Local Rules Modeling of Nucleation-Limited Virus Capsid Assembly

MIT-LCS-TM-584

Russell Schwartz[1], Peter E. Prevelige, Jr.[2], Bonnie Berger[1,3]

August 7, 1998

[1]Laboratory for Computer Science
[3]Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139

[2]Department of Microbiology
University of Alabama at Birmingham
Birmingham, AL 35294

Abstract

We describe an application of computer modeling to the study of the kinetics of virus capsid (protein shell) assembly. We examine two proposed models of the source of nucleation-limited growth, an observed growth pattern in which initiation of new capsids occurs significantly more slowly than subunit addition onto initiated capsids. We apply an abstract computer model of capsid assembly, based on the principle of local rules, to support a theoretical argument for favoring a two-conformation model over a one-conformation model. The theoretical analysis examines expected relative growth and nucleation rates and concludes that the two-conformation model should be able to support faster growth following nucleation for any fixed nucleation rate. Based on the theoretical argument, we develop predictions which are then supported by computer simulation results on a model of $T = 1$ capsid assembly. In addition to strengthening the argument for a two-conformation model, our results demonstrate the potential value of computer simulations in comparing hypothetical models for observed biochemical behaviors.

Keywords: kinetics, self-assembly, computer simulation, conformational switching.

1 Introduction

The kinetics of icosahedral virus capsid assembly have proven difficult to resolve. Icosahedral capsids have traditionally been described by the quasi-equivalence theory of Caspar and Klug (1962). But while this theory provides a description of the final assembled structure, it does not provide much direct insight into the process of assembly. Although some questions about capsid assembly kinetics have been answered, other key problems remain unresolved. Examples include constraints on the orders of assembly, the timing of conformational switching, and the role of scaffolding proteins in enforcing size-determination. It is likely that some aspects of assembly kinetics differ from one virus to another. For example, the T=7 phage HK47 appears to build pentamers and hexamers first, then assemble these completed capsomers to form a capsid (Xie & Hendrix, 1995), while P22, another T=7 phage, appears to assemble its capsid directly from individual coat proteins (Prevelige *et al.*, 1993).

One important open question concerns the source of nucleation-limited growth in icosahedral capsids. Nucleation-limited growth has been observed in P22 (Prevelige *et al.*, 1993). However, it has not yet been possible to establish the source of this behavior. A variation of the argument of Oosawa and Kasai (1962), proposed to explain helical self-assembly, could provide one explanation for the source of nucleation-limited

Authors' email addresses: rsschwar@mit.edu, prevelig@uab.edu, bab@lcs.mit.edu.

behavior in virus capsid assembly. Oosawa and Kasai noted that in a helix, a nucleation complex consisting of an initial turn contains fewer binding interactions per subunit than larger assemblies; subunits would therefore have more difficulty overcoming the configurational entropy penalty of binding when they nucleate a new structure than when they add additional subunits to an existing structure, explaining why nucleation would be relatively unfavorable compared to growth following nucleation. Caspar (1980) specifically described how this argument might apply to helical viruses. A similar argument might also be used to explain nucleation-limited behavior in icosahedral virus capsids; in the icosahedral case, the number of binding interactions per subunit grows from only one when two initially unbound subunits first collide, to three or more in a completely closed capsid, as each subunit contacts, at a minimum, its two neighbors in its capsomer and a subunit in at least one neighboring capsomer. An alternative hypothesis, the autostery model of Caspar (1980), involves a similar entropy penalty, but conjectures that it is produced in part by conformational switching. In the autostery model, conformational switching occurs after binding. We discuss a slightly different two-conformation model in which proteins occupy binding and non-binding conformations at various times when free in solution; binding of two free proteins is then unfavorable because both proteins must happen to be in the binding conformation simultaneously, requiring two entropy penalties for a single binding interaction, while completing a pentamer is more favorable because it requires five entropy penalties for five binding interactions. Both one- and two-conformation theories could explain observed kinetic behavior, and it has proven difficult to decide between them. The first theory has the advantage of greater simplicity, requiring no conformational shifting. However, it has been argued that configurational entropy alone could not be sufficient to explain the magnitude of the entropy penalty that would be required to give the observed behavior (Caspar, 1980). So far, it has not been possible to demonstrate the validity of either theory experimentally.

Such questions often cannot be adequately addressed through current laboratory techniques; we have therefore developed a computer simulator in the hopes of providing additional insight. The basis of the simulator is the local rules model of Berger *et al.* (1994), which describes capsid assembly in terms of "local rules" specifying binding patterns for different conformations of coat proteins. Local rules models can offer some predictions about assembly kinetics without computer simulations, as demonstrated by Berger and Shor (1995), although there are limits to the predictions that can be made directly from the theory. Our simulator combines a molecular dynamics-like approach with the local rules model, allowing it to abstract away many details of binding interactions, making the problem of capsid assembly computationally tractable (Schwartz *et al.*, 1998). The simulator is described in more detail in the methods section.

2 Theory

This section provides a theoretical argument for favoring the two-conformation model of nucleation-limited behavior over the one-conformation model. The argument is based on relating entropy penalties due to configurational specificity of binding to entropy penalties due to conformational switching during both the nucleation phase and the growth phase of capsid assembly. The argument here is described assuming pentameric nucleation, however it applies qualitatively to nucleation complexes of any size greater than one protein. In addition, the argument assumes that growth occurs by addition of individual coat proteins following nucleation, as as been observed for P22 (Prevelige *et al.*, 1993), rather than through the addition of larger pre-assembled structures, such as capsomers. The argument concludes that if nucleation-limited growth is promoted by the existence of an entropy penalty to binding, then having a component of this entropy penalty come from conformational switching should promote a greater ratio of growth rate to nucleation rate compared to having an entropy penalty derived purely from configurational entropy. This implies an advantage to the existence of a non-binding conformation in promoting rapid growth and a high yield of correct, complete capsids.

We can propose some reasons why evolution might select for a high ratio of growth rate to nucleation rate in virus capsid assembly. One potential advantage is that by limiting the frequency of nucleation events and insuring that growth proceeds rapidly following nucleation, a virus avoids the possibility of exhausting all available subunits while many partially formed capsids are being constructed. If nucleation occurred too often or growth following nucleation were too slow, it might be that many capsids would nucleate, absorbing many subunits, before more than a small fraction had completed, thus slowing the growth process or reducing the overall yield. A secondary benefit would be reducing the probability of collisions between partially formed capsids, which might be less stable than fully formed capsids, thus possibly reducing the incidence of malformations.

Both configurational specificity and the existence of a non-binding conformation can provide large entropy penalties, which could promote nucleation-limited growth. However, the two sources of entropy penalty differ in the ratio of the entropy penalty for nucleation to the entropy penalty for subunit additions following nucleation. If it is assumed that viruses evolve to optimize some ratio of nucleation rate to growth rate following nucleation, then the differing ratios provide an argument for favoring the existence of a conformational component to any entropy penalty. The two ratios can be found by examining the number of times the two types of entropy penalty, configurational and conformational, are incurred in nucleation and in subsequent subunit additions.

The effect of a configurational entropy penalty is considered first. For nucleation to occur, it is necessary for four coat protein subunits to converge in the correct configuration relative to some initial subunit. Suppose the entropy penalty implies some probability p_1 that two proteins happen to be in the correct relative configurations for binding. In order for nucleation to occur, four proteins must converge relative to one protein of arbitrary initial configuration, giving a contribution of $(p_1)^4$ to the binding probability. Following nucleation, each additional particle to be added must end up in the correct position relative to the existing partial shell, giving a contribution of p_1. This means that the contribution of configurational entropy to the nucleation probability varies with the fourth power of the contribution to the probability of incorporating a subunit after nucleation. It can be noted that this relationship is only approximate, since the five subunits may be able to attach sequentially in a short window of time, rather then all converging at precisely the same time. However, the fifth order dependence of nucleation rate on concentration observed in P22 (Prevelige *et al.*, 1993) suggests the approximation is reasonable.

A conformational entropy penalty would be expected to behave slightly differently. If it is assumed that there is one binding conformation and one non-binding conformation, then the proportion of time spent in the binding conformation determines some conformational entropy penalty of binding. This entropy penalty can be interpreted as a probability, p_2, that a protein happens to be in the binding conformation at any given time. In order for nucleation to occur, it is necessary for all five coat protein subunits to be in the binding conformation when they converge into the correct relative positions for binding. This implies that a total conformational contribution of $(p_2)^5$ to binding probability. Adding an additional subunit after nucleation requires only that the new subunit be in the binding conformation, since all subunits already contained in the partial capsid will already be fixed in the binding conformation. This implies a conformational contribution of p_2 to binding probability for each subunit addition following nucleation. Thus, the contribution of the conformational entropy penalty to probability of nucleation varies approximately with the fifth power of the contribution to the probabilities of subsequent subunit additions. The relationship is approximate for the same reasons as with the configurational entropy penalty, but should likewise be reasonable given the data on P22.

These two analyses suggest that conformational entropy should be more effective than configurational entropy at producing nucleation-limited growth while allowing a maximum rate of growth following nucleation. If there is some entropy-induced component of nucleation probability, p, required to insure nucleation-limited growth, and it comes entirely from a configurational entropy of binding, then it will be approximately true that subsequent subunit additions occur at a rate proportional to the fourth root of the nucleation rate. On the other hand, if the entropy penalty derives entirely from a conformational entropy penalty of binding, then subunit additions following nucleation will occur at a rate approximately proportional to the fifth root of the nucleation rate. This implies an advantage to having a conformational component to the entropy penalty: the more a required entropy penalty is dominated by conformational rather than configurational entropy, the less effect this entropy penalty will have on slowing growth after nucleation has occurred. Thus, evolving a non-binding conformation to create nucleation-limited behavior should provide an advantage in assuring rapid growth.

3 Methods

3.1 Computer Model

We conducted computer simulations using a molecular dynamics-like simulator incorporating the local rules model. The simulator implements a "soup" of free-floating particles, representing individual coat proteins. These particles are capable of forming and breaking bonds to other particles in accordance with a local rules model. In addition, they have many adjustable parameters, allowing users to fine-tune simulations to particular tasks or models of growth.

The local rules model provides a means of abstracting away many aspects of inter-subunit binding interactions, making a molecular dynamics-like model of capsid self-assembly computationally feasible. Under this model, coat proteins are represented abstractly by subunits with user-specified binding properties. These binding properties determine to which other subunits a particular subunit may bind. In addition, they specify the activation energies for the association and dissociation reactions and the mechanical force the binding interactions exert on bound particles. Potential binding interactions also have angle and distance tolerances which enforce how close to their optimal relative positions two particles must be before they can bind to each other. Under this model, particles can move freely throughout a simulated solution and assemble into large structures without requiring a low-level model of the specific forces creating binding interactions in actual viruses.

In addition to binding properties, there are many other user-specifiable properties of coat proteins and the simulation environment. Coat proteins can have different masses and radii as well as different shapes, created from unions of spheres. Furthermore, binding properties and other physical properties of coat proteins can change probabilistically over time through a model of conformational switching, in which users can assign different potential energies to different conformations, controlling the probability of subunits occupying each conformation at a given time. In addition, users can specify parameters controlling some aspects of the behavior of the simulated solution, such as temperature and viscosity.

Several forces act on particles over the course of a simulation. The forces particles exert on each other through binding interactions are modeled through three springs: a translational spring, which pulls particles towards their ideal translational offsets; a bending spring, which straightens binding interactions that are skewed; and a rotational spring, which limits rotations around binding interactions. In addition, forces are exerted due to collisions between pairs of particles, or between particles and an artificial boundary created around the simulated solution. Finally, forces are exerted due to a model of Brownian motion which combines damping with small random perturbations to keep average kinetic energy close to a fixed value over time. Integrating the equations of motion given these forces causes a simulation to evolve over time.

Combining these details creates a dynamic simulation of self-assembly kinetics. This simulation allows for particles that can form into multiple clusters and allows such clusters to break apart or rearrange their binding patterns over time. In addition, the model allows for malformations, as binding interactions can form in non-ideal positions and can be stretched from those positions. For further details on the implementation of the simulator, the reader is referred to Schwartz *et al.* (1998).

3.2 Experimental Design

We have created four simulations to compare the different theories on the source of nucleation-limited capsid assembly. All four involve T=1 capsids composed entirely of a single coat protein. We define the four simulations to be considered as follows:

- Simulation A: The coat protein takes on two conformations, one binding and one non-binding, with a 3:1 ratio of non-binding to binding proteins.
- Simulation B: Coat proteins have a single conformation, identical to the binding conformation of simulation A, with concentration reduced by a factor of four.
- Simulation C: Coat proteins have a single conformation, identical to the binding conformation of simulation A, except that binding tolerances are restricted to reduce the favorable configuration space of binding for one protein by a factor of four.

- Simulation D: Coat proteins have a single conformation, identical to the binding conformation of simulation A, except that binding tolerances are altered to reduce the favorable configuration space of binding for one protein by a factor of two.

The four simulations described above demonstrate different aspects of these two models. Simulation A represents the two-conformation model, in which nucleation-limited behavior derives from an entropy penalty of conformational switching. Simulation B is used to show that the effect of the non-binding conformation derives from more than just the reduced concentration of active subunits. Simulations C and D demonstrate the one-conformation model, in which nucleation-limited behavior derives solely from an entropy penalty from the configurational constraints of binding. Simulation A was conducted first, to measure how long it required to grow one complete T=1 capsid from 120 particles randomly distributed throughout the simulated solution, testing for completion at intervals of 1000 time steps. The mapping between these time steps and actual time should not be interpreted other than qualitatively, as it has not been possible to measure how reliably simulations capture crucial time-related simulation properties, such as the diffusion rate. However, capsid parameters were biased to allow for rapid growth due to computational limits on running large numbers of time steps, and the time period of formation of these capsids should therefore correspond to a significantly shorter time than the time required for the formation of an actual virus capsid. Once the first simulation was completed, the other simulations were conducted, using the closure time in simulation A as a base and examining the simulation states at multiples of this base. From the theoretical model it can be predicted that simulation B should have approximately the same initial nucleation rate as simulation A, while simulations C and D should have higher rates. After nucleation, capsids in simulation C should grow at a similar rate to nucleated capsids in simulation A, given equal numbers of free monomers, while nucleated capsids in simulation B should grow at a slower rate and those in D grow at a higher rate.

It is possible to quantify the predicted effects of the different models in terms of the entropy penalty implemented by conformational switching in simulation A, by concentration in simulation B, and by configurational entropy in simulations C and D. We will describe this in terms of the probability p that a free protein in simulation A is in the binding conformation at any given point in time (chosen here to be .25). In terms of p, the relative contributions of these entropy penalties to nucleation and growth rates for the four simulations, and the ratios of the contributions, should be approximately:

	nucleation	growth	nucleation/growth
A	p^5	p	p^4
B	p^5	p^2	p^3
C	p^4	p	p^3
D	p^2	$p^{\frac{1}{2}}$	$p^{\frac{3}{2}}$

Simulation A should therefore have the highest ratio of growth rate to nucleation rate of the four simulations.

4 Results

We now discuss the specific results of our simulation experiments. Our results show correct, complete, capsid growth only in simulation A. The other simulations only produced partially formed or malformed capsids.

In simulation A, a completed capsid was first visible at 9000 time steps. The simulation at this point is shown in figure 1. In this picture, large spheres represent proteins in the binding conformation while small spheres represent proteins in the non-binding conformation. The single completed capsid is visible. No other nucleation can be seen.

Simulation B did not produce any completed capsids. The status of simulation B at four multiples of 9000 time steps is shown in figure 2. One feature that stands out in contrast to simulation A is that multiple nucleations have occurred independently. This becomes an obstacle to production of a completed capsid late in the simulation, as the nucleated clusters each use a large fraction of the proteins, preventing any individual cluster from gathering enough proteins to form a complete capsid. Another feature of this simulation is that the partially formed capsids interact with each other producing malformed clusters of particles, such as that visible near the center of figure 2C.

Simulation C also produced no completed capsids. Simulation C is shown at four multiples of 9000 time steps in figure 3. This simulation exhibits several of the same features as simulation B. It has multiple nucleations by the first 9000 time steps, and again, this seems ultimately to prevent completed capsid growth

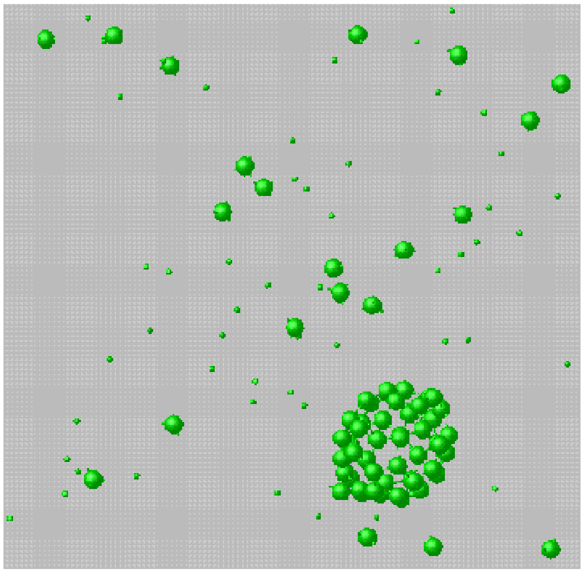

Figure 1: Simulation A at 9000 time steps. This simulation implements the two-conformation model. The figure shows a single completed capsid in the lower-right corner. Other subunits are visible, although no other nucleation events have occurred. The larger particles represent subunits in the binding conformation, while the smaller particles represent subunits in the non-binding conformation.

by causing all free proteins to be absorbed into incomplete shells. By figure 3C, one capsid is nearly complete, but is unable to proceed further on the time scale examined due to the low numbers of free proteins. In figure 3B, it can be observed that two partially formed capsids have collided, producing malformed growth as they interact. However, local rearrangements are able to correct the malformation and separate the two partial capsids by figure 3C.

Simulation D was characterized by the early production of an uncorrected malformation. The status of simulation D at four multiples of 9000 time steps is shown in figure 4. As of 9000 time steps, shown in figure 4A, simulation D seems similar to simulation C, although there are still some free proteins available in simulation D, unlike in simulation C. By 27000 time steps, shown in figure 4B, a malformed capsid has been created by a collision between two partially formed capsids that became "stuck" to one another. Unlike the malformation in simulation C, this malformation is never corrected by local rearrangements of the proteins. This malformed capsid consists of two layers, each similar to a correctly formed partial capsid. Over the course of the simulation, the malformed capsid gradually accumulates more of the free proteins.

5 Discussion

Overall, several predicted aspects of capsid behavior were observed. Parameters were chosen to promote correct, rapid growth for simulation A, so it is not surprising that that was observed. In itself, simulation A therefore tells us very little about the success of our model. Simulation B showed the predicted reduced rate of post-nucleation growth compared to simulation A, and the resultant higher rate of nucleation relative to capsid growth. Simulation C displayed the predicted higher nucleation rate, but was unable to complete any capsids. Simulation D also displayed a higher nucleation rate than simulation A, as well as a growth rate that appears to have been comparable to that of simulation A until free proteins were exhausted. The simulation experiment thus results in the same conclusion as the theoretical argument: that the two-conformation model can achieve a higher rate of growth for a given nucleation rate than the one-conformation model.

In some cases it was not possible to evaluate a predicted effect. In simulation C, capsid growth may have been slowed by two competing factors: a lower attachment probability relative to simulation A, and the drop in free protein concentration due to multiple nucleations. It cannot be definitively concluded that either of these factors alone acted as predicted. Conversely, it cannot be determined with certainty if the multiple

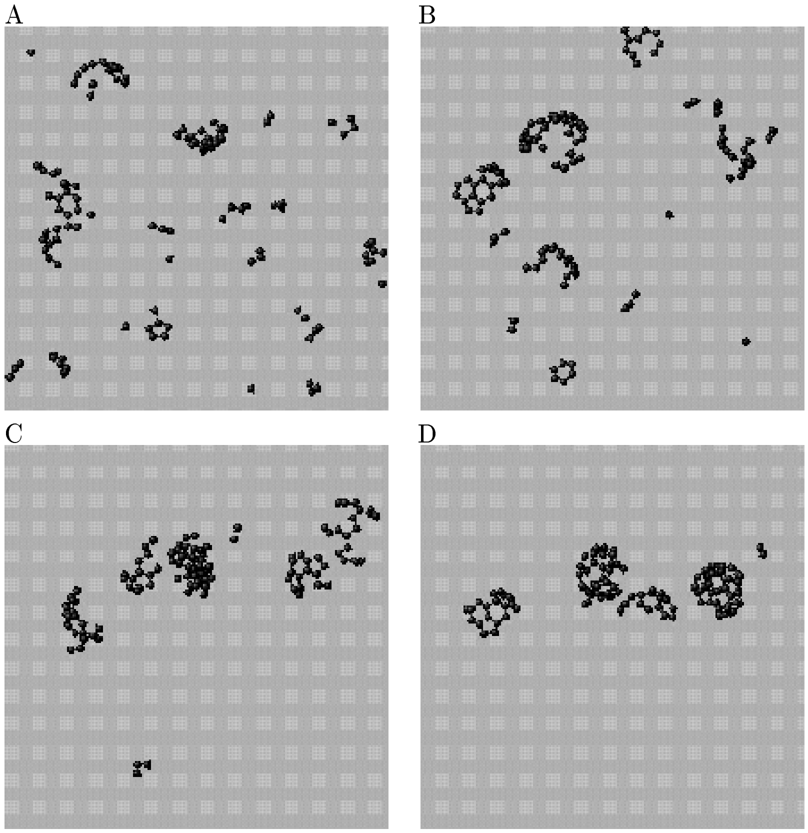

Figure 2: Simulation B at (A) 9000, (B) 18000, (C) 27000, and (D) 72000 time steps. This simulation implements a one-conformation model in which concentration is reduced by increasing volume to produce an effect similar to an additional entropy penalty of binding. The figure shows many nucleation events early in the simulation but no nucleated capsids growing fast enough after nucleation to form a completed capsid. By time step 72000, all nucleated capsids are incomplete or malformed.

nucleations in simulation B by 9000 time steps reflect a higher nucleation rate than A or if they reflect the greater number of free proteins available early in simulation B due to its slower growth rate. Furthermore, because free proteins are rapidly exhausted in simulations B, C, and D, we cannot clearly evaluate whether rates of assembly following nucleation are consistent with the theoretical predictions.

The nature of malformations observed may also be significant in understanding why one model of nucleation might be favored over another. The malformations arising from interactions of partially formed capsids, which occurred in all simulations except A, can be interpreted to be indirectly due to the higher ratio of growth rate to nucleation rate compared to simulation A; whereas simulation A had only one partially formed capsid present during the simulation, the others all had multiple nucleated capsids coexisting, leaving open the possibility of interactions. Extrapolating to larger solutions, it can be hypothesized that a correct balance is needed between these two rates to nucleate capsids sufficiently fast to have a high overall yield per unit time, while preventing the concentration of partially formed capsids from ever being sufficiently high that malformations due to collisions begin to dominate the growth process. This implies a lower bound on the ratio of growth rate to nucleation rate if a virus is to generate a large number of correctly formed capsids. A similar lower bound on the ratio of growth rate to nucleation rate is created by the constraint that nucleated capsids must have time to complete before too many others have nucleated; the consequence of too many nucleations occurring too quickly relative to growth rate is seen in simulations B and C, in which all free subunits are used up by partial capsids before any one capsid can complete. These lower bounds in turn suggest an evolutionary advantage to the two-conformation model, which, according to the arguments presented here, should be able to support a higher relative growth rate than the one-conformation model if

A B C D

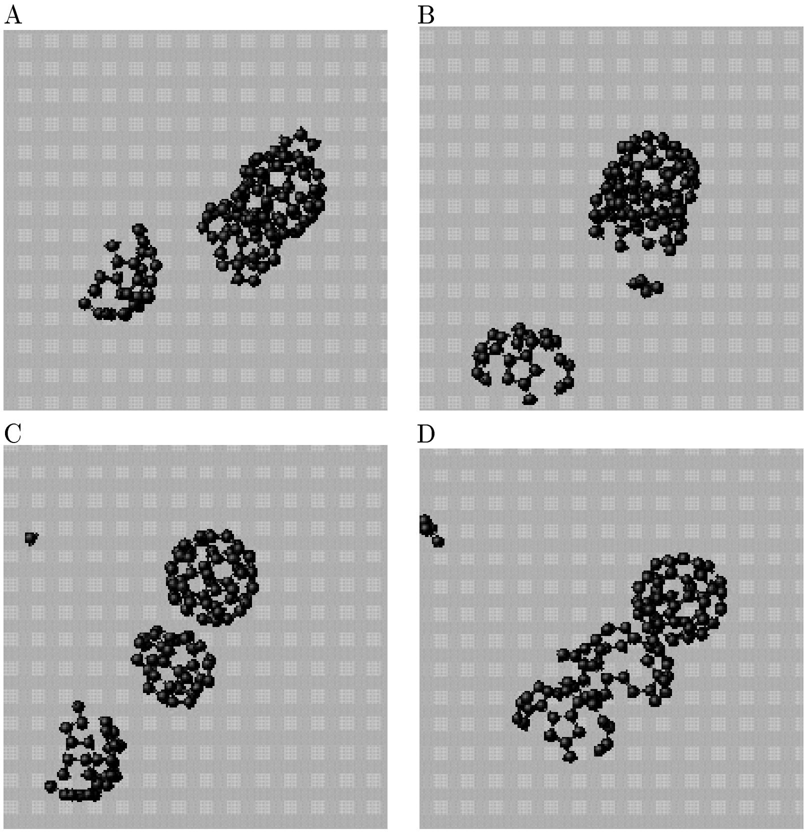

Figure 3: Simulation C at (A) 9000, (B) 27000, (C) 36000, and (D) 72000 time steps. This simulation implements a one-conformation model in which the binding tolerance, representing configurational entropy, as been made one fourth as lenient.

the ratio between nucleation and growth rates is bounded by other evolutionary constraints.

While neither our theoretical argument nor our simulation results can definitively resolve the source of nucleation-limited behavior, together they support the argument for the two-conformation model. The theoretical argument provides a rationale for an evolutionary advantage to two-conformation growth. The simulations demonstrate how the proposed rationale might lead to the predicted effects in practice. In addition, they allow observation of other, unanticipated effects of the two models, which could form the basis for new experiments or theoretical analyses that might aid in distinguishing between the two models. These unanticipated effects include the observations on the nature of malformations described above.

This application also helps to demonstrate the value of simulation work as a tool for modeling and evaluating theoretical predictions. In this case, computer simulation provides a means of comparing proposed models that is not available through any other method. While laboratory work can provide a means for testing hypotheses based on theoretical models, there are some cases in which it proves difficult or impossible to generate testable predictions that distinguish between even significantly different models. The nucleation-limited growth problem studied in the present work is such an example, in which it has so far proven impossible to decide between two proposed theoretical models on the basis of the available evidence. In such cases, simulation work can provide a means for evaluating the behavior of the models and possibly discovering unanticipated aspects of their behavior which could then be applied to distinguish the models in the laboratory. In the present work, it could be argued that simulation was not strictly necessary, since it was possible to develop a purely mathematical argument for favoring one proposed model over another. In other cases, however, models may be so hard to analyze or important aspects of their behavior so difficult to predict that computer simulations may provide the only available means of comparison.

Acknowledgments. This material is based upon work supported under a National Science Foundation

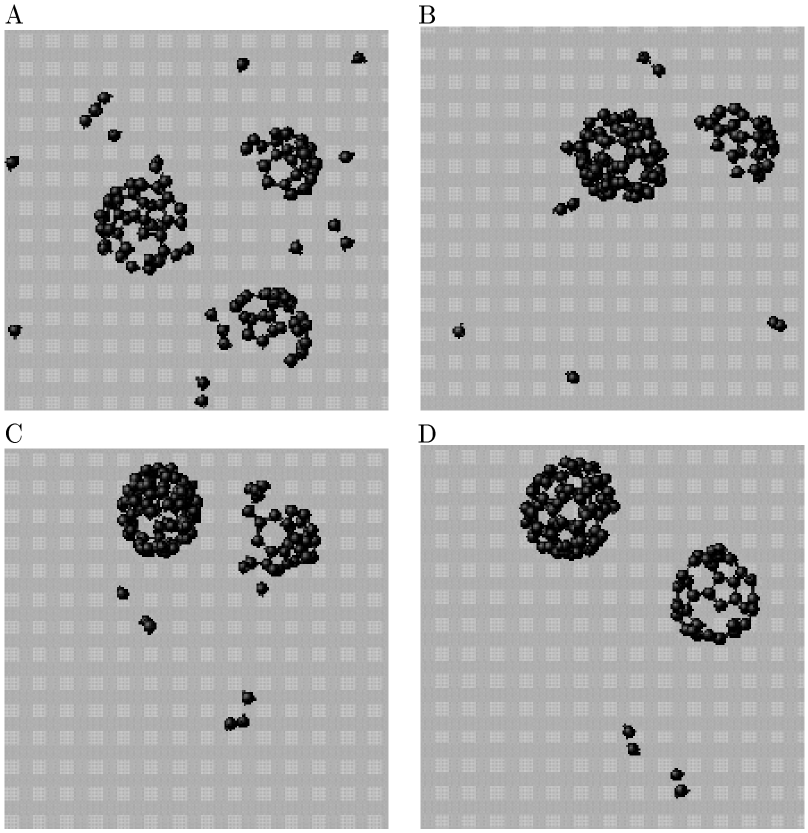

Figure 4: Simulation D at (A) 9000, (B) 18000, (C) 27000, and (D) 72000 time steps. This simulation implements a one-conformation model in which the binding tolerances have been made one half as lenient.

Graduate Research Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. Bonnie Berger is supported by NSF Career Award 95019997-CCR and DOE grant DEFG0295ER25253. Bonnie Berger and Peter Prevelige are supported by NSF Computational Biology grant DBI9711234. We would like to thank Charles Leiserson and the Cilk group for their suggestions on using Cilk efficiently and for allowing us use of their computers. Finally, we would like to thank Seth Teller for allowing us the use of his workstations.

References

Berger, B. & Shor, P. W. (1995). MIT Laboratory for Computer Science Technical Memo 527.

Berger, B., Shor, P. W., Tucker-Kellog, L., & King, J. (1994). *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7732–7736.

Caspar, D. L. D. (1980). *Biophys. J.* **232**, 103–138.

Caspar, D. L. D. & Klug, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

Oosawa, F. & Kasai, M. (1962). *J. Mol. Biol.* **4**, 10–21.

Prevelige, P. E., Thomas, D., & King, J. (1993). *Biophys. J.* **64**, 824–835.

Schwartz, R., Shor, P. W., Prevelige, P. E., & Berger, B. (1998). submitted for publication.

Xie, Z. & Hendrix, R. W. (1995). *J. Mol. Biol.* **253**, 74–85.