doi:10.1016/j.jmb.2010.05.059 J. Mol. Biol. (2010) 401, 298–308

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Impact of Viral RNA on Assembly Pathway Selection

Victoria L. Morton[1]†, Eric C. Dykeman[2]†, Nicola J. Stonehouse[1], Alison E. Ashcroft[1], Reidun Twarock[2] and Peter G. Stockley[1]*

[1]*Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK*

[2]*York Centre for Complex Systems Analysis, University of York, York, YO10 5DD, UK*

Received 20 March 2010; received in revised form 21 May 2010; accepted 22 May 2010
Available online 1 June 2010

Many single-stranded RNA viruses self-assemble their protein containers around their genomes. The roles that the RNA plays in this assembly process have mostly been ignored, resulting in a protein-centric view of assembly that is unable to explain adequately the fidelity and speed of assembly in such viruses. Using bacteriophage MS2, we demonstrate here via a combination of mass spectrometry and kinetic modelling how viral RNA can bias assembly towards only a small number of the many possible assembly pathways, thus increasing assembly efficiency. Assembly reactions have been studied *in vitro* using phage coat protein dimers, the known building block of the $T=3$ shell, and short RNA stem–loops based on the translational operator of the replicase cistron, a 19 nt fragment (TR). Mass spectrometry has unambiguously identified two on-pathway intermediates in such reactions that have stoichiometry consistent with formation of either a particle 3-fold or 5-fold axis. These imply that there are at least two sub-pathways to the final capsid. The flux through each pathway is controlled by the length of the RNA stem–loop triggering the assembly reaction and this effect can be understood in structural terms. The kinetics of intermediate formation have been studied and show steady-state concentrations for intermediates between starting materials and the $T=3$ shell, consistent with an assembly process in which all the steps are in equilibrium. These data have been used to derive a kinetic model of the assembly reaction that in turn allows us to determine the dominant assembly pathways explicitly, and to estimate the effect of the RNA on the free energy of association between the assembling protein subunits. The results reveal that there are only a small number of dominant assembly pathways, which vary depending on the relative ratios of RNA and protein. These results suggest that the genomic RNA plays significant roles in defining the precise assembly sub-pathway followed to create the final capsid.

Keywords: virus assembly; mass spectrometry; kinetics; modelling; bacteriophage MS2

Edited by M. F. Summers

Introduction

Viruses are important examples of biological self--assembly. These assembly mechanisms have evolved to be extremely robust and efficient, allowing large numbers of progeny particles to emerge from infection of a host cell often by just a single virion. For most spherical viruses, the genomic nucleic acid is surrounded by protein containers (capsids). These typically follow icosahedral surface lattices and display the property of quasi-equivalence, in which coat proteins adopt symmetry-related conformations or positions that define the size and shape of the capsid.[1] This allows the virus to build a larger container from smaller subunits and therefore use less genetic material to encode this structure.[2] The precise molecular mechanism by which coat protein subunits adopt the correct quasi-conformer with high efficiency during assembly has been a puzzle since the first atomic resolution structure of a spherical virus was reported over 30 years ago.[3] Similarly, the detailed assembly pathway(s) taken towards the final capsid has not been described in molecular detail for any viral system. For single-stranded (ss) RNA viruses, which constitute a major class of all viruses,[4] the RNA plays several important roles during assembly.

*Corresponding author. E-mail address: stockley@bmb.leeds.ac.uk.

† V.L.M. and E.C.D. contributed equally to this work.

In the MS2 $T=3$ capsid (Fig. 1) 180 coat protein subunits pack as 90 non-covalent dimers, defining two quasi-equivalent conformers: an asymmetric (A/B) dimer in which the folds of the loops connecting the F and G β-strands in each subunit differ, and a symmetric (C/C) dimer in which the FG-loops have identical folds.[5,6] B subunit FG-loops surround the particle 5-fold axes and those in the A and C subunits interdigitate at the 3-fold axes. Thus, efficient formation of a capsid of the correct size and symmetry is determined by the accurate placement of these protein quasi-conformers. We have shown recently that the quasi-equivalent conformation of the capsomere, a coat protein dimer, can be switched by binding an RNA stem–loop.[7–9] Both a sequence-specific, tight-binding RNA stem–loop of just 19 nt from within the genome (TR) and a number of non-TR RNA stem–loops can act as the allosteric trigger, resulting in formation of the asymmetric, A/B-like structure (CP_2:TR) from the RNA-free symmetric structure (C/C-like).[7–9] The results imply that many different stem–loops within the genome could provide position-specific information for such switching events throughout assembly of the protein shell. This idea is consistent with the layer of RNA lying directly underneath the coat proteins seen in cryo-electron microscopy reconstructions of both the virion and capsids reassembled around sub-genomic RNA fragments.[10–12,30] The assembling capsid also appears to promote structural rearrangements within the genomic RNA, one consequence of which is to reduce its persistence length so that it fits within the confines of the particle. It appears that there is a mutually-induced conformational switching between the RNA and its coat protein.[7,8,30]

We show here that the roles of the RNA are even more intricate than previously appreciated. In addition to the conformational switching, RNA stem-loops can bias assembly of the MS2 capsid towards a limited number of sub-assembly pathways. This occurs via the influence of bound RNA on the association energies between protein building blocks, resulting in a dramatic reduction of the complexity of the assembly process. Combining mass spectrometry and kinetic modelling we have quantified how the viral RNA impacts assembly pathway selection, ensuring efficient formation of a protein container of the correct size and symmetry. In particular, via mass spectrometry we have characterised the dominant, on-pathway assembly intermediates and determined the kinetics of their formation. These data were used to develop a kinetic

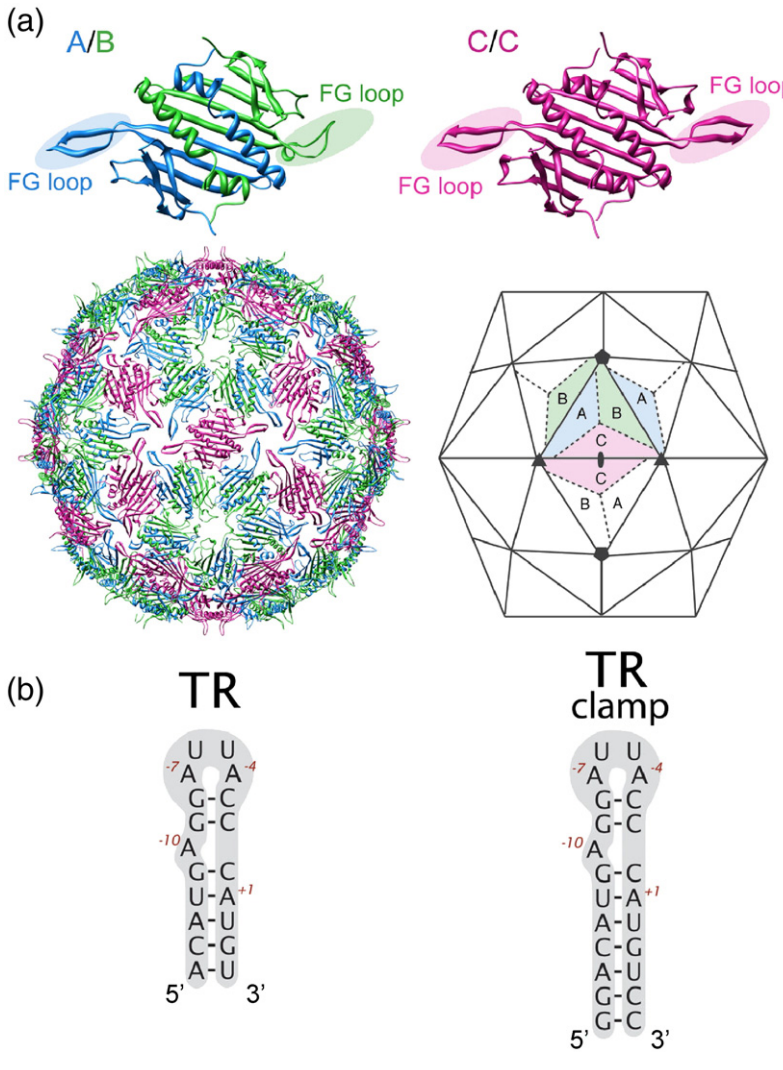

Fig. 1. Structures of the components used in MS2 bacteriophage assembly. (a) Ribbon models of the quasi-equivalent CP_2 conformers (A/B and C/C) and their relationship within the capsid (PDB 2MS2). (b) The sequence and secondary structure of the TR RNA translational operator and the TR clamp variant used here for assignment of peaks in mass spectra. The latter is based on the structure used by Grahn et al.[28]

model that accounts for the observed intermediate assembly kinetics and also predicts a number of features that could not be determined previously. These include details of how the RNA affects the association of coat protein dimers in the assembling viral capsid, as well as its influence on the nature of the most probable assembly pathways. These results imply a paradigm shift in our view of RNA virus assembly away from a protein-centric view to one that recognizes the important co-operative roles of the genomic RNA.

Results and Discussion

Identifying assembly intermediates and their kinetics of formation via mass spectrometry

Earlier, we used electrospray ionisation-mass spectrometry (ESI-MS) to study capsid assembly *in vitro* from mixtures of coat protein dimers (CP_2) and TR stem–loops.[9] The spectra showed a broad peak consistent with a fully completed $T=3$ shell, confirming the assembly-competence of the protein dimers and TR, and several lower order species that were assigned unambiguous stoichiometry. These included a species at 67.0 kDa consistent with [2(CP_2:TR)] and two significantly populated higher mass intermediates (Fig. 2). The first of these has a mass of 182.9 kDa consistent with a species with the stoichiometry of [3(CP_2:TR) + 3CP_2], i.e. six coat protein dimers, three with TR bound, termed the hexamer. This would correspond to a 3-fold axis from the $T=3$ capsid (Fig. 1a). The second intermediate had a mass of 290~300 kDa, and we were initially unable to assign its stoichiometry unambiguously due to its low-intensity broad peaks.

To confirm the stoichiometry of both of these high mass intermediates in the present study, we used both TR and a variant encompassing two additional base pairs at the bottom of its stem (TR clamp; Fig. 1b) to trigger assembly. This was combined with use of a newer instrument that allowed improved spectra to be obtained (Fig. 2b and d). The addition of extra bases increases the mass of the RNA from 6.0 kDa for TR to 7.3 kDa for TR clamp. Assembly reactions at a 2:1 ratio of CP_2:RNA(R) were monitored using TR or TR clamp and the spectra compared (Fig. 2b; Table 1). For [3(CP_2:R) + 3CP_2], the measured mass with TR was 182.9 kDa (expected 182.9 kDa) and that with TR clamp was 186.8 kDa (expected 186.8 kDa). This difference in mass of 3.9 kDa is consistent with the presence of three RNA molecules in each case, which corroborates its assignment as the hexameric species. For the larger intermediate formed with TR, the observed mass was 304.8 kDa. Considering all possible stoichiometries, this value corresponds best to a complex of ten coat protein dimers containing five RNAs, which is a species with a stoichiometry of [5(CP_2:R) + 5CP_2], with an expected mass of 304.8 kDa, termed the decamer.[13] In the TR clamp experiment, the mass measured was 312.1 kDa, a difference in measured mass between the TR and TR clamp of 7.3 kDa, ~0.8 kDa higher than expected (311.3 kDa) for the above stoichiometry. However, the suggested stoichiometry is still the best fit for this species because alternative stoichiometries have far larger mass errors; e.g. a stoichiometry of [6(CP_2:R) + 4CP_2], with an expected mass of 318.7 kDa for the TR clamp species, would have a very significant mass error (6.6 kDa; Table 1). The slight discrepancy between observed and expected values for the decamer formed with TR clamp, which is <3 % of the total mass, can easily be accounted for. The measured mass of non-covalent complexes is frequently higher than expected from the sum of their components, a phenomenon previously attributed to incomplete removal of buffer ions during the electrospray process (Fig. 2b).[14–17] Increasing the concentration of the coat protein in these reactions resulted in the intensities of these two high mass intermediates decreasing at the same rate, implying that they are both on-pathway to the $T=3$ capsid.

Interestingly, the relative amounts of these two intermediates are different for the TR and TR clamp reactions. The amount of the hexamer complex remains largely unchanged but the amount of decamer is decreased significantly in TR clamp reactions. This is not simply the result of differences in protein affinity between TR and TR clamp (Supplementary Data Fig. S1). Simple structural modelling of the minimal 5-fold complex ([5(CP_2:RNA)]), the pentamer of coat protein dimers that is part of the decamer, offers an explanation of this effect (Fig. 2b). The stems of the bound RNAs in the model all meet at the 5-fold axis, presumably leading to steric/repulsive clashes. Electrostatic effects could be exaggerated in the gas phase, which lacks screening counter ions. We are assuming that the species seen in the gas phase are representative of those that occur in solution but there is no easy way to prove that this is true. Whatever the origin of their relative stability the TR clamp decamer is less stable than its TR equivalent. Indeed, in other spectra with even longer versions of TR the amount of the decamer is reduced even further.[7] This would explain why the pentamer is never seen, but begs the obvious question of how the decamer assembles, which we address below.

Having assigned the stoichiometry of the higher order intermediates unambiguously, we studied the kinetics of their formation (Fig. 2c and d) using the procedure summarised in Materials and Methods. Assembly reactions were done at CP_2:TR ratios of 1:1, 2:1 and 3:1. Representative spectra of each reaction after 1 h are shown in Fig. 2d. The relative levels of the four dominant species were monitored; the initiation complex [CP_2:TR] of mass 33.5 kDa (black), the [2(CP_2:TR)] complex of mass 67.0 kDa (red), and the two intermediates discussed above, [3(CP_2:TR) + 3CP_2] of mass 182.9 kDa (blue) and [5(CP_2:TR) + 5CP_2] of mass 304.8 kDa (green). We have not seen species that would correspond to higher multiples of the initiation complex other than

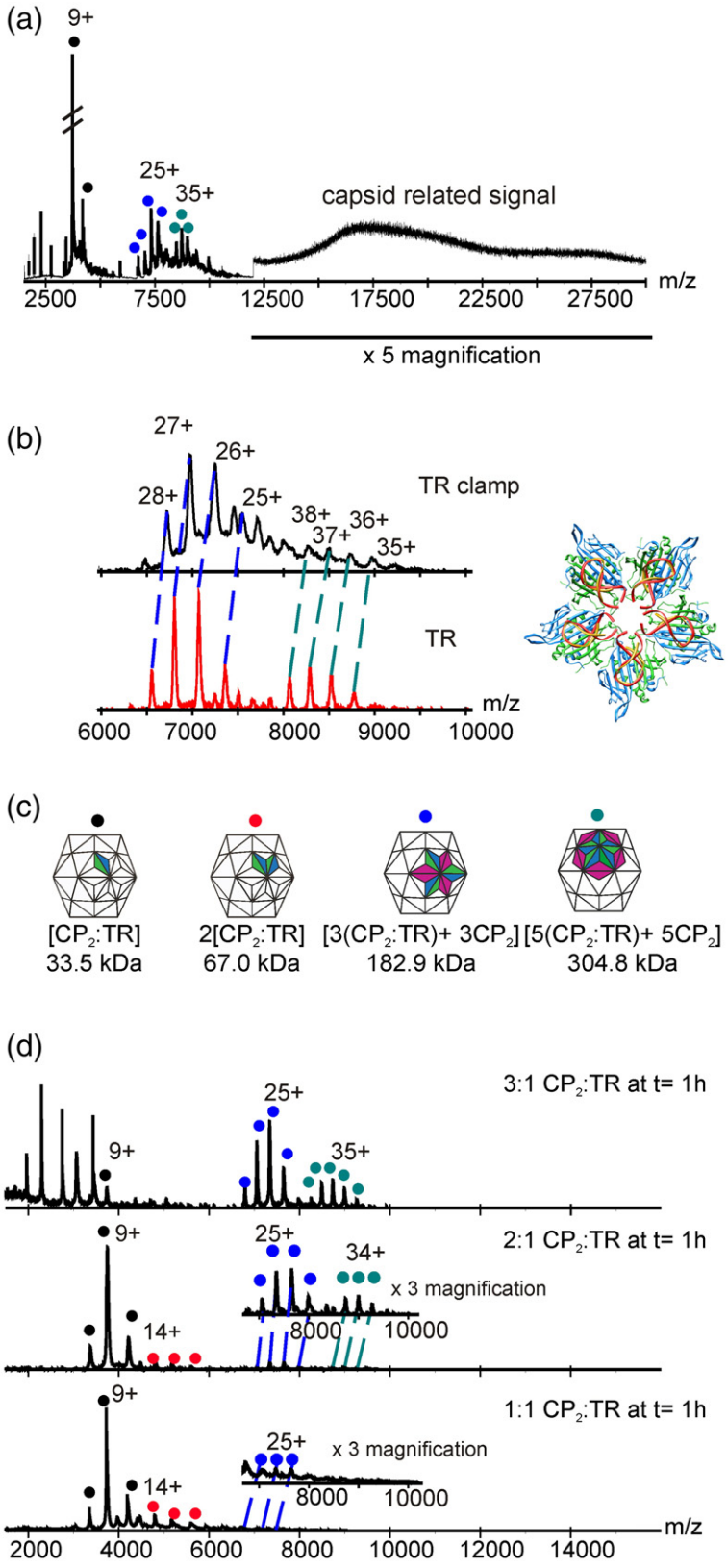

Fig. 2. Mass spectrometry of assembly reactions. (a) Mass spectrum of a TR-mediated assembly reaction in the CP_2:TR ratio 2:1 after 30 min. The observed intermediates are [CP_2:TR] (black dots), [3(CP_2:TR)+3CP_2] (blue dots) and [5(CP_2:TR)+5CP_2] (green dots); the capsid-related signal is evident. (b) Confirmation of the stoichiometry of the decameric intermediate. Assembly reaction of CP_2 with TR clamp (top) and with TR (bottom) in a 2:1 ratio at $t = 120$ min. The charge state distributions corresponding to the intermediates are indicated for [3(CP_2:TR)+3CP_2] and for [5(CP_2:TR)+5CP_2]. The right-hand panel shows a model of RNA stem–loops binding to all dimers adjacent to a particle 5-fold axis. The RNA stem–loop shown in yellow represents TR (1AQ3 [28]) and that shown in red is a 23 nt stem–loop (1TXS, truncated to the correct length[29]), which is representative of the TR clamp and has been manually positioned over the TR stem–loop. (c) The proposed arrangement of the CP dimer building blocks in each dominant intermediate based on the assigned mass and its location in the final capsid. (d) Monitoring intermediates of virus assembly of TR-mediated assembly reactions over time at ratios of 1:1, 2:1 and 3:1 of CP_2:TR. Mass spectra shown are at $t = 1$ h. The observed intermediates are labelled as in c.

$n = 2$, i.e. [2(CP_2:TR)], and the pentamer has not been detected in any spectra. The reproducibility of the spectra used for the kinetic studies for the 1:1 and 2:1 reactions was very high with SE±6%. This was not true for the 3:1 reaction, which is significantly more variable (Supplementary Data Fig. S2). This might reflect the fact that the dissociated CP_2 starting material is in 20 mM acetic acid, so that as the protein concentrations increase the pH of the reaction decreases. This is consistent with the increased amounts of CP monomer seen in these reactions (Fig. 2d). In addition, at assembly ratios of 3:1 CP_2:TR, the

Table 1. The measured and calculated masses for the [3(CP_2:RNA)+3CP_2] and [5(CP_2:RNA)+5CP_2] intermediates assembled with either TR or TR clamp

Stoichiometry	Measured mass (kDa)	Expected mass (kDa)
[3(CP_2:TR)+3CP_2]	182.9	182.9
[3(CP_2:TR clamp)+3CP_2]	186.8	186.8
[5(CP_2:TR)+5CP_2]	304.8	304.8
[5(CP_2:TR clamp)+5CP_2]	312.1	311.3
[6(CP_2:TR clamp)+4CP_2]	312.1	318.7

The calculated mass for an alternative intermediate [6(CP_2:TR clamp)+4CP_2] rules out this option as a possible stoichiometry.

capsid forms rapidly and a capsid-related signal is evident at $t=10$ min. All these factors contribute to variation in ionisation efficiency (±15%) that could make the data less reproducible.[18] In the 1:1 and 2:1 reactions, the initiation complex dominates and the higher order intermediates appear at much lower levels (Fig. 3). For the 3:1 reaction, however, the intermediates form rapidly and, in most spectra, appear at least as abundant as the CP_2:TR complex (Supplementary Data Fig. S2). The intermediate concentrations for the 1:1 and 2:1 reactions were therefore normalised with respect to the amount of initiation complex (100%) and the results are shown in Fig. 3a and b. Also shown is the 2:1 result with TR clamp (Fig. 3c). For both RNAs, and for these reaction stoichiometries, the relative concentrations of each of these species remained fairly constant from 10 min to over 3 h. Their relative abundance was also similar, with the amounts being:

$$[CP_2{:}TR] \gg [2(CP_2{:}TR)] > [3(CP_2{:}TR) + 3CP_2] > [5(CP_2{:}TR) + 5CP_2]$$

The amount of the capsid-related peak increased during these reactions, confirming that there was net flux through the pathway in favour of assembly and that the reaction was not at equilibrium. Over the course of the experiment, however, the intermediates appeared to have reached approximate steady-state concentrations; i.e. all the intermediate species smaller than the $T=3$ final product were in pseudo-equilibrium.

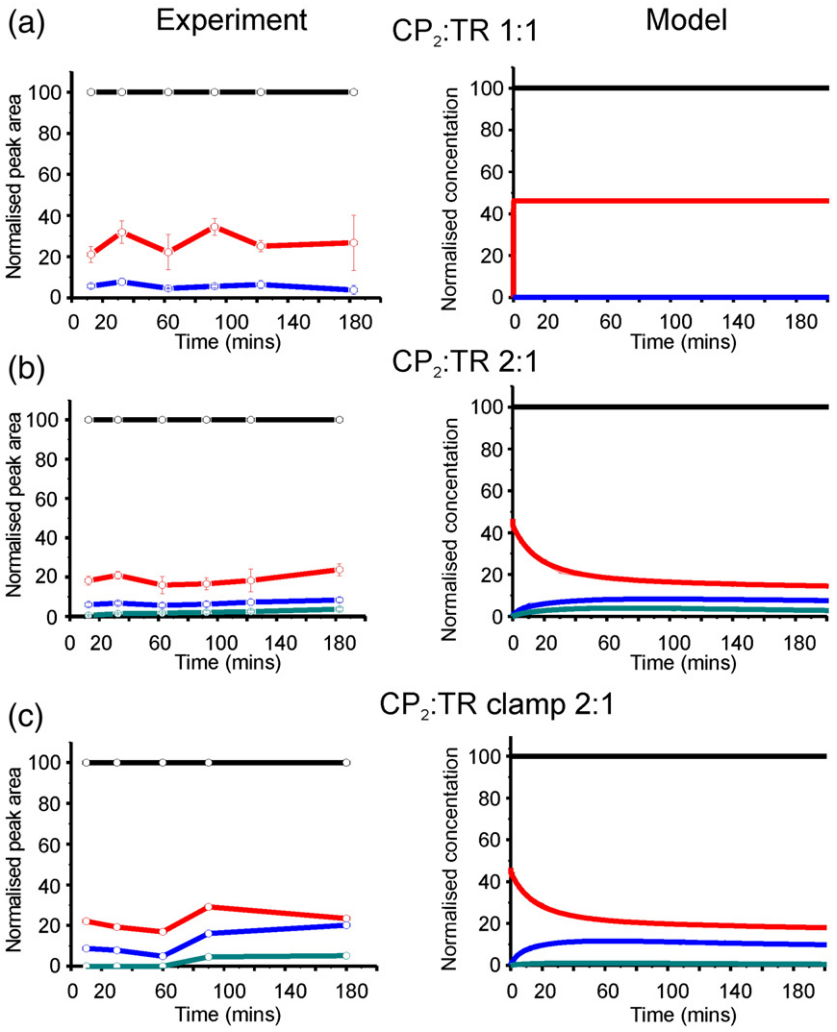

Fig. 3. Kinetics of virus assembly. Experiment (left-hand panels) and kinetic modelling (right-hand panels) of MS2 capsid assembly in the presence of TR stem–loops. In the experiment, intermediates were monitored over time using mass spectrometry at different initial starting ratios of CP_2:TR (a, 1:1; b, 2:1). The dominant intermediates are shown, and correspond to [CP_2:TR] (black); [2(CP_2:TR)] (red); [3(CP_2:TR)+3CP_2] (blue); [5(CP_2:TR)+5CP_2] (green). Experimental plots show the ratio of intermediates relative to [CP_2:TR], assuming similar ionisation efficiencies between the different intermediates. Concentrations of intermediates in the modelled data were normalized to [CP_2:TR] to facilitate comparison with the experimental data. (c) Experimental 2:1 assembly reaction of CP_2:TR clamp and modelling of the TR clamp experiment. The stronger repulsion of the TR clamp stem–loops around the 5-fold axis of the capsid is due to the presence of two extra base pairs, and is modelled by increasing ΔG_α by 0.2 kcal mol to −0.6 kcal mol. Experimental analysis of triplicate 1:1 and 2:1 reassembly reactions showed a standard deviation across the data set of less than 6%, demonstrating the high reproducibility of these experiments. The 3:1 data, however, showed significantly more variability, with a higher standard deviation of ~20% (see Supplementary Data). Error bars shown on plots represent the standard error of the mean.

Modelling the assembly kinetics

Previously, others have described kinetic models for viral capsid assembly that do not include the influence of the viral RNA being packaged.[19,20] The MS2 system is an ideal system in which to study such effects.[7–9,30] In particular, the mass spectrometry data on the TR-triggered assembly reactions allow the influence of viral RNA to be modelled for the first time. The modelling provides information that is difficult to obtain experimentally, such as the impact of RNA binding on coat protein association energies and predictions about detailed assembly pathways.

Mass spectrometry has shown that the unit of capsid growth is a coat protein dimer,[9] so in the model the two building blocks present in the fully assembled capsid (A/B and C/C) are used to form the capsid. Our model allows only RNA-free C/C-like dimers and TR stem–loops to react to form the A/B-like CP_2:TR building block (Fig. 1). Although A/A-like dimers might exist in solution, they have a conformation very similar to that of the C/C-like dimer and are therefore not treated as a separate species. In addition to the A/B and C/C quasi-conformers, we included a third assembly-incompetent species, B/B, in the model. This species is inferred to exist from the results of stopped-flow kinetic measurements of TR binding to coat protein dimers.[21] The B/B dimer is believed to be assembly-incompetent due to its inability to bind TR because of a steric clash that would occur between RNA and the polypeptide backbone in the *B* conformation at residue Ala80.[5] Since the other five dimer combinations of the A, B and C subunits act essentially as A/B and C/C dimers in their ability to bind TR, we assume the B/B dimer to occur in a 1:5 ratio with respect to the A/B- and C/C-like dimers. We assume also that B/B dimers can interconvert into an assembly-competent C/C dimer, which is consistent with experiment.[21] Note also that in higher order intermediates TR might be able to bind to C/C dimers that are "locked" in their conformation by contacts with neighbouring dimers, accounting for the eventual disappearance of the kinetic trap seen when TR saturates CP_2.[7] We have not considered this pathway here.

Experimental evidence indicates that assembly takes place by association of a single dimeric building block with a growing shell per reaction;[9] concatenations of higher order intermediates with multiple A/B and/or C/C dimers are therefore not permitted in the model. Thus, a higher order intermediate with n dimers, denoted as I_n, reacts with either a C/C dimer or a CP_2:TR complex (i.e. a dimer in A/B conformation in complex with TR) to form an intermediate with $n+1$ dimers, I_{n+1}. The assembly process can be described by the following assembly reactions:

$$\begin{array}{ccc} \mathrm{B/B} & \rightleftharpoons & \mathrm{C/C} \\ \mathrm{C/C + TR} & \rightleftharpoons & \mathrm{CP_2{:}\,TR} \\ I_n + \mathrm{D} & \rightleftharpoons & I_{n+1} \end{array} \qquad (1)$$

where D is either C/C or CP_2:TR.

We deduced a set of differential equations describing the time evolution of the concentrations of the assembly intermediates. Using the conditions of the experiments described above, we chose a CP_2 to TR ratio of either 1:1 or 2:1 as starting concentrations and solved the system numerically. This allowed us to determine the expected kinetic concentration profiles of the assembly intermediates. Note that the concentration profiles depend on the forward and backward rates of the assembly reactions listed in Eq. (1). The forward rate for conversion of a B/B dimer to a C/C dimer is taken from previous stopped-flow kinetic measurements, $k_f = 0.23\ s^{-1}$,[21] and the backward rate was inferred as described in Materials and Methods. Similarly, both the free energy of activation for TR binding ($\Delta G_{TR}^{++} = -4.3$ kcal/mol) and reaction rates are known for the second reaction (i.e. for TR binding to C/C) from experiment;[21] specifically $k_f = 1.98 \times 10^9\ mol^{-1}\ s^{-1}$ and $k_b = 2.2\ s^{-1}$. However, the rates for the third reaction in Eq. (1) and, in particular, the association free energies of a CP_2:TR complex with either a C/C dimer or another CP_2:TR complex, are not known from experiment. Current computational methods are not able to determine accurate values for such free energies.[22] Endres et al. have estimated the protein–protein association energies in viral capsids to be approximately −3.4 kcal mol^{-1} using a combination of experiment and assembly kinetic models.[19] However, the effect that RNA binding would have on association energies is not known. We therefore treated the free energy of association, ΔG, for CP_2:TR binding to CP_2:TR (ΔG_α) or to C/C (ΔG_β), as two free parameters and determined the values for which modelling is in good agreement with experiment.

Fine-tuning the assembly models

We used the mass spectrometry data on the assembly kinetics in Fig. 2 to establish values for ΔG_α and ΔG_β. Interestingly, we found that if all the species above the CP_2:TR complex are considered to be assembly competent, there is no value of the association energies for which model and experiment agree (Supplementary Data Fig. S3). The paradox in the modelling is the observed concentration of [2(CP_2:TR)] (~40% of the normalised [CP_2:TR] in 1:1 and ~20% in 2:1 reactions, respectively; Fig. 3a and b), because it does not seem to contribute to the assembly of larger intermediates by addition of another coat protein dimer or CP_2:TR complex. Its observed concentrations cannot be reconciled with the relative abundances of the [3(CP_2:TR) + 3CP_2] and [5(CP_2:TR) + 5CP_2] intermediates. This is primarily because free C/C-like dimers would be expected to bind to the [2(CP_2:TR)] intermediates preferentially, due to the potential double contact that can be formed ($2\Delta G_\beta$), leading to an assembly pathway dominated by the decamer rather than the hexamer (Fig. 4). This effect would strongly bias the [2(CP_2:TR)] species, as well as a number of other two-dimer intermediates, and favour assembly along the

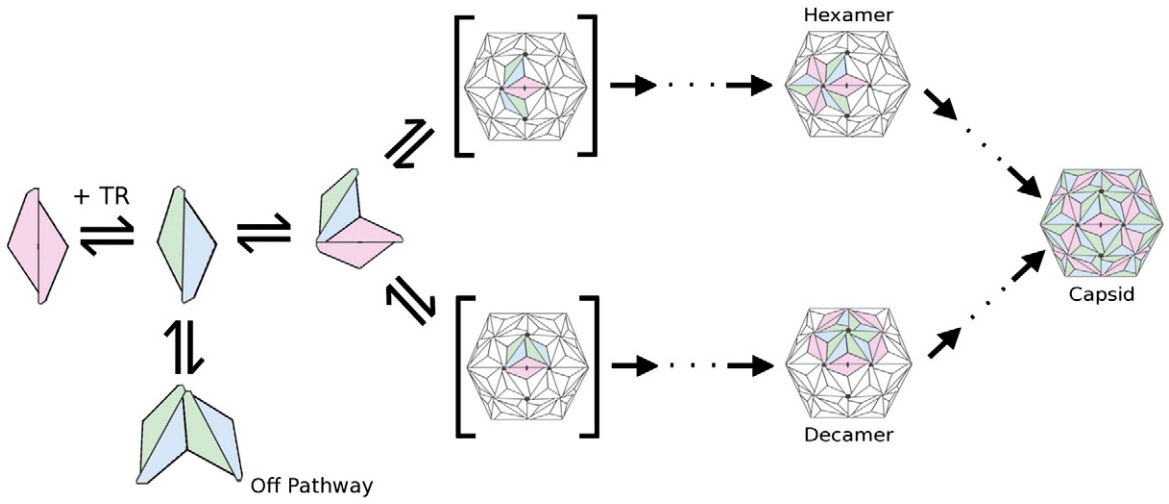

Fig. 4. MS2 assembly model. The figure shows a cartoon illustrating the presumed assembly pathways revealed by mass spectrometry and kinetic modelling. Species not observed directly are shown in brackets. The existence of the off-pathway species has been concluded as a consequence of the high concentration of [2(CP_2:TR)].

decamer pathway, producing relatively high concentrations of [5(CP_2:TR) + 5(CP_2)] with essentially no [3(CP_2:TR) + 3(CP_2)]. This result is opposite to that obtained in experiment (Fig. 3). The resolution of this paradox requires us to add an additional assumption to our model; namely, that the [2(CP_2:TR)] species is not assembly-competent but instead constitutes a dead-end species in these reactions that has to dissociate into two CP_2:TR species in order to fuel further capsid formation (Fig. 4). Indeed, a Brownian dynamics simulation of MS2 dimer interactions in the presence of TR has shown the existence of an encounter complex formed from two CP_2:TR that has a lower energy than the native state, but does not form an assembly-competent dimer.[31] In molecular terms, this effect might indicate that an isolated [2(CP_2:TR)] species is dynamic and that the binding site for an incoming C/C dimer is poorly formed.

Whatever the basis for the behaviour of [2(CP_2:TR)], modelling this species as a trap immediately allows unique, defined values for ΔG_α and ΔG_β to be determined that yield kinetic profiles in excellent agreement with experiment. In particular, for free energies of association of $\Delta G_\alpha = -0.86$ kcal mol^{-1} (CP_2:TR with CP_2:TR) and $\Delta G_\beta = -5.72$ kcal mol^{-1} (CP_2:TR with C/C dimer), the model predicts the occurrence of decameric and hexameric intermediates and their relative concentrations with respect to the initiation complex and the kinetic trap for both 1:1 and 2:1 reactions (Fig. 3a and b). The large difference in the bond energies ΔG_α and ΔG_β implied by these results can be understood in terms of the role of the TR stem–loop. Electrostatic interactions between the two negatively charged TR stem–loops at the CP_2:TR–CP_2:TR interface reduce the overall free energy of these contacts (ΔG_α), whilst having a positive effect on the binding energy between CP_2:TR and C/C (ΔG_β). Both these values are significantly different from those derived from models that take account of only the binding free energy between protein building blocks.[23] Due to the similarity of the dimer interfaces, we can assume that the energetic contributions of protein–protein interaction $\Delta G_{protein/protein}$ is approximately identical at these interfaces. Based on a value of $\Delta G_{protein/protein}$ of –3.4 kcal mol^{-1}, (chosen as an estimate following Ref.19) we can compute the impact of the RNA on the free energies of association:

$$\begin{aligned} \Delta G_\alpha &\approx \Delta G_{protein\,/\,protein} + \Delta G_{RNA\,/\,RNA} \\ \Delta G_\beta &\approx \Delta G_{protein\,/\,protein} + \Delta G_{RNA\,/\,protein} \end{aligned} \quad (2)$$

where $G_{RNA/RNA}$ and $\Delta G_{RNA/protein}$ are energy terms related to the repulsive forces between the two RNA stem–loops at a CP_2:TR–CP_2:TR interface, and the attractive forces between the RNA in the CP_2:TR complex and a neighbouring RNA-free C/C dimer, respectively. We obtain values of:

$$\begin{aligned} \Delta G_{RNA\,/\,RNA} &\approx 2.5\ \text{kcal mol}^{-1} \\ \Delta G_{RNA\,/\,protein} &\approx -2.3\ \text{kcal mol}^{-1} \end{aligned}$$

which reveal the strong influences of the RNA on the overall association energies.

Modelling the effect of TR clamp

The mass spectra of assembly reactions triggered by TR or TR clamp suggest (Fig. 2b) that the hexameric intermediate forms equally well with both types of RNA. The signals for the TR-containing decameric intermediate are also strong but significantly weaker with TR clamp. This effect is even more marked with longer RNAs,[7] and is captured in our kinetic model. Increasing ΔG_α by only 0.2 kcal mol^{-1} (0.05 kcal mol^{-1}/additional nucleotide) to account for an additional repulsive impact on the free energy, leads to a prediction of a significant drop in the concentration of the decamer species (Fig. 3c). This result suggests that the model accounts for the essential features of the effects RNA has on the assembly process. The RNA controls both the

relative concentrations of A/B and C/C dimers as well as the interaction energies between them, and in this way biases assembly towards pathways favouring the hexameric and decameric species built around 3-fold and 5-fold axes, respectively. Moreover, larger steric hindrance with increasing RNA size biases assembly towards the hexamer pathway for genomic length RNA.

Identifying the most probable assembly pathways

Earlier, we argued that the presence of a hexamer species corresponding to a 3-fold axis and the absence in any spectra of an equivalent pentamer species representing a 5-fold axis ([5(CP_2:TR)]) guaranteed the assembly of a capsid with $T=3$ quasi-symmetry.[9] The identification of a species with apparent 5-fold symmetry ([5(CP_2:TR) + 5CP_2]) appears at variance with this argument. There is still, however, no sign of ions corresponding to a pentameric complex that we can now be confident should ionise successfully under these conditions, given that the larger decamer is detected. This raises two questions: why does the pentamer species not form? and how does the decameric intermediate assemble in its absence? The kinetic assembly model provides answers to both questions. We computed the concentrations of all theoretically possible intermediates on the pathway to the decamer in the 1:1, 2:1 and 3:1 reactions. All of these intermediates occur at much smaller concentrations than the decamer itself, and would therefore be difficult to detect experimentally. However, theory shows a clear difference in their relative probabilities of occurrence, which allows us to predict the putative assembly pathways to the decameric species (Fig. 5). Inspection of the pathways in Fig. 5 shows assembly taking place via a series of A/B-like to C/C-like coat protein dimer interactions. The two types of dimer are attached to the growing capsid in an alternating fashion (apart from the second to last step in Fig. 5, where a second CP_2:TR complex is attached before completion with a C/C dimer). This suggests that the additional free energy of interaction with the outer layer of C/C dimers in the decamer compensates for the destabilising effects of TR binding on each A/B dimer of a fully formed 5-fold axis. This explains why the extended decameric complex but not the pentameric species is observed. Since the decamer is on the pathway to $T=3$ capsid formation, this result is consistent with observation.

The conclusion that assembly of capsid protein dimers in the presence of TR favours alternate association of CP_2:TR with C/C over CP_2:TR self-association is consistent with Brownian dynamics simulations of association rates of MS2 dimers in the absence and in the presence of TR stem–loops.[31] In addition, our kinetic analysis quantifies the dependence of the dominant assembly pathways on the relative ratio of TR to CP_2 and reveals how the alternating mode of assembly is realised for different relative concentration scenarios. The kinetic modelling also provides an explanation for why the hexamer and decamer species are dominant. The presence of RNA enhances the association energy of CP_2:TR with a C/C-like dimer, whilst weakening CP_2:TR self-association, making the formation of CP_2:TR-C/C contacts the preferred mode of assembly. Since both hexamer and decamer species consist of an equal number of A/B:TR and C/C-like dimers, both can be assembled via pathways of alternating A/B-like to C/C-like dimer additions, allowing formation of the preferred contact at each step. This is a simple and effective way for the particle to reduce the complexity of capsid formation to a small number of dominant pathways and avoid aberrant assemblies of smaller protein containers.

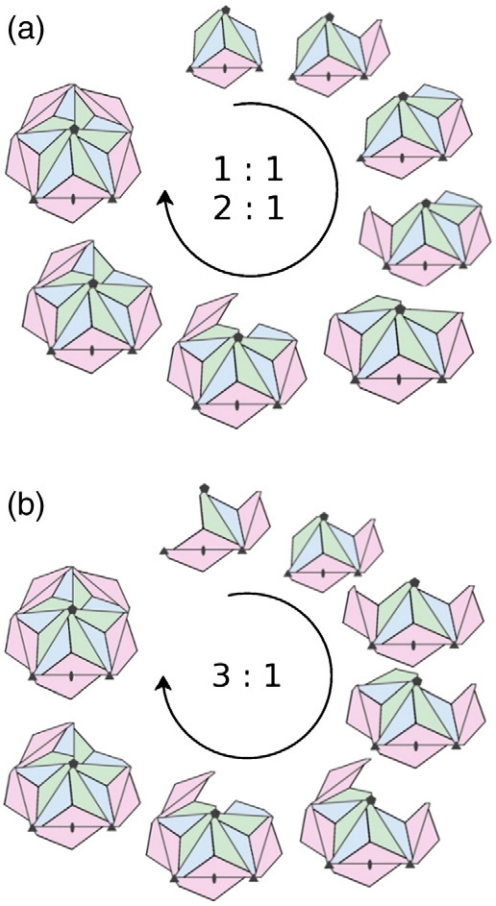

Fig. 5. Assembly of the decamer. A cartoon of the most probable pathways towards the decamer derived from the kinetic model for assembly at CP_2:TR ratios of 1:1 and 2:1 (a) or 3:1 (b). Note the prediction in each case of a species with the same stoichiometry as the hexamer intermediate, although this species, along with the other intermediates shown here, is predicted to occur in much smaller amounts than the highly populated hexamer and decamer (see Results and Discussion).

Virus assembly pathways involve conserved, exposed protein surfaces and should therefore be excellent targets for anti-viral drugs. However, no such clinically approved drug has been identified to date. The work described above with MS2 provides one explanation for this. Assembly of this ssRNA virus is not driven simply by protein–protein

interactions, but is instead a highly co-operative process in which the genome being packaged plays a very active role. This does not mean that the coat proteins are unable to assemble in the absence of RNA, or around non-cognate RNAs or other poly-anionic molecules. What it does imply is that the fidelity and efficiency of assembly of the cognate components has evolved to be extremely high. The inherent cooperativity between protein and genome in the MS2 system allows the RNA to both regulate the attachment of appropriate quasi-conformers in the growing protein shell, and determine the preferred sub-pathways followed towards the fully formed $T=3$ capsid. A drastic reduction of the complexity of the allowed assembly pathways is a major advantage for the virus, and is part of a much wider phenomenon controlling the assembly of many molecular machines.[24] Here, we have been able both to observe the consequences of such cooperativity directly in assembly intermediates and to model its effects. In particular, we have identified two dominant assembly pathways, one through a hexameric intermediate forming a particle 3-fold axis and the other through a decameric species comprising an extended 5-fold axis, that can both be extended into a fully formed $T=3$ capsid. Experiment confirms that the RNA–protein interactions bias assembly away from creation of a simple pentamer that could assemble erroneously into a $T=1$ shell. Modelling allows these effects to be quantified for the first time. It also allows us to predict assembly pathways and the relative concentrations of all the intermediates on each pathway. For instance, the model explains how the decamer can form even though the pentamer does not. It also adds to the interpretation of the experimental data by showing that although decamer assembly goes via asymmetric intermediates with the stoichiometry of hexamers, the concentration of such species is vanishingly small ($\sim$1000:1) compared to the significantly populated 3-fold intermediate.

When the phage packages its 3569 nt long genome, the relative stabilities of pentameric, hexameric and decameric species will obviously help to determine the most efficient assembly pathway. This will also impose some constraints on the RNA–coat protein interactions that are formed. We have shown that short genomic extensions of the TR stem–loop increase the rate and yield of $T=3$ capsid formation,[7] and that these effects are sequence-dependent. We have shown also that assembly efficiency is inversely proportional to RNA fragment length with RNAs$>$1000 nt long, and that there is good evidence that these RNAs participate in coat protein conformer switching.[30] It is clear from cryo-EM maps that the RNA cannot participate in these multiple contacts with its coat protein shell by forming only decameric or hexameric species, so some hybrid of the two pathways established here for assembly with short stem–loops must be followed in assembly of wild type phage.

The assembly cooperativity revealed by these results has clear implications for the design of anti-viral drugs directed at assembly pathways. Ablating protein–protein interactions alone may not be enough to overcome the formative influence of the genomic RNA on the assembly pathway. As we have shown, multiple RNA–protein binding events can participate in viral assembly in MS2 by acting as allosteric switches of the protein conformation.[7,8] These may be relatively low-affinity contacts to non-TR RNA stem–loops and might contribute maximally to capsid assembly only in the context of the wild type genome. The presence of the high-affinity TR site in the genome automatically provides the mechanism for genome packaging selectivity; a fact exploited to package non-cognate RNAs in MS2 capsids.[25] The fact that TR has only a modest effect on packaging efficiency of long RNAs shows how important the cooperativity of binding other stem-loops along the RNA must be.[26,27] This leads to a new, less protein-centric view of the assembly process, and implies an important prediction: reagents inhibiting the roles of the genomic RNA in assisting the coat proteins to assemble into a particle of the correct size and symmetry should be particularly powerful anti-viral agents.

Materials and Methods

Assembly reactions

These assembly reactions were performed as described.[9] Briefly, CP_2 (8 μM in 20 mM acetic acid) was incubated with an equimolar concentration of TR in 40 mM ammonium acetate (pH 5.2–5.7) at 4 °C. This formed the basis of the 1:1 reaction. Following an equilibration period of 10 min, more CP_2 was added to form the 2:1 and 3:1 assembly reactions. Reactions were sampled from an initial time point of 10 min post mixing to 3 h.

Quantification of the virus intermediates

For quantification of the virus intermediates des-Arg bradykinin, a small peptide of mass 904 Da, similar to that used earlier as an internal standard,[18] was tried. Electrospray ionisation-mass spectrometry has been used to determine the concentration of protein monomers using either an isoform of the target protein or a small peptide. des-Arg bradykinin is an ideal internal standard because it ionises well and, under the conditions used in these experiments, does not form adducts with CP_2 or the TR RNA components used in the virus assembly reactions. It is detected as singly charged ions at m/z 904, which do not overlap with any of the intermediates that lie within the m/z range 3000–10,000. Initially, a concentration calibration curve from 2.5 μM to 50 μM CP_2 was determined in the same buffer used to monitor virus assembly. Accurate measurements of ion concentration can be made by calibration of the peak areas of CP_2 and bradykinin. The concentration calibration curve (peak area *versus* protein concentration) was found to be linear over the range analysed and was used to convert experimental peak areas of the intermediates directly into protein concentrations. Unfortunately, the conditions optimised in the mass spectrometer for ionisation of the large non-covalent virus intermediates resulted in fragmentation of the

peptide. Alcohol dehydrogenase, a non-covalently bound tetrameric protein complex of 147.5 kDa and with ionisation efficiency expected similar to the virus intermediates was also tried as an internal standard. However, the charge states of the tetramer overlapped with those of the 3-fold intermediate. It was therefore decided to run reactions without an internal standard and to normalise virus intermediate species to the initiation complex [CP_2:TR], which ionises well.

Mass spectrometry

The concentration of CP_2 in solution was calculated from the UV absorbance. Stock solutions were diluted into buffer to give a final volume of reaction mixture of 25 μl. Samples were analysed by positive ionisation nano electrospray using a Synapt HDMS, quadrupole-IMS-orthogonal acceleration time-of-flight mass spectrometer (Waters Corp., Manchester, UK), equipped with a Nanomate (Advion Biosciences, Inc., Ithaca, NY, USA) temperature-controlled, automated sample handling and ionisation interface, maintained at 4 °C. A capillary voltage of 1.9 kV was set with nitrogen gas flow of 0.5 psi (1 psi ≈ 6.9 kPa) for sample introduction and ionisation. A source temperature of 60 °C and desolvation temperature of 100 °C were used. Data were acquired in continuum mode over the m/z range 500–30,000 and data processing was performed using the MassLynx software supplied with the mass spectrometer. An external calibration using CsI clusters was applied to the data. Each intermediate within the mass spectrum was identified by observation of at least two consecutive charge state peaks from which the mass was measured. The areas of all peaks (including salt adducts) for a given intermediate species were then summed and expressed as a ratio of the total peak area of the initiation complex. The signal for the initiation complex was defined as 100% and each intermediate was normalised to this value.

Kinetic modelling

The assembly reactions in Eq. (1) were converted into a set of differential equations describing the change in concentration of an assembly intermediate, I_n. The time evolution of the concentration of a single unique assembly intermediate containing n dimers, $[I_n]$, was expressed in terms of the concentrations of all species differing from I_n by exactly one dimer that can either enhance (via association of I_{n-1} with free dimer D or dissociation of I_{n+1}) or deplete (via dissociation of I_n or association of I_n with D) its concentration:

$$\frac{d[I_n]}{dt} = \sum_{C_{n-1}} \left(\sum_D \left(k_{f(n-1)}[I_{n-1}][D] - k_{b(n)}[I_n] \right) \right) + \sum_{C_{n+1}} \left(\sum_D \left(k_{b(n+1)}[I_{n+1}] - k_{f(n)}[I_n][D] \right) \right) \quad (3)$$

Here, $k_{f(n)}$ and $k_{b(n)}$ denote the forward and backward rates, respectively, of the corresponding reaction involving the intermediate I_n with n coat protein dimers. The initial concentration of coat protein dimers is considered in the model to occur in the ratio B/B to C/C of 1:5. To enforce this ratio, we set the value of the backwards rate for the conversion of a B/B dimer to a C/C dimer (k_b) to 0.046 s^{-1}. Apart from the rate of TR binding to C/C, which has been determined experimentally,[21] the rates of the reactions are not known. Brownian dynamics simulations of the association of two dimeric building blocks provided insights into the relative rates of association for CP_2:TR with C/C *versus* CP_2:TR with CP_2:TR.[31] The forward rates $k_{f(n)}$ of the assembly reactions have therefore been chosen as:

$$k_{f(n)} = k_f = 1.0 \times 10^5 \text{mol}^{-1}\text{s}^{-1}$$

for all reactions. In any case, a different choice for k_f, to a first approximation, would result in only a shift of the time frame (i.e. the speed of reaching equilibrium) and the concentration profiles would remain unaffected. Backwards rates are inferred from the relation

$$k_{f(n)} / k_{b(n)} = S_1 S_n \exp(\Delta G_n / RT) \quad (4)$$

which constrains the ratio of forward and backward rates of a reaction. This relation depends on the symmetry of the incoming subunit (S_1), the number of equivalent ways in which the subunit can associate with reactant I_n (S_n), as well as the free energy ΔG_n of the reaction (composed of energetic contributions from bond formation and interactions with TR). Together, the product S_1 S_n in Eq. (4) accounts for the degeneracy of the reaction; i.e. the number of different ways in which the reaction can be realized due to the symmetry of the intermediate and/or incoming subunit. R is the gas constant (1.98×10^{-3} kcal mol K) and T is temperature (in Kelvin), chosen here as $T = 277.15$ K (4 °C), the temperature at which the experiments were done. We solved Eq. (3) using an adaptive Runge-Kutta 45 algorithm‡ for the ratios 1:1, 2:1 and 3:1 (in multiples of 8 μM) of coat protein *versus* TR concentration. Following experiment, we started with equal amounts of coat protein and TR, then added further aliquots of coat protein after 10 min for the 2:1 and 3:1 reactions. We determined values for the two free energy parameters ΔG_α and ΔG_β in the model by sampling a variety of possible parameters within the range −0.1 to −10.1 kcal mol^{-1}. We first varied ΔG_α and ΔG_β in increments of 2 kcal mol^{-1} between −10.1 and −0.1 kcal mol^{-1} for both free energy parameters. A comparison of the predicted concentration profiles of [2(CP_2:TR)], [3(CP_2:TR)+3CP_2] and [5(CP_2:TR)+5CP_2] with the experimental data allowed us to choose new upper and lower bounds for these parameters. We then repeated the procedure iteratively using ever smaller increments. In the final round of sampling, we varied the parameters in increments of 0.02 kcal mol^{-1} with ΔG_α ranging from −0.9 to −0.8 kcal mol^{-1} and ΔG_β ranging from −5.80 to −5.70 kcal mol^{-1}. Different outcomes were compared using a minimal RMSD over the cumulative deviations for the concentration profiles of the dominant intermediates at 180 min for the 1:1 and 2:1 reactions, ignoring the 3:1 reaction because factors contributing to variation in ionisation efficiency could make these data less reproducible than those for the 1:1 and 2:1 reactions.

Acknowledgements

We thank Dr Neil Ranson, Dr Katerina Toropova, Dr Karim Elsawy and Dr Óttar Rolfsson for many helpful discussions of these data and assistance in preparing the structural figures. We are grateful for financial support of this work from the UK BBSRC, the Leverhulme Trust, and the University of Leeds.

‡ Adaptive Runge-Kutta-Fehlberg from http://www.netlib.org

R.T. thanks the Leverhulme Trust for a Research Leadership Award.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.05.059

References

1. Caspar, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.
2. Crick, F. H. & Watson, J. D. (1956). Structure of small viruses. *Nature*, **177**, 473–475.
3. Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature*, **276**, 368–373.
4. Schneemann, A. (2006). The structural and functional role of RNA in icosahedral virus assembly. *Annu. Rev. Microbiol.* **60**, 51–67.
5. Golmohammadi, R., Valegard, K., Fridborg, K. & Liljas, L. (1993). The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* **234**, 620–639.
6. Valegard, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, **345**, 36–41.
7. Basnak, G., Morton, V. L., Rolfsson, Ó., Stonehouse, N. J., Ashcroft, A. E. & Stockley, P. G. (2010). Viral genomic single-stranded RNA directs the pathway toward a $T = 3$ Capsid. *J. Mol. Biol.* **395**, 924–936.
8. Dykeman, E. C., Stockley, P. G. & Twarock, R. (2010). Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J. Mol. Biol.* **395**, 916–923.
9. Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J. *et al.* (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a $T=3$ viral capsid. *J. Mol. Biol.* **369**, 541–552.
10. Koning, R., van den Worm, S., Plaisier, J. R., van Duin, J., Pieter Abrahams, J. & Koerten, H. (2003). Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *J. Mol. Biol.* **332**, 415–422.
11. Toropova, K., Basnak, G., Twarock, R., Stockley, P. G. & Ranson, N. A. (2008). The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* **375**, 824–836.
12. van den Worm, S. H., Koning, R. I., Warmenhoven, H. J., Koerten, H. K. & van Duin, J. (2006). Cryo electron microscopy reconstructions of the Leviviridae unveil the densest icosahedral RNA packing possible. *J. Mol. Biol.* **363**, 858–865.
13. Rolfsson, O., Toropova, K., Morton, V. L., Francese, S., Basnak, G., Thompson, G. S. *et al.* (2008). RNA packing specificity and folding during assembly of the bacteriophage MS2. *Comput. Math. Methods Med.* **9**, 339–349.
14. Nettleton, E. J., Sunde, M., Lai, Z., Kelly, J. W., Dobson, C. M. & Robinson, C. V. (1998). Protein subunit interactions and structural integrity of amyloidogenic transthyretins: evidence from electrospray mass spectrometry. *J. Mol. Biol.* **281**, 553–564.
15. Pinkse, M. W., Maier, C. S., Kim, J. I., Oh, B. H. & Heck, A. J. (2003). Macromolecular assembly of Helicobacter pylori urease investigated by mass spectrometry. *J. Mass Spectrom.* **38**, 315–320.
16. Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U. *et al.* (2000). Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc. Natl Acad. Sci. USA*, **97**, 5185–5190.
17. Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. (2000). Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *J. Am. Chem. Soc.* **122**, 3550–3551.
18. Smith, A. M., Jahn, T. R., Ashcroft, A. E. & Radford, S. E. (2006). Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J. Mol. Biol.* **364**, 9–19.
19. Endres, D. & Zlotnick, A. (2002). Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys. J.* **83**, 1217–1230.
20. Zlotnick, A. (1994). To build a virus capsid. An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59–67.
21. Lago, H., Parrott, A. M., Moss, T., Stonehouse, N. J. & Stockley, P. G. (2001). Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: identification of a protein conformer unable to bind RNA. *J. Mol. Biol.* **305**, 1131–1144.
22. Reddy, V. S., Natarajan, P., Okerberg, B., Li, K., Damodaran, K. V., Morton, R. T. *et al.* (2001). Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.* **75**, 11943–11947.
23. Johnson, J. M., Tang, J., Nyame, Y., Willits, D., Young, M. J. & Zlotnick, A. (2005). Regulating self-assembly of spherical oligomers. *Nano Lett.* **5**, 765–770.
24. Williamson, J. R. (2008). Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* **4**, 458–465.
25. Pasloske, B. L., Walkerpeach, C. R., Obermoeller, R. D., Winkler, M. & DuBois, D. B. (1998). Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. *J. Clin. Microbiol.* **36**, 3590–3594.
26. Beckett, D. & Uhlenbeck, O. C. (1988). Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. *J. Mol. Biol.* **204**, 927–938.
27. Beckett, D., Wu, H. N. & Uhlenbeck, O. C. (1988). Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* **204**, 939–947.
28. Grahn, E., Stonehouse, N. J., Murray, J. B., van den Worm, S., Valegard, K., Fridborg, K. *et al.* (1999). Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements. *RNA*, **5**, 131–138.
29. Du, Z., Yu, J., Ulyanov, N. B., Andino, R. & James, T. L. (2004). Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. *Biochemistry*, **43**, 11959–11972.
30. Rolfsson, O., Toropova, K., Ranson, N. A. & Stockley, P. G. (2010). Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid. *J. Mol. Biol.* **401**, 309–322.
31. ElSawy, K. M., Caves, L. & Twarock, R. (2010). The impact of viral RNA on the association rates of capsid protein assembly: bacteriophage MS2 as a case study. *J. Mol. Biol.* This issue. doi:10.1016/j.jmb.2010.05.037.