Author for correspondence:
Giuliana Indelicato
e-mail: giuliana.indelicato@york.ac.uk

A coarse-grained model of the expansion of the human rhinovirus 2 capsid reveals insights in genome release

Giuliana Indelicato[1], Paolo Cermelli[2] and Reidun Twarock[1]

[1]Department of Mathematics, University of York, York, UK
[2]Department of Mathematics, University of Turin, Turin, Italy

GI, 0000-0002-3445-2473; PC, 0000-0002-2181-8800; RT, 0000-0002-1824-2003

Human rhinoviruses are causative agents of the common cold. In order to release their RNA genome into the host during a viral infection, these small viruses must undergo conformational changes in their capsids, whose detailed mechanism is strictly related to the process of RNA extrusion, which has been only partially elucidated. We study here a mathematical model for the structural transition between the native particle of human rhinovirus type 2 and its expanded form, viewing the process as an energy cascade, i.e. a sequence of metastable states with decreasing energy connected by minimum energy paths. We explore several transition pathways and discuss their implications for the RNA exit process.

1. Introduction

Human rhinovirus (HRV) is the cause of the common cold that affects billions of people each year. The small viral particle, about 30 nm in diameter, has a protein shell, called the viral capsid, that encapsulates and thus protects its single-stranded RNA genome. An essential step in the infection process is the structural rearrangement of the proteins in the capsid shell, as this rearrangement results in the formation of pores through which the genomic RNA is extruded during the infection. A better understanding of this mechanism may therefore point to novel targets for anti-viral therapy and prevention.

Conformational changes occur in a number of viruses during infection [1–4]. In each case, the spatial rearrangement of coat protein or coat protein domains results in the expansion of the capsid and the opening of pores through which the genomic material, and in some cases also other viral components, are released. We focus here on the pseudo $T = 3$ icosahedral capsid of HRV type 2 (HRV2), which is made of 60 protomers, each composed of four coat proteins VP1 to VP4 (figure 1*a*). While the three larger proteins (VP1, VP2, VP3) form the exterior surface of the particle, the smaller VP4 is located at the interior capsid surface. A characteristic feature of the protomer is the presence of a hydrophobic pocket located at the core of VP1, that in the native virus is occupied by the so-called pocket factor, presumably a fatty acid, which is believed to stabilize the capsid [5–10].

During infection, HRV attaches to the membrane of the cell at a receptor site, is internalized into the endosome, where it loses its pocket factors and the capsid expands, leading to the formation of the pores. The RNA exits the capsid through pores at the twofold axes and enters the host cell through channels in the endosomal membrane [11]. The mechanism by which RNA is released has not been completely elucidated, but experimental evidence suggests that exit occurs by an organized mechanism and is preceded by a substantial reorganization of the RNA inside the capsid [12]. The details of the conformational changes leading to the expansion of the capsid and the opening of the pores are likely to be related to the mechanism by which RNA is released.
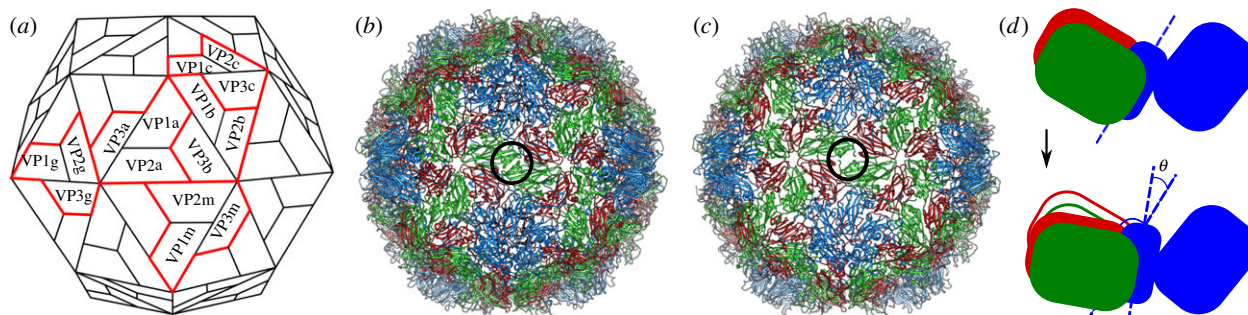
Figure 1. The capsid. (*a*) Sketch of the arrangement of the coat proteins VP1–VP3 in the native virion. Small-case Latin letters label coat proteins in the same protomer: 5 protomers are highlighted, labelled by (*a*, *b*, *c*, *g*, *m*). (*b*) The HRV2 native particle (PDB ID 1fpn). VP1s are displayed in blue, VP2s in green and VP3s in red. The black circle at the centre of the figure highlights a twofold position where pores are going to open as a consequence of protein displacement. (*c*) The expanded particle (PDB ID 3tn9—same colour coding as in (*b*) applies). (*d*) Details of the hinge movement within the protomer due to the relative rotation of the subunits of VP1 (adapted from [10]): VP1, VP2 and VP3 are coloured in blue, green and red, respectively. (Online version in colour.)

Moreover, it is known that for some viruses belonging to the same genus of HRVs, i.e. the Enterovirus genus, the engagement with the receptor triggers the conformational change [13–16].

In addition to the phylogenetic classification HRVs are divided into a major group and a minor group based on which cellular receptor they use for viral entry: major-group HRVs bind to the intercellular adhesion molecule 1 receptor, while the minor-group HRVs bind the low-density lipoprotein receptor. In the major group, the conformational changes of the capsid leading to RNA release are triggered by the interaction with the cellular receptor, whereas in the minor group, the low endosomal pH induces the removal of the pocket factor and the consequent viral structural rearrangements that leads to the expansion of the capsid.

We focus here on HRV2, a minor group rhinovirus, and we study the transition pathways between the native particle (figure 1*b*) and its genome-containing expanded form, the A-particle, which is a metastable intermediate on the pathway to the empty B (80S) particle (figure 1*c*). During this transition, individual protomers of the capsid undergo a conformational change, during which the two domains of VP1 move apart by an angle θ (figure 1*d*; called opening event in the following), and VP2 and VP3 move in concert with one of the domains. Opening of all protomers collectively results in expansion of the native particle by about 4% in radius and release of the VP4s.

The mathematical model used here is based on the coarse-grained approach developed in [17], in which the capsid proteins (CPs) are viewed as elementary units interacting by weak bonds, and the resulting energy landscape is explored by determining the paths and the energy barriers joining the metastable states. This allows to study the order in which the individual protomer transitions might occur, in order to identify the likely transition kinetics of the capsid expansion event.

In particular, we have addressed two issues that are relevant for understanding genome release, an essential part of the infection process (for more details, see the Discussion section).

First, we have investigated whether the structural transition of the capsid during its expansion is governed by diffuse nucleation events, with no regularity, or by a more organized domino effect, as suggested by the fact that interactions between the capsomers are relatively weak (the energy cascade hypothesis).

Second, and more importantly, given that it has been experimentally established that (i) RNA exit is directional, with the 3'-end exiting first, and (ii) the positions of the 5'- and 3'-ends inside the capsid are fixed in areas of the capsid that are roughly opposite to each other [18,19], the following question arises: What is the relation between the localization of the 3'-end and the opening of the pores in the capsid and in the endosome through which the 3'-end exits? Our model has been designed to address this issue. In particular, we explore whether there are pathways that induce a preferential opening of the pores near the 3' site, and we show that there is a reasonable parameter range in which this is the case. In fact, there is a clear separation in the parameter space between three possible modes of opening, and this suggests that a fine regulation of the RNA-extrusion process (i.e. the preferential opening of a twofold channel in the capsid near the attachment site of the 3'-end of the RNA) is possible without requiring regulation via the action of receptors. This is consistent with the fact that some strains of human rhinovirus, such as HRV2 that is under consideration here, do not require receptors for genome release [11], and suggests a principle of economy in the release mechanism.

Note that, in order to be effective, the above mechanism requires that the cascade is triggered at the protomer at which the 3'-end of the RNA is bound. This could perhaps be triggered by a specific interaction between genomic RNA and CP at that site, that impacts on that CP's conformation and its interactions with surrounding CPs. Conformational changes in CP in response to contact with an RNA stem-loop have been reported in other viruses before, such as the allosteric conformer switch in MS2 that is a prerequisite for capsid assembly [20,21]. Also, in human parechovirus (HPeV), a different picornavirus that does not cleave its VP0 into VP2 and VP4, we have recently shown that there are multiple dispersed RNA sequence/ structure motifs in the viral genome with affinity for CP that we termed packaging signals due to their role in capsid formation [22,23]. Cryo-electron microscopy studies of rhinovirus also show multiple dispersed contacts between genomic RNA and capsid, with a change in the contact pattern upon expansion [12]. In particular, in addition to the contacts close to the twofold axes that are present already in the native HRV2 particle, there are new contacts around the fivefold axes in the expanded A-particle (cf. fig. 2 in

[12]). It is, therefore, possible that contacts between genomic RNA and CP play a role in the release mechanism.

2. A model of the structural transition

During infection, HRV2 is internalized by the cell within endosomes, and the decrease in pH triggers a series of structural transformations of the capsid, leading to the expansion of the particle, the formation of pores and to the exit of the viral RNA into the cytosol. The particle has been imaged at different stages of the expansion: the structure of the native particle has been determined at 2.6 Å resolution by X-ray crystallography [6], while X-ray crystal structures at 6.4 Å and cryo-electron microscopy studies [12] showed that A (the genome-containing expanded form—see the Introduction) and B (the expanded and empty form) particles are almost identical, and an X-ray structure has been determined at 3.0 Å resolution for the empty particle [10]. Henceforth, in what follows, we identify the A capsid with the B capsid.

The full dynamics of the process is still unclear, although some of the occurring structural transformations have been in part elucidated. It is generally acknowledged that, as a consequence of the decrease in pH, the pocket factor at the core of VP1 is released from its location, and this allows the relative rotation of two domains of VP1 (the α helix and the C-terminus move away from the β-barrel), resulting in a change in the conformation of the monomer and the collapse of the pocket. This hinge movement affects VP2 and VP3 positions as they displace in concert with one of VP1 domains [7,8,11].

The relative motion of the subunits of VP1 together with VP2, VP3 leads to an increase of the diameter of the capsid, and results in the opening of three types of pores, two of which at the quasi-threefold and twofold axes. The other channels are located at symmetry-related positions (around the fivefold axes) and supposedly allow to externalize a portion of VP1 (N-terminal residues) that is thought to be instrumental in the adhesion of the capsid to the interior of the endosomal membrane, through which the RNA must pass in order to enter the cytosol.

The pores at the twofold axes are relevant for infection because they are wide enough to allow the transit and exit of VP4 and the RNA. With reference to figure 1*a* they are located at the interfaces VP2a:VP2m and at positions related by icosahedral symmetry, and their formation is caused by the relative clockwise motion of VP2a and VP2m that move away from each other as a result of the hinge movement within the protomers [10]. There is evidence that the RNA leaves the capsid at one of the pores at the twofold axes [24,25], and the process starts at its 3' end [18]. Furthermore, in the native virion, the 5' end, which is the end at which RNA synthesis begins, is bound inside the capsid to a CP at a threefold site roughly opposite the twofold exit side.

2.1. The coarse-grained model

The expansion of the capsid is the result of a collective rearrangement induced by rigid body motions of protomer domains (figure 1*b,c,d*). These rigid motions will be parametrized by a single angular variable, one for each protomer, denoted by $\theta_i \in [0, 1]$ for $i = 1, \ldots, 60$, with the convention that $\theta_i = 0$ and $\theta_i = 1$ correspond to the closed and open configuration of the protomer, respectively. Also, we assume that the protomers are labelled as in figure 2*a* on a Schlegel diagram of the capsid. The numbering scheme has been

chosen such that protomers in the same pentamer are consecutively numbered, and that the numbers of opposite protomers add up to 61. Each capsid configuration is associated with a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{60})$. Hence, the two states $\theta_i = 0 \ \forall i$ and $\theta_i = 1 \ \forall i$ correspond to the native (closed) and expanded (open; A particle) capsid.

The changes in the capsid may occur either simultaneously or as a cascade of subsequent events, and we model it as a sequence of elementary transitions between metastable states, i.e. local minima, of a suitable energy function. The energy cascade mechanism has been demonstrated for the maturation of HK97 in [26,27], and mathematical models based on these ideas have been developed in [17,28].

We assume that the free energy is the sum of three components, that take protein–protein interactions into account, and model protein–RNA interactions as parameters modifying these energy terms. The first term, that we call the intra-protomer energy, accounts for the forces that drive the structural transition of a protomer after the removal of the pocket factor, as well as the barrier that has to be overcome to break the bonds necessary for this process. The second term is the intra-pentamer energy, that captures contributions from the bonds between the adjacent protomers within the same pentamer. Finally, the inter-pentamer energy accounts for the bonds between adjacent protomers belonging to different pentamers.

The above terms only account for the energy barriers related to the breaking of bonds between different capsomers (or domains of the same protomer, as is the case for VP1). However, the interactions between the capsomers are much more sophisticated, and certainly include mechanical knock-on effects due to their mutual push and pull during the opening of the protomers and the expansion. To include these effects, we introduce a weak constraint on the motion of adjacent capsomers via a penalization term in the total energy.

2.2. Intra-protomer interactions

The relative rotation of the subunits of VP1 upon removal of the pocket factor has the effect of widening the angle between the domains by θ, that we will call the opening angle in the following, thereby disrupting and forming a number of bonds between the subunits. The impact of this conformational change on the protein interfacial energies has been quantified in [10] (see also table 1).

We describe the switch between the two configurations of the protomer via a two-well energy function, with minima at the closed and open configuration of the protomer. It is this energetic term that drives the expansion of the capsid in our model.

To motivate our choice of two-well energy, we note that while in other viruses (cowpea chlorotic mottle virus, equine rhinitis A virus (ERAV)) there is evidence that the confinement of the RNA inside the capsid induces an internal pressure that drives the expansion, no evidence for such a mechanism is present for HRV2. Rather, the idea here is that, after the removal of the pocket factor, the domains of VP1 within a protomer are free to attain a preferred stable configuration, which suggests that the intra-protomer energy must have a minimum $h < 0$ in correspondence of this metastable state ($\theta = 1$).

However, as mentioned above, in order to reach this state the two domains of VP1 have to break a number of
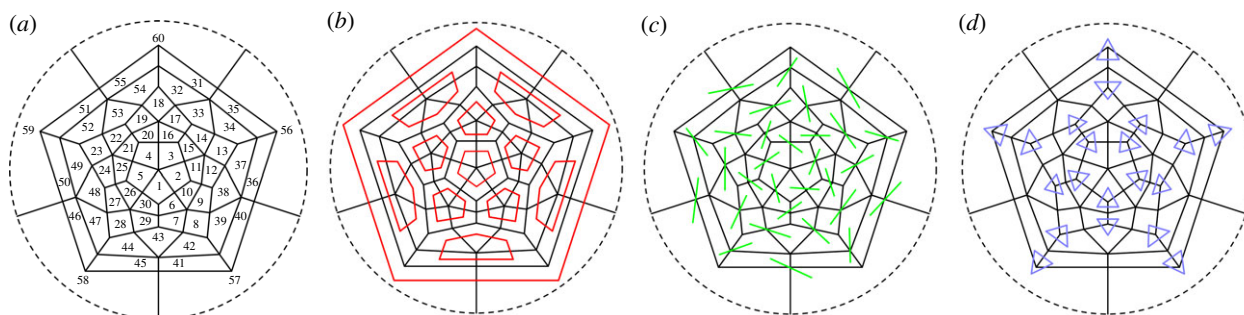
Figure 2. (*a*) Schlegel diagram of the capsid architecture, showing the numbering scheme used to identify individual protomers in the viral capsid. (*b*) Intra-pentamer interactions around the particle fivefold axes (adjacency matrix A_{ij}) are shown as pentagons (red in the online version). (*c*) Inter-pentamer interactions across the particle twofold axes (adjacency matrix B_{ij}) are shown as lines (green in the online version). (*d*) Inter-pentamer interactions around the particle threefold axes (adjacency matrix C_{ij}) are shown as triangles (blue in the online version). (Online version in colour.)

bonds, which suggests that the closed form of the protomer is metastable (hence the first minimum at $\theta = 0$) and that there is a barrier k between these two wells whose size is proportional to the energy necessary to break those bonds.

Finally, we know that (for values of the pH as in the endosome and in the absence of the pocket factor) the open capsid is more stable than the closed one, which suggests that the minimum at $\theta = 1$ is deeper than the minimum at $\theta = 0$, so that the reverse transition is more difficult than the forward one.

Let henceforth $f(\theta)$ be a two-well energy as in figure 3*a*, parametrized by two coefficients k and h, such that

— f has two minima, one at $\theta = 0$ with $f(0) = 0$ and the other at $\theta = 1$, that correspond to the closed and open configurations of the protomer (after the pocket factor has been removed), respectively; the quantity $f(1) = h < 0$ is the depth of the energy well corresponding to the open configuration of the protomer.
— f has a single maximum at $\theta_m \in (0, 1)$, whose value is $f(\theta_m) = k > 0$. Thus k represents the energy barrier between the wells and reflects the number of bonds that must be broken in order to complete the rotation of the VP1 subunits. According to table 1, the relevant proteins interacting are (see also figure 1):

$$VP1a{:}VP2a, \quad VP1a{:}VP3a, \quad VP2a{:}VP3a. \qquad (2.1)$$

We assume that the total intra-protomer energy of the capsid has the form

$$E_{\text{protomer}}(\boldsymbol{\theta}) = \sum_{i=1}^{60} f(\theta_i). \qquad (2.2)$$

In the simulations described later, we have tested four different explicit forms for the function f, without finding any significant difference. The simplest one is the fourth-order polynomial

$$f(x) = hx^2 \frac{(3x^2 - 4(p+1)x + 6p)}{(2p - 1)},$$

with p a solution of $p^3(p-2)/(2p-1) = -k/h$, where k and h are as in figure 3*a*. We have also tried piecewise smooth functions obtained by interpolating fixed points with splines and, finally, a Gaussian mixture model

$$f(x) = -\ln\left(a_1\, e^{-(x^2/s_1)} + a_2\, e^{-((x-1)^2/s_2)}\right),$$

with $a_1 = 1$, $a_2 = 400$, $s_1 = 1/20$, $s_2 = 1/15$, inspired by recent techniques developed to reconstruct low-dimensional dynamical transition networks from high-dimensional static samples [29].

2.3. Intra-pentamer interactions

This term refers to the cohesive interactions among two adjacent protomers in the same pentamer. Specifically, according to table 1, the intra-pentamer interactions are

$$VP1a{:}VP1b, \; VP1a{:}VP3b, \; VP2a{:}VP1b$$
$$VP2a{:}VP3b, \; VP3a{:}VP1b, \; VP3a{:}VP3b, \qquad (2.3)$$

that, referring to the Schlegel diagram in figure 2*a*, correspond to the red segments in figure 2*b*.

The corresponding intra-pentamer energy is the energy that is required in order to break the intra-pentamer bonds that block the opening of each protomer. We assume it to have the general form

$$E_{\text{pentamer}}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i,j=1}^{60} A_{ij} g(\theta_i, \theta_j), \qquad (2.4)$$

where A_{ij} is the adjacency matrix of the graph in figure 2*b*, whose nodes are the protomers and whose edges are the intra-pentamer bonds (red segments). We require that the function g has the form $g(\theta_i, \theta_j) = \tilde{g}(a(\theta_i^2 + \theta_j^2))$ where $\tilde{g}$ is smooth and monotonically increasing from 0 to 1 in the interval $[0, +\infty)$, $\tilde{g}(0) = 0$, and $\tilde{g}(x) \equiv 1$ for $x \geq 1$; here $1/a > 0$ is a measure of a typical interaction radius. In our simulations, we will use

$$g(\theta_i, \theta_j) = 1 - e^{-a(\theta_i^2 + \theta_j^2)}, \qquad (2.5)$$

with $a \gg 1$ sufficiently large (cf. also [30] and figure 3*b*).

The interaction energy $g(\theta_i, \theta_j)$ has a sharp minimum at $\theta_i = \theta_j = 0$, i.e. when both adjacent protomers are closed and undeformed. However, when either $\theta_i = 1$ or $\theta_j = 1$, so that at least one of the protomers is open, the energy is maximal, since it is sufficient that one of the protomers is open to break the bonds involved in the interaction with its neighbours.

2.4. Inter-pentamer interactions

This term accounts for the contributions of the cohesive interactions among protomers belonging to different pentamers,
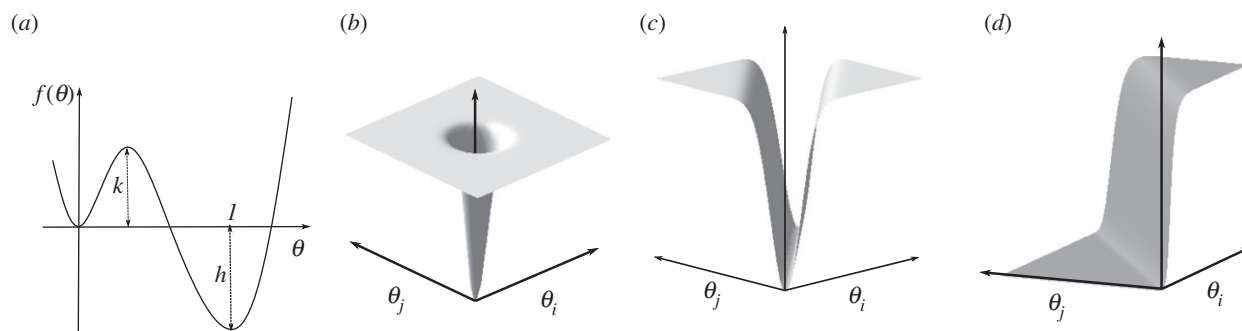
Figure 3. The individual energy terms for intra-protomer, intra-pentamer, inter-pentamer interactions and the terms for the mechanical and steric interactions. (a) Bistable intra-protomer energy $f(\theta)$; (b) plot of the function $g(\theta_i, \theta_j)$ in the intra-pentamer and inter-pentamer interaction energies; (c) plot of the function $r(\theta_i, \theta_j)$ accounting for the steric and mechanical interactions around a threefold axis; (d) plot of the function $s(\theta_i, \theta_j)$ accounting for the steric and mechanical interactions around a fivefold axis. (Online version in colour.)

Table 1. Changes in protein–protein interactions during the transition from native to 80S particle. The table and its description are adapted from the electronic supplementary material in Garriga et al. [10], and the labelling of the proteins is as in figure 1a. The number of unique residue pairs that contact each other at the subunit interface is listed for each interface in the native and the 80S particles. The numbers of mutual interactions in both capsids are listed in the fifth column. The sixth and seventh column indicate the number of interactions that are only present in the native and in the 80S particle, respectively. The association energies (kcal mol^{-1}) for each interface of each capsid type are stated in the eighth and ninth columns. The interacting residue pairs, association energies for each interface were calculated using VIPERdb tools.

		interactions					association energy	
		native	80S	conserved	lost	new	native	80S
intra-protomer	VP1a–VP2a	82	75	71	11	4	−84.7	−70.2
	VP1a–VP3a	151	110	96	55	14	−188.8	−142
	VP2a–VP3a	66	62	57	9	5	−68.9	−68.5
intra-pentamer	VP1a–VP1b	44	26	19	25	7	−40.7	−33.4
	VP1a–VP3b	42	28	20	22	8	−56.9	−38.8
	VP2a–VP1b	14	0	0	14	0	−16.7	0
	VP2a–VP3b	33	20	16	17	4	−28.1	−27.5
	VP3a–VP1b	10	7	7	3	0	−12.5	−3.9
	VP3a–VP3b	33	30	29	4	1	−41.1	−38
	VP3a–VP3c	4	4	4	0	0	−3.6	−3.1
inter-pentamer	VP1a–VP2g	20	0	0	20	0	−20.3	0
	VP3a–VP2g	63	49	33	30	16	−54.6	−47.4
	VP2a–VP2m	14	6	1	13	5	−26.5	−11.4

that we identify with the energy of the bonds between these protomers that have to be disrupted in order to expand the capsid. We distinguish between interactions around the threefold and twofold axes. According to table 1, the inter-pentamer interactions around the threefold axes are

$$VP1a{:}VP2g, \ VP3a{:}VP2g, \ VP3a{:}VP3g. \quad (2.6)$$

Referring to the Schlegel diagram in figure 2a, the inter-pentamer interactions around the threefold axes are indicated by the blue segments in figure 2d. Analogously, the interactions across twofold axes, i.e.

$$VP2a{:}VP2m, \quad (2.7)$$

correspond to the green segments in figure 2d. Consistent with this, we assume that the inter-pentamer energy is the sum of

two contributions,

$$E_{3\text{fold}}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i,j=1}^{60} B_{ij}g(\theta_i, \theta_j) \quad \text{and}$$

$$E_{2\text{fold}}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i,j=1}^{60} C_{ij}g(\theta_i, \theta_j), \quad (2.8)$$

where B_{ij} and C_{ij} are the adjacency matrices of the graphs in figures 2c,d respectively, i.e. the graphs whose nodes are the protomers and whose edges are the intra-pentamer bonds, and the function g is as in (2.5).

2.5. Mechanical constraints
Neighbouring proteins tend to impact on each others motion via steric and mechanical constraints. To simplify, we assume that protomers push and pull each other across interfaces

within the pentamer (figure 2b/ fivefold axes), and interfaces between different pentamers (figure 2d/ threefold axes). Because of the relative positions of the VP1 domains involved in the openings of the individual protomers, we only need to consider interactions at the *ab* (cf. (2.3)) and *ag* interfaces (cf. (2.6)), neglecting possible steric interactions across the *am* interface (VP2a:VP2m). This is because the interface lies at the twofold axis where the main pore opens as protomer domains move away from each other; in fact, the expansion of the capsid tends to separate these interfaces, after breaking the cohesive bonds involved in the inter-protomer energy.

We account for the steric and mechanical interactions by energetic terms that penalize the separation between adjacent protomers. In our simplified model, we do not study the actual motion of the protomers in detail, but we assume that it can be represented by a function of the variables θ_i, so that it is reasonable to enforce the constraints by penalizing the difference $\theta_i - \theta_j$, where i and j are adjacent protomers.

To describe the steric and mechanical interactions between different pentamers around the threefold axes, we penalize the relative motion of the protomers through a term of the form

$$E_{S3}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{ij=1}^{60} C_{ij} r(\theta_i, \theta_j), \qquad (2.9)$$

where C_{ij} is the adjacency matrix of the graph in figure 2d and moreover $r(\theta_i, \theta_j) = \tilde{r}(b(\theta_i - \theta_j)^2)$ (cf. figure 3c). Here, the function $\tilde{r}(x)$ satisfies the same requirements as the function $\tilde{g}$ in §2.3, and $b \gg 1$ is a real constant measuring the rigidity of the proteins, with small values meaning soft and large values meaning rigid. In our simulations, we shall use

$$r(\theta_i, \theta_j) = 1 - e^{-b(\theta_i - \theta_j)^2}. \qquad (2.10)$$

As to the intra-pentamer interfaces, since the opening of a protomer induces a clockwise rotation of the pentamers, we can assume that the opening results in a push of the adjacent clockwise protomer. This introduces a chirality in the corresponding penalization of the energy, that has the form

$$E_{S5}(\boldsymbol{\theta}) = \sum_{ij=1}^{60} \tilde{A}_{ij} s(\theta_i, \theta_j), \qquad (2.11)$$

where

$$\tilde{A}_{ij} = \begin{cases} A_{ij} & \text{if } i > j \bmod 5 \\ 0 & \text{otherwise} \end{cases}$$

is the matrix that takes into account only the clockwise adjacencies within the pentamer, and the push force from i to j is accounted for by the function

$$s(\theta_i, \theta_j) = \begin{cases} r(\theta_i, \theta_j) & \text{if } \theta_i > \theta_j, \\ 0 & \text{otherwise.} \end{cases}$$

2.6. The total energy
In summary, the total energy is

$$E(\boldsymbol{\theta}) = E_{\text{protomer}}(\boldsymbol{\theta}) + c_1 E_{\text{pentamer}}(\boldsymbol{\theta}) + c_2 E_{\text{2fold}}(\boldsymbol{\theta}) \\ + c_3 E_{\text{3fold}}(\boldsymbol{\theta}) + c_4 E_{S3}(\boldsymbol{\theta}) + c_5 E_{S5}(\boldsymbol{\theta}), \qquad (2.12)$$

where c_i, $i = 1 \ldots 5$ are parameters indicating the relative strengths of individual energy contributions. In particular, the functions E_{pentamer}, E_{2fold}, E_{3fold}, E_{S3} and E_{S5} are all

normalized to the same height of the plateau, so that the energy barriers corresponding to the number of bonds lost in the rearrangement of the capsid across the interfaces (2.3), (2.6), and (2.7), respectively, are accounted for by the parameters c_1, c_2 and c_3.

Recall now that the intra-protomer energy E_{protomer} involves two parameters k and h, with k the energy barrier necessary to break the bonds within the protomer and allow for its opening (cf. figure 3a). The actual value of the barrier is 89.5 kcal mol^{-1} (cf. table 1), and corresponds to the number of bonds broken during the transition according to table 1. We normalize the energy so that $k = 1$ by dividing all constants by this value, i.e.

$$c_1 = \frac{92.8}{89.5}, \quad c_2 = \frac{24.6}{89.5} \quad \text{and} \quad c_3 = \frac{48.9}{89.5}. \qquad (2.13)$$

The constants c_4 and c_5 represent values of the energy required to violate the steric constraints. In the absence of experimental values, we will discuss below their role and the impact of different choices for these values. As to the constant h, which is the depth of the well of the intra-protomer energy corresponding to the open state of the protomer, we have chosen it to be $h = -6$ in order to ensure that the total energy decreases at each transition event (cf. §3.1).

3. The expansion of the capsid
Our model describes the expansion of the capsid via the sequential opening of one or more protomers as a sequence of transitions between metastable states, i.e. local minima, of the total energy. To explore the energy landscape of the capsid we use a technique borrowed from the theory of large deviations [31], based on the assumption that the transitions occur along minimum energy paths (MEPs), which are paths connecting minima that occur with maximum probability when small stochastic fluctuations are included into the model [32].

3.1. Minimum energy paths
We briefly review below the above technique, referring to [32–35] for a detailed motivation and description.

To describe MEPs between two minima of the energy function, assume that small fluctuations around a minimum are described by a stochastic dynamical system of the form $d\boldsymbol{\theta} = -\nabla_{\boldsymbol{\theta}} E \, dt + \epsilon \, d\boldsymbol{W}$, where ε tends to zero ($\varepsilon \ll 1$) and $\boldsymbol{W}$ is a multidimensional Brownian motion. Every trajectory of this system starting near a minimum exits its basin of attraction with probability 1 and enters the basin of attraction of another minimum, and the trajectories concentrate with high probability near a special path that connects the two minima, called the MEP. The MEP between two minima can be computed by a technique invoking a curve evolution equation [32–34].

Consider now the collection of all metastable states of the capsid, called the state space $\mathcal{S}$. The cardinality of $\mathcal{S}$ is 2^{60}, implying that the problem of studying all possible transitions is computationally intractable. Therefore, we have restricted our study only to transitions that involve the opening of a single protomer at a time. We have also numerically explored some reverse transitions, corresponding to the closing of some of the protomers, as well as some of those involving the simultaneous opening of at least two protomers, but we

have consistently found larger energy barriers than those corresponding to single protomer openings.

Hence, we define a discrete-time Markov chain with state space $\mathcal{S}$ such that the transition probability from state $\boldsymbol{\theta}$ to state $\boldsymbol{\theta}'$ is

$$P(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{\exp(-2\beta(\boldsymbol{\theta}, \boldsymbol{\theta}')/\epsilon^2)}{\sum_{\tilde{\boldsymbol{\theta}} \in \mathcal{S}} \exp(-2\beta(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})/\epsilon^2)}, \tag{3.1}$$

where the states $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are such that there exists i for which $\theta_i = 0$ and $\theta'_i = 1$, and $\theta_j = \theta'_j \in \{0, 1\}$ for $j \neq i$, and there is a MEP between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ such that the energy computed along the path has a single maximum at a point $\boldsymbol{\theta}^*$. We set $P(\boldsymbol{\theta}, \boldsymbol{\theta}') = 0$ otherwise. If the transition probability is not zero, we define the barrier between the two states as [35]

$$\beta(\boldsymbol{\theta}, \boldsymbol{\theta}') = E(\boldsymbol{\theta}^*) - E(\boldsymbol{\theta}). \tag{3.2}$$

Note that our choice of the transition probabilities implies that we are forcing the transition to be irreversible, which is not usually the case in large deviations theory. Our choice follows from the fact that closed empty HRV2 capsids have not been observed experimentally, which suggests that the actual process of expansion is indeed not reversible, even though there are some viruses, such as ERAV, for which closed empty capsids are observed *in vitro* under specific experimental conditions.

Irreversibility is also implied by the fact that the barriers for the forward transitions (closed to open) in our model are consistently smaller than those for the reverse transition. This is indeed a consequence of our choice of the intra-protomer energy f. In fact, according to our general scheme, the expansion of the capsid is due to the protomer reaching a more stable position, when the obstruction due to the pocket factor is removed. This is modelled by the deep well in the function $f(\theta)$. All other energy terms are cohesive and counteract the expansion. The likelihood of the reverse process (i.e. closing versus opening) is related to the relative depths of the energy wells: if the 'open' well is deeper than the 'closed' well, it is less probable to go back to the closed state than to go forward from the closed to the open state. Hence, in order to account for the difficulty (if not impossibility) of the reverse transition, we have assumed the energy well in the intra-protomer energy f to be deep enough so that the global energy (the sum of all contributions) actually decreases in each protomer opening, and the barrier for closing is higher than the barrier for opening. Note finally that this fact also implies that the energy decreases at each transition step: the energy cascade.

3.2. An approximate expression of the barriers

We prove in appendix A three basic facts that facilitate the analysis of the transition mechanism. In particular, for large a and b, we have the following results:

(a) The configurations in which each protomer is either closed ($\theta_i \sim 0$) or open ($\theta_i \sim 1$) are local minima of the energy.
(b) The MEPS are approximately rectilinear paths between minima: in particular, focusing on transitions that correspond to the opening of a single protomer, let $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ be two metastable states such that $\theta_j = \theta'_j$ if $j \neq i$ and $\theta_i = 0$, $\theta_i' = 1$ for some protomer i. Then, if a MEP

connecting $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ exists, its parametric expression as $b \to +\infty$ is

$$\theta_i(t) \sim t, \quad \theta_j(t) \sim \theta_j \equiv \theta'_j \quad \text{for } j \neq i, \ t \in [0, 1]. \tag{3.3}$$

(c) Finally, the barriers can be estimated in terms of the constants k and $c_1, \ldots, c_5$ in the expression of the total energy (2.12). On MEP (3.3), corresponding to the opening of protomer i, the barriers are

$$\beta(\boldsymbol{\theta}, \boldsymbol{\theta}') \sim k + \sum_{j/\theta_j=0} (c_1 A_{ij} + c_2 B_{ij} + c_3 C_{ij}) \\ + \sum_{j/\theta_j=0} (c_4 C_{ij} + c_5 \tilde{A}_{ij}). \tag{3.4}$$

4. Results

4.1. Transition pathways

We first address the issue of whether the expansion of the capsid is likely to occur either by a sequential process of successive openings of adjacent protomers, or by a disordered generalized opening of many or all protomers at different unrelated sites.

To assess the likely transition mechanism, we note that the barrier to opening a single protomer in the native configuration of the capsid according to (3.4) is

$$\beta(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) = k + 2c_1 + c_2 + 2c_3 + 2c_4 + c_5, \tag{4.1}$$

where $\boldsymbol{\theta}_0 = (0, 0, \ldots, 0)$ and, without loss of generality due to the symmetry of the capsid, $\boldsymbol{\theta}_1 = (1, 0, \ldots, 0)$. Now consider the second transition event: if the protomer that opens is not adjacent to protomer $i = 1$, the energy barrier is still (4.1); otherwise, if it is adjacent to 1, then the barrier is smaller than (4.1) (see below). Hence, the most probable transition path is a sequence of openings of adjacent protomers initiated at a single nucleation site, rather than a sequence of random openings at unrelated protomers.

Next, we study how the initial stages of the expansion depend on the parameters b, c_4, and c_5 in the mechanical constraint energy. Here we interpret b as playing the role of an elasticity modulus, with larger b meaning more rigid proteins, and c_4, c_5 indicating the strength of the mechanical constraints on proximal protomers.

Numerical simulations (cf. table 2) and the argument presented in appendix A show that the initial stages of the transition belong to a small number of different classes, depicted in figure 4 and labelled according to their positions relative to five-, three- and twofold axes of the particle:

— **Class** 5: After the opening of the first protomer, the transition proceeds by opening adjacent protomers in the same pentamer, either clockwise or anti-clockwise around the fivefold axis. This occurs when $c_5 = 0$ and c_4 is small, so that the mechanical constraint around the threefold axis is weak and there is no chiral constraint around the fivefold axes.
— **Class** 5^-: The transition proceeds by opening adjacent protomers belonging in the same pentamer, in counter-clockwise order. This occurs when the chiral intra-pentamer constant c_5 is large and much larger than c_4.
— **Class** 3: The transition proceeds by opening adjacent protomers in the same trimer, either clockwise or anti-clockwise around the threefold axis. This occurs when

the constraint around the threefold axis is strong, i.e. when the constant c_4 is large and much larger than c_5.

— **Classes** 2^-, 2^+: After the opening of the first protomer, the transition proceeds by opening its neighbour in the same pentamer, either in clockwise (+) or counter-clockwise (−) order, according to whether $c_5 = 0$ or not. The third protomer to open is then the one located across the twofold axis of the dimer, and this leads to the opening of a pore at this twofold axis. Using the values (2.13) for c_1, c_2, c_3, this happens when the constraints have comparable strength, i.e. when $0.21 < c_4 - c_5 < 0.49$.

We now show how the above classification can also be derived from (3.4). Assume, without loss of generality, that protomer 1 has opened, and consider now the second opening event. The protomers that are adjacent to $i = 1$ are 2, 5, 6, 10, 30 (figure 2a), and the corresponding energy barriers are

$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2}) = k + c_1 + c_2 + 2c_3 + 2c_4,$$
$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,5}) = k + c_1 + c_2 + 2c_3 + 2c_4 + c_5,$$
$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,6}) = \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,30}) = k + 2c_1 + c_2 + c_3 + c_4 + c_5$$
$$\text{and} \quad \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,10}) = k + 2c_1 + 2c_3 + 2c_4 + c_5,$$

where we have denoted by $\boldsymbol{\theta}_{a,b,\dots,c}$ the state in which $\theta_a = \theta_b = \dots = \theta_c = 1$, and $\theta_j = 0$ for all $j \neq a, b, \dots, c$. Therefore,

$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,5}) \Leftrightarrow 0 < c_5,$$
$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,6}) = \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,30}) \Leftrightarrow c_3 + c_4 < c_1 + c_5$$
$$\text{and} \quad \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,10}) \Leftrightarrow c_2 < c_1 + c_5$$
$$\text{(always true since } c_2 < c_1\text{).}$$

The first observation is that, when $c_5 > 0$, the opening of protomer 2, adjacent to 1 in the counter-clockwise direction, is preferred to the opening of the protomer 5, adjacent to 1 in the clockwise position. Hence, the classes 5 and 2^+ can only be realized when $c_5 = 0$. If also c_4 is small, the sequence of the openings is then determined just by the competition between the bond energies c_1, c_2 and c_3. Since $c_1 > c_2$, c_3 the energy decrease due to the disruption of the intra-pentamer interface 1–2 or 1–5 dominates and the barrier for the opening of 2 or 5 is lower than all other barriers.

However, when $c_5 > 0$, the intra-pentamer penalization term (2.11) is such that the opening of protomer 1 generates a push on the clockwise adjacent protomer 5 (when this is in the closed state) that costs energy. If protomer 5 is in the open state, there is no push and no corresponding energy penalization. Hence, assume that 1 is open. Then by the above argument the opening of protomer 2 would not generate a push on its clockwise neighbour 1 and would not require energy. On the other hand, the opening of 5 would generate a push on protomer 4 (figure 2a), which is closed, and this would require to overcome an energy barrier corresponding to c_5. Hence, even though the opening of the protomers generates a clockwise push, the sequence by which they open propagates counter-clockwise, as in rarefaction waves.

Actually, the opening of protomer 5 also should release the energy accumulated in the push from protomer 1 at its counter-clockwise side, so that the net gain of energy should be zero, as for the opening of protomer 2. However, by our assumptions on the form of the chiral constraint energies, the relaxation of the constraint is felt only when the opening of protomer 5 is almost complete, while the expenditure of energy due to the push on 4 operates already at the first stages of the opening,

so that it does represent a barrier to overcome to initiate the process (cf. expression (A 4)).

A stronger statement is that, if the steric constraint energy thresholds c_4 and c_5 are such that

$$c_4 - c_5 < c_1 - c_3 \quad \text{and} \quad c_5 > 0, \tag{4.2}$$

then, with highest probability, the second protomer to open is necessarily protomer 2 (recall that $c_3 < c_1$ by (2.13)).

Assume now that the second protomer that opens is protomer 2, and compute the barriers required to open any of the protomers adjacent to the pair 1–2, i.e. 3, 6, 10, 11, 15, 30. We get

$$\beta(\boldsymbol{\theta}_{1,2}, \boldsymbol{\theta}_{1,2,10}) = k + 2c_1 + c_3 + c_4 + c_5,$$
$$\beta(\boldsymbol{\theta}_{1,2}, \boldsymbol{\theta}_{1,2,3}) = k + c_1 + c_2 + 2c_3 + 2c_4,$$
$$\beta(\boldsymbol{\theta}_{1,2}, \boldsymbol{\theta}_{1,2,15}) = k + 2c_1 + 2c_3 + 2c_4 + c_5,$$
$$\text{and} \quad \beta(\boldsymbol{\theta}_{1,2}, \boldsymbol{\theta}_{1,2,11}) = \beta(\boldsymbol{\theta}_{1,2}, \boldsymbol{\theta}_{1,2,6}) = \beta(\boldsymbol{\theta}_{1,2}, \boldsymbol{\theta}_{1,2,30})$$
$$= k + 2c_1 + c_2 + c_3 + c_4 + c_5.$$

Therefore, we have first that always

$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,10}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,15}),$$
$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,10}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,11}) = \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,6}) = \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,30}),$$

and

$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,10}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,3}) \Leftrightarrow c_1 + c_5 < c_2 + c_3 + c_4,$$
$$\beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,3}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,11}) \Leftrightarrow c_3 + c_4 < c_1 + c_5$$
$$\text{and} \quad \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,3}) < \beta(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{1,2,15}) \Leftrightarrow c_2 < c_1 + c_5$$
$$\text{(always true since } c_2 < c_1\text{).}$$

Hence, if

$$0 < c_1 - c_2 - c_3 < c_4 - c_5 < c_1 - c_3, \tag{4.3}$$

the third protomer that opens is protomer 10, with highest probability, and we recover class 2^-, which requires the interplay of both mechanical constraints to be operative.

Analogously, always in the hypothesis that the second protomer to open is 2, the opening proceeds along the pentamer in a counter-clockwise direction and belongs to class 5^- when

$$c_4 - c_5 < c_1 - c_2 - c_3, \tag{4.4}$$

since in this case c_4 is too small to force the opening along a trimer.

Finally, a similar analysis shows that the transition proceeds by opening a full trimer, i.e. it belongs to class 3, if

$$c_1 - c_3 < c_4 - c_5 \quad \text{and} \quad c_1 + c_2 - c_3 < c_4. \tag{4.5}$$

Hence, as expected, the competition between the steric constraint energies determines the kinetics of the opening sequence of the protomers.

4.2. The pores at the twofold axes

As mentioned in §2, during the expansion a number of pores open in the capsid: the pores at the twofold axes seem to be the most important, because they are the widest, and allow both for the VP4 and the RNA to exit from the capsid. The opening of these pores is due to the clockwise rotation of two facing protomers, for instance those labelled a and m in figure 1a. In particular, the pore is formed when VP2a moves away from VP2m.

In our schematics in figure 4 showing the first three steps of the transition pathways, this only occurs at the

Table 2. Results of numerical simulations showing the initial modes of transition for $a = b = 100$, classified according to the classes in figure 4 for different choices of c_4 (rows) and c_5 (columns).

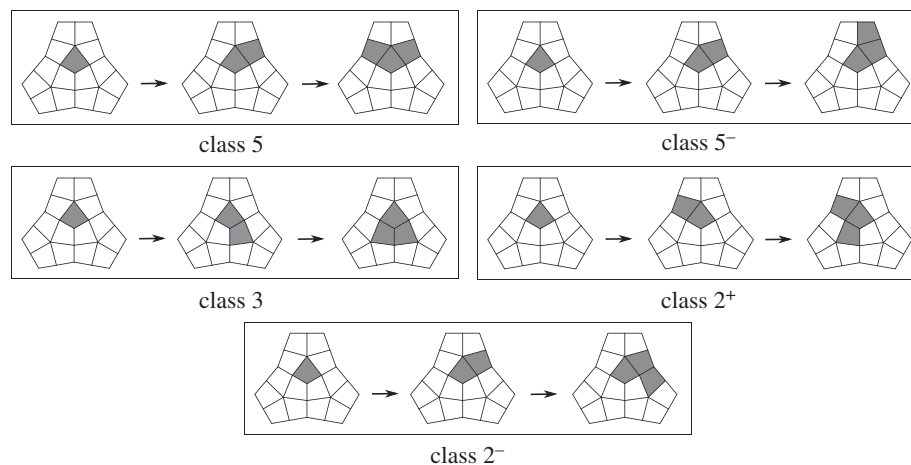| c_4 \ c_5	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
0	5	5	2+	2+,2 –	3,2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.1	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.2	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.3	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.4	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.5	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.6	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.7	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.8	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
0.9	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3	3	3
1	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3	3
1.1	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3	3	3
1.2	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3	3
1.3	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3	3	3
1.4	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3	3
1.5	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3	3
1.6	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3	3	3
1.7	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3	3
1.8	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	2 –	3
1.9	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –
2	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –	5 –

Figure 4. Classes of transition pathways. Different transition pathways are classified according to the positions of the second and the third protomer undergoing an opening event.

third stage of classes 2^+ and 2^-. In all other transition paths, this generally occurs at a later stage.

Furthermore, in classes 2^+ and 2^-, the position of the first pore to open is uniquely determined, while in the other expansion modes the location of the first twofold pore to open is random.

5. Discussion

Viral capsids are metastable structures—stable enough to provide protection for the viral genome, yet able to release their genomic cargo upon environmental cues. An important feature regulating this balance is the cooperative effects of multiple, relatively weak interactions between neighbouring coat proteins.

Previous studies demonstrate that the most energetically favourable mechanism for the global conformational change of the capsid during infection is a sequence of elementary conformational changes of single or small groups of protomers, that weaken the cooperative effect of the cohesive energies and destabilize adjacent protomers [28]. At each stage, the total energy decreases, as well as the threshold for each individual transition (figure 5). This mechanism has been experimentally observed in the maturation of HK97 in [26,27], and studied on simple capsid models in [17,28]. Each elementary change is viewed as a path between two energy minima on the energy landscape, and the precise sequence of openings can be determined by exploring the energy landscape via MEPs.

In this work, we have applied the above approach to HRV2, with the purpose of contributing some working hypotheses to the research on the infection mechanism of HRV2, and improve the understanding of the factors that promote the release of the genome in the host cell.

Specifically, for this type of virus, no receptor is involved in the externalization of the RNA and its inclusion into the host cell. Once the virus is internalized into the early endosome, the change of pH within the endosome induces the structural modifications leading to the expansion of the capsid described in §2. As discussed there, as a result of the expansion, pores open at twofold axes of the capsid, the N-terminal arm of VP1 is externalized and VP4 exits the capsid: the N-terminal arm of VP1 is hydrophobic and

is thought to be important in the adhesion of the capsid to the interior of the endosomal membrane.

Furthermore, the RNA exit from the capsid starts by extruding the 3′-end from a pore at one of the twofold axes [18,24,25].

Experimental evidence shows that the endosome is not disrupted during the exit of the RNA into the cell [36]. Hence, it is likely that the RNA exits from the endosome into the cell through pores in the endosomal membrane, and it seems that VP4 is instrumental in the formation of these pores.

A first hypothesis is that, in the absence of any regulatory mechanism, the capsid expands and all pores at the particle twofold axes open simultaneously. In this case, the 3′-end of the RNA is free to exit the capsid via any available pore (most likely the one which is nearest to its location in the interior of the capsid), but it faces the problem of exiting the endosome and entering the cell through one of the pores in the endosomal membrane. There are two possible ways to achieve this: either the RNA is extruded into the endosome and must then find a pore (which is an unlikely scenario as the genome would be degraded in the endosome), or the pore in the capsid is directly adjacent to a pore in the endosomal membrane. However, if the pore through which the RNA exits is not defined, any pores at any of the 30 particle twofold axes could contact the endosomal membrane, making this process inefficient and thus unlikely.

This suggests the following alternative. If the expansion is initiated at the site at which the 3′-end of the RNA is located, and if the expansion pathway is such that it preferentially opens the pore nearest to the 3′-end, then VP4 would be externalized at that capsid pore first, enabling it to open a pore in the endosomal membrane just facing this pore. This would then result in formation of a channel through which the RNA could exit both the capsid and the endosome and enter the cell. Then the expansion would proceed, but there would be no need for the opening of further pores in the endosomal membrane, and the problem of the localization and matching of the three players in this game (the 3′-end of the RNA, the pore in the capsid and the pore in the endosomal membrane) would be solved. Also, if the first twofold pore that opens during the transition pathway was uniquely determined in this way in proximity to the binding site at the 3′-end of the RNA, this would ensure that the RNA could exit

even if capsid expansion was stopped or slowed down after a few steps.

We have shown here that a simple coarse-grained model can account for initial pore opening immediately near the 3′-end of the RNA, and that this automatically occurs for every energy-decreasing transition pathway, if the constants c_4 and c_5 satisfy reasonable constraints, i.e. $c_4 \sim c_5$. This is highly likely, given that c_4 and c_5 are related to the mechanical response of two different interfaces of the same protein.

However, our coarse-grained model cannot explain how binding of the 3′-end of the RNA to a protomer may promote the initiation of the structural capsid transition at this protomer. There are precedent cases for RNA–CP contacts changing the conformation of capsomers and their ability to interact with other capsomers in the capsid shell, such as the allosteric conformers switch of the MS2 dimer triggered by contact with the RNA. Our model suggests that, similarly, the RNA–CP interactions at the protomer in contact with the 3′-end of the RNA might account for its role in initiating the transition pathway. However, this suggestion requires experimental validation.

In summary, our model is designed to demonstrate that it is possible, for a realistic range of the structural parameters describing the capsid, to envisage a simple mechanism by which the three fundamental players (the 3′-end of the RNA, the pore in the capsid and the pore in the endosomal membrane) act in a coordinated fashion without the need for further regulation or any indeterminacy to ensure an efficient and fast release of the genome into the host cell.

Appendix A. Metastable states, minimum energy paths and barriers

We provide here a justification for assertions (a), (b) and (c) in §3.2.

To prove (a), note that when $\theta_i \in \{0, 1\}$ for all i, each term in the total energy (2.12) either attains a local minimum or is locally constant. Therefore, since the minimum of a sum is greater than or equal to the sum of the minima, it follows that all $\boldsymbol{\theta} \in \{0, 1\}^{60}$ are minimizers. Other minimizers could in principle exist, but they cannot be found analytically. We have therefore performed a set of global numerical searches (trust-region algorithm with Matlab Global Optimization Toolbox), which showed no evidence of other local minima.

Assertion (b) has been also validated by numerical simulations, but an analytical proof can be sketched using an asymptotic expansion as $b \to \infty$ of the action functional based on the notion of Γ-convergence [37]. The argument is as follows: it is known that the MEP connecting two
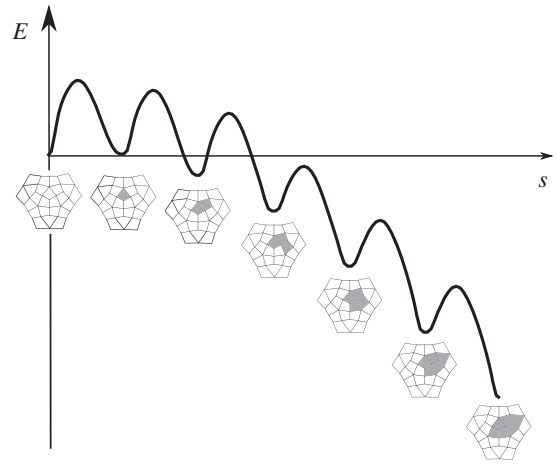

Figure 5. A typical energy cascade for the opening of the first 6 protomers in mode 2^-. The energy is plotted along a path in $\mathbb{R}^{60}$ that is the union of the first 6 MEPs, and s is the arc parameter along the path.

minima $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ is a minimizer of the geometric action functional (cf. [34])

$$
\begin{aligned}
J(\boldsymbol{\theta}) &= \int_0^1 \{|\nabla E(\boldsymbol{\theta}(s))||\boldsymbol{\theta}'(s)| + \nabla E(\boldsymbol{\theta}(s)) \cdot \boldsymbol{\theta}'(s)\} \, ds \\
&= \int_0^1 |\nabla E(\boldsymbol{\theta}(s))||\boldsymbol{\theta}'(s)| \, ds + E(\boldsymbol{\theta}(1)) - E(\boldsymbol{\theta}(0)),
\end{aligned}
\quad \text{(A 1)}
$$

among all smooth curves $\boldsymbol{\theta}(s)$ such that $\boldsymbol{\theta}(0) = \boldsymbol{\theta}_0$ and $\boldsymbol{\theta}(1) = \boldsymbol{\theta}_1$. Since the last term is independent of the path, the MEP is also a minimizer of the functional $I(\boldsymbol{\theta}) = \int_0^1 |\nabla E||\boldsymbol{\theta}'| \, ds$.

For simplicity, we write $b = n \in \mathbb{N}$, E_n in place of E, and split the energy as the sum of two terms, the first of which is independent of n:

$$
\begin{aligned}
E_n &= E_0 + c_4 E_{S3} + c_5 E_{S5}, \\
E_0 &= E_{\text{protomer}} + c_1 E_{\text{pentamer}} + c_2 E_{2\text{fold}} + c_3 E_{3\text{fold}}.
\end{aligned}
$$

Introducing the functionals

$$
I_n(\boldsymbol{\theta}) = \int_0^1 |\nabla E_n(\boldsymbol{\theta}(s))||\boldsymbol{\theta}'(s)| \, ds, \quad \tilde{I}_0(\boldsymbol{\theta}) = \int_0^1 |\nabla E_0(\boldsymbol{\theta}(s))||\boldsymbol{\theta}'(s)| \, ds
$$

and

$$
I_0(\boldsymbol{\theta}) = \tilde{I}_0(\boldsymbol{\theta}) + \Phi(\boldsymbol{\theta}(0), \boldsymbol{\theta}'(0)) + \Psi(\boldsymbol{\theta}(1), \boldsymbol{\theta}'(1)),
$$

with Φ and Ψ depending only on the boundary values of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, we first note that the MEP relative to $\tilde{I}_0$ is a straight line segment. In fact, let $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ be two metastable states such that $\theta_j^0 = \theta_j^1 \in \{0, 1\}$ if $j \neq i$ and $\theta_i^0 = 0$, $\theta_i^1 = 1$ for some protomer i. We claim that the minimizer $\boldsymbol{\theta}(s)$, $s \in [0, 1]$, of $\tilde{I}_0$ connecting $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}_1$ is

$$
\begin{aligned}
\theta_i(s) = s, \quad &\theta_j(s) \equiv \theta_j^0 = \theta_j^1 \in \{0, 1\} \\
&\text{for } j \neq i, \ s \in [0, 1].
\end{aligned}
\quad \text{(A 2)}
$$

To see this, we resort to the characterization of the MEPs in [34] as parametrized curves $\boldsymbol{\theta} = \boldsymbol{\theta}(s)$ with the property that $\nabla E_0(\boldsymbol{\theta}(s))$ is parallel to $\boldsymbol{\theta}'(s)$ for $s \in [0, 1]$. For $j \neq i$, we have that

$$
\frac{\partial E_0}{\partial \theta_j} = f'(\theta_j) + 2a \sum_k (c_1 A_{jk} + c_2 B_{jk} + c_3 C_{jk}) \theta_j \exp(-a(\theta_j^2 + \theta_k^2)),
$$

(A 3)

and since $f'(0) = f'(1) = 0$ and $\theta_j \exp(-a(\theta_j^2 + \theta_k^2)) \sim 0$ whenever $\theta_j \in \{0, 1\}$, this term vanishes on (A 2), which is therefore the MEP for $\tilde{I}_0$. This implies that also the

minimizers of I_0 are the straight-line segments (A 2): consider in fact a putative minimizer of I_0 that is not a straight line segment, and modify it on the interval $(\delta/2,\ 1-\delta/2)$ by a smooth curve that coincides with (A 2) on $(\delta,\ 1-\delta)$, so that the boundary values of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ do not change. Since (A 2) is a minimizer of $\tilde{I}_0$, as $\delta \to 0$ this would decrease I_0, which is a contradiction.

We now turn to the asymptotics of the action functional. We say that I_n Γ-converges to I_0 if (cf. [37])

g_1: for every converging sequence $\boldsymbol{\theta}^n \to \boldsymbol{\theta}$ (in a sense to be specified below) then $I_0(\boldsymbol{\theta}) \leq \liminf_n I_n(\boldsymbol{\theta}^n)$;

g_2: for every $\boldsymbol{\theta}$ there exists a sequence $\boldsymbol{\theta}^n \to \boldsymbol{\theta}$ such that $I_0(\boldsymbol{\theta}) = \lim_n I_n(\boldsymbol{\theta}^n)$.

This notion of convergence for the action functional has the advantage that, if a sequence of minimizers of I_n converges to some $\boldsymbol{\theta}$, then this is a minimizer of I_0 (cf. [37]). In terms of the MEPs, this means that we can approximate the MEP for large b in terms of the minimizer of the action functional I_0, which is a rectilinear path.

The choice of the topology that defines the convergence of the sequences $\boldsymbol{\theta}^n$ must guarantee that sequences of minimizers converge, and, at the same time, that g_1 and g_2 hold. To simplify matters, we choose here the $C^1([0, 1], \mathbb{R}^{60})$ norm, which is admittedly too strong to ensure the convergence of minimizers, but makes g_1 and g_2 simpler to prove. With some additional technical effort the result below can be extended to weak convergence in $W^{1,\infty}([0, 1], \mathbb{R}^{60})$, that guarantees the convergence of minimizers.

We sketch here a proof of the fact that I_n Γ-converges to I_0 for a simplified problem in two dimensions, with $\boldsymbol{\theta} = (\theta_1, \theta_2)$, assuming that the energy reduces to

$$E_n(\boldsymbol{\theta}) = f(\theta_1) + f(\theta_2) + K(1 - \exp(-n(\theta_1 - \theta_1)^2),$$

with $E_0(\boldsymbol{\theta}) = f(\theta_1) + f(\theta_2)$ and K a positive constant.

Consider now any sequence $\boldsymbol{\theta}^n$ such that $\boldsymbol{\theta}^n \to \boldsymbol{\theta}$ and $(\boldsymbol{\theta}^n)' \to \boldsymbol{\theta}'$ uniformly on $[0, 1]$, such that, to fix ideas, $\boldsymbol{\theta}_0 = (0, 0)$ and $\boldsymbol{\theta}_1 = (1, 0)$: expanding the square of the first factor in the action functional I_n and letting $\rho_n(s) = \theta_1^n(s) - \theta_2^n(s)$ and $R_n(s) = 2nK\rho_n(s) \exp(-n\rho_n^2(s))$, we find

$$|\nabla E_n(\boldsymbol{\theta}^n(s))|^2 = (f'(\theta_1^n(s)) + R_n(s))^2 + (f'(\theta_2^n(s)) - R_n(s))^2.$$

The idea is now to decompose the integration domain $[0, 1]$ into the union of $I_\delta^n = \{s \in [0, 1] : |\rho_n(s)| < \delta\}$ and its complement. On $[0, 1] \setminus I_\delta^n$ it straightforward to see that R_n converges uniformly to zero, while on I_δ^n the sequence R_n is not bounded in the sup-norm and dominates the terms

involving f'. Hence, I_n can be approximated by

$$I_n(\boldsymbol{\theta}^n) \sim \int_{I_\delta^n} \sqrt{2}|R_n(s)||(\boldsymbol{\theta}^n)'(s)|\,\mathrm{d}s + \int_{[0,1]\setminus I_\delta^n} |\nabla E_0(\boldsymbol{\theta}^n(s))||(\boldsymbol{\theta}^n)'(s)|\,\mathrm{d}s$$

for n large. Now, $\nabla E_0(\boldsymbol{\theta}^n(s))$ and $(\boldsymbol{\theta}^n)'(s)$ converge uniformly to $\nabla E_0(\boldsymbol{\theta}(s))$ and $\boldsymbol{\theta}'(s)$ on $[0, 1]$, and therefore, by the arbitrariness of δ, the second term above converges to $\tilde{I}_0(\boldsymbol{\theta})$. On the other hand, if we assume for simplicity that the only point at which $\rho_n(s)$ vanish is $s = 0$, on I_δ^n we have that $\rho_n(s) = \theta_1^n(s) - \theta_2^n(s) \sim (c_1^n - c_2^n)s$, with $c_i^n = (\theta_i^n)'(0) \to c_i = \theta_i'(0)$, so that the first term above becomes, if $c_1 \neq c_2$,

$$\int_{I_\delta^n} 2\sqrt{2}nK|c_1^n - c_2^n|s\ \mathrm{e}^{-n(c_1^n-c_2^n)^2s^2} \sqrt{(c_1^n)^2 + (c_2^n)^2}\,\mathrm{d}s$$

$$\sim K\frac{\sqrt{2((c_1^n)^2 + (c_2^n)^2)}}{|c_1^n - c_2^n|} \to K\frac{\sqrt{2(c_1^2 + c_2^2)}}{|c_1 - c_2|},$$

which is the function $\Phi(\boldsymbol{\theta}(0),\ \boldsymbol{\theta}'(0))$ in the expression for I_0.

Third, we prove (c). Given that, as $b = n \to \infty$ the MEP is approximately the straight line segment (A 2), we can compute the barriers by evaluating the energy on it:

$$E(t) = f(t) + \sum_j (c_1 A_{ij} + c_2 B_{ij} + c_3 C_{ij})(1 - \exp(-a(t^2 + \theta_j^2)))$$

$$+ \sum_j (c_4 C_{ij} + c_5 \tilde{A}_{ij} H(t - \theta_j) + c_5 \tilde{A}_{ji} H(\theta_j - t))$$

$$\times (1 - \exp(-b(t - \theta_j)^2)) + \mathrm{const.},$$

where H is the Heaviside step function. Therefore,

$$E(t) - E(0) = f(t) + \sum_{j/\theta_j=0} (c_1 A_{ij} + c_2 B_{ij} + c_3 C_{ij})(1 - \exp(-at^2))$$

$$+ \sum_{j/\theta_j=0} (c_4 C_{ij} + c_5 \tilde{A}_{ij})(1 - \exp(-bt^2))$$

$$- \sum_{j/\theta_j=1} (c_4 C_{ij} + c_5 \tilde{A}_{ji})(1 - \exp(-b(1-t)^2)).$$

$$(A\ 4)$$

The above function is the sum of f and three monotone functions, two increasing and one decreasing. For b and a large enough, E has a single maximum in $[0, 1]$, at the point at which f has its maximum, and the value of the maximum of E is exactly the sum of the maxima of its constituents, which yields the relation (3.4).

Finally, we list in table 2 the results of numerical simulations showing how the initial opening mode depends on the constants c_4 and c_5.

References

1. Zandi R, Reguera D, Bruinsma R, Gelbart W, Rudnick J. 2005 Assembly and disassembly of viral capsids. *J. Theor. Med.* **6/2**, 69–72. (doi:10.1080/10273 660500149166)

2. Tuthill TJ, Harlos K, Walter TS, Knowles NJ, Groppelli E, Rowlands DJ, Stuart DI, Fry EE. 2009 Equine rhinitis A virus and its low pH empty particle: clues towards an aphthovirus entry mechanism? *PLoS Pathog.* **5**, e100062. (doi:10.1371/journal.ppat. 1000620)

3. Bakker SE, Groppelli E, Pearson AR, Stockley PG, Rowlands DJ, Ranson NA. 2014 Limits of structural plasticity in a picornavirus capsid revealed by a massively expanded equine rhinitis A virus particle. *J. Virol.* **88**, 6093–6099. (doi:10.1128/ JVI.01979-13)

4. Speir J, Munshi S, Wang G, Baker TS, Johnson JE. 1995 Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy.

Structure **3**, 63–78. (doi:10.1016/S0969-2126(01)00135-6)

5. Zhao R, Pevear DC, Kremer MJ, Giranda VL, Kofron JA, Kuhn RJ, Rossmann MG. 1996 Human rhinovirus 3 at 3.0 A resolution. *Structure* **4**, 1205–1220. (doi:10.1016/S0969-2126(96)00128-1)

6. Verdaguer N, Blaas D, Fita I. 2000 Structure of human rhinovirus serotype 2 (HRV2). *J. Mol. Biol.* **300**, 1179–1194. (doi:10.1006/jmbi. 2000.3943)

7. Hewat EA, Neumann E, Blaas D. 2002 The concerted conformational changes during human rhinovirus 2 uncoating. *Mol. Cell* **10**, 317–326. (doi:10.1016/S1097-2765(02)00603-2)

8. Hewat EA, Blaas D. 2004 Cryoelectron microscopy analysis of the structural changes associated with human rhinovirus type 14 uncoating. *J. Virol.* **78**, 2935–2942. (doi:10.1128/JVI.78.6.2935-2942.2004)

9. Liu Y, Hill MG, Klose T, Chen Z, Watters K, Bochkov YA, Jiang W, Palmenberg AC, Rossmann MG. 2016 Atomic structure of a rhinovirus C, a virus species linked to severe childhood asthma. *Proc. Natl Acad. Sci. USA* **113**, 8997–9002. (doi:10.1073/pnas.1606595113)

10. Garriga D, Pickl-Herk A, Luque D, Wruss J, Castón JR, Blaas D, Verdaguer N. 2012 Insights into minor group rhinovirus uncoating: the X-ray structure of the HRV2 empty capsid. *PLoS Pathog.* **8**, e1002473. (doi:10.1371/journal.ppat.1002473)

11. Fuchs R, Blaas D. 2012 Productive entry pathways of human rhinoviruses. *Adv. Virol.* **2012**, 1–13. (doi:10.1155/2012/826301)

12. Pickl-Herk A, Luque D, Vives-Adrián L, Querol-Audí J, Garriga D, Trus BL, Verdaguer N, Blaas D, Castón JR. 2013 Uncoating of common cold virus is preceded by RNA switching as determined by X-ray and cryo-EM analyses of the subviral A-particle. *Proc. Natl Acad. Sci. USA* **110**, 20 063–20 068. (doi:10.1073/pnas.1312128110)

13. Strauss M, Levy HC, Bostina M, Filman DJ, Hogle JM. 2013 RNA transfer from poliovirus 135S particles across membranes is mediated by long umbilical connectors. *J. Virol.* **87**, 3903–3914. (doi:10.1128/JVI.03209-12)

14. Butan C, Filman DJ, Hogle JM. 2014 Cryo-electron microscopy reconstruction shows poliovirus 135S particles poised for membrane interaction and RNA release. *J. Virol.* **88**, 1758–1770. (doi:10.1128/JVI.01949-13)

15. Strauss M, Filman DJ, Belnap DM, Cheng N, Nuoel RT, Hogle JM. 2015 Nectin-like interactions between poliovirus and its receptor trigger conformational changes associated with cell entry. *J. Virol.* **89**, 4143–4157. (doi:10.1128/JVI.03101-14)

16. Lee H, Shingler KL, Organtini LJ, Ashley RE, Makhov AM, Conway JF, Hafenstein S. 2016 The novel asymmetric entry intermediate of a picornavirus captured with nanodiscs. *Sci. Adv.* **2**, e1501929. (doi:10.1126/sciadv.1501929)

17. Cermelli P, Indelicato G, Zappa E. 2017 Minimum energy paths for conformational changes of viral capsids. *Phys. Rev. E* **96**, 012407. (doi:10.1103/PhysRevE.96.012407)

18. Harutyunyan S, Kumar M, Sedivy A, Subirats X, Kowalski H, Köhler G, Blaas D. 2013 Viral uncoating is directional: exit of the genomic RNA in a common cold virus starts with the poly-(A) tail at the 39-end. *PLoS Pathog.* **9**, e1003270. (doi:10.1371/journal.ppat.1003270)

19. Harutyunyan S, Kowalski H, Blaas D. 2014 The rhinovirus subviral A-particle exposes 3′-terminal sequences of its genomic RNA. *J. Virol.* **88**, 6307–17. (doi:10.1128/JVI.00539-14)

20. Stockley PG, Rolfsson O, Thompson GS, Basnak G, Francese S, Stonehouse NJ, Homans SW, Ashcroft AE. 2007 A simple, RNA-mediated allosteric switch controls the pathway to formation of a $T = 3$ viral capsid. *J. Mol. Biol.* **369**, 541–552. (doi:10.1016/j.jmb.2007.03.020)

21. Dykeman EC, Stockley PG, Twarock R. 2010 Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J. Mol. Biol.* **395**, 916–923. (doi:10.1016/j.jmb.2009.11.016)

22. Twarock R, Leonov G, Stockley PG. 2018 Hamiltonian path analysis of viral genomes. *Nat. Commun.* **9**, 2021. (doi:10.1038/s41467-018-03713-y)

23. Shakeel S, Dykeman EC, White SJ, Ora A, Cockburn JJB, Butcher SJ, Stockley PG, Twarock R. 2017 Genomic RNA folding mediates assembly of human parechovirus. *Nat. Commun.* **8**, 5. (doi:10.1038/s41467-016-0011-z)

24. Levy HC, Bostina M, Filman DJ, Hogle JM. 2010 Catching a virus in the act of RNA release: a novel poliovirus uncoating intermediate characterized by cryo-electron microscopy. *J. Virol.* **84**, 4426–4441. (doi:10.1128/JVI.02393-09)

25. Bostina M, Levy H, Filman DJ, Hogle JM. 2010 Poliovirus RNA is released from the capsid near a twofold symmetry axis. *J. Virol.* **85**, 776–783. (doi:10.1128/JVI.00531-10)

26. Cardone G, Duda RL, Cheng N, You L, Conway JF, Hendrix RW, Steven AC. 2014 Metastable intermediates as stepping stones on the maturation pathways of viral capsids. *mBio* **5**, e02067-14. (doi:10.1128/mbio.02067-14)

27. Ross PD, Conway JF, Cheng N, Dierkes L, Firek BA, Hendrix RW, Steven AC, Duda RL. 2006 A free energy cascade with locks drives assembly and maturation of bacteriophage HK97 capsid. *J. Mol. Biol.* **364**, 512–525. (doi:10.1016/j.jmb.2006.08.048)

28. Cermelli P, Indelicato G, Twarock R. 2013 Nonicosahedral pathways for capsid expansion. *Phys. Rev. E* **88**, 032710. (doi:10.1103/PhysRevE.88.032710)

29. Pearce P, Woodhouse FG, Forrow A, Kelly A, Kusumaatmaja H, Dunkel J. 2018 Learning dynamical information from static protein and sequencing data. *bioRxiv* 401067. (doi:10.1101/401067)

30. Zandi R, van der Schoot P, Reguera D, Kegel W, Reiss H. 2006 Classical nucleation theory of virus capsids. *Biophys. J.* **90**, 1939–1948. (doi:10.1529/biophysj.105.072975)

31. Freidlin MI, Wentzell AD. 2012 *Random perturbations of dynamical systems*. Dordrecht, The Netherlands: Springer.

32. Weinan E, Ren W, Vanden Eijnden E. 2002 String method for the study of rare events. *Phys. Rev. B* **66**, 052301. (doi:10.1103/physrevb.66.052301)

33. Weinan E, Ren W, Vanden-Eijnden E. 2004 Minimum action method for the study of rare events. *Commun. Pure Appl. Math.* **57**, 637–656. (doi:10.1002/cpa.20005)

34. Heymann M, Vanden Eijnden E. 2008 Pathways of maximum likelihood for rare events in nonequilibrium systems: application to nucleation in the presence of shear. *Phys. Rev. Lett.* **100**, 140601. (doi:10.1103/PhysRevLett.100.140601)

35. Cameron MK. 2013 Computing Freidlin's cycles for the overdamped Langevin dynamics. Application to the Lennard-Jones-38 cluster. *J. Stat. Phys.* **152**, 493–518. (doi:10.1007/s10955-013-0770-4)

36. Ren J *et al.* 2013 Picornavirus uncoating intermediate captured in atomic detail. *Nat. Commun.* **4**, 1929. (doi:10.1038/ncomms2889)

37. Braides A. 2002 *Γ-convergence for beginners*. Oxford, UK: Oxford University Press.