

# Model-Based Analysis of Assembly Kinetics for Virus Capsids or Other Spherical Polymers

Dan Endres\* and Adam Zlotnick†

\*Department of Mathematics and Statistics, University of Central Oklahoma, Edmond, Oklahoma 73034; and †Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190 USA

**ABSTRACT** The assembly of virus capsids or other spherical polymers—empty, closed structures composed of hundreds of protein subunits—is poorly understood. Assembly of a closed spherical polymer is unlike polymerization of a filament or crystal, examples of open-ended polymers. This must be considered to develop physically meaningful analyses. We have developed a model of capsid assembly, based on a cascade of low-order reactions, that allows us to calculate kinetic simulations. The behavior of this model resembles assembly kinetics observed in solution (Zlotnick, A., J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres. 1999. *Biochemistry*. 38:14644–14652). We exhibit two examples of this general model describing assembly of dodecahedral and icosahedral capsids. Using simulations based on these examples, we demonstrate how to extract robust estimates of assembly parameters from accessible experimental data. These parameters, nucleus size, average nucleation rate, and average free energy of association can be determined from measurement of subunit and capsid as time and concentration vary. Mathematical derivations of the analyses, carried out for a general model, are provided in an Appendix. The understanding of capsid assembly developed in this paper is general; the examples provided can be readily modified to reflect different biological systems. This enhanced understanding of virus assembly will allow a more quantitative analysis of virus stability and biological or antiviral factors that affect assembly.

## INTRODUCTION

Virus capsid assembly is an example of protein polymerization. Most spherical virus capsids exhibit icosahedral quaternary structure. They are constructed from multiples of 60 monomeric or oligomeric structural subunits, arranged in equivalent or quasi-equivalent environments (Caspar and Klug, 1962). To be viable, virus capsids must assemble efficiently and with high fidelity. The specific mechanisms of capsid assembly are poorly understood for spherical viruses. There is little experimental information describing intersubunit binding energies, rates, and orders for assembly reactions, and formation of nucleating structures.

In vitro capsid assembly has been observed for many viruses. Cowpea chlorotic mottle virus (CCMV) provided the first example of in vitro assembly of a spherical virus (Bancroft et al., 1967). Both empty CCMV capsids and infectious RNA-filled virions have been assembled (Bancroft and Hiebert, 1967; Bancroft et al., 1968; 1969; Adolph and Butler, 1976; Zhao et al., 1995; Fox et al., 1998). Characteristics common to capsid assembly were first observed with CCMV, including extreme sensitivity to solution conditions, concentration dependence, and an apparent critical concentration (Adolph and Butler, 1976). Since 1967 many other systems have been studied. Polyoma virus assembly was pleiomorphic, resulting in the production of icosahedral  $T = 1$ , octahedral, and  $T = 7$  particles, as well

as sheets and other polymers (Salunke et al., 1989). Poliovirus assembly from oligomeric intermediates demonstrated the fragility of the provirion form of picornaviruses (Rom-baut et al., 1990; Basavappa et al., 1994; Rueckert, 1996). Assembly of related bacteriophages MS2 and R17 required a specific nucleic acid oligomer (Beckett et al., 1988; van den Worm et al., 1998). In vitro assembly of bacteriophage P22 demonstrated a relationship between scaffold and capsid proteins; these experiments also demonstrated sigmoidal capsid assembly kinetics for the first time (Prevelige et al., 1993). We present a framework linking experimental data to physically meaningful assembly parameters.

Previously, we have developed models based on thermodynamics and kinetics (Zlotnick, 1994; Zlotnick et al., 1999). Capsid accumulation exhibits the sigmoidal kinetics typical of multistep reactions (c.f. discussion of consecutive reactions in Fersht (1999)). During the initial lag phase, successive intermediates transiently accumulate and are consumed in turn by subsequent reactions. However, without regulation, the assembly reaction is susceptible to kinetic trapping (see Caspar (1980); Zlotnick (1994); Zlotnick et al. (1999, 2000)). Several regulating mechanisms have been proposed that decrease susceptibility to kinetic trapping. Autostery, in which free subunits equilibrate between assembly-competent and -incompetent states, has been proposed (Caspar, 1980); a form of it was used in the local rules model of quasi-equivalent assembly to decrease susceptibility to trapping (Berger et al., 1994; Schwartz et al., 1998). We have focused on regulation through nucleation by incorporating a “kinetically limiting” early step in the reaction, described as KL assembly (Zlotnick et al., 1999), wherein the first few intermediates either assemble slowly or are unstable. Nucleation is common with other biopoly-

Submitted December 20, 2001, and accepted for publication April 17, 2002.

Address reprint requests to Adam Zlotnick, P.O. Box 26901, BRC464, Oklahoma City, OK 73190. Tel.: 405-271-9030; Fax: 405-271-3910; E-mail: adam-zlotnick@ouhsc.edu.

© 2002 by the Biophysical Society

0006-3495/02/08/1217/14 \$2.00

**TABLE 1** Naming conventions for constants and variables

$N$	Number of subunits in a complete capsid
$[N]$	Capsid concentration
$[m]$	Concentration of the $m$ th species
$[u]$	Subunit concentration
$u_0$	Initial subunit concentration
$nuc$	Number of subunits composing the nucleating structure
$j$	Degeneracy (number of planar rotational symmetries) of an individual subunit
$c$	Number of contact sites on an individual subunit
$c_m$	Number of intersubunit contacts in the $m$ th intermediate
$C$	Number of intersubunit contacts in a complete capsid
$K_{Acap}$	Association constant for a complete capsid, in units of $M^{-N}$
$K_{Acon}$	Association constant for a single intersubunit contact
$f_{elong}$	Microscopic forward rate constant for elongation reactions
$f_{nuc}$	Microscopic forward rate constant for nucleation reactions
$s_m$	Statistical factor, degeneracy of the forward reaction leading to the $m$ th species
$\sigma_m$	Statistical factor, degeneracy of dissociation of the $m$ th species
$b_m$	Backward rate constant for dissociation of the $m$ th species
$\nu$	Ratio $f_{nuc}/f_{elong}$ of forward rate constants, for nucleated reactions $\nu \ll 1$
$[m]_\infty$	Concentration of the $m$ th species at $t = \infty$ , equilibrium
$K_{Dapp}$	Apparent dissociation constant, $[u]_\infty$ value for which $[N]_\infty = [u]_\infty$
$K_m$	Equilibrium constant for $[m]/([m-1][u])$

mers and has been observed in virus capsid assembly (Sorger et al., 1986; Beckett et al., 1988; Bartenschlager and Schaller, 1992), where a nucleating structure might be a completed vertex or a closed ring of subunits, much as the completion of the first turn nucleates a helical filament (Oosawa and Asakura, 1975). These and other regulatory mechanisms can be incorporated into the general models discussed here.

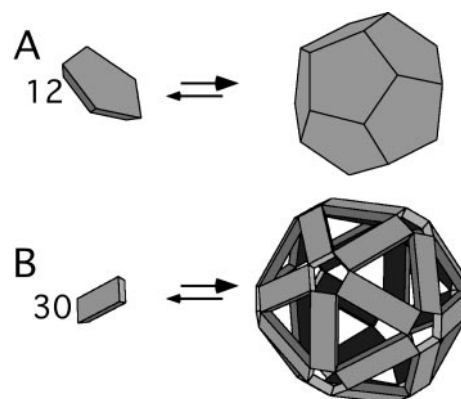
In this paper we further develop techniques for analyzing assembly (Zlotnick, 1994; Zlotnick et al., 1999) using two specific model systems to illustrate the kinetic analysis derived more generally in the Appendix. The example systems show two specific cases of the differential rate equations (A2) for subunit, intermediate species, and capsid concentrations. In its complete generality the model system allows the assignment of different rates and/or orders for each of the cascade of assembly reactions and is in some respects the spherical analog of a helical filament model (Flyvbjerg et al., 1996). Fundamental geometric differences between filament and shell assemblies require distinct mathematical treatments, as the former is open-ended and the latter is closed.

## METHODS

See Table 1 for naming conventions.

### Modeling capsid assembly at equilibrium

Two model capsids were used for calculating simulations (Fig. 1). The first, initially developed without nucleation as the equilibrium (EQ) model



**FIGURE 1** Capsid geometries. (A) The dodecahedral model capsid has 12 pentagonal subunits. (B) The icosahedral model capsid has 30 twofold symmetrical subunits. The subunits for the icosahedral model are tetravalent, with binding sites at each corner; subunits for the dodecahedral model are pentavalent, with a binding site at each edge.

in Zlotnick (1994), consists of 12 pentavalent pentagonal subunits assembling as faces of a dodecahedron by making 30 intersubunit contacts (Fig. 1 A). This geometry resembles picornavirus quaternary structure (Rueckert, 1996). The second system is built from 30 tetravalent rectangular subunits (making 60 intersubunit contacts), which can be thought of as a  $T = 1$  icosahedron assembled from “dimeric” subunits (Fig. 1 B). The tetravalent subunit resembles the geometry used in schematic representations of hepatitis B virus capsid protein dimer (Zlotnick et al., 1996).

To describe capsid polymerization we develop the system of rate equations for the concentrations of species formed in the assembly cascade detailed in the Appendix (Eq. A2). The equations are determined by capsid geometry, assembly path, and intersubunit binding energies. At equilibrium, the differential rate equations reduce to a series of algebraic equations from which it is possible to determine equilibrium values for subunit, capsid, and intermediate species. For a simplest model, at equilibrium, only monomer and capsid are present at appreciable concentrations (Zlotnick, 1994). The equilibrium relationship between the reaction cascade endpoints is:

$$K_{Acap} = \frac{[N]_\infty}{[u]_\infty^N} \quad (1)$$

where  $N$  is the number of subunits in a complete capsid, e.g.,  $N = 12$  for the dodecahedral and  $N = 30$  for the icosahedral model.

Equation 1 results in the unwieldy units of  $M^{1-N}$  for the association constant  $K_{Acap}$ . Two expressions that are more convenient are  $K_{Dapp}$ , the apparent dissociation constant defined where equilibrium concentrations of subunit and capsid are equal (Eq. 2) and  $K_{Acon}$ , the pairwise (per-contact) association constant between subunits (see Eq. 3) (Zlotnick, 1994). Because  $N$  is typically large, it is more practical to use the logarithmic form of equations involving  $K_{Acap}$  when manipulating numerical data.

$$K_{Dapp} = (K_{Acap})^{1/1-N} \quad (2)$$

Successive substitutions through the sequence of model equations at equilibrium give  $K_{Acap}$  as the product of 1) a statistical coefficient accounting for the overall degeneracy of the reaction, and 2) a power of  $K_{Acon}$  accounting for each contact formed in capsid assembly. This can be written generally for the assembly of a polyhedral structure of  $N$  subunits with  $c$  contacts per subunit where each subunit has  $j$ -fold degeneracy. Such an  $N$ -hedral structure will have a total of  $C = cN/2$  pairwise contacts, leading

to the general form where  $j^{N-1}/N$  is the product of the statistical factors for the degeneracy of the component reactions (Eq. 3).

$$K_{\text{Acap}} = \frac{j^{N-1}}{N} (K_{\text{Acon}})^c \quad (3)$$

Individual statistical factors can be determined explicitly for each assembly step as previously described (Zlotnick, 1994). Applying Eq. 3 to our dodecahedral and icosahedral models gives Eqs. 4.

$$K_{\text{Acap}} = \frac{5^{11}}{12} K_{\text{Acon}}^{30} \quad \text{dodecahedral}$$

$$K_{\text{Acap}} = \frac{2^{29}}{30} K_{\text{Acon}}^{60} \quad \text{icosahedral} \quad (4)$$

In general, equilibrium is described for a system of rate equations by setting the derivative equal to zero. For models like those discussed here (Eqn A2), having a single representative conformation for each size of intermediate, this algebraic system is easily solved by back-substitution, giving the equilibrium concentrations of all species in terms of subunit

$$[m]_{\infty} = \frac{s_2 \cdots s_m}{\sigma_2 \cdots \sigma_m} K_{\text{Acon}}^{c_m} [u]_{\infty}^m \quad (5)$$

where  $c_m$  gives the total number of intersubunit contacts formed in assembling the  $m$ th species (composed of  $m$  subunits). For models incorporating several paths, the expression (Eq. 5) exhibits a complicated dependence on  $K_{\text{Acon}}$  and degeneracy.

In principle,  $K_{\text{Acon}}$  need not be the same for all interactions. For example, the  $K_{\text{Acon}}$  used for the nucleation interactions might be less than that used for contacts established during elongation. If this were a simple entropy penalty, one would expect the penalty to be recouped once the nucleus is assembled (Oosawa and Asakura, 1975). For simplicity, we use the same  $K_{\text{Acon}}$  for all reactions.

## Modeling assembly kinetics

The above model of assembly allows calculation of assembly kinetics with few additional assumptions. We assume for simplicity that every contact scores the same per-contact association energy,  $\Delta G_c$ , from which we define  $K_{\text{Acon}}$  via the Arrhenius relation  $\Delta G_c = -RT \ln K_{\text{Acon}}$ . Second, we limit our simulations to only two microscopic forward rate constants, for nucleation ( $f_{\text{nuc}}$ ) and elongation ( $f_{\text{elong}}$ ), which are modified by appropriate statistical factors. Third, we have restricted assembly to proceed only by addition of individual subunits, without high-order or pseudo-high-order elongation steps. Finally, some consideration must be given to the choice of assembly paths that are used in kinetic simulations: a typical capsid may have hundreds or thousands of possible paths through tens or hundreds of intermediates. This is analogous to an energy landscape model of protein folding (Dinner et al., 2000; Dill, 1999). However, the first and last few reactions in capsid assembly have little degeneracy; the greatest degeneracy is encountered when assembly is half complete.

There are two obvious strategies for limiting the number of paths used in the model: 1) use the “best” paths, and 2) use the “average” path. In the first strategy, we define a measure of “significance” or the likelihood of a path. In the second, we specify only one intermediate of each size, the “average” intermediate of that size. Both of these paths follow steep energy descents with no intrinsic energy minima other than capsid; these paths are akin to a smooth, steep trajectory on an energy landscape.

One realization of the best-path strategy is to use paths through maximally stable intermediates, an *msi-measure*. For example, there are typically two configurations for the trimer: open (linear) and closed (triangular). Given that subunit geometry admits both, closed trimers are a lower-

energy intermediate than open trimers because they gain one additional unit of contact energy from closure. The *msi-measure* assigns a measure of one to paths rooted in the most stable intermediates and zero to all others, deleting them. We have used the *msi-measure* to choose intermediates for our dodecahedral model. Statistical coefficients are still required to reflect degeneracy of the *msi-path(s)*.

In the “average-path” strategy, energy and degeneracy are distributed evenly among the steps along a single assembly path, which may or may not be an actual path. Although we developed a version of our icosahedral model using the *msi-measure*, the behavior of the *msi* model was not significantly different from that of the average-path version with regard to the concentrations of capsid ( $[N]$ ) and subunit ( $[u]$ ). In the average-path construction for the icosahedral model there are 29 reactions over which we distribute the energy from the 60 contacts. The dimerization step yielded only one energy unit because there is only one contact. The next 26 steps were assigned two energy units each (two intersubunit contacts per subunit association) because actual paths have from one to three new contacts formed per step. The final two subunits make three and four contacts on association and were assigned three and four energy units, respectively. Reaction degeneracy was distributed in a similar manner.

Assembly kinetics are simulated as species concentrations  $[u]$ ,  $[2]$ ,  $\dots$ ,  $[N]$  evolve according to the differential rate equations (Eq. A2) from which we extract a typical example for convenient reference as Eq. 6.

$$\frac{d[m]}{dt} = f_m s_m [u][m-1] - f_{m+1} s_{m+1} [u][m] + b_{m+1}[m+1] - b_m[m] \quad (6)$$

This equation takes account of both subunit-intermediate association (forward “ $f$ -terms”) and dissociation (backward “ $b$ -terms”) contributing to the kinetics of the  $m$ th intermediate. The statistical factors  $s_m$  and  $\sigma_m$  adjust the rates to account for the degeneracy of the reactions; examples of their calculation for the dodecahedral model are given in Zlotnick (1994). The reverse rate constant for the  $m$ th reaction

$$b_m = \sigma_m \frac{f_m}{(K_{\text{Acon}})^r} \quad (7)$$

is calculated from the forward rate ( $f_m$ ), the per-contact energy (from  $K_{\text{Acon}}$ ), the number ( $r$ ) of contacts broken in dissociating the  $m$ th subunit, and the reverse reaction’s degeneracy  $\sigma_m$ .

The system of rate equations (A2) developed from the above formalism is not amenable to exact solution, unlike filament assembly (Flyvbjerg et al., 1996). Solutions were calculated using BERKELEY MADONNA (Berkeley Software, Berkeley, CA), with either RK4 or Rosenbrock numerical integration methods, and Waterloo MAPLE (Waterloo, ON, Canada) implementations of the RKF45 integrator. MAPLE simulations were calculated using 14 of the 15 digits available in standard floating-point precision, consuming no more than a few minutes of CPU-time per initial concentration. In no case was high precision required to obtain accurate values for parameter estimators. Different models were fit to synthetic data sets by minimizing the RMSD, as implemented in BERKELEY MADONNA.

## RESULTS AND DISCUSSION

### An overview of simulated assembly kinetics

Sigmoidal kinetics such as those exhibited by capsid assembly (Fig. 2) are typical for the synthesis of product by any multistep reaction (Fersht, 1999). For capsid assembly, the three stages of kinetics (lag, rapid assembly, and asymp-

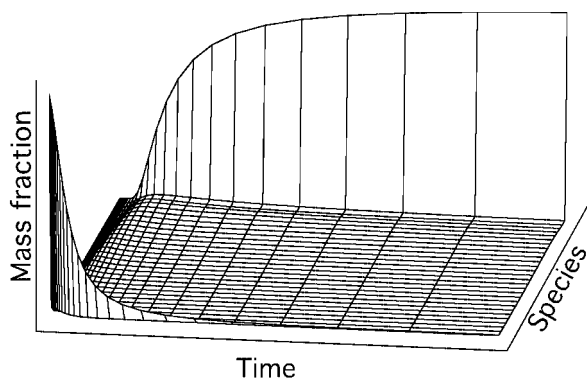


FIGURE 2 Assembly kinetics calculated for the icosahedral model. The cascade of reactions exhibits the sigmoidal kinetics experimentally observed for *in vitro* capsid assembly. Subunit concentration declines. Intermediates successively rise during lag phase and then slowly decrease to extremely low equilibrium concentrations. After all intermediates are synthesized, rapid production of capsid begins.

otic) can be associated with well-defined characteristics of the reaction cascade.

The initial lag phase is characterized by the sequential building-up of assembly intermediates. Each intermediate reaches a peak concentration in its turn, then slowly relaxes toward its eventual equilibrium concentration. Until the final intermediate begins to approach its peak concentration, capsid production is very slow. The lag phase can be characterized as the initial interval during which a near steady-state “assembly line” of intermediates is constructed. In reactions with nucleation and elongation components, the elongation rate has its clearest effect on the lag duration (see assertion 4).

During the rapid stage of kinetics there is a burst of capsid synthesis as free subunits react with intermediates, effecting a nearly steady conversion of mass in free subunit into mass in capsid state. The rapid rate of capsid synthesis is supported by the relatively high concentrations of reactants, both free subunit and intermediate. During this stage, intermediate concentrations remain almost constant. Both nucleation and elongation rates make contributions to the slope of this part of the kinetic trajectory.

It turns out that assembly path degeneracy is not an important issue for simulations where intermediate concentrations are low. Different assembly paths lead to small changes in the concentrations of different intermediates, but not subunit or capsid. We found that the back reactions are insignificant in the elongation steps until reactions approach equilibrium. Because all assembly reactions proceed one subunit at a time at nearly identical forward rates, all possible paths will take roughly the same time to complete, although some paths are more likely than others.

The asymptotic phase occurs when the depletion of free subunit becomes significant. Low subunit concentration causes the slow rate of nucleus formation to be the major regulating factor in capsid synthesis (see assertion 2).

Conceptually, nucleation occurs when the first few intermediates in a series of reactions are less stable than subsequent products, as might occur in a helical polymer prior to the first interactions between turns (Oosawa and Asakura, 1975; Flyvbjerg et al., 1996). Nucleation of a spherical polymer might depend on formation of a primitive closed structure, such as the first face or vertex of a polyhedron (e.g., an HBV trimer (Zlotnick et al., 1999)). Energetically, nucleation reactions can be differentiated from subsequent elongation reactions by a slower forward rate and/or weaker association energy. For simplicity, we have chosen to focus on a rate-delimited nucleation step, observing that our results do not significantly depend on the mechanism by which nucleation differs from elongation.

Reaction simulations are less susceptible to kinetic traps when assembly is regulated by a nucleation step (Zlotnick et al., 1999). Nucleated reactions can yield near-equilibrium concentrations of capsid without trapping despite relatively high association energies, high initial concentrations of subunit, and fast rates (see assertion 1). Resistance to kinetic trapping results from prevention of over-initiation of assembly. Because the forward rate is proportional to subunit concentration, when subunit concentration is too high, virtually all subunits are assembled into intermediates in the absence of a regulatory step. In nucleated reactions, or in reactions “auto-sterically” regulated by the persistence of assembly-competent and -incompetent forms of the subunit (Caspar, 1980; Schwartz et al., 1998), unbound subunit is held in reserve. The introduction of an assembly-incompetent subspecies amounts to no more than the formal addition of a slow, initial, first-order reaction step. Mathematically speaking, it is not surprising that either mechanism resists trapping because both introduce slow coordinates in the early equations of the assembly cascade.

### Nucleation and the steady-state assembly line

In the absence of a regulating step, successful assembly requires a balance between protein concentration and association energy (Zlotnick, 1994). As we increase either contact affinity,  $K_{Acon}$ , or initial subunit concentration,  $u_0$ , the amount of subunit in the assembly line increases. During the rapid-production phase, subunit can be thought of as being converted, via the assembly line, “directly” from free to bound-in-capsid state: the concentrations of intermediates approximate a steady state. When  $K_{Acon}$  or  $u_0$  is too large, assembly is over-initiated; too much mass has entered the assembly line, leaving the reactions starved for available subunits: a kinetic trap. Conversely, too little mass in the assembly line will result in a slow reaction that is starved for intermediates. Because capsid production is slow for either extreme, there is an optimal steady-state concentration of intermediates for  $K_{Acon}$  at some value of  $u_0$  (cf. assertion 1 and Fig. 4).

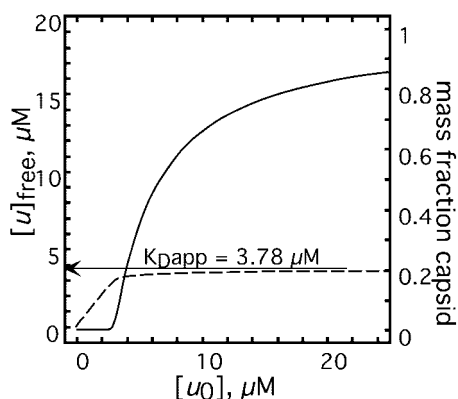


FIGURE 3 An assembly isotherm for the icosahedral (30-mer) model. The mass fraction capsid (solid line, right axis) as a function of the initial subunit concentration,  $[u_0]$ . Capsid concentration increases as a 30th power of free subunit (Eq. 3). Free subunit concentration (dashed line) varies little when it approaches  $K_{Dapp}$ , but it is not constant; i.e., there is no true critical concentration. Subunit concentration  $[u]$  will reach  $K_{Dapp}$ , the horizontal line at  $3.78 \mu\text{M}$   $[u]$ , when  $[u]_\infty = [N]_\infty$ ; on a mass fraction basis, this will occur at  $117 \mu\text{M}$  total subunit with  $\sim 97\%$  of the mass as capsid. The isotherm was calculated for  $\log(K_{Acon}) = 2.5$ , corresponding to  $\Delta G_c = -3.4 \text{ kcal/mol}$ .

In nucleated assembly the concentration of free subunit falls much more gradually than during assembly lacking a regulating step (Zlotnick et al., 1999). Nucleated assembly is resistant to kinetic trapping, but not immune to it. The degree of resistance depends on the distinction between nucleation and elongation phases. For example, when the ratio of the forward rate constants,  $f_{nuc}/f_{elong}$ , is  $< \sim 0.001$ , assembly may be extremely resistant to trapping; when  $f_{nuc}/f_{elong} > 0.05$ , the reaction resembles EQ assembly and may be easily trapped. A weaker contact energy for nucleation reactions enhances the resistance of nucleated assembly to kinetic trapping.

### Apparent $K_D$ and pseudo-critical concentrations

As  $u_0$  is increased, the equilibrium capsid concentration increases rapidly because of the high algebraic degree of Eq. 1. As Fig. 3 shows, the point of greatest sensitivity to change in  $u_0$  is well approximated by  $K_{Dapp}$  (Eq. 2), the subunit concentration for which equilibrium subunit and capsid concentrations will be equal. Because  $K_{Dapp}$  is calculated on a molar basis, the initial concentration required to achieve it is  $(N + 1)K_{Dapp}$ . In these models  $K_{Dapp}$  is a pseudo-critical concentration (Fig. 3). From Eq. 1 it is clear that free subunit concentration  $[u]_\infty$  will vary, albeit slowly, for  $u_0 \gg (N + 1)K_{Dapp}$ . Because  $[u]_\infty$  increases so slowly as a function of  $u_0$ , there is the appearance of a constant  $[u]_\infty$ , approximated by  $K_{Dapp}$ . This same effect was predicted and observed with micelle formation (Tanford, 1980). It is not surprising that critical concentrations have been indicated for CCMV (Adolph and Butler, 1976) and

HBV (Seifer et al., 1993), whose capsids have 90 and 120 subunits, respectively, as CCMV and HBV would be expected to have much steeper isotherms than the 30-subunit model (Fig. 3).

### Scaling assembly simulations

We searched for intrinsic units of concentration and time for kinetic simulations (see Eq. A5) similar to those observed in the case of filament assembly (Flyvbjerg et al., 1996). Approximately the same trajectories are observed for reactions where concentration is expressed in units of  $u_0 K_{Acon}$  and time in units of  $t/K_{Acon}$ . This scale invariance is intrinsic to the model equations A2, simplifying comparison of different kinetic simulations. The rescaling is most nearly exact during the lag and rapid-assembly stages when the forward-reaction terms dominate the kinetics and all cascade reactions are essentially unidirectional.

### Assertion 1: $K_{Acon}$ can be estimated from kinetic trajectories at long times

A reaction must be at equilibrium for precise determination of association energy, yet kinetic simulations of virus assembly reactions (Fig. 2) approach equilibrium asymptotically. We examined simulations over long times to observe how closely they approached equilibrium and to test our estimation of the pairwise association energy,  $\Delta G_c$ . Reactions were simulated for a range of  $K_{Acon}$  values. For any fixed concentration, reactions at modest  $K_{Acon}$  approached their equilibria rapidly; at higher values reactions were susceptible to stalling. Stalling occurs when the concentration of reactants is so low that bimolecular reactions do not proceed at an appreciable rate.

As assembly kinetics approach equilibrium, capsid concentration changes very slowly. Approximate values for  $\Delta G_c$  were calculated using the concentrations of capsid  $[N]$  and subunit  $[u]$  measured at long times in place of their equilibrium values in Eqs. 1, 3, and 4. Intermediates were not included in these calculations to make our results more “realistic” because intermediate concentrations are expected to be so low that they are experimentally undetectable. In long-time simulations with modest  $K_{Acon}$ , the estimated per-contact energy  $\Delta G_c$  was well within 10% of the exact value. Estimates for reactions with stronger association energies yielded worse approximations because these reactions stall when they became starved for free subunits and intermediates at long times. This is because a large  $K_{Acon}$  corresponds to a very low equilibrium concentration of subunit,  $[u]_\infty$  (Eqs. 1 and 3) so reactions reached a starvation condition and stalled before approaching equilibrium. Estimation of  $\Delta G_c$  was slightly more accurate at low initial subunit, even though higher initial subunit concentrations yielded a larger mole fraction of capsid. This seems coun-

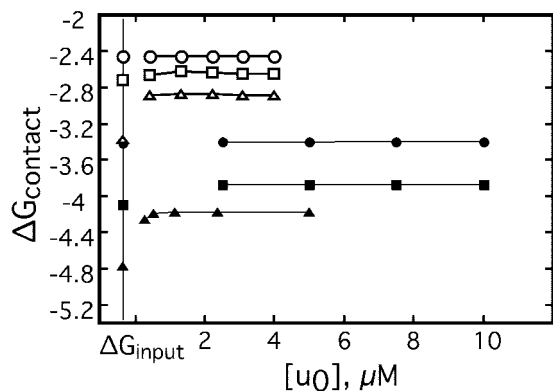


FIGURE 4 Energy per contact,  $\Delta G_c$ , calculated from the  $K_{Acap}$  observed at the end of kinetic simulations of a 24-h assembly experiment for the dodecahedral (*open symbols*) and icosahedral (*closed symbols*) models. The simulations all use  $f_{nuc} = 100 \text{ M}^{-1} \text{ s}^{-1}$ ,  $f_{elong} = 10^6 \text{ M}^{-1} \text{ s}^{-1}$ . Association constants were calculated using Eqs. 1 and 2, and converted into energy units using  $\Delta G_c = -RT \ln K_{Acon}$ , where  $T$  is set to 298 K and  $R$  is the gas constant,  $1.987 \text{ cal (deg mol)}^{-1}$ . Simulations for the 12-mer model correspond to  $-2.5$  (*circles*),  $-2.7$  (*squares*), and  $-3.4$  kcal/(mol contact) (*triangles*). Simulations for the 30-mer were calculated for  $\Delta G_c$  of  $-3.4$  (*circles*),  $-4.1$  (*squares*), and  $-4.8$  kcal/(mol contact) (*triangles*). Though the ranges of  $\Delta G_c$  are different, the  $\Delta G_{app}$  values are similar due to capsid and subunit geometry. Ignoring the relatively small statistical component, for the 12-mer,  $\Delta G_{app}$  is  $\sim 30/11(\Delta G_c)$ ; for the 30-mer,  $\Delta G_{app}$  is  $\sim 60/29(\Delta G_c)$  (see Eq. 7).

terintuitive in light of stalling due to starvation; however, this is simply an effect of  $u_0$  being closer to  $[u]_{\infty}$ , when  $u_0$  is small. The dodecahedral model proved to be more stringent than the icosahedral model for this estimation (Fig. 4) because subunit geometry results in a lower  $K_{Dapp}$  for a given  $\Delta G_c$ , as can be seen by combining Eqs. 2 and 4 to generate the approximations of Eq. 8.

$$K_{Dapp} \approx 0.27 K_{Dcon}^{5.5} \quad \text{dodecahedral}$$

$$K_{Dapp} \approx 0.21 K_{Dcon}^2 \quad \text{icosahedral} \quad (8)$$

In either case, a  $K_{Dapp}$  much lower than  $\sim 0.2 \mu\text{M}$  will result in a stalled reaction.

### Assertion 2: Nucleation rate can be determined from the rate of capsid formation

The regulatory step has a profound effect on assembly kinetics. For nucleated assembly, both the size of the nucleus and the rate of nucleation must be considered. We consider the case where nucleation proceeds from slow association of subunits to form a metastable nucleus.

When a reaction enters the asymptotic phase, the assembly line of intermediates is still in steady state, though the rate of capsid production is attenuated. Because every new nucleus enters the assembly line and proceeds to form a capsid, we assert that the rate of capsid production is pro-

portional to the rate of nucleus formation. More accurately, the average forward rate constant for the nucleation steps is proportional to the rate of capsid formation divided by  $[u]^{nuc}$ . This is expressed in

$$\frac{d[N]}{dt} \approx f_{nuc} s_{nuc} K_{subnuc} [u]^{nuc} \quad (9)$$

where  $K_{subnuc}$  is the equilibrium constant for the formation of  $[nuc - 1]$ . The value for  $K_{subnuc}$  is the coefficient in Eq. 5. The derivation of Eq. 9 is provided in the Appendix (see A7–A12). We are able to arrive at Eq. 9 because in nucleated assembly 1) elongation reactions are essentially unidirectional, and 2) from relatively early times forward, subnuclear intermediates are already near their equilibrium concentrations. Equation 9 holds at early times in the asymptotic phase, before intermediates in the assembly line have begun to disassemble. Under these conditions, almost every nucleus produced leads to the formation of a complete capsid, i.e., the rate of capsid production is approximately the same as the forward-rate of nucleus production.

Given the size of the nucleus (see assertion 3), nucleus geometry and statistical factors can be deduced. The value for  $K_{subnuc}$  can be calculated from  $K_{Acon}$  (see assertion 1) and the statistical factors. Equation 9 is then an equation in only one unknown,  $f_{nuc}$ . Fig. 5 shows plots of the ratio

$$\frac{d[N]}{dt} / [u]^{nuc} \quad (10)$$

calculated for simulations using three values of  $f_{nuc}$  reflecting the impact that decreasing the forward nucleation rate has on the ratio in Eq. 10. The value of the ratio is concentration-independent after the reaction has reached steady state. The plateau height in each case is equal to the forward rate constant,  $f_{nuc}$ , multiplied by  $K_{subnuc}$ . We emphasize that the estimator depends only on the nucleation rate and not on the elongation rate. In simulations where the nucleation rate was fixed and the elongation rate varied, the plateau height remained fixed.

In general, concentration independence of a quantity calculated from concentration-dependent data indicates that the calculated quantity is a parameter of the system. In this case, the concentration independence in the plots illustrates the dependence of the ratio (Eq. 10) on the coefficients in the governing equations (Eq. A2) and not on the initial conditions.

### Assertion 3: Nucleus size can be determined from the concentration dependence of the extent of assembly

Simulations show that nucleus size cannot be determined accurately by searching for a best fit to the exponent in Eq. 9. In the analysis given in the Appendix (Eqs. A13–A18) we develop a sensitive estimator (Eq. 12) for nucleus size. This

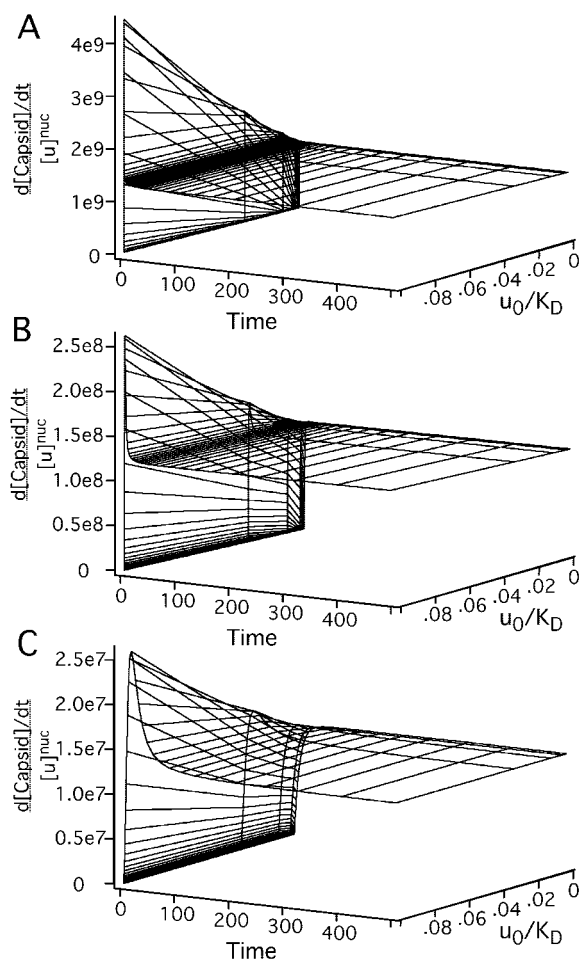


FIGURE 5 The rate of capsid formation can be related the nucleation rate,  $f_{nuc}$ . Time and concentration series for the ratio (Eq. 10) of capsid production rate,  $d[N]/dt$ , to free subunit concentration raised to the nucleus size power,  $[u]^{nuc}$ . The concentration-independent plateau height is directly proportional to the nucleation rate. The value of  $f_{nuc}$  was set to  $10^4 \text{ M}^{-1} \text{ s}^{-1}$  (A),  $10^3 \text{ M}^{-1} \text{ s}^{-1}$  (B), and  $10^2 \text{ M}^{-1} \text{ s}^{-1}$  (C). For these three families of reactions,  $f_{elong} = 10^6 \text{ M}^{-1} \text{ s}^{-1}$ , and  $\Delta G_c = -4.1 \text{ kcal/mol}$ .

analysis rests on two assumptions: 1) the ratio in Eq. 10 is independent of  $u_0$ , and 2) the reactions considered are in a transient steady state where formation of capsid is equal to the depletion of free subunit (Eq. 11).

$$N \frac{d[N]}{dt} = - \frac{d[u]}{dt} \tag{11}$$

Our results (see Appendix) show that the nucleus size appears as the temporal minimum value of the function

$$L = \left( \frac{\partial \ln[N]}{\partial u_0} / \frac{\partial \ln[u]}{\partial u_0} \right) \tag{12}$$

The function  $L$  is the slope of a plot of  $\ln[N]$  versus  $\ln[u]$  at a single time across a range of initial protein concentrations,  $u_0$  (Zlotnick et al., 1999). The minimum will only be found

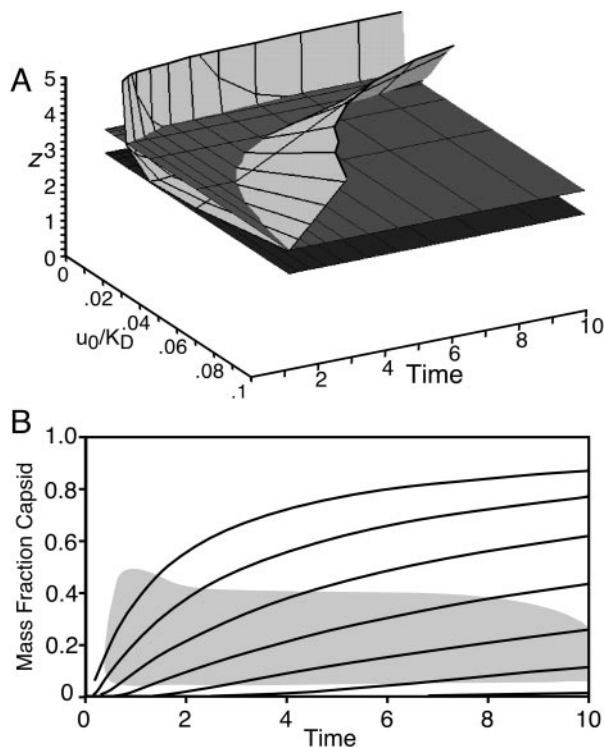


FIGURE 6 Determination of nucleus size for the icosahedral model. The minimum of the function  $L$  (Eq. 16) gives the nucleus size estimate. (A) The surface of  $L$  (light gray) is shown for the  $nuc = 3$ ,  $\nu = 7 \times 10^{-5}$ , icosahedral model. The minimum of the surface is partially obscured by the error plane  $z = 3.33$  and lies above the error plane  $z = 2.67$ , which are dark and medium gray, respectively. (B) The corresponding kinetic simulations (black) show where (shaded region) nucleus size is estimated to within  $\pm 0.3$ .

for a family of reactions that are simultaneously in steady state. This means that there is a range of times and concentrations for which the function  $L$ , near its minimum, approximates nucleus size.

Fig. 6 A shows the graph of the function  $L$  over a range of times and concentrations for our icosahedral model with nucleus size  $nuc = 3$ . The minimum value of  $L$  varies little from the nucleus size. Fig. 6 B shows the series of simulations used to calculate the surface in Fig. 6 A, with the shaded region delineating the domain in which nucleus size is approximated to within 0.33. Nucleus size is unambiguously approximated over a fairly wide range of initial concentrations at early time. For lower concentrations, the approximation holds at later times. Fig. 7 shows the surfaces  $z = L(u_0, t)$  for nucleus sizes 3, 4, and 5 in the dodecahedral model.

**Assertion 4: Elongation kinetics are most apparent in the early phases of assembly reactions**

Examination of kinetic simulations (Fig. 2) shows that the lag phase is due to the time required to assemble

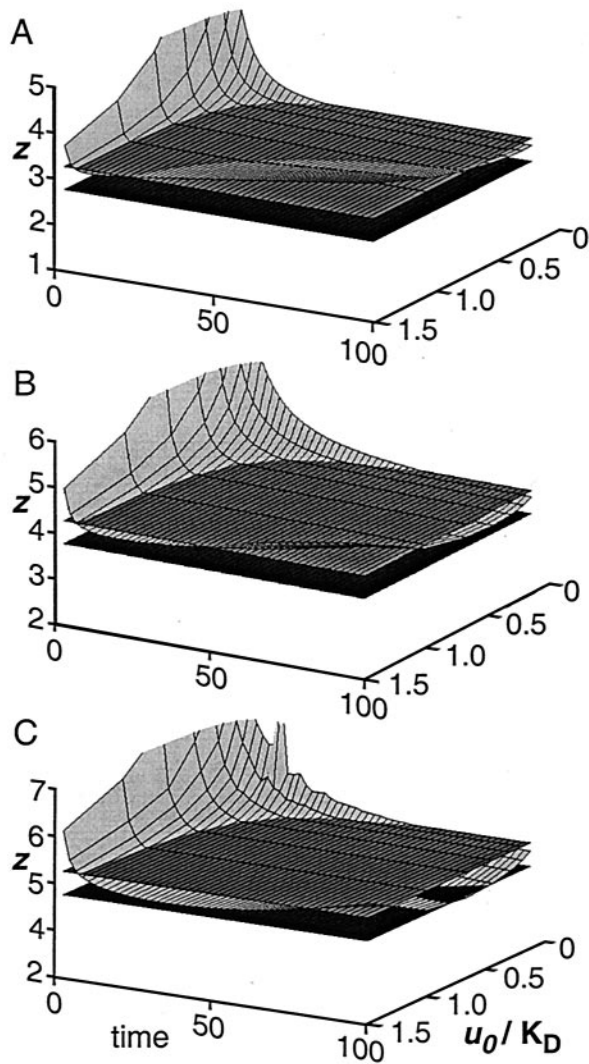


FIGURE 7 Determination of nucleus size for three different families of dodecahedral simulations. Nucleus size estimation surfaces  $z = L(u_0, t)$  for the dodecahedral model with nucleus sizes of 3 (A), 4 (B), and 5 (C). The figure follows the use of grayscales as in Fig. 6.

intermediates. The duration of this phase is affected by nucleus size, nucleation rate, the stability of subnuclear intermediates (i.e.,  $K_{Acon}$ ), and elongation rate. Other parameters remaining constant, simulations indicate a localized dependence of the lag duration on the elongation rate,  $f_{elong}$  (Fig. 8). This is clearest when comparing a series of kinetic trajectories where only  $f_{elong}$  was varied (Fig. 8 A). Note that the different simulations coalesce at longer times, where the nucleation rate regulates the rate of capsid formation (assertion 2). When examining a range of subunit concentrations, the effect of  $f_{elong}$  on the lag duration is most evident at lower  $[u_0]$  (Fig. 8, B–E). The variation in lag duration makes it clear that it should be possible to approximate  $f_{elong}$  by curve-fitting.

For small nuclei in large capsids  $f_{elong}$  becomes more important because of the many post-nuclear steps. The elongation rate principally regulates the lag phase while the nucleation rate regulates all phases. Because the nucleation rate affects all phases (Eq. A2), it approximates a global time rescaling factor.

Because the rapid phase includes contributions from  $f_{nuc}$  and  $f_{elong}$ , we examined the concentration dependence of  $\ln(d[N]/dt)$  against  $\ln u_0$ , a classical kinetic approach for estimating reaction order in single-step reactions. We observe the concentration dependence of maximal rate is  $\sim 1$ . This is consistent (though not conclusive) with a pseudo-first-order reaction, as previously suggested (Zlotnick et al., 1999). There is no simple relationship between the dependence of  $\ln(d[N]/dt)$  on  $\ln u_0$  and nucleus size.

### Sensitivity of kinetic trajectories to model and parameters

How reliable are parameters estimated from assembly kinetics? How well can we discriminate between the correct model of an assembly reaction and a model that mimics the data but does not represent a physically reasonable description of the reaction?

We find that the models are sensitive to variations in parameters. A simulation produced using the wrong parameter values, though it may mimic data at one concentration, could not fit a concentration series. We demonstrate this by examining two reactions (Fig. 9). In the first case we fit  $K_{Acon}$ ,  $f_{nuc}$ , and  $f_{elong}$  for an icosahedral (30-mer) model to data from a dodecahedral simulation. In the second case, the correct model was used to fit simulated data for an icosahedron, but we held  $K_{Acon}$  to an incorrect value. In both experiments residuals showed a systematic error that was more pronounced when calculated for a concentration series, though the initial fit may have looked attractive.

Calculated kinetic trajectories are sensitive to model geometry and kinetic-thermodynamic parameters. Fitting experimental data requires an appropriate model, e.g., a 120-mer *msi*-path model for hepatitis B assembly. A close fit across a concentration series indicates that the assembly is well represented by the model and that the parameters extracted from the experimental data (as outlined in our assertions 1-3) give a fair representation of the true parameters. We stress, however, that there is no need for such a comparison to extract parameter values from experimental data. A poor fit calculated with parameters extracted from the experimental data indicates that mechanisms not included in the initial model need to be included in further refinements. In this way the extraction of parameters followed by model fitting serves as a basic probe of the assembly process.



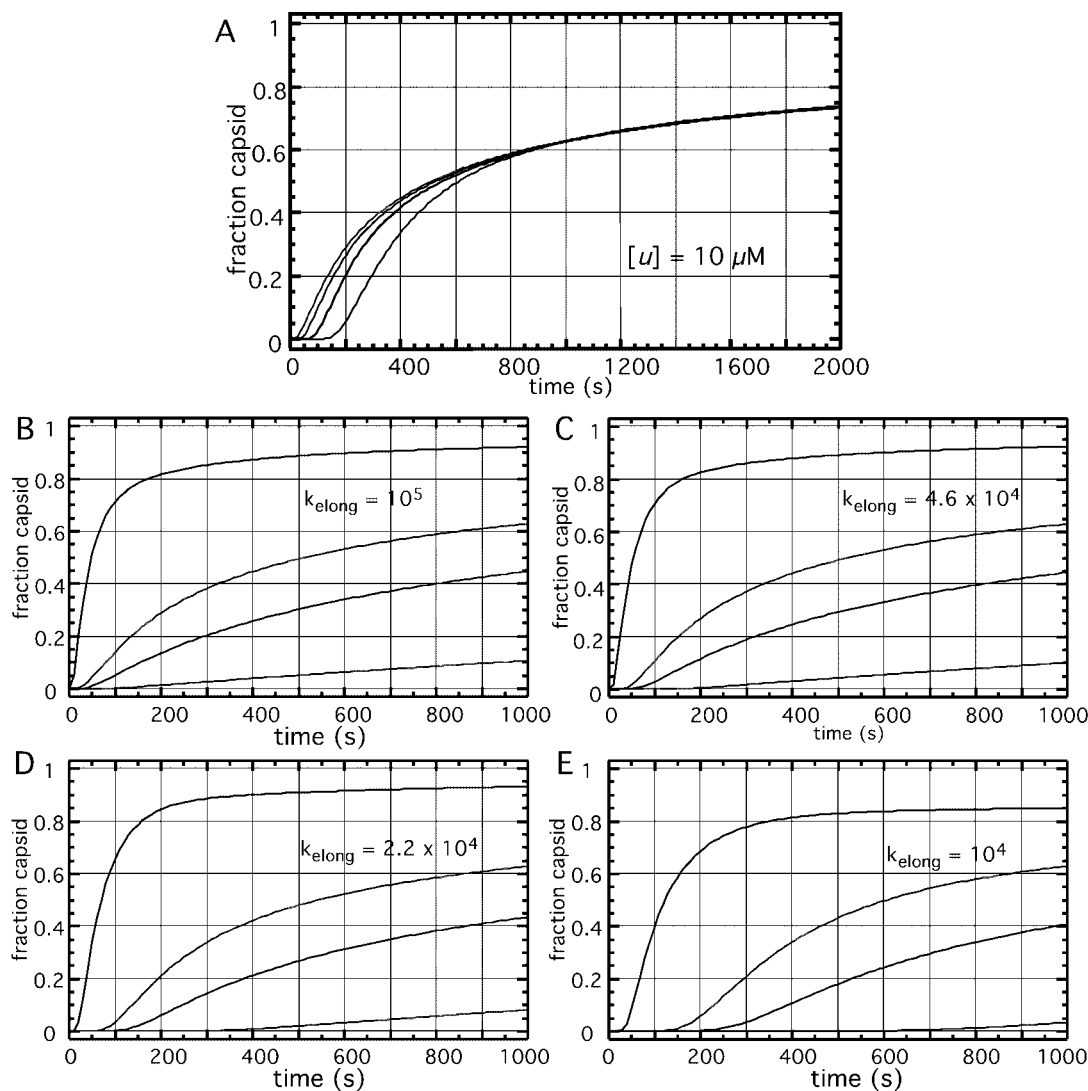


FIGURE 8 The effect of elongation rate is most obvious in the early stages of assembly kinetics. (A) Elongation rate is varied for a constant initial subunit concentration. The elongation rates are  $10^5$ ,  $4.6 \times 10^4$ ,  $2.2 \times 10^4$ , and  $10^4 \text{ M}^{-1} \text{ s}^{-1}$ , from left to right; the nucleation rate is  $100 \text{ M}^{-1} \text{ s}^{-1}$ . At long times, where the nucleation governs the rate of capsid formation, the four curves are coincident, as predicted for a common nucleation rate. (B–E) Families of kinetic simulations where  $[u_0]$  is 20, 10, 6, and  $2 \mu\text{M}$ , corresponding to  $[u_0]/K_D$  of 0.063, 0.032, 0.019, and 0.0063. For these reactions,  $f_{\text{elong}}$  is  $10^5$  (B),  $4.6 \times 10^4$  (C),  $2.2 \times 10^4$  (D), and  $10^4 \text{ M}^{-1} \text{ s}^{-1}$  (E). All simulations in this figure were calculated using a 30-mer model with a trimeric nucleus,  $\Delta G_c = -4.1$ , and  $f_{\text{nuc}} = 100 \text{ M}^{-1} \text{ s}^{-1}$ .

## CONCLUSIONS

In conjunction with studies of the assembly of CCMV and HBV, we have developed a simple model of capsid assembly that makes limited assumptions. Based on this mathematically expressed model, we have derived techniques to determine reaction parameters from kinetic data.

The basic assembly model consists of defined capsid geometry and four parameters: the microscopic, per-contact equilibrium constant ( $K_{\text{Acon}}$ ), the nucleus size ( $nuc$ ), the nucleation on-rate ( $f_{\text{nuc}}$ ), and the elongation on-rate ( $f_{\text{elong}}$ ). We have shown how good approximations of three of these parameters can be extracted from the kinetics of capsid

formation. In our experience, these are the most accessible data from assembly experiments. We have illustrated these methods by extracting parameters from simulated data calculated from 12- and 30-subunit models. Practically, the larger model gave more robust results, as mathematically expected. The mathematical analysis was carried out for an arbitrary capsid size and spherical capsid geometry consistent with equivalent subunits. Estimation of other assembly parameters will most likely require fitting observed kinetics with a more detailed model of the assembly reaction.

In the first few minutes of a typical in vitro assembly reaction,  $\sim 10^{12}$  nuclei form and go on to yield capsids (Zlotnick et al., 1999, 2000); this is unlike polymeriza-

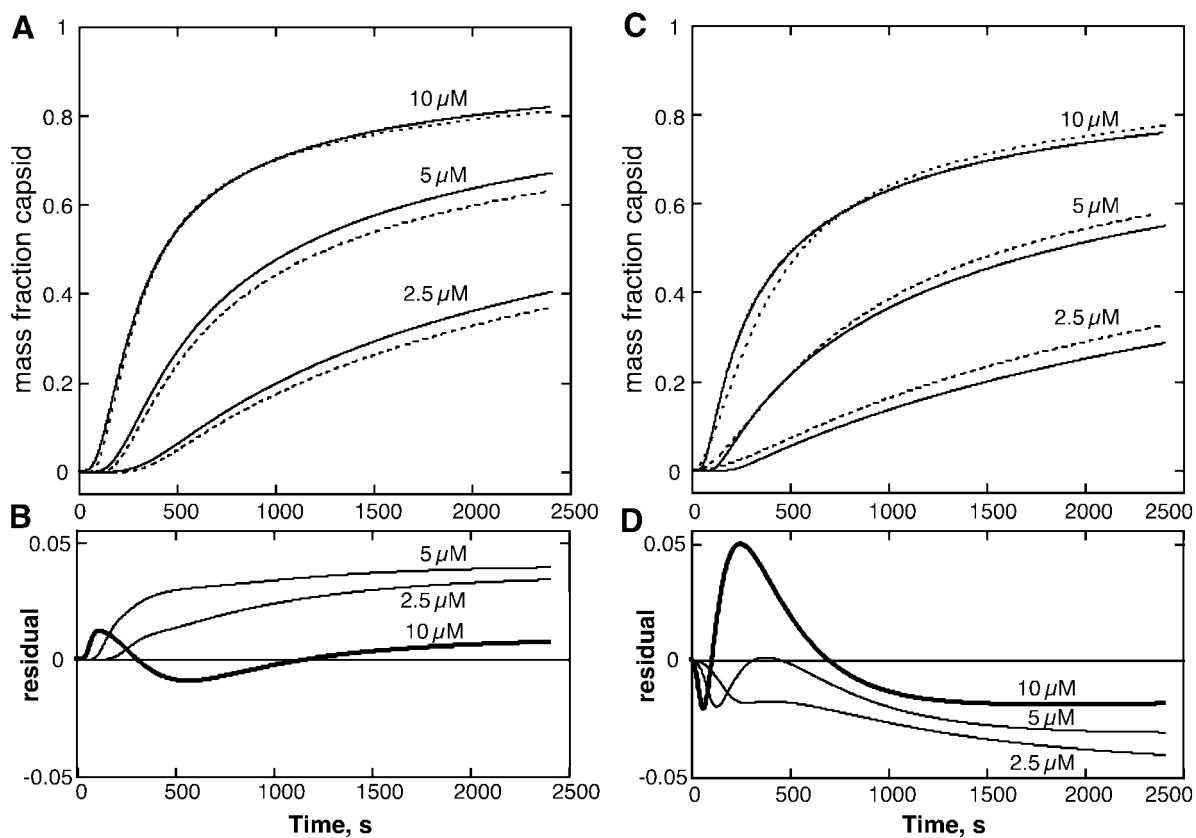


FIGURE 9 Fitting models to synthetic data. Data calculated for a dodecahedron (solid line,  $\log K_{Acon} = -3.3$ ;  $f_{nuc} = 30 \text{ s}^{-1}$ ;  $f_{elong} = 1500 \text{ s}^{-1}$ ) with a 10- $\mu\text{M}$  subunit were fit with an icosahedral model (dashed line,  $\log K_{Acon} = -3.84$ ;  $f_{nuc} = 80 \text{ s}^{-1}$ ;  $f_{elong} = 21,760 \text{ s}^{-1}$ ) (A, B); data calculated for an icosahedron with a 10- $\mu\text{M}$  subunit (solid line,  $\log K_{Acon} = -3.5$ ;  $f_{nuc} = 100 \text{ s}^{-1}$ ;  $f_{elong} = 30,000 \text{ s}^{-1}$ ) were fit with an icosahedral model with the  $\log(K_{Acon})$  restrained to  $-4$  (dashed line,  $\log K_{Acon} = -4.0$ ;  $f_{nuc} = 44 \text{ s}^{-1}$ ;  $f_{elong} = 440,000 \text{ s}^{-1}$ ) (C, D). Kinetic trajectories of the starting model and fit model were also compared for 5  $\mu\text{M}$  and 2.5  $\mu\text{M}$  subunits. Residuals are expressed as the difference between the correct mass fraction capsid and the mass fraction calculated for the wrong model.

tion of a crystal, where a single nucleus leads to formation of a polymer of up to  $10^{12}$  subunits. The general model of capsid assembly described in this paper is a logical outgrowth of our previous work (Zlotnick, 1994; Zlotnick et al., 1999, 2000). To our knowledge, it is the only rigorous description of the thermodynamics and kinetics specific to spherical polymerization. Using quasi-molecular dynamics simulations to examine assembly of individual capsids, Berger and co-workers (Schwartz et al., 1998) were able to reproduce the sigmoidal assembly curves expected for assembly of large populations, essentially a statistical mechanical approach to capsid assembly, though their system was not suitable for developing analyses of experimental data. However, our results are distinctly different from analyses of capsid assembly based on formation of open-ended polymers such as crystals, sheets, or filaments (Oosawa and Asakura, 1975; Prevelige et al., 1993).

The model developed here is completely general. For clarity, we have not incorporated many possible features into the

model. Such mechanisms include autostery (Schwartz et al., 1998; Caspar, 1980), energy penalties for nucleation, cooperativity of association, cooperativity of dissociation, scaffolding proteins, protein-nucleic acid interaction, and the role of post-assembly conformation changes (c.f. Conway et al., 2001). These modifications to the model represent directions for future investigation. Divergence between predicted assembly kinetics and experimental observation is expected if one fails to account for biologically relevant reactions. It is this divergence that will allow investigators to progressively develop and refine accurate descriptions of virus assembly.

These models of assembly improve our ability to understand the behavior of free capsid protein during an infection and the effect of nucleating factors. They are readily applied to analysis of in vitro assembly reactions for virions and other spherical nanoparticles. We can speculate that some day there will be antiviral drugs directed at interfering with normal virus assembly; evaluating such antivirals will require accurate models of the assembly reactions that they inhibit.

## APPENDIX

### Mathematical details

In its most general setting, the model kinetics are described by a system of equations

$$\frac{dX}{dt} = V(X) \quad (\text{A1})$$

where  $X = (X_1, \dots, X_N)$  gives the concentrations of reactants and products and  $V = (V_1, \dots, V_N)$  gives rates of change for the species. The specific form of  $V$  is determined by details of the reaction cascade. We solve Eq. A1 for a time series  $X(t, X_0)$  whose coordinates are the concentrations at a given time, subject to the vector of initial concentrations,  $X_0$ . Generally, Eq. A1 cannot be solved explicitly in terms of elementary functions, so solutions are obtained numerically.

To facilitate developing a simplest-case generalizable model we assume that 1) subunits add one at a time, 2) each species contains a different number of subunits, 3) there is perfect geometry of association where iso-energetic contacts lead to a single product, and 4) all energies are additive. With these assumptions the system of equations A1 becomes

$$\begin{aligned} \frac{d[u]}{dt} &= -f_{\text{nuc}}(2s_2[u]^2 + \dots + s_{\text{nuc}}[u][\text{nuc} - 1]) \\ &\quad - f_{\text{elong}}(s_{\text{nuc}}[u][\text{nuc}] + \dots + s_N[u][N - 1]) \\ &\quad + \nu(2b_2[2] + \dots + b_{\text{nuc}}[\text{nuc}]) \\ &\quad + b_{\text{nuc}+1}[\text{nuc} + 1] + \dots + b_N[N] \\ \frac{d[2]}{dt} &= f_{\text{nuc}}(s_2[u]^2 - s_3[u][2]) + \nu(b_3[3] - b_2[2]) \\ &\quad \vdots \\ \frac{d[\text{nuc}]}{dt} &= f_{\text{nuc}}s_{\text{nuc}}[u][\text{nuc} - 1] - f_{\text{elong}}s_{\text{nuc}+1}[u][\text{nuc}] \\ &\quad + b_{\text{nuc}+1}[\text{nuc} + 1] - \nu b_{\text{nuc}}[\text{nuc}] \\ &\quad \vdots \\ \frac{d[m]}{dt} &= f_{\text{elong}}(s_m[u][m - 1] - s_{m+1}[u][m]) \\ &\quad + b_{m+1}[m + 1] - b_m[m] \\ &\quad \vdots \\ \frac{d[N]}{dt} &= f_{\text{elong}}s_N[u][N - 1] - b_N[N]. \end{aligned} \quad (\text{A2})$$

Back-reaction coefficients  $b_m$  are determined by

$$b_m = \sigma_m f_m K_{\text{Dcon}}^r \quad (\text{A3})$$

with  $r = c_m - c_{m-1}$  giving the number of contacts that must be broken to dissociate a subunit from the  $m$ th intermediate and  $K_{\text{Dcon}} = (K_{\text{Acon}})^{-1}$ .

Equations A2 are solved using the initial condition  $X_0 = (u_0, 0, \dots, 0)$ , i.e., the reaction begins with exclusively free subunits. For each  $u_0 > 0$ , the solution approaches an equilibrium  $X_{\infty}(u_0)$  obtained explicitly by solving the algebraic system  $V(X) = 0$  subject to the equation of conservation of mass

$$u_0 = [u] + \sum_{m=2}^N m[m]. \quad (\text{A4})$$

### Scaling assembly kinetics

The conservation equation A4 is invariant under arbitrary time and mass rescaling. The solutions to the differential equations A2 are not, although they are *nearly* invariant under the transformation

$$(t, X) \mapsto \left( K_{\text{Dcon}} t, \frac{X}{K_{\text{Dcon}}} \right) \quad (\text{A5})$$

(i.e., where time  $t$  is replaced with  $K_{\text{Dcon}} t$ , and concentration  $X$  with  $X/K_{\text{Dcon}}$ ). The transformed equations are identical to Eq. A1 except that an extra factor of  $K_{\text{Acon}} = 1/K_{\text{Dcon}}$  appears in each back-reaction term. For example, the result of this transformation applied to a typical equation from A2,

$$\begin{aligned} \frac{d[m]}{dt} &= f_m(s_m[u][m - 1] - s_{m+1}[u][m]) \\ &\quad + b_{m+1}[m + 1] - b_m[m] \end{aligned}$$

gives rise to

$$\begin{aligned} \frac{d[m]}{dt} &= f_m(s_m[u][m - 1] - s_{m+1}[u][m]) \\ &\quad + K_{\text{Acon}}(b_{m+1}[m + 1] - b_m[m]). \end{aligned} \quad (\text{A6})$$

The back-terms in the transformed equations are least significant when 1) the assembly path distributes at least two contacts to each step, 2)  $K_{\text{Dcon}} \ll 1$ , 3) subunit is not too depleted, and 4)  $\nu \leq K_{\text{Dcon}}$ . The first point follows because the back-terms will contain at least one extra factor of  $K_{\text{Dcon}}$  to cancel the  $K_{\text{Acon}}$  coming from the transformation, whenever there are  $r \geq 2$  contacts formed in a given reaction (see Eq. A3). Consequently, points 2) and 3) follow because the forward terms—invariant under the transformation (identical in A2 and A6)—dominate the rate equations. The requirement described in point 4) is specific for the initial dimerization step, where there is typically only one contact,  $r = 1$ .

Rescaling A2 facilitates comparison of time-series and related quantities for reactions with differing contact energies by using  $K_{\text{Dcon}}$  as the unit of concentration. The rescaling is most nearly exact during the lag, rapid-assembly, and early asymptotic phases, when forward-terms dominate and subunit is not too depleted. After subunit has become depleted—as the reaction approaches equilibrium—back-terms become more significant and the invariance of A2 under the rescaling A5 becomes less exact.

### Isolation of the nucleation rate

After peaking during the reaction's rapid-assembly phase, intermediate concentrations decay exponentially (by standard linearization theorems (Hartman, 1982)) to their equilibrium values. For the  $m$ th intermediate this implies

$$([m] - [m]_{\infty}) \approx [m]_{t_m} e^{-r_m(t-t_m)} \quad (\text{A7})$$

where  $[m]_{\infty}$  and  $[m]_{t_m}$  indicate concentrations at equilibrium and at a time  $t_m$ , after which the approximation is valid to within some prescribed

relative error. Combining the corresponding approximations (Eq. A7) for consecutive intermediates gives (using Eq. 5 and letting the net change in the number of intersubunit contacts  $r = c_m - c_{m-1}$ )

$$\frac{[m]}{[m-1]} \approx \frac{[m]_\infty}{[m-1]_\infty} = K_m[u]_\infty = \frac{s_m}{\sigma_m} K_{\text{Acon}}^r [u]_\infty \quad (\text{A8})$$

with the approximation improving in time  $t > t_m$ . Standard multistep assembly data depicted in Fig. 2 illustrate why  $t_m$  rapidly increases with  $m$ , i.e., higher intermediates can only be built from lower ones, so the approximation (Eq. A8) is accurate much sooner for early intermediates. In particular, Eq. A8 is valid for subnuclear intermediates long before it holds for the nucleus and elongation intermediates. We find then, that upon stabilization of the assembly line, the subnuclear intermediate concentrations will maintain ratios approximating equilibrium conditions that can be written as

$$[nuc - j] \approx K_{\text{nuc}-j} [u] [nuc - 1 - j], \quad j = 1 \cdots nuc - 2. \quad (\text{A9})$$

In addition, for a strongly forward reaction, while  $[u]$  is not too small, i.e., before the later part of the asymptotic phase, the forward reaction terms will tend to dominate the right-hand side of

$$\begin{aligned} \frac{d[nuc]}{dt} &= f_{\text{nuc}} s_{\text{nuc}} [u] [nuc - 1] - f_{\text{elong}} s_{\text{nuc}+1} [u] [nuc] \\ &\quad + b_{\text{nuc}+1} [nuc + 1] - b_{\text{nuc}} [nuc]. \end{aligned} \quad (\text{A10})$$

From the onset of the rapid assembly phase, as long as the mass bound in the assembly line remains approximately constant, intermediate concentrations are in their steady state and, in particular,  $[nuc]$  is nearly constant. Under these conditions the back-terms are negligible. This means that almost every nucleus produced leads to the formation of the next larger intermediate. The same can be said of each supernuclear intermediate as well, so each nucleus leads to a complete capsid. The rate of capsid production is then approximately the same as the forward-rate of nucleus production, and we have

$$\frac{d[N]}{dt} = f_{\text{nuc}} s_{\text{nuc}} [u] [nuc - 1]. \quad (\text{A11})$$

Sequential substitution for  $[nuc-1]$ ,  $[nuc-2]$ ... in Eq. A11 using the approximations A9 gives

$$\frac{d[N]}{dt} \approx f_{\text{nuc}} s_{\text{nuc}} K_{\text{nuc}-1} \cdots K_2 [u]^{nuc}. \quad (\text{A12})$$

The constants  $K_m$  are determined by the capsid geometry and are obtainable by back-substitution in A2 at equilibrium (as given in Eq. 5; also in Zlotnick, 1994). When subunit becomes depleted, the steady-state assembly line and the estimate A11 gradually fail.

## Estimation of nucleus size

Our remaining work concerns dependence of  $[u]$  and  $[N]$  on both  $t$  and  $u_0$ . In the interest of clarity, we will begin to indicate differentiations with respect to time as partial derivatives. In keeping with convention we will not explicitly indicate such dependence (as  $[u](u_0, t)$  and  $[N](u_0, t)$ ), because it is understood implicitly.

We observe that simulated reactions, initiated with an ample supply of subunit, rapidly enter a steady state in which intermediate concentrations are slowly varying. Under such conditions A12 holds and the ratio

$$\lambda_p = \frac{\partial [N]}{\partial t} / [u]^p \quad (\text{A13})$$

is independent of  $u_0$  for  $p = nuc$ . Furthermore, the mass bound in the assembly line is nearly constant so that the mass not in the assembly line (capsid and subunit) is nearly constant as well, giving

$$\frac{\partial}{\partial t} (N[N] + [u]) \approx 0. \quad (\text{A14})$$

In simulations the accuracy of the approximation A14 is also independent of  $u_0$  for  $u_0$  sufficiently large to drive assembly (for reactions with  $u_0 \leq K_{\text{Dapp}}$  there is no stable size for the assembly line as there is no rapid-assembly phase). Solving A14 for  $\partial[u]/\partial t$ , dividing through by  $[u]^p$  and differentiating with respect to  $u_0$ , we obtain

$$\frac{\partial}{\partial u_0} \left( N \frac{\partial [N]}{\partial t} / [u]^p \right) = - \frac{\partial}{\partial u_0} \left( \frac{\partial [u]}{\partial t} / [u]^p \right) = 0$$

leading to the conditions

$$\begin{aligned} N [u]^p \frac{\partial}{\partial u_0} \frac{\partial [N]}{\partial t} &= N p [u]^{p-1} \frac{\partial [u]}{\partial u_0} \frac{\partial [N]}{\partial t} \\ &= -p [u]^{p-1} \frac{\partial [u]}{\partial u_0} \frac{\partial [u]}{\partial t} \\ &= -[u]^p \frac{\partial}{\partial u_0} \frac{\partial [u]}{\partial t}. \end{aligned}$$

Under these conditions we also have

$$\frac{\partial}{\partial u_0} \left( \frac{\partial [N]}{\partial t} / [N] \right) = \frac{1}{[N]^2} \left( p [N] \frac{\partial [u]}{\partial u_0} - \frac{\partial [N]}{\partial u_0} \right) \frac{\partial [u]}{\partial t}$$

and

$$\frac{\partial}{\partial u_0} \left( \frac{\partial [u]}{\partial t} / [u] \right) = \frac{p-1}{[u]^2} \frac{\partial [u]}{\partial u_0} \frac{\partial [u]}{\partial t}.$$

Forming their ratio and exchanging the order of mixed partials we find

$$\begin{aligned} \left( \frac{\partial^2 \ln [N]}{\partial t \partial u_0} / \frac{\partial^2 \ln [u]}{\partial t \partial u_0} \right) &= \frac{[u]}{(p-1)[N]} \left( p - \left( \frac{\partial \ln [N]}{\partial u_0} / \frac{\partial \ln [u]}{\partial u_0} \right) \right). \end{aligned} \quad (\text{A15})$$

Suppose for the moment that for each  $u_0$ , at a particular time  $t = t(u_0)$ , that

$$\left( \frac{\partial^2 \ln [N]}{\partial t \partial u_0} / \frac{\partial^2 \ln [u]}{\partial t \partial u_0} \right) = \frac{\partial \ln [N]}{\partial u_0} / \frac{\partial \ln [u]}{\partial u_0}. \quad (\text{A16})$$

As a consequence, momentarily at  $t = t(u_0)$ , by virtue of A16, Eq. A15 becomes

$$\left( \frac{\partial \ln[N]}{\partial u_0} / \frac{\partial \ln[u]}{\partial u_0} \right) = \frac{[u]}{(p-1)[N]} \left( p - \left( \frac{\partial \ln[N]}{\partial u_0} / \frac{\partial \ln[u]}{\partial u_0} \right) \right).$$

which can be solved for

$$\begin{aligned} \frac{\partial \ln[N]}{\partial u_0} / \frac{\partial \ln[u]}{\partial u_0} &= p \frac{[u]}{[u] + (p-1)[N]} \\ &= p \left( 1 - \frac{(p-1)[N]}{[u]} + \left( \frac{(p-1)[N]}{[u]} \right)^2 - \dots \right) \end{aligned}$$

so that the approximation

$$L = \left( \frac{\partial \ln[N]}{\partial u_0} / \frac{\partial \ln[u]}{\partial u_0} \right) \approx p \quad (\text{A17})$$

holds to within  $p(p-1)[N]/[u]$ , if  $[u] < (p-1)[N]$  at  $t = t(u_0)$ . The value  $p = nuc$  is the desired nucleus size.

How can we find this special time  $t(u_0)$  when this approximation holds? As it turns out,  $p$  is the temporal minimum value of the function

$$L = \left( \frac{\partial \ln[N]}{\partial u_0} / \frac{\partial \ln[u]}{\partial u_0} \right)$$

achieved during steady-state at the time  $t = t(u_0)$ . To see that this is the case we differentiate  $L$  with respect to  $t$ :

$$\frac{\partial L}{\partial t} = \frac{\frac{\partial \ln[u]}{\partial u_0} \frac{\partial^2 \ln[cap]}{\partial t \partial u_0} - \frac{\partial \ln[cap]}{\partial u_0} \frac{\partial^2 \ln[u]}{\partial t \partial u_0}}{\left( \frac{\partial \ln[u]}{\partial u_0} \right)^2} \quad (\text{A18})$$

This derivative (A18) will equal 0 at the time  $t = t(u_0)$  when the temporal minimum of  $L$  is reached by the reaction with initial concentration  $u_0$ . Consequently, the numerator of A18 must be zero, leaving A16, concluding our analysis.

Conditions under which  $\lambda_p$  (A13) is independent of  $u_0$  persist over a wide temporal interval, but the onset of this interval is delayed as  $u_0$  decreases. When  $u_0$  is small compared to  $K_{Dapp}$ , the assembly line fails to persist. For A14 to hold, the total mass bound in intermediate species must be held constant.

Assembly cascades are susceptible to both starvation and kinetic trapping. For  $[u] < K_{Dapp}$  reactions are starved for available subunits; for  $[u]$  sufficiently large, reactions are well-supplied with available free subunits. For  $u_0 \gg K_{Dapp}$ , free subunits are abundant ab initio unless assembly is over-initiated, causing a kinetic trap. In a starved reaction there is no surplus of subunit mass to be transferred through the assembly line, so dissociation of intermediates becomes a significant source of mass for capsid production. Fortunately, concentrations for which A14 can be expected to hold are just those concentrations that lead to the experimentally measurable values of capsid and subunit required to calculate the value of  $L$ .

This work was supported by the American Cancer Society (Research Project Grant RPG-99-339-01) and the National Science Foundation (MCB0111025) to A.Z. D.E. was also supported by CGS&R at the University of Central Oklahoma.

## REFERENCES

- Adolph, K. W., and P. J. Butler. 1976. Assembly of a spherical plant virus. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 276:113–122.
- Bancroft, J. B., and E. Hiebert. 1967. Formation of an infectious nucleoprotein from protein and nucleic acid isolated from a small spherical virus. *Virology*. 32:354–356.
- Bancroft, J. B., E. Hiebert, and C. E. Bracker. 1969. The effects of various polyanions on shell formation of some spherical viruses. *Virology*. 39:924–930.
- Bancroft, J. B., G. J. Hills, and R. Markham. 1967. A study of the self-assembly process in a small spherical virus. Formation of organized structures from protein subunits in vitro. *Virology*. 31:354–379.
- Bancroft, J. B., G. W. Wagner, and C. E. Bracker. 1968. The self-assembly of a nucleic-acid free pseudo-top component for a small spherical virus. *Virology*. 36:146–149.
- Bartenschlager, R., and H. Schaller. 1992. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* 11:3413–3420.
- Basavappa, R., R. Syed, O. Flore, J. P. Icenogle, D. J. Filman, and J. M. Hogle. 1994. Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: structure of the empty capsid assembly intermediate at 2.9 Å resolution. *Protein Sci.* 3:1651–1669.
- Beckett, D., H. N. Wu, and O. C. Uhlenbeck. 1988. Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* 204:939–947.
- Berger, B., P. W. Shor, L. Tucker-Kellogg, and J. King. 1994. Local rule-based theory of virus shell assembly. *Proc. Natl. Acad. Sci. U.S.A.* 91:7732–7736.
- Caspar, D. L. 1980. Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophys. J.* 32:103–138.
- Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* 27:1–24.
- Conway, J. F., W. R. Wikoff, N. Cheng, R. L. Duda, R. W. Hendrix, J. E. Johnson, and A. C. Steven. 2001. Virus maturation involving large subunit rotations and local refolding. *Science*. 292:744–748.
- Dill, K. A. 1999. Polymer principles and protein folding. *Protein Sci.* 8:1166–1180.
- Dinner, A. R., A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* 25:331–339.
- Fersht, A. R. 1999. Structure and Mechanism in Protein Science. W. H. Freeman and Co., New York.
- Flyvbjerg, H., E. Jobs, and S. Leibler. 1996. Kinetics of self-assembling microtubules: an “inverse problem” in biochemistry. *Proc. Natl. Acad. Sci. U.S.A.* 93:5975–5979.
- Fox, J. M., G. Wang, J. A. Speir, N. H. Olson, J. E. Johnson, T. S. Baker, and M. J. Young. 1998. Comparison of the native CCMV virion with in vitro assembled CCMV virions by cryoelectron microscopy and image reconstruction. *Virology*. 244:212–218.
- Hartman, P. 1982. Ordinary Differential Equations. Birkhauser, Boston.
- Oosawa, F., and S. Asakura. 1975. Thermodynamics of Polymerization of Protein. Academic Press, London.
- Prevelige, P. E. J., D. Thomas, and J. King. 1993. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* 64:824–835.
- Rombaut, B., R. Vrijnsen, and A. Boeye. 1990. New evidence for the precursor role of 14 S subunits in poliovirus morphogenesis. *Virology*. 177:411–414.
- Rueckert, R. R. 1996. Picornaviridae: the viruses and their replication. *In* Fields Virology. B. N. Fields, D. M. Knipe, P. M. Howley, R. M.

- Chanock, J. L. Melnick, T. P. Monath, B. Roizman, and S. E. Straus, editors. Lippincott-Raven, Philadelphia. 609–654.
- Salunke, D., D. L. Caspar, and R. L. Garcea. 1989. Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* 56:887–900.
- Schwartz, R., P. W. Shor, P. E. J. Prevelige, and B. Berger. 1998. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* 75:2626–2636.
- Seifer, M., S. Zhou, and D. N. Standring. 1993. A micromolar pool of antigenically distinct precursors is required to initiate cooperative assembly of hepatitis B virus capsids in *Xenopus* oocytes. *J. Virol.* 67:249–257.
- Sorger, P. K., P. G. Stockley, and S. C. Harrison. 1986. Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *J. Mol. Biol.* 191:639–658.
- Tanford, C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*. John Wiley and Sons, Inc., New York.
- van den Worm, S. H., N. J. Stonehouse, K. Valegard, J. B. Murray, C. Walton, K. Fridborg, P. G. Stockley, and L. Liljas. 1998. Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. *Nucleic Acids Res.* 26:1345–1351.
- Zhao, X., J. M. Fox, N. H. Olson, T. S. Baker, and M. J. Young. 1995. In vitro assembly of cowpea chlorotic mottle virus from coat protein expressed in *Escherichia coli* and in vitro-transcribed viral cDNA. *Virology.* 207:486–494.
- 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., R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young. 2000. Mechanism of capsid assembly for an icosahedral plant virus. *Virology.* 277:450–456.
- Zlotnick, A., N. Cheng, J. F. Conway, F. P. Booy, A. C. Steven, S. J. Stahl, and P. T. Wingfield. 1996. Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry.* 35:7412–7421.
- Zlotnick, A., J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres. 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry.* 38:14644–14652.