

Local rule simulations of capsid assembly

R. SCHWARTZ*†, P. W. SHOR‡¶ and B. BERGER‡||

†Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15217, USA

‡Department of Mathematics and Computer Science, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Local rules theory describes virus capsid self-assembly in terms of simple local binding preferences of discrete subunit conformations. The theory offered a parsimonious explanation for the complexity and regularity of virus capsids that resolved several inconsistencies between experimental observations and prior theories. Simultaneously, it provided a valuable practical abstraction for simulating the assembly process. Local rule models offered important advantages over other methods in rapidly developing simulations of complex geometries and assembly processes. Subsequent extensions considerably augmented the range of assembly phenomena amenable to simulation studies. Here we review the development of local rule-based simulation models and their applications to basic biology and medicine. We conclude by discussing the future prospects of local rule models for *in silico* experimentation with capsid assembly.

Keywords: Capsid; Simulation; Self-assembly; Kinetics; Reaction pathway

Msc: 92B99; 92E99; 92-08

1. Introduction

Spherical virus capsid assembly is perhaps the most complex and sophisticated example of self-assembly known, typically involving the spontaneous interaction of hundreds of distinct coat proteins to produce a complicated but regular structure under unpredictable and adverse conditions. In addition to their medical relevance to many important human and agricultural diseases, viruses are therefore an important model system for understanding self-assembly in general. Spherical capsids exhibit icosahedral symmetry but typically have chemically identical proteins occupying several geometrically distinct binding environments (symmetry groups under rotations of the capsid). Spherical capsid geometries are generally classified in terms of “*T* numbers” [1], which are based on the relative locations of pentamers, or pentons, and hexamers, or hexons, of capsid proteins in the shell. Figure 1 shows some capsid geometries observed in nature: the $T = 1$, 7 and 13. Despite extensive study of the capsid assembly phenomenon, important open questions remain about such issues as the constraints on assembly pathways, the role and timing of conformational switching, the

prospects for disrupting the assembly process, and the lessons virus assembly can teach us about designing novel self-assembly systems.

Because of the difficulties of directly observing or precisely manipulating the nanometer-scale assembly process, simulation methods have emerged as a valuable tool for studying virus capsid assembly. Simulations allow us to compare models of the assembly process and test their abilities to explain experimental data. Simulations similarly allow us to evaluate possible strategies for disrupting capsid assembly prior to their experimental realization, potentially accelerating drug discovery and optimization. Furthermore, simulation methods that prove able to match the extensive data available on capsid assembly may provide a basis for tools for rapid *in silico* prototyping of novel systems.

In the remainder of this paper, we review the development of capsid assembly simulation methods based on the local rules theory Berger *et al.* 1994, a model of assembly directed by simple local interactions between assembling coat proteins. We consider the progress from simple local rule models through increasingly sophisticated variants. We conclude by discussing recent

*Corresponding author. Tel.: +1-412-268-3971. Fax: +1-412-268-7129. Email: russells@andrew.cmu.edu

¶Tel.: +1 617 253 1827. Fax: +1 617 258 5429. Email: shor@math.mit.edu

||Corresponding author. Tel.: +1 617 253 4368. Fax: +1 617 253 4358. Email: bab@mit.edu

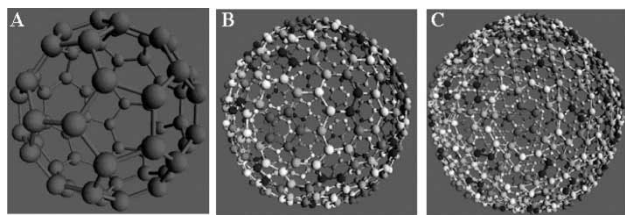


Figure 1. Examples of naturally-occurring capsid geometries produced by basic local rule simulations. A: $T = 1$ geometry. B: $T = 7$ geometry. C: $T = 13$ geometry.

theoretical advances in local rule simulation and considering the likely value of and challenges to local rule methods in future studies of capsid assembly.

2. Basic local rule models

The local rules model of Berger *et al.* [2] proposed that capsid assembly could be understood in terms of local protein–protein interactions between a small set of discrete coat protein conformations, each with a characteristic shape and set of binding sites. In this model, coat proteins select their conformations as they polymerize by examining only their immediate local binding environments in the shell. Local rules theory thus provided a sharp contrast to the prevailing “quasi-equivalence” model [1], which explained the existence of distinct binding environments in terms of elastic deformations of capsid proteins to minimize long-range, global stresses in the shell. In a local rule model, the behaviour of a given assembly system is described by a “rule set”, such as that shown in figure 2A, which encodes the geometry and binding specificities of each conformation. Assembly is proposed to proceed by repeated addition of coat proteins consistent with the rules. This process for the $T = 3$ example is shown in figure 2B–D and would ultimately terminate uniquely in the $T = 3$ geometry of figure 2E. A need for multiple conformations may explain why icosahedral capsids are often nearly spherical during assembly, as is possible when distinct binding environments have distinct conformations, but become very angular during post-assembly maturation, as would be necessary if the number of conformations decreased once the full set was

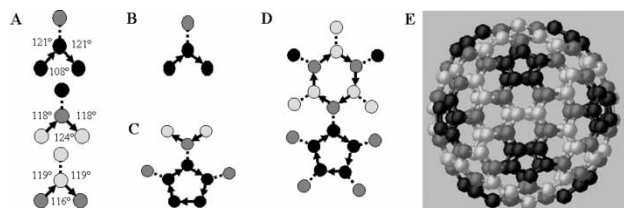


Figure 2. Illustration of local rules using a $T = 3$ geometry. A: A rule set specifying the shapes and binding configurations of three coat protein conformations. B–D: Expansion of growing a shell from a single coat protein through formation of, for example, a penton and a hexon. E: A complete $T = 3$ shell generated by the rule set. Different shades correspond to different conformations.

no longer needed for assembly. Note that a local rule model need not use monomers as its basis; rules describing assembly in terms of capsomer interactions might be more appropriate for some systems.

Although, the basic local rules model was intended as an explanation for capsid assembly in nature, it quickly proved valuable as a technique for simulating the assembly process. An initial local rule simulator allowed for the rapid development of many simplified models of different geometries, including some not explicable by quasi-equivalence theory. Figure 3 shows examples from early work with local rules: The previously known “ $T = 6$ ” geometry of the papova viruses [3], the “ $T = 2$ ” since observed in the Reoviridae family [4], and the so far unobserved “ $T = 5$ ” and “ $T = 8$ ”. Basic local rule simulations also allowed exploration of the robustness of assembly under the model, showing, for example, that a single large but local deformation can reliably lead to a dramatic global assembly error, while numerous smaller local errors could be tolerated without disrupting shell closure [2]. Local rule simulations also provided a basis for understanding aberrant structures formed from capsid proteins, helping to explain the molecular mechanisms by which altered assembly conditions might manipulate the rules of assembly [5].

3. Non-deterministic local rules

One important extension to local rules arose from the realization that rule sets might be ambiguous, allowing multiple possible partners for some binding sites. Figure 4A shows an example of an ambiguous rule set [6]. For the most part, real viruses under normal conditions produce a single defined structure, though, so we must assume that additional constraints reimpose determinism on the system. Nonetheless, these disambiguated rule sets can provide novel insights about capsid assembly that the original local rules models could not. For example, they can suggest possible constraints on assembly pathways needed to allow a well-defined structure to grow from a partially undefined rule set.

These simulations can also be revealing about relationships between distinct geometries. The rules from figure 4A can produce either $T = 4$ or $T = 7$ shells based on a minor change in how they are disambiguated (Figure 4C), potentially explaining several puzzling indications that $T = 7$ shells have a $T = 4$ “potential” built into them. It may also explain the hexon skew observed in $T = 7$ bacteriophages, in which hexons appear to have 180° rotational symmetry with a noticeable skew along one axis, consistent with the notion that they are made of a repeat of three conformations. Furthermore, we interpret the disambiguation rules (figure 4C) to reflect the action of scaffolding proteins, consistent with various indications that scaffolding proteins do select between $T = 4$ and $T = 7$ geometries under various circumstances (see [6]).

Ambiguous rule sets were also useful in early simulations of strategies for disrupting capsid assembly [13]. Ambiguous

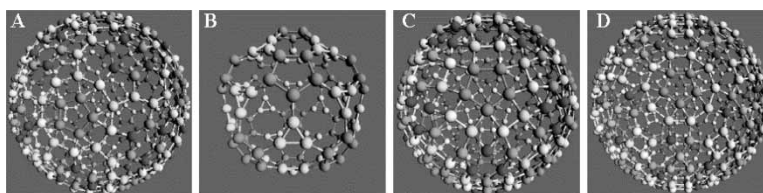


Figure 3. Examples of non-quasi-equivalent geometries produced by local rule simulations. A: ‘ $T = 6$ ’ geometry of papovaviruses. B: ‘ $T = 2$ ’ geometry of reoviruses. C: hypothetical ‘ $T = 5$ ’ geometry. D: hypothetical ‘ $T = 8$ ’ geometry.

rules were used to allow a probability of incorporating incorrectly functioning “poisoned subunits” into otherwise correctly assembling structures. Figure 5A shows a simulation using a poisoned subunit that distorts a growing capsid in a local region, leading to an overall failure of the capsid to close.

4. The local rules dynamics model

The next significant conceptual advance came from the recognition that local rules could be extended to describe interactions among an ensemble of partially formed subunits by encoding parameters of a binding kinetics model in addition to discrete binding specificities. The result was the local rules dynamics (LRD) model [7,8]. In an LRD simulation, we begin with a pool of subunits with defined positions and orientations in a three-dimensional space. Subunits diffuse via a model of Brownian motion and can bind with those that are physically close to them and have compatible binding sites according to a transition state model of energetics.

Extending local rules with a model of kinetics enabled various experiments impossible with more abstract local rules representations. LRD was used to examine the sensitivity of a model $T = 1$ system to variations in binding strengths, the tolerance of binding to deviations from ideal relative binding positions, and the subunit concentration [7]. Other experiments compared proposed models for the source of nucleation-limited assembly kinetics [9]. LRD simulations also revealed mechanisms by which interactions between partially formed capsids could lead to malformations. In simulations, such errors are favoured by high concentrations and lenient binding

tolerances, factors that also promote rapid growth, suggesting that a trade-off between rate and fidelity of assembly may be a limiting factor on the absolute efficiency possible for self-assembly systems [7].

The LRD model also allowed more detailed studies of strategies for capsid assembly targeted anti-virals. A significant hope of prior theoretical work on capsid assembly targeted drugs was that drugs could convert growing capsids into off-pathway “monster” assemblies that absorb far more coat proteins than do correctly formed shells. Such drugs could therefore be effective at stoichiometries of much less than one drug molecule per capsid. LRD simulations supported this strategy, showing how a hypothetical “poisoned subunit” with a distorted geometry could significantly disrupt overall on-pathway growth even at low concentrations (figure 5B and C).

5. Recent advances in local rules simulation methods

A new theoretical method for local rules modelling [10] was recently developed to address two key problems with the LRD model: it is too computationally intensive to simulate large numbers of subunits over long time scales and its energy model depends on parameters that are difficult to determine experimentally. This new method uses an N-fold way discrete-event model [11] to simulate the self-assembly process, similar in concept to recent work on fast, quantitatively accurate simulations of amorphous self-assembly systems [12]. Algorithmic improvements make this local rules N-fold way approach computationally tractable for even very large, complex systems. In this new work, as in a standard N-fold way simulation, systems are parameterized in terms of on- and off-rates of reactions. The

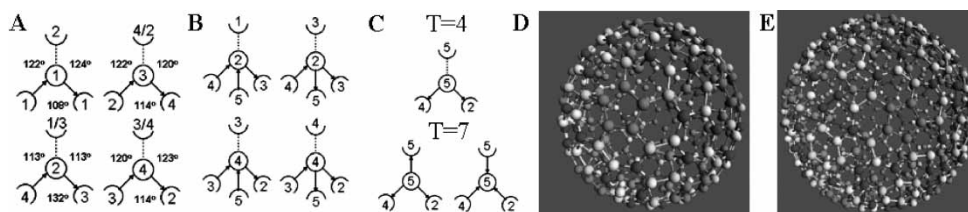


Figure 4. An ambiguous rule set that explains $T = 4$ and $T = 7$ geometries, where different conformations are numbered and a colour is associated with each number. A: Four rules, three of which have ambiguous binding specificities. B: Disambiguation of the rules by a fifth (scaffolding) conformation. C: Two possible rules for the scaffold conformation five, yielding either a symmetric dimer that forces a $T = 4$ geometry or an asymmetric dimer that forces a $T = 7$ geometry. D: $T = 4$ scaffold-containing shell produced by the first resolution. E: $T = 7$ scaffold-containing shell produced by the second resolution.

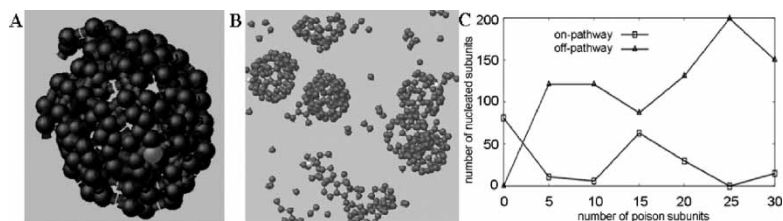


Figure 5. Applications of local rule-based simulations to evaluating strategies for capsid assembly targeted anti-virals. A: Shell disrupted by random addition of a “poisoned subunit”. B: LRD simulation of a pool of shells growing in the presence of a small number of poisoned subunits. C: Graph of on- and off-pathway growth as functions of numbers of poisoned subunits in a series of LRD simulations.

rates are encoded as parameters of exponential random variables representing the waiting time until a single pair of binding sites either bind or break apart. A simulation is then a form of continuous-time Markov model (CTMM) that steps between discrete binding and unbinding interactions, avoiding computations of particle trajectories between these discrete events. This concept generalizes easily to local rules by assuming that the rules encoding a conformation’s binding specificities also encode the parameters of the waiting time distributions. Unlike with the more sophisticated energy parameters of the LRD model, though, the waiting time parameters in the N-fold way approach are directly proportional to standard chemical rate constants and are thus comparatively easy to determine in practice.

While this new method has been implemented in a simulator, it has not yet been applied to realistic capsid assembly models. Nonetheless, work to date on simpler polymer systems confirms the model’s quantitative accuracy and efficiency [10]. We expect this work to be the basis for the next stage of local rule-based simulations of capsid assembly, by allowing for rapid simulations of complex capsids that can be matched to the detailed experimental data on capsid assembly kinetics that is now becoming available.

6. Discussion

Given their simplicity and computational tractability, local rule models still have important practical advantages in addressing unresolved issues in capsid assembly studies. Examples include the constraints on assembly pathways, the role of interactions between sub-assemblies, and the mechanisms of geometry selection. These questions remain difficult to approach experimentally and would greatly benefit from simulation tools that can implement proposed models and compare their behaviours to experimental measurements. The ability to rapidly implement such simulations for diverse geometries and assembly models has been a key strength of local rule-based methods. Recent theoretical work on fast, quantitative local rule simulation can be expected to lead to continued advances in theoretical studies of capsid assembly.

Nonetheless, it appears likely that there are limits to the local rules methodology. Local rule models may not be adaptable to some systems that lack the symmetry of icosahedral capsids or that rely on inherently non-local mechanisms, such as interactions between coat proteins and nucleic acid. As we move to greater realism in models, other non-local effects, such as a progressive reduction of diffusion rates of sub-assemblies as their sizes increase, may begin to have noticeable effects on simulation accuracy. The virtues of local rule models for simulation prototyping nonetheless suggest that we should continue to approach assembly as an emergent product of simple local interactions, compromising that model only as the experimental results demand it.

Acknowledgements

We thank Reidun Twarock for her support and the participants in the Mathematical Virology Workshop for helpful comments on this work. R. S. was supported in this work by US National Science Foundation grants EIA-0320595 and DBI-0346981. We also thank Jonathan King, Peter Prevelige and Robert Garcea for their involvement in various stages of this work.

References

- [1] 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.
- [2] Berger, B., Shor, P.W., Tucker-Kellogg, L. and King, J., 1994, Local rule-based theory of virus shell assembly. *Proc. Natl Acad. Sci. USA*, **91**, 7732–7736.
- [3] Rayment, I., Baker, T.S. and Caspar, D.L.D., 1982, Polyoma virus capsid structure at 22.5 Å resolution. *Nature*, **295**, 110–115.
- [4] Grimes, J.M., Burroughs, J.N., Gouet, P., Dirpose, J.M., Malby, R., Zientara, S., Mertens, P.P.C. and Stuart, D.I., 1998, The atomic structure of the bluetongue virus core. *Nature*, **395**, 470–478.
- [5] Schwartz, R., Garcea, R.L. and Berger, B., 2000, ‘Local rules’ theory applied to polyoma virus polymorphic capsid assemblies. *Virology*, **268**, 461–470.
- [6] Berger, B., King, J., Schwartz, R. and Shor, P.W., 2000, Local rule mechanism for selecting icosahedral shell geometry. *Discrete Appl. Math.*, **104**, 97–111.
- [7] Schwartz, R., Shor, P.W., Prevelige, P.E., Jr. and Berger, B., 1998, Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.*, **75**, 2626–2636.

- [8] Schwartz, R., 2000, The local rules dynamics model for self-assembly simulation, Massachusetts Institute of Technology Laboratory for Computer Science Technical Report 800.
- [9] Schwartz, R., Prevelige, P.E., Jr., and Berger, B., 1998, Local rules modeling of nucleation-limited virus capsid assembly, MIT Laboratory for Computer Science Technical Memo 584.
- [10] Jamalyaria, F., Rohlfis, R. and Schwartz, R., 2005, Queue-based method for efficient simulation of biological self-assembly systems. *J. Comput. Phys.*, **204**, 100–120.
- [11] Gillespie, D.T., 1976, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Chem. Phys.*, **81**, 2340–2361.
- [12] Laurenzi, I.J., Bartels, J.D. and Diamond, S.L., 2002, A general algorithm for exact simulation of multicomponent aggregation processes. *J. Comput. Phys.*, **177**, 418–449.
- [13] Prevelige, P., 1998, Inhibiting virus-capsid assembly by altering the polymerisation pathway. *Trends Biotechnol.*, **16**, 61–65.