A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures

DAN ENDRES,¹ MASAKI MIYAHARA,^{1,2} PAUL MOISANT,^{1,2} AND ADAM ZLOTNICK²

¹Department of Mathematics and Statistics, University of Central Oklahoma, Edmond, Oklahoma 73034, USA

²Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73190, USA

(RECEIVED December 24, 2004; FINAL REVISION December 24, 2004; ACCEPTED March 24, 2005)

Abstract

The capsids of spherical viruses may contain from tens to hundreds of copies of the capsid protein(s). Despite their complexity, these particles assemble rapidly and with high fidelity. Subunit and capsid represent unique end states. However, the number of intermediate states in these reactions can be enormous—a situation analogous to the protein folding problem. Approaches to accurately model capsid assembly are still in their infancy. In this paper, we describe a sail-shaped reaction landscape, defined by the number of subunits in each species, the predicted prevalence of each species, and species stability. Prevalence can be calculated from the probability of synthesis of a given intermediate and correlates well with the appearance of intermediates in kinetics simulations. In these landscapes, we find that only those intermediates along the leading edge make a significant contribution to assembly. Although the total number of intermediates grows exponentially with capsid size, the number of leading-edge intermediates grows at a much slower rate. This result suggests that only a minute fraction of intermediates needs to be considered when describing capsid assembly.

Keywords: capsid assembly; virus assembly; protein polymerization; protein folding; energy landscape

Supplemental material: see www.proteinscience.org

Reprint requests to: Adam Zlotnick, Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, P.O. Box 26901, BRC464, Oklahoma City, OK 73190, USA; e-mail: adam-zlotnick@ouhsc.edu; fax: (405) 271-3910.

Abbreviations: N , number of subunits in a complete capsid; n , number of subunits in an intermediate; $stat$, the statistical factor reflecting assembly degeneracy over a whole capsid; $\Delta G_{contact}$, pairwise association energy between subunits; $\Delta G_{n,j}$, the overall association energy for the j -th intermediate of n subunits; P , probability of the specified intermediate; μ , a weighting factor for reaction chemistry; s , the statistical factor reflecting assembly degeneracy for a specific reaction; f , the forward reaction rate for a specified reaction, a function of s and a microscopic rate; b , the backward rate for a specified reaction, a function of f and $\Delta G_{n,j}$; t , time.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.041314405>.

Complicated reactions, such as protein folding, are often described in terms of energy landscapes (Bryngelson et al. 1995; Wolynes 1996; Brooks et al. 1998; Chan and Dill 1998; Onuchic et al. 1998; Dobson and Karplus 1999; Dinner et al. 2000). This representation of the reaction allows a quantitative description of the multiplicity of intermediates, the many paths that can be followed to local and global minima, and the different kinetic barriers that each path must overcome. A similar descriptive approach has also been applied to association reactions (Wales 1987, 1996; Ball et al. 1996; Wolynes 1996; Kumar et al. 2000). In these representations, a higher-dimensional problem must be compressed into a

low-dimensional representation while still conveying the complexity of the landscape. Additionally, the information conveyed by the landscape must be related to fundamentally incomplete experimental data. In this paper we describe a reaction landscape for the assembly of the capsids of icosahedral viruses that applies equally well to assembly of any closed structure constructed through a cascade of association reactions.

About half of all known virus families have spherical morphology, and the overwhelming majority of these are icosahedral (Fields et al. 1996). Icosahedral viruses have capsids constructed of 60 identical asymmetric units. Most capsids are constructed from T multiples of 60 proteins, where there are T identical proteins in the asymmetric unit in distinct but quasiequivalent environments (Caspar and Klug 1962). The individual proteins making up the capsid may be grouped together to form protomers, the fundamental unit of the assembly reaction. For example, hepatitis B virus has a $T = 4$ capsid assembled from 120 homodimers (Zhou and Standring 1992; Crowther et al. 1994; Wingfield et al. 1995), and $T = 1$ picornaviruses (e.g., poliovirus) are assembled from 12 pentamers of heterotrimers (Rueckert 1996). Capsids play diverse roles in virus lifecycles: Depending on the virus, capsids may protect the viral genome, bind receptors, direct intracellular transport, and serve as compartments for nucleic acid metabolism (Fields et al. 1996). Consequently, interfering with capsid assembly may be a potent, though as yet unutilized, target for antiviral therapeutics (Prevelige 1998; Zlotnick et al. 2002; Deres et al. 2003; Hacker et al. 2003).

Several approaches have been taken to describe capsid assembly reactions. The most stable intermediates were defined for two 180-subunit viruses, using a structure-based identification of the most stable intermediate of each size (Horton and Lewis 1992). A more sophisticated analysis, investigating three virus structures, examined a broader range of intermediates (up to 2500 of each size) to consider multiple paths for assembly reactions and found that for a given capsid protein, a subset of intermediates was extremely stable (Reddy et al. 1998). Based on this combinatorial approach to identifying the most stable intermediates, it was suggested that assembly follows energetically preferred paths and that morphological switches could occur where different types of intermediates are isoenergetic (Reddy et al. 1998). Systems of differential rate equations (so-called master equations [Wales 1996]) have been developed for highly simplified model systems comprised of a single assembly path for building a polyhedron (Fig. 1A) (Zlotnick 1994; Endres and Zlotnick 2002). Analyses based on this approach were used to identify the nucleating reactions for hepatitis B virus (Zlotnick et al. 1999), cowpea chlorotic mottle virus (Zlotnick et al. 2000), and polyoma virus (Casini et al. 2004). Though simple compared to the *in vitro* reactions

Figure 1. The paths and intermediates for assembly of a dodecahedral capsid. (A) A single path sequence using only the most stable intermediates. (B) The complete path sequence showing all 73 species of the dodecahedron as two-dimensional projections (Schlegel diagrams), where shaded faces correspond to assembled subunits. The 11-mer and 12-mer look the same because the 12th face is implicit. The segments joining species identify possible reactions paths. The horizontal axis n is size (number of assembled subunits); the vertical axis j is type as ordered by the enumeration algorithm.

they model, these simulations have also helped explain rapid equilibration of assembly reactions (Ceres and Zlotnick 2002; Zlotnick 2003), hysteresis to dissociation (Singh and Zlotnick 2003), assembly inhibition/misdirection (Zlotnick et al. 2002), and the generation of kinetic traps (Stray et al. 2004). Coarse-grained molecular dynamics simulations have also been applied to assembly, leading to the suggestion that the assembly rate need not slow dramatically as the capsid approaches completion (Rapaport et al. 1999). A molecular dynamics-like approach was also used to examine a “local rules” mechanism for quasiequivalence and its regulation of assembly (Berger et al. 1994; Schwartz et al. 1998). “Local rules” was also used to emulate aberrant assembly, suggesting assembly misdirection as an antiviral approach (Prevelige 1998). However, there is an inherent stability to icosahedral symmetry (Zandi et al. 2004). Descriptions of potential surfaces for assembly of capsid-like molecules emphasize that this reaction could be described as a relatively smooth assembly funnel (Wales 1987, 1996; Wolynes 1996).

A useful representation of virus assembly must include features dictated by biology and allow other features of the assembly reaction to be identified. Geometrically, a wide variation in model structure is available (Twarock 2004). Using models, methods have been developed for recovering physically meaningful reaction parameters

(Endres and Zlotnick 2002). The goal of accurately describing assembly of capsids is complicated by the number of possible intermediates, which grows dramatically with the number of subunits from which the capsid is built. This is analogous to the Levinthal paradox: The number of possible conformations available to a protein grows exponentially with the number of residues, making it impossible for a protein of even modest size to search all conformational space to identify the lowest energy fold (Levinthal 1969). Landscape approaches have shown how proteins will follow an energetically preferred constellation of paths resulting in a much smaller and faster search of conformational space (Bryngelson et al. 1995; Wolynes 1996; Brooks et al. 1998; Chan and Dill 1998; Onuchic et al. 1998; Dobson and Karplus 1999; Dinner et al. 2000). In the same way, we find that a landscape approach shows that only a small population of intermediates is expected to make significant contributions to virus assembly, reducing a computationally intractable problem to a more practical size. Sparse assembly models that include appropriate intermediates yield simulations almost identical to those calculated with complete models and will be suitable for fitting experimental observations of assembly reactions.

Results

From sparse to complete models

For a capsid having N subunits, the minimal assembly model has one species for each size $n = 1, \dots, N$. It is natural to incorporate the most stable intermediate of each size in a minimal model and expand it with additional species along excursions from the minimal path to obtain a still sparse but more representative model. The most stable intermediates are those with the largest number of inter-subunit contacts; Figure 1A shows them for assembly of the Platonic dodecahedron, $N = 12$. Complete models include all intermediates; for the dodecahedron (Figs. 1B, 2), there are 73 possible species—from the pentagonal subunit to the complete capsid.

The total number of species in the complete model grows rapidly, probably exponentially, with N . For the Platonic polyhedra with $N = 4, 6, 8, 12$, and 20 faces, there are 4, 8, 14, 73, and 2649 species, respectively. The species for the dodecahedron are shown as an array of Schlegel diagrams in Figure 1B, where we observe the size n having the largest number of species, 20, is $n = 6$, i.e., the hexamers. The icosahedron's species count distribution maximum, 548, is realized for 13-mers (see Fig. 3). The shift of the maximum to a value of $n > N/2$ is a result of the many larger species with several holes.

We have developed algorithms (to be described elsewhere) that generate all intermediates in a capsid

Figure 2. A landscape representation for assembly of a dodecahedron. (A) In a plot of $\Delta G^\circ/n$ vs. n , the most stable intermediate of each size has the lowest average energy. The species are color-coded by size (number, n , of assembled subunits). (B) The stability-probability landscape populated by the 73 dodecahedral assembly species resembles a sail. The “leading edge” of the sail approximates a line from the subunit ($n = 1, j = 1$) to capsid ($n = 12, j = 1$). The horizontal coordinates are number of subunits (n) and log probability [$-\log(P)$]; the vertical coordinate is the average energy per subunit, $\Delta G^\circ/n$. Segments show association/dissociation paths. (C) A cross-section of B along $n = 6$, hexamers. Representative Schlegel diagrams for the different hexamer types are shown in a plot of $\Delta G^\circ/n$ vs. $-\log(P)$. Points occupied by multiple species (in this case, enantiomorphs) have only one Schlegel shown, but are labeled with the list of indices from Figure 1B. The most probable hexamer types, ($n = 6, j = 3, 5$), are not the most stable ($n = 6, j = 1$). The least stable, chain-like species, ($n = 6, j = 17, 20$) are also the least probable.

assembly system. For a capsid of N subunits, the complete set of intermediates is generated by building up from a first subunit (e.g., Wales 1987), adding one subunit at a time to each site and then deleting duplicates. These algorithms also identify all reaction paths between species, which must be counted to evaluate reaction degeneracy (Zlotnick 1994). For the dodecahedron, there are 263 different segments joining the 73

Figure 3. Two reaction landscapes for assembly of a 20-subunit icosahedron (*inset*). All 2649 intermediates are shown. (A) The leading edge comprises few intermediates when $\mu = 0.1$ for reactions that only form a single bond. Intermediates cover a broad range of $-\log(P)$, but the most probable ones have $\log(P)$ close to 0. A pathway through the most probable intermediates is traced by a dashed line from subunit to capsid. (B) The landscape where $\mu = 1$ is much more compressed. The most probable species are not always the most stable ones. The most probable species are traced by a thin solid line (the dashed line from A is shown for comparison).

species (Figs. 1B, 2A); for the icosahedron ($N = 20$), there are 17,241 segments joining the 2649 species (Fig. 3). Thus, even for $N = 12$, a capsid may be built by many possible routes.

Using libraries of intermediates, we can describe assembly of a capsid by treating addition of each subunit as an independent association reaction. The evolution of species concentrations over time is governed by a system of quadratically nonlinear differential rate equations (e.g., Equation 2). The system has as many equations as there are species participating in the reaction as a whole—from subunit through intermediates to capsid (Zlotnick et al. 1994, 1999; Endres and Zlotnick 2002).

The reaction landscape

The numerous intermediate species that might form during capsid assembly can be compared on the basis of

stability (Fig. 2A). To facilitate comparison, species are designated by (n, j) , where n is the number of subunits and j is arbitrarily assigned by the enumeration algorithm. The average free energy per subunit, $\Delta G^\circ_{(n,j)}/n$, of the species (n, j) , is based on scoring one unit of association energy for each edge–edge intersubunit contact ($\Delta G^\circ_{\text{contact}}$) and accounting for reaction degeneracy. For c contacts, the association energy is $c\Delta G^\circ_{\text{contact}} + RT \ln(\text{stat})$, where stat is a statistical factor derived from reaction degeneracy, R is the gas constant, and T is the temperature in kelvins. The value of $RT \ln(\text{stat})$ is small compared to $c\Delta G^\circ_{\text{contact}}$ (see Equation 5 in Endres and Zlotnick 2002). As the capsid nears completion, there is a progressive increase in stability (Fig. 2A). In general, the most stable species are those that are most compact, such as those shown for a dodecahedron in Figure 1A corresponding to the lowest energy path in Figure 2A.

The two-dimensional representation of Figure 2A fails to account for the importance of kinetics in regulating capsid assembly reactions (cf. Stray et al. 2004). In developing kinetics simulations, we use reaction degeneracy to modify the microscopic forward rate constant (Zlotnick 1994). These same factors can be used to calculate a probability distribution (Equation 1) that predicts the prevalence for each species. The probability, $P(n, j)$, takes into account (1) forward reaction degeneracy ($s_{n,j,i}$), i.e., the number of ways for each progenitor species $(n-1, i)$ to form (n, j) ; (2) the concentration of each progenitor as predicted by its probability $P(n-1, i)$; and (3) a weighting factor $\mu_{n,j,i}$ related to the chemistry of the given association reaction (discussed below). $P(n, j)$ varies from 0 to 1. In accord with the approximate unidirectionality of these association reactions at early times, P ignores any contribution of dissociation to the concentration of (n, j) . When $\mu = 1$, forward reaction degeneracy is the only factor needed to calculate P , so that $\ln(P)$ is related to the entropy associated with the forward reactions.

$$P(n, j) = \frac{\sum_i s_{n,j,i} P(n-1, i) \mu_{n,j,i}}{\sum_j \sum_i s_{n,j,i} P(n-1, i) \mu_{n,j,i}} \quad (1)$$

A landscape showing the participants in an assembly reaction can be constructed using probability [$P(n, j)$], stability ($\Delta G^\circ_{\text{contact}}/n$), and the reaction coordinate (n). We use $\log(P)$ to construct the landscape because P is extremely small ($P < 1$) for most intermediates. At either extreme of n , subunit ($n = 1$) and capsid ($n = N$), only one possible form of n -mer exists so that $\log[P(1)] = \log[P(N)] = 0$. The values of $\log(P)$ for intermediate species cover a broad range, which suggests that many of these species will never be present at a high concentration—in an assembly reaction, the bulk of the population will exist in a small selection of states.

Examination of a section (Fig. 2C) of the landscape in Figure 2B shows that the most stable intermediate is not always the most probable one. The most stable intermediate is compact, with the largest number of contacts. The most probable intermediates allow the greatest degeneracy of assembly. When all forward reactions are identical, except for degeneracy, the paths to the most stable intermediate may encounter a bottleneck. For example, the most stable hexamer (6,1) is formed only by docking a subunit in one site of the most stable pentamer (5,1) or in one site of the improbable ring (5,8), while the more probable hexamers types are each formed from several pentamer types in various ways (Fig. 1B). The least stable and least probable intermediates are the long chains, e.g., enantiomeric hexamer types 17 and 20, which have only five contacts (Fig. 2C).

When $\mu = 1$, all reactions are postulated to have the same microscopic rate constant. However, under this constraint assembly is prone to kinetic traps (Zlotnick et al. 1999). A nucleation phase, where selected reactions are either slower or weaker, results in assembly that is much more robust to extremes in association energy and subunit concentration (Zlotnick et al. 1999). In the calculation of probability, P , setting the appropriate $\mu_{n,j,i} < 1$ can account for nucleation and also reduce the contribution of relatively unstable intermediates.

A landscape where nucleation-like reactions have $\mu < 1$ shows distinct differences from a $\mu = 1$ landscape (Fig. 3, cf. A and B). In Figure 3A, μ is set to 0.1 for all reactions where the newly added subunit is connected by a single edge–edge contact. Since they only generate a single weak contact, these reactions are similar to reactions leading to nucleus formation. As a result, P values for the most stable intermediates increase and the P values for chain-like species become much smaller. When nucleation-like reactions are weighted with $\mu < 1$ in the dodecahedron (Supplemental figure) and icosahedron landscapes (Fig. 3A), the most stable intermediates become the most probable ones. When P includes a component that reflects reaction chemistry, it is no longer strictly “probability.”

In assembly landscapes (Figs. 2B, 3), a small population of species lies near the high probability edge of the energy landscape, i.e. where $\log(P)$ is nearest to 0. This high-probability “leading edge” of the reaction landscape should include the most stable species even when they are not the most probable. Examination of P reveals that only a few intermediates of each size are predicted to be dominant and that most intermediates make very little contribution to the reaction. In short, the leading-edge species are those that contribute most significantly to the assembly reaction as measured energetically by $\Delta G^\circ_{\text{contact}}/n$ and kinetically by their probability P . Precise definition of the leading edge depends on the choice of

μ and an ad hoc cutoff value. Consider the 548 types of 13-mers in Figure 3 (green ovals). In the $\mu = 1$ landscape (Fig. 3B), the 63 most probable 13-mers contribute 75% of the probability; in the landscape where μ is set to 0.1 for reactions that form single edge–edge contacts, only seven 13-mers are needed to account for 75%. Probability P correlates with the mole fractions of the different species at early times when stability has little effect on prevalence. At later times, when stability is a greater factor in intermediate concentration, fewer intermediates (the most stable ones) are important—as demonstrated in kinetics simulations described below.

The prevalence of intermediates changes over time

The reaction landscapes indicate that some intermediates are more “probable” than others. To investigate this, the relative concentrations of different intermediates were observed in kinetics simulations. A system of rate equations describing assembly can be generated from a set of intermediates and paths (Fig. 1, A or B) together with the corresponding reaction degeneracies. In these kinetics simulations, all forward reactions have the same microscopic association rate k_{elong} multiplied by the individual forward reaction degeneracy $s_{n,j,i}$ to yield the forward rate $f_{n,j,i}$ of Equation 2 (Zlotnick 1994). For simplicity, the simulations described in this paper do not have a separate nucleation phase (which would require a separate rate and impose constraints on the reactions paths). The dissociation reaction rate $b_{n,j,i}$ is the product of the forward rate and the dissociation constant determined from the association energy. A separate equation is developed for each species concentration $[n,j]$ following the general form:

$$\frac{d[n,j]}{dt} = \sum_i f_{n,j,i}[1][n-1,i] - \sum_k f_{n+1,k,j}[1][n,j] - \sum_i b_{n,j,i}[n,j] + \sum_k b_{n+1,k,j}[n,k] \quad (2)$$

The summations represent buildup to (n,j) from $(n-1)$ -mers, buildup to $(n+1)$ -mers, dissociation of (n,j) , and dissociation to (n,j) from $(n+1)$ -mers, respectively. There is no equivalent to μ built into the equations used to generate the simulations discussed in this paper.

Simulations for the assembly of dodecahedral capsids showed the characteristic sigmoid kinetics of capsid formation (cf. Fig. 4, inset, and Fig. 5 to typical experimental results) (Zlotnick et al. 1999). The evolution of hexamer concentrations was examined at three times, $t = 6, 60$, and 600 sec, selected to reflect different phases of the association reaction (Fig. 4). Hexamers were chosen as the most diverse selection of intermediates and because the most probable hexamer is not the most

Figure 4. The relative concentrations of intermediates evolve over time. The normalized fractions for hexamer are shown at $t = 6$ sec (white circles), $t = 60$ sec (gray circles), and $t = 600$ sec (black circles). As the reaction progresses, more of the mass is concentrated in the most stable species; some intermediates become extremely rare. The weighting factor μ , which down-weights the probability of unstable intermediates, recapitulates the clock. The probabilities of hexamers for $\mu = 0.1$ (white squares) closely match the simulations at 6 sec. The probabilities for $\mu = 0.01$ (gray squares) match the 60-sec data. Shaded regions group the different time and μ data sets. (Inset) The kinetics simulation from which these data are extracted; trajectories are shown for subunit, hexamer, and capsid mass fractions; vertical lines identify times of 6, 60, and 600 sec. For this simulation, the initial subunit concentration was $10 \mu\text{M}$, $\Delta G^\circ_{\text{contact}} = -3.5 \text{ kcal/mol}$ [a typical value for per contact association energy in viruses (Ceres and Zlotnick 2002; Zlotnick 2003)], and the microscopic forward rate constant was $1000 \text{ M}^{-1} \text{ sec}^{-1}$. Simulations used the Rosenbrock 4th–5th order, stiff system integrator in MAPLE (Waterloo Software) with $\text{abserr} = \text{relerr} < 1.0\text{E}-9$ to control fluctuations across 0 concentration at early times.

stable. Because of the breadth of the concentration range, we express the concentration of a given species (n, j) as the log of the mole fraction of (n, j) among n -mers.

$$\log P(n, j) = \log \frac{[n, j]}{\sum_j [n, j]} \quad (3)$$

The earliest time examined ($t = 6$ sec) is during the lag phase of the reaction, where populations of intermediates are initially generated but are far from steady state. At $t = 6$ sec, hexamer concentration is dominated by the six most probable intermediates (Fig. 4). Even at early times, the hexamer distribution is not based entirely on intermediate probability but is modulated by stability. As the reaction progresses, through the rapid assembly of capsid ($t = 60$) and enters the plateau phase ($t = 600$), where intermediates approximate a steady state that slowly approaches equilibrium, stability becomes the dominant factor in hexamer distribution. At $t = 600$, the most stable hexamer accounts for 73% of all

hexamers. This observation also demonstrates that though assembly is macroscopically unidirectional and despite the hysteresis to capsid dissociation (Singh and Zlotnick 2003), back reactions are critical to describing the changing concentrations of intermediates. The importance of back reactions is also supported by coarse-grained molecular dynamics simulations (Schwartz et al. 1998; Rapaport 2004).

The evolution of species over time can be recapitulated by changing μ ; μ mimics the clock. The parameter μ acts to decrease the probability (P) of less stable intermediates formed by making a single contact. For $\mu = 1$, the fastest forming species of n -mer are nearest the high-probability edge of the landscape (Fig. 2C). When μ is decreased for single-contact reactions, more stable species become more probable. Correspondingly, in simulations, the concentrations of intermediates evolve to favor progressively more stable n -mers over time. This approximation is robust, holding for simulations with many values of $\Delta G^\circ_{\text{contact}}$ and initial subunit concentrations (data not shown).

The $\log(P)$ coordinate emphasizes that assembly reactions are kinetic events, during which some species may achieve transient dominance because of their high probability of formation, independent of their stability. This is a key feature of the probability distribution: P estimates the distribution of mass among the various configurations of n -mer, early in the reaction.

Comparing sparse and complete model systems

Given that very few intermediates of size n contribute significantly to the mole fraction of all n -mers, how

Figure 5. Successful assembly reactions do not require low probability (P) intermediates. Four models of dodecahedron assembly were compared (using the same reaction parameters described for Fig. 4): (1) a complete model with 73 species (heavy solid line); (2) a 22-species model that includes the 12 most stable species and 10 intermediates that are one contact less stable (heavy dashed line); (3) a 21-species model that uses the most probable intermediates required for $>75\%$ probability, cumulative P , for each size (thin solid line); and (4) the 12 most stable species (dashed line). All four reactions begin with $10 \mu\text{M}$ subunit and eventually reach nearly the same concentrations of capsid at equilibrium. If reactions were quantitative, they would yield $0.833 \mu\text{M}$ capsid.

important are the improbable and less stable n -mers to assembly kinetics? To investigate this question, we compared assembly simulations for the dodecahedron for four basic models using (1) the 12 most stable species, (2) the 21 species required to account for at least 75% of the single-contact $\mu = 0.1$ probability (Supplemental figure), (3) the 22 species filling the two lowest energy levels in every size, and (4) the complete set of 73 species.

Successful reactions, for which the reaction produced an abundance of capsid without a kinetic trap, were simulated for 10 μM subunit (see legend for Fig. 4). The differences between simulations for the 21-, 22-, and 73-species models were so small that they would not be experimentally detectable (Fig. 5). The 12-species model showed a significant shortfall throughout the rapid assembly phase (Fig. 5). Equilibrium concentrations of capsid and free subunit would be experimentally indistinguishable for all four models.

Reactions where there was a substantial accumulation of intermediates due to a kinetic trap were more sensitive to the different intermediates included in the simulation (data not shown). Nonetheless, the sparse models give excellent approximations of the kinetic trajectory of the 73-species model.

Conclusions

We have developed a landscape representation for a complex association reaction: assembly of polyhedral structures such as virus capsids. The representation indicates that few of the many possible intermediates make a significant (measurable) contribution to assembly. This result is consistent with earlier descriptions of likely assembly intermediates based exclusively on their stability (Reddy et al. 1998). Our landscape is a function of species size, stability, and probability. The probability parameter emphasizes how assembly kinetics affects the assembly path: the most stable intermediate is not always the most likely. Building an accurate model of assembly must reflect kinetics and thermodynamics of the intermediate reactions. Though the calculated landscape is time- and concentration-independent, this representation emphasizes the earliest times in the reaction when the concentration of an intermediate is more strongly affected by the rate of its formation than by its stability. The distribution of species in an association reaction varies in time, thus a static calculation seeking to characterize the entire reaction must represent both transient and equilibrium states.

Isolation and characterization of every intermediate in a virus assembly reaction is experimentally inaccessible because of intermediate heterogeneity and rarity.

Intermediates have been observed as heterogeneity in small angle scattering (Cuillel et al. 1983), examples in an electron micrographs (Silva and Weber 1988; Prevelige et al. 1993; Kainov et al. 2003), in trapped reactions (Stray et al. 2004), or a disparity between different spectroscopic signals (Kainov et al. 2003; P. Ceres and A. Zlotnick, unpubl.). However, a description of representative intermediates is vital to interpreting experimental observation of both successful and unsuccessful reactions. The reaction landscape suggests a means of identifying the "leading edge" species critical to successful assembly.

The prevalent intermediates will be those species at (or near) the leading edge of the reaction landscape; other intermediates are expected to be rare. Our simulations indicate that a model restricted to leading edge species provides an excellent representation of the complete model. This is an important observation for developing models of assembly. The total number of possible intermediates appears to explode exponentially, whereas the number of species in the leading edge of an assembly landscape will grow much more slowly. We suggest that the leading edge may be definable so that the number of species in it grows nearly linearly with capsid size. Because a small number of significant intermediates are sufficient for accurate assembly simulations, these simulations will be readily calculable. The approach we have described is general; the particulars of geometry and specific features of the assembly reaction must be determined experimentally for the landscape to be relevant to a specific structure or virus. The peculiar features of an assembly system will allow the resulting landscapes and simulations to be unique and predictive.

Acknowledgments

This work was supported by NSF grant MCB-011025 to A.Z. In addition, D.E. and M.M. thank the Jackson College of Graduate Studies and Research at the University of Central Oklahoma for their support. Dr. Michael Thorpe provided helpful discussion. We also thank Heather Hardway, Toshihisa Kubo, and Ho Kong ("Henry") So for technical assistance.

References

- Ball, K.D., Berry, R.S., Kunz, R.E., Li, F.Y., Proykova, A., and Wales, D.J. 1996. From topographies to dynamics on multidimensional potential energy surfaces of atomic clusters. *Science* **271**: 963–966.
- Berger, B., Shor, P.W., Tucker-Kellogg, L., and King, J. 1994. Local rule-based theory of virus shell assembly. *Proc. Natl. Acad. Sci.* **91**: 7732–7736.
- Brooks III, C.L., Gruebele, M., Onuchic, J.N., and Wolynes, P.G. 1998. Chemical physics of protein folding. *Proc. Natl. Acad. Sci.* **95**: 11037–11038.
- Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21**: 167–195.
- Casini, G.L., Graham, D., Heine, D., Garcea, R.L., and Wu, D.T. 2004. In vitro papillomavirus capsid assembly analyzed by light scattering. *Virology* **325**: 320–327.

- Caspar, D.L.D. and Klug, A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**: 1–24.
- Ceres, P. and Zlotnick, A. 2002. Weak protein–protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**: 11525–11531.
- Chan, H.S. and Dill, K.A. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins* **30**: 2–33.
- Crowther, R.A., Kiselev, N.A., Bottcher, B., Berriman, J.A., Borisova, G.P., Ose, V., and Pumpens, P. 1994. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**: 943–950.
- Cuillet, M., Berthet-Colominas, C., Krop, B., Tardieu, A., Vachette, P., and Jacrot, B. 1983. Self-assembly of brome mosaic virus capsids. Kinetic study using neutron and X-ray solution scattering. *J. Mol. Biol.* **164**: 645–650.
- Deres, K., Schroder, C.H., Paessens, A., Goldmann, S., Hacker, H.J., Weber, O., Kramer, T., Niewohner, U., Pleiss, U., Stoltefuss, J., et al. 2003. Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science* **299**: 893–896.
- Dinner, A.R., Sali, A., Smith, L.J., Dobson, C.M., and Karplus, M. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**: 331–339.
- Dobson, C.M. and Karplus, M. 1999. The fundamentals of protein folding: Bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**: 92–101.
- Endres, D. and Zlotnick, A. 2002. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**: 1217–1230.
- Fields, B.N., Knipe, D.M., Howley, P.M., Chanock, R.M., Melnick, J.L., Monath, T.P., Roizman, B., and Straus, S.E. 1996. *Virology*, 3rd ed. Lippincott-Raven, Philadelphia.
- Hacker, H.J., Deres, K., Mildenerberger, M., and Schroder, C.H. 2003. Antivirals interacting with hepatitis B virus core protein and core mutations may misdirect capsid assembly in a similar fashion. *Biochem. Pharmacol.* **66**: 2273–2279.
- Horton, N. and Lewis, M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**: 169–181.
- Kainov, D.E., Butcher, S.J., Bamford, D.H., and Tuma, R. 2003. Conserved intermediates on the assembly pathway of double-stranded RNA bacteriophages. *J. Mol. Biol.* **328**: 791–804.
- Kumar, S., Ma, B., Tsai, C.J., Sinha, N., and Nussinov, R. 2000. Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci.* **9**: 10–19.
- Levinthal, C. 1969. How to fold graciously. In *Mossbauer spectroscopy in biological systems* (eds. P. DeBrunner et al.), pp. 22–24. University of Illinois Press, Allerton House, Monticello, IL.
- Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z., and Socci, N.D. 1998. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci.* **92**: 3626–3630.
- Prevelige, P.E.J. 1998. Inhibiting virus-capsid assembly by altering the polymerisation pathway. *Trends Biotech.* **16**: 61–65.
- Prevelige, P.E., Thomas, D., and King, J. 1993. Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells. *Biophys. J.* **64**: 824–835.
- Rapaport, D.C. 2004. Self-assembly of polyhedral shells: A molecular dynamics study. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **70**: <http://pre.aps.org>.
- Rapaport, D.C., Johnson, J.E., and Skolnick, J. 1999. Supramolecular self-assembly: Molecular dynamics modeling of polyhedral shell formation. *Comp. Phys. Comm.* **121**: 231–235.
- Reddy, V.S., Giesing, H.A., Morton, R.T., Kumar, A., Post, C.B., Brooks III, C.L., and Johnson, J.E. 1998. Energetics of quasiequivalence: Computational analysis of protein–protein interactions in icosahedral viruses. *Biophys. J.* **74**: 546–558.
- Rueckert, R.R. 1996. Picornaviridae: The viruses and their replication. In *Fields virology* (eds. B.N. Fields et al.), pp. 609–654. Lippincott-Raven, Philadelphia, PA.
- Schwartz, R., Shor, P.W., Prevelige, P.E.J., and Berger, B. 1998. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.* **75**: 2626–2636.
- Silva, J.L. and Weber, G. 1988. Pressure-induced dissociation of brome mosaic virus. *J. Mol. Biol.* **199**: 149–159.
- Singh, S. and Zlotnick, A. 2003. Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. *J. Biol. Chem.* **278**: 18249–18255.
- Stray, S.J., Ceres, P., and Zlotnick, A. 2004. Zinc triggers conformational change and oligomerization of hepatitis B virus capsid protein. *Biochemistry* **43**: 9989–9998.
- Twarock, R. 2004. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.* **226**: 477–482.
- Wales, D.J. 1987. Closed-shell structures and the building game. *Chem. Phys. Lett.* **141**: 478–484.
- . 1996. Structure, dynamics, and thermodynamics of clusters: Tales from topographic potential surfaces. *Science* **271**: 925–929.
- Wingfield, P.T., Stahl, S.J., Williams, R.W., and Steven, A.C. 1995. Hepatitis core antigen produced in *Escherichia coli*: Subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry* **34**: 4919–4932.
- Wolynes, P.G. 1996. Symmetry and the energy landscapes of biomolecules. *Proc. Natl. Acad. Sci.* **93**: 14249–14255.
- Zandi, R., Reguera, D., Bruinsma, R.F., Gelbart, W.M., and Rudnick, J. 2004. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci.* **101**: 15556–15560.
- Zhou, S. and Strandberg, D.N. 1992. Hepatitis B virus capsid particles are assembled from core-protein dimer precursors. *Proc. Natl. Acad. Sci.* **89**: 10046–10050.
- 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.
- . 2003. Are weak protein–protein interactions the general rule in capsid assembly? *Virology* **315**: 269–274.
- Zlotnick, A., Reddy, V.S., Dasgupta, R., Schneemann, A., Ray Jr., W.J., Rueckert, R.R., and Johnson, J.E. 1994. Capsid assembly in a family of animal viruses primes an autoproteolytic maturation that depends on a single aspartic acid residue. *J. Biol. Chem.* **269**: 13680–13684.
- Zlotnick, A., Johnson, J.M., Wingfield, P.W., Stahl, S.J., and Endres, D. 1999. A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**: 14644–14652.
- Zlotnick, A., Aldrich, R., Johnson, J.M., Ceres, P., and Young, M.J. 2000. Mechanism of capsid assembly for an icosahedral plant virus. *Virology* **277**: 450–456.
- Zlotnick, A., Ceres, P., Singh, S., and Johnson, J.M. 2002. A small molecule misdirects assembly of hepatitis B virus capsids. *J. Virol.* **76**: 4848–4854.