CHAPTER THREE

CryoEM reconstruction approaches to resolve asymmetric features

Daniel J. Goetschius[a,b], Hyunwook Lee[a,c], Susan Hafenstein[a,b,c,*]

[a]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, United States
[b]Department of Medicine, College of Medicine, The Pennsylvania State University, Hershey, PA, United States
[c]Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, United States
*Corresponding author: e-mail address: shafenstein@psu.edu

Contents

Abstract

Although icosahedral viruses have highly symmetrical capsid features, asymmetric structural elements are also present since the genome and minor structural proteins are usually incorporated without adhering to icosahedral symmetry. Besides this inherent asymmetry, interactions with the host during the virus life cycle are also asymmetric. However, until recently it was impossible to resolve high resolution asymmetric features during single-particle cryoEM image processing. This review summarizes the current approaches that can be used to visualize asymmetric structural features. We have included examples of advanced structural strategies developed to reveal unique features and asymmetry in icosahedral viruses.

Advances in Virus Research, Volume 105
ISSN 0065-3527
https://doi.org/10.1016/bs.aivir.2019.07.007

Abbreviations

AP protein	maturation or A-protein (MS2 phage)
CAR	coxsackie- and adenovirus receptor
CryoEM	cryogenic electron microscopy
CVB3	coxsackievirus B3
Fab	antigen-binding fragment
HSV	herpes simplex virus
MS2	bacteriophage MS2
MurA	UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase (*E. coli*)
Murine CMV	murine cytomegalovirus
Protein A2	maturation protein 2 (Q beta phage)
RdRp	RNA-dependent RNA polymerase
SH1	*Haloarcula hispanica* SH1 virus
T number	triangulation number
TBEV	tick-borne encephalitis virus
VP4	viral protein 4 (rotovirus)

1. Icosahedral symmetry

Throughout the history of cryogenic electron microscopy (cryoEM) 3D image reconstruction, icosahedral viruses have long played a prominent role due to the symmetry of the repeating protein subunits that make up the capsid. Early work by Caspar and Klug established the "quasi-equivalence" principle, which is the foundation for modern structural virology (Caspar and Klug, 1962). The most common form for small spherical viruses, the icosahedron, has five-, three- and twofold rotational symmetry axes, and the icosahedral lattice presents the most efficient organization of repeating subunits. The simplest architecture is comprised of only 12 pentamers, but larger icosahedral shells are constructed with additional hexamers in arrangements that depend on quasi-equivalent interactions due to variable environments for the chemically identical structural proteins. Most icosahedral viruses follow the quasi-equivalence principle of Caspar and Klug, defined by a triangulation number, or T number, which describes how many subunits make up the capsid. Assembling an icosahedral capsid from many of the same building blocks, allows the virus to package a small genome that encodes the main structural proteins, which are expressed in multiple copies.

The structural biologist has benefited by using icosahedral structural redundancy to gain resolution by averaging all the many subunits together. However, imposing icosahedral symmetry averaging results in the loss of any

unique feature, which is overwhelmed by signal of the predominant features. Thus, even in the case of structures known to have unique features, strict imposition of icosahedral averaging can still result in an atomic resolution map (Kostyuchenko et al., 2016; Lee et al., 2016; Sirohi et al., 2016; Therkelsen et al., 2018), albeit with the asymmetry obliterated.

2. Asymmetry in icosahedral viruses

Upon packaging, parts of the genome may become icosahedrally-ordered and thus partially visualized during a traditional reconstruction process, but for the most part the virus genome folds into an energetically favorable conformation allowed by interactions with the internal capsid surface. Thus, the genome with or without its packaging proteins does not completely follow the entire symmetry of the capsid and represents an inherent asymmetry that has seldom been resolved. Besides the asymmetry sometimes required for genome packaging, there are "minor" structural proteins, which are present in smaller numbers per capsid than the proteins that assemble the main capsid shell. Many minor structural proteins are incorporated asymmetrically into the mature virus particle.

Many steps during the virus life cycle involve asymmetric changes. Viruses recognize host molecules to facilitate attachment and entry. Genome release may only be successful if the changes leading up to release are asymmetric. Conversely, host immune molecules are triggered that recognize the virus. In a physiological setting these interactions are almost certainly asymmetric or stoichiometric, but have rarely been visualized at atomic resolution using a structural approach. However, with recent advances in cryoEM, asymmetric features can be resolved that were previously missed. This is clearly an exciting advance in structural virology since all viruses contain or develop asymmetric features that play important functions. Furthermore, since asymmetry has been so difficult to study using a structural approach, most of these features are generally poorly understood. Those structures often perform essential functions for viral infection, replication, assembly, and transmission, so these recent advances will provide us with crucial information for future efforts to understand virus infections and to develop new methods for their control.

3. Reconstruction approaches

Standard icosahedral refinement obliterates any underlying asymmetric features due to the averaging process. In the best case, heterogeneous

features may leave partial density or poor local resolution as a clue to their existence. Once one has been clued into the potential existence of structural heterogeneity, the next step is to attempt one of the various asymmetric reconstruction approaches. So far in the field, many different terms have been used to describe techniques solving symmetry mismatch in icosahedral viruses. Building on efforts to categorize this diversity of terms (Huiskonen, 2018), the basic asymmetric approaches can be classified into the following categories:

(a) naive asymmetric refinement.
(b) symmetry relaxation.
(c) symmetry expansion and focused classification/refinement.
(d) subparticle classification/refinement.

Within these categories, additional subroutines can be implemented to visualize asymmetric features. A good example is partial image subtraction. Density may be subtracted from the particle images via a 2D projection of a masked reconstruction to enhance the relative strength of the asymmetric viral features. For instance, the capsid density may be subtracted as a first step to solve the fundamentally asymmetric structure of the viral genome (Liu and Cheng, 2015; Li et al., 2016; Zhang et al., 2015).

Reconstruction terms

Focused = masked
Classification = no further changes in angles or offsets
Localized = subvolume extraction
Blockbased = subvolume extraction
Refinement = includes changes in angles and offsets
Local refinement = changes in angles and offsets applied to masked or extracted subvolumes
Symmetry expansion = reorientation of individual particles
Partial image subtraction = removes densities through use of a masked map

3.1 Standard icosahedral refinement

Even when asymmetry is suspected, an icosahedral refinement is often conducted as a starting point. Importantly, icosahedrally averaging an asymmetric particle can result in atomic resolution (Lee et al., 2016) of a map that has incorporated clues to asymmetry. Details from the icosahedral map may provide additional information on the asymmetry to fuel further decisions about the reconstruction process. Depending on the context, weak density or poor local resolution may suggest which of several potential problems

need to be addressed, such as subunit heterogeneity, partial occupancy of a binding partner, and local or global capsid flexibility. Additional clues to a heterogeneous dataset include a high b-factor or poorer than expected resolution given specific parameters of the dataset, especially particle number and pixel size. Such an example would be having a million particles with a fine pixel size, but when reconstructing the map the resolution stalls out at a modest value. As a matter of practical workflow, the particle orientations solved in the icosahedral refinement serve as an initial starting point for the more advanced asymmetric techniques. Thus, in an "Icosahedral plus" approach, once an initial icosahedral reconstruction has been completed to align the data according to the global capsid symmetry, subsequent operations identify local deviations from that symmetry. These sampling methods currently take the form of single-particle (masked, i.e., focused classification) or subparticle techniques that rely on the extraction of smaller subvolumes (localized classification and block-based reconstruction).

3.2 Naïve asymmetric refinement

In an icosahedral reconstruction, each particle contributes to averaging in 60 different orientations, described by a set of Euler angles defined as rot, tilt, and psi (Scheres, 2012). The simplest approach to resolve asymmetry in otherwise icosahedral viruses is omitting the imposition of symmetry averaging during data processing, which has sometimes been termed a C1 reconstruction approach. This process is computationally expensive, as it requires sampling the entire globe of possible orientations. The naïve asymmetric refinement also sacrifices resolution, because each icosahedral particle contributes to the final reconstruction only once rather than multiple times (i.e., once for every symmetry-related orientation). Additionally, there is a risk of becoming trapped in a local minimum, because each particle contributes only one orientation. However, this orientation may not be the one that emphasizes the asymmetric feature, which allows the signal from local asymmetric features to be overwhelmed by the global icosahedral architecture of the virus capsid during sampling. Thus the algorithm may prematurely settle on a "good" solution while precluding the opportunity to find the ideal solution that emphasizes the underlying asymmetric feature.

The naïve asymmetric approach is especially error prone for deeply heterogeneous datasets, such as an undersaturated complex of an icosahedral virus that only a few binding sites occupied by a host protein molecule. However, if the naïve asymmetric reconstruction is able to reveal unique features, the risks and trade-offs are clearly warranted. Such an approach

was used to reconstruct MS2 phage to 8.7 Å resolution, revealing the location and structure of the asymmetrically incorporated AP protein and genome, whereas an icosahedral reconstruction of the same data reached 4.1 Å while obscuring these asymmetric features (Koning et al., 2016). Likewise, asymmetric reconstructions of immature and mature Kunjin virus (20 Å and 35 Å, respectively) showed an icosahedral configuration of the glycoprotein shell with variable positioning of the nucleocapsid core that was eccentrically positioned in the immature virus, but which became centered in the mature virus (Therkelsen et al., 2018). Additionally, an otherwise naïve asymmetric refinement may be purposefully biased by introducing an extra density to the reference model at the site of the expected asymmetry, as was done in the case of the RNA polymerase within cytoplasmic polyhedrosis virus (Zhang et al., 2015).

3.3 Symmetry relaxation

Symmetry relaxation begins with an icosahedral refinement in order to assign orientations, which allows subsequent iterative sampling of the 60 symmetry-related orientations per particle. This approach effectively "relaxes" the higher-order symmetry and allows for the emergence of underlying asymmetric features. In addition, the process of symmetry relaxation circumvents the high computational costs associated with naive asymmetric refinements because it removes the need to sample the full globe of conformation space.

Liu and Cheng used a symmetry-relaxation approach to identify and reconstruct the 10 RdRp molecules of cypovirus to 4 Å resolution, along with their associated segmented RNA genome (Li et al., 2017; Liu and Cheng, 2015) (Fig. 1). This approach was performed for both non-transcribing and transcribing virus, providing the first in situ structural details of genome replication in cypovirus. In this case, the genome and RdRps were found to exhibit pseudo-D3 symmetry, compared to the overall icosahedral symmetry of the capsid itself. A separate paper (Li et al., 2016) details the software implementation of their reconstruction protocol.

Lee et al. used a similar symmetry-relaxation approach to solve the 7.8 Å asymmetric structure of coxsackievirus B3 (CVB3) interacting with the host receptor, coxsackie- and adenovirus receptor (CAR) inserted in nanodiscs. That structure identified a local reorganization of the viral genome upon receptor engagement (Lee et al., 2016) and formation of a unique pore, likely for genome release. Thus, the 7.8 Å symmetry-mismatch reconstruction revealed details that were abolished by symmetry averaging in the 3.7 Å icosahedral map.

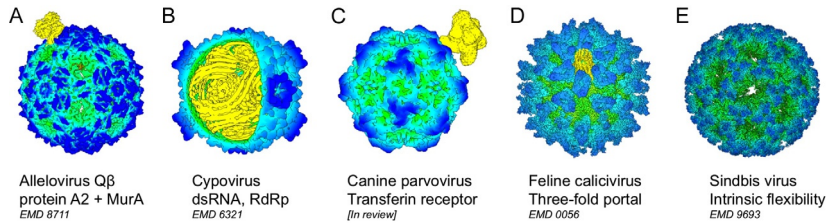

Fig. 1 Recent examples of asymmetric features resolved in icosahedral viruses: (A) Alleovirus Qβ, featuring asymmetrical incorporated maturation protein A2 and MurA (Cui et al., 2017), was reconstructed using a focused (masked) approach. Previous studies resolved protein A2 via naïve asymmetric refinement (Gorzelnik et al., 2016). (B) Cypovirus (reoviridae), with asymmetries including the segmented dsRNA genome and RdRp, responded well to a symmetry-relaxation reconstruction approach (Liu and Cheng, 2015). (C) Canine parvovirus interacting asymmetrically with transferrin-bound transferrin receptor, solved using focused classification (in review). (D) Feline calicivirus, with a unique threefold portal vertex, presumably for genome release, was likewise solved using a masked classification approach (Conley et al., 2019). (E) Sindbis virus (alphaviridae), a large 700 Å diameter virus with intrinsic flexibility that was refined to high resolution using a subparticle approach (Chen et al., 2018).

3.4 Symmetry expansion and focused classification/refinement

3.4.1 Symmetry expansion

Enforcement of icosahedral symmetry during refinement generally limits the computational search range for particle orientations to that of a single asymmetric unit (RELION) (Scheres, 2012), though some programs omit this step (cryoSPARC) (Punjani et al., 2017). Fundamentally, this means that every asymmetric unit contributes equally to the idealized reconstruction, irrespective of unique features that a given portion of the capsid might possess.

To allow interrogation of unique features each particle is assigned every possible symmetry-related orientation, called symmetry expansion. This process aligns each asymmetric unit within the dataset, allowing for subsequent operations to investigate deviation from the idealized icosahedral symmetry. Completing symmetry expansion is an essential first step to adequately perform functions such as focused classification, localized reconstruction, or local refinement.

3.4.2 Discrete vs. continuous heterogeneity

The type of heterogeneity in the target particle will dictate whether to do classification or refinement. Structural heterogeneities may be either discrete or continuous. Discrete heterogeneities include the presence of a unique vertex (HSV) (McElwee et al., 2018), incorporation of minor capsid proteins (A-protein in MS2) (Koning et al., 2016), or partial occupancy

of a binding partner (TBEV) (Füzik et al., 2018). Such heterogeneities can be readily addressed with focused classification, without the need for local refinement beyond the initial icosahedral refinement and symmetry expansion operations.

Continuous heterogeneities may be more difficult to deal with. For instance, a large viral particle may not exhibit perfect structural rigidity, either as a result of the high forces experienced during purification (centrifugation), or as an intrinsic characteristic. Thus any given viral particle may exhibit anisotropy, with deviations in diameter depending on the axis that is chosen. In the case of such squishy particles, local deviations from rigid icosahedral symmetry (x-, y-offsets, orientation angles) exist on a continuous spectrum, rather than as discrete states. Thus a perfect focused classification would result in infinitely many classes, with each possessing a vanishingly small number of particles. In such cases, local refinement of angles and offsets is a more sensible approach. Successful application of local searches during the reconstruction process resulted in high resolution maps for HSV (Dai and Zhou, 2018), sindbis virus (Chen et al., 2018), murine CMV (Liu et al., 2019), and SH1 (De Colibus et al., 2019).

3.4.3 Focused classification

During masked or focused classification, the data underlying the icosahedral reconstruction are expanded to generate all symmetry-related orientations of individual particles (Bai et al., 2015). A mask is then used to define a smaller subvolume within the refined structure, for example, an individual capsomer, unique feature, or capsid region attached to its ligand. These masked volumes can then be subjected to 3D classification as if they were individual particles, without further refinement of angles or x-, y-offsets. This approach allows for identification of distinct classes based solely on the characteristics of the masked volumes and within the context of the otherwise icosahedral structure of the virus.

Although used initially to identify subunit variation in multimeric proteins such as γ-secretase (Bai et al., 2015) as well as particles with higher-order symmetry including GroEL (seven-fold symmetry) (Roh et al., 2017), focused classification has also been used recently to identify deviations from symmetry within icosahedral viruses. This approach was used to reconstruct herpes simplex virus 1 (HSV-1) to 8 Å resolution, identifying a unique five-fold vertex corresponding to the portal as well as the portal-vertex-associated tegument proteins (McElwee et al., 2018). Since the portal itself contains

fivefold symmetry, the final reconstruction could also be accomplished with fivefold symmetry imposed (C5) rather than as a C1 reconstruction, retaining the resolution-boosting benefits of higher-order symmetry. This same approach was used to reveal the threefold vertex portal of calicivirus to high resolution (Conley et al., 2019). Recently, a similar approach was used to reconstruct the structure of the canine parvovirus capsid bound to TfR, which revealed distinct variations in the orientation of the bound receptor with respect to the viral capsid (Lee et al., 2019). In this case, the virus-receptor complex possesses no higher symmetry, limiting the final resolution of the map.

3.4.4 Focused refinement

Once an initial icosahedral refinement followed by symmetry expansion has been completed, local refinement may be used to permit subregions of the viral capsid to make small deviations from strict icosahedral symmetry. These deviations take the form of changes to the x-, y-offset and orientations (angles). To define the subvolume of interest, a soft-edged mask must be generated by a process that currently requires user intervention and optimization. Sjors Scheres (MRC, Cambridge, UK) provides a helpful rule of thumb for successfully implementing masked refinement, namely, that the masked region must contain at least as much mass as would be required to refine an isolated complex. Thus a mass of about 150 kDa (Scheres, 2016) or greater is typically required. It is likely that this size constraint will continue to decrease with further improvements in hardware and software.

3.4.5 Multibody refinement

Multibody refinement (Nakane et al., 2018) is a further development of focused refinement, though application to many asymmetric questions with icosahedral viruses is currently problematic. The technique allows the user to define multiple distinct volumes within the reconstruction via masking. Each masked region is locally refined independently, with the other regions subtracted at the end of each iteration. This allows the image subtraction to become more precise with each subsequent iteration, improving the quality of the local refinement. Practically speaking, the approach can be used to investigate interactions between capsid and a binding partner, by defining a few masked volumes for refinement. Currently; however, it is impractical to use multibody refinement to allow independent, simultaneous iterative local refinement of 60 symmetry-related subvolumes, due to computational limits.

3.4.6 Software

Symmetry expansion, focused classification, focused refinement, and multi-body refinement are currently implemented in RELION 3 (Zivanov et al., 2018). Symmetry expansion is trivial to accomplish in RELION using the command relion_particle_symmetry_expand. At an algorithmic level, each particle orientation (Psi, Rot, Tilt in RELION) is transformed by the selected icosahedral rotation matrix, and a new expanded star file is written with 60 sets of rlnAnglePsi, rlnAngleRot, and rlnAngleTilt values for every original particle. For custom purposes, the same results can be achieved using the freely available python module pyquaternion (http://kieranwynn.github.io/pyquaternion), or the scipy.spatial.transform sub-module (https://docs.scipy.org/doc/scipy/reference/spatial.transform.html).

Focused refinement is also implemented as a beta feature in cryoSPARC v2 (Punjani et al., 2017), where it is termed local refinement. However, symmetry expansion is not currently implemented as a native feature in cryoSPARC, making it challenging to achieve focused refinement of icosahedral viruses with this program. Custom scripts have been developed to implement symmetry expansion in cryoSPARC, however, this process is not currently user-friendly (http://php.scripts.psu.edu/dept/hafenstein/software). A unique feature of cryoSPARC is in allowing the user to define a point within the reconstruction volume as a customized fulcrum, effectively assigning a new center point to the focused region. This advantage allows for much more precise adjustments to orientation angles, as the rotations can be performed about the new fulcrum, rather than the center point of the viral capsid. Liu and Cheng's symmetry-mismatch program is provided as a supplement to their paper describing the technique (Li et al., 2016). Implementation has also been achieved with a modified version of RELION, symbreak (Lee et al., 2016) (http://php.scripts.psu.edu/dept/hafenstein/software).

3.5 Subparticle classification/refinement

Subparticle approaches are fundamentally similar to focused classification and refinement, with a few additional benefits: the ability to adjust for the local defocus of the subparticle of interest (Caspar and Klug, 1962), lower computational burden due to reextract at a smaller box size (Lee et al., 2016), and finer angular adjustment due to reassignment of the center point from bulk capsid to the subparticle of interest (Therkelsen et al., 2018). Computational benefits of a smaller box size are partially offset by the increased storage requirements from extraction of up to 60 subparticles for every 1 particle in the original dataset. Subparticle approaches are frequently coupled with partial image subtraction.

3.5.1 Localized reconstruction

Although focused approaches use a mask to narrow the region of investigation, localized reconstruction targets the volume via extraction of smaller subparticles from the larger virus particle (Ilca et al., 2015). The extracted subvolumes are centered and reextracted from the original micrographs, with or without partial image subtraction. Recentering and reextracting allows for finer adjustments to be made to the angles and offsets during localized refinement, compared to a focused refinement (c. Section 3.4.4). However, it is possible merely to classify the subvolumes without making further refinement to angles and offsets. This localized classification of the extracted subvolumes makes for faster computation and is preferred when the subvolume is too small to allow for further refinement. The final result is a small output subvolume map, which is out of context from the entire map.

Ilca et al. used localized reconstruction to identify the symmetry-mismatched P2 polymerase that is located within the capsid of bacteriophage phi6 under the icosahedral threefold vertex. In the same work, this approach was also used to analyze the partial occupancy of the VP4 spike in rotavirus (Ilca et al., 2015). Localized reconstruction approaches were also used to identify and characterize bound antibody fragments (Fabs) in a partially occupied complex that had about 11 Fab bound per virus. Dai and Zhou also used llca's approach with local refinements to account for local capsid flexibility and refine the HSV-1 capsid to 3.5 Å (Dai and Zhou, 2018). Phi6, rotovirus, and the virus-Fab projects all benefited from removing the density of the bulk particle (partial image subtraction) to aid the process of local angle searches, thus improving the resolution of the subvolume of interest.

3.5.2 Block-based reconstruction

To achieve atomic resolution reconstructions for large icosahedral virus capsids (greater than 50 nm diameter) there are two significant bottlenecks inherent to the molecular mass. First, the capsid shell is a multiple-protein complex, composed of hundreds to thousands of copies of protein subunits. The large size and multimeric nature causes intrinsic structural flexibility that contributes to overall particle deformation, limiting the attainable resolution. Second, and more importantly, the large capsid generates a defocus gradient across the particle because the thickness of the specimen exceeds the microscope depth of field. Thus, the grounding assumption for 3D reconstruction, the projection theorem, does not hold true at high resolution for large viruses because the Fourier transform of the 2D projection cannot truly represent the central section of the Fourier transform of the 3D object (DeRosier, 2000).

This physical phenomenon can be described by the Ewald sphere effect in X-ray crystallography (Ewald, 1921). Normally, particles smaller than 30 nm are not affected by the Ewald sphere effect (Zhu et al., 2018).

Both of these two problems can be resolved by using a "block-based" reconstruction method, in which the object is divided into multiple blocks and each block is reconstructed separately with its own local mean defocus value (Zhu et al., 2018). The separation of the structural components resolves the structural flexibility and can solve the Ewald problem (Russo and Henderson, 2018).

The process for the block-based reconstruction is fundamentally similar to the localized reconstruction approach. After conventional icosahedral refinement, each asymmetric unit is split into multiple "blocks," or sub-volumes, by determining the block positions and centers. The 2D images for all blocks are clipped, or reextracted, from the raw cryoEM images by using the calculated 2D centers. The local defocus of individual blocks is calculated accordingly. The clipped 2D images are used for local refining and reconstructing of the blocks. The culmination of this process is when the reconstructions of all block are combined to generate a full capsid structure, which is unique to the block-based reconstruction method.

The blockbased method was successfully applied to the of HSV-2 B- and C-capsids, producing 3.1 Å and 3.75 Å resolution maps, respectively, as well as atomic models (Wang et al., 2018; Yuan et al., 2018). The HSV capsid shell is 1250 Å in diameter and composed of about 3000 proteins, forming extensive intermolecular interactions. Conventional icosahedral reconstruction limits the resolution to ~4 Å for these large diameter capsids. Subsequently, the same technique was used to solve the structure of Sindbis virus to 3.5 Å (alphaviridae, ~700 Å diameter) (Chen et al., 2018). Future advances in image processing algorithms might improve the achievable resolution even further for the large icosahedral capsids.

3.5.3 Software

A software package, localized reconstruction (Ilca et al., 2015), was written to be used in conjunction with scipion (de la Rosa-Trevín et al., 2016) and RELION (Scheres, 2012). It is available for download at github (https://github.com/OPIC-Oxford/localrec). Block-based reconstruction software is likewise available at github (https://github.com/homurachan/Block-based-recontruction), with the local refinement and reconstruction being accomplished using EMAN (Ludtke et al., 1999), JSPR (Guo and Jiang, 2014), or RELION.

4. Complementary methods

4.1 Molecular dynamics

Another area that shows promise for the analysis of imperfect icosahedral symmetry is molecular dynamics simulations. Some large icosahedral viruses such as hepatitis B virus exhibit deviations from symmetry, resulting in distorted capsids that respond poorly to symmetry averaging, limiting the attainable resolution (Schlicksup et al., 2018). All-atom molecular dynamics simulations have helped define the extent of capsid flexibility and distortion, providing constraints on the resolution achievable by strict symmetry averaging of cryoEM specimens (Hadden et al., 2018). This analysis further supports the utility of subparticle-based local refinement techniques, which compensate for global flexibility. It should also be possible for local subparticle refinement to inform molecular dynamics simulations by comparing parameters from individual locally refined subparticles to their original icosahedral consensus.

4.2 Correlation of local classification

Local classification provides an opportunity to identify structurally distinct regions that are otherwise geometrically identical. This could take the form of an asymmetrically incorporated minor capsid protein (Gorzelnik et al., 2016; Koning et al., 2016), polymerase (Li et al., 2017; Liu and Cheng, 2015; Zhang et al., 2015), or an interacting host protein such as a receptor or antibody fragment (Füzik et al., 2018). Although subvolume classification can be used for purely 3D structural reasons, namely, to reduce heterogeneity and achieve high resolution structures for the contents of each classified volume, the underlying data retains potentially valuable information regarding the original setting of the subvolumes in the context of the individual particles. By returning the distinct subvolume structures to their original whole-particle environments, relationships between components of the entire map can be revealed.

4.2.1 Population histograms

At the most basic level, one can identify the number of occurrences of subvolume class A vs. class B at both a population and a per-particle level. For example, a polymerase might be reasonably expected to be present once-and-only-once in any given capsid. By querying the underlying 3D classification files for the chosen subvolumes (e.g., the relion .star file) one at a time for both class A and class B, one can generate histograms for the number of times each particle identifier is present for a given class. For the case of a polymerase, if

class A corresponds to P+ (polymerase present) and class B corresponds to P− (polymerase absent), this should result in a strong 1:59 ratio in the histograms. Some complexity will necessarily arise if overlapping subvolumes are chosen for the analysis. For instance, in the case of a polymerase under the threefold symmetry axis, three distinct P+ classes will likely be present, one for each 120° rotation about the local symmetry axis, thus resulting in a 1:20 polymerase present to absent ratio. While this seemingly adds complexity, it is nevertheless trivial to compensate for in the analytical pipeline.

4.2.2 Analysis of cooperativity

More advanced analytical techniques can assess interactions between subvolumes on individual capsids, i.e., the relative spatial distribution of subvolume structures in class A vs. class B. We have developed this approach to assess the spatial distribution of antibody fragments (Fab) in undersaturated canine parvovirus + Fab datasets (unpublished). This relies on the *localized reconstruction* package (Ilca et al., 2015) to first extract and recenter all symmetry-related subvolumes. Subsequently, localized 3D classification is performed to distinguish Fab occupied vs. Fab unoccupied sites of the capsid on a per-particle basis. Each individual viral particle is then assigned a nondegenerate fingerprint consisting of an integer string detailing the 3D configuration of Fab fragments. Crucially, this integer string contains the inter-subvolume distances between each occupied site, such that a particle with n occurrences of class A is represented by a non-degenerate string of length $\{n(n-1)/2\}$. These fingerprints are then pooled over the entire dataset and normalized against a fingerprint representing all valid symmetry-related relationships. This process reveals deviations in the distribution of the chosen subvolume from random-chance behavior, specifically, whether certain configurations are favored, under-represented, or forbidden in the dataset. At a biochemical level, this analysis reveals information about underlying positive or negative cooperativity of ligand binding within the sample. As currently implemented, the correlative classification approach relies on binary discrimination between two distinct states, but could be adapted to more complex situations.

5. Summary

Although we have long known that there are many important functions controlled by asymmetric structures in icosahedral viruses, they have been difficult to analyze at a structural level. The development of key new technologies has made it possible to gain high enough resolution to provide a detailed analysis (Fig. 2; Table 1). Undoubtedly, there will be

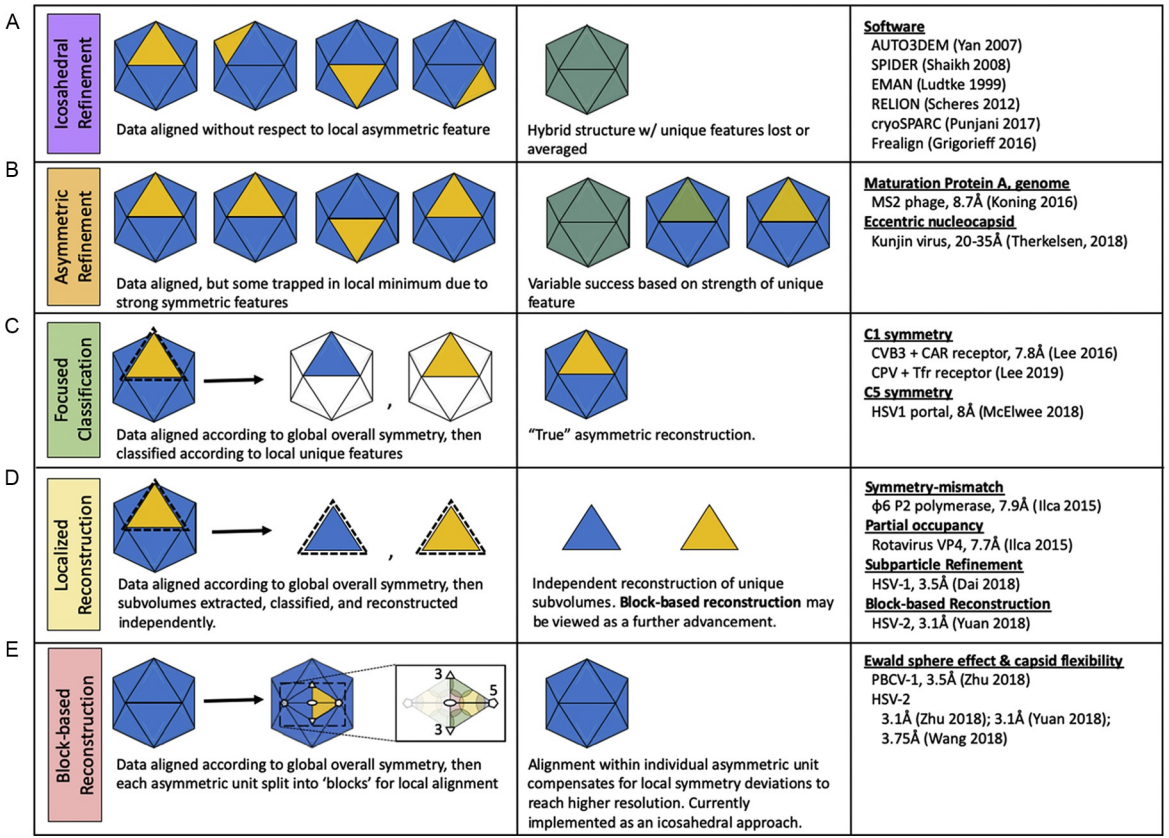

Fig. 2 See legend on next page.

Table 1 Approaches to asymmetry in icosahedral viruses with examples from the literature.

Asymmetric refinement			
• Cytoplasmic polyhedrosis virus	dsRNA genome, RdRp	PMID 26503045	2015
• MS2	ssRNA genome, A-protein	PMID 27561669	2016
• Q-beta	ssRNA genome, A_2	PMID 27671640	2016
• Kunjin virus (flaviviridae)	Eccentric core position	PMID 30348794	2018
Symmetry mismatch			
• Cypovirus (reoviridae)	dsRNA genome, RdRp	PMID 26383954	2015
• Coxsackievirus	CAR receptor	PMID 27574701	2016
• Cypovirus (reoviridae)	dsRNA genome, RdRp	PMID 27914893	2017

Fig. 2 Reconstruction strategies with examples from the literature. (A) In traditional *icosahedral refinement* (I) the higher symmetry is useful for reaching higher resolution, but any asymmetric features are lost. A smaller range of angles is sampled, aiding computation. Multiple software suites have been developed for cryoEM (Grigorieff, 2016; Ludtke et al., 1999; Punjani et al., 2017; Scheres, 2012; Shaikh et al., 2008; Yan et al., 2007) (B) In traditional *asymmetric refinement* (C1) (Koning et al., 2016; Therkelsen et al., 2018), the full range of angles (rot, tilt, psi) must be sampled during refinement. Failure to impose symmetry effectively reduces particle number 60-fold, reducing overall resolution. Refinement may or may not resolve asymmetric features, as strong features present in the otherwise icosahedral structure may trap the algorithm in a local (false) minimum. Success depends on the relative strength of the asymmetric feature. (C) *Focused classification* begins with a traditional icosahedral refinement, and then subsequently implements a mask to sample and 3D-classify symmetry-related locations within the map. Following classification, a full reconstruction can be completed, revealing the asymmetric feature within the context of the overall map (Lee et al., 2016; McElwee et al., 2018). (D) *Localized reconstruction* is similar to focused classification, but subvolumes of interest are centered and extracted before 3D-classification. The process is compatible with partial image subtraction and local refinement to boost the resolution of the unique feature of interest (Dai and Zhou, 2018; Ilca et al., 2015; Yuan et al., 2018). (E) After initial icosahedral refinement, *block-based reconstruction* instead divides the asymmetric unit into smaller blocks for local refinement. Following local refinement, blocks are rebuilt into an icosahedral map, thus compensating for local imperfections of the icosahedral symmetric due to intrinsic capsid flexibility of large viruses (Wang et al., 2018; Yuan et al., 2018; Zhu et al., 2018).

Table 1 Approaches to asymmetry in icosahedral viruses with examples from the literature.—cont'd

Focused/masked approaches			
• HIV-1	Intasome	PMID 28059769	2017
• Q-beta	ssRNA genome, A_2	PMID 29078304	2017
• Herpes Simplex Virus	Portal-vertex-assoc. tegument	PMID 29924793	2018
• Calicivirus	Portal	PMID 30626974	2019
Symmetry expansion and classification			
• Bacteriophage IKe <not Icos>	ssDNA genome	PMID 30819888	2019
Subparticle approaches			
• Bacteriophage phi6	RdRp	PMID 26534841	2015
• Adenovirus D26	Fiber	PMID 28508067	2017
• Tick-borne encephalitis virus	Fab, partial occupancy	PMID 29382836	2018
• Rift Valley fever virus	Capsomers	PMID 29367607	2018
• HSV-1	Capsid "blocks"	PMID 29622628	2018
• HSV-2	Capsid "blocks"	PMID 29674632	2018
• Sindbis virus (alphaviridae)	Capsid "blocks"	PMID 30552337	2018
• Murine CMV	Capsid "blocks"	PMID 30779794	2019
• SH1 virus	Capsid "blocks"	PMID 30926810	2019
• HCIV-1, HHIV-2	Spikes	PMID 30862777	2019

additional refinements on the methods outlined here. These approaches provide new opportunities to gain a more complete understanding of asymmetric structure-function relationships. The many dynamic, partially occupied, and asymmetric features of viruses will also certainly provide important new targets for antiviral interventions.

Acknowledgments

This work was funded in part by the Pennsylvania Department of Health CURE funds, Huck Institutes at The Pennsylvania State University, and by the Office of the Director, National Institutes of Health, under Award Numbers 1 R01 AI107121-01 (SH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Bai, X., Rajendra, E., Yang, G., Shi, Y., Scheres, S.H.W., 2015. Sampling the conformational space of the catalytic subunit of human γ-secretase. eLife 4, e11182. https://doi.org/10.7554/eLife.11182.

Caspar, D.L., Klug, A., 1962. Physical principles in the construction of regular viruses. Cold Spring Harb. Symp. Quant. Biol. 27, 1–24.

Chen, L., et al., 2018. Implication for alphavirus host-cell entry and assembly indicated by a 3.5Å resolution cryo-EM structure. Nat. Commun. 9, 5326.

Conley, M.J., et al., 2019. Calicivirus VP2 forms a portal-like assembly following receptor engagement. Nature 565, 377–381.

Cui, Z., et al., 2017. Structures of Qβ virions, virus-like particles, and the Qβ-MurA complex reveal internal coat proteins and the mechanism of host lysis. Proc. Natl. Acad. Sci. U. S. A. 114, 11697–11702.

Dai, X., Zhou, Z.H., 2018. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science 360 eaao7298. https://doi.org/10.1126/science.aao7298.

De Colibus, L., et al., 2019. Assembly of complex viruses exemplified by a halophilic euryarchaeal virus. Nat. Commun. 10, 1456.

de la Rosa-Trevín, J.M., et al., 2016. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. 195, 93–99.

DeRosier, D.J., 2000. Correction of high-resolution data for curvature of the Ewald sphere. Ultramicroscopy 81, 83–98.

Ewald, P., 1921. Evaluation of optical and electrostatic lattice potentials. Ann. Phys. 64, 253–287.

Füzik, T., et al., 2018. Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody. Nat. Commun. 9, 436.

Gorzelnik, K.V., et al., 2016. Asymmetric cryo-EM structure of the canonical Allolevivirus Qβ reveals a single maturation protein and the genomic ssRNA in situ. Proc. Natl. Acad. Sci. U. S. A. 113, 11519–11524.

Grigorieff, N., 2016. Frealign: an exploratory tool for single-particle cryo-EM. Methods Enzymol. 579, 191–226.

Guo, F., Jiang, W., 2014. Single particle cryo-electron microscopy and 3-D reconstruction of viruses. Methods Mol. Biol. 1117, 401–443.

Hadden, J.A., et al., 2018. All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. eLife 7, e32478. https://doi.org/10.7554/eLife.32478.

Huiskonen, J.T., 2018. Image processing for cryogenic transmission electron microscopy of symmetry-mismatched complexes. Biosci. Rep. 38, BSR20170203. https://doi.org/10.1042/BSR20170203.

Ilca, S.L., et al., 2015. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nat. Commun. 6, 8843.

Koning, R.I., et al., 2016. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. Nat. Commun. 7, 12524.

Kostyuchenko, V.A., et al., 2016. Structure of the thermally stable Zika virus. Nature 533, 425–428.

Lee, H., et al., 2016. The novel asymmetric entry intermediate of a picornavirus captured with nanodiscs. Sci. Adv. 2, e1501929.

Lee, H., et al., 2019. Transferrin receptor binds virus capsid with dynamic motion. Proc. Natl. Acad. Sci. (in press; accepted August 29, 2019).

Li, X., Liu, H., Cheng, L., 2016. Symmetry-mismatch reconstruction of genomes and associated proteins within icosahedral viruses using cryo-EM. Biophys. Rep. 2, 25–32.

Li, X., et al., 2017. Near-atomic resolution structure determination of a cypovirus capsid and polymerase complex using cryo-EM at 200 kV. J. Mol. Biol. 429, 79–87.
Liu, H., Cheng, L., 2015. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. Science 349, 1347–1350.
Liu, W., et al., 2019. Atomic structures and deletion mutant reveal different capsid-binding patterns and functional significance of tegument protein pp150 in murine and human cytomegaloviruses with implications for therapeutic development. PLoS Pathog. 15, e1007615.
Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. 128, 82–97.
McElwee, M., Vijayakrishnan, S., Rixon, F., Bhella, D., 2018. Structure of the herpes simplex virus portal-vertex. PLoS Biol. 16, e2006191.
Nakane, T., Kimanius, D., Lindahl, E., Scheres, S.H., 2018. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. eLife 7, e36861. https://doi.org/10.7554/eLife.36861.
Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296.
Roh, S.-H., et al., 2017. Subunit conformational variation within individual GroEL oligomers resolved by cryo-EM. Proc. Natl. Acad. Sci. U. S. A. 114, 8259–8264.
Russo, C.J., Henderson, R., 2018. Ewald sphere correction using a single side-band image processing algorithm. Ultramicroscopy 187, 26–33.
Scheres, S.H.W., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530.
Scheres, S.H.W., 2016. Processing of structurally heterogeneous cryo-EM data in RELION. Methods Enzymol. 579, 125–157.
Schlicksup, C.J., et al., 2018. Hepatitis B virus core protein allosteric modulators can distort and disrupt intact capsids. eLife 7, e31473. https://doi.org/10.7554/eLife.31473.
Shaikh, T.R., et al., 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protoc. 3, 1941–1974.
Sirohi, D., et al., 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. Science 352, 467–470.
Therkelsen, M.D., et al., 2018. Flaviviruses have imperfect icosahedral symmetry. Proc. Natl. Acad. Sci. U. S. A. 115, 11608–11612.
Wang, J., et al., 2018. Structure of the herpes simplex virus type 2 C-capsid with capsid-vertex-specific component. Nat. Commun. 9, 3668.
Yan, X., Sinkovits, R.S., Baker, T.S., 2007. AUTO3DEM—an automated and high throughput program for image reconstruction of icosahedral particles. J. Struct. Biol. 157, 73–82.
Yuan, S., et al., 2018. Cryo-EM structure of a herpesvirus capsid at 3.1 Å. Science 360, eaao7283. https://doi.org/10.1126/science.aao7283.
Zhang, X., et al., 2015. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. Nature 527, 531–534.
Zhu, D., et al., 2018. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle cryo-EM reconstructions. Nat. Commun. 9, 1552.
Zivanov, J., et al., 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife 7, e42166. https://doi.org/10.7554/eLife.42166.