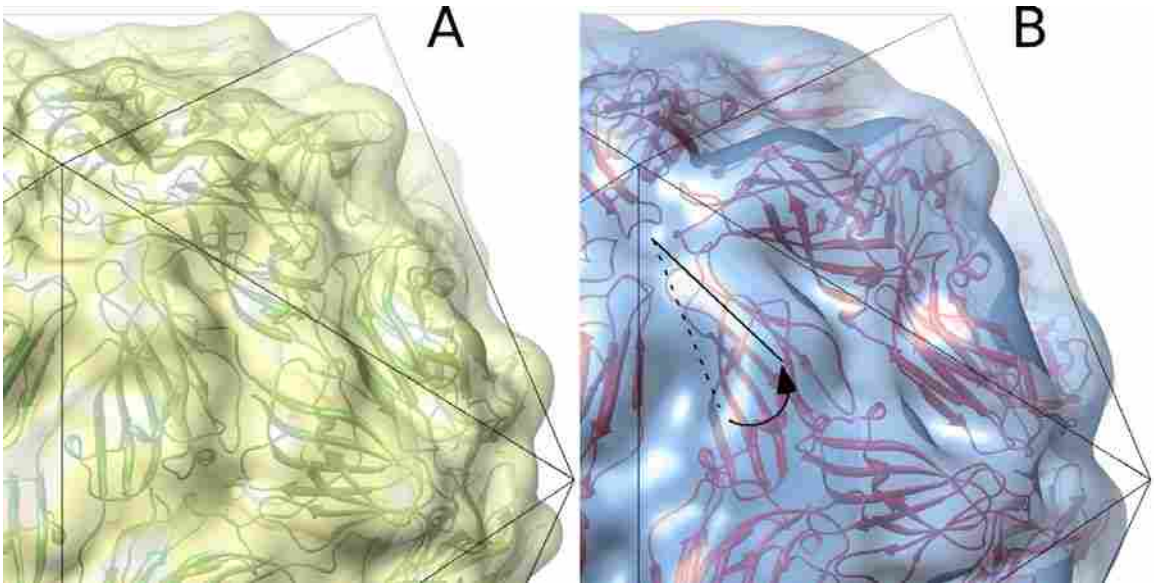


Figure 6.4 STMV crystal coordinates fitted into a cryoEM reconstruction. Panel A, crystal coordinates (PDB 1A34 [57]) depicted in green and derived electron density map, in yellow. Panel B, crystal structure, in orange, fitted into the cryoEM reconstruction, depicted in blue.

loop conformations with the lowest discrete optimized protein energy value. This conformation, and others with similar energy values, coincided with the 2-fold axis bulge (Fig. 6.5C). Potential interactions between two DE loops were revealed after icosahedral symmetry was applied to the remodeled loop. The loop consists of residues that could form salt bridges (Lys101, Asp104, Glu107, Glu108) and other hydrogen bonds (Thr102, Ser105, Thr106) with another DE loop. We refined the modeling of the DE loop as to maximize interactions at the 2-fold axis. The limited number of inter-subunit interactions at the 2-fold axis is in contrast with those found in the crystal structure. In the crystal structure, at least 18 unique hydrogen bonds stabilize the dimer, not including additional water-mediated hydrogen bonds.

The new fitting also resulted in the merging of three capsid protein subunits at the 3-fold axes (Fig. 6.6). The convergence of the proteins suggests that more intermolecular interactions occur at the 3-fold axes than previously thought. Indeed,

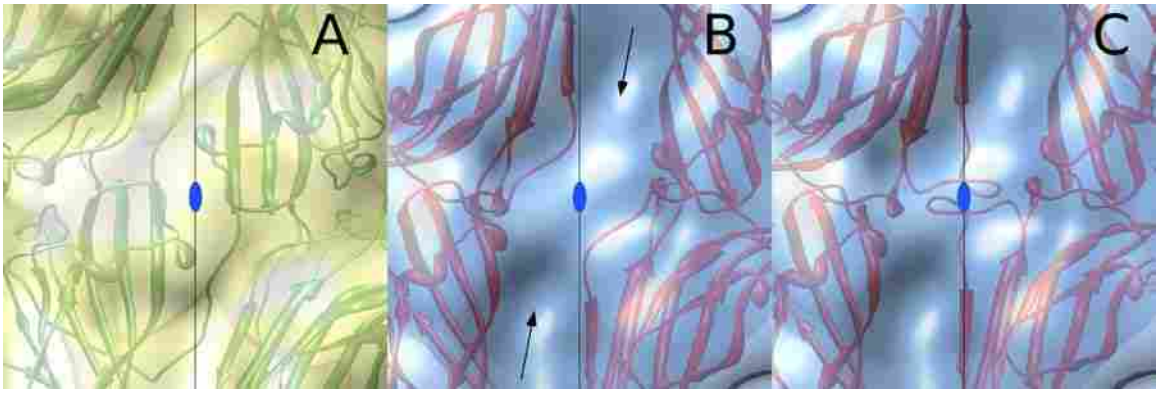


Figure 6.5 Structural differences of STMV crystal and cryoEM structures at the 2-fold axis. Panel A, original STMV crystal structure (PDB 1A34 [57]. Panel B, rigid-body fitting of the crystal structure into the cryoEM reconstruction. The fitting accounts for the lack of density at both sides of the 2-fold axis, indicated by the arrows. Panel C, loop DE was remodeled to explain a bulge that coincide with the 2-fold axis (oval). Views are from outside the particle.

a closer investigation of plausible interactions at the 3-fold axes revealed increased hydrogen bond interactions and hydrophobic interactions (Fig. 6.7). The interactions at the 3-fold axis largely occur among flexible loops, which may be further rearranged or remodeled to yield additional favorable interactions. The collective differences between the two structures at the 5-, 3-, and 2-fold axes suggest a conformational switch from a 2-fold dominated arrangement in the crystal structure to a 3-fold dominated arrangement in the cryoEM reconstruction.

Although the new fitting relied on rotational and translational adjustments about the 5-fold axis, the subunit interactions at the axis appear to be unchanged. The shallow cavity formed at the 5-fold axis is present in both structures, indicating favorable interactions formed among the five subunits at that location. In the crystal structure, the convergence of the five subunits at the icosahedral vertices is marked by a network of asparagine side chains (Asn115 and Asn117), which forms a ten-ligand cage about the axis. The cage was shown to surround a single sulfate ion [57], which

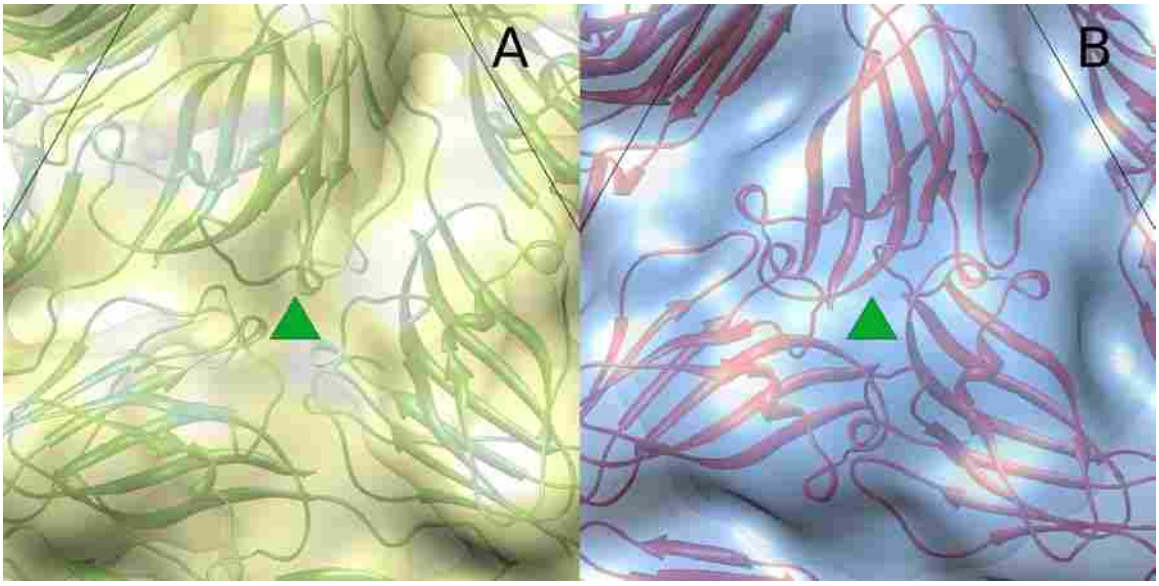


Figure 6.6 STMV cryoEM reconstruction suggests an increased interaction at the 3-fold axis. Panel A, STMV crystal structure (PDB 1A34 [57]). Panel B, rigid-body fitting of the crystal structure into the cryoEM reconstruction. The fitting accounts for the increase of density at the 3-fold axis.

presumably facilitates hydrogen-bonding with the asparagine cage. We note that the buffers used in this study lacked sulfates and phosphates, therefore suggesting that the presence of large anions locks the capsid in a conformation consistent with the crystal structure. Nevertheless, the caged anion is not requisite for the overall structural integrity, as previously suggested by Larson *et al.* [57] and supported by our fitting. In the new conformation, hydrogen bonding interactions among the ten asparagines (∼3 Å apart) may be sufficient to stable the interactions among the five subunits.

The relatively modest resolution of our reconstruction (14 Å) was prohibitory in fitting the coordinates of the amino-terminal "arm" into the density map. Nevertheless, the rigid-body fitting of the major fold maintained the general direction of the "arm" into the capsid interior. The crystal structure of STMV revealed that several basic residues of the "arm" interact with the RNA core in a non-sequence specific

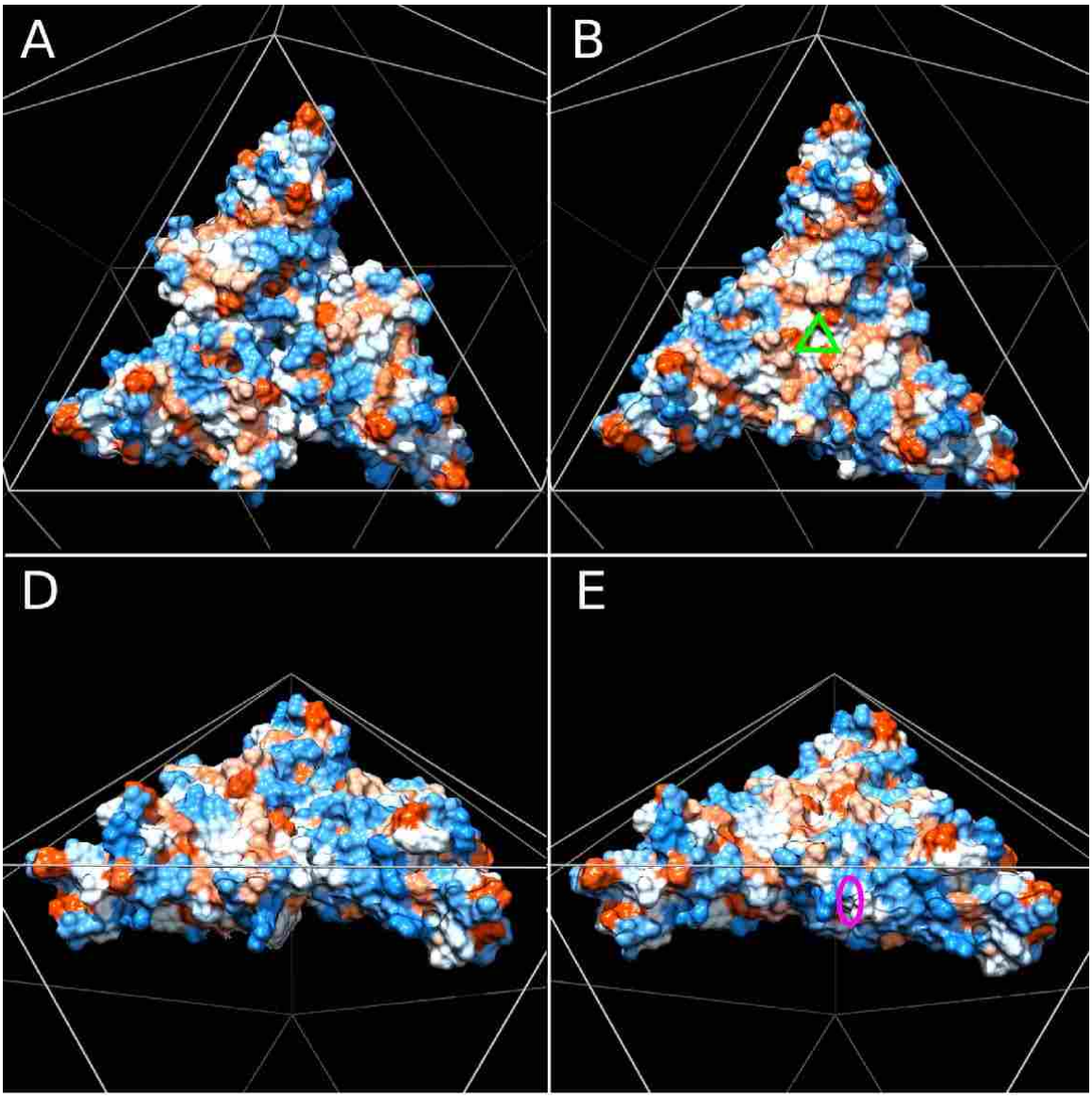


Figure 6.7 STMV hydrophobic interactions at the 3-fold axis. Panel A, molecular surface representation of three STMV major capsid proteins in the crystal structure (PDB 1A34 [57]). Residues are colored according to their hydrophobicity character [173]. The scale ranges from red (apolar residues) to blue (polar residues), with intermediate values in white. Panel B, fitting of the crystal structure into the cryoEM map produces hydrophobic residues (in red) to cluster at the 3-fold axis (open triangle), leaving hydrophilic residues (in blue) at the 2-fold axis (open oval). Panels C and the D are the same as A and B from a different view. All views are from the outside of the particle.

manner. Our fitting of the coordinates upholds the tethering nature of the "arm", assuming that the "arm" is flexible.

Whether our cryoEM structure represents a more physiological state of the virus is unknown. However, we do note that the STMV crystals were formed in ∼11% ammonium sulfate (w/v) [67]. We investigated whether our cryoEM structure was a result of a favorable conformation due to components in our buffering solution. Our attempts to record cryoEM images of STMV in a solution containing similar amounts of ammonium sulfate were unsuccessful due to refractory conditions in forming thin, vitreous ice. However, cryoEM of STMV in pure water or Buffer A at physiological pH values yielded similar 3D STMV reconstructions (Table 6.1, Fig. 6.1). Therefore, the crystal structure of the virion may represent a natural conformation when STMV is in complex with large anions. On the other hand, our cryoEM reconstruction may represent the STMV structure at a more physiological anion concentration.

We studied whether the conformation of the STMV crystal structure is affected by crystal contacts between STMV subunits. Crystal contacts were formed entirely by interactions between flexible loops that extend out of the 3-fold face, namely loops AB, BC, and FG. Therefore, interactions between these loops from adjacent virions may have dominated the quaternary structure shown in the crystal structure, which is supported by the fact that the atomic coordinates of the BC loop completely protrude from the cryoEM reconstruction.

6.3.2 Capsid expansion at high pH

STMV capsid experiences an expansion when it is exposed to high pH [174]. We investigated this capsid expansion by cryoEM by acquiring images of STMV and poliovirus in Buffer A at pH 10 (Table 6.1). Poliovirus was used as internal standard to calibrate the reconstruction of STMV under this new pH condition. For unknown

reasons, poliovirus assumed a 80S conformation [106], therefore, we calculated the correct pixel size of the images by comparing poliovirus reconstruction with a previously calibrated 80S poliovirus map. After correcting the pixel size of STMV images accordingly, we observed a radial expansion of STMV capsid of ∼10 Å, double than previously reported [174], without a significantly change in the overall aspect of the capsid structure (Fig. 6.8). Nevertheless, the RNA expands substantially less, ∼3.5 Å radially. This uncoordinated expansion of the RNA and the capsid suggests that the RNA structure is inherently more stable than the capsid. Because the RNA is a single molecule has less degrees of freedom than the capsid, formed by the cooperative arrangement of the 60 smaller-independent proteins. This fact could explained the increase flexibility associated to the capsid shell. The discrepancy between the capsid expansion we reported (∼10 Å) and previously observed (∼5 Å [174]) is likely due to methodological difference, the later being an estimate using AFM.

6.3.3 CryoEM reconstructions represent a single conformation

To rule-out the possibility of model bias, we computed 10 independent reconstructions using the icosahedral RMM [129]. Two out of the ten maps showed identical arrangement of the capsid as the described cryoEM structure. Only these two maps presented a RNA structure in agreement with the crystal structure. The rest of the structures either had large voids on the capsid, presented spiky structures, or contained large protrusions on the 5-fold axes. Different classification schemes were tested to group putative image subpopulation, all of which resulted in identical 3D reconstructions. Further, separate data sets were recorded using varying acceleration voltages and defocus ranges, but 3D reconstructions of these data sets yielded similar

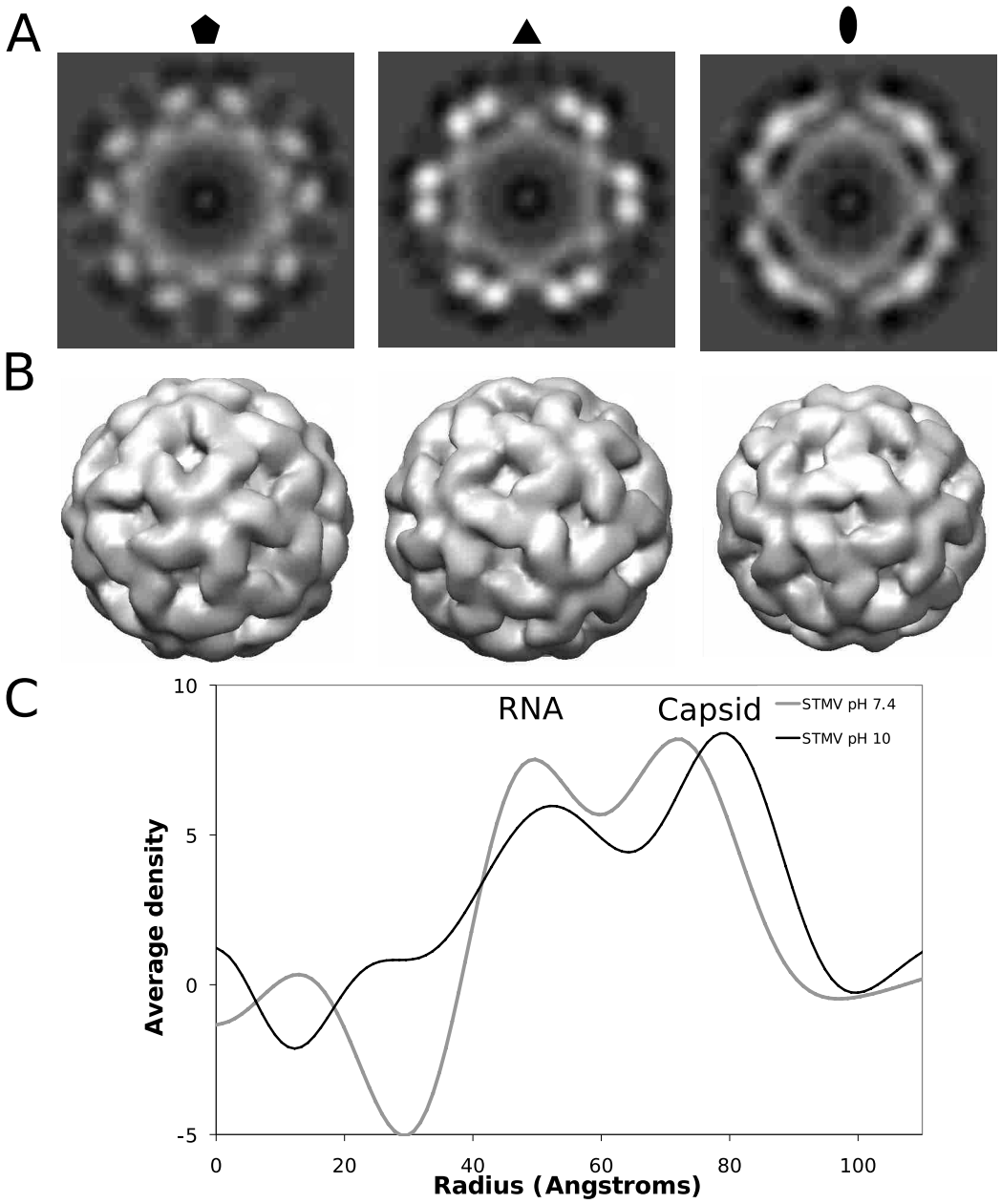


Figure 6.8 STMV reconstruction at pH 10. The reconstruction was computed from 1,948 particle images ($\times 59,000$ magnification) acquired on July 30, 2010 (Table 6.1). Panel A, shows three different central sections from the capsid oriented with a 5-, 3- and 2-fold axis toward the reader, respectively. Panel C, same as panel A but surface renderings. Panel C, spherically-averaged density profiles of cryoEM structures at pH 7.4 and 10. The capsid expanded by $\sim$10 Å, conserving the overall structure (compared to the capsid structure at physiological pH values; Fig. 6.1). The RNA appears to be more stable because of the comparatively reduced expansion.

structures as our initial results (Table 6.1, Fig 6.1).

6.3.4 The RNA core is stable

Although the capsid arrangement was noticeably different between density maps from crystallography and cryoEM, the densities corresponding to RNA duplexes were surprisingly unchanged. We fitted the atomic coordinates of the RNA duplexes into the STMV cryoEM reconstruction without need of neither rotational nor translational adjustments (Fig. 6.9). Only the five central nucleotides of the nine-bases RNA helixes could be fitted in the cryoEM reconstruction. These nucleotides make the highest number of contacts with the protein capsid and have the least thermal motion (lowest B value), as determined in the crystal structure. We did not observe the inner portion of the RNA helices, which appear disordered in the cryoEM reconstruction even though full occupancy of the helix has been suggested at each dyad in the icosahedrically averaged crystal structure.

These results indicate that the RNA core is remarkably stable and suggests the core has a scaffolding function upon which the capsid proteins are assembled. Thus, the structural role of STMV RNA may be as important as the protein information encoded by the genome. The stability of the RNA structure is consistent with previous reports by atomic force microscopy which indicated that extracted RNA adopts a globular state with a diameter consistent with the inner radius of the capsid shell [76]. Kuznetsov *et al.* reported that rigorous denaturing conditions were needed to disrupt the secondary structure of the RNA [76]. A molecular dynamic simulation study has also revealed the stability of the RNA core alone [65].

The ability to visualize specimens in a vitreous, native-like state is a paramount reason for the growing popularity of cryoEM. Here, we demonstrated that even modest-resolution structures derived from cryoEM can be used in complement with

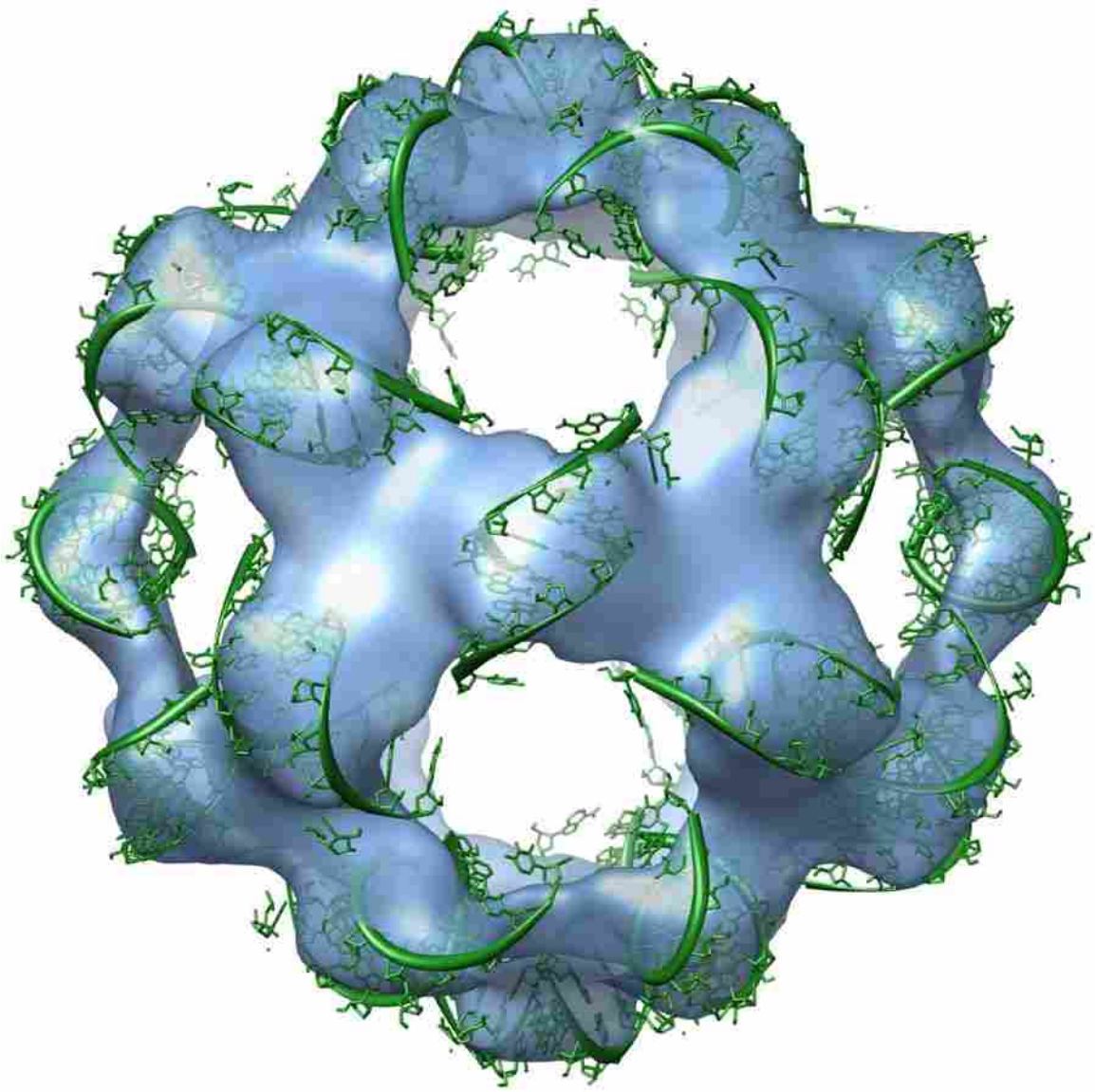


Figure 6.9 Superimposed RNA coordinates into the cryoEM reconstruction. Five central nucleotides out of the nine-bases helices were fitted in high contour level densities in the cryoEM reconstruction.

crystallography data. Our cryoEM reconstruction of STMV, resolved to 14 Å resolution, revealed a significantly different quaternary structure in capsid organization compared to the structure solved by X-ray crystallography. However, the major fold of the capsid protein appears unchanged, as demonstrated by rigid-body fitting of the atomic coordinates into the cryoEM reconstruction. Removal of all buffer components in the STMV solution did not affect the cryoEM reconstruction. Therefore, deviations between our cryoEM structure and the crystal structure are unlikely triggered by the composition of our STMV solution. Yet, capsid variation and dynamics appear to be independent of the RNA structure. Indeed, the ability of STMV capsids to assume various conformations suggests that interactions among the sixty-related subunits are inherently flexible, whereas the RNA core appears to be more rigid. Our results support evidence by others [67, 76] that at least part of the RNA core consistently folds into a stable structure, which presumably drives the assembly process of the virion. The scaffolding nature of STMV RNA is in line with the model that the RNA core is at least partially folded before capsid proteins tile onto the core. Whether the RNA fully folds into a globular state before assembly or whether assembly occurs in concert with RNA folding is unknown. The consistencies of the RNA structures reported here and by others [57, 66] suggest that the full structure of the STMV genome could be resolved by means of cryoEM, given sufficient particle images.

STMV is a challenging virus to be reconstructed by cryoEM and 3D reconstruction. Because of the small size of the virion, ice thickness has to finely tuned. Furthermore, the smoothness and roundness of the capsid shell could result in incorrect assignment of orientations. Validation efforts are ongoing in the lab to reconcile some inconsistent results (see differences in Fig. 6.1, top panel). We also noted that the resolution of the reconstructions didn't grow as the expected rate in relationship of the number of particles. As part of these efforts, we are planning a tilting experi-

ment to verify the accuracy of the assigned orientations [149]. Additionally, we are considering the option of conjugating Fabs to the viral capsid to increase the SNR and help in refinement of orientations,[1] and we are currently testing other cryoEM reconstruction packages.

6.4 Asymmetric reconstruction

A number of asymmetric reconstructions were computed using different data sets and reconstruction parameters (Table 6.2). After alignment of the reconstructions, the structures showed low homology. Furthermore, reconstructions were assessed quantitatively by PCA resulting in no obvious clustering (Section 3.3.7).

Direct imaging of STMV RNA *ex virio* has been delayed because of problems related to its extraction. Some of the challenges include, avoiding degradation of the RNA and fine tuning of the cation concentration to avoid aggregation and disruption of the structure. Fluorescence labeling experiments of STMV RNA are ongoing to test the accessibility to electron-dense probes and assess the feasibility of labeling the RNA (see Section 7.1).

[1] This idea was explored in Section 3.3.4.

Table 6.2 Asymmetric reconstructions of STMV[a]

Date	Capsid-subtracted images	Method	Number of particles[b]	Resolution cut-offs (Å)[c]	Number of repetitions
April/2009	Yes	ARMM	25,000	25, 90	10[d]
Feb/2010	Yes	ARMM	47,500	25, 90	10
Feb/2010	ARMM	47,500	25, 90	10	
Feb/2010	Yes	ARMM	47,500	18, 90	10
Feb/2010	Yes	SMRMM	47,500	25, 90	10
Feb/2010	SMRMM	47,500	25, 90	10	
May/2010	Yes	ARMM	47,500	25, 90	10[e]
Jun/2010	Yes[f]	ARMM	47,500	25, 90	10
Nov/2010	Yes	ARMM	47,500	25, 90	51

[a] All data sets were CTF corrected for the phase component only: inversion of the phases caused by aberrations in the imaging system [131, 136]. The nominal magnification was ×59,000, and 200 keV the accelerating voltage.
[b] Refers to number of defocal pairs
[c] Low and high resolution cut-offs for orientation and origin determination. 3D reconstructions were typically computed at a maximum resolution equal to the low resolution - 2 Å.
[d] Particles were subtracted, but jobs were aborted.
[e] For each iteration only the particles that were less than 1° apart from the icosahedral orientations were included in the reconstruction. In this way the icosahedral orientations acted as a filter. The asymmetric orientations were moved to the icosahedral asymmetric subunit before they were compared.
[f] Particle origins were further refined using a reconstruction from original particles (capsid + RNA).

6.4.1 Asymmetric random model method

All efforts to reconstruct STMV RNA using the asymmetric RMM failed. We limited the orientation refinement to a radii that encompassed only the RNA component. We tested subtracted ("capsid-less"), as well as, not subtracted particles (Section 4.2.2, Table 6.2). For a set of jobs we used the icosahedral orientations as a way to filter correct orientations. After each iteration, the asymmetric orientations were transformed to equivalent orientations in the icosahedral asymmetric unit. Particles with orientations less than 1° apart from the icosahedral orientations were included in the reconstruction. These orientations were not changed in subsequent iterations. For another set of reconstructions, refinement of the origins was performed independently with particles that contained the capsid component. Lastly, a massive amount of reconstructions were computed with the hope that the we could detect successful reconstructions in case success rate was very low. To facilitate analysis of such amount of reconstructions, the last reconstruction was computed using complete particles (capsid + RNA). If the reconstructions showed an icosahedral capsid then we were confident about the success of the asymmetric reconstruction. None of the models exhibited an icosahedrally arrangement of the capsid shell.

6.4.2 Symmetry-mismatch random-model method

As with the asymmetric RMM, we failed to reconstruct STMV RNA using the SM-RMM. These results could be explained by one or more of the following reasons: (i) RNA doesn't possess any order; (ii) the RNA core adopts a limited amount of alternative conformations; (iii) the RNA is ordered, but does not assume a fixed orientation relative to the capsid; (iv) deficiencies on the methodology (capsid subtraction, or orientation refinements); (v) inaccuracies on the orientation determination in

the icosahedral reconstructions.

We tested two different subtraction approaches that appeared to perform equally well for removing the capsid component of the images. In the first approach we computed a different density scale and shift factor per micrograph (Section 4.2.2). In the second approach, because the images were previously normalized (by *X3DPREPROCESS* during extraction), we applied the same density scale and shift factor to all the particles (independently of micrograph). We are currently experimenting all the subtraction ideas we mentioned for poliovirus (Section 5.4.2). Additionally we are planning on subtracting the icosahedral portion of the RNA.

Chapter 7

Conclusion

In this work I have presented the efforts to characterize the 3D structure of the genome of small single-stranded icosahedral viruses by cryoEM and 3D reconstruction. In Chapter 1, I introduced the current knowledge about the structural organization of viral genomes and the biological implication of these structures in transcription, replication, capsid assembly, capsid stability and evasion from the immune system. I indicated the limitations of the current structural methods, such as, X-ray crystallography, NMR, and AFM, emphasizing the potential of cryoEM to allow the structural determination of viral genomes.

In Chapters 2 and 3, I presented the asymmetric random-model method. We demonstrated how suitable starting models can be derived directly from un-stained, un-averaged cryoEM images of asymmetric or symmetric structures. In addition, we explored important factors for the success of the asymmetric random-model method: variations in angular step-size, presence of prominent structural features and handedness in the structure. We concluded that any single validation criterion can be misleading, and the most reliable way to validate cryoEM structures is to use multiple criteria.

Chapter 4 showed that asymmetric structures within icosahedral shells are resolved by the symmetry-mismatch, random-model method. We also observed that the success rate and accuracy of the orientations improved with increased number of particle images.

We applied the asymmetric random-model and symmetry-mismatch, random-model methods to STMV and poliovirus (Chapters 5 and 6). No consistent structure of the genome was found. However, for poliovirus, we observed common densities among different asymmetric reconstruction, suggesting perhaps, that only certain portions of the RNA are ordered and assume specific positions inside the capsid. For STMV, we found an alternative capsid conformation to the previously described crystal structure. Remarkably, the icosahedral portion of the RNA seemed to adopt the same conformation in both the crystal and the cryoEM structures. This suggests that the RNA of STMV is very stable and plays a role as scaffold, allowing the capsid shell to adopt a different conformation.

7.1 Future directions

As mentioned in Sections 5.4.2 and 6.4.2, we are pursuing different capsid subtracting approaches to mask the capsid component—leaving the genome component unaffected. These new approaches take into account possible negative effects resulting from inaccurate determination/correction of the CTF. The purpose of the capsid subtraction is to increase the likelihood to orient only the asymmetric densities corresponding to the RNA.

We are studying in more detail the structural heterogeneity of poliovirus data set. We would like to find all the possible alternative conformation for the RNA using 2D reference-free classification procedures.

Because it has been showed that the genome of STMV and poliovirus could adopt the same 3D structure in the absence of coat proteins [73, 76], I suggest the structural determination of *ex virio* genomes. After digesting the capsid with proteinases, the RNA could be gently extracted (phenol/chloroform/isoamyl alcohol) in the presence of buffers containing monovalent and divalent cations in order to maintain the secondary and tertiary structure.

One of the most important challenges of cryoEM-based reconstructions is the validation of the final 3D model. The lack of rigorous proof of correctness is a serious hindrance for newly characterized structure resolved by cryoEM. Because clusters of heavy atoms, can be conjugated to designated chemical groups on nucleotides, I propose to resolve the structure of single-stranded viral genomes by labeling specific sequences of the viral genome with a cluster of electrodense atoms. These labeled genomes might greatly facilitate i) the orientation determination of the 2D images via cryoEM reconstruction, ii) the localization of an initial position to start tracing the genome, and iii) the validation of the resulting 3D cryoEM reconstruction.

Two commercially available gold clusters could be assayed, undecagold, an 11-gold-atom cluster with a diameter of 8 Å, and nanogold, a 67-gold-atom cluster with a diameter of 14 Å. Unlike nanogold, undecagold is only visible in 3D density maps calculated from cryoEM micrographs. Several studies have proved the usefulness of these probes [13, 175–177].

Three different strategies could be taken to localize specific portions of the genome. These alternative strategies are presented in order of complexity.

1. Mapping of the 3' end (Fig. 7.1A). The genomic RNA could be coupled with undecagold or nanogold on their 3' terminal ends based upon an oxidation of the ribose *cis* diols (2' and 3' position) to aldehydes with periodate. The aldehydes will react with nano or undecagold to form the Schiff's base, that will be reduced

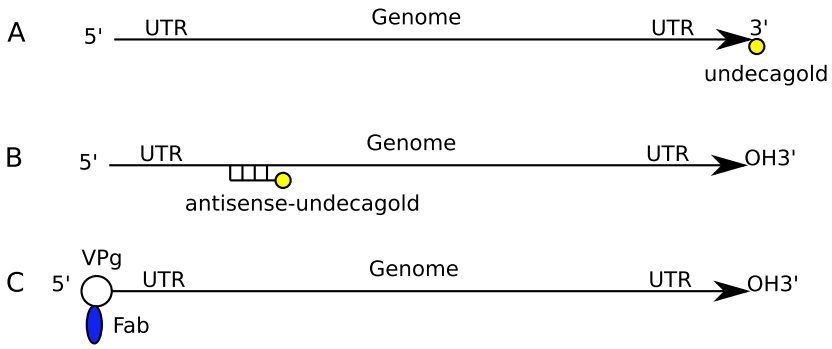


Figure 7.1 Strategies to localize specific portions of extracted genomes. A. Mapping of the 3' end. B. Hybridization of accessible sequences using antisense DNA/RNA gold probes. C. Inmunolocalization of VPg in picornaviruses.

to form a stable bond [178, 179]. Alternative protocols have been described to functionalize RNA with undecagold or nanogold [180–186]

2. Hybridization of accessible sequences of the viral genome by antisense DNA/RNA-gold probes (Fig. 7.1B). This strategy involves i) the identification of viral accessible sequences using a hybridization accessibility assay [73], ii) synthesis of a probe, and iii) subsequent hybridization.

3. Immunolocalization of VPg (Fig. 7.1C). The 5' end of the viral genomes of poliovirus is covalently linked to a small peptide, VPg. If VPg is exposed, it could be localized immunologically using monoclonal antibodies. For fine localization, only the antigen binding fragments (Fabs) of antibodies should be used. Antibodies digested by papain yields three fragments: two Fab fragments and one Fc fragment.

After complexing the viral genome with the label, the sample should be applied to an EM grid with a holey support film, blotted and vitrified. It is expected that the antisense probe would specifically bind to the complementary viral sequence and

help to unambiguously localize this region of interaction. After asymmetrically reconstruct the images, if the final structure consistently shows a single high dense spot (corresponding to the probe) located always in the same position, the structure is likely ordered and the reconstruction is correct. On the other hand, a number of electrodense spots will mean either i) that the structure is not ordered, ii) a failure of the reconstruction technique, or iii) that the genome is unstructured at that position. In this last case, labeling alternative sequences will help to determine the extent of the unstructured region. The nucleic acid backbone should be traced starting from an area close to the probe.

Bibliography

[1] J. Adamcik, D. V. Klinov, G. Witz, S. K. Sekatskii, and G. Dietler, "Observation of single-stranded DNA on mica and highly oriented pyrolytic graphite by atomic force microscopy," FEBS Lett. **580,** 5671–5675 (2006).

[2] E. Schultes, A. Spasic, U. Mohanty, and D. Bartel, "Compact and ordered collapse of randomly generated RNA sequences," Nat. Struct. Mol. Biol. **12,** 1130–1136 (2005).

[3] C. Hyeon, R. I. Dima, and D. Thirumalai, "Size, shape, and flexibility of RNA structures," J. Chem. Phys. **125,** 194905–1949015 (2006).

[4] J. S. Pfingsten, D. A. Costantino, and J. S. Kieft, "Structural basis for ribosome recruitment and manipulation by a viral IRES RNA," Science **314,** 1450–1454 (2006).

[5] I. Goodfellow, Y. Chaudhry, A. Richardson, J. Meredith, J. W. Almond, W. Barclay, and D. J. Evans, "Identification of a cis-acting replication element within the poliovirus coding region," J. Virol. **74,** 4590–4600 (2000).

[6] V. D'Souza and M. Summers, "Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus," Nature **431,** 586–590 (2004).

[7] J. M. Watts, K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess Jr, R.

Swanstrom, C. L. Burch, and K. M. Weeks, "Architecture and secondary structure of an entire HIV-1 RNA genome," Nature **460,** 711–716 (2009).

[8] J. Speir, S. Munshi, G. Wang, T. Baker, and J. Johnson, "Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy," Structure **3,** 63–78 (1995).

[9] Z. Chen, C. Stauffacher, Y. Li, T. Schmidt, W. Bomu, G. Kamer, M. Shanks, G. Lomonossoff, and J. Johnson, "Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution," Science **245,** 154–159 (1989).

[10] J. Huiskonen, F. de Haas, D. Bubeck, D. Bamford, S. Fuller, and S. Butcher, "Structure of the bacteriophage $^{\phi}$6 nucleocapsid suggests a mechanism for sequential rna packaging," Structure **14,** 1039–1048 (2006).

[11] K. Dryden, S. Wieland, C. Whitten-Bauer, J. Gerin, F. Chisari, and M. Yeager, "Native hepatitis B virions and capsids visualized by electron cryomicroscopy," Mol. Cell **22,** 843–850 (2006).

[12] A. M. Roseman, J. A. Berriman, S. A. Wynne, P. J. G. Butler, and R. A. Crowther, "A structural model for maturation of the hepatitis B virus core," P. Natl. Acad. Sci. USA **102,** 15821–15826 (2005).

[13] A. Zlotnick, N. Cheng, S. J. Stahl, J. F. Conway, A. C. Steven, and P. T. Wingfield, "Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: Implications for morphogenesis and organization of encapsidated RNA," P. Natl. Acad. Sci. USA **94,** 9556–9561 (1997).

[14] F. Booy, W. Newcomb, B. Trus, J. Brown, T. Baker, and A. Steven, "Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus," Cell **64,** 1007–1015 (1991).

[15] S. van den Worm, R. Koning, H. Warmenhoven, H. Koerten, and J. van Duin, "Cryo electron microscopy reconstructions of the *Leviviridae* unveil the densest icosahedral RNA packing possible," J. Mol. Biol. **363,** 858–865 (2006).

[16] R. Golmohammadi, K. Valegård, K. Fridborg, and L. Liljas, "The refined structure of bacteriophage MS2 at 2.8 Å resolution," J. Mol. Biol. **234,** 620–639 (1993).

[17] R. Koning, S. van den Worm, J. R. Plaisier, J. van Duin, J. P. Abrahams, and H. Koerten, "Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2," J. Mol. Biol. **332,** 415–422 (2003).

[18] K. Valegård, J. Murray, P. Stockley, N. Stonehouse, and L. Liljas, "Crystal structure of an RNA bacteriophage coat protein operator complex," Nature pp. 623–626 (1994).

[19] R. McKenna, D. Xia, P. Willingmann, L. Hag, S. Krishnaswamy, M. Rossmann, N. Olson, T. Baker, and N. Incardona, "Atomic structure of single-stranded DNA bacteriophage $_{\Phi}$X174 and its functional implications," Nature **355,** 137–143 (1992).

[20] N. Olson, M. Gingery, F. Eiserling, and T. Baker, "The structure of isometric capsids of bacteriophage T4," Virology **279,** 385–391 (2001).

[21] A. Fokine, P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao, and M. G. Rossmann, "Molecular architecture of the prolate head of bacteriophage T4," P. Natl. Acad. Sci. USA **101,** 6003–6008 (2004).

[22] A. Fokine *et al.*, "A three-dimensional cryo-electron microscopy structure of the bacteriophage $^{\phi}$KZ head," J. Mol. Biol. **352,** 117–124 (2005).

[23] J. Wery, V. Reddy, M. Hosur, and J. Johnson, "The refined three-dimensional structure of an insect virus at 2.8 Å resolution," J. Mol. Biol. **235,** 565–586 (1994).

[24] A. Fisher and J. Johnson, "Ordered duplex RNA controls capsid architecture in an icosahedral animal virus," Nature **361,** 176–179 (1993).

[25] M. Tihova, K. Dryden, T. Le, S. Harvey, J. Johnson, M. Yeager, and A. Schneemann, "Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length," J. Virol. **78,** 2897–2905 (2004).

[26] L. Tang, K. Johnson, L. Ball, T. Lin, M. Yeager, and J. Johnson, "The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA," Nat. Struct. Mol. Biol. **8,** 77–83 (2001).

[27] J. Griffith, D. Griffith, I. Rayment, W. Murakami, and D. Caspar, "Inside polyomavirus at 25-Å resolution," Nature **355,** 652–654 (1992).

[28] J. Tsao *et al.*, "The three-dimensional structure of canine parvovirus and its functional implications," Science **251,** 1456–1464 (1991).

[29] M. Agbandje-McKenna, A. Llamas-Saiz, F. Wang, P. Tattersall, and M. Rossmann, "Functional implications of the structure of the murine parvovirus, minute virus of mice," Structure **6,** 1369–1381 (1998).

[30] E. Hendry, H. Hatanaka, E. Fry, M. Smyth, J. Tate, G. Stanway, J. Santti, M. Maaronen, T. Hyypiä, and D. Stuart, "The crystal structure of coxsackievirus A9: new insights into the uncoating mechanisms of enteroviruses," Structure **7,** 1527–1538 (1999).

[31] J. Muckelbauer, M. Kremer, I. Minor, G. Diana, F. Dutko, J. Groarke, D. Pevear, and M. Rossmann, "The structure of coxsackievirus B3 at 3.5 Å resolution," Structure **3,** 653–667 (1995).

[32] C. Xiao, C. Bator-Kelly, E. Rieder, P. Chipman, A. Craig, R. Kuhn, E. Wimmer, and M. Rossmann, "The crystal structure of coxsackievirus A21 and its interaction with ICAM-1," Structure **13,** 1019–1033 (2005).

[33] A. Hadfield, W. Lee, R. Zhao, M. Oliveira, I. Minor, R. Rueckert, and M. Rossmann, "The refined structure of human rhinovirus 16 at 2.15 Å resolution: implications for the viral life cycle," Structure **5,** 427–441 (1997).

[34] E. Arnold and M. Rossmann, "Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å," J. Mol. Biol. **211,** 763–801 (1990).

[35] R. Zhao, D. Pevear, M. Kremer, V. Giranda, J. Kofron, R. Kuhn, and M. Rossmann, "Human rhinovirus 3 at 3.0 Å resolution," Structure **4,** 1205–1220 (1996).

[36] N. Verdaguer, D. Blaas, and I. Fita, "Structure of human rhinovirus serotype 2 (HRV2)," J. Mol. Biol. **300,** 1179–1194 (2000).

[37] S. Krishnaswamy and M. Rossmann, "Structural refinement and analysis of Mengo virus," J. Mol. Biol. **211,** 803–844 (1990).

[38] K. Lentz *et al.*, "Structure of poliovirus type 2 lansing complexed with antiviral agent SCH48973: comparison of the structural and biological properties of the three poliovirus serotypes," Structure **5,** 961–978 (1997).

[39] D. Filman, R. Syed, M. Chow, A. Macadam, P. Minor, and J. Hogle, "Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus," EMBO J. **8,** 1567–1579 (1989).

[40] E. Fry, N. Knowles, J. Newman, G. Wilsden, Z. Rao, A. King, and D. Stuart, "Crystal structure of swine vesicular disease virus and implications for host adaptation," J. Virol. **77,** 5475 (2003).

[41] J. Chang, P. Weigele, J. King, W. Chiu, and W. Jiang, "Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery," Structure **14,** 1073–1082 (2006).

[42] G. Lander, L. Tang, S. Casjens, E. Gilcrease, P. Prevelige, A. Poliakov, C. Potter, B. Carragher, and J. Johnson, "The structure of an infectious P22 virion shows the signal for headful DNA packaging," Science **312,** 1791–1795 (2006).

[43] Z. Zhang, B. Greene, P. Thuman-Commike, J. Jakana, P. Prevelige Jr, J. King, and W. Chiu, "Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy," J. Mol. Biol. **297,** 615–626 (2000).

[44] X. Agirrezabala, J. Martín-Benito, J. Castón, R. Miranda, J. Valpuesta, and J. Carrascosa, "Maturation of phage T7 involves structural modification of both shell and inner core components," EMBO J. **24,** 3820–3829 (2005).

[45] M. E. Cerritelli, N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, and A. C. Steven, "Encapsidated conformation of bacteriophage T7 DNA," Cell **91,** 271–280 (1997).

[46] M. Cerritelli, J. Conway, N. Cheng, B. Trus, and A. Steven, "Molecular mecha-

nisms in bacteriophage T7 procapsid assembly, maturation, and DNA containment," Adv. Protein. Chem. **64,** 301–323 (2003).

[47] L. Comolli, A. Spakowitz, C. Siegerist, P. Jardine, S. Grimes, D. Anderson, C. Bustamante, and K. Downing, "Three-dimensional architecture of the bacteriophage ϕ29 packaged genome and elucidation of its packaging process," Virology **371,** 267–277 (2008).

[48] A. Petrov, K. Lim-Hing, and S. Harvey, "Packaging of DNA by bacteriophage epsilon15: structure, forces, and thermodynamics," Structure **15,** 807–812 (2007).

[49] Y. Xiang, M. Morais, A. Battisti, S. Grimes, P. Jardine, D. Anderson, and M. Rossmann, "Structural changes of bacteriophage ϕ29 upon DNA packaging and release," EMBO J. **25,** 5229–5239 (2006).

[50] G. Effantin, P. Boulanger, E. Neumann, L. Letellier, and J. Conway, "Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships," J. Mol. Biol. **361,** 993–1002 (2006).

[51] C. Martín, R. Burnett, F. de Haas, R. Heinkel, T. Rutten, S. Fuller, S. Butcher, and D. Bamford, "Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions," Structure **9,** 917–930 (2001).

[52] D. Doan, K. Lee, P. Laurinmäki, S. Butcher, S. Wong, and T. Dokland, "Three-dimensional reconstruction of hibiscus chlorotic ringspot virus," J. Struct. Biol. **144,** 253–261 (2003).

[53] M. Sherman, R. Guenther, F. Tama, T. Sit, C. Brooks, A. Mikhailov, E. Orlova,

T. Baker, and S. Lommel, "Removal of divalent cations induces structural transitions in red clover necrotic mosaic virus, revealing a potential mechanism for RNA release," J. Virol. **80,** 10395–10406 (2006).

[54] R. Aramayo, C. Merigoux, E. Larquet, P. Bron, J. Perez, C. Dumas, P. Vachette, and N. Boisset, "Divalent ion-dependent swelling of tomato bushy stunt virus: a multi-approach study," BBA-Gen. Subjects **1724,** 345–354 (2005).

[55] S. Larson, J. Day, M. Canady, A. Greenwood, and A. McPherson, "Refined structure of desmodium yellow mottle tymovirus at 2.7 Å resolution," J. Mol. Biol. **301,** 625–642 (2000).

[56] B. Böttcher and R. Crowther, "Difference imaging reveals ordered regions of RNA in turnip yellow mosaic virus," Structure **4,** 387–394 (1996).

[57] S. Larson, J. Day, A. Greenwood, and A. McPherson, "Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution," J. Mol. Biol. **277,** 37–59 (1998).

[58] P. Gouet, J. Diprose, J. Grimes, R. Malby, J. Burroughs, S. Zientara, D. Stuart, and P. Mertens, "The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography," Cell **97,** 481–490 (1999).

[59] K. Reinisch, M. Nibert, and S. Harrison, "Structure of the reovirus core at 3.6 Å resolution," Nature **404,** 960–967 (2000).

[60] J. Lawton, M. Estes, and B. Prasad, "Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles," Nat. Struct. Mol. Biol. **4,** 118–121 (1997).

[61] J. B. Pesavento, J. A. Lawton, M. K. Estes, and B. V. V. Prasad, "The reversible condensation and expansion of the rotavirus genome," P. Natl. Acad. Sci. USA **98,** 1381–1386 (2001).

[62] B. Prasad, R. Rothnagel, C. Zeng, J. Jakana, J. Lawton, W. Chiu, and M. Estes, "Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus," Nature **382,** 472–473 (1996).

[63] N. Ban and A. McPherson, "The structure of satellite panicum mosaic virus at 1.9 Å resolution," Nat. Struct. Mol. Biol. **2,** 882–890 (1995).

[64] D. Makino, J. Day, S. Larson, and A. McPherson, "Investigation of RNA structure in satellite panicum mosaic virus," Virology **351,** 420–431 (2006).

[65] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, "Molecular dynamics simulations of the complete satellite tobacco mosaic virus," Structure **14,** 437–449 (2006).

[66] S. Larson, S. Koszelak, J. Day, A. Greenwood, J. Dodds, and A. McPherson, "Double-helical RNA in satellite tobacco mosaic virus," Nature **361,** 179–182 (1993).

[67] S. B. Larson and A. McPherson, "Satellite tobacco mosaic virus RNA: structure and implications for assembly," Curr. Opin. Struc. Biol. **11,** 59–65 (2001).

[68] W. Jiang, J. Chang, J. Jakana, P. Weigele, J. King, and W. Chiu, "Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus," Nature **439,** 612–616 (2006).

[69] N. Opalka, M. Tihova, C. Brugidou, A. Kumar, R. Beachy, C. Fauquet, and

M. Yeager, "Structure of native and expanded sobemoviruses by electron cryomicroscopy and image reconstruction," J. Mol. Biol. **303,** 197–211 (2000).

[70] K. Tars, A. Zeltins, and L. Liljas, "The three-dimensional structure of cocksfoot mottle virus at 2.7 Å resolution," Virology **310,** 287–297 (2003).

[71] K. Wilkinson, R. Gorelick, S. Vasa, N. Guex, A. Rein, D. Mathews, M. Giddings, and K. Weeks, "High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states," PLoS Biol. **6,** e96 (2008).

[72] P. Simmonds, A. Tuplin, and D. Evans, "Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence," RNA **10,** 1337–1351 (2004).

[73] M. Davis, S. M. Sagan, J. P. Pezacki, D. J. Evans, and P. Simmonds, "Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses," J. Virol. **82,** 11824–11836 (2008).

[74] A. M. Yoffe, P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, "Predicting the sizes of large RNA molecules," P. Natl. Acad. Sci. USA **105,** 16153–16158 (2008).

[75] P. Simmonds and D. B. Smith, "Structural constraints on RNA virus evolution," J. Virol. **73,** 5787–5794 (1999).

[76] Y. G. Kuznetsov, S. Daijogo, J. Zhou, B. L. Semler, and A. McPherson, "Atomic force microscopy analysis of icosahedral virus RNA," J. Mol. Biol. **347,** 41–52 (2005).

[77] T. Shimoike, S. Mimori, H. Tani, Y. Matsuura, and T. Miyamura, "Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation," J. Virol. **73,** 9718–9725 (1999).

[78] T. L. Tellinghuisen and R. J. Kuhn, "Nucleic acid-dependent cross-linking of the nucleocapsid protein of sindbis virus," J. Virol. **74,** 4302–4309 (2000).

[79] A. Schneemann, "The structural and functional role of RNA in icosahedral virus assembly," Ann. Rev. Microbiol. **60,** 51–67 (2006).

[80] A. Bertolotti-Ciarlet, L. J. White, R. Chen, B. V. V. Prasad, and M. K. Estes, "Structural requirements for the assembly of norwalk virus-like particles," J. Virol. **76,** 4044–4055 (2002).

[81] D. Marshall and A. Schneemann, "Specific packaging of nodaviral RNA2 requires the N-terminus of the capsid protein," Virology **285,** 165 – 175 (2001).

[82] K. Owen and R. Kuhn, "Identification of a region in the Sindbis virus nucleocapsid protein that is involved in specificity of RNA encapsidation," J. Virol. **70,** 2757–2763 (1996).

[83] E. Frolova, I. Frolov, and S. Schlesinger, "Packaging signals in alphaviruses," J. Virol. **71,** 248–258 (1997).

[84] A. Kramvis and M. Kew, "Structure and function of the encapsidation signal of hepadnaviridae," J. Viral Hepatitis **5,** 357–367 (1998).

[85] F. Qu and T. Morris, "Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size," J. Virol. **71,** 1428–1435 (1997).

[86] K. N. Johnson, L. Tang, J. E. Johnson, and L. A. Ball, "Heterologous RNA encapsidated in pariacoto virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions," J. Virol. **78,** 11371–11378 (2004).

[87] J. M. Johnson, D. A. Willits, M. J. Young, and A. Zlotnick, "Interaction with capsid protein alters RNA structure and the pathway for *in vitro* assembly of cowpea chlorotic mottle virus," J. Mol. Biol. **335,** 455–464 (2004).

[88] X. F. Dong, P. Natarajan, M. Tihova, J. E. Johnson, and A. Schneemann, "Particle polymorphism caused by deletion of a peptide molecular switch in a quasiequivalent icosahedral virus," J. Virol. **72,** 6024–6033 (1998).

[89] M. A. Krol, N. H. Olson, J. Tate, J. E. Johnson, T. S. Baker, and P. Ahlquist, "RNA-controlled polymorphism in the in vivo assembly of 180-subunit and 120-subunit virions from a single capsid protein," P. Natl. Acad. Sci. USA **96,** 13650–13655 (1999).

[90] A. C. Smith, K. L. Poulin, and R. J. Parks, "DNA genome size affects the stability of the adenovirus virion," J. Virol. **83,** 2025–2028 (2009).

[91] A. J. Bett, L. Prevec, and F. L. Graham, "Packaging capacity and stability of human adenovirus type 5 vectors," J. Virol. **67,** 5911–5921 (1993).

[92] J. E. Johnson and R. R. Rueckert, "Packaging and release of the viral genome," In *Structural biology of viruses*, 1st ed., W. Chiu, R. Burnett, and R. Garcea, eds., pp. 271–274 (Oxford University Press New York, 1997).

[93] J. Dubochet, F. Booy, R. Freeman, A. Jones, and C. Walter, "Low temperature electron microscopy," Annu. Rev. Biophys. Bio. **10,** 133–149 (1981).

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

[95] M. Adrian, J. Dubochet, J. Lepault, and A. McDowall, "Cryo-electron microscopy of viruses," Nature pp. 32–36 (1984).

[96] R. Grassucci, D. Taylor, and J. Frank, "Preparation of macromolecular complexes for cryo-electron microscopy," Nat. Protoc. **2,** 3239–3246 (2007).

[97] K. A. Taylor and R. M. Glaeser, "Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future," J. Struct. Biol. **163,** 214–223 (2008).

[98] D. De Rosier and A. Klug, "Reconstruction of three dimensional structures from electron micrographs," Nature **217,** 130–134 (1968).

[99] A. C. Steven and D. M. Belnap, "Electron microscopy and image processing: an essential tool for structural analysis of macromolecules," In *Current Protocols in Protein*, G. P. Taylor, ed., pp. 17.12.11–17.12.39 (John Wiley and Sons, Hoboken, New Jersey, 2005).

[100] Y. Zhang, V. A. Kostyuchenko, and M. G. Rossmann, "Structural analysis of viral nucleocapsids by subtraction of partial projections," J. Struct. Biol. **157,** 356–364 (2007).

[101] J. A. Briggs, J. T. Huiskonen, K. V. Fernando, R. J. Gilbert, P. Scotti, S. J. Butcher, and S. D. Fuller, "Classification and three-dimensional reconstruction of unevenly distributed or symmetry mismatched features of icosahedral particles," J. Struct. Biol. **150,** 332–339 (2005).

[102] H. T. Jãälinoja, E. Roine, P. Laurinmãki, H. M. Kivelã, D. H. Bamford, and S. J. Butcher, "Structure and host-cell interaction of SH1, a membrane-containing, halophilic euryarchaeal virus," P. Natl. Acad. Sci. USA **105,** 8008–8013 (2008).

[103] J. T. Huiskonen, V. Manole, and S. J. Butcher, "Tale of two spikes in bacteriophage PRD1," P. Natl. Acad. Sci. USA **104,** 6666–6671 (2007).

[104] R. R. Rueckert, "Picornaviridae: the viruses and their replication," In *Fields virology*, 3th ed., B. N. Fields, D. M. Knipe, and P. M. Howley, eds., pp. 609–654 (Lippincott-Raven, Philadelphia, Pa, 1996).

[105] J. M. Hogle, M. Chow, and D. J. Filman, "Three-dimensional structure of poliovirus at 2.9 Å resolution," Science **229,** 1358–1365 (1985).

[106] D. M. Belnap *et al.*, "Molecular tectonic model of virus structural transitions: the putative cell entry states of poliovirus," J. Virol. **74,** 1342–1354 (2000).

[107] G. Harauz and F. P. Ottensmeyer, "Direct three-dimensional reconstruction for macromolecular complexes from electron micrographs," Ultramicroscopy **12,** 309–319 (1984).

[108] R. A. Crowther, N. A. Kiselev, B. Böttcher, J. A. Berriman, G. P. Borisova, V. Ose, and P. Pumpens, "Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy," Cell **77,** 943–950 (1994).

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

[110] M. Schatz, E. V. Orlova, P. Dube, J. Jäger, and M. van Heel, "Structure of

Lumbricus terrestris hemoglobin at 30 Å resolution determined using angular reconstitution," J. Struct. Biol. **114,** 28–40 (1995).

[111] T. S. Baker and R. H. Cheng, "A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy," J. Struct. Biol. **116,** 120—130 (1996).

[112] S. J. Ludtke, P. R. Baldwin, and W. Chiu, "EMAN: Semiautomated software for high-resolution single-particle reconstructions," J. Struct. Biol. **128,** 82–97 (1999).

[113] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state* (Oxford University Press, USA, 2006).

[114] D. Bhella, I. G. Goodfellow, P. Roversi, D. Pettigrew, Y. Chaudhry, D. J. Evans, and S. M. Lea, "The structure of echovirus type 12 bound to a two-domain fragment of its cellular attachment protein decay-accelerating factor (CD 55)," J. Biol. Chem. **279,** 8325–8332 (2004).

[115] X. Zhang, E. Settembre, C. Xu, P. R. Dormitzer, R. Bellamy, S. C. Harrison, and N. Grigorieff, "Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction," P. Natl. Acad. Sci. USA **105,** 1867–1872 (2008).

[116] M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank, "Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*," J. Microsc.-Oxford **146,** 113–136 (1987).

[117] A. E. Leschziner and E. Nogales, "The orthogonal tilt reconstruction method: An approach to generating single-class volumes with no missing cone for *ab*

initio reconstruction of asymmetric particles," J. Struct. Biol. **153,** 284–299 (2006).

[118] J. Walz, T. Tamura, N. Tamura, R. Grimm, W. Baumeister, and A. J. Koster, "Tricorn protease exists as an icosahedral supermolecule *in vivo*," Mol. Cell **1,** 59–65 (1997).

[119] P. A. Thuman-Commike and W. Chiu, "Improved common line-based icosahedral particle image orientation estimation algorithms," Ultramicroscopy **68,** 231–255 (1997).

[120] J. R. Castón, D. M. Belnap, A. C. Steven, and B. L. Trus, "A strategy for determining the orientations of refractory particles for reconstruction from cryo-electron micrographs with particular reference to round, smooth-surfaced, icosahedral viruses," J. Struct. Biol. **125,** 209–215 (1999).

[121] M. van Heel, "Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction," Ultramicroscopy **21,** 111–123 (1987).

[122] X. Liu, W. Jiang, J. Jakana, and W. Chiu, "Averaging tens to hundreds of icosahedral particle images to resolve protein secondary structure elements using a Multi-path Simulated Annealing optimization algorithm," J. Struct. Biol. **160,** 11–27 (2007).

[123] T. Ogura and C. Sato, "A fully automatic 3D reconstruction method using simulated annealing enables accurate posterioric angular assignment of protein projections," J. Struct. Biol. **156,** 371–386 (2006).

[124] H. Elmlund, J. Lundqvist, S. Al-Karadaghi, M. Hansson, H. Hebert, and M. Lindahl, "A new cryo-EM single-particle *ab initio* reconstruction method visu-

alizes secondary structure elements in an ATP-fueled AAA+ motor," J. Mol. Biol. **375,** 934–947 (2008).

[125] S. J. Ludtke, D.-H. Chen, J.-L. Song, D. T. Chuang, and W. Chiu, "Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy," Structure **12,** 1129–1136 (2004).

[126] F. Cantele, S. Lanzavecchia, and P. L. Bellon, "The variance of icosahedral virus models is a key indicator in the structure determination: a model-free reconstruction of viruses, suitable for refractory particles," J. Struct. Biol. **141,** 84–92 (2003).

[127] M. van Heel, "Three-dimensional reconstruction from projections with unknown angular relationships," In *Eighth European Congress on Electron Microscopy*, Csanády, Á, ed.,**2,** 1347–1348 (Eighth European Congress on Electron Microscopy, Budapest, Hungary, 1984).

[128] M. van Heel and G. Harauz, "Direct 3D reconstruction from projections with initially unknown angles," In *Pattern Recognition in Practice II*, E. S. Gelsema and L. N. Kanal, eds., pp. 279–288 (Elsevier, North-Holland Publishing, Amsterdam, 1985).

[129] X. Yan, K. A. Dryden, J. Tang, and T. S. Baker, "*Ab initio* random model method facilitates 3D reconstruction of icosahedral particles," J. Struct. Biol. **157,** 211–225 (2007).

[130] J. B. Heymann, M. Chagoyen, and D. M. Belnap, "Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology," J. Struct. Biol. **151,** 196–207 (2005).

[131] J. B. Heymann and D. M. Belnap, "Bsoft: Image processing and molecular modeling for electron microscopy," J. Struct. Biol. **157,** 3–18 (2007).

[132] U. Rawat, A. Zavialov, J. Sengupta, M. Valle, R. Grassucci, J. Linde, B. Vestergaard, M. Ehrenberg, and J. Frank, "A cryo-electron microscopic study of ribosome-bound termination factor RF2," Nature **421,** 87–90 (2003).

[133] J. F. Conway, B. L. Trus, F. P. Booy, W. W. Newcomb, J. C. Brown, and A. C. Steven, "The effects of radiation damage on the structure of frozen hydrated HSV-1 capsids," J. Struct. Biol. **111,** 222–233 (1993).

[134] M. Valle, A. Zavialov, W. Li, S. Stagg, J. Sengupta, R. Nielsen, P. Nissen, S. Harvey, M. Ehrenberg, and J. Frank, "Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy," Nat. Struct. Mol. Biol. **10,** 899–906 (2003).

[135] S. M. Stagg, G. C. Lander, J. Quispe, N. R. Voss, A. Cheng, H. Bradlow, S. Bradlow, B. Carragher, and C. S. Potter, "A test-bed for optimizing high-resolution single particle reconstructions," J. Struct. Biol. **163,** 29 – 39 (2008).

[136] J. F. Conway and A. C. Steven, "Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution," J. Struct. Biol. **128,** 106–118 (1999).

[137] T. S. Baker, J. Drak, and M. Bina, "Reconstruction of the three-dimensional structure of simian virus 40 and visualization of the chromatin core," P. Natl. Acad. Sci. USA **85,** 422–426 (1988).

[138] D. Belnap, N. Olson, N. Cladel, W. Newcomb, J. Brown, J. Kreider, N. Christensen, and T. Baker, "Conserved features in papillomavirus and polyomavirus capsids," J. Mol. Biol. **259,** 249–263 (1996).

[139] D. Salunke, D. Caspar, and R. Garcea, "Polymorphism in the assembly of polyomavirus capsid protein VP1," Biophys. J. **56,** 887–900 (1989).

[140] D. Bubeck, D. J. Filman, N. Cheng, A. C. Steven, J. M. Hogle, and D. M. Belnap, "The structure of the poliovirus 135S cell entry intermediate at 10-angstrom resolution reveals the location of an externalized polypeptide that binds to membranes," J. Virol. **79,** 7745–7755 (2005).

[141] W. O. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," J. Microsc.-Oxford **127,** 127–138 (1982).

[142] J. Frank, A. Verschoor, and M. Boublik, "Computer averaging of electron micrographs of 40S ribosomal subunits," Science **214,** 1353–1355 (1981).

[143] M. Unser, B. L. Trus, J. Frank, and A. C. Steven, "The spectral signal-to-noise ratio resolution criterion: Computational efficiency and statistical precision," Ultramicroscopy **30,** 429–434 (1989).

[144] D. A. Winkelmann, T. S. Baker, and I. Rayment, "Three-dimensional structure of myosin subfragment-1 from electron microscopy of sectioned crystals," J. Cell Biol. **114,** 701–713 (1991).

[145] T. D. Goddard, C. C. Huang, and T. E. Ferrin, "Visualizing density maps with UCSF Chimera," J. Struct. Biol. **157,** 281 – 287 (2007).

[146] *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009.

[147] M. Valle, A. Zavialov, J. Sengupta, U. Rawat, M. Ehrenberg, and J. Frank, "Locking and unlocking of ribosomal motions," Cell **114,** 123–134 (2003).

[148] S. J. Ludtke, M. L. Baker, D.-H. Chen, J.-L. Song, D. T. Chuang, and W. Chiu, "*De novo* backbone trace of GroEL from single particle electron cryomicroscopy," Structure **16,** 441–448 (2008).

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

[150] M. G. Rossmann and Y. Tao, "Cryo-electron-microscopy reconstruction of partially symmetric objects," J. Struct. Biol. **125,** 196–208 (1999).

[151] M. C. Morais, Y. Tao, N. H. Olson, S. Grimes, P. J. Jardine, D. L. Anderson, T. S. Baker, and M. G. Rossmann, "Cryoelectron-microscopy image reconstruction of symmetry mismatches in bacteriophage ϕ29," J. Struct. Biol. **135,** 38–46 (2001).

[152] E. Sanz-García, A. B. Stewart, and D. M. Belnap, "The random-model method enables *ab initio* 3D reconstruction of asymmetric particles and determination of particle symmetry," J. Struct. Biol. **171,** 216–222 (2010).

[153] T. S. Baker, W. W. Newcomb, F. P. Booy, J. C. Brown, and A. C. Steven, "Three-dimensional structures of maturable and abortive capsids of equine herpesvirus 1 from cryoelectron microscopy," J. Virol. **64,** 563–573 (1990).

[154] J. Cello, A. Paul, and E. Wimmer, "Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template," Science **297,** 1016–1018 (2002).

[155] B. Brandenburg, L. Y. Lee, M. Lakadamyali, M. J. Rust, X. Zhuang, and J. M. Hogle, "Imaging poliovirus entry in live cells," PLoS Biol. **5,** e183 (2007).

[156] H. C. Levy, M. Bostina, D. J. Filman, and J. M. Hogle, "Catching a virus in the act of RNA release: a novel poliovirus uncoating intermediate characterized by cryo-electron microscopy," J. Virol. **84,** 4426–4441 (2010).

[157] M. Ida-Hosonuma *et al.*, "The alpha/beta interferon response controls tissue tropism and pathogenicity of poliovirus," J. Virol. **79,** 4460–4469 (2005).

[158] M. Hindiyeh, Q.-H. Li, R. Basavappa, J. M. Hogle, and M. Chow, "Poliovirus mutants at histidine 195 of VP2 do not cleave VP0 into VP2 and VP4," J. Virol. **73,** 9072–9079 (1999).

[159] M. Rossmann *et al.*, "Structure of a human common cold virus and functional relationship to other picornaviruses," Nature **317,** 145–153 (1985).

[160] M. Larkin *et al.*, "Clustal W and Clustal X version 2.0," Bioinformatics **23,** 2947–2948 (2007).

[161] V. R. Racaniello and D. Baltimore, "Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome," P. Natl. Acad. Sci. USA **78,** 4887–4891 (1981).

[162] N. De Jesus, "Epidemics to eradication: the modern history of poliomyelitis," Virol. J. **4,** 70 (2007).

[163] L. P. Beales, A. Holzenburg, and D. J. Rowlands, "Viral internal ribosome entry site structures segregate into two distinct morphologies," J. Virol. **77,** 6574–6579 (2003).

[164] S. Curry, M. Chow, and J. Hogle, "The poliovirus 135S particle is infectious," J. Virol. **70,** 7125–7131 (1996).

[165] S. D. Fuller, S. J. Butcher, R. H. Cheng, and T. S. Baker, "Three-dimensional reconstruction of icosahedral particles—the uncommon line," J. Struct. Biol. **116,** 48–55 (1996).

[166] R. A. Valverde and J. A. Dodds, "Evidence for a satellite RNA associated naturally with the U5 strain and experimentally with the U1 strain of tobacco mosaic virus," J. Gen. Virol. **67,** 1875–1884 (1986).

[167] J. A. Dodds, "Satellite tobacco mosaic virus," Ann. Rev. Phytopath. **36,** 295–310 (1998).

[168] J. Day, Y. G. Kuznetsov, S. B. Larson, A. Greenwood, and A. McPherson, "Biophysical studies on the RNA cores of satellite tobacco mosaic virus," Biophys. J. **80,** 2364 – 2371 (2001).

[169] E. Mancini, M. Clarke, B. Gowen, T. Rutten, and S. Fuller, "Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus," Mol. Cell **5,** 255–266 (2000).

[170] H. Gao, M. Valle, M. Ehrenberg, and J. Frank, "Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset," J. Struct. Biol. **147,** 283–290 (2004).

[171] W. Wriggers, "Using Situs for the integration of multi-resolution structures," Biophys. Rev. **2,** 21–27 (2010).

[172] A. Fiser, R. Do, and A. Šali, "Modeling of loops in protein structures," Protein Sci. **9,** 1753–1773 (2000).

[173] J. Kyte and R. Doolittle, "A simple method for displaying the hydropathic character of a protein," J. Mol. Biol. **157,** 105–132 (1982).

[174] Y. Kuznetsov, S. Larson, J. Day, A. Greenwood, and A. McPherson, "Structural transitions of satellite tobacco mosaic virus particles," Virology **284,** 223–34 (2001).

[175] P. P. Datta, M. R. Sharma, L. Qi, J. Frank, and R. K. Agrawal, "Interaction of the G' domain of elongation factor G and the C-terminal domain of ribosomal protein L7/L12 during translocation as revealed by cryo-EM," Mol. Cell **20,** 723–731 (2005).

[176] H. Jeon and G. G. Shipley, "Localization of the N-terminal domain of the low density lipoprotein receptor," J. Biol. Chem. **275,** 30465–30470 (2000).

[177] T. Makino, M. Tomishige, and M. Kikkawa, "Unique conformation of kinesin-1's neck linker in the nucleotide-free state," Biophys. J. **96,** 135a–135a (2009).

[178] H. Gu, P. Ho, K. Tsang, C. Yu, and B. Xu, "Using biofunctional magnetic nanoparticles to capture Gram-negative bacteria at an ultra-low concentration," Chem. Commun. **2003,** 1966–1967 (2003).

[179] J. Lipka, J. Hainfeld, and J. Wall, "Undecagold labeling of a glycoprotein: STEM visualization of an undecagoldphosphine cluster labeling the carbohydrate sites of human haptoglobin-hemoglobin complex," J. Ultra. Mol. R. **84,** 120–129 (1983).

[180] B. Blechschmidt, V. Shirokov, and M. Sprinzl, "Undecagold cluster modified $tRNA^{Phe}$ from *Escherichia coli* and its activity in the protein elongation cycle," Eur. J. Biochem. **219,** 65–71 (1994).

[181] J. Hainfeld, W. Liu, and M. Barcena, "Gold–ATP," J. Struct. Biol. **127,** 120–134 (1999).

[182] B. Blechschmidt, W. Jahn, J. Hainfeld, M. Sprinzl, and M. Boublik, "Visualization of a ternary complex of the *Escherichia coli* Phe-tRNA^{Phe} and Tu·GTP from *Thermus hermophilus* by scanning transmission electron microscopy," J. Struct. Biol. **110,** 84–89 (1993).

[183] J. Hainfeld, M. Sprinzl, V. Mandiyan, S. Tumminia, and M. Boublik, "Localization of a specific nucleotide in yeast tRNA by scanning transmission electron microscopy using an undecagold cluster," J. Struct. Biol. **107,** 1 (1991).

[184] E. Skripkin, G. Yusupova, M. Yusupov, P. Kessler, C. Ehresmann, and B. Ehresmann, "Synthesis and ribosome binding properties of model mRNAs modified with undecagold cluster," Bioconjugate Chem. **4,** 549–553 (1993).

[185] V. Pavlov, Y. Xiao, B. Shlyahovsky, and I. Willner, "Aptamer-functionalized Au nanoparticles for the amplified optical detection of thrombin," J. Am. Chem. Soc. **126,** 11768–11769 (2004).

[186] A. Alivisatos, K. Johnsson, X. Peng, T. Wilson, C. Loweth, M. Bruchez, and P. Schultz, "Organization of "nanocrystal molecules" using DNA," Nature **382,** 609–611 (1996).

Index