A Local Rule Based Theory of Virus Shell Assembly

Bonnie Berger[*†] Peter W. Shor[‡] Lisa Tucker-Kellogg[†] Jonathan King[§]

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

Abstract

A local rule 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. *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.

1 Introduction

The study of virus shell structure and assembly is crucial for understanding how viruses reproduce and how anti-viral drugs might interfere with the assembly of virus shells. One of the most notable aspects of virus shells is their highly regular structure: they are generally spherical and possess strong symmetry properties. Almost all human viruses, and many plant and animal viruses, are icosahedral [11]. These include the rhinovirus, poliovirus, and herpesvirus, all of which have rounded icosahedral shells. 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 are believed to assemble with only limited aid from cellular machinery; they appear to "self-assemble," or spontaneously polymerize and take shape, in the host cell environment. This was originally established for the rod shaped tobacco mosaic virus [12] and has since been shown for many spherical viruses [5]. 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, the assembly of the shells seems easy to understand, because the structure is so regular. In fact, it has been difficult to determine the actual pathway through which the subunits interact to form a closed shell composed of hundreds of subunits [27]. In icosahedral viruses this has been particularly difficult to explain because very often the same protein occurs in non-symmetric positions [27].

Previous attempts at explaining the formation of closed icosahedral shells from subunits have focused on the icosahedral symmetry, through the Caspar and Klug theory of *quasi-equivalence* [7]. 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 subunit 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 structures [13, 24, 25]. This suggests that the emphasis on the final symmetry of the structure has been a barrier to understanding how these proteins assemble into such complex structures.

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 [15, 21, 23], as has been proposed by [6, 27]. This important observation informs the approach to virus shell assembly presented below.

It is useful to consider what an individual protein "knows," from an information-theoretic point of view. It might appear that each protein needs to know something about the global structure. In fact, if the proteins assume different conformations during the assembly process depending on their relative positions, each protein has enough local information to "know" where to bond. Thus, a protein needs to know nothing about what is going on in the rest of the structure in order to form icosahedral shells.

In particular, we show that possible assembly pathways can be given that depend on only the interactions of a protein with its immediate neighbors, rather than on larger structural building blocks. This leads to simple, local rule theories that suffice to explain assembly. The shell proteins of each virus obey a set of local rules which direct assembly. From a mathematical point of view, these rules can be applied in any order, and the same shell results; thus, many assembly pathways could work. For the $T = 7$ shell (the T number classifies the combinatorial structure of the shell), two combinatorially different sets of local rules are given. One uses seven conformations of the coat protein; the other uses only four. Currently there is not adequate experimental evidence to say that one of these sets directs assembly in any particular virus.

A complete set of local rules for a virus, in addition to the combinatorial structure of the rules, must also specify interaction angles in 3D space, torsional angles, and interaction lengths. These will determine the exact shape of the assembled structure. Two sets of rules that are approximately the same tend to produce nearly identical final shapes, whereas those that are very different from each other produce different shapes. Computer simulations can show the relationship between the angles and bond lengths of the local rules and the consequent virus shape. For the $T = 7$ shell, computer investigations of the interaction angles and lengths indicate that if these are changed by less than 8%, the final shapes will be nearly identical.

Local rule theories are not limited to viruses that conform to the theory of quasi-equivalence. Recently, investigators have discovered viruses with unusual icosahedral symmetries [21, 26, 22], such as the polyomavirus, which causes tumors in mice. These have have five-sided building blocks surrounded by *six* neighboring building blocks. Little is understood about the assembly of these "non-quasi-equivalent" viruses. The distinction between quasi-equivalence and non-quasi-equivalence is not so important if viewed in terms of local rules for assembly. Because local rules break the quasi-symmetry during the assembly process in any case, the fact that this symmetry is broken in non-quasi-equivalent viruses does not affect the local rule hypothesis. In particular, a set of local rules is presented that completely and uniquely determines the final conformation of the polyomavirus.

For a number of viruses, incorrect or malformed assemblages of coat protein subunits have been described, including tubular and spiral variants. The coat subunits in these structures are normal, but there have been errors in their interactions with each other [9, 20]. Preliminary computer simulations show that if the local rules are distorted in certain ways, or if certain mistakes are made in the assembly process, spiral structures can form.

2 Definitions and Motivation

One striking aspect of virus shells is their strong symmetry properties. All the viruses discussed in this paper have what is called *icosahedral structure.* They are based on the icosahedron (figure **??**a), a mathematical solid, and they have the same symmetry properties as the icosahedron.

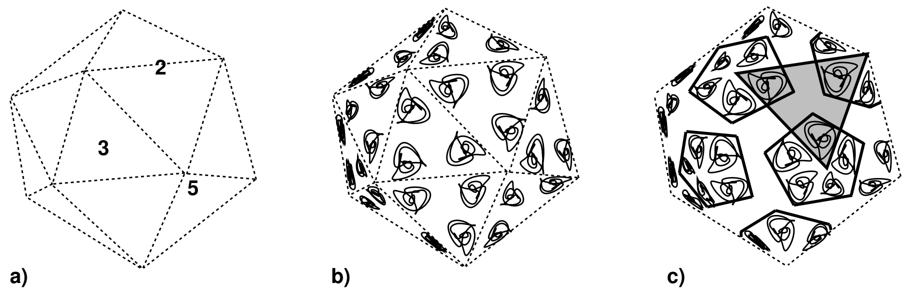

Figure 1. a) An icosahedron has 5-fold rotational symmetry at its 12 vertices, 3-fold rotational symmetry at its 20 triangular faces, and 2-fold rotational symmetry at its 30 edges. There are 60 symmetric regions in an icosahedron, each one lying in a third of a triangular face. b) Each triangular face has three proteins, one in each symmetric region. This icosahedral virus shell is made of 20 triangular faces and 60 identical proteins. c) The same icosahedral structure as in (b) but with pentameric clustering. Heavy lines group the proteins into pentamers around the 5-fold axes of symmetry. One triangular group is shaded for contrast.

Caspar and Klug [7] were the first to point out the link between icosahedra and icosahedral virus shells in their theory of quasi-equivalence. They classify icosahedral shells according to their *T number.* Their definition of T number is equivalent to the following: The T number of an icosahedral shell is the number of proteins per corner of each triangular face.

The theory of quasi-equivalence classifies icosahedral shells whose protein subunits all have very similar (quasi-equivalent) neighborhoods. Caspar and Klug assume that every protein subunit is a member of a hexamer or a pentamer, that the hexamers are arranged in a hexagonal lattice, and that there are exactly 12 pentamers, which lie on the five-fold axes of symmetry of the icosahedral shell. These assumptions are derived from the hypothesis that all subunits lie in nearly identical neighborhoods. Given these assumptions, it follows that there are only a limited number of possibilities for shell structures. A mathematical consequence is the restriction of the possible set of T numbers: the only ones allowed are $1, 3, 4, 7, 9, 12, 13, 16, 19, 21, 25, \ldots$. These are numbers of the form $f^2(h^2 + hk + k^2)$, where f, h, and k are non-negative integers [7, 4, 3].

This paper represents these shells in a way that better illustrates local rules. For example, the $T = 1$ shell of satellite tobacco necrosis virus [4] 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 (figure 1b). The same structure can be redrawn by grouping the proteins at each vertex into pentamers, as in figure 1c. A graph representation of an icosahedral structure can be obtained by replacing the proteins with nodes and drawing an edge between two nodes when there is a bond interaction[1] between the two proteins (Figure 2). All these structures still have icosahedral symmetry.

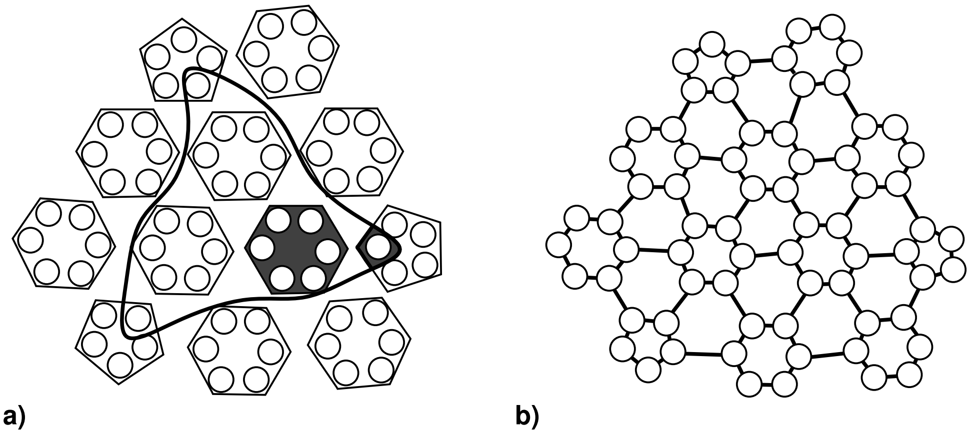

Figure 2.a) A portion of a $T = 7$ virus shell with the seven subunits in a corner unit shaded, the pentamers and hexamers drawn in light lines, and the triangular face in a heavy curved line. The protein subunits are depicted as circles. b) The same overall structure as (a), but redrawn in a graph representation to emphasize bond interactions, rather than the pentameric and hexameric building blocks. Every protein is a node and every bond is an edge.

The focus on quasi-equivalence has led to the restriction on T numbers, which may be somewhat

[1] For the purposes of abstraction, we refer to the bond interactions between two proteins, comprised of electrostatic, van der Waals, or other non-covalent chemical interactions, as a single bond.

artificial. These constraints have grounding in mathematical symmetry, but they are not required by nature, because not all viruses satisfy Caspar and Klug's hypotheses. There is no physical reason for ruling out $T=2$, for example. Indeed, $T=6$, which will be discussed later, occurs in nature but does not fit the quasi-equivalent format. However, most viruses do fit into this format.

3 Local Rules

An alternative hypothesis is described for how icosahedral structures form, based solely on simple local rules for determining how proteins interact. For simplicity, we will assume virus shells contain a single type of coat protein; the theory of assembly presented here works in all cases.

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; that is, there are seven proteins in positions that are not equivalent under icosahedral symmetry, giving 420 proteins overall.

Seven distinct conformations of the coat protein, or shapes, have been observed in the P22 precursor capsid [23]. One of these seven conformations will be considered first. Figure 3 gives the rules for how the first conformation chemically bonds. The type 1 conformation is in the center.

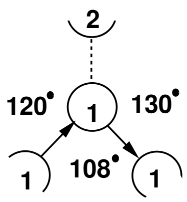

Figure 3. A representation of a local rule for the type 1 conformation of P22. Each protein subunit is represented as a circle or a part of a circle labeled with its conformation. Angles between bond interactions are the approximate number of degrees between the centers of the protein subunits, obtained from computer simulations. Bonds are represented as unit length, possibly with an associated direction. One way to think about the bonds is that the protein is sticking its arm out into another protein, like a key into a lock. For many viruses, this interpretation is consistent with experimental observations.

Given the bond to the type 2 neighbor, then, at a position clockwise from this bond at about an angle of 130°, only a type 1 conformation can attach. Similarly, only a type 1 conformation can attach at an angle of about 108° from this latter bond. We call this representation the type 1 local rule.

Similar local rules can be constructed for all the seven conformations in bacteriophage P22 (figure 4). For these rules, all subunits with the same conformation will have the same bond

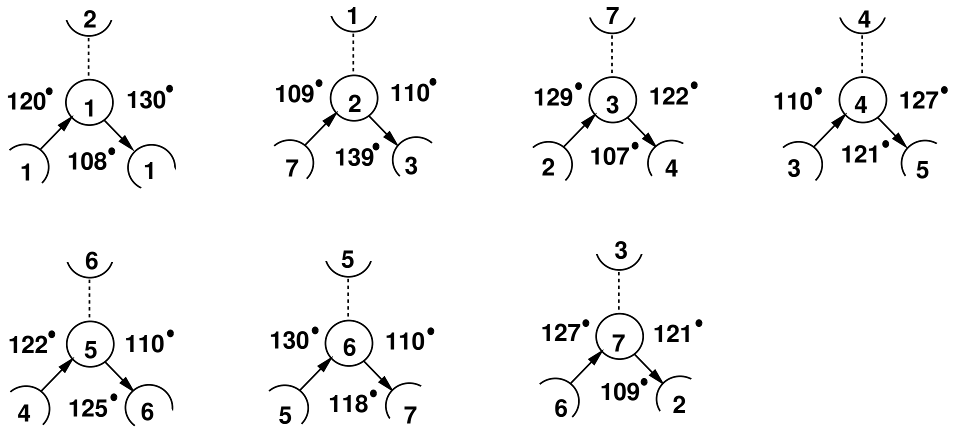

Figure 4. Possible local rules for a left-handed $T=7$ virus. Angles are not based on any particular virus, but are derived from a computer simulation.

interactions. The conformations and bonds in the local rules can be determined experimentally through micrographs of the shell [23].

As soon as a subunit has at least one bond interaction, these rules can be applied unambiguously to determine the subunit's remaining neighbors. The different orders in which local rules can be applied give all the possible ways in which the assembly process might proceed. In figure 5, two such possible examples are illustrated. It is also conceivable that the rules can be applied simultaneously.

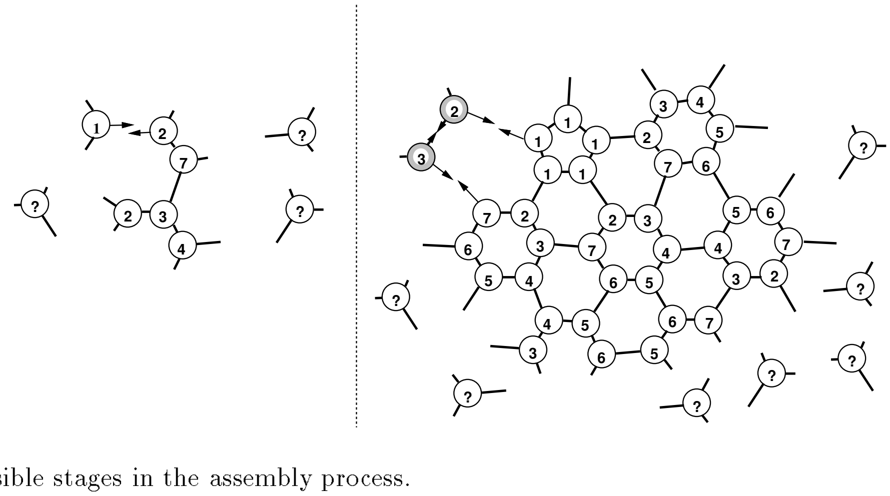

Figure 5. Possible stages in the assembly process.

While it would be consistent with the local rules that pentamers and hexamers initially form and later bond 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 and then forming bond interactions as specified, a protein acquiring a bond interaction and then being forced into the corresponding conformation, or some cooperative combination. It is also possible that a protein does not adopt the conformation specified by the rules until after several of its neighbors have arrived.

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 structure resembling the $T = 7$ shell, or some subset of the shell, but nothing else (figure 6). This has been verified by computer simulations.

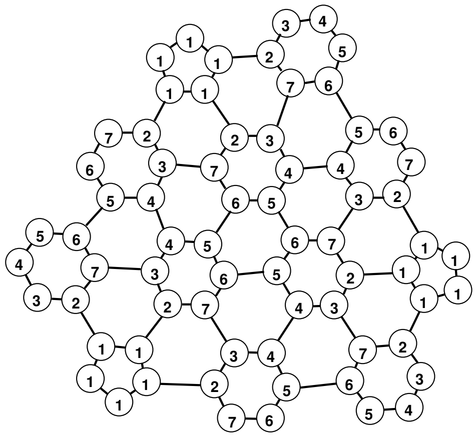

Figure 6. The same overall structure as in figure 2b, but redrawn to emphasize local rules.

For the computer simulations, an energy model was set up assuming a quadratic penalty for deviations from the interaction angles, torsional angles, and bond lengths given in the rules. The proteins were added to existing bonding sites; if there were no candidates able to attach in the existing structure within one protein diameter of the bonding site, a new protein was added. The local rules were used to determine 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 the proteins were moved in accordance with the forces and torques computed from the energy model. For the set of rules in Figure 4, the angles were at first derived from a physical model of a spherical $T = 7$ structure and subsequently refined by using the results of the first computer simulation. The rule angles and lengths could likewise be derived from experimental data or refinement of hypothetical values through computer simulation. The bonding sites were examined both in random and breadth-first orders, in each case resulting in the formation of a closed shell.

Computer simulations thus far show that the local rules are relatively robust. Even initial rules offset from the rules in figure 4 by a randomly selected amount of up to 9.6° (about 8%) for each rule angle and 8% for each bond length lead to the formation of a closed shell in 3D space, which looks nearly identical to the one formed by the original rules. If the angles were changed by up to 10%, the shell failed to close in approximately half the trials; but when it closed, it still looked very similar to the original shell. (Clearly, these numbers depend on the underlying assumptions of the algorithm.) Through more substantial (non-random) changes in the local rules, a virus' shell can vary between spherical and polyhedral shapes.

In a local rule theory, a protein "knows" where it is in the shell by looking at the conformations of its neighbors. Thus, any set of local rules for structures with $T > 1$ requires a number of different conformations; otherwise, a protein bonding to the structure will not "know" in which of the various non-equivalent positions it will be and thus whether it should be part of a pentamer or a hexamer. These different conformations need not be maintained in the mature form of the virus. For P22, the differences between the protein conformations in non-equivalent positions are noticeably less in the mature form than in the precursor form [23]. Perhaps functionally different protein conformations are required for assembly, while in the mature form all the proteins assume the same functionality and need only be different enough to hold the shell together stably. This may also be the reason for the substantial changes that other bacteriophages undergo between their precursor and mature forms [16].

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. However, in the case of higher T number structures, this many conformations would be infeasibly large. The number of conformations can sometimes be reduced by using rules that assign the same conformation to non-equivalent positions. An alternate set of local rules for $T = 7$, using only four conformations, is given in figure 7. There is a trade-off for using fewer conformations in that this set of rules is not as robust as the set given in Figure 4, and the last rule (the forbidden hexagon) may require a more complicated control mechanism.

In the alternate set of $T = 7$ rules, the hexamers are symmetric under rotations of 180°. This is intriguing because the micrographs of P22 show near-symmetry of the hexamers under 180° rotations [23]. It is possible that this alternate set of rules is responsible for determining the assembly of P22, and that the reason seven different conformations of the protein are observed is that functionally equivalent conformations in non-equivalent positions are responding to different

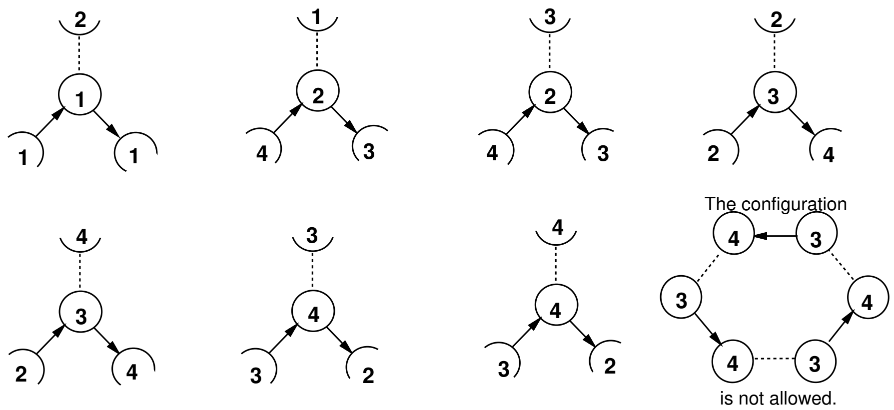

Figure 7. A second set of local rules for assembly of a left-handed $T = 7$ virus. Solid lines are bonds within capsomeres; dotted lines are bonds between capsomeres. We assume that the shell is initiated at a pentamer and that a protein does not assume its final configuration until there is at least one other protein in the same capsomere. The disallowed configuration might be impossible because the sum of the angles around the hexagon is too large or too small to allow it to close. Alternatively, since the three type 4 conformations in the hexagon are spatially adjacent, they may form a trimer that has higher energy than the 4-4-2 trimer. These rules produce the structure in Figure 6 but with conformations 5, 6, and 7 replaced by 2, 3, and 4, respectively.

stresses.

This alternate set of local rules for $T=7$ is nearly the same as a set of local rules for $T=4$. The forbidden-hexagon rule can be changed to give a set of rules for a $T=4$ shell. In fact, for three $T=7$ bacteriophages the coat proteins can also form $T=4$ shells. One of the mistakes observed in the assembly of P22 in the absence of scaffolding proteins is the formation of a $T=4$ shell instead of a $T=7$ structure [9]. In bacteriophage λ, mutations of the coat protein exist that form functional $T=4$ shells [18]. Finally, the $T=7$ bacteriophage P2 has a satellite virus P4 that forms $T=4$ shells using the P2 coat protein [8, 1].

Another well-studied class of icosahedral viruses are the $T=3$ plant viruses [15, 17, 31]. All of their structures are similar, and several theories for their assembly have been advanced [28, 32]. The theory proposed by Sorger *et al.* [32] can be cast in a local rule framework. Although these virus shells have three non-equivalent positions, the proteins in two of these positions assume quite similar conformations [32, 31]. These are labeled 1 in the graph representation in figure 8, while proteins in the third position are labeled 2. From the rules implied by this representation, both $T=3$ and $T=1$ shells can be constructed; however, as noted by [32], if assembly is initiated by a structure containing a type 2 conformation, these will propagate during assembly to uniquely determine the $T=3$ structure. The coat proteins of many of these viruses in fact can be induced to form $T=1$ shells [28].

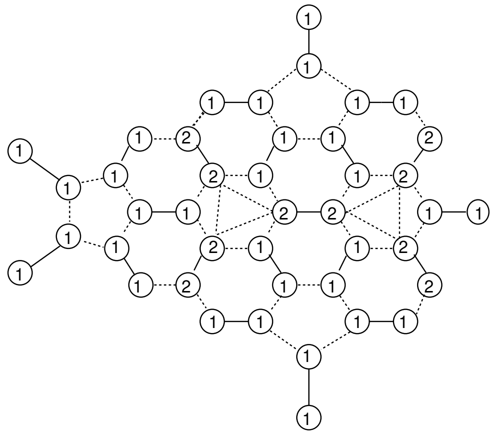

Figure 8. A graph representation for $T=3$ plant viruses. Not all bonds are shown, but the bonds shown are sufficient to abstract a set of local rules that direct assembly of virus shells. The shell proteins are believed to form dimers in solution, which are represented by solid lines in the figure.

In cases of higher T number, an important question is how scaffolding proteins assist coat proteins in following the local rules. For P22, scaffolding proteins are required for proper shell formation, and they assemble at the same rate as the coat-protein subunits. Mutations of the scaffolding proteins often cause malformations of the virus shell. However, scaffolding proteins are not required for this virus to retain its correct shape after the assembly process is completed [19, 20]. One role scaffolding proteins could be playing is enforcing or reinforcing the local rules. For example, if the set of local rules in Figure 7 controlling 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 [19, 10], is that four scaffolding proteins are associated with each hexamer, and five with each pentamer. In this case, the role of the scaffolding proteins could be to break six-fold symmetry in hexamers.

Local Rules for a Non-Quasi-Equivalent Virus. The polyomaviruses are simple DNA animal viruses. Their failure to fit into Caspar and Klug's quasi-equivalent framework has been one of the most debated points in structural virology. Based on recent results obtained from x-ray crystallography, Liddington *et al.* [21] postulated that assembly occurs by forming pentamers and that these building blocks are subsequently tied together. In what follows, we apply the local rule theory to a polyomavirus, simian virus 40, 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 where they first assemble into substructures, which then come together to form the shell.

Simian virus 40 (SV40) is a 360-subunit spherical virus with a shell consisting entirely of pentamers, some of which contact five other pentamers and some of which contact *six* other pentamers (figure 9). This structure still has 5-fold, 3-fold, and 2-fold symmetry and is therefore icosahedral.

SV40 has been considered an anomaly, because the theory of quasi-equivalence assumes that all icosahedral shells have exactly 12 equally-spaced *penta*valent (i.e. have five neighboring capsomeres) pentamers and that all other subunits are packed into hexavalent hexamers. Clearly this assumption does not hold for SV40. Research [21, 26, 22] on the structure of SV40 has focused on how pentamers could be hexavalent and on how the same protein could occupy very asymmetric environments.

Another way to describe this anomaly is in terms of T numbers; SV40 would correspond to a $T = 6$ in that there are six proteins per corner of each triangular face. However, as the theory of quasi-equivalence does not allow 6 as a possible T number, it was classified as an anomalous $T = 7$. The capsomeres are indeed arranged in much the same pattern as for a $T = 7$ virus.

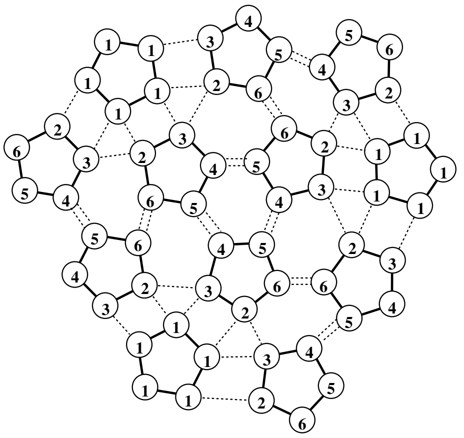

Figure 9. A simplified diagram of how the coat proteins in the SV40 shell connect with each other.

Local rules for SV40 can be constructed that are not substantially different than for other icosahedral viruses. It could simply have six local rules, one for each of its conformations. The local rules in figure 10 describe all the types of protein-protein interactions in SV40.[2]

These simple rules for six conformations guarantee the final form: from a mathematical point of view, applying the rules in figure 10 in any order to an initial subunit will result in the same pattern of interconnectivity as in figure 9. It is true that the rules are somewhat different than those in figure 4, since a protein can have four bonds instead of three.

The function of the C-terminal arms has been described as "tying together" the pentameric building blocks [21]. While the C-terminal arms clearly do span the spaces between pentamers, they may also play the dual role of enforcing the bond interactions of the local rules. The N-terminal arms (which tie together the subunits within a pentamer) may force the subunits to form pentamers through the rotational angles of the bonds.

As has been observed in the crystal structure, the cores of the six conformations of SV40 are essentially the same, but in each conformation the two "arms" are in different places. Thus, the differences in the rules for SV40 are reflected in the bonds, whereas for P22, for example, these differences are reflected in the protein subunit conformations.

[2]Purified polyoma virus subunits will assemble into shells *in vitro* without assistance from other proteins, but these shells vary greatly in size and shape, whereas assembly *in vivo* is much more homogeneous [29, 30]. It thus seems likely that the local rules are not completely determined by the shell protein of polyoma virus but require aid from some other protein. This may apply also to SV40.

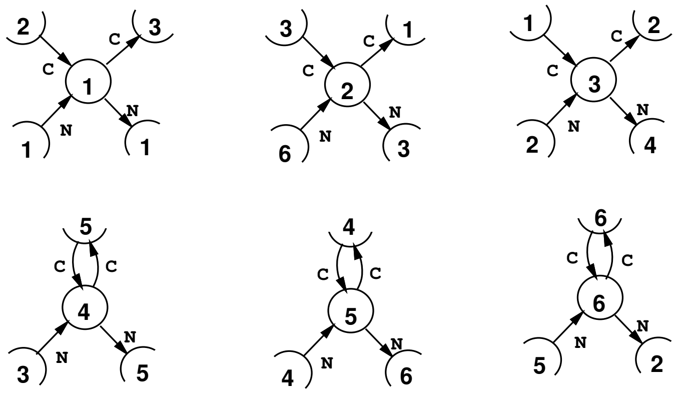

Figure 10. Local rules for the SV40 virus. Each protein subunit is labeled with the type of its conformation. The double, directed edges could be simplified to either a single, directed or undirected, edge; we have drawn them as double edges so as to correspond to the known biological structure [21]. Each inter-pentameric bond interaction is a C–terminal arm of the protein subunit, labeled with a direction to indicate which subunit it is from and a "C" to indicate which arm. Each intra-pentameric bond interaction is an N–terminal arm, labeled also with a direction and "N."

4 Closure and Malformation

Although the above discussion might suggest that closure is easily assured, simulations show that a spiraling malformation can occur if the local rules are "broken" just once. Such incorrectly polymerized spiral structures have been observed with P22 and other viruses [20, 9, 16].

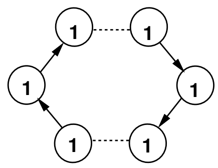

Figure 11. An example of how the local rules can be "broken": the connected components with solid directed edges between them are assumed to form before the dashed edges arise.

Suppose the two distinct connected components in figure 11, drawn as solid lines, were formed independently. When they encounter each other (i.e. the dashed lines are added), the local energy may be minimized by their forming a hexamer, when they really should form a pentamer. This distortion of the rules allows six type 1 subunits instead of five to fit together to form a capsomere at what “should” be a 5-fold axis of symmetry. If the local rules are correctly followed thereafter, this hexamer would next be surrounded with six capsomeres instead of five. Since hexagons tile a plane, a region of an icosahedral surface having a hexamer in place of a pentamer 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 because they will be separated by the length of the planar region (figure 12). One side may curl inward, and the second may form another shell layer around it. Computer experiments show that if local rules are broken in this way, spiraling can indeed occur. This exact malformation may not occur in nature, but other mistakes in formation could result in large planar sections of a shell which would likewise spiral.

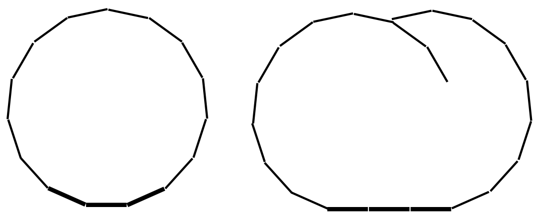

Figure 12. A cross-sectional diagram of spiraling. On the left, a spherical shape is constructed from segments with regularly-spaced curvature. On the right, a region without curvature is created at the bottom of the sphere, but subsequent growth retains its regularly-spaced curvature. The resulting structure does not achieve closure.

Another common malformation of virus shells is the formation of open tubes. These can be viewed as planar sheets of hexamers which have been curved until two of their edges meet. This kind of malformation has been observed in many viruses, including SV40 [2] and bacteriophages λ and T4 [16]. The formation of these tubes may require some bonds between two conformations which are not allowed in the local rules. Possibly these bonds have higher energy than the bonds in closed shells, but once a tube has begun to form, the lowest energy additions to this structure are a continuation of the tube. The structure of these tubes might further illuminate the rules for assembly.

5 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 T_m of 87°C [14]. 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, J. King, 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 are essentially templates for energetically favorable arrangements. An individual protein can adhere to a slightly asymmetric location, and then the surrounding structure would readjust to find a low energy arrangement. Implementing these flexible templates was the main difficulty in doing a computer simulation of the assembly process.

Previous attempts at interfering with the infection process have mainly focused on interrupting infection by a fully-formed shell at the binding site. There have also been attempts at preventing the scaffolding proteins from exiting the interior of a fully-formed shell so that there is no room for the genomic information of a virus. The local rules tell us that if we can interfere with a single bond interaction, the shells may not close.

The local rule theory suggests the way shell assembly is initiated, maintained, and terminated. It elucidates a mechanism for the spiraling malformation of some viruses and is a promising method for explaining other malformations. We expect local rule methods to be useful for exploring the structure and assembly of virus shells, malformed viruses, and perhaps even non-icosahedral viruses.

Acknowledgements

Dan Kleitman has been generous with his time, ideas, and support. Discussions with Peter Prevelige have been most informative. Doug Muir's programming assistance has been a great help. B. B. was supported in part by the Radcliffe Bunting Institute through an ONR grant and by MIT grants AF #78181 and DARPA #N00014-92-J-1799; L. T.-K. by an NSF Predoctoral Fellowship and AF grant #78181; and J. K. by NIH GM 17,980.

References

[1] M. Agarwal, M. Arthur, R. D. Arbeit, & R. Goldstein (1990) Regulation of icosahedral virion capsid size by the *in vivo* activity of a cloned gene product, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2428–2432.

[2] T. S. Baker, L. D. Caspar, & W. T. Murakami (1983) Polyoma virus "hexamer" tubes consist of paired pentamers, *Nature* **303**, 446–448.

[3] C. Branden & J. Tooze (1991) *Introduction to Protein Structure* (Garland Publishing, New York) chapter 11.

[4] S. Casjens (1985) An introduction to virus structure and assembly, in *Virus Structure and Assembly*, ed. S. Casjens (Jones and Bartlett, Boston, MA) pp. 1–28.

[5] S. Casjens & J. King (1975) Virus assembly, *Annu. Rev. Biochem.* **44**, 555–611.

[6] D. L. D. Caspar (1980) Movement and self-control in protein assemblies: Quasi-equivalence revisited, *Biophys. J.* **232**, 103–138.

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

[8] T. Dokland, B. H. Lindqvist, & S. D. Fuller (1992) Image reconstruction from cryo-electron micrographs reveals the morphopoietic mechanism in the P2–P4 bacteriophage system, *EMBO J.* **11**, 839–846.

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

[10] K. Eppler, E. Wyckoff, J. Goates, R. Parr, & S. Casjens (1991) Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging, *Virology* **183**, 519–538.

[11] B. N. Fields, ed. (1985) *Virology* (Raven Press, New York).

[12] H. Fraenkel-Conrat & R. C. Williams (1955) Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components, *Proc. Natl. Acad. Sci. U.S.A.* **41**, 690–698.

[13] M. T. Fuller & J. King (1982) Assembly *in vitro* of bacteriophage P22 procapsids from purified coat and scaffolding subunits, *J. Mol. Biol.* **156**, 633–665.

[14] M. Galisteo & J. King (1991) Conformational transformations in the protein lattice of phage P22 procapsids, *Annu. Rev. Biochem.* **60**, 321–348.

[15] S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler, & G. Bricogne (1978) Tomato bushy stunt virus at 2.9 Å resolution, *Nature* **276**, 368–373.

[16] R. W. Hendrix (1985) Shape determination in virus assembly: The bacteriophage example, in *Virus Structure and Assembly*, ed. S. Casjens (Jones and Bartlett, Boston, MA) pp. 169–204.

[17] J. M. Hogle, A. Maeda, & S. C. Harrison (1986) Structure and assembly of turnip crinkle virus I: X-ray crystallographic structure analysis at 3.2 Å resolution, *J. Mol. Biol.* **191**, 625–638.

[18] I. Katsura (1983) Structure and inherent properties of the bacteriophage lambda head shell IV: Small-head mutants, *J. Mol. Biol.* **171**, 297–317.

[19] J. King & S. Casjens (1974) Catalytic head assembling protein in virus morphogenesis, *Nature* **251**, 112–119.

[20] J. King, R. Griffin-Shea, & M. T. Fuller (1980) Scaffolding proteins and the genetic control of virus shell assembly, *Quart. Rev. Biol.* **55**, 369–393.

[21] R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, & S. C. Harrison (1991) Structure of simian virus 40 at 3.8-Å resolution, *Nature* **354**, 278–284.

[22] C. T. Patch, A. S. Levine, & A. M. Lewis, Jr. (1979) The adenovirus-SV40 hybrid viruses, in *Comprehensive Virology*, eds. H. Fraenkel-Conrat & R. C. Wagner (Plenum Press) Vol. 13, pp. 459–541.

[23] B. V. V. Prasad, P. E. Prevelige, Jr., E. Marietta, R. O. Chen, D. Thomas, J. King, & W. Chiu (1993) Three-dimensional transformation of capsids associated with genome packing in a bacterial virus, *J. Mol. Biol.* **231**.

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

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

[26] I. Rayment, T. S. Baker, & D. L. D. Caspar (1982) Polyoma virus capsid structure at 22.5 Å resolution, *Nature* **295**, 110–115.

[27] M. G. Rossmann (1984) Constraints on the assembly of spherical virus particles, *Virology* **134**, 1–13.

[28] M. G. Rossmann & J. W. Erickson (1985) Structure and assembly of icosahedral shells, in *Virus Structure and Assembly*, ed. S. Casjens (Jones and Bartlett, Boston, MA) pp. 29–74.

[29] D. M. Salunke, D. L. D. Caspar, & R. L. Garcea (1986) Self-assembly of purified polyomavirus capsid protein VP_1, *Cell* **46**, 895–904.

[30] D. M. Salunke, D. L. D. Caspar, & R. L. Garcea (1989) Polymorphism in the assembly of polyomavirus capsid protein VP_1, *Biophys. J.* **56**, 887–900.

[31] A. M. Silva & M. G. Rossmann (1987) Refines structure of southern bean mosaic virus at 2.9 Å resolution, *J. Mol. Biol.* **197**, 69–87.

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