

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15

## **Virus assembly, allostery, and antivirals**

Adam Zlotnick<sup>1,2\*</sup> & Suchetana Mukhopadhyay<sup>2\*</sup>

1 – Department of Molecular and Cellular Biochemistry, Indiana University  
Bloomington, IN 47405, USA

2 – Department of Biology, Indiana University, Bloomington, IN 47405, USA

\* Corresponding authors: Zlotnick, A. ([azlotnic@indiana.edu](mailto:azlotnic@indiana.edu)), Mukhopadhyay, S. ([sumukhop@indiana.edu](mailto:sumukhop@indiana.edu))

1 **Assembly of virus capsids and surface proteins must be regulated to ensure that the**  
2 **resulting complex is an infectious virion. Here we examine assembly of virus capsids,**  
3 **focusing on hepatitis B virus and bacteriophage MS2, and formation of glycoproteins in**  
4 **the alphaviruses. These systems are structurally and biochemically well-characterized**  
5 **and are simplest-case paradigms of self-assembly. Published data suggest that capsid**  
6 **and glycoprotein assembly is subject to allosteric regulation, that is, regulation at the**  
7 **level of conformational change. The hypothesis that allostery is a common theme in**  
8 **viruses suggests that deregulation of capsid and glycoprotein assembly by small**  
9 **molecule effectors will be an attractive antiviral strategy, as has been demonstrated with**  
10 **hepatitis B virus.**

## 12 **Is regulation of self-assembly required?**

13 The common structural denominator for a typical small virus is a genome surrounded by a shell  
14 composed of dozens of copies of the capsid protein. Many small viruses also have lipid  
15 envelopes studded with glycoproteins that can facilitate cell entry. Even a small virus is  
16 complex, yet its formation is a canonical example of self-assembly. Given the right conditions,  
17 the capsid proteins of many viruses will assemble to capsid-like structures, rapidly, in high yield,  
18 and with high fidelity (Table 1). This observation opens up two distinct fields of study: the  
19 process of self-assembly and regulation of assembly. The process of self-assembly is best  
20 described by physical chemistry; for example, self-assembly can be modeled by a system of  
21 differential equations or emulated with small molecules. Regulating assembly, on the other  
22 hand, is fundamentally biochemical and ultimately biological in nature. Consider our hypothetical  
23 virus where unregulated assembly would yield capsids that did not contain the viral genome and  
24 toxic accumulation of fusion protein. These spontaneous assembly reactions must be delayed  
25 until the right time and place. Regulation is likely to be at the level of allostery, which is defined  
26 as a conformational change in a molecule, usually a protein, that alters its activity, induced by  
27 an effector molecule. In addition to allostery, regulation could require viral or host factors.  
28 Regulation of the large event of virus assembly by a small reaction, such as inducing  
29 conformational change of a single protein, provides the leverage that allows virus replication *in*  
30 *vivo*. Disruption of regulation is an ideal target for antiviral therapeutics. Defining the regulation  
31 that directly controls the physical chemistry of assembly is a field in its infancy and the focus of  
32 this review.

33 In this review, we will separately examine the assembly of virus capsids and the membrane-  
34 bound glycoprotein complexes present on the exterior of enveloped viruses. Both the capsid  
35 and glycoprotein are protein oligomers; from a physical-chemical perspective, their assembly is  
36 similar; from a biological perspective they demonstrably have similar elements of regulation.  
37 Though there are numerous experimental systems (Tables 1 and 2), the discussion of capsid  
38 assembly will focus on hepatitis B virus (HBV) with significant references to retroviruses and  
39 bacteriophage MS2. Furthermore, small molecules that dramatically affect HBV assembly have  
40 demonstrable antiviral activity. The discussion on glycoprotein assembly will focus on the  
41 formation of the spike complex of alphaviruses as the spike is composed of two proteins which  
42 intricately interact and are required for viral entry. Finally, we will discuss potential steps in  
43 assembly that could be targeted for designing antiviral therapeutics.

## 45 **Assembly of virus capsids**

46 The capsids of icosahedral viruses have tens to hundreds of copies of the capsid protein(s).  
47 The simplest vision of virus capsid assembly is one where rigid assembly units (AUs) collide by

1 Brownian motion, interact with perfect geometry, and associate irreversibly. Like any utopia, this  
2 vision fails examination: assembly simulations are incredibly sensitive to kinetic traps consisting  
3 of partial capsids that have a negligible chance of completion due to depletion of subunits  
4 (Figure 1).

5 Instead, assembly simulations, from master equations that treat assembly reactions as a well-  
6 mixed solution<sup>2,3</sup> to molecular dynamics that describe stochastic formation of single capsids<sup>4,5</sup>,  
7 concur on three generalizations (Box 1). First, weak interactions are necessary to minimize  
8 errors and kinetic traps. Weak interactions also contribute to a defined nucleation step. Second,  
9 nucleation minimizes the initiation of assembly, decreasing kinetic trapping of intermediates due  
10 to depletion of AUs. Third, the initial kinetic phase where there is little capsid formation is due to  
11 the time required to build the steady state of intermediates that supports subsequent assembly.  
12 More sophisticated mathematical models that incorporate the biological details of individual  
13 viruses (e.g. nucleic acid, scaffolding, and allostery) will help to identify opportunities to interfere  
14 with assembly.

15 Experimental observation of assembly of empty HBV capsids agrees well with the predictions of  
16 theoretical models<sup>6</sup> (Box 1). HBV has a T=4 icosahedral capsid composed of 120 assembly  
17 units, the homodimeric core protein (Cp)<sup>7</sup>. Cp can be assembled *in vitro* in response to ionic  
18 strength<sup>6</sup> and the kinetics are sigmoidal<sup>6</sup>. The average association energy between two subunits  
19 is -3 to -4 kcal/mol, corresponding to a millimolar dissociation constant. Because subunits are  
20 tetravalent, the weak association energy corresponds to a micromolar pseudo-critical  
21 concentration<sup>8</sup>. *In vitro* HBV assembly is resistant to, but not altogether immune to, kinetic  
22 traps<sup>9,10</sup>.

23 HBV is a simplest case system, a homopolymer of dimeric AUs. In cowpea chlorotic mottle virus  
24 and bacteriophage HK97, different complexes of AUs participate in assembly<sup>11,12</sup>. In many  
25 viruses, scaffold proteins support assembly. Bacteriophage P22 has a scaffold protein that  
26 thermodynamically and kinetically contributes to assembly<sup>13,14</sup>. Excess P22 scaffold can  
27 actually block assembly by trapping numerous intermediates<sup>14</sup>. Scaffolds can play roles in  
28 switching morphologies, as in bacteriophages P2 and P4<sup>15</sup>, and complex roles in subsequent  
29 maturation<sup>16</sup>. Thus, scaffold proteins could direct geometry, stability, and a measure of  
30 regulation by imposing stepwise assembly.

31 Viral nucleic acid can also serve as a molecular scaffold. Viral RNA has been considered as an  
32 'antenna' to attract free AUs<sup>17</sup> and as a solid that attracts and organizes AUs on its surface<sup>18</sup>.  
33 Assembly driven by RNA can be thought of in terms of the McGhee-von Hippel model of non-  
34 specific protein binding to a surface<sup>19</sup>, which considers that the association for nucleic acid (NA),  
35  $K_{NA}$ , is modified by a cooperativity coefficient,  $\omega$ , based on the protein-protein association  
36 constant. An  $\omega$  value of 1 indicates no cooperativity; an  $\omega$  value greater than 1000 results in  
37 steep cooperativity and reactions that appear to be two-state. Cowpea chlorotic mottle virus,  
38 which does not assemble under RNA-binding conditions, binds RNA with low cooperativity,  
39 displaying gradual assembly and partial capsids<sup>20</sup>. HBV, where the reciprocal of the pseudo-  
40 critical concentration is  $\sim 10^5$  (equivalent to  $\omega$ ) under physiological conditions, binds RNA  
41 tenaciously and with high cooperativity, resulting in quantitative assembly under mild  
42 conditions<sup>21</sup>. A nucleic acid scaffold can thus concentrate the capsid protein and provide  
43 additional association energy.

44

#### 45 **Assembly and nucleic acids as allosteric effectors**

46 Simple theoretical models fail to describe experimental results where induced conformational  
47 changes activate assembly. This behavior fits the definition of allostery. Nucleic acid-regulated

1 allostery is observed with bacteriophage MS2 and retroviruses, e.g. human immunodeficiency  
2 virus (HIV) and Rous sarcoma virus (RSV).

3 Retroviral Gag, the 'capsid protein' of immature retroviruses, is a multidomain protein. The  
4 elements of Gag that are most important for this discussion are the two-domain capsid (CA)  
5 segment (the primary mediator of Gag-Gag interactions), a spacer peptide that follows CA, and  
6 the RNA-binding nucleocapsid (NC) segment. The earliest suggestion of allostery in retrovirus  
7 assembly came from *in vitro* assembly studies using DNA oligomers and Gag<sup>22, 23</sup>. Gag  
8 dimerization was critical for assembly -- HIV Gag would not assemble with very short oligos nor  
9 if the oligo had high affinity sequences separated by a low affinity sequence, suggesting that  
10 protein-protein interaction was required to activate assembly<sup>24</sup>. Replacing NC with a leucine  
11 zipper promoted dimerization of Gag and was sufficient to drive assembly in cells<sup>25</sup>, though  
12 other factors were required *in vitro*<sup>26</sup>. What transitions are driven by protein oligomerization and  
13 DNA-binding? Small angle neutron scattering (SANS) indicated that HIV Gag had a propensity  
14 to fold on itself<sup>17</sup>, thus, straightening might be a critical assembly transition. NMR studies have  
15 indicated that the spacer sequence C-terminal to Gag (in HIV and RSV) can refold<sup>27, 28</sup>. This  
16 refolding transition, driven by binding nucleic acid, might expose surfaces on CA to allow  
17 interaction<sup>28</sup> and generate a new, potentially helix-rich interaction between the spacer  
18 peptides<sup>29, 30</sup>. Recent tomographic studies of immature HIV demonstrate CA-CA interactions  
19 and spacer peptide-spacer peptide interactions<sup>30, 31</sup>. Thus, it appears that Gag undergoes one or  
20 more conformational transitions to assemble into an immature lattice. These conformational  
21 transitions are involved in Gag assembly and are consistent with a regulatory effect, i.e. they  
22 prevent assembly by obstructing interactions<sup>28</sup> and this obstruction can be removed by binding a  
23 co-factor<sup>26</sup>.

24 In the RNA bacteriophage MS2, the role of allostery mediated by nucleic acid is unambiguous  
25 (Figure 2). These phages are organized with T=3 quasi-symmetry; their capsids are comprised  
26 of 90 chemically identical dimers that fit into AB and CC environments<sup>32</sup>. B half-dimers cluster  
27 around icosahedral fivefold vertices, A and C half-dimers alternate around icosahedral  
28 threefolds (quasi-sixfolds). The A and C half-dimers are very similar, while a loop on the B  
29 subunits has a different conformation, suggesting a structural switch. Uhlenbeck and co-workers  
30 showed that a specific RNA stem-loop from the genome, TR, was necessary and by itself  
31 sufficient to induce capsid assembly<sup>33, 34</sup>. By NMR, TR-free dimer was shown to be symmetrical  
32 in solution, suggestive of a CC dimer, but TR-bound dimer was asymmetric<sup>35</sup>. Pure RNA-free  
33 dimers or pure TR-saturated dimers do not assemble rapidly, i.e. they appeared kinetically  
34 trapped<sup>35</sup>. Addition of the other form of coat protein to trapped species resulted in rapid  
35 assembly<sup>36</sup>. These results showed that both forms of dimer, symmetric and RNA-bound  
36 asymmetric, are needed for efficient assembly and that RNA was an allosteric effector. Non-TR  
37 RNA stem-loops also trigger these effects, although they are much weaker binders, implying  
38 that multiple RNA-coat protein interactions within the capsids also contribute to conformer  
39 switching<sup>36, 37</sup>. Normal mode analysis has suggested that the RNA-induced conformational  
40 change might be one of dynamics, rather than static dimer conformations<sup>38</sup>, and is consistent  
41 with the lack of sequence-specificity. Analyzing mixed assembly reactions by mass  
42 spectrometry has provided a clear example of assembly by one dimer at a time rather than via  
43 coalescence of oligomers<sup>35</sup>. Examination of assembly reactions with TR incorporated into a  
44 longer RNA have lead to the surprising conclusion that assembly might be directed by an RNA  
45 scaffold<sup>36</sup>, funneling the system along only a few of the very many possible assembly paths<sup>36</sup>.  
46 Thus, initiation by an allosteric event and limitation by allosteric responses to protein-RNA and  
47 protein-protein interaction gives MS2 assembly the appearance of following a deterministic path.  
48 Interestingly, recent theoretical studies suggest that only a relatively small number of

1 intermediates will ever be used during assembly as a consequence of kinetic availability and  
2 thermodynamic stability<sup>39</sup>.

3

#### 4 **Allostery in capsid construction**

5 HBV Cp assembly is also an example of allosterically controlled assembly, even though its  
6 assembly superficially resembles association of rigid bodies. For example, HBV Cp binds  
7 cooperatively to Zn<sup>++</sup>, resulting in increased intrinsic fluorescence (indicating conformational  
8 change), and faster assembly kinetics<sup>10</sup>. Conformational differences between free Cp dimer and  
9 the same protein in the context of capsid are evident in the crystal structure of an assembly  
10 defective mutant (Figure 3b)<sup>40</sup>. Dimers from capsid are symmetrical about the dimer interface<sup>41</sup>,  
11 while free dimer is asymmetric. Comparison of the structures suggests a mechanical linkage of  
12 sub-domains connected to a central chassis; these sub-domains displace one another like so  
13 many molecular dominos, leading to movements of up to 9 Å and a conformation that is  
14 geometrically incompatible with an icosahedral structure. In comparison to the crystal structure,  
15 an NMR study under non-assembly conditions showed a symmetric dimer that was readily  
16 asymmetricized by interaction with a Cp-binding peptide<sup>42</sup>. Thus, the dimer is able to adopt  
17 multiple conformations, or families of conformations, tentatively categorized as assembly-active  
18 and assembly-inactive.

19

#### 20 **Allostery and antiviral strategies based on assembly**

21 For allostery, there must be a conformational change and an allosteric effector. Synthetic HBV  
22 assembly effectors have now been identified, the heteroaryldihydropyrimidines (HAPs) and  
23 phenylpropanamides. These show promise as antiviral agents, though the identity of the natural  
24 allosteric effector still remains unknown.

25 HAPs were discovered in a search for non-nucleoside inhibitors of HBV replication and only  
26 later shown to affect HBV production by a core protein-dependent mechanism<sup>43</sup>. *In vitro*  
27 experiments showed that HAPs could misdirect assembly and, more critically, HAPs strengthen  
28 dimer-dimer association and accelerate assembly kinetics, sometimes by orders of magnitude<sup>9</sup>.  
29 <sup>44</sup>. A crystal structure of a HAP-capsid complex (Figure 3a) shows that HAPs lead to a change  
30 in capsid quaternary structure (but not Cp tertiary structure), bowing out icosahedral fivefolds  
31 while flattening hexagonal arrangements of dimers<sup>41</sup>. These two effects explain the basis of  
32 assembly misdirection by destabilizing fivefolds while favoring sheets of Cp dimers<sup>9</sup>. The HAP  
33 molecule fills a hydrophobic pocket at the dimer interface, increasing buried surface area and  
34 presumably the association constant while slightly distorting the geometry of Cp-Cp  
35 interaction<sup>41</sup>.

36 Using a series of HAPs, a strong correlation was observed between antiviral effect and the rate  
37 of capsid formation, and a negligible correlation between antiviral effect and HAP stabilization of  
38 Cp-Cp interaction<sup>44</sup>. The low concentrations of the most effective HAPs required to suppress  
39 HBV replication suggest that misdirection (which requires stoichiometric concentrations of HAP)  
40 is not central to antiviral activity<sup>9</sup>. Similarly, phenylpropanamides, which were recently shown to  
41 accelerate assembly with little capsid stabilization and no assembly misdirection<sup>45</sup>, inhibit virus  
42 production in culture to yield empty capsids<sup>46</sup>.

43 We propose that the kinetic effects of HAPs and phenylpropanamides are the critical predictor  
44 of antiviral activity<sup>44, 45</sup>. This hypothesis leads to predictions about the nature of assembly and its  
45 allosteric activation. First, as per assembly models, a kinetic effect will be most evident in the  
46 nucleation step. Starting capsid assembly at an inappropriate time (e.g. without the viral genetic

1 material) is likely to be bad for the virus. Second, faster kinetics indicate that effectors decrease  
2 the energy barrier to assembly, which can be best explained by invoking an effector-induced  
3 conformational change. The graphic of the HAP site in the context of free and the capsid  
4 conformations (Figure 3c) supports this hypothesis. The key point is that only the nucleation  
5 step of an assembly reaction needs be affected to gain an overall enhancement of assembly  
6 kinetics (Figure 3d). Thus, the few effectors needed for nucleation can be leveraged to consume  
7 many copies of Cp for the production of failed virus.

8 As a point of comparison to assembly effectors, there are molecules that inhibit other allosteric  
9 transitions in viruses, such as stabilization of picornaviruses to uncoating and inhibition of HIV  
10 maturation. The WIN compounds (e.g. pleconaril) bind to a hydrophobic pocket in rhinoviruses  
11 and polioviruses and entropically inhibit the structural transition that enables release of the viral  
12 genome<sup>47, 48</sup>. A peptide inhibitor of HIV, 'CAI', inhibits the structural transitions between the N-  
13 terminal and C-terminal domains of CA that are associated with maturation, possibly favoring  
14 inappropriate geometries<sup>49</sup>. CAI also blocks assembly of immature Gag. However, consideration  
15 of assembly models suggests that simply inhibiting assembly could be losing strategy as it is  
16 easily overcome by overproduction of the capsid protein(s)<sup>50</sup>.

## 18 **Viral glycoprotein spikes**

19 While capsid and envelope glycoproteins have different functions, they share common  
20 principles that regulate their assembly. Viral glycoproteins, or spikes, are multimers that are  
21 assembled sequentially before becoming competent for catalyzing cell entry. The protein-  
22 protein interactions within a spike and between spikes are analogous to the interactions  
23 between AUs that drive capsid assembly. The processing events of the spikes parallel the  
24 nucleation events that initiate capsid assembly. Furthermore, factors such as environmental  
25 pH and peptides digested from the spikes as they are folded and refolded function as allosteric  
26 regulators and sensors during assembly.

27 The assembly of viral glycoprotein spikes requires a coordinated, highly regulated, processive  
28 mechanism to be successful. The alphavirus system, which is discussed below, demonstrates  
29 that even with less than a half-dozen proteins, the virus has developed an assembly mechanism  
30 that relies on cellular signals and the environment to initiate conformational or allosteric changes  
31 in the assembly process. Stabilizing any of the discrete conformations along the spike  
32 assembly path by binding of a small molecule would have tremendous potential for disrupting  
33 the timing and compartmentalization of this complex series of reactions.

## 35 **Assembly of alphavirus spike complexes**

36 Alphaviruses have 80 trimeric spikes, each ~70 Å in height. Each spike is a trimer of an E2-E1  
37 heterodimer, with the E2 protein at the top of the spike and the E1 protein lying underneath,  
38 close to the lipid membrane (Figure 4)<sup>51</sup>. In addition to interactions within each heterodimer,  
39 there are interactions between heterodimers. An E1 molecule from one heterodimer contacts  
40 the stalk of an adjacent E2 molecule, which is part of another heterodimer<sup>51</sup>. Each spike  
41 interacts with the nucleocapsid core via the cytoplasmic region of E2 and the E2 and E1  
42 proteins interact uniquely with each other via their transmembrane regions<sup>51</sup>.

43 The E2 protein binds to host-cell receptor and the E1 protein mediates fusion between the viral  
44 membrane and the host cell membrane. The receptor binding domain in E2 is positioned above  
45 the fusion peptide of E1<sup>51</sup>. One could speculate that upon receptor-binding, unzipping of the  
46 E2-E1 dimer must be initiated to ultimately expose the fusion protein to the host membrane.

1 What might control this association and subsequent disassociation? Viral fusion occurs in the  
2 endosome and is pH dependent<sup>52</sup>. pH sensors or amino acid residues that destabilize the E2-  
3 E1 dimer or that stabilize the post-fusion E1 trimer conformation have been identified<sup>53</sup>.  
4 Interfering with the assembly of the E2-E1 dimer, either through stabilization or destabilization,  
5 would severely affect the allosteric regulation by the sensor residues during fusion and, as a  
6 result, cell entry would be blocked.

7 The association of the alphavirus E2-E1 dimer is a regulated process that occurs early in  
8 assembly. The alphavirus structural proteins are translated from a sub-genomic RNA as a  
9 single polyprotein of capsid-E3-E2-6K-E1. The capsid protein autoproteolytically cleaves itself  
10 from the rest of the polyprotein in the cytoplasm<sup>54</sup>. The remaining structural polyprotein is  
11 targeted to the endoplasmic reticulum and cleaved to pE2 (or p62, corresponding to E3+E2),  
12 6K, and E1 proteins by cellular proteases<sup>55</sup>. In the ER, E1 undergoes several conformations  
13 ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) each differing in the number of disulfide bond rearrangements. pE2 associates only  
14 with the E1 $\beta$  intermediate. As a pE2-E1 dimer and during trimerization, E1 continues to  
15 undergo conformational changes until a stable E1 $\epsilon$  conformation is reached. Late in the  
16 secretory pathway E1 $\epsilon$  becomes metastable, possibly in response to cleavage of pE2 during  
17 virus maturation (see below)<sup>56</sup>.

18 One could speculate that pE2 undergoes similar disulfide bond rearrangement as E1 during  
19 assembly. Both the E1 and E2 proteins have 12 cysteine residues in their ectodomain, and  
20 both interact with resident disulfide isomerases<sup>57</sup>. pE2 has several CXXC motifs, characteristic  
21 of disulfide isomerase substrates<sup>58</sup>. It has been suggested that E3 might function as a disulfide  
22 isomerase during alphavirus assembly<sup>59</sup>. In the absence of pE2, E1 is not transported to the  
23 plasma membrane; in the absence of E1, low amounts of pE2 are transported to the plasma  
24 membrane<sup>56</sup>. These results emphasize the necessity of forming a specific pE2-E1 dimer early in  
25 assembly.

26 Recently, a translational frameshift was identified in the 6K protein coding region, producing a  
27 TF (for transframe) protein<sup>60</sup>. The synthesis of TF presents a conundrum for maintaining a  
28 balance of pE2 and E1 protein during assembly. When the 6K protein is translated, equal  
29 amounts of pE2 and E1 are translated. However, when TF is translated, E1 translation is  
30 abrogated, presumably drastically impacting the assembly of the pE2-E1 heterodimer.

31

## 32 **Activation of viral fusion proteins**

33 Viral glycoproteins fill many roles; the most common is mediating fusion between the viral and  
34 host membrane via the fusion protein. The atomic structures of viral fusion proteins in the pre-  
35 and post-fusion conformations have been determined and led to their classification based on  
36 structural similarities: Class I (orthomyxo-, paramyxo-, retro-, filo-, arena-, and coronaviruses),  
37 Class II (alpha- and flaviviruses), and Class III (rhabdo-, baculo- and herpesvirus)<sup>61, 62</sup>. The  
38 three classes are trimers in their fusion-active conformation. For most viruses, the final step of  
39 glycoprotein spike assembly is the activation of the fusion protein. This step brings the fusion  
40 protein into a metastable state, priming it for rapid fusion with the host cell upon proper  
41 activation.

42 The requirement of a chaperone protein and its cleavage during assembly is a property of the  
43 Class II fusion proteins. The cleavage of E3 from pE2 occurs in the trans-Golgi<sup>63</sup> and marks the  
44 transition from an immature to mature or fusogenic alphavirus particle. Immature and mature  
45 alphaviruses are structurally similar, but immature particles have reduced infectivity levels  
46 compared to mature particles because of the presence of E3 and/or because E1 is not in a  
47 metastable state able to facilitate fusion<sup>63</sup>. Retaining the E3 protein until the trans-Golgi might

1 serve to prevent premature fusion of the assembled spikes with membranes in the host cell  
2 secretory pathway. Thus, E3 is an effector of spike activity. At the plasma membrane the spikes  
3 are arranged in large hexagonal arrays<sup>64</sup>, creating a platform for virus budding. Host proteins  
4 have not been identified within the alphavirus glycoprotein array, suggesting lateral interactions  
5 between the spikes are important, in contrast to vesicular stomatitis virus<sup>65</sup>, HIV<sup>66</sup>, and influenza  
6 virus<sup>67</sup> which also bud from the plasma membrane. Though not usually described as such, pE2  
7 acts as an effector of E1 activation at two different times, as a chaperone and by releasing E3  
8 for activation.

9 Similar to pE2, the flavivirus prM protein chaperones folding of the flavivirus E protein, and  
10 cleavage of the pr peptide is required to transition from the immature to the mature flavivirus. In  
11 contrast to alphaviruses, immature and mature flaviviruses have radically different structures<sup>68</sup>.

12 Class I fusion proteins do not have a distinct chaperone. Activation of the fusion proteins occurs  
13 by cleaving the fusion protein itself. Multiple extracellular and cellular proteases have been  
14 proposed to cleave influenza HA0 both intra- and extracellularly, and we suggest that additional  
15 proteases for other viruses will also be identified<sup>69</sup>. The class III fusion proteins do not have a  
16 chaperone protein nor do they require a cleavage step to activate the protein into a fusion-active  
17 form. The vesicular stomatitis virus G protein is found in three reversible conformations in a pH-  
18 dependent equilibrium. The reversible conformations are a unique property among Class III  
19 fusion proteins, suggestive of allosteric conformational changes, which also prevents premature  
20 fusion during assembly<sup>70</sup>. Thus, the classification of fusion proteins is dependent on the  
21 structure of the fusion protein as well as the regulatory mechanism for its activation. Both  
22 processes are likely controlled through allosteric mechanisms.

23 We suggest that most viral families rely on a series of 'sensing' events that control protein  
24 disulfide bond linkages, glycosylation, acylation, phosphorylation, oligomerization, and transport  
25 to the appropriate cellular membrane for viral budding. These sensors respond to pH, redox,  
26 and lipid environment as the proteins are assembled and transported through the host secretory  
27 pathway<sup>53</sup> to allosterically enhance assembly and ensure productive spike formation.

28

## 29 **Conclusions and future directions: interfering with conformational change and control** 30 **during viral assembly**

31 While targeting specific conformations is an attractive antiviral strategy, we further propose that  
32 one should consider the dynamic nature of virus assembly and that interfering with the  
33 environmental sensors and their targets could be equally beneficial. Both capsid and  
34 glycoprotein assembly are carefully orchestrated events requiring activation, by known and  
35 unidentified effectors. For example, we have learned from the HAP molecules that interfering  
36 with assembly is not equivalent to inhibiting assembly. In addition, weak interactions between  
37 protein subunits or glycoprotein complexes minimize production of misassembled spikes. Thus,  
38 altering the kinetics, misdirecting the intermediates, stabilizing an interaction to prevent  
39 disassembly are all viable options that disrupt a highly evolved set of reactions to achieve the  
40 same end goal: reduced virus propagation.

41

## 42 **Acknowledgements**

43 The lists of assembly systems in Table 1 and structures in Table 2 are necessarily incomplete;  
44 we apologize for examples of excellent work that we have failed to include. We thank Drs. Alan  
45 Rein, Thomas Gallagher, Peter Stockley, and Stephen Stray for their comments. This effort was  
46 supported by NIH grants R01-AI067417 and R01-AI077688 to AZ.

47



## 1 References

- 2 1. Porterfield, J.Z., and Zlotnick, A. (2010) An Overview of Capsid Assembly Kinetics. In  
3 *Emerging Topics in Physical Virology* (Stockley, P.G., and Twarock, R., eds), Imperial College  
4 Press
- 5 2. Endres, D., and Zlotnick, A. (2002) Model-based analysis of assembly kinetics for virus  
6 capsids or other spherical polymers. *Biophys. J.* 83, 1217-1230
- 7 3. Keef, T., *et al.* (2006) Master equation approach to the assembly of viral capsids. *J*  
8 *Theor Biol* 242, 713-721
- 9 4. Elrad, O.M., and Hagan, M.F. (2008) Mechanisms of size control and polymorphism in  
10 viral capsid assembly. *Nano Lett* 8, 3850-3857
- 11 5. Rapaport, D.C. (2008) Role of reversibility in viral capsid growth: a paradigm for self-  
12 assembly. *Phys Rev Lett* 101, 186101
- 13 6. Zlotnick, A., *et al.* (1999) A theoretical model successfully identifies features of hepatitis  
14 B virus capsid assembly. *Biochemistry* 38, 14644-14652
- 15 7. Seeger, C., *et al.* (2007) Hepadnaviruses. In *Fields Virology* (Knipe, D.M., *et al.*, eds),  
16 2977-3029, Lippincott Williams & Wilkins
- 17 8. Ceres, P., and Zlotnick, A. (2002) Weak protein-protein interactions are sufficient to drive  
18 assembly of hepatitis B virus capsids. *Biochemistry* 41, 11525-11531
- 19 9. Stray, S.J., *et al.* (2005) A heteroaryldihydropyrimidine activates and can misdirect  
20 hepatitis B virus capsid assembly. *Proc Natl Acad Sci U S A* 102, 8138-8143
- 21 10. Stray, S.J., *et al.* (2004) Zinc ions trigger conformational change and oligomerization of  
22 hepatitis B virus capsid protein. *Biochemistry* 43, 9989-9998
- 23 11. Zlotnick, A., *et al.* (2000) Mechanism of capsid assembly for an icosahedral plant virus.  
24 *Virology* 277, 450-456
- 25 12. Hendrix, R.W., and Duda, R.L. (1998) Bacteriophage HK97 head assembly: a protein  
26 ballet. *Adv Virus Res* 50, 235-288
- 27 13. Prevelige, P.E., *et al.* (1993) Nucleation and growth phases in the polymerization of coat  
28 and scaffolding subunits into icosahedral procapsid shells. *Biophys.J.* 64, 824-835
- 29 14. Parent, K.N., *et al.* (2005) Electrostatic interactions govern both nucleation and  
30 elongation during phage P22 procapsid assembly. *Virology* 340, 33-45
- 31 15. Wang, S., *et al.* (2006) Assembly of bacteriophage P2 and P4 procapsids with internal  
32 scaffolding protein. *Virology* 348, 133-140
- 33 16. Fane, B.A., and Prevelige, P.E., Jr. (2003) Mechanism of scaffolding-assisted viral  
34 assembly. In *Virus Structure* (Chiu, W., and Johnson, J.E., eds), 259-299, Academic Press
- 35 17. Datta, S.A., *et al.* (2007) Conformation of the HIV-1 Gag protein in solution. *J Mol Biol*  
36 365, 812-824
- 37 18. Hagan, M.F. (2009) A theory for viral capsid assembly around electrostatic cores. *J*  
38 *Chem Phys* 130, 114902
- 39 19. McGhee, J.D., and von Hippel, P.H. (1974) Theoretical aspects of DNA-protein  
40 interactions: cooperative and non-cooperative binding of large ligands to a one dimensional  
41 homogeneous lattice. *J Mol Biol* 86, 469-489
- 42 20. Johnson, J.M., *et al.* (2004) Interaction with capsid protein alters RNA structure and the  
43 pathway for in vitro assembly of cowpea chlorotic mottle virus. *J. Mol. Biol.* 335, 455-464
- 44 21. Porterfield, J.Z., *et al.* (2010) Full-length HBV core protein packages viral and  
45 heterologous RNA with similar high cooperativity. *J Virol* 84, 7174-7184
- 46 22. Campbell, S., and Rein, A. (1999) In vitro assembly properties of human  
47 immunodeficiency virus type 1 Gag protein lacking the p6 domain. *J Virol* 73, 2270-2279
- 48 23. Campbell, S., and Vogt, V.M. (1995) Self-assembly in vitro of purified CA-NC proteins  
49 from Rous sarcoma virus and human immunodeficiency virus type 1. *J Virol* 69, 6487-6497.

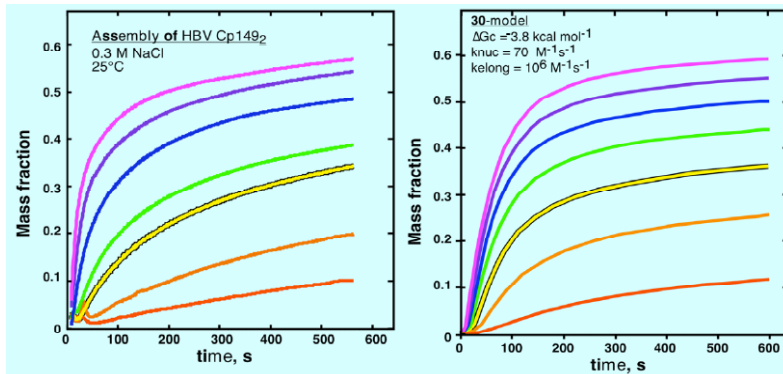
- 1 24. Feng, Y.X., *et al.* (2002) Reversible binding of recombinant human immunodeficiency  
2 virus type 1 Gag protein to nucleic acids in virus-like particle assembly in vitro. *J Virol* 76,  
3 11757-11762.
- 4 25. Johnson, M.C., *et al.* (2002) Nucleic acid-independent retrovirus assembly can be driven  
5 by dimerization. *J Virol* 76, 11177-11185.
- 6 26. Crist, R.M., *et al.* (2009) Assembly properties of human immunodeficiency virus type 1  
7 Gag-leucine zipper chimeras: implications for retrovirus assembly. *J Virol* 83, 2216-2225
- 8 27. Newman, J.L., *et al.* (2004) Flexibility in the P2 domain of the HIV-1 Gag polyprotein.  
9 *Protein Sci* 13, 2101-2107
- 10 28. Taylor, G.M., *et al.* (2010) NMR relaxation studies of an RNA-binding segment of the  
11 rous sarcoma virus gag polyprotein in free and bound states: a model for autoinhibition of  
12 assembly. *Biochemistry* 49, 4006-4017
- 13 29. Liang, C., *et al.* (2002) Characterization of a putative alpha-helix across the capsid-SP1  
14 boundary that is critical for the multimerization of human immunodeficiency virus type 1 Gag. *J*  
15 *Virol* 76, 11729-11737
- 16 30. Wright, E.R., *et al.* (2007) Electron cryotomography of immature HIV-1 virions reveals  
17 the structure of the CA and SP1 Gag shells. *EMBO J* 26, 2218-2226
- 18 31. Briggs, J.A., *et al.* (2009) Structure and assembly of immature HIV. *Proc Natl Acad Sci U*  
19 *S A* 106, 11090-11095
- 20 32. Valegard, K., *et al.* (1997) The three-dimensional structures of two complexes between  
21 recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA  
22 interactions. *J Mol Biol* 270, 724-738
- 23 33. Beckett, D., *et al.* (1988) Roles of operator and non-operator RNA sequences in  
24 bacteriophage R17 capsid assembly. *J Mol Biol* 204, 939-947.
- 25 34. Carey, J., *et al.* (1983) Sequence-specific interaction of R17 coat protein with its  
26 ribonucleic acid binding site. *Biochemistry* 22, 2601-2610
- 27 35. Stockley, P.G., *et al.* (2007) A simple, RNA-mediated allosteric switch controls the  
28 pathway to formation of a T=3 viral capsid. *J Mol Biol* 369, 541-552
- 29 36. Basnak, G., *et al.* (2010) Viral genomic single-stranded RNA directs the pathway toward  
30 a T=3 capsid. *J Mol Biol* 395, 924-936
- 31 37. Toropova, K., *et al.* (2008) The three-dimensional structure of genomic RNA in  
32 bacteriophage MS2: implications for assembly. *J Mol Biol* 375, 824-836
- 33 38. Dykeman, E.C., *et al.* (2010) Dynamic allostery controls coat protein conformer switching  
34 during MS2 phage assembly. *J Mol Biol* 395, 916-923
- 35 39. Moisant, P., *et al.* (2010) Exploring the paths of (virus) assembly. *Biophys J* 99, 1350-  
36 1357
- 37 40. Packianathan, C., *et al.* (2010) Conformational changes in the Hepatitis B virus core  
38 protein are consistent with a role for allostery in virus assembly. *J Virol* 84, 1607-1615
- 39 41. Bourne, C., *et al.* (2006) Global structural changes in hepatitis B capsids induced by the  
40 assembly effector HAP1. *J Virol* 80, 11055-11061
- 41 42. Freund, S.M.V., *et al.* (2008) Moving towards high-resolution descriptions of the  
42 molecular interactions and structural rearrangements of the human hepatitis B core protein. *J*  
43 *Mol Biol* 384, 1301-1313
- 44 43. Deres, K., *et al.* (2003) Inhibition of hepatitis B virus replication by drug-induced  
45 depletion of nucleocapsids. *Science* 299, 893-896
- 46 44. Bourne, C., *et al.* (2008) Small-molecule effectors of hepatitis B virus capsid assembly  
47 give insight into virus life cycle. *J Virol* 82, 10262-10270
- 48 45. Katen, S.P., *et al.* (2010) Trapping of hepatitis B virus capsid assembly intermediates by  
49 phenylpropanamide assembly accelerators. *ACS Chem Biol* DOI: 10.1021/cb100275b
- 50 46. Feld, J.J., *et al.* (2007) The phenylpropanamide derivative AT-130 blocks HBV  
51 replication at the level of viral RNA packaging. *Antiviral Res* 76, 168-177

- 1 47. Tsang, S.K., *et al.* (2000) Stabilization of poliovirus by capsid-binding antiviral drugs is  
2 due to entropic effects. *J Mol Biol* 296, 335-340
- 3 48. Speelman, B., *et al.* (2001) Molecular dynamics simulations of human rhinovirus and an  
4 antiviral compound. *Biophys J* 80, 121-129
- 5 49. Bartonova, V., *et al.* (2008) Residues in the HIV-1 capsid assembly inhibitor binding site  
6 are essential for maintaining the assembly-competent quaternary structure of the capsid protein.  
7 *J Biol Chem* 283, 32024-32033
- 8 50. Zlotnick, A., *et al.* (2002) A small molecule inhibits and misdirects assembly of hepatitis  
9 B virus capsids. *J Virol* 76, 4848-4854
- 10 51. Mukhopadhyay, S., *et al.* (2006) Mapping the Structure and Function of the E1 and E2  
11 Glycoproteins in Alphaviruses. *Structure* 14, 63-73
- 12 52. Kielian, M., and Helenius, A. (1985) pH-induced alterations in the fusogenic spike protein  
13 of Semliki Forest virus. *J Cell Biol* 101, 2284-2291
- 14 53. Sanchez-San Martin, C., *et al.* (2009) Dealing with low pH: entry and exit of alphaviruses  
15 and flaviviruses. *Trends Microbiol* 17, 514-521
- 16 54. Aliperti, G., and Schlesinger, M.J. (1978) Evidence for an autoprotease activity of sindbis  
17 virus capsid protein. *Virology* 90, 366-369
- 18 55. Garoff, H., *et al.* (1974) Isolation and characterization of the membrane proteins of  
19 Semliki Forest virus. *Virology* 61, 493-504
- 20 56. Carleton, M., *et al.* (1997) Role of glycoprotein PE2 in formation and maturation of the  
21 Sindbis virus spike. *J Virol* 71, 1558-1566
- 22 57. Molinari, M., and Helenius, A. (2000) Chaperone selection during glycoprotein  
23 translocation into the endoplasmic reticulum. *Science* 288, 331-333
- 24 58. Woycechowsky, K.J., and Raines, R.T. (2003) The CXC motif: a functional mimic of  
25 protein disulfide isomerase. *Biochemistry* 42, 5387-5394
- 26 59. Parrott, M.M., *et al.* (2009) Role of conserved cysteines in the alphavirus E3 protein. *J*  
27 *Virol* 83, 2584-2591
- 28 60. Firth, A.E., *et al.* (2008) Discovery of frameshifting in Alphavirus 6K resolves a 20-year  
29 enigma. *Virol J* 5, 108
- 30 61. Harrison, S.C. (2008) Viral membrane fusion. *Nat Struct Mol Biol* 15, 690-698
- 31 62. White, J.M., *et al.* (2008) Structures and mechanisms of viral membrane fusion proteins:  
32 multiple variations on a common theme. *Crit Rev Biochem Mol Biol* 43, 189-219
- 33 63. Zhang, X., *et al.* (2003) Furin processing and proteolytic activation of Semliki Forest  
34 virus. *J Virol* 77, 2981-2989
- 35 64. von Bonsdorff, C.H., and Harrison, S.C. (1978) Hexagonal glycoprotein arrays from  
36 Sindbis virus membranes. *J Virol* 28, 578-583
- 37 65. Brown, E.L., and Lyles, D.S. (2005) Pseudotypes of vesicular stomatitis virus with CD4  
38 formed by clustering of membrane microdomains during budding. *J Virol* 79, 7077-7086
- 39 66. Hammarstedt, M., *et al.* (2000) Minimal exclusion of plasma membrane proteins during  
40 retroviral envelope formation. *Proceedings of the National Academy of Sciences of the United*  
41 *States of America* 97, 7527-7532
- 42 67. Nayak, D.P., *et al.* (2009) Influenza virus morphogenesis and budding. *Virus Research*  
43 143, 147-161
- 44 68. Mukhopadhyay, S., *et al.* (2005) A structural perspective of the flavivirus life cycle. *Nat*  
45 *Rev Microbiol* 3, 13-22
- 46 69. Steinhauer, D.A. (1999) Role of Hemagglutinin Cleavage for the Pathogenicity of  
47 Influenza Virus. *Virology* 258, 1-20
- 48 70. Roche, S., and Gaudin, Y. (2002) Characterization of the equilibrium between the native  
49 and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is  
50 made of several trimers. *Virology* 297, 128-135

- 1 71. Bancroft, J.B. (1970) The self-assembly of spherical plant viruses. *Adv Virus Res* 16, 99-  
2 134
- 3 72. Bancroft, J.B., and Hiebert, E. (1967) Formation of an infectious nucleoprotein from  
4 protein and nucleic acid isolated from a small spherical virus. *Virology* 32, 354-356
- 5 73. Savithri, H.S., and Erickson, J.W. (1983) The self-assembly of the cowpea strain of  
6 southern bean mosaic virus: formation of T = 1 and T = 3 nucleoprotein particles. *Virology* 126,  
7 328-335
- 8 74. Lokesh, G.L., *et al.* (2002) A molecular switch in the capsid protein controls the particle  
9 polymorphism in an icosahedral virus. *Virology* 292, 211-223
- 10 75. Sastri, M., *et al.* (1999) Identification of a discrete intermediate in the  
11 assembly/disassembly of physalis mottle tymovirus through mutational analysis. *J Mol Biol* 289,  
12 905-918
- 13 76. Newcomb, W.W., *et al.* (1999) Assembly of the herpes simplex virus procapsid from  
14 purified components and identification of small complexes containing the major capsid and  
15 scaffolding proteins. *J Virol* 73, 4239-4250
- 16 77. Tellinghuisen, T.L., *et al.* (1999) In vitro assembly of alphavirus cores by using  
17 nucleocapsid protein expressed in *Escherichia coli*. *J Virol* 73, 5309-5319
- 18 78. Mukhopadhyay, S., *et al.* (2002) In vitro-assembled alphavirus core-like particles  
19 maintain a structure similar to that of nucleocapsid cores in mature virus. *J Virol* 76, 11128-  
20 11132.
- 21 79. Kirnbauer, R., *et al.* (1993) Efficient self-assembly of human papillomavirus type 16 L1  
22 and L1-L2 into virus-like particles. *J Virol* 67, 6929-6936
- 23 80. Salunke, D.M., *et al.* (1986) Self-assembly of purified polyomavirus capsid protein VP1.  
24 *Cell* 46, 895-904
- 25 81. Colomar, M.C., *et al.* (1993) Opening and refolding of simian virus 40 and in vitro  
26 packaging of foreign DNA. *J Virol* 67, 2779-2786
- 27 82. Rombaut, B., *et al.* (1984) In vitro assembly of poliovirus empty capsids: antigenic  
28 consequences and immunological assay of the morphopoietic factor. *Virology* 135, 546-550
- 29 83. Goodwin, S., *et al.* (2009) Foot-and-mouth disease virus assembly: processing of  
30 recombinant capsid precursor by exogenous protease induces self-assembly of pentamers in  
31 vitro in a myristoylation-dependent manner. *J Virol* 83, 11275-11282
- 32 84. Li, T.C., *et al.* (2005) Essential elements of the capsid protein for self-assembly into  
33 empty virus-like particles of hepatitis E virus. *J Virol* 79, 12999-13006
- 34 85. Gaussier, H., *et al.* (2006) Building a virus from scratch: assembly of an infectious virus  
35 using purified components in a rigorously defined biochemical assay system. *J Mol Biol* 357,  
36 1154-1166
- 37 86. Dryden, K.A., *et al.* (2006) Native hepatitis B virions and capsids visualized by electron  
38 cryomicroscopy. *Mol Cell* 22, 843-850
- 39 87. Huiskonen, J.T., *et al.* (2010) Electron cryotomography of Tula hantavirus suggests a  
40 unique assembly paradigm for enveloped viruses. *J. Virol.* 84, 4889-4897
- 41 88. Grunewald, K., *et al.* (2003) Three-dimensional structure of herpes simplex virus from  
42 cryo-electron tomography. *Science* 302, 1396-1398
- 43 89. Forster, F., *et al.* (2005) Retrovirus envelope protein complex structure in situ studied by  
44 cryo-electron tomography. *Proc Natl Acad Sci U S A* 102, 4729-4734
- 45 90. Zhang, W., *et al.* (2003) Visualization of membrane protein domains by cryo-electron  
46 microscopy of dengue virus. *Nat Struct Biol* 10, 907-912
- 47 91. Freiberg, A.N., *et al.* (2008) Three-dimensional organization of Rift Valley fever virus  
48 revealed by cryoelectron tomography. *J. Virol.* 82, 10341-10348
- 49 92. Ge, P., *et al.* Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science*  
50 327, 689-693

- 1 93. Beniac, D.R., *et al.* (2006) Architecture of the SARS coronavirus prefusion spike. *Nat*  
2 *Struct Mol Biol* 13, 751-752  
3 94. Cyrklaff, M., *et al.* (2005) Cryo-electron tomography of vaccinia virus. *Proc Natl Acad Sci*  
4 *U S A* 102, 2772-2777  
5 95. Harris, A., *et al.* (2006) Influenza virus pleiomorphy characterized by cryoelectron  
6 tomography. *Proc Natl Acad Sci U S A* 103, 19123-19127  
7 96. Welsch, S., *et al.* Electron tomography reveals the steps in filovirus budding. *PLoS*  
8 *Pathog* 6, e1000875  
9 97. Loney, C., *et al.* (2009) Paramyxovirus ultrastructure and genome packaging: cryo-  
10 electron tomography of sendai virus. *J Virol* 83, 8191-8197  
11 98. Rolfsson, O., *et al.* (2010) Mutually-induced conformational switching of RNA and coat  
12 protein underpins efficient assembly of a viral capsid. *J Mol Biol* 401, 309-322  
13 99. Zhang, W., *et al.* (2002) Placement of the structural proteins in Sindbis virus. *J Virol* 76,  
14 11645-11658  
15 100. Zlotnick, A. (1994) To build a virus capsid. An equilibrium model of the self assembly of  
16 polyhedral protein complexes. *J Mol Biol* 241, 59-67  
17  
18  
19

20 **Box 1. An analytical description of capsid assembly**



21  
22 Figure I. Comparison of HBV capsid assembly with simulations of capsid assembly. Assembly  
23 of a concentration series of HBV core protein (left) compared to assembly simulations of a 30-  
24 mer model (right), where subunits have an HBV-like geometry and the nucleus is trimeric<sup>1</sup>. The  
25 protein concentration increases from 4  $\mu\text{M}$  (red) to 10  $\mu\text{M}$  (purple) in 1  $\mu\text{M}$  increments. The  
26 calculation is parametrized to resemble HBV assembly. For the simulation, the bimolecular rates  
27 for nucleation ( $k_{\text{nuc}}$ ) and elongation ( $K_{\text{elong}}$ ) were  $70 \text{ M}^{-1}\text{s}^{-1}$  and a diffusion-limited  $10^5 \text{ M}^{-1}\text{s}^{-1}$ ,  
28 respectively; the pairwise association energy was  $-3.8 \text{ kcal/mol}$ .  
29

30  
31  
32  
33 Probably the simplest way to describe assembly of a capsid of N assembly units (AUs) is as a  
34 progressive series of intermediates<sup>100</sup>. To quantify this description, one can assume that all  
35 contacts between assembly units have the same energy, energies are additive, microscopic  
36 forward rates are identical (subject to statistical considerations), and that backward rates are the  
37 product of forward rates, dissociation constants and statistical terms. The result is a series of

1 rate equation each with terms for assembly and disassembly of  $n$ -mer and  $(n+1)$ mer,  
2 respectively. Two equations in the series are unique: the monomer equation references all  
3 intermediates and the final equation is a dead-end with only one build-up and one build-down  
4 term.

$$5 \quad d[n\text{-mer}]/dt = k_{\text{forward},n}[(n-1)\text{mer}][\text{monomer}] - k_{\text{forward},n+1}[n\text{-mer}][\text{monomer}] \\ 6 \quad \quad \quad - k_{\text{backward},n}[n\text{-mer}] + k_{\text{backward},n+1}[(n+1)\text{mer}]$$

7 Incorporating weak interactions between AUs and a slow nucleation step to limit the initiation of  
8 assembly results in robust reactions with minimal kinetics traps where errors and (meta)stable  
9 intermediates accumulate<sup>4, 39</sup>. This model necessarily leads to sigmoidal kinetics and, at  
10 equilibrium, a pseudo-critical concentration of free AU<sup>5, 100</sup>. The resulting model recapitulates  
11 most of the features observed *in vitro* (Figure I).

12 Assembly can be described in much greater detail as stochastic reactions using discrete event  
13 simulators or coarse-grained molecular dynamics<sup>4</sup>. These molecular simulations provide detail,  
14 showing conditions where association of intermediates could be an important path, where  
15 reversibility is critical<sup>5</sup>, and how subunit geometry can critically affect the size and shape of  
16 assembly products<sup>4</sup>. This statistical mechanical view of assembly leads to a view of assembly  
17 that is fundamentally similar to the thermodynamic-kinetic view of the previous paragraph<sup>4</sup>.  
18 However, the detail provides additional insights into the reactions and the behaviors that can be  
19 anticipated by biological molecules.

20

21

## 22 **Figure legends**

23 **Figure 1. Successful assembly and kinetic traps.** Assembly of a dodecahedron from  
24 pentagonal assembly units (faded colors) is more robust if regulated by nucleation. In nucleated  
25 assembly (top path), the initial association reactions are relatively slow, resulting in rare  
26 nucleation centers (brightly colored oligomer) and plenty of assembly units to allow completion  
27 of capsids. If assembly is unregulated (lower path), many partial capsids will form and very few  
28 assembly units will remain, kinetically trapping these intermediates. Of course, this trap might  
29 resolve if stochastic dissociation of some intermediates allows others to go to completion;  
30 alternatively, partial capsids can associate incorrectly, resulting in aberrant structures.

31

32 **Figure 2. Structure and assembly of bacteriophage MS2.** (a) The T=3 capsid of MS2 has 60  
33 AB dimers (blue and green, respectively) and 30 CC dimers (red), so that A and C subunits  
34 alternate around the quasi-sixfold axes. Icosahedral fivefold and quasi-sixfold (icosahedral  
35 threefold) vertices are shown in this representation. (b) The CC dimers are necessarily  
36 symmetric with an extended loop (highlighted in the open triangles) connecting the F and G  $\beta$   
37 strands at either end. Bound RNA, the TR stem-loop (in yellow), induces asymmetry in the AB  
38 dimers so that the FG loop of the B subunit (in the pentagon) adopts a unique conformation that  
39 allows it to pack around the fivefold vertices. (c) Pure CC dimers (red) or pure AB dimers (green  
40 and blue), with bound RNA, assemble rapidly when mixed, leading to the hypothesis that  
41 assembly proceeds in a stepwise reaction. The hexameric threefold cluster is one of the major  
42 complexes identified by mass spectrometry, suggesting a highly deterministic assembly path.  
43 This hypothesis has been strongly supported by further examination of RNA-driven assembly<sup>98</sup>.  
44 This figure was adapted from Ref. <sup>35</sup>.

45

1 **Figure 3. Structure an assembly of hepatitis B virus core.** (a) The HBV capsid is  
2 constructed from 120 dimers arrayed with T=4 icosahedral symmetry. One of the 60 icosahedral  
3 asymmetric units, an AB and CD dimer, are highlighted against on the 20 triangular facets that  
4 surface an icosahedron. Shown are two overlaid crystal structures of HBV (blue, PDB 2G33)  
5 and HBV with bound HAP1 assembly effector (red, PDB 2G34). More red is visible as HAPs  
6 cause the capsid to expand. Sixty HAPs are bound with high occupancy to a pocket in the C  
7 subunits (two are shown as white squares)<sup>41</sup>. (b) Dimers from the capsid (yellow) are  
8 symmetrical and identical. Asymmetric free dimers (blue, PDB 3KXS) are presumably in an  
9 assembly-inactive conformation. (c) A free dimer is color coded to show the chassis (blue) that  
10 is identical to the dimer from capsid and the three mobile subdomains (red, green, and orange).  
11 The subdomains are connected to the chassis by flexible joints, either Gly or Pro. Movement of  
12 one subdomain will force the others to shift their positions. A HAP molecule (magenta), modeled  
13 into its binding site, coordinates mobile subdomains with the chassis, facilitating the transition to  
14 an assembly-active state<sup>40</sup>. (d) HAPs accelerate assembly, even without misdirecting the  
15 reaction. Assembly of 10  $\mu\text{M}$  Cp initiated by ionic strength in the presence of 0 to 15  $\mu\text{M}$  HAP1<sup>9</sup>.  
16 The inset micrographs show that 2  $\mu\text{M}$  HAP1 yields spherical virus-like particles very fast,  
17 whereas higher HAP concentrations yield aberrant structures that scatter much more light.

18

19 **Figure 4. Structure of Sindbis virus.** (a) Surface view, looking down the icosahedral 2-fold  
20 axis, of Sindbis virus at 20 Å resolution. The triangular spikes are trimers of the E2-E1  
21 heterodimer. (b) Cross-section through an 11 Å resolution structure of Sindbis virus. The  
22 glycoprotein spikes are on the surface of the virus and their transmembrane region is seen  
23 traversing the lipid bilayer (transmem. helices). The nucleocapsid core is internal to the lipid  
24 bilayer with the RNA appearing disordered. The capsid and glycoproteins of alphaviruses both  
25 have icosahedral symmetry and their symmetry axes are aligned. This is in contrast to most  
26 enveloped viruses. (c) A surface view of the glycoprotein spikes showing the arrangement of  
27 the E2 and E1 proteins. Four monomers of E1 have been fit into the cryo-electron microscopy  
28 density of the 9 Å resolution Sindbis map and are shown in yellow, blue, red, and green. The  
29 yellow monomers surround the icosahedral three-fold axis while the blue, red, and green  
30 monomers surround the quasi-three-fold axis. The density corresponding to the E1 proteins has  
31 been subtracted from the density of the entire spike. The remaining density, which corresponds  
32 to the E2 protein, is shown in gray. The E2 protein covers the E1 protein and makes up the top  
33 portion of the spike. Interactions between E2-E1 heterodimers, along with those within the  
34 heterodimer, are thought to stabilize the spike complex. (d) A model of the allosteric changes  
35 that occur during assembly of the viral glycoproteins. The polyprotein is cleaved by cellular  
36 proteases to the individual proteins (green, E3; blue, E2; red, E1; 6K is not shown for clarity).  
37 Dimerization of pE2 and E1 to (pE2-E1) is followed by trimerization of the heterodimer, (pE2-  
38 E1)<sub>3</sub>. Last, cleavage of E3 from pE2 leaves a fusion competent trimer, (E2-E1)<sub>3</sub>. Changes in  
39 color indicate conformational changes in addition to changes in aggregation state. Panels (a),  
40 (b), and (c) were adapted from Ref.<sup>51, 99</sup>

41

42