JOURNAL OF MOLECULAR RECOGNITION
J. Mol. Recognit. 2005; **18**: 479–490
Published online 29 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/jmr.754

Review

Theoretical aspects of virus capsid assembly

Adam Zlotnick*

Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

A virus capsid is constructed from many copies of the same protein(s). Molecular recognition is central to capsid assembly. The capsid protein must polymerize in order to create a three-dimensional protein polymer. More than structure is required to understand this self-assembly reaction: one must understand how the pieces come together in solution. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords: **virus assembly; self-assembly; protein polymerization; landscape; capsid protein**

Received 12 July 2005; revised 8 August 2005; accepted 19 August 2005

INTRODUCTION AND SCOPE

Problem—role of assembly and stability for a virus

A virus is a tremendous example of the selfish gene hypothesis (Dawkins, 1989). It has only one task—producing copies of itself—which it must accomplish by evading host defenses and taking advantage of its host's reluctant hospitality. The virus's nucleic acid is protected by its capsid, a protein complex sometimes enveloped by lipid. Virus capsids are invariably multimers so that many copies of capsid protein(s), encoded by a relatively short nucleic acid sequence, can enclose, or encapsidate, a relatively large volume of nucleic acid. This strategy reduces the burden of encoding a large protein but generates a new problem of assembling multimers, with tens, hundreds or thousands of subunits, in a biologically realistic time frame.

This primary problem is compounded by the peculiarities of different viruses. In some cases, the nucleic acid must be uncoated during the course of infection, requiring the stability of the capsid to be adjustable. For some viruses, the capsid must incorporate specific machinery for packaging nucleic acid (e.g. bacteriophages and herpesviruses) or for replicating the already packaged nucleic acid (reoviruses and hepadnaviruses).

In vitro assembly systems have given us a window on virus assembly. *In vitro* assembly immediately tells us that these complex structures can self-assemble: all the information required for assembly is built into the tertiary structure of the capsid protein(s). Because we can observe assembly *in vitro*, and thence in progressively more physiological environments, we have an opportunity to understand an important biological process. This information can be used to therapeutically interfere with assembly, to subvert assembly to generate nanostructures (Falkner *et al.*, 2005; Johnson *et al.*, 2005; Prevelige, 1998; Stray *et al.*, 2005; Wang *et al.*, 2002; Zlotnick *et al.*, 2002), or applied to understanding assembly of nanostructures with virus-like symmetry (Bowden *et al.*, 2001; Duan *et al.*, 2004; Schweiger *et al.*, 2005; Seidel and Stang, 2002).

Virus assembly has been compared with crystallization (Caspar, 1980). The obvious similarity is that both viruses and crystals are symmetrical supramolecular structures. The obvious difference is that viruses have a defined size whereas crystals are conceptually an infinite lattice. Consequently, crystallization and capsid assembly must be described differently. Compare a typical protein crystallization experiment with an *in vitro* assembly reaction. In the successful crystallization experiment, a single nucleation event will lead to one crystal composed of 10^{12} or more proteins in a matter of days to weeks. In a typical *in vitro* assembly experiment for hepatitis B virus (HBV) capsids, over a period of seconds to hours there will be 10^{11} or more nucleation events leading to a like number of capsids, each comprising exactly 240 proteins (Zlotnick *et al.*, 1999). The sigmoidal kinetics for both processes only look similar; they disguise very different physical behavior.

Viruses come in a variety of shapes: filaments, bullet-shaped, irregular and spherical. Roughly half of all virus families are spherical, usually icosahedral; those viruses are the focus of this review. Understanding the assembly of spherical/icosahedral viruses requires an understanding of the behavior of this peculiar class of assembly reaction, formation of large defined-size oligomers, and experimental systems that allow one to observe assembly. Simple models of assembly will be discussed to enumerate their starting assumptions, describe their 'predictions', and evaluate their limitations.

THE SHAPE OF SPHERICAL VIRUSES

Crick and Watson (1956) proposed icosahedral geometry for spherical capsids, reasoning that the most effective way of

*Correspondence to: A. Zlotnick, Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, PO Box 26901, BRC464, Oklahoma City, OK 73104, USA.
E-mail: adam-zlotnick@ouhsc.edu

Contract/grant sponsor: American Cancer Society; *contract/grant number:* RSG-99-339-04-MBC.
Contract/grant sponsor: NIH; *contract/grant number:* 5R01-EB000432.

Abbreviations used: asu, asymmetric unit; CCMV, cowpea chlorotic mottle virus; HBV, hepatitis B virus; HIV, human immunodefficiency virus.

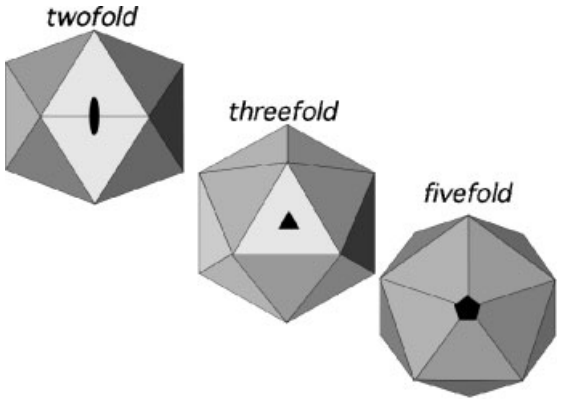

Figure 1. Icosahedra and icosahedral symmetry. Views are down respective symmetry axes. All symmetry axes cross in the center of the icosahedron.

enclosing a large volume with a small protein is to arrange it in a regular geometric repeat. Corroboration for this hypothesis was provided by D.L.D. Caspar, who noted that the arrangement of intense reflections in a precession X-ray photograph of Bushy Stunt Virus corresponded to the arrangement of icosahedral five-folds (Caspar, 1956), an interpretation that presaged rotation function searches for non-crystallographic symmetry (Rossmann and Blow, 1962). Of the Platonic solids, solids composed of identical subunits, icosahedra have the largest number of subunits. An icosahedron has 20 facets, where each facet is an equilateral triangle (Fig. 1). Since the facets each have three-fold symmetry, the resulting solid has $3 \times 20 = 60$ identical asymmetric units. Icosahedra can also be defined in terms of their 532-point symmetry. Each of the 12 icosahedral vertexes is coincident with a five-fold symmetry axis ($5 \times 12 = 60$); of course, symmetry axes run through one five-fold, through the center of the icosahedron, and out the other side. Three-fold axes pass through the center of the facets. There are 30 two-fold axes ($2 \times 30 = 60$) which run through the edge–edge contacts between facets.

The advantage of icosahedral geometry was that a relatively short RNA/DNA sequence could code for a protein that would encapsidate the complete genome. The theory of quasi-equivalence describes how multiples of 60 proteins can be arranged with icosahedral symmetry to enclose an even larger volume. It has been one of the great unifying themes of structural virology since it was introduced by Caspar and Klug (1962). The underlying postulate of quasi-equivalence is that the viral coat protein can form pentamers and hexamers using the same intersubunit contacts. The only requirement for an icosahedral facet is that it be an equilateral triangle with three identical asymmetric units arranged with threefold symmetry. Using a hexagonal grid (Fig. 2), one can see that there is an infinite series of equilateral triangles that enclose an integral number of smaller triangles.

A facet can be divided into three equivalent asymmetric units (asu). The area of an asu for a triangle with one vertex on the origin (0, 0) and a second vertex at a point (h, k) is:

$$T = h^2 + hk + k^2 \quad (1)$$

In the context of the icosahedron, the corners of the triangle are the icosahedral five-folds. The hexagonal grid points in the facet become the quasi-six-fold vertices. Quasi-symmetry arises because the subunits within an asu are not equivalent.

Determining the T number from a structure is straightforward in most cases. One identifies the icosahedral facet by its five-folds and counts the vertices from one five-fold to the next. The largest facet in the left of Fig. 2 would generate a $T = 16$ virus such as Herpes Simplex I (Booy *et al.*, 1988). In our hexagonal coordinate system the first five-fold is at the

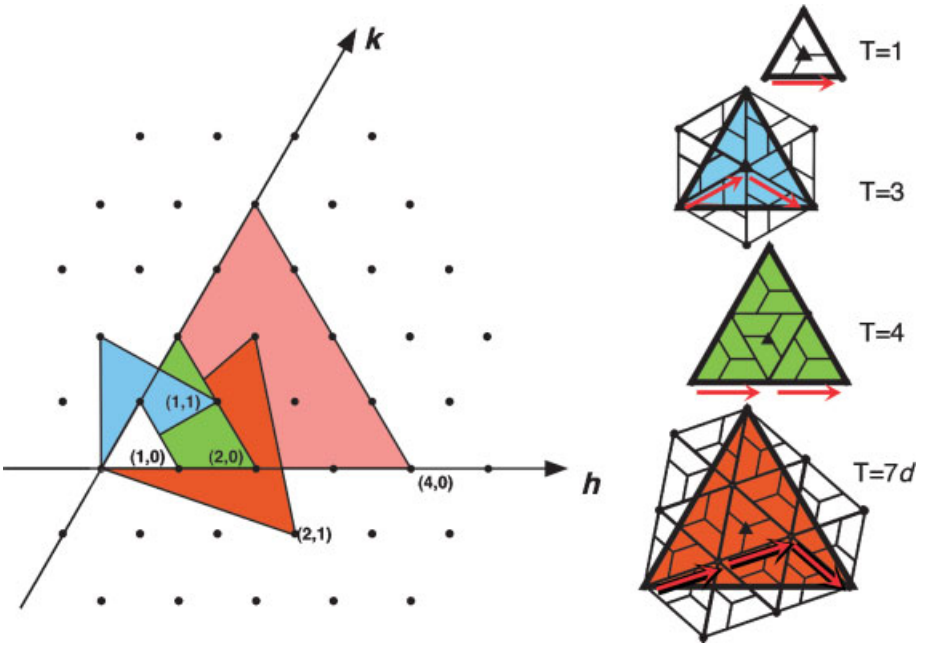

Figure 2. Facets for icosahedra with quasi-symmetry. The facets are equilateral triangles with an integral number of subunits. The geometry of each facet is described by the hexagonal coordinate system (left). On the right are selected facets, drawn with trapezoidal subunits. The arrows show how to determine the *h, k* index by counting vertices from five-fold to five-fold.

origin, $h = k = 0$, which places the second five-fold at $h = 4$, $k = 0$ to give $T = 16$. Consider two other examples: the facet for a $T = 3$ virus (e.g. CCMV; Speir *et al.*, 1995) has one quasi-six-fold vertex; therefore the coordinate for the second five-fold is $h = 1$, $k = 1$. $T = 7$ lattices (second from largest in the left panel of Fig. 2), the shortest path between five-folds first goes through two quasi-six-fold vertices ($h = 2$) and then requires a right turn to the five-fold ($k = 1$). The $T = 7$ lattice is an example of a chiral 'skew' lattice: a right-handed (*dextro*) $T = 7d$ lattice is observed in all known papovaviruses (Belnap *et al.*, 1996; Liddington *et al.*, 1991; Stehle and Harrison, 1996).

The greatest success of the theory of quasi-equivalence has been the prediction of lattices for the arrangement of capsid proteins. The 1962 prediction of structural identity for quasi-equivalent interactions (Caspar and Klug, 1962) has not been borne out; this can be seen by comparing quasi-equivalent interactions using the contact tables developed in VIPER (Reddy *et al.*, 2001). However, the geometry of quasi-equivalent lattices is prevalent: the icosahedral asu of some virus capsids mimics a quasi-equivalent lattice though it is made of different proteins. Two notable examples are the picornaviruses (cf. rhinovirus, poliovirus, and foot and mouth disease) and the comoviruses. The asu for both of these virus families contains three homologous ß-barrels arranged on a $T = 3$ lattice. For convenience this arrangement is referred to as $P = 3$ (Rossmann and Johnson, 1989). The $\sim$1900 Å diameter capsid of iridoviruses, amongst the largest known viruses, is arranged with $T = 169d$ ($h = 8$, $k = 7$) skew lattice; careful structural analysis shows that the lattice is misleading as the quasi-six-fold 'capsomeres' are really trimers (Yan *et al.*, 2000). A better known example of using the hexagonal lattice predicted by quasi-equivalence with non-hexagonal capsomeres is found with papovaviruses (Belnap *et al.*, 1996; Liddington *et al.*, 1991; Rayment *et al.*, 1982; Salunke *et al.*, 1989), where pentameric capsomeres fill both pentavalent (five-fold) and hexavalent positions in a $T = 7d$ lattice (Fig. 3).

Of course, the icosahedral asu may also be an asymmetric oligomer with no trace of quasi-equivalence. A common expression of this is in double-stranded RNA viruses capsids where the asu is composed of two copies of the capsid protein in distinctly different environments and conformations (Caston *et al.*, 1997; Cheng *et al.*, 1994; Grimes *et al.*, 1997, 1998; Hill *et al.*, 1999; Naitow *et al.*, 2002). A similar arrangement is observed when bromovirus capsid protein (CP) is overexpressed (Krol *et al.*, 1999) or the *in vitro* assembly of the related Cowpea Chlorotic Mottle Virus (CCMV) is perturbed (Tang *et al.*, 2004; Zlotnick *et al.*, 2000). The resulting $T = 1$ icosahedron, sometimes referred to as pseudo-$T = 2$ particle, can be described as an association of 12 star-like pentamers of CP dimers. The A half of each dimer is closely associated with the pentamer and icosahedral five-fold. The B half of each dimer extends outward to interdigitate between the B subunits of neighboring pentamers.

Some spherical viruses are not icosahedral. Retroviruses appear to incorporate some elements of symmetry but are not icosahedral. The capsids of retrovirus provirions are roughly spherical structures comprised of $\sim$2000 copies of the Gag polyproteins. Although this structure has well-defined regions of hexagonal symmetry, this may represent only local order (Fuller *et al.*, 1997; Nermut *et al.*, 1998). After proteolytic maturation, HIV capsid protein becomes a Fullerene cone (Benjamin *et al.*, 2005). Elongated, distorted but geometrical icosahedra are observed in bacteriophage T4 (Black and Showe, 1983). Many bacteriophages have their icosahedral symmetry disrupted by a unique vertex, which is occupied by a portal for their DNA packaging machinery (e.g. lambdoid phages, Hendrix and Garcea, 1994; and T7, Cerritelli and Studier, 1996).

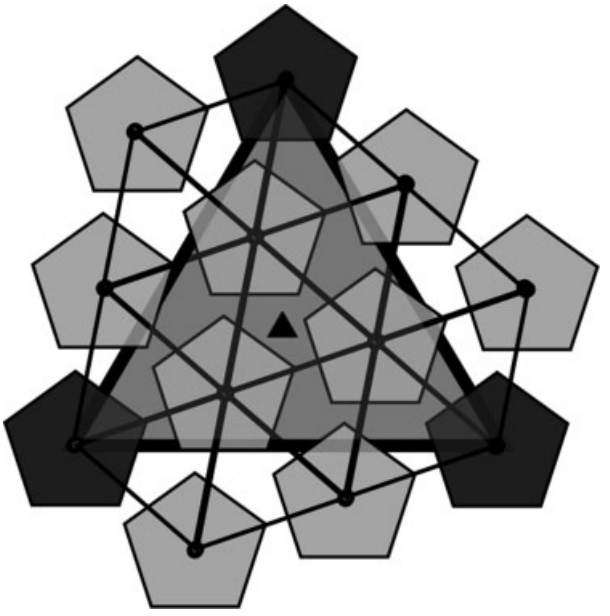

Figure 3. The breakdown and preservation of quasi-equivalence in papovaviruses. Papovaviruses (papillomaviruses, polyomaviruses, and SV40) are assembled from 72 pentamers of the major capsid protein. The 12 pentamers are located on five-fold icosahedral symmetry axes. The remaining pentamers are located at positions predicted for hexamers in a $T = 7d$ lattice. The hexavalent pentamers have distinctly non-equivalent interactions with their neighbors. Only the pentamers within the facet comprise the icosahedral repeat. The pentamers outside of the outline of the facet were included to parallel Fig. 2 and to help illustrate the hexavalent pentamers; they are not consistent with icosahedral geometry.

New outlooks on quasi-symmetry

Three recent re-examinations of physical basis for icosahedral symmetry and quasi-equivalence bear examination. A statistical examination of the arrangement of subunits tiling a sphere, forming weak non-specific interactions, suggests that quasi-equivalent lattices are the lowest energy icosahedral arrangement of subunits (Marzec and Day, 1993). Maintaining icosahedral constraints, this approach was able to generate correct packing for all T numbers up to 48 ($h = 4$, $k = 4$), including skew lattices.

A statistical mechanical examination of assembly is more informative. Polymerization of geometric solids without icosahedral constraints demonstrates that non-icosahedral particles can yield quasi-stable structures (Bruinsma *et al.*, 2003). However, at equilibrium, icosahedral structures are more stable (Zandi *et al.*, 2004). To achieve these results, interactions between circular capsomeres were minimized

in Monte-Carlo simulations. In order to favor icosahedral symmetry, it was necessary to impose a structural difference between five-fold and quasi-six-fold capsomeres. These results are contrary to the assumptions described in the first paragraph of this section (Caspar and Klug, 1962; Crick and Watson, 1956) and leads to the predictions of (i) non-icosahedral spherical viruses (such as retroviruses; Yeager *et al.*, 1998) and (ii) a specific evolutionary advantage to icosahedral geometry. Examination of real icosahedral structures shows that there are different geometrical manifestations of quasi-equivalence (Damodaran *et al.*, 2002). This quantitative re-examination of structure allows segregation of the numerous $T=3$ structures into distinct architectural categories.

Tesselation: a different view of a spherical particle

A virus capsid can be considered a sphere tiled with capsid protein. This view has led to a novel treatment of capsid structure based on tiling theory (Twarock, 2004). The results of this approach are particularly illuminating when applied to papovaviruses, which are assembled from pentavalent and hexavalent capsomeres. Twarock focused on the topology of inter-capsomere contacts, rather than the proteins. Trivalent contacts were described as kites, bivalent contacts as rhoms. Thus, although each capsomere was comprised of five proteins, it interacted with five or six neighbors, generalizing Caspar–Klug quasi-equivalence to a situation where it seemingly broke down. The same classes of contacts were also sufficient to describe tubular structures constructed of polyoma's capsomeres (Twarock, 2005).

MODELING CAPSID ASSEMBLY

Models of assembly are basically mathematical expressions of simple association reactions. While the individual reactions may be simple, the large number of reactions and the rules that govern how one reaction relates to the next can make the overall model rather imposing. The models described below fit into two general categories. Thermodynamic–kinetic models, which describe the behavior of large populations of particles, are based on simpler computations and can be directly applied to *in vitro* assembly systems. Statistical mechanical models generally make use of molecular dynamics to examine small populations of subunits and capsids. These models have been used to examine details of assembly reactions, including regulation quasi-equivalence, and the behavior of assembled capsids.

Assembly of large populations of capsids

How relevant is *in vitro* assembly to the biological process? In terms of concentration, *in vitro* reactions compare well. For many virus infections, the major function of the host cell becomes the manufacture of progeny virions. It is not unusual to see electron micrographs of a cell so loaded with virions that the particles are packed in crystal-like arrays. A crude estimate of cell volume can be based on water content: 10^{-15} ml for the bacteria *E. coli* and 10^{-12} ml for eukaryotic HeLa cells (Darnell *et al.*, 1990). A typical *in vitro* reaction may start with 20 μM capsid protein and produce 5×10^{13} particles per ml. At that particle density, there would be 50 particles in the volume of a bacteria, which compares well with the ca. 150 particles formed in an infection by a *T*-even phage. There would be 50 000 *in vitro* assembled particles in the volume of a HeLa cell, which compares well with the yield of 25 000–100 000 virion per cell for picornaviruses (Rueckert, 1996); similar yields of 100 000 virion per cell are found with reoviruses (Ross and Subramanian, 1981). Eukaryotic cells are highly compartmentalized, meaning that *in vivo* viral concentrations may be much higher. Conversely, many viruses produce far less progeny so that concentrations *in vivo* may be much lower. The point is that these simplest-case *in vitro* reactions occur at biologically relevant concentrations.

Of course, this argument does not address the effects of intracellular compartmentalization nor macromolecular crowding. Within a cell, the concentration of protein and nucleic acid is so high that it favors protein association by osmotic effects and altering water activity, and decreases diffusion, leading to compartmentalization (Bernado *et al.*, 2004; Ellis, 2001). Crowding may have the effect of favoring assembly or, alternatively, may impede assembly by separating reactants.

Modeling assembly of filaments and crystals

A good starting point for discussing capsid formation is to consider the simpler problem of crystallization (Fig. 4). Crystals, sheets, and filaments are all examples of open-ended polymers. The dominant feature of kinetics for these polymers is initiation of assembly, generally a nucleation event. Once assembly begins, it utilizes faster and/or more stable reactions to proceed until equilibrium between bound and free subunit is achieved (Frieden, 1985; Oosawa and Asakura, 1975). Because subunits can freely dissociate from or add to the periphery of these open-ended polymers, the polymers behave as a different phase with a critical concentration of free subunit in solution (Tanford, 1980).

Protein polymerization can proceed by a series of identical reactions so that all subunits and intersubunit contacts are equal independent of polymer size, an isodesmic reaction. Or, there may be a nucleation step where the initial reactions are slower, weaker or otherwise regulated (Frieden, 1989) to be distinct from subsequent elongation reactions. In such a polymer, the growing 'ends' are distinct from internal subunits. For simplicity, let us consider assembly of a filament beginning with nucleation. Formation of nucleus of n subunits will be treated as an nth-order reaction [eqn (2)]. The reality is that there are many different mechanisms of slow nucleation (consider actin; Matsudaira *et al.*, 1987). Even when thermodynamically favored, by definition nucleation will be slow at low concentrations. Crystallographers take advantage of this phenomenon to generate super-saturated protein solutions that result in only one or a few nuclei.

$$\mathrm{d}[\text{nucleus}]/\mathrm{d}t = k_{\text{nucleation}}[\text{subunit}]^n - k_{\text{dissoc}}[\text{nucleus}] \quad (2)$$

Equation (2) indicates that nucleation will depend on the nth power of subunit concentration and assumes that dissociation

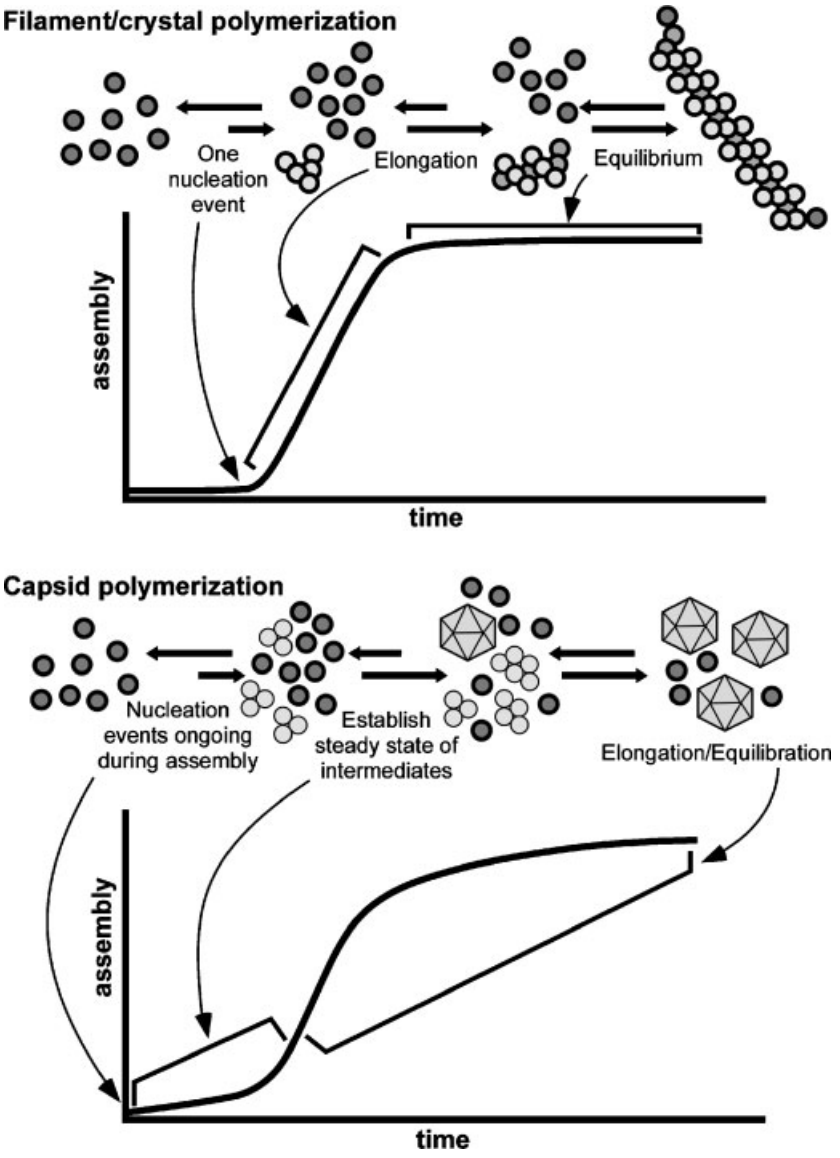

Figure 4. Kinetics of polymerization for infinite polymers (top panel) and capsids (bottom). For infinite polymers (filament, sheet or crystal), association may begin with single nucleus and lead to formation of a huge polymer; the lag phase is the time required to form nuclei. For a polymer of defined size (capsid), one nucleus is required for each capsid. Nuclei form continuously during the reaction. The lag phase is the time required to form a steady state of intermediates. In the kinetic trajectories, assembly is most easily considered in terms of average molecular weight. Note that there is no assembly during the lag phase for infinite polymer formation, whereas there is assembly (of intermediates) during the lag phase of the capsid assembly reaction. In both types of reaction, nucleation by kinetically and/or thermodynamically disfavored.

of the complex is a first order reaction. The nucleus is a unique starting point, but subunits adding to the growing polymer are equivalent. Once assembly begins, it proceeds with a simple rate equation [eqn (3)]. In this case, subunits add to the free 'ends' one at a time. For conceptual ease, the rate equation is in terms of 'ends' because all open sites are considered as equivalent, no matter the size of the polymer.

$$\mathrm{d[ends]}/\mathrm{d}t = k_{\mathrm{assoc}}[\mathrm{ends}][\mathrm{subunit}] - k_{\mathrm{dissoc}}[\mathrm{ends}] \qquad (3)$$

When polymerization reaches equilibrium, the rate of growth, d[ends]/dt, is 0, i.e. the association and dissociation rates are equal. Under conditions of equilibrium, we can see that the dissociation expression [eqn (4), rearranged from eqn (3)] indicates that a *critical concentration* of free subunit will remain in solution.

$$[\mathrm{ends}][\mathrm{subunit}]/[\mathrm{ends}] = k_{\mathrm{dissoc}}/k_{\mathrm{assoc}} \qquad (4)$$

Open-ended polymers will follow variations of this theme. Polymers with a defined number of subunits, such as capsids, will display similar behavior but for different kinetic and thermodynamic reasons. At equilibrium, a graph of subunit concentration vs assembly will show a breakpoint at the critical concentration. The breakpoint will not be

perfect, due in part to the difficulty in initiating and propagating assembly at low subunit concentrations, a kinetic effect.

The kinetics of assembly will be sigmoidal. Because nuclei form slowly, there will be a lag before any assembly is observed. Once one or more nuclei form, further assembly is fast and usually limited by the concentration of nuclei. At high initial concentrations, many nuclei will form, resulting in little or no lag and many polymers, e.g. a shower of crystals.

Modeling Assembly of large populations of capsids

At a superficial level, capsid assembly resembles crystal assembly (Fig. 4). Both sorts of polymers comprise (mainly) identical subunits in (mainly) identical environments. Both sorts of polymers show sigmoidal assembly kinetics. When both sorts of reactions reach equilibrium, the polymers show what appears to be a critical concentration for assembly.

The basis for capsid assembly kinetics and thermodynamics can be demonstrated by simulations (Endres and Zlotnick, 2002; Zlotnick, 1994; Zlotnick *et al.*, 1999).

$$\text{subunits} \Longleftrightarrow \text{nuclei} + \text{subunits} \Longleftrightarrow \text{capsids} \qquad (5)$$

Consider eqn (5); nuclei are presumed to contain a few subunits and assemble slowly compared with elongation. Alternatively, nucleating reactions may involve weaker association than subsequent reactions or require an exogenous effector (Beckett *et al.*, 1988; Choi *et al.*, 2002; Bartenschlager and Schaller, 1992; Stray *et al.*, 2004). In our simulations, nucleation rate constants (k_{nuc}) are typically at least an order of magnitude slower than elongation (k_{elong}). All association reactions are considered as second order with one subunit associating at a time to the growing polymer.

Sigmoidal kinetics arise from the many steps required to complete a capsid. Because simulations (and *in vitro* reactions) start with free subunits, nucleation rate (d[nuclei]/dt, not k_{nuc}) is fastest at time 0. This does not translate into the immediate appearance of many capsids because elongation reactions take time! However, as nuclei accumulate they begin to be consumed in the production of the $nuc+1$ intermediate. It is clear in equation 6 that the $nuc+2$ intermediate cannot accumulate until there is an appreciable concentration of $nuc+1$, where 'an appreciable concentration' may be sub-nanomolar. At very low $nuc+1$, the rate of $nuc+2$ formation will be inconsequential, no matter how high k_{elong} and/or [subunit].

$$\mathrm{d}[nuc+2]/\mathrm{d}t = k_{\text{elong1}}[nuc+1][\text{subunit}] + (\text{other terms}) \qquad (6)$$

As $nuc+2$ accumulates, it in turn will become fuel for formation of $nuc+3$. The concentration of $nuc+2$ does not fall to zero, it approaches a steady state. This means that eqn (6) must be modified to account for these subsequent reactions.

$$\mathrm{d}[nuc+2]/\mathrm{d}t = k_{\text{elong1}}[nuc+1][\text{subunit}] - k_{\text{elong2}}[nuc+2][\text{subunit}] + (\text{other terms}) \qquad (7)$$

The 'other terms' in eqn (7) describe dissociation reaction, which are required for simulation of a reversible reaction. The elongation rates for different intermediates (k_{elong1}, k_{elong2}, etc.) must reflect the microscopic forward rate and the degeneracy of the reaction. The degeneracy term accounts for the symmetry of the reactants and products (Zlotnick, 1994), analogous to binding to multiple identical sites (Cantor and Schimmel, 1980).

Each intermediate accumulates in turn until the next intermediate can be efficiently produced. Unlike crystallization, the presence of a lag phase in capsid formation has little to do with the process of nucleation and everything to do with forming a steady state of intermediates. Once the steady-state pipeline of intermediates is generated, then the formation of each new nucleus will correlate with the formation of a capsid.

Simulations indicate that assembly reactions rapidly approach equilibrium (Endres and Zlotnick, 2002). However, unlike the crystallization reaction, the equilibrium does not reflect the ability of subunits to freely diffuse between phases, i.e. a critical concentration. The forward reaction slows to a standstill as [capsid] asymptotically approaches equilibrium. For capsid assembly, the equilibrium depends on stability of the capsid, composed of N subunits:

$$K_{\text{capsid}} = [\text{capsid}]/[\text{subunit}]^N \qquad (8)$$

Equation 8 demonstrates the steep Nth power concentration dependence of assembly, which can be mistaken for a critical concentration. Unlike a critical concentration, there is no maximum subunit concentration, although a small increase in [subunit] can translate to a large increase in [capsid]. The association constant, K_{capsid}, is a difficult term to work with. Because the value of N may be in the hundreds, K_{capsid} may be absurdly large. In practice, we use $\log(K_{\text{capsid}})$ for calculations.

However, K_{capsid} can be partitioned into its components [eqn (9)], which include the association constant per subunit–subunit contact, K_{contact}, and a relatively small term for the product of all degeneracy terms ($\boldsymbol{s}$) involved in the association reaction, $\pi\boldsymbol{S}$ (Ceres and Zlotnick, 2002; Zlotnick, 2003). The exponential term, $Nc/2$, is the number of intersubunit contacts forming the capsid, where N is the number of subunits and c is the number of contact surfaces per subunit. To simplify examination of real data, the term $K_{\text{Dapparent}}$, which can be derived from K_{contact} or K_{capsid}, is defined as the concentration of where [subunit] = [capsid] [eqn (10)] (Ceres and Zlotnick, 2002; Zlotnick, 1994, 2003).

$$K_{\text{capsid}} = (\pi\boldsymbol{S}K_{\text{contact}}^{Nc/2}) \qquad (9)$$

$$\pi\boldsymbol{S} = \boldsymbol{s}^{N-1}/\boldsymbol{N} \qquad (10)$$

$$K_{\text{Dapparent}} = K_{\text{capsid}}^{1/1-N} = \pi\boldsymbol{S}^{1/1-N}K_{\text{contact}}^{(Nc/2)/(1-N)} \qquad (11)$$

Equation (9) is scalable to all intermediates, using an appropriate $\pi\boldsymbol{S}$. For model calculations we score one unit of association energy per subunit–subunit contact. The association constant for the nth intermediate and the corresponding forward rate, $k_{\text{elong},n}$ (the elongation rate for the nth

Table 1. Methods for extracting kinetic and thermodynamic parameters

1. Nucleus size	d[capsid]/d[CP] at one time: slope of a graph of In[capsid] vs In[CP] for several initial CP concentrations, measured at the same time while the rate is at its fastest
2. Nucleation rate	Proportional to (d[capsid]/dt)/[nucleus-1 measured when the reaction is in the plateau phase
3. $\Delta G_{contact}$	Measured when the reaction has equilibrated, see eqns (8), (9) and (11)
4. Elongation rate	A value for k_{elong} must be estimated from curve fitting

The numbering refers to Fig. 3. The analyses are developed in Endres and Zlotnick (2002).

reaction), allows calculation of the dissociation rate from $K_{A,n} = k_{elong,n}/k_{dissociation,n}$.

The preceding equations allow the simulation of assembly reactions based on fundamental thermodynamic and kinetic considerations with a minimal number of assumptions. Based on this mathematical construction, analyses were developed to extract important kinetic and thermodynamic parameters (Table 1, Fig. 5). In order to judge the validity of the models for a particular system, it is important to understand what assumptions are made and what behavior the models predict.

The starting assumptions built into the models are reductionist. All subunit–subunit contacts are treated as equivalent at all times during the reaction. All microscopic forward rates are also considered equivalent, independent of the size of the reactants, excepting nucleation rates which are imposed on the first two, three or five reactions in selected simulations. All reactions are restricted to a limited number of pathways. Restricting the number of paths for an assembly reaction greatly simplifies and speeds up calculations, but it is a simplification that merits careful attention (section 3.5).

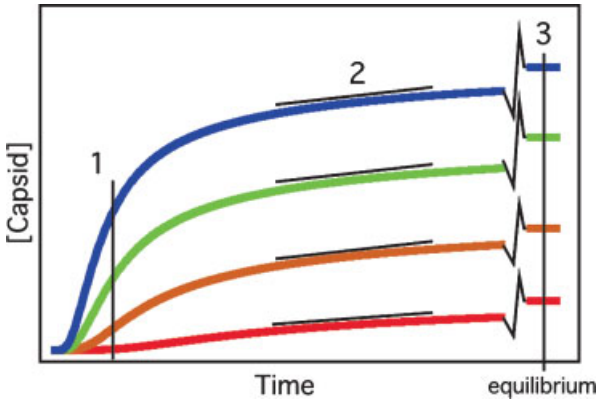

Figure 5. A simulation of a concentration series. As initial [capsid protein] increases, the lag becomes shorter, the slope steeper and the proportion of protein that forms capsid greater. The numbers refer to regions of the kinetic series that allow determination of (1) nucleus size, (2) nucleation rate and (3) association energy (see Table 1).

Prediction 1: nucleation minimizes kinetic traps. Given enough time, all reactions should come to equilibrium. However, on a biological time scale, many reactions become kinetically trapped far from equilibrium. In assembly reactions, most kinetic traps occur when there are too many starting points for the reaction and insufficient free subunits to complete them. This trap results in the accumulation of many intermediates and few capsids. In short, an incomplete picture is drawn from eqn (8), which only describes equilibrium.

Consider an assembly reaction without a nucleation step (Zlotnick *et al.*, 1999). Such a reaction can work under limiting conditions of low subunit concentration and weak association energy. Under the limiting conditions, these early intermediates are relatively rare. Yet equilibrium-limited reactions are not robust. If the limiting conditions are not met, many intermediates—and a trap—will result. A means of extending the concentration range suitable for assembly is controlling the availability of subunits by autostery (Caspar, 1980), where free subunits are generally in an assembly-incompetent conformation either in equilibrium with the competent form or induced into a competent conformation by contact with the growing virus.

With nucleation limiting the initial steps of the reaction, assembly is successful under a much broader range of conditions. When the nucleation rate sufficiently limits the reaction, only the most kinetically accessible intermediates will enter the steady state 'pipeline'. Equation (6) will have to be modified somewhat if j different intermediates of a given size are accessible and can form one of k possible subsequent products [eqn (12)]:

$$d[nuc+2]_k/dt = \Sigma k_{elong1j}[nuc+1]_j[subunit] + (other\ terms) \quad (12)$$

The rate of each of the j reactions will be slower because the concentration $[nuc+1]_j$ is smaller than the sum. However, the sum of these individual reactions will equal the rate for the single pathway. The multiplicity of paths will be important for accurately describing kinetic traps and peculiarities in the early time points of the reaction (Endres *et al.*, 2005).

Prediction 2: weak association energy is sufficient for formation of stable capsids. (Zlotnick, 1994, 2003). For reversible reactions in biology, we are used to thinking of dissociation constants in a nanomolar to micromolar range. However, that is for a single interaction. To form a capsid, each subunit must interact with at least three other subunits, i.e. they must be at least trivalent. Assuming additivity of association energy, this means that a small per contact energy can result in a substantial stability for a subunit bound in a capsid. Both the per contact energy and the number of contacts per subunit affect $K_{Dapparent}$ [eqn (11)]. For simulations based on micromolar starting concentrations of subunit, a millimolar $K_{contact}$ is sufficient for appreciable assembly. A weak $K_{contact}$ has the additional advantage that capsid assembly and stability are more susceptible to regulation.

Prediction 3: dissociation of capsids will display a marked hysteresis. Virus capsid dissociation is known as an example of hysteresis (Weber *et al.*, 1996). Typical

dissociation experiments induced by high pressure or denaturants resemble the highly cooperative transitions observed with protein unfolding. The same simulations that were developed to mimic assembly are able to recap features of hysteresis, including weak concentration dependence and a greater than expected stability (Singh and Zlotnick, 2003). The basis of hysteresis in these simulations is that the early intermediates in the dissociation reaction, capsid lacking one or two subunits, are stable and incorporate a high affinity (multivalent) binding site so that they are more likely to reassemble than continue dissociating. This creates a kinetic barrier to dissociation that cannot be overcome except if there is a catastrophic decrease in $K_{contact}$ or a biological mechanism exists for removing subunits from bulk solution. An implication of predictions 2 and 3 is that capsids may be metastable for viruses where uncoating is required to release the genome.

An alternative explanation for hysteresis is that capsid assembly is irreversible (Weber *et al.*, 1996). This hypothesis was based on the observed irreversible dissociation of a number of protein oligomers, including virus capsids, under high pressure with and without denaturants. The breadth of the dissociation transition was attributed to molecular heterogeneity of capsids. Irreversibility of capsid dissociation cannot be ruled out, especially in cases where there are post-assembly transitions or 'decorative' proteins in a virus that does not normally dissociate to release its genome (as in bacteriophages).

Assembly of individual particles

Perhaps the most accessible molecular dynamic simulation on the web is one of virus assembly (www.ph.biu.ac.il/~rapaport/images/capsid_anim.gif). In this simulation (Rapaport *et al.*, 1999), geometric subunits assemble into about a dozen $T = 1$ icosahedra, where each subunit has the trapezoidal shape common to capsids based on beta barrel architecture (Rossmann and Johnson, 1989). In order to allow productive assembly reactions, the number of nucleation sites was restricted. These studies show that the forward rate will only decrease slightly as a particle approaches completion, as suggested by the statistical terms in the thermodynamic–kinetic simulations.

A molecular dynamics approach was also taken to investigate regulation of quasi-equivalence (Fig. 6). Using trivalent subunits that could incorporate slight variations in their geometry, a system of 'local rules' was developed (Berger *et al.*, 1994). Local rules is based on the observation that there are distinct differences between quasi-equivalent environments in a capsid. In its most complete form, 'local rules' postulate that free subunits also adopt corresponding conformations with a high-energy barrier preventing fast exchange between conformations. In this scheme, an A subunit binds tightly to an A site. An A subunit binds poorly to non-A sites and can actually act as a competitive inhibitor. Local rules are redundant for some quasi-equivalent positions—only four rules were required to generate a $T = 7$ particle. This complete form of 'local rules' directs quasi-equivalence and acts a governor on the assembly reaction.

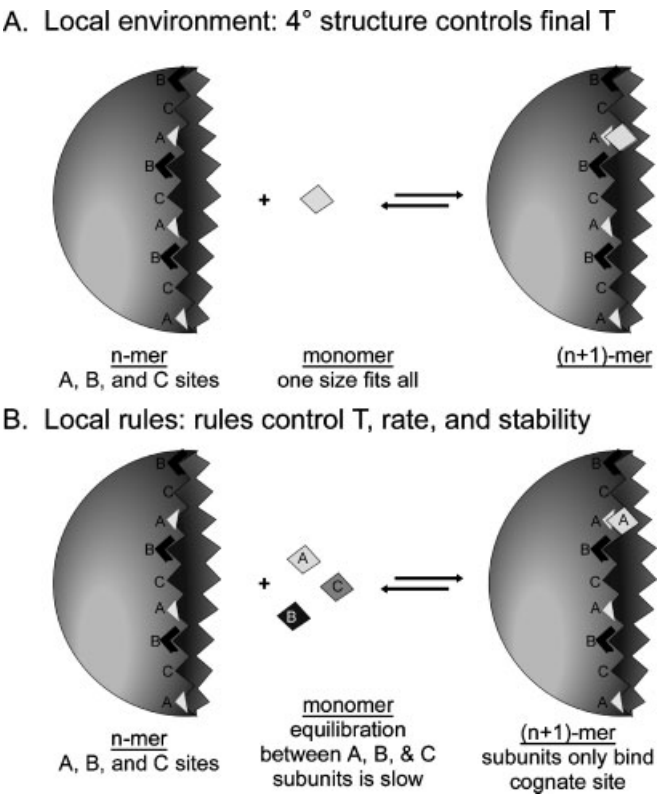

Figure 6. Local environments and local rules. A growing capsid will create new sites on its leading edge each of which generates a local environment. For small T numbers, local environment may be sufficient. For larger T numbers, tape measure proteins (Abrescia *et al.*, 2004) or local rules (Berger *et al.*, 1994) may be required. The local rules model constrains the structure of the available sites for addition of subunits and the structure of the subunits themselves. Because of competition between correct and incorrect subunits, local rules can provide regulation of assembly.

Local rules simulations followed the assembly of a single capsid at a time. A surprising level of flexibility in individual subunits was possible while still allowing correct assembly. When 'local rules' were relaxed (e.g. an A subunit bound to a non-A site) assembly of 'monsters' occurred, leading to the hypothesis that small molecules, antivirals, could be developed that would alter capsid polymerization (Prevelige, 1998). A more relaxed form of local rules eliminates the structural differences between free subunits with a concomitant loss of regulation of assembly (Schwartz *et al.*, 1998).

What makes local rules particularly attractive is that, like the thermodynamic–kinetic model, it suggests testable predictions for the behavior and structure of capsid proteins in solution. The more rigorous exposition of 'local rules' requires a discernable structural difference between different classes of free subunits. Differences must also be apparent in the structures of subunits when assembled. Finally, the different classes of binding site must also be distinct. Given these constraints, we can begin to identify which viruses are most likely to make use of 'local rules' regulation.

How general are the predictions of the local rules model? The prediction of distinct structural differences for assembled proteins is observed in most virus structures, but

with some important caveats. Notable examples of such differences are seen in bacteriophages PhiX174 and $\alpha 3$, where the external scaffold protein adopts distinct structures when bound to distinctly different sites on the provirus (Bernal *et al.*, 2003; Dokland *et al.*, 1997). One can easily envision this structural difference persisting in the unassembled protein, consistent with the first prediction of the local rule hypothesis. Distinct structural differences are also observed in the subunits of dsRNA viruses with pseudo-$T=2$ lattices (i.e. a $T=1$ lattice with two chemically identical subunits per icosahedral asu; Caston *et al.*, 1997; Grimes *et al.*, 1998; Hill *et al.*, 1999). However, in many cases the structural differences between subunits are limited to loops and N- and C-terminal structures that are probably disordered in solution. Perhaps the clearest example of this behavior is seen in the $T=7$ papovaviruses, where the C-terminus of each protein is in a distinct extended structure that weaves the capsid together but is almost certainly unfolded prior to assembly (Liddington *et al.*, 1991; Stehle *et al.*, 1994).

One assembly path or many paths

Examination of assembly by describing a single path is computationally straightforward but obviously artificial. Describing assembly in coarse-grained simulations allows examination of as many paths as one runs simulations. For even a small virus, there are too many possible intermediates to allow a complete description of assembly with a single path. In this regard, virus assembly is a special case of the Levinthal paradox (Levinthal, 1969) and should be describable in terms of a statistical landscape (Brooks *et al.*, 2001). A number of approaches have been taken to identify the important intermediates in an assembly reaction.

Build-up models (Wales, 1987) are one approach to identifying intermediates. Reddy *et al.* (1998) combined this approach with estimations of the free energy of interaction at protein–protein interfaces in several viruses to identify the most stable interactions; using this library of interactions it was possible to build a ladder of the most stable intermediates culminating in complete capsids. This approach successfully identified known stable intermediates for picornavirus assembly (corresponding to the 12S pentamer) and suggested stable intermediates in the assembly of nodaviruses (corresponding to the quasi-three-fold trimer). A most stable intermediate path could then be assembled by considering a combinatorial build-up strategy. Graph-theory was also applied to this problem (Sitharam and Agbandje-McKenna, 2005), focusing on local interactions and formation of sub-assemblies as critical components in forming the larger particle.

However, the most stable intermediates may not always be the most common intermediates (Endres *et al.*, 2005). Kinetic simulations based on complete arrays of intermediates for assembly of a dodecahedron from 12 pentagonal subunits indicated that the intermediates that assembled fastest were most common at early times in the reaction. At later times, the reaction tended to steady state so that the most stable species were dominant. The kinetically favored transient species had a dramatic effect on assembly kinetics.

By considering the potential (not free) energy surface, it is possible to identify the important general characteristics of assembly intermediates and assembly paths (Wales, 2005). One notable prediction of this approach is the observation that five-fold vertices must be 'spikey' in agreement with the free energy minima identified in Monte-Carlo simulations (Zandi *et al.*, 2004) and suggestive of experimental observation of distorted five-folds (e.g. bacteriophage HK97 maturation; Conway *et al.*, 2001). A second important point is that assembly energy surfaces must have a steep funnel shape in order to yield successful assembly, this is in agreement with the energy surface generated from a successful assembly simulation (Endres *et al.*, 2005). In both predictions, the same result was achieved by distinctly different approaches.

CLASH OF REALITIES: COMPARING CALCULATIONS TO BIOLOGICAL VIRUSES

The structures of viruses, VIPER

The object of any theoretical approach is to mirror and then predict experimental results. A central repository of virus structures and structural information has been compiled.

The structure of a virus capsid is the end point for assembly and a clue to the important features of the reaction. At this time, tens of capsid structures have been solved and are available at the Protein Data Bank (www.rcsb.org/pdb/) or, more conveniently, the virus particle explorer, VIPER (http://mmtsb.scripps.edu/viper/; Reddy *et al.*, 2001). VIPER has functions to generate virus fragments and a database describing features of the structures, such as protein–protein interactions and buried surface area.

The stability of viruses

A likely explanation for a disagreement between theory and experimental result is that the theory has neglected some feature of the reaction it describes. Frequently these are opportunities for discovery. One such opportunity is described below.

One of the analyses included in VIPER is an estimation of subunit–subunit association energies, analogous to K_{contact}. This energy is estimated from buried polar and non-polar surface area (Eisenberg and McLachlan, 1986). There are a number of similar methods (Ayala *et al.*, 1995; Baker and Murphy, 1998; Horton and Lewis, 1992). These methods work reasonably well at estimating the association energy of small complexes. All of them include the assumption that when considered over an area, interprotein contacts have evolved to the same quality of interaction. Consequently, these methods fail when a non-typical interaction exists, or a mutation disturbs the average environment, OR when there are thermodynamic states, other than those considered in the calculation.

The predicted values for intersubunit contact energies for viruses are uniformly extremely high. For example, the intersubunit contact energy for HBV is estimated to be on

the order of −20 to −30 kcal/mol (other algorithms give similar values), which is dramatically different from the experimentally measured value of −3 to −4 kcal/mol (Ceres and Zlotnick, 2002). Similarly, the average interaction energy between dimers in CCMV is about −3 kcal/mol (Johnson *et al.*, 2005) compared with the extraordinary −30 to −50 kcal/mol association energies calculated from structure. For comparison the association energy of biotin for avidin is on the order of −20 kcal/mol, one of the strongest non-covalent bonds.

I suggest that there is a structural transition between the assembly unit in solution and the same protein in the context of a capsid. This would allow capsid protein to accumulate in solution prior to assembly yet form highly ordered structures due to complementarity between subunits. This hypothesis resembles Caspar's autostery (Caspar, 1980), except that autostery was primarily supposed to control kinetics of assembly. In support of this hypothesis, we know that the HBV capsid protein undergoes a conformational transition from an assembly-inactive state prior to assembly (Stray *et al.*, 2004, 2005); thus, the HBV calculations are incomplete because they did not include the energetic contribution from that transition.

CONCLUSION

Understanding of the basis of capsid assembly and stability is still in its infancy. The simple models of assembly now available are only a starting point for understanding assembly reactions *in vitro* and *in vivo*. We also need more detailed descriptions of reactants and products. Complicating any analysis is the indication that capsid proteins may undergo conformational changes concomitant with assembly. Furthermore, we now know that we cannot treat capsids as rigid structures (Phelps and Post, 1995, 1999; Speelman *et al.*, 2001). Although we have no complete stories, by comparing experimental systems and increasingly complex models we will gain a more complete picture of the complexity of assembly.

Acknowledgments

I wish to thank Dr Stephen Stray, Dr Christina Bourne, Dr Santanu Mukherjee, Pablo Ceres, and Jennifer Johnson for their comments and discussion. The Zlotnick laboratory is supported by grants from the American Cancer Society (RSG-99-339-04-MBC) and the NIH (5R01-EB000432).

REFERENCES

Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martin C, Burnett RM, Stuart DI, Bamford DH, Bamford JK. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* **432**: 68–74.

Ayala YM, Vindigni A, Nayal M, Spolar RS, Record MT, Jr, Di Cera E. 1995. Thermodynamic investigation of hirudin binding to the slow and fast forms of thrombin: evidence for folding transitions in the inhibitor and protease coupled to binding. *J. Mol. Biol.* **253**: 787–798.

Baker BM, Murphy KP. 1998. Prediction of binding energetics from structure using empirical parameterization. *Meth. Enzymol.* **295**: 294–315.

Bartenschlager R, Schaller H. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* **11**: 3413–3420.

Beckett D, Wu HN, Uhlenbeck OC. 1988. Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* **204**: 939–947.

Belnap DM, Olson NH, Cladel NM, Newcomb WW, Brown JC, Kreider JW, Christensen ND, Baker TS. 1996. Conserved features in papillomavirus and polyomavirus capsids. *J. Mol. Biol.* **259**: 249–263.

Benjamin J, Ganser-Pornillos BK, Tivol WF, Sundquist WI, Jensen GJ. 2005. Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography. *J. Mol. Biol.* **346**: 577–588.

Berger B, Shor PW, Tucker-Kellogg L, King J. 1994. Local rule-based theory of virus shell assembly. *Proc. Natl. Acad. Sci. USA* **91**: 7732–7736.

Bernado P, Garcia de la Torre J, Pons M. 2004. Macromolecular crowding in biological systems: hydrodynamics and NMR methods. *J. Mol. Recognit.* **17**: 397–407.

Bernal RA, Hafenstein S, Olson NH, Bowman VD, Chipman PR, Baker TS, Fane BA, Rossmann MG. 2003. Structural studies of bacteriophage alpha3 assembly. *J. Mol. Biol.* **325**: 11–24.

Black LW, Showe MK. 1983. Morphogenesis of the T4 head. In *Bacteriophage T4*, Mathews CK, Kutter EM, Mosig G, Berget PB (eds). American Society for Microbiology: Washington, DC; 219–245.

Booy FP, Newcomb WW, Brown JC, Steven AC. 1988. Herpesvirus nucleocapsids visualized in the frozen-hydrated state. In *Proceedings of the 46th Annual Meeting of the Electron Microscopy Society of America*. San Francisco Press: San Francisco, CA; 164–165.

Bowden NB, Weck M, Choi IS, Whitesides GM. 2001. Molecule-mimetic chemistry and mesoscale self-assembly. *Acc. Chem. Res.* **34**: 231–238.

Brooks CL III, Onuchic JN, Wales DJ. 2001. Statistical thermodynamics. Taking a walk on a landscape. *Science* **293**: 612–613.

Bruinsma RF, Gelbart WM, Reguera D, Rudnick J, Zandi R. 2003. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* **90**: 248101.

Cantor CR, Schimmel PR. 1980. *Biophysical Chemistry*. WH Freeman: San Francisco, CA.

Caspar DL. 1980. Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys. J.* **32**: 103–138.

Caspar DLD. 1956. Structure of tomato bushy stunt virus. *Nature* **177**: 476–477.

Caspar DLD, Klug A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**: 1–24.

Caston JR, Trus BL, Booy FP, Wickner RB, Wall JS, Steven AC. 1997. Structure of L-A virus: a specialized compartment for the transcription and replication of double-stranded RNA. *J. Cell Biol.* **138**: 975–985.

Ceres P, Zlotnick A. 2002. Weak protein–protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**: 11525–11531.

Cerritelli ME, Studier FW. 1996. Assembly of T7 capsids from independently expressed and purified head protein and scaffolding protein. *J. Mol. Biol.* **258**: 286–298.

Cheng RH, Caston JR, Wang GJ, Gu F, Smith TJ, Baker TS, Bozarth RF, Trus BL, Cheng N, Wickner RB. 1994. Fungal

virus capsids, cytoplasmic compartments for the replication of double-stranded RNA, formed as icosahedral shells of asymmetric Gag dimers. *J. Mol. Biol.* **244**: 255–258.
Choi YG, Dreher TW, Rao AL. 2002. tRNA elements mediate the assembly of an icosahedral RNA virus. *Proc. Natl Acad. Sci. USA* **99**: 655–660.
Conway JF, Wikoff WR, Cheng N, Duda RL, Hendrix RW, Johnson JE, Steven AC. 2001. Virus maturation involving large subunit rotations and local refolding. *Science* **292**: 744–748.
Crick FHC, Watson JD. 1956. The structure of small viruses. *Nature* **177**: 473–475.
Damodaran KV, Reddy VS, Johnson JE, Brooks CL III, 2002. A general method to quantify quasi-equivalence in icosahedral viruses. *J. Mol. Biol.* **324**: 723–737.
Darnell J, Lodish H, Baltimore D. 1990. *Molecular Cell Biology*, 2nd edn. Scientific American Books: New York.
Dawkins R. 1989. *The Selfish Gene*, 2nd edn. Oxford University Press: Oxford.
Dokland T, McKenna R, Ilag LL, Bowman BR, Incardona NL, Fane BA, Rossmann MG. 1997. Structure of a viral procapsid with molecular scaffolding. *Nature* **389**: 308–313.
Duan HW, Kuang M, Wang J, Chen DY, Jiang M. 2004. Self-assembly of rigid and coil polymers into hollow spheres in their common solvent. *J. Phys. Chem. B* **108**: 550–555.
Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature* **319**: 199–203.
Ellis RJ. 2001. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **11**: 114–119.
Endres D, Zlotnick A. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**: 1217–1230.
Endres D, Miyahara M, Moisant P, Zlotnick A. 2005. A reaction landscape for identifying important intermediates in the self-assembly of virus capsids and other polyhedra. *Protein Sci.* **14**: 1518–1525.
Falkner JC, Turner ME, Bosworth JK, Trentler TJ, Johnson JE, Lin T, Colvin VL. 2005. Virus crystals as nanocomposite scaffolds. *J. Am. Chem. Soc.* **127**: 5274–5275.
Frieden C. 1985. Actin and tubulin polymerization: the use of kinetic methods to determine mechanism. *A. Rev. Biophys. Biophys. Chem.* **14**: 189–210.
Frieden C. 1989. The regulation of protein polymerization. *Trends Biochem. Sci.* **14**: 283–286.
Fuller SD, Wilk T, Gowen BE, Krausslich HG, Vogt VM. 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr. Biol.* **7**: 729–738.
Grimes JM, Jakana J, Ghosh M, Basak AK, Roy P, Chiu W, Stuart DI, Prasad BV. 1997. An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. *Structure* **5**: 885–893.
Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Zientara S, Mertens PP, Stuart DI. 1998. The atomic structure of the bluetongue virus core. *Nature* **395**: 470–478.
Hendrix RW, Garcea RL. 1994. Capsid assembly of dsDNA viruses. *Sem. Virol.* **5**: 15–26.
Hill CL, Booth TF, Prasad BV, Grimes JM, Mertens PP, Sutton GC, Stuart DI. 1999. The structure of a cypovirus and the functional organization of dsRNA viruses. *Nat. Struct. Biol.* **6**: 565–568.
Horton N, Lewis M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**: 169–181.
Johnson JM, Tang J, Nyame Y, Willits D, Young MJ, Zlotnick A. 2005. Regulating self-assembly of spherical oligomers. *Nano Lett.* **5**: 765–770.
Krol MA, Olson NH, Tate J, Johnson JE, Baker TS, Ahlquist P. 1999. RNA-controlled polymorphism in the *in vivo* assembly of 180-subunit and 120-subunit virions from a single capsid protein. *Proc. Natl Acad. Sci. USA* **96**: 13650–13655.
Levinthal C. 1969. *Mossbauer Spectroscopy in Biological Systems*. Allerton House: Monticello, IL.
Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC. 1991. Structure of simian virus 40 at 3.8-Å resolution. *Nature* **354**: 278–284.
Marzec CJ, Day LA. 1993. Pattern formation in icosahedral virus capsids: the papova viruses and Nudaurelia capensis beta virus. *Biophys. J.* **65**: 2559–2577.
Matsudaira P, Bordas J, Koch MH. 1987. Synchrotron x-ray diffraction studies of actin structure during polymerization. *Proc. Natl Acad. Sci. USA* **84**: 3151–3155.
Naitow H, Tang J, Canady M, Wickner RB, Johnson JE. 2002. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nat. Struct. Biol.* **9**: 725–728.
Nermut MV, Hockley DJ, Bron P, Thomas D, Zhang WH, Jones IM. 1998. Further evidence for hexagonal organization of HIV gag protein in prebudding assemblies and immature virus-like particles. *J. Struct. Biol.* **123**: 143–149.
Oosawa F, Asakura S. 1975. *Thermodynamics of Polymerization of Protein*. Academic Press: London.
Phelps DK, Post CB. 1995. A novel basis of capsid stabilization by antiviral compounds. *J. Mol. Biol.* **254**: 544–551.
Phelps DK, Post CB. 1999. Molecular dynamics investigation of the effect of an antiviral compound on human rhinovirus. *Protein Sci.* **8**: 2281–2289.
Prevelige PEJ. 1998. Inhibiting virus-capsid assembly by altering the polymerisation pathway. *Trends Biotechnol.* **16**: 61–65.
Rapaport DC, Johnson JE, Skolnick J. 1999. Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation. *Comp. Phys. Commun.* **121**: 231–235.
Rayment I, Baker TS, Caspar DL, Murakami WT. 1982. Polyoma virus capsid structure at 22.5 Å resolution. *Nature* **295**: 110–115.
Reddy VS, Giesing HA, Morton RT, Kumar A, Post CB, Brooks CL III, Johnson JE. 1998. Energetics of quasiequivalence: computational analysis of protein–protein interactions in icosahedral viruses. *Biophys. J.* **74**: 546–558.
Reddy VS, Natarajan P, Okerberg B, Li K, Damodaran KV, Morton RT, Brooks CL III, Johnson JE. 2001. Virus Particle Explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.* **75**: 11943–11947.
Ross PD, Subramanian S. 1981. Thermodynamics of protein association reactions: forces contributing to stability. *Biochemistry* **20**: 3096–3102.
Rossmann MG, Blow DM. 1962. The detection of subunits within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**: 24–31.
Rossmann MG, Johnson JE. 1989. Icosahedral RNA virus structure. *A. Rev. Biochem.* **58**: 533–573.
Rueckert RR. 1996. Picornaviridae: the viruses and their replication. In *Fields Virology*, Fields BN, Knipe DM, Howley PM, Chanock RM, Melnick JL, Monath TP, Roizman B, Straus SE (eds). Lippincott-Raven: Philadelphia, PA; 609–654.
Salunke D, Caspar DL, Garcea RL. 1989. Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* **56**: 887–900.
Schwartz R, Shor PW, Prevelige PEJ, Berger B. 1998. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* **75**: 2626–2636.
Schweiger M, Yamamoto T, Stang PJ, Blaser D, Boese R. 2005. Self-assembly of nanoscale supramolecular truncated tetrahedra. *J. Org. Chem.* **70**: 4861–4864.
Seidel SR, Stang PJ. 2002. High-symmetry coordination cages via self-assembly. *Acc. Chem. Res.* **35**: 972–983.
Singh S, Zlotnick A. 2003. Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. *J. Biol. Chem.* **278**: 18249–18255.
Sitharam M, Agbandje-McKenna M. 2005. Modeling virus self-assembly pathways: avoiding dynamics using geometric constraint decomposition. *J. Comp. Biol.* in press.
Speelman B, Brooks BR, Post CB. 2001. Molecular dynamics simulations of human rhinovirus and an antiviral compound. *Biophys. J.* **80**: 121–129.
Speir JA, Munshi S, Wang G, Baker TS, Johnson JE. 1995. 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.

Stehle T, Harrison SC. 1996. Crystal structures of murine polyomavirus in complex with straight-chain and branched-chain sialyloligosaccharide receptor fragments. *Structure* **4**: 183–194.

Stehle T, Yan Y, Benjamin TL, Harrison SC. 1994. Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature* **369**: 160–163.

Stray SJ, Ceres P, Zlotnick A. 2004. Zinc ions trigger conformational change and oligomerization of hepatitis B virus capsid protein. *Biochemistry* **43**: 9989–9998.

Stray SJ, Bourne C, Punna S, Lewis WG, Finn MG, Zlotnick A. 2005. A Heteroaryldihydropyrimidine enhances and can misdirect assembly of hepatitis B virus capsid protein. *Proc. Natl Acad. Sci. USA* **102**: 8138–8143.

Tanford C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2nd edn. Wiley: New York.

Tang J, Johnson JM, Dryden K, Young MJ, Zlotnick A, Johnson JE. 2005. Genetic and structural analysis of subunit hinges and switches and their control of viral capsid polymorphism. *J. Struct. Biol.* in press.

Twarock R. 2004. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* **226**: 477–482.

Twarock R. 2005. Mathematical models for tubular structures in the family of Papovaviridae. *Bull. Math. Biol.* **67**: 973–987.

Wales DJ. 1987. Closed-shell structures and the building game. *Chem. Phys. Lett.* **141**: 478–484.

Wales DJ. 2005. The energy landscape as a unifying theme in molecular science. *Phil. Trans. R. Soc. A* **363**: 357–377.

Wang Q, Lin T, Tang L, Johnson JE, Finn MG. 2002. Icosahedral virus particles as addressable nanoscale building blocks. *Angew. Chem. Int. Ed. Engl.* **41**: 459–462.

Weber G, Da Poian AT, Silva JL. 1996. Concentration dependence of the subunit association of oligomers and viruses and the modification of the latter by urea binding. *Biophys. J.* **70**: 167–173.

Yan X, Olson NH, Van Etten JL, Bergoin M, Rossmann MG, Baker TS. 2000. Structure and assembly of large lipid-containing dsDNA viruses. *Nat. Struct. Biol.* **7**: 101–103.

Yeager M, Wilson-Kubalek EM, Weiner SG, Brown PO, Rein A. 1998. Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy: implications for retroviral assembly mechanisms. *Proc. Natl Acad. Sci. USA* **95**: 7299–7304.

Zandi R, Reguera D, Bruinsma RF, Gelbart WM, Rudnick J. 2004. Origin of icosahedral symmetry in viruses. *Proc. Natl Acad. Sci. USA* **101**: 15556–15560.

Zlotnick A. 1994. To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**: 59–67.

Zlotnick A. 2003. Are weak protein–protein interactions the general rule in capsid assembly? *Virology* **315**: 269–274.

Zlotnick A, Johnson JM, Wingfield PW, Stahl SJ, Endres D. 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**: 14644–14652.

Zlotnick A, Aldrich R, Johnson JM, Ceres P, Young MJ. 2000. Mechanism of capsid assembly for an icosahedral plant virus. *Virology* **277**: 450–456.

Zlotnick A, Ceres P, Singh S, Johnson JM. 2002. A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *J. Virol.* **76**: 4848–4854.

