Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 7732–7736, August 1994
Applied Mathematics

Local rule-based theory of virus shell assembly

BONNIE BERGER*†‡, PETER W. SHOR§, LISA TUCKER-KELLOGG†, AND JONATHAN KING¶

*Mathematics Department, †Laboratory for Computer Science, and ¶Biology Department, Massachusetts Institute of Technology, Cambridge, MA 02139; and §AT&T Bell Laboratories, Murray Hill, NJ 07974

Communicated by R. L. Graham, April 1, 1994

ABSTRACT **A local rule-based theory is developed which shows that the self-assembly of icosahedral virus shells may depend on only the lower-level interactions of a protein subunit with its neighbors—i.e., on local rules rather than on larger structural building blocks. The local rule theory provides a framework for understanding the assembly of icosahedral viruses. These include both viruses that fall in the quasi-equivalence theory of Caspar and Klug and the polyoma virus structure, which violates quasi-equivalence and has puzzled researchers since it was first observed. Local rules are essentially templates for energetically favorable arrangements. The tolerance margins for these rules are investigated through computer simulations. When these tolerance margins are exceeded in a particular way, the result is a "spiraling" malformation that has been observed in nature.**

The study of virus-shell structure and assembly is crucial for understanding how viruses reproduce. One notable aspect of virus shells is their highly regular structure: they are generally spherical and possess strong symmetry. Almost all human viruses and many plant and animal viruses have icosahedral shells (1, 2). These shells are constructed of repeated protein subunits, or coat proteins, which surround their condensed DNA or RNA genomes. A given shell usually consists of hundreds of copies of one protein, but sometimes copies of two or three different proteins.

Many of these viral shells appear to "self-assemble," or spontaneously polymerize in the host cell environment, with only limited aid from cellular machinery (3, 4). Sometimes assembly is assisted by scaffolding proteins, which assemble with the coat proteins to form a precursor shell but are removed before the shell matures. At first glance, shell assembly seems easy to understand because the structure is so regular. In fact, it has been difficult to determine the actual pathway through which hundreds of subunits interact to form a closed shell (5). This has been particularly difficult to explain for icosahedral viruses because often the same protein occurs in nonsymmetric positions.

Previous attempts at explaining the assembly process have focused on the icosahedral symmetry through the Caspar and Klug theory of "quasi-equivalence" (6). This theory classifies icosahedral shells whose protein subunits all have very similar (quasi-equivalent) neighborhoods and form hexamers and pentamers in the virus shell. The general belief was that shells were formed by assembly of these pentamer and hexamer building blocks. However, in the most closely analyzed experimental system for studying the assembly process, the bacterial virus P22, closed icosahedral shells assemble efficiently from purified monomeric protein subunits, even though the subunits are arranged as pentamers and hexamers in the final shell (7–9). This suggests that the emphasis on the final symmetry of the structure has been a barrier to understanding shell assembly.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

It was also generally believed that proteins took on only one conformation, particularly very stable proteins such as those that form virus shells. Recent evidence indicates that virus-shell proteins in fact take on several conformations (10–12) as has been proposed (5, 13). This important observation informs the approach to virus-shell assembly presented below.

The primary idea behind a local rule-based theory is that, if the protein subunits assume different conformations during the assembly process depending on their relative positions, a protein binding to the structure has enough local information to "know" where to bind. In particular, possible assembly pathways can be given that depend only on the interactions of a protein with its immediate neighbors rather than on larger structural building blocks.

Icosahedral Structure

All of the viruses discussed in this paper have what is called "icosahedral structure" (Fig. 1 *Left*). Caspar and Klug (6) pointed out the link between icosahedra and virus shells in their theory of quasi-equivalence, which classifies icosahedral shells according to their T number. Their definition of T number is equivalent to the number of subunits per corner of each triangular face; a virus thus has 60 T subunits altogether. Caspar and Klug assumed that these shells were formed of a hexagonal lattice with pentamers at the fivefold axes of symmetry and with the remaining subunits arranged in hexamers. A mathematical consequence of these assumptions is the restriction of the possible set of T numbers to the sequence 1, 3, 4, 7, 9, 12, 13, 16, 19, 21, 25, . . . ; these are the numbers of the form $f^2(h^2 + hk + k^2)$, where f, h, and k are nonnegative integers (6, 14, 15).

This paper represents these shells in a way that better illustrates local rules. For example, a $T = 1$ shell is typically viewed as an icosahedron except that, instead of having one protein at each vertex, it has a protein at each corner of each triangular face (Fig. 1 *Center*). The same structure can be redrawn by grouping the proteins at each vertex into pentamers (Fig. 1 *Right*). A graph representation of an icosahedral structure can be obtained by replacing the proteins with vertices and drawing an edge between two vertices when there is a binding interaction between the two proteins (Fig. 2). (For the purposes of abstraction, we refer to the interactions between two proteins, comprising electrostatic, van der Waals, or other noncovalent chemical interactions, as a single binding interaction.)

Local Rules

The local rule theory as applied to icosahedral structures is now described. For simplicity, we will assume virus shells contain a single kind of coat protein; the theory of assembly presented here works in all cases.

Abbreviation: SV40, simian virus 40.

‡To whom reprint requests should be addressed.

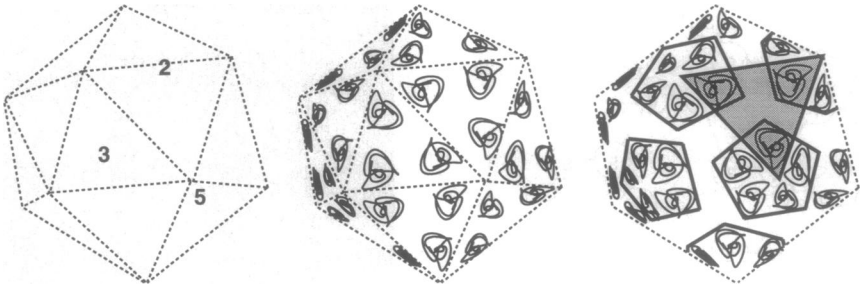

FIG. 1. (*Left*) An icosahedron has fivefold rotational symmetry at its 12 vertices, threefold rotational symmetry at its 20 triangular faces, and twofold rotational symmetry at its 30 edges. There are 60 symmetric regions in an icosahedron, each one lying in a third of a triangular face. (*Center*) Each triangular face has three proteins, one in each symmetric region. (*Right*) An icosahedral structure with the same symmetry as in *Center* but with pentameric clustering. One triangular group is shaded for contrast.

For each possible T number or shell size of an icosahedral virus, a set (or several possible alternative sets) of local rules exist that build the corresponding shell. These local rules are of the following form. We assume that identical protein subunits take on a small number of distinct conformations. The local rules then specify, for each conformation, which other conformations it can bind to and the approximate interaction angles, interaction lengths, and torsional angles that can occur between them. By following this local information, the subunits will form a closed icosahedral shell with the desired T number. Some sets of local rules require the assembly process to start with a given initiation complex to guarantee formation of the desired structure.

Local Rules for Quasi-Equivalent Viruses. The local rule theory can be illustrated through the example of the bacteriophage P22 virus shell, which is a $T = 7$ virus. Seven conformations of the coat protein, or shapes, have been observed in the P22 precursor capsid (12); however, it is not clear that these are all truly distinct. Let us first suppose that there are seven conformations. Fig. 3 *Top Left* gives the rules for how one of these, the type 1 conformation, chemically binds. A type 1 conformation has a binding site for a type 2 conformation and two binding sites for type 1 conformations. Similar local rules can be constructed for all the seven conformations in P22 (Fig. 3). The binding interactions in the local rules are present in micrographs of the shell; however, additional interactions may also be present (12), which may have only a secondary effect on the assembly process.

As soon as a subunit has at least one binding interaction, these rules can be applied unambiguously to determine the subunit's remaining neighbors. The different orders of applying local rules give the possible ways in which the assembly process might proceed. While it would be consistent with the local rules that pentamers and hexamers initially form and then bind together as previously believed, this is not required by the theory. Chemically speaking, the local rules do not dictate which event comes first: a protein adopting a conformation or a protein acquiring a binding interaction.

The question remains, what structures can be built if these local rules must be respected? Applying the local rules to an arbitrary starting protein can result in a $T = 7$ shell or some subset of the shell, but nothing else (Fig. 4). Computer simulations verified this fact for the local rules in Fig. 3. The simulations worked as follows: An energy model was set up by assuming a quadratic penalty for deviations from the interaction angles, torsional angles, and interaction lengths given in the rules. An existing binding site was chosen as the site to attach the next protein; if no candidate proteins able to attach were in the existing structure within one-protein diameter of the binding site, a new protein was added. The local rules determined the conformation and location of each new protein. After a protein was added, the resulting structure was optimized to minimize energy by iterating optimization steps. In each step, all of the proteins were moved in accordance with the forces and torques computed from the energy model. The binding sites were examined in both

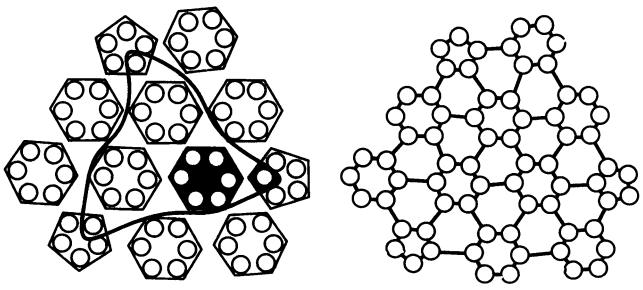

FIG. 2. (*Left*) A portion of a $T = 7$ virus shell, with the seven subunits in a corner unit shaded, the pentamers and hexamers drawn in light lines, and the triangular face in a heavy curved line. The protein subunits are depicted as circles. (*Right*) The same overall structure as in *Left* but redrawn in a graph representation to emphasize binding interactions. Every protein is a vertex, and every binding interaction is an edge.

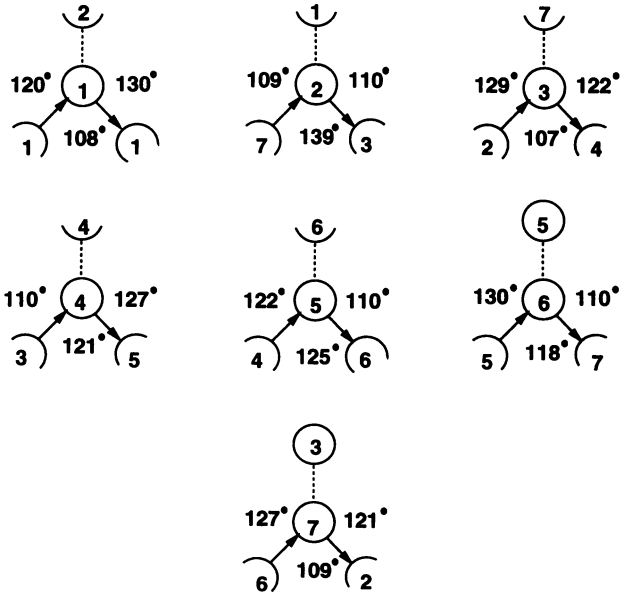

FIG. 3. Possible local rules for a left-handed $T = 7$ virus. Each protein subunit is represented as a circle or part of a circle labeled with its conformation. There is sometimes a direction associated with each edge. Angles between binding interactions are the approximate number of degrees between the centers of the protein subunits in three dimensions. Angles were not based on any particular virus but were first derived from a physical model of a spherical $T = 7$ structure and were subsequently refined by using the results of a computer simulation.

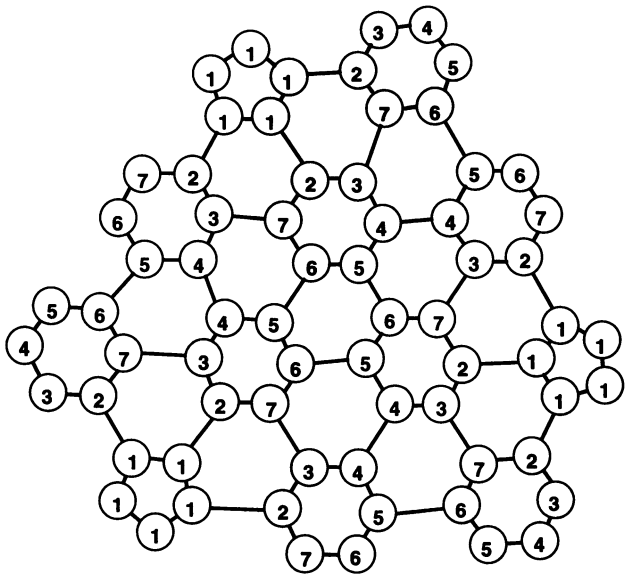

FIG. 4. Same overall structure as in Fig. 2 *Right* but redrawn to emphasize local rules.

random and breadth first orders, in each case resulting in the formation of a closed shell.

Computer simulations show that the local rules are relatively robust. Even initial rules offset from the rules in Fig. 3 by a randomly selected amount of up to 9.6° (about 8%) for each rule angle and 8% for each interaction length led to the formation of a nearly identical closed shell in three-dimensional space (Fig. 5). If the angles were changed by up to 10%, the shell failed to close in approximately half the trials; but it still looked very similar when it closed. Through more substantial (nonrandom) changes in the local rules, a virus' shell can vary between spherical and polyhedral shapes.

Local rule theories can be constructed for all T numbers. There is always a set of local rules with the number of conformations equal to the T number. Sets of local rules that use fewer conformations also exist; these sets assign the same conformation to nonequivalent positions. An alternate set of local rules for $T = 7$, using only four conformations, is given in Fig. 6. In this set of $T = 7$ rules, the hexamers are symmetric under rotations of 180°. Micrographs of P22 precursor capsids similarly show the near-symmetry of the hexamers under 180° rotations (12). Always allowing the disallowed hexagon in Fig. 6 would give a set of rules for a $T = 4$ shell. In fact, the coat proteins of three $T = 7$ bacteriophages can also form $T = 4$ shells (16–19).

Another well-studied class of icosahedral viruses are the $T = 3$ plant viruses (10, 20, 21). Several theories for their assembly have been advanced (22, 23). Although these $T = 3$ virus shells have three nonequivalent positions, the proteins in two of these positions assume quite similar conformations (21, 23). These are labeled "1" in the graph representation in Fig. 7, while proteins in the third position are labeled "2." A set of rules can be extracted from this representation that permits both $T = 3$ and $T = 1$ shells. In fact, the coat proteins of many of these viruses can form $T = 1$ shells (22). However, as similarly noted (23), if assembly is initiated by a structure containing a type 2 conformation, these will propagate during assembly to uniquely determine the $T = 3$ structure.

Local Rules for a Non-Quasi-Equivalent Virus. The failure of the tumor-linked *Polyomavirus* species to fit into the quasi-equivalent framework has been a much debated point in structural virology. These viruses have 360-subunit shells consisting entirely of pentamers, some of which contact five

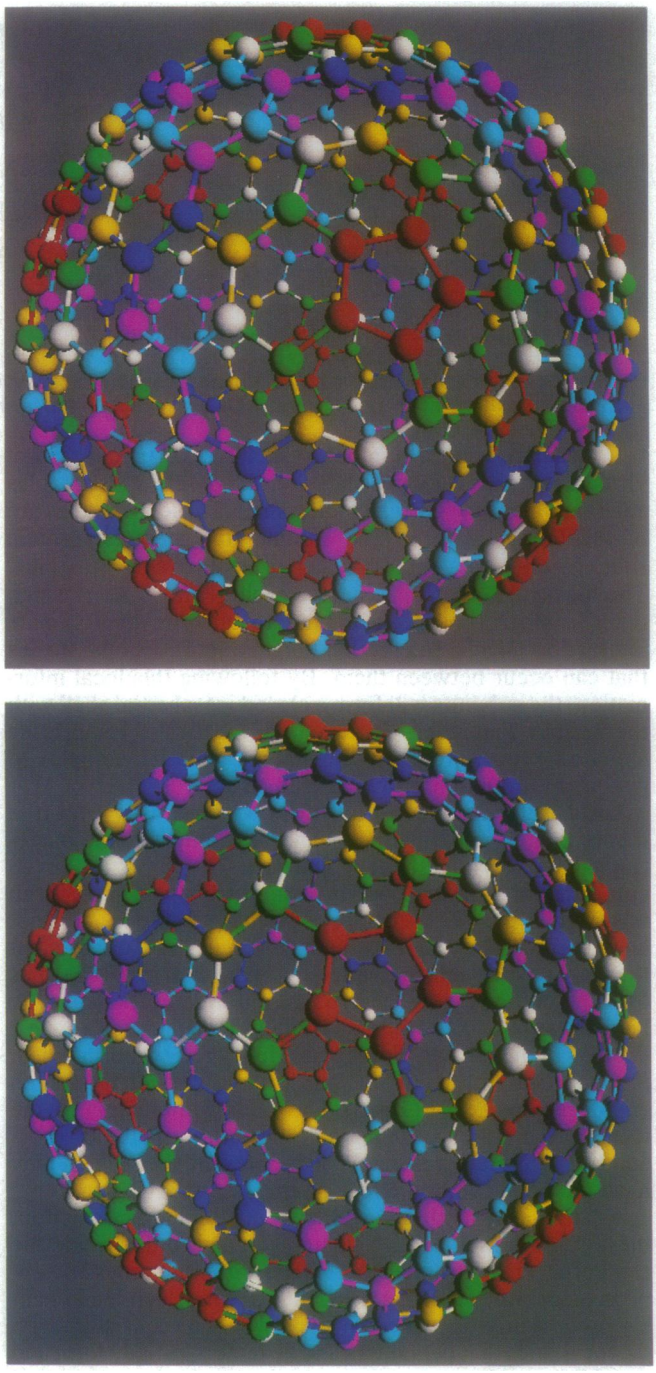

FIG. 5. (*Upper*) Silicon Graphics INDIGO 2 image of the shell resulting from the rules in Fig. 3. (*Lower*) The same figure as in *Upper* except that the structure was formed from randomly perturbed rules, offset up to 8% from the rules that formed the structure in *Upper*. Note that the two structures look nearly identical.

other pentamers and some of which contact six other pentamers (Fig. 8). One can thus view this as a $T = 6$ structure, a T number disallowed by the theory of quasi-equivalence. Research (11, 24, 25) on this structure has focused on how pentamers could be hexavalent and on how the same protein can occupy very asymmetric environments. Liddington *et al.* (11) postulated that assembly occurs by forming pentamers that are subsequently tied together. In what follows, we apply the local rule theory to a polyoma virus, simian virus 40 (SV40), to produce a new hypothesis for its assembly. As remarked, local rule theories could apply both in the case where monomers assemble directly to form the shell and

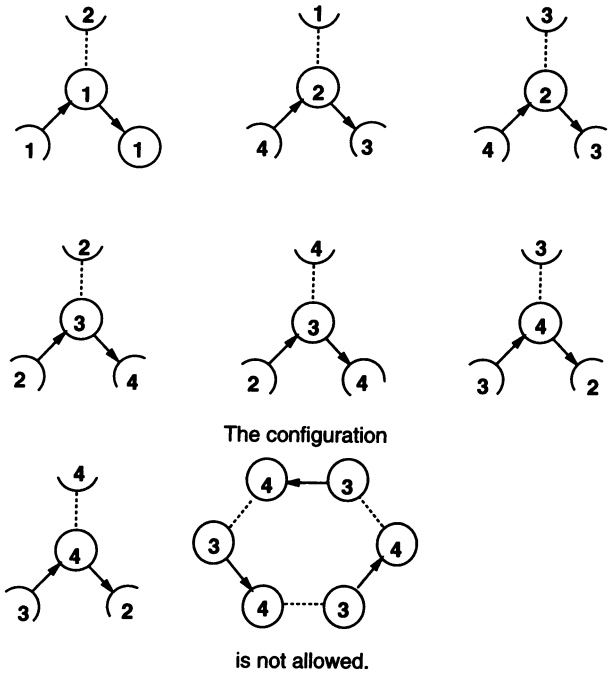

FIG. 6. A second set of local rules for assembly of a left-handed $T = 7$ virus. These rules produce the structure in Fig. 4 with conformations 5, 6, and 7 replaced by 2, 3, and 4, respectively. Solid arrows are binding interactions within capsomeres; dotted lines, between capsomeres. We assume that the shell is initiated at a pentamer and that a protein does not assume its final configuration until there is an adjacent protein in its capsomere. A possible mechanism for the disallowed-hexagon rule is that the three type 4 conformations, which could be spatially adjacent, form a trimer of higher energy than a trimer of two type 4 and one type 2 conformations; alternatively, the rule could be enforced through interactions with the scaffolding proteins.

where they first assemble into substructures, which then come together to form the shell.

Local rules for SV40 can be constructed that are not substantially different than for other icosahedral viruses. It could simply have six local rules (Fig. 9), one for each of its conformations. These rules guarantee the final form: computer simulations show that applying the rules in Fig. 9 in any order to an initial subunit will result in the same pattern of interconnectivity as in Fig. 8.

For SV40, six protein conformations have been observed, but the binding interactions are more complicated than as indicated by the local rule theory (11). The function of the C-terminal arms of the SV40 coat protein has been described

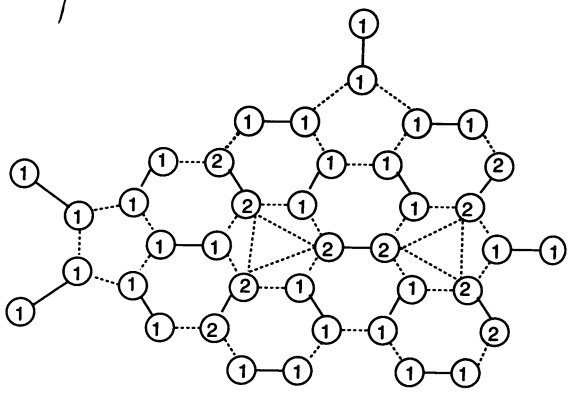

FIG. 7. A graph-based representation for $T = 3$ plant viruses. Not all binding interactions are shown, but the interactions shown are sufficient to abstract a set of local rules.

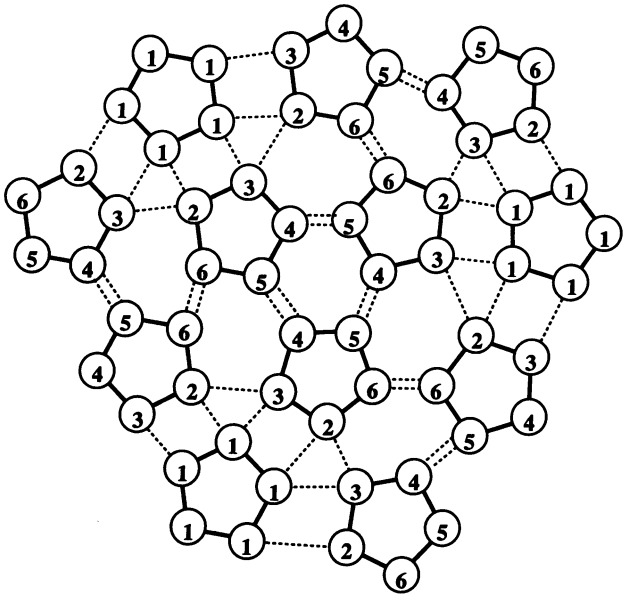

FIG. 8. A simplified diagram of how the coat proteins in the polyoma virus SV40 shell connect with each other.

as "tying together" the pentameric building blocks (11); these arms may also play the dual role of enforcing the binding interactions of the local rules.

Closure and Malformation. Although the above discussion might suggest that closure is easily assured, simulations show that a spiraling malformation can occur if local rules are "broken" just once. Such incorrectly polymerized spiral structures have been observed for P22 and other viruses (16, 26, 27).

This work provides a possible explanation for spiral structures. Suppose that somehow a P22 shell starts with six type 1 subunits, instead of five, fitting together to form a capsomere. If the local rules were correctly followed thereafter, this hexamer would next be surrounded with six hexamers instead of five. This region of the shell consisting solely of hexamers will be relatively planar, but the regions growing around it will have the normal radius of curvature. When the sides have curved 180°, they will not be near enough to close (Fig. 10). One side may curl inward, and the second may form an outer layer around it. Computer experiments show that if local rules are broken in this way, spiraling can indeed occur.

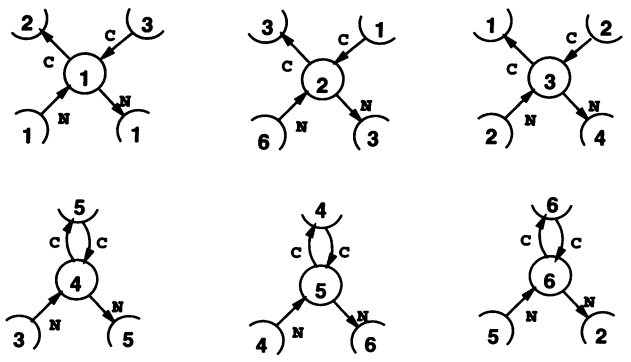

FIG. 9. Local rules for the SV40 virus. Each protein subunit is labeled with the type of its conformation. The double, directed edges could be simplified to a single edge; they are drawn as double edges to correspond with the known biological structure (11). Each inter-pentameric-binding interaction is a C-terminal arm of the protein subunit, labeled with a direction to indicate which subunit it is from. Each intra-pentameric-binding interaction is an N-terminal arm.

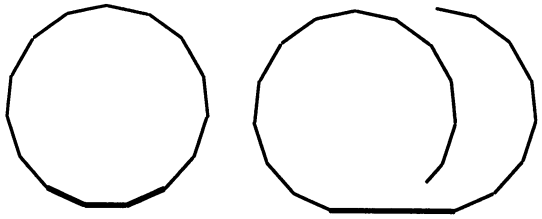

FIG. 10. A cross-sectional two-dimensional analog of spiraling. (*Left*) A spherical shape is constructed from segments with regular curvature. (*Right*) A region without curvature is created at the bottom of the sphere, but subsequent growth retains the regular curvature. The resulting structure does not close.

DISCUSSION

The local rule formulation implies conformational flexibility for the precursor subunits that are going to assemble into a shell. Since multiple conformations of the same subunit have now been observed for many mature virions, subunit flexibility is not unpalatable. Until recently, it has not been possible to directly measure the stability of precursor subunits. However, this has recently been determined for the P22 coat subunit. The assembled capsid lattice is very stable with a melting temperature (T_m) of 87°C (28). However, the subunit prior to polymerization is only marginally stable, with a broad melting transition with a midpoint at 40°C (M. Galisteo, C. Gordon, and J.K., unpublished results). This low thermal stability is consistent with conformational flexibility; alternatively, the actual precursor states of coat subunits may resemble a folding intermediate, not yet locked into a mature conformation.

Local rules may help in the determination of virus structures. A virus might be hypothesized to obey a given combinatorial set of local rules for assembly. This could imply that certain non-quasi-equivalent proteins are in similar conformations, knowledge that could aid in the determination of structure. Local rules may also help in identifying likely positions for scaffolding proteins. For instance, if the local rules in Fig. 6 control assembly in P22, it seems likely that the positions of scaffolding proteins are nearly symmetric under 180° rotation of the hexamers. A hypothesis that achieves this and is consistent with current estimates for the number of scaffolding proteins (29, 30) is that four scaffolding proteins are associated with each hexamer and five with each pentamer.

Previous attempts at interfering with the infection process have mainly focused on interrupting infection by a fully formed shell at the binding site. The local rules tell us that if we can interfere with a single binding interaction, the shells may not close. Recent experiments indicate that the subunit assembly process may be a sensitive locus for inhibitors of virus assembly (31).

Dan Kleitman has been generous with his time, ideas, and support. Discussions with Peter Prevelige and Wah Chiu were informative. Doug Muir's programming assistance has been a great help. B.B. was supported in part by the Radcliffe Bunting Institute through an Office of Naval Research grant and by Massachusetts Institute of Technology Grants AF 78181 and DARPA N00014-92-J-1799; L.T.-K., by a National Science Foundation predoctoral fellowship and Air Force Grant 78181; and J.K., by National Institutes of Health Grant GM 17980.

1. Baker, T. S., Caspar, D. L. D. & Murakami, W. T. (1983) *Nature (London)* **303**, 446–448.
2. Fields, B. N., ed. (1985) *Virology* (Raven, New York).
3. Fraenkel-Conrat, H. & Williams, R. C. (1955) *Proc. Natl. Acad. Sci. USA* **41**, 690–698.
4. Casjens, S. & King, J. (1975) *Annu. Rev. Biochem.* **44**, 555–611.
5. Rossmann, M. G. (1984) *Virology* **134**, 1–13.
6. Caspar, D. L. D. & Klug, A. (1962) *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.
7. Fuller, M. T. & King, J. (1982) *J. Mol. Biol.* **156**, 633–665.
8. Prevelige, P., Thomas, D. & King, J. (1988) *J. Mol. Biol.* **202**, 743–757.
9. Prevelige, P. E., Jr., Thomas, D. & King, J. (1993) *Biophys. J.* **64**, 824–835.
10. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978) *Nature (London)* **276**, 368–373.
11. Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991) *Nature (London)* **354**, 278–284.
12. Prasad, B. V. V., Prevelige, P. E., Jr., Marietta, E., Chen, R. O., Thomas, D., King, J. & Chiu, W. (1993) *J. Mol. Biol.* **231**, 65–74.
13. Caspar, D. L. D. (1980) *Biophys. J.* **232**, 103–138.
14. Casjens, S. (1985) in *Virus Structure and Assembly*, ed. Casjens, S. (Jones and Bartlett, Boston), pp. 1–28.
15. Branden, C. & Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York), Chap. 11, pp. 161–177.
16. Earnshaw, W. & King, J. (1978) *J. Mol. Biol.* **126**, 721–747.
17. Katsura, I. (1983) *J. Mol. Biol.* **171**, 297–317.
18. Dokland, T., Lindqvist, B. H. & Fuller, S. D. (1992) *EMBO J.* **11**, 839–846.
19. Agarwal, M., Arthur, M., Arbeit, R. D. & Goldstein, R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2428–2432.
20. Hogle, J. M., Maeda, A. & Harrison, S. C. (1986) *J. Mol. Biol.* **191**, 625–638.
21. Silva, A. M. & Rossmann, M. G. (1987) *J. Mol. Biol.* **197**, 69–87.
22. Rossmann, M. G. & Erickson, J. W. (1985) in *Virus Structure and Assembly*, ed. Casjens, S. (Jones and Bartlett, Boston), pp. 29–74.
23. Sorger, P. K., Stockley, P. G. & Harrison, S. C. (1986) *J. Mol. Biol.* **191**, 639–658.
24. Rayment, I., Baker, T. S. & Caspar, D. L. D. (1982) *Nature (London)* **295**, 110–115.
25. Patch, C. T., Levine, A. S. & Lewis, A. M., Jr. (1979) in *Comprehensive Virology*, eds. Fraenkel-Conrat, H. & Wagner, R. C. (Plenum, New York), Vol. 13, pp. 459–541.
26. King, J., Griffin-Shea, R. & Fuller, M. T. (1980) *Q. Rev. Biol.* **55**, 369–393.
27. Hendrix, R. W. (1985) in *Virus Structure and Assembly*, ed. Casjens, S. (Jones and Bartlett, Boston), pp. 169–204.
28. Galisteo, M. & King, J. (1993) *Biophys. J.* **65**, 227–235.
29. King, J. & Casjens, S. (1974) *Nature (London)* **251**, 112–119.
30. Eppler, K., Wyckoff, E., Goates, J., Parr, R. & Casjens, S. (1991) *Virology* **183**, 519–538.
31. Teschke, C. M., King, J. & Prevelige, P. (1993) *Biochemistry* **132**, 10658–10665.