ELSEVIER

Discrete Applied Mathematics 104 (2000) 97–111

DISCRETE APPLIED MATHEMATICS



# Local rule mechanism for selecting icosahedral shell geometry

B. Berger[a,*,1], J. King[b,2], R. Schwartz[c,3], P.W. Shor[d]

[a] *Department of Mathematics and Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[b] *Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[c] *Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[d] *AT&T Labs, Florham Park NJ 07932, USA*

**Abstract**

Local rules are given for various virus capsid, or protein shell, geometries. Local rules specify local interaction patterns that can direct the self-assembly of coat and scaffolding subunits into different capsid geometries. They provide a mathematical abstraction allowing combinatorial analysis of possible assembly pathways. Local rules and simulations are provided for a geometry called "$T = 7$" such that a small change in the rules produces an alternate "$T = 4$" geometry. This is intriguing since evidence has been accumulating that some $T = 7$ capsids are closely related to $T = 4$ capsids. 

*Keywords*: Self-assembly; Virus capsid; Local rules; Conformational switching; Shell geometry

## 1. Introduction

Virus capsids are constructed of repeated protein subunits, or coat proteins, which surround their condensed DNA or RNA genomes. Many of these capsids are believed to assemble with only limited aid from cellular machinery; they "self-assemble", or spontaneously polymerize and take shape, in the host cell environment. Sometimes, assembly is assisted by scaffolding proteins, which assemble with the coat proteins to form a precursor capsid, but are removed before the capsid matures. At first glance, the

* Corresponding author.

[1] Bonnie Berger was supported in part by NSF Career Award 95019997-CCR, DOE grant DEFG0295ER25253 and NSF Computational Biology grant DBI9711234.

[2] Jonathan King was supported in part by NIH Gm 17, 980

[3] This material is based upon work supported under a National Science Foundation 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.

0166-218X/00/$ - see front matter 
PII: S0166-218X(00)00187-6

assembly of the capsids 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. An important subclass of viruses are the icosahedral viruses, whose coats exhibit icosahedral symmetry patterns. Icosahedral viruses occur in a variety of geometries, which are designated by "$T$ numbers", where a $T=n$ virus typically has $60n$ coat proteins [6]. The assembly pathways of icosahedral viruses have been particularly difficult to explain because very often the same coat protein occurs in them in non-symmetric positions [19].

In Berger et al. [3], a hypothesis for icosahedral viral capsid assembly mechanisms was introduced. This hypothesis, the local rules theory of virus capsid assembly, proposes that the assembly of a virus capsid is directed by the local interactions of its coat and scaffolding protein subunits. The hypothesis requires that these subunits can assume different conformations. During assembly, the conformation of a protein subunit in the capsid dictates the conformation of its neighboring subunits when they join the capsid. It was shown that these local interactions could be designed so as to generate capsids having geometric structures observed in actual virus capsids. Different sets of local interactions, or local rules, can be designed to produce different final capsids, as well as malformed capsids. Varying the angles and interaction lengths associated with a set of local rules by small amounts can change the shape of the final capsid while keeping the same basic geometric structure. The nature of the rules is such that by changing which conformations bind to each other, one can change the geometric structure of the capsid, and thus produce capsids of various sizes. The local rules model has previously been applied to exploring capsid size selection and examining potential sources of capsid malformations [3,21]. In addition, the local rules model formed the basis of a molecular dynamics-like simulation model developed for exploring the self-assembly kinetics of small populations of capsids [20].

For some time, evidence has been accumulating that $T=7$ and 4 capsids are in some cases closely related. Indications that this is true arise in experiments with bacteriophage P2 and its satellite phage P4 [16]. The capsid of bacteriophage P2 is a $T=7$ icosahedral structure formed from a phage protein gpN. When the bacteria is coinfected with P4, P4 capsids are also formed which have $T=4$ icosahedral structure but also use the P2 coat protein gpN. The P4 protein that influences size was determined to be gpSid [22,1]. Since gpSid is present in immature P4 particles, but not in the mature virus, it is believed to act as a scaffolding protein [2,8]. Similarly, a P2 protein gpO is believed to act as a scaffolding protein [4,5,13]. Marvik et al. [16] showed that the capsid size is indeed determined by these auxiliary proteins gpO and gpSid. When P2 gene N was cloned and inserted into bacteria without these auxiliary proteins, gpN produced predominantly irregularly shaped malformations, but also a small number of spherical capsids of both $T=4$ and 7 structure. When the P2 protein gpO was also introduced, gpN formed predominately $T=7$ structures. When a P4 bacteriophage infection was present, but gpO was absent, the protein gpN formed predominately $T=4$ capsids. Finally, with both gpO and P4 infection present, both size capsids were formed, but with many fewer malformations than with gpN alone.

Another indication that $T = 4$ and 7 capsids are related is that in bacteriophage $\lambda$, which also has a $T = 7$ capsid, mutations of the coat protein exist that form functional $T = 4$ capsids [11]. A third indication of this relationship is that one of the common mistakes observed in the assembly of the phage P22 in the absence of scaffolding proteins is the formation of a $T=4$ capsid instead of a $T=7$ structure [7]. The structures of hexons and pentons (hexagonal and pentagonal substructures of icosahedral capsids that are also referred to as hexamers and pentamers) have been found to be very similar between the normal $T = 7$ and the abnormal $T = 4$ structures [24]. In addition, the structures resulting from assembly of pure P22 coat protein are quite similar to those resulting from pure P2 coat protein [7,16,17], possibly indicating that these two viruses have similar assembly mechanisms, even though the sequences of these two proteins do not appear very similar [14]. Finally, a mutation in the $T = 4$ Sinbis virus has been found to result in what appear to be $T = 7$ capsids [12].

We have developed local rule theories for these capsid geometries describing the necessary conformations for each geometry and the immediate neighbor conformations of each conformation in the shell. If these icosahedral capsids indeed form by following a set of local rules, then this set of local rules should have the property that a relatively small perturbation of the rules directing a $T = 7$ capsid should produce a $T = 4$ capsid. In this paper, we give such a set of rules. It would be nice if this set of rules could completely predict the structure of the virus. Unfortunately, these rules are at too high a level of abstraction to predict exactly where the scaffolding molecules should be in the structure. However, these rules do imply that the coat and scaffolding should have certain properties which have been observed in several $T = 7$ viruses as well as the P2 and P4 viruses. The local rules given here also account for the elongation of hexamers observed in phage P22 procapsids [17]. We demonstrate the feasibility of these rules through computer simulations. In addition, we describe applications of the principles behind the rules to other geometries.

## 2. The local rules model

### 2.1. Basic model and assumptions

The *theory of quasi-equivalence* [6] observed that chemically identical protein subunits may not be entirely functionally equivalent in a virus capsid due to differences in the global symmetry, but it was still assumed that they really have almost equivalent conformations and nearly identical chemical properties. Since then, evidence has been accumulating that the subunit conformations can be much more flexible than the theory of quasi-equivalence suggested [9]. A local rule theory postulates that the coat and scaffolding proteins can assume a number of distinct conformations, and thus it is not bound by the notion that these proteins have identical properties.

A local rule theory [3] is a set of local interactions specifying which conformations are allowed to bind to each other, as well as the admissible interaction lengths and relative binding and torsional angles between these interactions. For every $T$ number, it

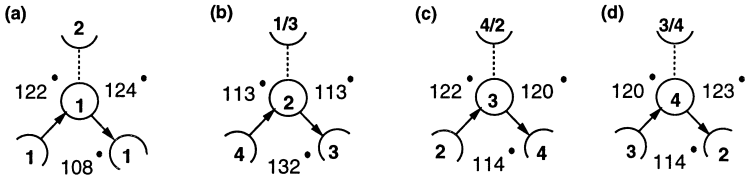


Fig. 1. Possible local rules for a left-handed $T = 7$ virus. Each protein subunit is represented as a circle or part of a circle labeled with its conformation(s). The notation "i/j" means that the subunit can take on conformation i or j. Binding interactions are represented possibly with an associated direction, indicating an interaction that is asymmetric. Angles between binding interactions are the approximate numbers of degrees between the centers of the protein subunits. Angles are not based on any particular virus, but are derived from a computer simulation.

is possible to find a set of local rules producing that $T$ number and using a number of protein conformations that is equal to the $T$ number [3]. It is sometimes also possible to find a different set of local rules that results in the same capsid geometry, but requires fewer conformations, by using rules that assign the same conformation to non-equivalent positions.

The conformations used for the assembly rules need not be the same as the conformations in the final capsid. The set of rules given below for a $T = 7$ capsid assumes only four distinct conformations of the subunits. The local rules require that these four subunit conformations behave differently during the assembly process; however, it does not follow that there are four conformations in the final capsid. There could be more, since there are seven non-equivalent positions in the final capsid, and the resulting non-equivalent neighborhoods could induce two subunits that behaved similarly during assembly to assume different final conformations. There could also be fewer, since the distinctness of conformations during assembly need not be maintained in the final capsid. In phage P22, for example, the conformations of coat proteins in the final capsid, after it has undergone a maturation stage subsequent to the addition of all subunits, appear more similar than the conformations in the pre-maturation procapsid [17].

We now describe the set of local rules for $T = 7$ geometries with four conformations. One of these four conformations will be considered first. Fig. 1(a) gives the rule for how one of the four conformations, the type 1 conformation, chemically binds in three dimensions. A type 1 conformation has a binding site for a type 2 conformation, and two binding sites for type 1 conformations. Given the binding interaction to the type 2 neighbor, then, at a position clockwise from this at an angle of about $124^\circ$, only a type 1 conformation can attach. Similarly, only a type 1 conformation can attach at an angle of about $108^\circ$ from this latter binding interaction. We call this representation the type 1 local rule. Note that although type 1 and type 2 conformations are represented by circles (or parts of circles), their labels (i.e., 1 and 2) are intended to imply asymmetric conformations. Note also that the angles for the rule do not have to add up to $360^\circ$ because the rule is three dimensional.

Similar local rules can be constructed for all the four conformations in a $T = 7$ capsid (Fig. 1). Certain conformations have ambiguous edges, labeled with two possible

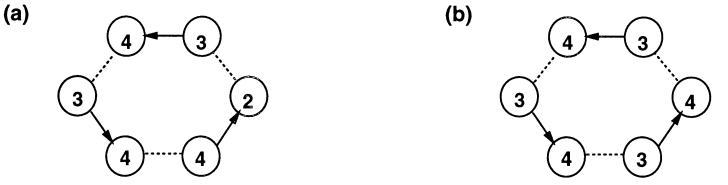


Fig. 2. (a) This configuration is allowed. (b) This configuration is not allowed.

neighbors each, because for this set of local rules one conformation may need to be able to bind to two alternate conformations in certain directions. The binding interactions in the local rules need to be present in the capsid; however, additional interactions may also be present which would have only a secondary effect on the assembly process.

There are various additional constraints imposed upon this set of rules. They assume that the capsid is initiated at a pentamer. They also assume that a protein does not assume its final configuration until there is at least one other protein in the same capsomere (penton or hexon). That is, when a protein attaches itself to the growing capsid at a point when it does not have an adjacent subunit in the same capsomere, it remains in a state of conformational flexibility and does not yet adopt a final conformation. When another subunit attaches next to it, thus creating two points of attachment for the new capsomere, the two proteins will fall into a low-energy configuration and stop fluctuating.

One additional constraint needs to be imposed in addition to the rules already given. During the application of these rules, there is one point during the growth of the capsid where there are two possible configurations which are both allowed by the rules given. In order to consistently build $T = 7$ capsids, it is necessary to disallow one of them and thus consistently choose the other configuration. In this section, we will not go into the possible biochemical mechanisms for choosing one of these two configurations. However, we will show that if the configuration in Fig. 2(a) is consistently chosen, $T = 7$ capsids will result, and if the one in Fig. 2(b) is consistently chosen, $T = 4$ capsids will result. Consequently, the Fig. 2(b) configuration is called the *disallowed configuration* for $T = 7$ capsids.

We now give a step-by-step explanation of the construction of these capsids. Suppose the rules in Fig. 1 have been applied to generate the portion of the capsid in Fig. 3(a). Then by applying the second option for each of the ambiguous edges of the rules in Figs. 1(c) and (d) to the capsid, the portion of the capsid in Fig. 3(b) results. Notice that the capsid now includes the configuration given in Fig. 2(a). Alternatively, by applying the first choice for each of the ambiguous edges of the rules in Figs. 1(c) and (d) to the capsid in Fig. 3(a), the portion of the capsid in Fig. 3(c) results. However, here the capsid includes the disallowed configuration for $T = 7$ from Fig. 2(b). This occurs at what is called a "quasi-three-fold" axis of symmetry. By consistently choosing between the two configurations in Fig. 2 in this manner, it is possible to generate either a $T = 7$ capsid (Fig. 4(a)) or a $T = 4$ capsid (Fig. 4(b)).

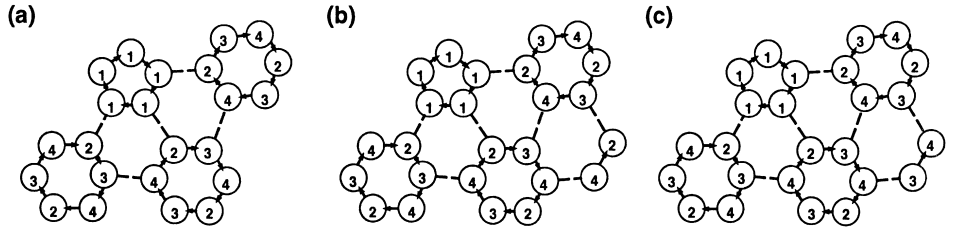


Fig. 3. (a) Partial growth of a capsid according to the rules in Fig. 1. (b) Application of the rules in Fig. 1 to the partial capsid in (a). (c) Another application of the rules in Fig. 1 to the partial capsid in (a) yields the disallowed configuration for $T = 7$ capsids.

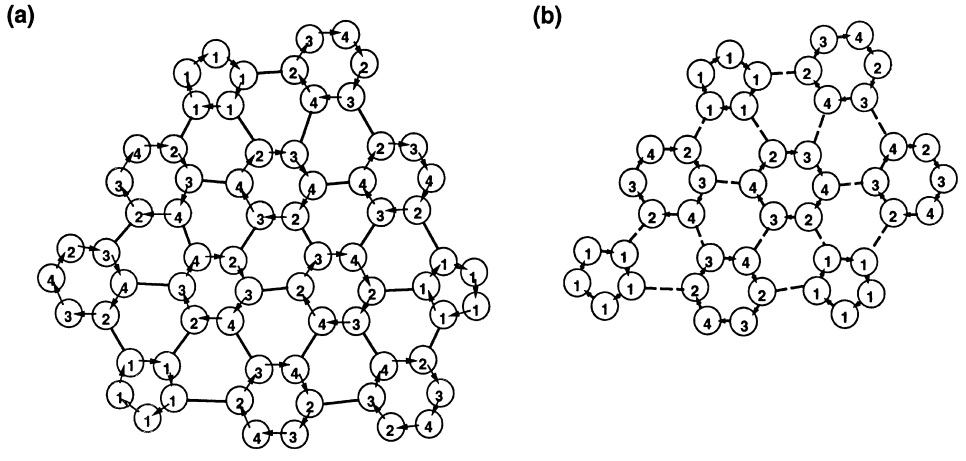


Fig. 4. (a) Portion of a $T = 7$ capsid produced by the rules in Fig. 1. (b) Portion of a $T = 4$ capsid produced by the rules in Fig. 1 when the disallowed configuration is consistently allowed.

Some care must be taken in the order of adding coat proteins to the virus capsid. If, for example, the proteins were added in an order to form a long chain with no interconnections, it would not produce the desired icosahedral capsid, even though this scenario is allowed by the rules. For the set of $T = 7$ rules, the following guidelines are adequate to consistently produce the correct final structure. The guidelines are:

1. Whenever a new capsomere (i.e., hexamer or pentamer) is started, it should be finished before new capsomeres adjacent to it are started. Note that the conformations of all the proteins in a capsomere are determined by the conformations of the first proteins added to this capsomere.
2. If possible, when starting a new capsomere, add it to a point on the capsid where it can attach to two finished capsomeres. This clearly is not possible for the first or second capsomeres added to the capsid, but is possible thereafter.

By adjusting the kinetics, it should be possible to make the coat and scaffolding proteins for the most part obey these guidelines. That is, reactions that add proteins to an existing capsomere should be favored over ones starting a new capsomere, and new

capsomeres should preferentially be started at spots adjoining two existing capsomeres. The requirement that proteins do not assume their final conformation until there are two proteins present in the same capsomere matches well with these requirements for kinetics.

For the set of rules in Fig. 1, orders of protein additions with a weaker set of guidelines will also produce the final structure. In particular, as long as proteins are only added onto a protein that is not in conformational flux, then $T = 7$ capsids are formed.

The set of local rules for $T = 7$ in Fig. 1 is nearly the same as a set of local rules for $T = 4$. In fact, if we allow only the first option for each of the ambiguous edges of the $T = 7$ rules, they become identically $T = 4$ rules. This is equivalent to consistently choosing the disallowed configuration from Fig. 2 (which is what prevents this set of rules from forming $T = 4$ capsids). This establishes that a small perturbation of the rules can switch between $T = 4$ and 7 geometries, as we required in the introduction.

### 2.2. Possible roles for the scaffolding proteins

One of the key properties of the rules given in Fig. 1 is that the conformations of the proteins in the hexamers have $180^\circ$ symmetry. This reduces the number of coat protein conformations necessary to implement local rules for a $T = 7$ capsid from seven to four by putting proteins on the opposite sides of a hexamer into the same conformation. Because of the symmetry, the conformation of any one protein in the hexamer prescribes the conformations for the remaining proteins in the hexamer; this enables the capsid to assemble without introducing too much ambiguity in the process. One possible role for the scaffolding protein is to provide this two-fold symmetry of hexamers; that is, to break a natural six-fold symmetry associated with coat protein hexamers.

Scaffolding molecules might break symmetry by enforcing the choice of configuration in Fig. 2. The experimental evidence on conformational switching in the P2/P4 virus system indicates that the scaffolding proteins in these viruses have this duty [16]. In the case of P2, the scaffolding must reach across the quasi three-fold axes to interact in some way with scaffolding or coat proteins around them and force the configuration in Fig. 2(a). In the case of P4, this would correspond to the scaffolding reaching across the three-fold axes of symmetry.

Evidence in P22 suggests that coat proteins form trimer interactions around the three-fold and quasi-three-fold symmetry points [23], providing a possible mechanism by which scaffolding proteins could enforce or break the three-fold symmetry. Fig. 5 shows the $T = 7$ and 4 lattices with these trimer interactions added. We can note that the patterns of conformations around the trimers differ between the two lattices. For instance, the $T = 7$ shell contains a 4-4-2 trimer around the quasi-three-fold symmetry point, while the $T = 4$ has a 4-4-4 trimer around its corresponding three-fold symmetry point. Scaffolding interactions at these points could potentially provide the additional information needed to force a symmetric 4-4-4 trimer given one type of scaffolding or

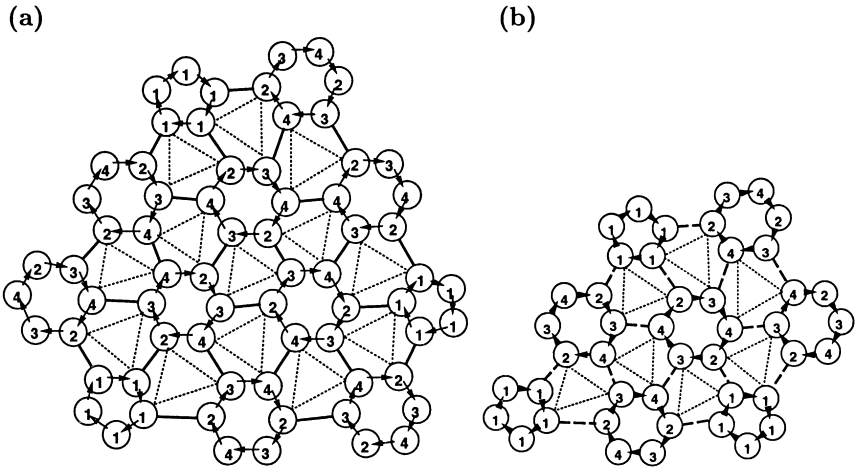


Fig. 5. (a) Portion of a $T=7$ capsid extended with trimer interactions. (b) Portion of a $T=4$ capsid extended with trimer interactions.

to block it given another type of scaffolding. A lack of scaffolding protein would then leave trimers unconstrained, leading to shells of either type as well as shells that mix the two kinds of trimers, providing a possible rationale for the observed mixture of $T=7$, 4, and malformed shells when P22 is grown without scaffolding [7,24].

However, scaffolding protein is not actually necessary to implement this set of rules. The chemical properties of the coat protein itself might force it into hexamers with two-fold but not six-fold symmetry. Since the three type 4 conformations in the disallowed configuration could easily be spatially adjacent, they may independently form a trimer that has higher energy than the 4-4-2 conformation trimer and thus is energetically unfavorable. Alternatively, the disallowed configuration might be impossible because the sum of the bonding angles around it is too large or too small to allow it to close, although this scenario seems less likely.

### 2.3. Switching mechanisms for other T numbers

Similar switching mechanisms seem to exist for other $T$ numbers. There is a set of local rules with seven conformations relating $T=13$ to 7 capsids in much the same way that the $T=7$ set of rules presented in this paper is related to $T=4$ capsids (Fig. 6). These rules have the same additional constraints as those imposed on the rules for the $T=7$ capsid (Fig. 1), except the disallowed configuration which occurs at a quasi-three-fold axis of symmetry is different. Consistently choosing the disallowed configuration will cause the first option on each ambiguous edge to be chosen every time, restricting the rules chosen to precisely those rules for a $T=7$ capsid with seven conformations [3].

This suggests that it may be useful to look for biological evidence for a relation between $T=13$ and 7 capsids, although to date we have no such evidence. We can

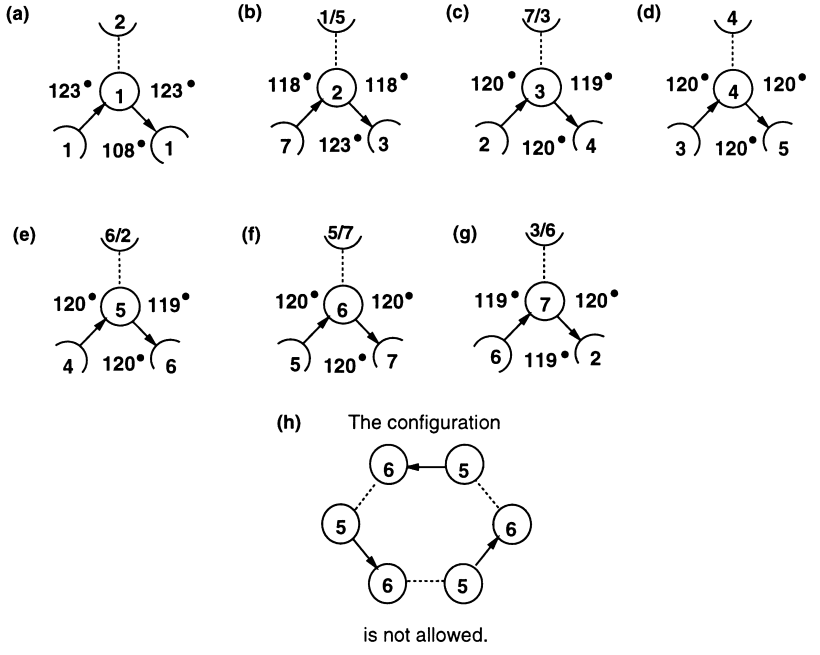


Fig. 6. Possible local rules for a left-handed $T = 13$ virus. The disallowed configuration is given with the rules.

also propose looking for similar mathematical relationships between other pairs of $T$-numbers. For example, we are currently examining rule sets relating $T = 9$ and $T = 3$ or $T = 4$ geometries.

## 3. Simulation methods

In order to demonstrate the feasibility of the $T = 7$ rule set discussed above, we have conducted computer simulations using two versions of the local rules simulator described in Berger et al. [3]. The simulator constructs abstract models of capsids from user-defined rule sets, attaching proteins one at a time according to the specified local rules. Proteins themselves are represented by simple spherical masses capable of binding to each other in accordance with the local rules' binding patterns. Bound proteins exert forces on each other according to a mass–spring model. The energy of the structure induced by the springs is reduced to a minimum after each subunit addition.

Rules are specified through files describing the binding patterns of each conformation of the protein. A rule file specifies, for each conformation, the conformations of its neighbors in the shell and their relative positions, represented as distances and angular offsets. In addition, binding interactions have assigned types, which specify which pairs of binding sites are compatible when a particular conformation attaches to two or

more neighbors that have the same conformation. For example, in the $T = 7$ rules of Fig. 1, rule 1(a) requires a protein in conformation 1 to connect to two other proteins in conformation 1. The different types of the edges, denoted by arrows of different directions, allow the simulator to determine which of a conformation 1 node's edges should connect to which edges of its conformation 1 neighbors.

The basic simulator was modified for the current application to allow a conformation to have multiple rules. This makes it possible to create ambiguities in determining the conformations of some neighbors of a subunit, as required by the $T = 7$ rules given above. The simulator was further modified to add a new subunit only when the conformation of the subunit can be determined unambiguously from the attachment site. This reflects the constraint that subunits must settle to a unique state before further growth can occur off of them. This places some limitations on the order of assembly.

The mass–spring model used to compute forces between neighboring particles treats each protein as a spherical unit mass with binding interactions represented by three springs. These springs create translational, bending, and rotational forces. The first spring creates a translation force on each of the two nodes connected to it given by

$$\boldsymbol{F} = \boldsymbol{d} - l_0 \frac{\boldsymbol{d}}{||\boldsymbol{d}||},$$

where $\boldsymbol{d}$ is the vector distance between the two nodes and $l_0$ is length of the spring in its relaxed state. This force also has a torque component, given by

$$\boldsymbol{T}_{\mathrm{f}} = \boldsymbol{V} \times \boldsymbol{F},$$

where $\boldsymbol{F}$ is the force from above, and $\boldsymbol{V}$ is the ideal edge vector from the node to its neighbor. A second spring creates a bending torque, which straightens edges that are connected on non-ideal angles. This torque is given by

$$\boldsymbol{T}_{\mathrm{b}} = \boldsymbol{V} \times \boldsymbol{V}',$$

where $\boldsymbol{V}$ is the same as in the previous spring, and $\boldsymbol{V}'$ is the ideal edge vector of the nodes neighbor. Finally, there is a rotational torque that corrects for rotations around a bond. This torque is given by

$$\boldsymbol{T}_{\mathrm{r}} = \boldsymbol{O} \times \boldsymbol{O}',$$

where $\boldsymbol{O}$ and $\boldsymbol{O}'$ are orientation vectors perpendicular to the ideal edge vectors which are aligned with each other in an ideal bond. Following each subunit addition, the forces on the springs are relaxed to a minimum energy state, reducing the stresses between bonds in the shell.

Simulations were run on Silicon Graphics Indigo 2 workstations. All screen snapshots in this paper were taken from these workstations. The different versions of the simulator used in the present work were written in C, with graphics produced by the IrisGL libraries.

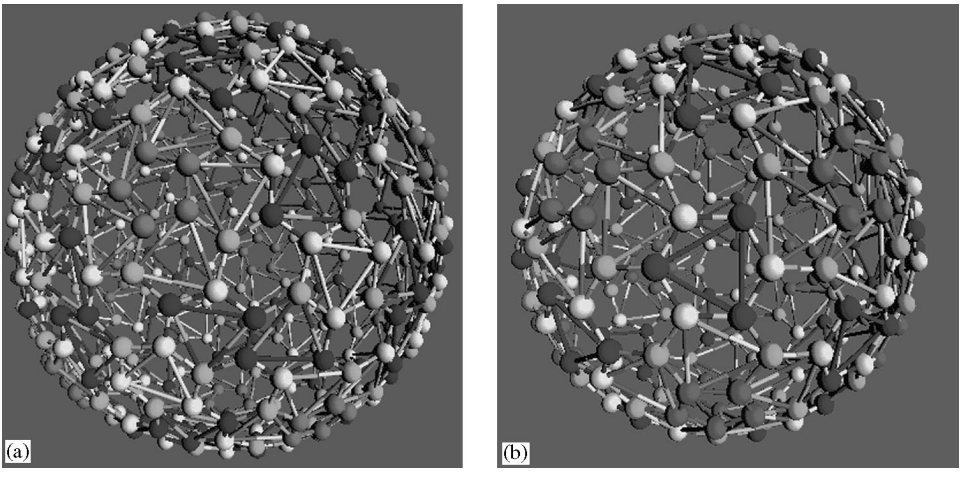


Fig. 7. (a) Screen snapshot of a $T = 7$ capsid produced by computer simulation of the proposed local rules. (b) Screen snapshot of a $T = 4$ capsid produced by computer simulation of the proposed local rules. The rules for coat proteins are identical between the two capsids except for restrictions on allowed trimer interactions. The capsids in parts (a) and (b) are not shown at the same scale.

## 4. Simulation results

We have applied the simulator described above to the generation of $T = 4$ and 7 capsids according to the proposed local rules. The four conformation rules for $T = 7$ capsids, shown in Fig. 1, were coded for use by the local rules computer simulator described in the previous section. The basic rules were extended with additional trimer constraints to impose the switching mechanism. These trimer constraints are modeled by two additional edges per node, corresponding to the additional edges shown in Fig. 5. Nodes were allowed to participate in only those trimers in which a node of that conformation can participate in the correct final shell. For example, a conformation 2 node could participate only in 1-2-3 and 4-4-2 trimers in a $T = 7$ shell, and only in 1-2-3 trimers in a $T = 4$ shell. It was possible to switch between the formation of $T = 4$ and 7 geometries by altering only the allowed trimers.

Fig. 7 shows two capsids generated by the simulator. The trimers used to provide the switching mechanism are visible in both shells in addition to the normal edges. Fig. 7(a) shows a $T = 7$ capsid. The use of two different possible trimers for each node of conformations 2, 3, and 4 breaks the two-fold symmetry around the the hexamers, creating a $T = 7$ geometry. It should be noted that this simulation model does not specify the exact positions of the scaffolding proteins in this $T = 7$ shell, only their action in enforcing possible trimers. This model would be consistent with any scaffolding positions capable of restricting trimers to the allowed configurations. Altering the scaffolding rules to produce a single possible trimer per conformation such that the disallowed configuration of Fig. 2(b) is enforced, rather than forbidden, results in

a $T = 4$ geometry, as shown in Fig. 7(b). This demonstrates how switching can be accomplished by using trimer configurations to disambiguate the rule set of Fig. 1 in two different ways.

## 5. Discussion

Experimentally, the P2 coat protein assembles into fewer malformations in the presence of both scaffolding proteins than it does in the presence of either alone [16]. This seems counter-intuitive, because it might be the case that these scaffolding molecules interfere with each other by both bonding to the same capsid. Since one type makes $T = 4$ capsids and the other $T = 7$ capsids, a capsid containing both types of scaffolding particles would seem very likely to be malformed. We previously predicted that this might be produced by having each type of scaffold protein form a connected inner scaffold, such that the two scaffolds compete with one another with the first one to attach blocking the other and selecting the geometry. However, Marvik et al. [15] have since determined that the gpSid molecules in P4 form a connected *outer* scaffold. This gives a better explanation of why there are fewer malformations in the presence of both scaffolding proteins than in the presence of either alone: in P4, both scaffolding proteins are simultaneously present during assembly, so the assembly is being stabilized by both scaffolds cooperatively (one internal and one external), which apparently works better together than either does alone. The hypothesis that scaffolding acts specifically to enforce trimer constraints may also provide a mechanism for this effect. Since P2 and P4 share one kind of trimer (the 1-2-3 trimer) it could be that P4 relies on P2's scaffolding protein gpO to enforce that trimer, rather than redundantly enforcing it itself, and that gpSid acts only on those trimers that differ between the two geometries. The rules would then be under-determined when enforced by gpSid alone, leading to many malformed structures. Furthermore, the gpSid molecules do indeed appear to be connected to the capsid at the three-fold axes of symmetry (as well as other points), which is where our theory predicts they should operate to direct the assembly of a $T = 4$ rather than a $T = 7$ capsid.

Recall that we needed the initiation complex to be a pentamer. This assumption seems realistic since for several bacteriophages, an initiation complex located at one of the pentamers is required to guarantee formation of the desired structure in vivo [25]. Further evidence in support of this assumption appears in Prevelige et al. [18], which gives evidence that in vitro initiation of P22 capsids occurs at a pentamer.

Conformational flexibility is required for the sets of rules given here, since proteins do not assume their final conformation until after having first bound and also waited for a neighboring protein to attach. However, this is quite reasonable for real viral coat proteins (at least for P22) since according to Galisteo et al.'s [9] results, these proteins are quite flexible in solution. The local rules theory also need not be restricted to assembly by conformationally flexible monomers. Whereas certain bacteriophages such as P22 and $\lambda$ are known to form from monomers, similar local rules can be given

that build capsids from the association of dimer and capsomere building blocks. These rules also make use of 2-fold symmetry in the hexamers to direct assembly.

The requirement in the rules for two-fold rotational symmetry of hexamers is also intriguing because the micrographs of P22 show near-symmetry of the hexamers under $180^\circ$ rotations [17]. This is seen in the elongation of the hexamers of P22 precursor virus along the axis connecting the two type 3 conformations.

For P22, the differences between the protein conformations in non-equivalent positions are noticeably less in the mature form than in the precursor form [17]. 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 capsid together stably. This may also be the reason for the substantial changes that other bacteriophages undergo between their precursor and mature forms [10].

Whereas in P4 and P2, scaffolding proteins appear responsible for the switch between $T = 4$ or 7 capsids, in $\lambda$, several mutations of the coat protein produce $T = 4$ capsids [11]. This may indicate that the role of scaffolding in the local rules differs in these viruses. Whereas in P22 and P2 the choice of hexamer as in Fig. 2 is enforced by the scaffolding, in $\lambda$ it may be enforced through interactions between coat proteins. Alternatively, the mutation of $\lambda$ producing $T = 4$ capsids may be in the section of the coat protein responsible for interacting with the scaffolding, in which case the scaffolding molecule could still be involved in directing the choice of $T = 7$ rather than $T = 4$ capsids.

The proposed set of local rules would also explain the fact that most $T = 7$ bacteriophages only have a single type of protein for pentameric and hexameric vertices (although some, such as T4, use separate proteins for the two kinds of vertices). If there were a separate coat protein for the pentameric vertices, this protein would have to be able to distinguish a potentially pentameric site from a hexameric site when joining to the forming capsid. For the local rule theory given in this paper, this distinction sometimes does not occur until another coat protein joins adjacent to the first. This would indicate that for this set of local rules, the same protein needs to be able to take on both hexameric and pentameric conformations.

The relationship between the two sets of rules for $T = 7$ and 4 may also indicate a pathway for the evolution of the assembly mechanism of a $T = 7$ virus. Since these two sets of rules are relatively similar, a $T = 4$ virus capsid protein with rules involving four conformations may have evolved to produce a larger $T = 7$ capsid by a mutation of the coat protein or the development of a scaffold protein to break some of the symmetries of the $T = 4$ geometry. The relationship between $T = 7$ and 13 capsids may similarly have led to the evolution of $T = 13$ bacteriophages.

Local rules may also provide a means for analyzing certain kinds of capsid malformations. The coat protein for P2, gpN, can form spiral structures and "telephone"-shaped structures in which two spirals, often having different diameters, are connected; these outcomes are especially common when gpN assembles in the absence of scaffolding protein [15]. The local rules provide a possible explanation for the standard spiral

structure; in particular, if a hexamer occurs at a 5-fold axis of symmetry, spiraling occurs in computer simulations [3]. One may also suggest that once this initial flat region is formed, it is more likely that this flat region gives rise to other such "mistakes", producing an even larger flat region. When pentamers are finally added to the flat region, they introduce curvature. Then if the local rules are correctly followed thereafter, the "telephone"-shaped structure could result. If pentamers are never added to the large flat region, then tubular malformations would be likely to result. Another possible explanation for both types of spirals arises when the constraint that the application of the rules in Fig. 1 must start with a pentamer is violated. Suppose that the capsid started building with a hexamer. Then it would be ambiguous how to build out from this initial hexamer, whereas it is not ambiguous how to build out from a pentamer. Furthermore, it would be ambiguous when to add a pentamer to the growing structure. Then large flat regions could result with the same consequences as those discussed above. Simulations conducted with a molecular dynamics-like local rule model have also shown how small local binding errors can induce capsid malformations resembling both spiral types [20].

Further analysis of the theoretical implications of the local rules may provide additional insights into the biological systems examined. We have derived a mathematical model for understanding some virus capsid assembly systems. We have also argued for the validity of this model by showing how properties derived from the mathematical local rules framework explain observed properties of the biological systems. It is likely that other mathematical properties of the local rules framework will reflect physical properties of the virus assembly systems studied. Combinatorial analysis of possible local rules given geometric constraints may provide insight into why only certain assembly strategies have yet been observed or may predict the existence of yet unobserved virus assembly systems. Locating further mathematical relationships between different capsid geometries, as we have done with the $T = 4/T = 7$ and $T = 7/T = 13$ systems, may be useful in studying evolutionary relationships between different viruses. Understanding constraints on assembly pathways is also an essential step to developing mathematical models of the dynamic assembly process which might be matched to data on capsid assembly kinetics. Finally, further analysis of the information content of local rules may be useful in discovering underlying symmetries and potential simplifications that could reveal still unrecognized mechanisms behind virus capsid assembly.

## Acknowledgements

We are grateful to Doug Muir for developing the basic simulator used in the simulation experiments and to Peter Prevelige for his advice and assistance with this work.

## References

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

[2] K.J. Barrett, M.L. Marsh, R. Calendar, Interactions between a satellite bacteriophage and its helper, J. Mol. Biol. 106 (1976) 683–707.
[3] B. Berger, P.W. Shor, L. Tucker-Kellogg, J. King, Local rule-based theory of virus shell assembly, Proc. Natl. Acad. Sci. USA 91 (16) (1994) 7732–7736.
[4] E. Bertani, E.W. Six, The P2-like phages and their parasite, P4, in: R. Calendar (Ed.), The Bacteriophages, Vol. 2, Plenum Press, New York, 1988, pp. 73–143.
[5] S. Casjens, R.W. Hendrix, Control mechanisms in dsDNA bacteriophage assembly, in: R. Calendar (Ed.), The Bacteriophages, Vol. 1, Plenum Press, New York, 1988, pp. 15–91.
[6] D.L.D. Caspar, A. Klug, Physical principles in the construction of regular viruses, Cold Spring Harbor Symp. Quant. Biol. 27 (1962) 1–24.
[7] W. Earnshaw, J. King, Structure of phage P22 coat protein aggregates formed in the absence of the scaffolding protein, J. Mol. Biol. 126 (1978) 721–747.
[8] C.G. Fossdal, GpSid Functions as a transient, P4 capsid size determining scaffold, Ph.D. Thesis, University of Oslow, Norway, 1991.
[9] M.L. Galisteo, C.L. Gordon, J. King, Stability of wild type and temperature sensitive protein subunits of the bacteriophage P22 capsid, J. Biol. Chem. 270 (1995) 16595–16601.
[10] R.W. Hendrix, Shape Determination in Virus Assembly: The Bacteriophage Example, in: S. Casjens (Ed.), Virus Structure and Assembly, Jones and Bartlett, Boston, MA, 1985, pp. 169–204.
[11] I. Katsura, Structure and inherent properties of the bacteriophage lambda head shell IV: small-head mutants, J. Mol. Biol. 171 (1983) 297–317.
[12] H. Lee, D.T. Brown, Mutations in an exposed domain of Sindbis virus capsid protein result in the production of noninfections virions and morphological variants, Virology 202 (1994) 390–400.
[13] J.A. Lengyel, R.N. Goldstein, M. Marsh, M.G. Sunshine, R. Calendar, Bacteriophage P2 head morphogenesis: cleavage of the major capsid protein, Virology 53 (1973) 1–23.
[14] N.A. Linderoth, R. Ziermann, E. Haggard-Ljunguist, G.E. Christie, R. Calendar, Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2, Nucl. Acids Res. 19 (1991) 7207–7214.
[15] O.J. Marvik, T. Dokland, R.H. Nøkling, E. Jacobsen, T. Larsen, B.H. Lindqvist, The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids, J. Mol. Biol. 251 (1995) 59–75.
[16] O.J. Marvik, P. Sharma, T. Dokland, B.H. Lindqvist, Bacteriophage P2 and P4 assembly: alternate scaffolding proteins regulate capsid size, Virology 200 (1994) 702–714.
[17] B.V.V. Prasad, P.E. Prevelige, Jr., E. Marietta, R.O. Chen, D. Thomas, J. King, W. Chiu, Three-dimensional transformation of capsids associated with genome packing in a bacterial virus, J. Mol. Biol. 231 (1993) 65–74.
[18] P.E. Prevelige, Jr., D. Thomas, J. King, Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells, Biophys. J. 64 (1993) 824–835.
[19] M.G. Rossmann, Constraints on the assembly of spherical virus particles, Virology 134 (1984) 1–13.
[20] R. Schwartz, P.W. Shor, P.E. Prevelige, Jr., B. Berger, Local rules simulation of the kinetics of virus capsid self-assembly, Biophys. J. 75 (1998) 2626–2636.
[21] R. Schwartz, R.L. Garrea, B. Berger, "Local rules" theroy applied to polyomavirus polymorphic capsid assemblies, Virology 268 (2000) 461–470.
[22] D. Shore, G. Deho, J. Tsipis, R.N. Goldstein, Determination of capsid size by satellite bacteriophage P4, Proc. Natl. Acad. Sci. USA 75 (1978) 400–404.
[23] P.A. Thuman-Commike, B. Greene, J. Jakana, B.V.V. Prasad, J. King, P.E. Prevelige Jr., W. Chiu, Three-dimensional structure of scaffolding-containing phage P22 procapsids by electron cryo-microscopy, J. Mol. Biol. 260 (1996) 85–98.
[24] P.A. Thuman-Commike, B. Greene, J.A. Malinski, J. King, W. Chiu, Role of the scaffolding protein in P22 procapsid size determination suggested by $T = 4$ and $T = 7$ procapsid structures, Biophys. J. 74 (1998) 559–568.
[25] J.M. Valpuesta, J.L. Carrascosa, Structure of viral connectors and their function in bacteriophage assembly and DNA packaging, Quart. Rev. Biophys. 27 (1994) 107–155.