*J. Mol. Biol.* (1994) **241**, 59–67

# To Build a Virus Capsid

## An Equilibrium Model of the Self Assembly of Polyhedral Protein Complexes

**Adam Zlotnick**

*Department of Biological Sciences*
*Purdue University*
*West Lafayette, IN 47907, U.S.A.*

The capsids of spherical (icosahedral) viruses are constructed of multiples of 60 subunits. The question of how these polymers assemble is basic to understanding the viral life cycle. A formalism describing virus assembly as an equilibrium between coat protein subunits, assembly intermediates and intact virus is presented. This equilibrium model of virus assembly is consistent with experimental observations of virus assembly. At equilibrium, either intact virus or free subunits are dominant species, assembly intermediates are predicted to be found only in trace concentrations. The concentration of assembled virus at equilibrium is expected to be extremely concentration-dependent and resemble a highly cooperative reaction although the model does not explicitly include cooperativity. For statistical assembly of a polyhedron, a nucleus is not necessarily required and polymerization can proceed through a cascade of bimolecular reactions rather than a single higher order reaction. Thus, kinetics of assembly do not necessarily show the extreme concentration dependence typical of nucleated protein polymerization. Modest intersubunit interaction energies result in a very stable capsid; consequently, a small change in this interaction energy can result in a considerable change in the capsid-subunit equilibrium. Some possible effects of nucleation and protein-nucleic acid interactions on virus assembly and capsid morphology are considered.

*Keywords:* virus assembly; protein polymerization; protein folding; morphogenesis

### 1. Introduction

Simple icosahedral viruses are complexes composed of the viral genome (or some fragment of nucleic acid) packaged in a spherical protein capsid. Efficient use is made of the viral genome coding for the capsid protein by using multiple copies of protein(s) in a geometric arrangement.

An icosahedral virus capsid is composed of 60 icosahedral asymmetric units each of which may be composed of one or several proteins. The triangulation number, $T$, is the number of identical proteins within each icosahedral asymmetric unit. Thus, there are $60T$ quasi equivalent subunits per virus capsid. The capsid surface is formed by an arrangement of protein pentamers and hexamers, resembling a geodesic dome. The $T$ proteins in each icosahedral asymmetric unit are situated in $T$ unique, yet quasi equivalent environments (Caspar & Klug, 1962).

Capsid proteins from many viruses can reversibly assemble and disassemble in response to solution conditions. *In vitro* capsid assembly was first observed in cowpea chlorotic mottle virus (CCMV†; Bancroft, 1970). The coat protein of CCMV, a $T = 3$ RNA virus, can form $T = 3$ particles, $T = 1$ particles, multilayered particles, and long tubes (Bancroft, 1970; Butler, 1979). Other examples include turnip crinkle virus (TCV) coat protein which reassembles into native-like $T = 3$ particles in the presence of RNA, or $T = 1$ particles if the basic, RNA binding N terminus is proteolytically removed (Sorger *et al.*, 1986). Southern bean mosaic virus switches between $T = 1$ and $T = 3$ particles during RNA-dependent reassembly in response to a change in pH (Savithri & Erickson, 1983). Pentamers of polyoma virus coat protein assemble into $T = 1$ particles, octahedral particles, native-like pseudo $T = 7$ particles, amongst a plethora of aggregates; the distribution depending on solution conditions (Salunke *et al.*, 1986, 1989). Bacteriophage P22 reassembles into a $T = 7$ particle in a kinetically

† Abbreviations used: CCMV, cowpea chlorotic mottle virus; TCV, turnip crinkle virus.



0022-2836/94/310059-09 $08.00/0

complex manner requiring a scaffolding protein (Prevelige *et al.*, 1993).

From an experimental standpoint, under given solution conditions, viruses are either intact or completely disassembled into stable poly- or monomeric subunits. Half assembled viruses are only found under peculiar circumstances. From a thermodynamic standpoint, a virus that can be reversibly assembled must be either at or approaching equilibrium with intermediates and component subunits.

Spontaneous assembly of viruses, *in vitro*, is often compared to nucleated assembly of linear polymers. Formation of the nucleus, consisting of several monomers, is the rate limiting step to polymerization. Addition of monomers to the nucleus can continue indefinitely. The end residues of such a polymer are equivalent, no matter how long the polymer. All of the interior residues are approximately equivalent. This is not the case during assembly of a spherical virus. The terminal residues of a partially built virus are not necessarily equivalent, an intact virus has no terminal termini, the size of the polymer is not indefinite, and there is not necessarily a well defined nucleus. Classical nucleated assembly seems to be an inappropriate description of virus assembly. This paper describes a formalism, derived for a spherical polymer and assembly intermediates at equilibrium, which may be appropriate for viruses and other topologically closed polymers. Unlike nucleated assembly, equilibrium assembly may not have a critical concentration and the kinetics of assembly do not necessarily show the extreme concentration dependence associated with nucleation.

## 2. Theory

### (a) *Basic model and assumptions*

The number of protein subunits and geometry of the building blocks is as important a consideration as the geometry of the final product when describing the assembly of an icosahedral virus. A dodecahedron built of 12 pentagonal subunits, reminiscent of picornaviruses such as rhinovirus and poliovirus (Rueckert, 1990), will be used to illustrate a basic equilibrium model of viral assembly. The picornaviruses are assembled from 12 pentamers. Each pentamer consists of five icosahedral asymmetric units, each asymmetric unit consisting of three proteins.

In the model system, a subunit is defined as a stable building block in the assembly process; in the context of a virus such an assembly subunit is often a protein oligomer. The contact between the edges of adjacent subunits, at a dodecahedral 2-fold axis, is considered to have $\Delta G_c^\circ$ free energy; all 30 edge to edge contacts are identical. Only ideal dodecahedral geometry is allowed for each intermediate. Neither cooperativity nor nucleation are built into this model.

Describing virus assembly as a series of equilibrium reactions allows evaluation of the concentration of each species in solution. For $\Delta G_c^\circ \ll 0$ kcal/mole, the concentration of only the most stable intermediates, those which have the greatest number of intersubunit contacts, must be evaluated to determine the distribution of species. The association constant, $K_n$, of a species $\boldsymbol{n}$, a specific arrangement of n subunits, can be separated into statistical components, $S_n$ and $S_{in}$, and a non-statistical association constant $K'_n$ (equation (1)). The statistical factor, $S_n$, can be treated as the ratio of two factors: the number of ways of forming a species, $\boldsymbol{n}$, from specific intermediates and of dissociating the product to produce equivalent reactants. Table 1 gives $S_n$ for formation of the most stable species $\boldsymbol{n}$ from its most stable predecessor, $(\boldsymbol{n-1})$, and $\boldsymbol{1}$, free subunit. For the dodecahedron used in this example, the five equivalent orientations for each incoming monomer create the second statistical factor describing the degeneracy of the incoming subunit, $S_{in} = 5$. $K'_n$ is a function of the number of contacts formed without including the degeneracy of the interaction (Table 1).

$$K_n = \frac{[\boldsymbol{n}]}{[\boldsymbol{n-1}][1]} = S_{in} S_n K'_n \qquad (1)$$

For example the evaluation of $K_3$ is described. There are two ways to add a subunit, *1*, to species *2* to form species *3* (Table 1) and there are three ways to dissociate *3* back to *1* and *2*, thus $S_3 = 2/3$. Two intersubunit contacts are made in forming *3*, resulting in a free energy of $2\Delta G_c^\circ$ and $K'_3 = e^{-2\Delta G_c^\circ/RT}$. Thus $K_3 = (5)(2/3)e^{-2\Delta G_c^\circ/RT}$, where $R$ is the gas constant, 1·987 cal deg$^{-1}$ mole$^{-1}$, and $T$ is the temperature in Kelvin, set to 298 K.

The equilibrium concentration of each species can be evaluated in terms of subunit concentration. Equation (1) can be rearranged to solve for the concentration of intermediate species $\mathbf{n}$ (equation (2)). The concentration of the species $(\mathbf{n-1})$ can be serially substituted by equations describing the concentration of the preceding intermediates.

$$[\boldsymbol{n}] = S_{in} S_n K'_n [\boldsymbol{n-1}][1] = \left( \prod_{i=2}^{n} S_{in} S_i K'_i \right) [1]^n \qquad (2)$$

In equation (2) it is apparent that the concentration of each species is a function of the concentration of monomer and free energy of contact formation. A striking result of this model is that for $\Delta G_c^\circ \ll 0$ kcal/mole either subunit or intact dodecahedron are the dominant species, intermediates are present in trace concentrations (Figure 1, Table 2). Three concentrations of free subunit were chosen to illustrate the paucity of intermediates as the equilibrium shifts from nearly all subunit to nearly all intact virus. The data presented in Figure 1, without knowledge of how they were derived, would falsely suggest that this is an equilibrium involving only a free subunit and dodecahedron, with no stable intermediates.

The concentration of dodecahedra is very sensitive to the concentration of free subunit (Figure 2).

**Table 1**
*Assembly intermediates and factors describing their assembly*

| n | Model | Build Up | Build Down | $S_n$ | c | $K'_n$ |
|---|---|---|---|---|---|---|
| 1 |  | — | — | — | — | — |
| 2 |  | 5/2† | 1 | 5/2 | 1 | $e^{-1\Delta G^\circ_c/RT}$ |
| 3 |  | 2 | 3 | 2/3 | 2 | $e^{-2\Delta G^\circ_c/RT}$ |
| 4 |  | 3 | 2 | 3/2 | 2 | $e^{-2\Delta G^\circ_c/RT}$ |
| 5 |  | 4 | 2 | 4/2 | 2 | $e^{-2\Delta G^\circ_c/RT}$ |
| 6 |  | 1 | 5 | 1/5 | 3 | $e^{-3\Delta G^\circ_c/RT}$ |
| 7 |  | 5 | 1 | 5/1 | 2 | $e^{-2\Delta G^\circ_c/RT}$ |
| 8 |  | 2 | 4 | 2/4 | 3 | $e^{-3\Delta G^\circ_c/RT}$ |
| 9 |  | 2 | 3 | 2/3 | 3 | $e^{-3\Delta G^\circ_c/RT}$ |
| 10 |  | 3 | 2 | 3/2 | 3 | $e^{-3\Delta G^\circ_c/RT}$ |
| 11 |  | 2 | 5 | 2/5 | 4 | $e^{-4\Delta G^\circ_c/RT}$ |
| 12 |  | 1 | 12 | 1/12 | 5 | $e^{-5\Delta G^\circ_c/RT}$ |

$S_n$, statistical factor for a particle of **n** subunits; c, contacts between intermediates.

Intermediates in viral assembly. Only the most stable intermediate is illustrated. The statistical weighting for [**n**] is related to the ratio of the number of ways to "build up" from and "build down" to defined reactants, in this case monomer and the next smallest illustrated species. The non-statistical association constant, $K'_n$, is a function of the number of contacts formed and $\Delta G^\circ_c$. A statistical factor, $S_{in} = 5$, for the degeneracy of the incoming subunit is also required for calculation of the association constant for formation of the illustrated species *n*, $K_n = S_{in}S_nK'_n$.

† For conceptual simplicity, the statistical weighting is described as the number of ways a given reaction may occur. The build-up factor of 5/2, instead of the intuitively obvious factor of 5, is required to accurately account for the concentrations of all species. The factor of 1/2, analogous to the symmetry factor, $\sigma$, in the calculation of the entropy of a diatomic gas, is due to the degeneracy of the dimer formed in this region. The degeneracy of other reactants and products in this Table is accurately defined in terms of the number of paths for a given reaction.

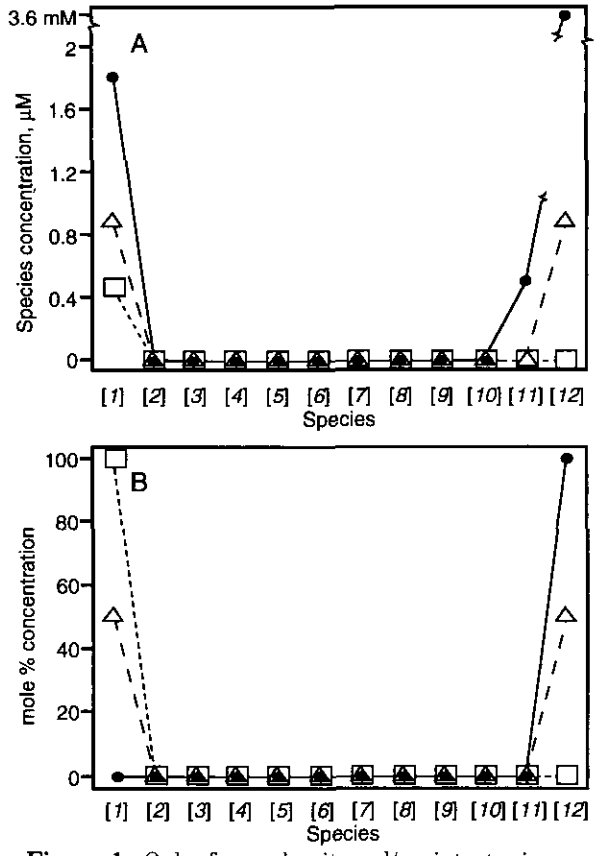


**Figure 1.** Only free subunit and/or intact virus are present in appreciable concentrations. Data are from Table 2, calculated according to equation (2) using a $\Delta G^\circ_c$ of −2·76 kcal/mole statistical weights from Table 1. The concentration of each species, in brackets, is calculated for free subunit concentrations of 0·44 μM (G), 0·88 μM (C), and 1·8 μM (J). A, The concentrations are plotted against a μM scale, note the change of scale necessary to include **12** for 1·8 μM free subunit. B, Plotting the mole fraction of each species reveals the near homogeneity of the mixtures with 0·44 μM and 1·8 μM free subunit. By definition at the $K_{d_{app}}$ 0·88 μM *1*, there are equal concentrations of *1* and *12*.

**Table 2**
*The concentration of assembly intermediates at three concentrations of* **1**

| Species | Concentration (M) | | |
|---|---|---|---|
| *1* | 0·44E-6 | 0·88E-6 | 1·8E-6 |
| *2* | 2E-10 | 1E-9 | 5E-9 |
| *3* | 4E-12 | 3E-11 | 2E-10 |
| *4* | 1E-13 | 2E-12 | 3E-11 |
| *5* | 5E-15 | 2E-13 | 5E-12 |
| *6* | 2E-15 | 1E-13 | 1E-11 |
| *7* | 2E-16 | 3E-14 | 4E-12 |
| *8* | 2E-16 | 7E-14 | 2E-11 |
| *9* | 5E-16 | 2E-13 | 1E-10 |
| *10* | 1E-15 | 1E-12 | 1E-9 |
| *11* | 1E-13 | 2E-10 | 5E-7 |
| *12* | 2E-10 | 0·88E-6 | 3·6E-3 |

Equilibrium concentration of the most stable example of each species calculated according to equation (3) using $\Delta G^\circ_c = -2{\cdot}72$ kcal/mole, corresponding to a $K_d$ of 10 mM for a single contact. Concentrations of *1* are approximately 1/2x, 1x and 2x the $K_{d_{app}}$ of 0·88 μM.

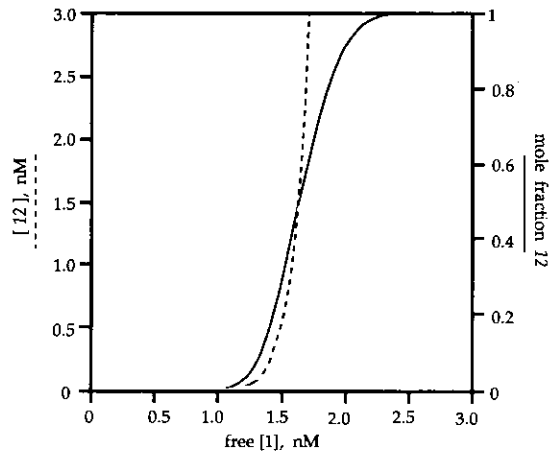


**Figure 2.** The effect of subunit concentration on assembly. The concentration of *12*, intact virus, is calculated for free subunit, *1*, according to equation (3). Data are plotted in terms of the molar concentration of *12* (broken line) and the mole fraction of *12* (continuous line) in solution. The free energy used in equation (3) is $\Delta G_c^\circ = -4{\cdot}08$ kcal/mole which corresponds to a $K_{d^{app}}$ of 1·65 nM. At 1·65 nM monomer, the $K_{d^{app}}$, intact virus makes up 50 mole % of the material in solution.

This is anticipated from equation (3) for the concentration of dodecahedron, [*12*], derived from equation (2).

$$[12] = \left(\prod_{i=2}^{12} S_{in} S_n K_i'\right)[1]^{12}$$

$$= 5^{11}\frac{1}{12}\,e^{-30\Delta G_c^\circ/RT}[1]^{12}. \quad (3)$$

It is convenient to define an operational relationship between subunit and dodecahedron as the apparent $K_d$, $K_{d^{app}}$, when [*1*] = [*12*]. The concentrations of free subunit in Figure 1 were chosen to be $1/2K_{d^{app}}$, $K_{d^{app}}$ and $2K_{d^{app}}$ for the selected $\Delta G_c^\circ$ of $-2{\cdot}72$ kcal/mole. In an experimental test, a value near $K_{d^{app}}$ could be mistaken for a critical concentration of nucleated assembly. $K_{d^{app}}$ is not the inverse of the association constant defined in equation (3). There is a linear relationship between the free energy of a contact, $\Delta G_c^\circ$, and the overall subunit-dodecahedron equilibrium described by $\Delta G_{app} = -RT\ln(K_{d^{app}})$. This relationship can be seen by rearranging the natural log of equation (3), setting [*12*] and [*1*] to $K_{d^{app}}$.

$$\Delta G_{app} = (30\Delta G_c^\circ - RT\ln(5^{11}/12))/11. \quad (4)$$

The denominator of equation (3) is the number of subunits in the complete polyhedron less one, the numerator is the total bonding free energy and statistical entropy. For a polyhedron with many subunits the value of $\Delta G_{app}$ approaches the average association energy per subunit of the intact capsid. Note that the energy for incorporating the last subunit into the dodecahedron is $5\Delta G_c^\circ$ multiplied by the statistical terms. Statistical terms will generally be small compared to the bonding term, thus $\Delta G_{app}$ is approximately one half the association energy holding a single subunit into the capsid ($\Delta G_{app} = 2{\cdot}73\Delta G_c^\circ - 0{\cdot}84$ kcal/mole). (The factor of 1/2 arises because each subunit contributes 1/2 the associative surface of each bond it forms.) Because of the steep concentration dependance of assembly, a small change in the concentration of free subunit has a dramatic effect on the concentration of intact virus. Conversely, a small change in $\Delta G_c^\circ$ can result in a significant change in virus stability.

A general formula for the concentration of a regular structure of $N$ monomers, which make $i$ equivalent contacts, is given in equation (5).

$$[N-\text{mer}] = i^{(N-1)}\frac{1}{N}\,e^{-\frac{iN\Delta G_c^\circ}{2RT}}[1]^N. \quad (5)$$

For the dodecahedron used in this example $N = 12$ and $i = 5$. The factor $1/N$ in equation (5) is the statistical weighting for platonic solids assembled from equivalent subunits, such as a dodecahedron built of pentagons. The statistical factors when there are non-equivalent subunits, different proteins, or a specific nucleus must be determined on a case by case basis.

Each intermediate in the assembly of a spherical polymer is different and generally more stable than its predecessor. In the intact dodecahedron, or virus, all subunits are fully ligated in a lowest energy environment. This is unlike linear polymers where the monomers at the ends of the polymer are equivalent and internal monomers are equivalent.

### (b) *Some kinetic implications of equilibrium assembly*

Assembly of viruses requires a cascade of bimolecular, or higher order, reactions. It is implicit in the basic assembly model that at very low initial subunit concentrations, the concentration of the early intermediates of assembly may be very low. Thus, a lag phase may be a dominant feature of assembly. The kinetic implications of equilibrium assembly are tested in a series of simulations, varying initial concentrations of free subunit.

Knowledge of the detailed pathway is not required in a discussion of the thermodynamics of assembly. In order to simulate the kinetics of assembly a pathway, or pathways, must be specified. In a simplest case simulation, assembly proceeds through the most stable intermediates by addition of monomer. Assembly by association of two or more intermediates is not considered; nor is the effect of other, less stable, intermediates. Thus, the kinetic simulation requires only 11 simultaneous bimolecular interactions for the assembly of dodecahedron from monomer. Assembly is treated as a diffusion limited reaction. All on-rates, $k_{on}$, are $10^8\ M^{-1}\ s^{-1}$, a value that is close to the diffusion limited association of two proteins. Off-rates, $k_{off}$, are determined from $k_{on}/K_n = k_{off}$, where $K_n$ is the association constant for species **n** defined in

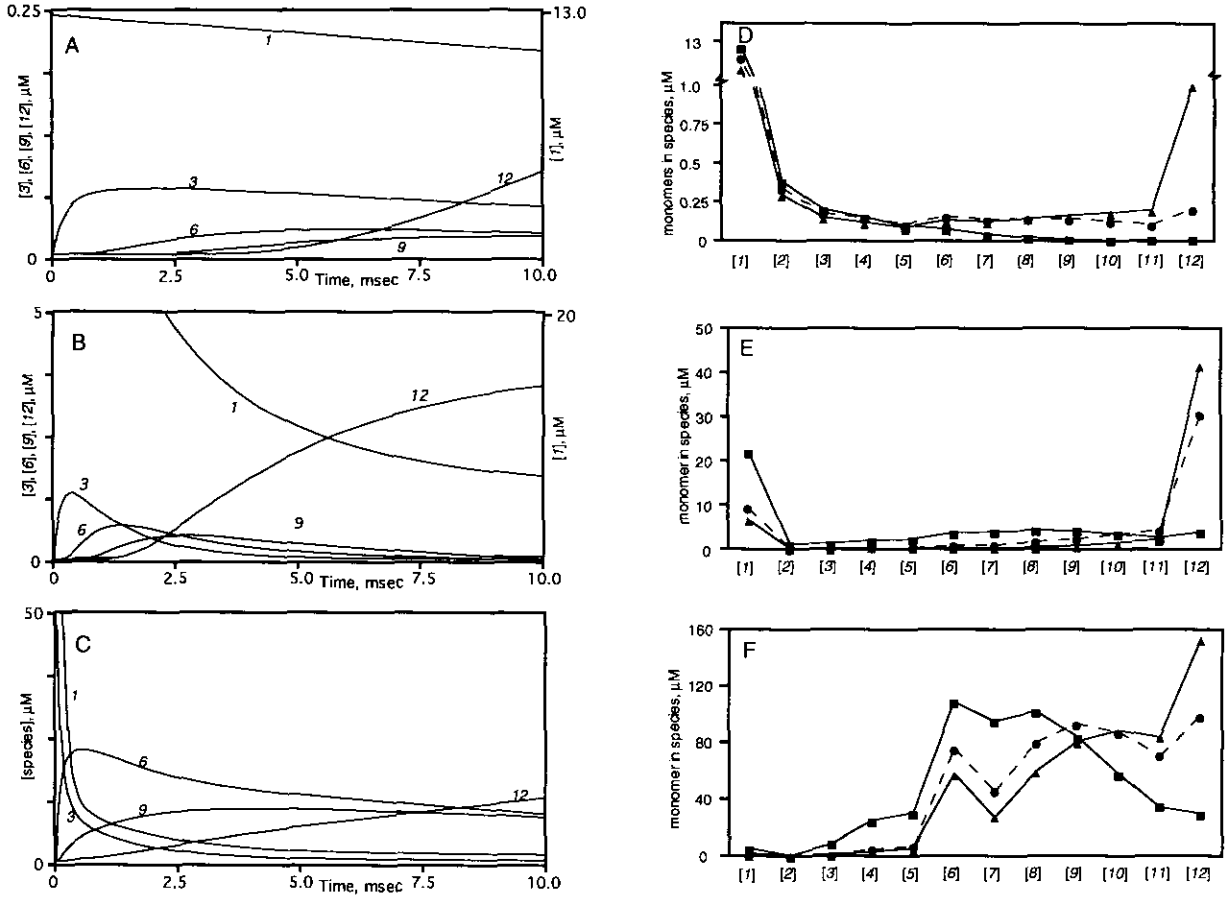


**Figure 3.** Kinetic simulations of icosahedron assembly. Assembly proceeds through only the most stable intermediates and only by addition of monomer, as described in the text. A, B, C, The change in concentration of some assembly reactants with time. Initial free subunit concentrations are 13 $\mu$M (A), 50 $\mu$M (B) and 500 $\mu$M (C). Only time courses for 5 species, free subunit (*1*), trimer (*3*), hexamer (*6*), nonamer (*9*) and dodecahedron (*12*), are shown for clarity. D, E, F, The concentration of each assembly intermediate at specific times during assembly: 2 ms (■), 4 ms (●) and 10 ms (▲). Initial free subunit concentrations are 13 $\mu$M (D), 50 $\mu$M (E) and 500 $\mu$M (F). In these last 3 graphs (D, E, F), concentrations are given as the number of subunits incorporated into each species; thus, the concentration of **12** in D, E and F is 12 times the molar concentration shown in A, B and C.

equation (2). The rate equations were solved by numerical integration using a fourth order Runge-Kutta approximation as implemented in the program STELLA II (High Performance Systems, Hartford, NH). Simulations run at three concentrations are shown in Figure 3; concentrations were chosen so that, at equilibrium, final dodecahedron concentrations were approximately one, four, and 40 times the $K_{d^{app}}$ of 0·88 $\mu$M for $\Delta G_c = -2{\cdot}72$ kcal/mole.

Diffusion limited assembly of dodecahedra is a fast reaction. At moderate initial subunit concentrations, 50 $\mu$M free subunit, a short lag time of about 1 ms is observed before the appearance of complete dodecahedron (Figure 3B). During the lag phase, smaller intermediates are assembled and progressively consumed in the assembly of larger species until finally dodecahedron accumulates (Figure 3E). Nearly 90% of the equilibrium concentration of dodecahedron is found after only 10 milliseconds. Equilibrium is approached asymptotically as the concentration of intermediates decreases. At lower initial concentration, the lag time is longer, about 5 ms, and the rate of appearance of dodecahedron is predictably slower (Figure 3A, D); 90% of the equilibrium concentration is not reached till 100 ms.

At high initial concentrations of free subunit the kinetics are more complicated. The lag time is shorter at 500 $\mu$M initial subunit concentration than at 50 $\mu$M, but it takes longer to approach equilibrium (Figure 3C). The concentration of dodecahedron reaches 90% of the equilibrium concentration only after 40 milliseconds. This is due to the limitations of the assembly pathway chosen. Moderate sized intermediates are assembled very rapidly, consuming most of the free subunit (Figure 3F). Further assembly requires scavenging free subunits by disassembly of some intermediates. If the assembly pathway had included a reaction allowing association of intermediates this kinetic barrier could have been avoided. Incorrect association of intermediates could result in the occurrence of the "monster" particles observed in some *in vitro* assembly reactions (Sorger *et al.*, 1986). *In vivo*,

where there is a supply of freshly translated coat protein subunits, a kinetic barrier due to lack of free subunit is probably irrelevant.

When $\Delta G_c$ is greater, $-5{\cdot}45$ kcal/mole, assembly can proceed, albeit slowly, at much lower subunit concentrations. At higher subunit concentrations, the problem where most of the free subunit becomes kinetically trapped in intermediates is more important. The greater energy of association decreases the *availability of free subunit very early in the* assembly pathway, similar to the example in Figure 3C and F. The higher association energy also slows the disassembly of the excess intermediates which is required to produce free subunit. Concomitantly the appearance of complete dodecahedra is necessarily slower.

The predictions of these simulations are in qualitative agreement with the kinetics of assembly of virus-like particles from brome mosaic virus coat protein, *in vitro* (Cuillel *et al.*, 1983). At protein concentrations of 5 to 8 mg/ml, approximately 50% of the material was assembled into $T = 3$ capsids or nearly complete capsids within one second. A multiplicity of unresolved intermediate states were observed during assembly. Additional capsid assembly occurred at a much slower rate. This is comparable to the simulation in Figure 3C and F where material that is kinetically trapped in intermediates slowly assembles into dodecahedra. Trypsinized capsid protein assembled into $T = 1$ particles with much slower kinetics including a lag phase of about one hour (Cuillel *et al.*, 1981). Measurable concentrations of intermediates were not observed during assembly of the $T = 1$ particles, this case is comparable to the simulation shown in Figure 3A and D.

These simulations are dependent on several assumptions. All on-rates are treated as equal and diffusion limited. The stipulated pathway proceeds only through the most stable intermediates (Table 1). Finally, free monomeric subunits are required as a reactant in each assembly reaction. These are not necessarily realistic assumptions; each on-rate may be unique, the diffusion limit is a limiting case. The choice of intermediate species and reactions was for convenience of calculation. The simulations do show that (1) equilibrium assembly need not be a slow process and (2) intermediates may transiently accumulate and be kinetically trapped.

### (c) *Cooperativity and nucleation*

Only two features of subunit interactions are defined in the basic equilibrium model: particle geometry and the energy of subunit–subunit association. Particle geometry, including subunit geometry, defines the number of subunits in the particle and the degeneracy of assembly. The only adjustable parameter for a given particle is the energy term. This simple equation can be modified to include other relevant features of assembly such as multiple subunit types and interactions, nucleation and cooperativity. In this paper, cooperativity is defined as some binding energy, positive or negative, in addition to the binding energy from pairwise interactions between adjacent subunits. By this definition, a possible source of cooperativity would be the formation of a $\beta$ annulus, observed in several $T = 3$ plant viruses including southern bean mosaic virus (Abad-Zapatero *et al.*, 1980) and tomato bushy stunt virus (Harrison *et al.*, 1978). A $\beta$ annulus is a $\beta$ sheet structure that is formed around icosahedral 3-fold (*quasi* 6-fold) axes by the N *termini of three* monomers which do not have a common edge. This binding energy is in addition to the more easily accountable pairwise interactions; without structural information the source of this cooperative binding energy would not be obvious. Incorporation of cooperativity into the equilibrium assembly model requires an understanding of the source of this energy and how it will affect the stability of different intermediates.

Independent of whether assembly is nucleated or not, the behavior of equilibrium assembly can suggest nucleation. A lag phase is often considered diagnostic of nucleation (Oosawa & Asakura, 1975), but this effect can be generated by equilibrium assembly (Figure 3). Because of the extreme concentration dependence for assembly of a virus, the $K_{d^{app}}$ can mimic a critical concentration (Figure 2). Below the $K_{d^{app}}$ there is little intact dodecahedron; above the $K_{d^{app}}$ most free protein will be incorporated into capsids at equilibrium. Oosawa & Asakura (1975) noted that a polymer forming a closed shell should not require nucleation. In spite of this, we find that polymerization of a shell may resemble nucleated assembly.

In the assembly of viruses, a hetero-oligomer of coat protein and a specific RNA or DNA structure often is implicated as the nucleus conferring specificity to choice of nucleic acid packaged (Wei *et al.*, 1990; Beckett *et al.*, 1988, Zhong *et al.*, 1993). Formation of a nucleus, particularly at low coat protein concentrations, is likely to affect assembly kinetics by providing an elevated initial concentration of otherwise rare intermediates. As suggested by Sorger *et al.* (1986), a nucleus may also have a role in dictating the morphology of the resulting capsid by either thermodynamically or kinetically directing the pathway of subunit association through self control (Caspar, 1980).

A nucleus need not be a complex of coat protein. For example, phage Φ29 capsid and scaffold proteins spontaneously assemble into a heterogeneous *mixture of capsid-like complexes with different sizes* and shapes when co-expressed in *Escherichia coli*. However, when the portal vertex protein is also co-expressed the synthetic procapsids are uniform, resembling the naturally occurring structure (Guo *et al.*, 1991). Scaffold proteins may regulate the geometry of association at many local sites, acting as a local nucleus or chaperone (Prevelige & King, 1993), in addition to providing association energy from the coat protein—scaffold interaction. In the absence of scaffold protein, phage P22 coat protein forms misorganized aggregates at high concentra-

tions (Earnshaw & King, 1978); assembly induced by mixing coat and scaffold proteins yields $T = 7$ particles (Prevelige *et al.*, 1988). An expression to describe the role of scaffolding protein in assembly, or for the assembly of any heteropolymer, must incorporate energy terms for each form of contact and concentration terms for each scaffold and coat proteins.

(d) *The role of nucleic acid in viral assembly*

The coat proteins of many icosahedral viruses have in common a basic N terminus as a nucleic acid binding motif (Argos & Johnson, 1984). Electrostatic interactions with nucleic acid could drive formation of coat protein-nucleic acid complexes independent of ordered capsid assembly. The local concentration of coat protein could thus be much higher than the average solution concentration, even driving the assembly of viruses that have a low $K_{d_{app}}$ in the absence of nucleic acids (*vis à vis* Pontius, 1993). In the limiting case, where the coat protein has an extremely strong affinity for nucleic acid, virus assembly will be essentially a concentration-independent intramolecular reaction. Such reactions are often treated in terms of effective concentration, compared to a bimolecular interaction, which may be orders of magnitude greater than possible in solution (Page & Jencks, 1971).

Proteins non-specifically bound to nucleic acids may slide along the nucleic acid or even jump from one non-specific site to another at no free energy cost (Von Hippel & Berg, 1989). Thus, subunits can readily rearrange to conform to viral geometry and still be bound to nucleic acid.

The above discussion is most appropriate for *in vitro* assembly reactions. *In vivo*, capsids may be assembled by incorporation of newly synthesized proteins (see section (b), above). Thus, the role of nucleic acids may be restricted to concentrating incoming capsid protein leaving the concentration dependence unaffected.

The bromovirus family provides an example of protein-nucleic acid interactions that are important for virus assembly, though not absolutely required. The N terminus of bromovirus coat proteins is very basic and is probably bound to RNA in the intact virus (van der Graaf *et al.*, 1992; Speir, unpublished results). Brome mosaic virus can be disassembled under high pressure under mild solution conditions (Silva & Weber, 1988). The dissociated virus readily reassembles in a concentration-independent manner, when the pressure is decreased. This indicates that the subunits are not free to diffuse away from each other but remain associated in a disassembled form. Though not unequivocal, this is the sort of behavior expected for viral assembly assisted by non-specific protein-RNA interactions. Concentration-independent reversible disassembly was also found with bacteriophage R17 and attributed to protein-RNA interactions (Da Poian *et al.*, 1993).

A high effective concentration of subunits can drastically affect the predicted behavior of assembly when the free energy of intersubunit association is very small. When $\Delta G_c^\circ$ is $\ll 0$ kcal/mole the most stable intermediates are the predominant species. But, when $\Delta G_c^\circ$ is close to 0 kcal/mole the energy term in equation (2) approaches unity; all possible species, 70 in the case of a dodecahedron, must be considered. For $\Delta G_c^\circ$ close to 0 kcal/mole, no association between subunits would be observed at the protein concentrations possible in bulk solution. In the special case where the effective concentration is near 1 M and $\Delta G_c^\circ$ is near 0 kcal/mole, a statistical distribution of intermediates is expected. This may only be possible when there is a combination of strong protein-nucleic acid interactions and an environment that promotes protein–protein dissociation by decreasing $\Delta G_c$. Broad distributions of intermediates have been observed in the case of R17 phage disassembly by high pressure (Da Poian *et al.*, 1993) and urea-induced disassembly of Nodamura virus (A.Z. *et al.*, unpublished results).

(e) *A thermodynamic basis for switching capsid size*

*In vitro*, it is not uncommon to find that a viral capsid protein can assemble into capsids of several different sizes and morphologies. The basic model suggests a thermodynamic basis for switching capsid size in addition to a conformational mechanism. For a larger capsid with more assembly subunits, the increased concentration dependence (equation (4)) creates a steeper assembly slope but offsets the curve to a slightly higher concentration (see Figure 2). Thus, at very low subunit concentrations small capsids will be prevalent, while at higher concentrations the larger capsids will be favored (see Oosawa & Asakura, 1975).

Switching mechanisms are likely to involve more factors than concentration alone. Construction of larger icosahedral capsids from a single type of subunit requires quasi equivalent binding. In a $T = 1$ capsid all subunits are equivalent. For $T > 1$, quasi equivalence leads to different binding environments and presumably different association energies; building a virus of one type of subunit with multiple binding modes is analogous to assembling a heteropolymeric virus. The different environments may require different protein conformations, with different energies, i.e. a conformational switch. Non-specific binding to RNA can increase the effective concentration of subunit and thus also affect capsid type. A more realistic switching mechanism must incorporate the effects of nucleation and association energy as well as concentration.

## 3. Discussion and Caveats

The formalism presented here describes an equilibrium relationship between monomers and a complete polymer that can account for many observations of viral assembly. It is applicable to the assembly of viruses, oligomeric proteins and prions, any polymer that has a definite number and

arrangement of subunits, as opposed to indefinitely long linear polymers. Specific behavior is predicted for viruses at equilibrium: generally, only monomer and intact virus are found in appreciable concentration, no nucleus is specifically required. The basic model cannot, *a priori*, predict the effect of multiple binding modes (conformational switching), nuclei, and scaffolding on polymerization. Crystal structures of viruses have shown that there are many strategies whereby a subunit may adapt to different environments (for some examples see Harrison, 1980; Liddington *et al.*, 1991; Wery *et al.*, 1994). The statistical approach to equilibrium assembly is similar to that taken by Caspar (1963) in considering the nucleation step in the kinetics of tobacco mosaic virus assembly.

The predictions of this model are based on the assumption that the polymer is approaching equilibrium with its subunits. This may not be a realistic description of viral assembly *in vivo*, yet equilibrium assembly mimics nature very well, in the simplest case using only one adjustable parameter, $\Delta G_c$, for a capsid with a given geometry. Treating virus assembly as a process at equilibrium provides a basis for analyzing virus stability and mechanisms of assembly.

This formalism suggests two basic parameters that may be experimentally accessible, $\Delta G_{app}$ and the power of the concentration dependence of assembly. It may be possible to directly observe $\Delta G_{app}$ and the power dependence in concentration titrations. $\Delta G_{app}$ is the average association energy of each assembly subunit in the intact capsid. The concentration dependence indicates the number of subunits the capsid is assembled from. This term may suggest the presence of some kinetically unstable species on which assembly is dependent, it may be attenuated by positive cooperativity, or it may not be accessible if assembly is dependent on association with a polynucleotide.

In the simplest case, comparison of equilibrium assembly and classical nucleated assembly show differences in kinetic and equilibrium behavior. In the equilibrium assembly of an n subunit polyhedron, the extent of polymerization is a function of the nth power of the concentration of free subunits, there is no formal critical concentration. However, the concentration dependence of assembly kinetics may be difficult to interpret. Conversely, studies on nucleated assembly of linear polymers have focused on the kinetics of polymerization (Oosawa & Asakura, 1975). In nucleated assembly, the rate-limiting step is nucleus formation and thus it is the kinetics that are dependent on the concentration of total subunit to the nth power, where n is the number of monomers in the nucleus. In an extreme case, the polymerization of the sickle cell mutant of hemoglobin was dependent on the 32nd power of the concentration, indicating the existence of a 32 protein nucleus (Hofrichter *et al.*, 1974). Furthermore, above the critical concentration of nucleation the extent of polymerization is linear with the total concentration of monomer; at equilibrium, the concentration of free monomer will not exceed the critical concentration.

In nucleated assembly it is implicit that polymerization is infinite in one, two, or three dimensions, limited only by the availability of monomer. As discussed in this paper and by Oosawa & Asakura (1975), a classical nucleus is not necessary for assembly of polyhedra where, in general, larger aggregates are more stable than smaller ones and aggregate size is limited by closure of the polyhedra.

Capsid assembly does not always lead to a single uniform product. Thus, there must be some manner of switching between different possible products. The choice or distribution of product can be regulated by nucleation, concentration, protein–nucleic acid interactions, or scaffolding protein, but, in the end, must depend on some conformational switch within the assembly subunit itself. Conformational switching, described as self control or autostery, was suggested by Caspar (1963, 1980) to explain the complexity of nucleation and assembly in tobacco mosaic virus. Regulation of the association of free protein for the growing polymer by existing protein–protein interactions could be a means of constraining the pathway of capsid assembly.

I have simplified many elements of the model systems discussed in this analysis, only considering infinitely stable, symmetrical subunits. Nature is not so convenient. Subunits may be asymmetric with several inter-related binding sites. While it may be impossible to analyze experimental data in detail, the equilibrium assembly model can provide a framework for qualitative interpretation of results. The model described in this paper is not necessarily applicable where assembly is based on kinetically trapped products rather than products and intermediates approaching equilibrium.

This formalism is specifically designed for topologically closed polymers, it is in no way appropriate for a linear polymer. On the other hand one can draw a parallel from equilibrium assembly to protein folding. In protein folding many (non-equivalent) interactions serve to stabilize a folded product which is in equilibrium with the unfolded form. The individual reactions are analogous to intersubunit contacts; it is the mathematical product of this constellation of relatively weak individual associations that may result in the apparently strong cooperativity and stability of many proteins. The unfolding of proteins has been described in terms of a system partition function (Murphy & Freire, 1992). The equilibrium assembly model allows explicit evaluation of the major components of the partition function for a self associating system.

Dr William Ray provided substantial help and criticism in developing this model. The model itself is a partial answer to questions of virus assembly that arose in numerous discussions with Dr John E. Johnson. Dr Mark Young and Dr Carol Post also provided many useful comments. This manuscript was prepared with the assistance of Sharon Fately. During the course of this work,

A.Z. was supported by a Biophysics training grant (GM08296), an NIH grant (GM34220) to John E. Johnson, and a Merck graduate fellowship.

## References

Abad-Zapatero, C., Abdel-Mequid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). Structure of southern beam mosaic virus at 2·8 Å resolution. *Nature (London)*, **286**, 33–39.

Argos, P. & Johnson, J. E. (1984). Chemical stability in simple spherical plant viruses. In *Biological Macromolecules and Assemblies* (McPherson, A., ed.), vol. 1, John Wiley and Sons, Inc., New York.

Bancroft, J. B. (1970). The self assembly of spherical plant viruses. *Advan. Virus Res.* **19**, 99–134.

Beckett, D., Wu, H.-N. & Uhlenbeck, O. C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* **204**, 939–947.

Butler, P. J. G. (1979). Assembly of regular viruses. *Int. Rev. Biochem.* **25**, 205–237.

Caspar, D. L. D. (1963). Assembly and stability of tobacco mosaic virus particles. *Advan. Protein Chem.* **18**, 37–118.

Caspar, D. L. D. (1980). Movement and self control in protein assemblies: quasi equivalence revisited. *Biophys. J.* **32**, 103–135.

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

Cuillel, M., Jacrot, B. & Zulauf, M. (1981). A $T = 1$ capsid formed by protein of brome mosaic virus in the presence of trypsin. *Virology*, **93**, 63–72.

Cuillel, M., Berthet-Colominas, C., Krop, B., Tardiet, A., Vachette, P. & Jacrot, B. (1983). Self assembly of brome mosaic virus capsids: kinetic studies using neutron and X-ray solution scattering. *J. Mol. Biol.* **164**, 645–650.

Da Poian, A. T., Oliveira, A. C., Gaspar, L. P., Silva, J. L. & Weber, G. (1993). Reversible pressure dissociation of R17 bacteriophage. The physical individuality of virus particles. *J. Mol. Biol.* **231**, 999–1008.

Earnshaw, W. & King, J. (1978). Structure of phage P22 coat protein aggregates formed in the absence of scaffolding protein. *J. Mol. Biol.* **126**, 721–747.

Guo, P., Erickson, S., Xu, W., Olson, N., Baker, T. S. & Anderson, D. (1991). Regulation of Φ29 prohead shape and size by the portal vertex. *Virology*, **183**, 366–373.

Harrison, S. C. (1980). Protein interfaces and intersubunit bonding: the case of tomato bushy stunt virus. *Biophys. J.* **32**, 139–153.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). Tomato bushy stunt virus at 2·9 Å resolution. *Nature (London)*, **276**, 368–373.

Hofrichter, J., Ross, P. D. & Eaton, W. A. (1974). Kinetics and mechanism of deoxyhemoglobin S gelation: a new approach to understanding sickle cell disease. *Proc. Nat. Acad. Sci., U.S.A.* **71**, 4864–4868.

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). Structure of simian virus 40 at 3·8 Å resolution. *Nature (London)*, **354**, 278–284.

Murphy, K. P. & Freire, E. (1992). Thermodynamics of structural stability and cooperative folding behavior in proteins. *Advan. Protein Chem.* **43**, 313–361.

Oosawa, F. & Asakura, S. (1975). *Thermodynamics of the Polymerization of Protein*. Academic Press, New York.

Page, M. I. & Jencks, W. P. (1971). Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Nat. Acad. Sci., U.S.A.* **68**, 1678–1683.

Pontius, B. W. (1993). Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *TIBS*, **18**, 181–186.

Prevelige, P. E., Jr & King, J. (1993). Assembly of bacteriophage P22: a model for dsDNA virus assembly. In *Prog. Med. Virol.* (Melnick, J. L., ed.), vol. 40, Korger, Basel.

Prevelige, P. E., Jr, Thomas, D. & King, J. (1988). Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells *in vitro*. *J. Mol. Biol.* **202**, 743–757.

Prevelige, P. E., Jr, Thomas, D. & King, J. (1993). Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**, 824–835.

Rueckert, R. R. (1990). Picornaviridae and their replication. In *Virology* (Fields, B. N. & Knipe, D. M., eds), Raven Press, Ltd., New York.

Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. (1986). Self-assembly of purified polyomavirus capsid protein VP1. *Cell*, **46**, 895–904.

Salunke, D. M., Caspar, D. L. D. & Garcea, R. L. (1989). Polymorphism in the assembly of polyomavirus capsid protein VP1. *Biophys. J.* **56**, 887–900.

Savithri, H. S. & Erickson, J. W. (1983). The self-assembly of the cowpea strain of southern bean mosaic virus: formation of $T = 1$ and $T = 3$ nucleoprotein particles. *Virology*, **126**, 328–335.

Silva, J. L. & Weber, G. (1988). Pressure-induced dissociation of brome mosaic virus. *J. Mol. Biol.* **199**, 149–159.

Sorger, P. K., Stockley, P. G. & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly *in vitro*. *J. Mol. Biol.* **191**, 639–658.

van der Graaf, M., Scheek, R. M., van der Linden, C. C. & Hemminga, M. A. (1992). Conformation of a pentacosapeptide representing the RNA-binding N-terminus of cowpea chlorotoic mottle virus coat protein in the presence of oligophosphates: a two-dimensional proton nuclear magnetic resonance and distance geometry study. *Biochemistry*, **31**, 9177–9182.

Von Hippel, P. H. & Berg, O. (1989). Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678.

Wei, N., Heaton, L. A., Morris, T. J. & Harrison, S. C. (1990). Structure and assembly of turnip crinkle virus VI. Identification of coat protein binding sites on the RNA. *J. Mol. Biol.* **214**, 85–95.

Wery, J.-P., Reddy, V. S., Hosur, M. V. & Johnson, J. E. (1994). The refined three-dimensional structure of an insect virus at 2·8 Å resolution. *J. Mol. Biol.* **235**, 565–586.

Zhong, W., Dasgupta, R. & Rueckert, R. R. (1993). Evidence that the packaging signal for nodaviral RNA2 is a bulged stem-loop. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 11146–11150.

*Edited by P. von Hippel*

*(Received 13 December 1993; accepted 4 May 1994)*