This is a repository copy of A modelling paradigm for RNA virus assembly.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/143320/

Version: Published Version

Article:

https://doi.org/10.1016/j.coviro.2018.07.003

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Introduction

The formation of a viral protein container encapsulating a virus’ genomic cargo is a prerequisite for the successful propagation of a viral infection. A better understanding of this process can therefore be exploited for therapy, either via the development of antiviral strategies inhibiting assembly, or the repurposing of the self-assembly process for the design of gene vectors and vaccines.

The initial focus in the study of virion assembly was directed towards in vitro studies of capsid self-assembly in the absence of other viral components. Models developed in tandem with such experiments provided an understanding of the kinetics [1–3] and thermodynamics [4,5] of spontaneous capsid self-assembly, and of the roles of protein–protein interactions in defining quasiequivalent capsid geometries [6,7]. They also elucidated the local rules underpinning coat protein (CP) self-association during capsid formation [8,9]. Many viruses, especially double-stranded DNA viruses, assemble their capsids prior to genome packaging via an ATP driven packaging motor. The protein-centric models, with the addition of scaffolding proteins in the case of larger capsid shells, are therefore an adequate context to study capsid assembly in these cases. By contrast, single-stranded RNA viruses, the largest family of viruses and containing many important human pathogens, package their genomes during capsid assembly, exhibiting a co-assembly process. For these viruses, capsid assembly has to be modelled in tandem with genome packaging. An important aspect of virus assembly in the presence of genomic RNA is the need for genome compaction [10], and several groups have made important contributions to the modelling of this aspect of virus assembly [11*,12,13*,14*]. The impact of non-specific electrostatic interactions between genomic RNAs and CP [15–18,19**] and of the stiffness of the RNA molecule on the assembly process [20*] have been analysed. It has also been shown that the secondary structure of the RNA molecules play an essential role in determining capsid morphology in the self-assembly of Cowpea Chlorotic Mottle Virus (CCMV)-like particles [21]. The roles of genomic RNA have been studied in the assembly of helical viruses [22**]. Moreover, molecular dynamics simulations of capsid assembly, both in the absence and presence of different types of cargoes, have made important contributions to our understanding of virus assembly [23**,24]. Indeed, viral capsids can be assembled in vitro around different types of cargoes, including anions [25–27]. The models presented here go one step further. Instead of viewing viral genomes as passive passengers with at most non-specific electrostatic contributions to the assembly process, they demonstrate the consequences of the cooperative action of multiple, sequence-specific contacts between genomic RNA and CP.

Genomic RNA is not a passive passenger

Even in the absence of the genomic RNA, the CP of most single-stranded RNA viruses can self-assemble in vitro, but the process is typically much faster and more efficient in the presence of genomic RNA. This is the case, for example, for the assembly of the MS2 capsid (Figure 1) in the presence of multiple copies of the translational
A modelling paradigm for RNA virus assembly Twarock et al. 2018, 31:74–81

Figure 1

Genomic RNA defines capsomer conformation in MS2. (a) The MS2 capsid (based on pdb-id 1ZDH) is formed from asymmetric (blue/green) and symmetric (pink) forms of the coat protein dimer in a 2:1 ratio. (b) The stem–loop TR triggers a conformational change from the symmetric to the asymmetric form of the coat protein dimer. The characteristic packaging signal recognition motif is given by (x)\textcircled{y}A in the apical loop of the stem–loop, and the A in the 5′ bulge. Other stem–loops in the viral genome sharing aspects of this motif can also function as packaging signals [29,32].

repressor (TR) [28], a stem–loop in the genomic RNA known to function also as a packaging signal. This observation suggests that the contributions from genomic RNA to the assembly process are significant and therefore cannot be neglected in the assembly models.

There is only one copy of TR in the MS2 genome. Binding of TR to the CP dimer triggers a conformational switch from the symmetric dimer, the dominant form in solution, to its asymmetric conformation [29] that is needed in a 2:1 ratio for the construction of the capsid (Figure 1a). Normal mode analysis has revealed the structural features of TR that are required for this allosteric effect [30,31], demonstrating that many other, multiple dispersed, stem–loops in the MS2 genome could trigger the same effect [32]. This has resulted in the packaging signal (PS) hypothesis: Multiple dispersed secondary structure elements in the genomic RNA, with CP recognition features akin to those of the known high affinity PS, also trigger conformational changes of the CP dimer to its asymmetric conformation. These multiple dispersed sites have been called PSs, in analogy to the high affinity PS with which they share their characteristic feature for CP recognition. In the case of MS2, assembly mediated by these multiple dispersed PSs is also known as the dimer-switching model [33]. In other viruses, PSs can play a number of different roles in promoting capsid formation [35**,36*,45]. However, these different scenarios all share the same basic mechanism of PS-mediated assembly, in which multiple dispersed sites in the (pre) genomic viral RNA with affinity for CP promote efficient formation of a viral capsid with the correct geometry.

A mathematical model of PS-mediated assembly

In order to investigate how such multiple dispersed PS sites mediate capsid assembly, we developed a mathematical model that captures their collective impact on virus assembly efficiency (Figure 2) [37,38**]. From a geometric point of view, the simplest model of an icosahedral capsid is a dodecahedral shell formed from 12 pentagonal capsid building blocks (pentamers). This is representative of small plant viruses (\(T = 1\) geometries in the Caspar–Klug classification [39]), or the structures of Picornaviruses ((Pseudo)\(T = 3\) structures in which pentamers are formed by five protomers, each consisting of different polypeptides corresponding to the structural protein (VP) units). The model captures the assembly of 12 pentamers into a dodecahedral shell according to a set of simple local rules (Figure 2b): pentamers associate with, and disassociate from, PSs on the genomic RNA with rates depending on CP:PS affinity. As the precise nucleotide sequences of the PSs vary around their shared recognition motif, their affinities for CP can be distinct. In our model, they fall into three categories, weak (from 0 to \(−4\) kcal/M), intermediate (from \(−4\) kcal/M to \(−8\) kcal/M), and strong (from \(−8\) kcal/M to \(−12\) kcal/M), reflecting affinities seen in MS2 [40,41]. If two pentamers are bound to adjacent PSs, they form (or subsequently break) CP–CP interactions with rates determined by the free energy of the CP:CP bonds, chosen to be \(−2.5\) kcal/M following estimates in Ref. [4]. This model allows us to study the determinants of PS-mediated assembly in a scenario of reduced computational complexity.

A systems approach is key

Assembly against a backdrop of cellular competitor RNAs (in a 1:300 ratio consistent with experimental studies) [38**] reveals relatively low yields of viral particles compared with an abundance of misencapsidated particles (Figure 3), implying that in this simple form the model would not account for the assembly efficiency expected in vivo. This suggests that a key feature of the assembly process in vivo is missing in the model. Bacteriophage Qβ
A modelling paradigm for packaging signal-mediated assembly. (a) A dodecahedral model system is used as a coarse-grained representation of capsid geometry. (b) The order in which the protein building blocks of the capsid (pentamers) are recruited is indicated by a connected line (path) that connects midpoints of adjacent pentamers. A connected subset of such a path is shown superimposed on capsid assembly intermediates formed from four pentamers; the two examples represent different assembly scenarios. (c) The assembly of the dodecahedral model system from 12 pentamers is modeled in the presence of RNAs, that are represented by 12 beads, each of which represents a PS. Beads are colour-coded according to their affinities for CP, as green (strong), blue (intermediate) and red (weak). (d) The system assembly based on a set of local rules that are formulated as assembly reactions, describing RNA:CP and CP:CP interactions.

A solution to a viral-equivalent of Levinthal's Paradox

The model also reveals the mechanism by which viruses efficiently navigate the landscape of possible assembly intermediates [38]. In protein folding, the ensemble of potential folding pathways of an amino acid sequence into its native conformation is so complex that a random exploration of different options would take longer than the known age of the universe. Despite this, proteins fold within biologically meaningful timeframes, a phenomenon known as Levinthal’s Paradox, which we now understand, because protein chains do not sample all possible conformations on their way to their folded state. Similarly, the number of geometrically distinct ways in which a viral capsid can be built from CP is vast, yet virus assembly must have evolved strategies to bias assembly to the most efficient assembly pathways in order to sustain a productive infection against host defence mechanisms. Our model of PS-mediated assembly demonstrates how multiple dispersed PSs with varying affinities for CP can achieve this under the condition of the protein ramp (Figure 3). In particular, variations in PS affinities for CP across the genomic sequence result in nucleation of assembly at specific sites, as opposed to nonlocalised nucleation across the full length of the RNA genome in the absence of the protein ramp, that is, PSs impact on nucleation behaviour. Only a small number of distinct assembly pathways from the ensemble of geometrically possible ones are actually realized during PS-mediated assembly, which are characterized by assembly intermediates that deviate only minimally from those maximising CP:CP contacts. This demonstrates that the PS distribution mitigates the combinatorial complexity of the assembly process. In short, it solves a virus-equivalent to Levinthal’s Paradox in protein folding.
Hamiltonian paths analysis

Different assembly scenarios can be encoded by geometric book-keeping devices that capture the order in which PSs make contact with CP during virus assembly. In particular, by connecting PS binding sites on the capsid interior in the order in which the corresponding PS:CP contacts are made, a connected string is obtained that provides a geometric representation of the assembly pathway. Superposition of all possible such strings results in a polyhedral shape with vertices at the PS binding sites at the capsid’s interior surface, and edges connecting vertices on neighbouring capsomers. From a mathematical point of view, each individual string corresponds to a Hamiltonian path on this polyhedron, that is, a connected path visiting every polyhedral vertex precisely once. They do not represent, however, the exact location of the viral genome, which can make excursions into the capsid interior (Figure 4a). The (local) geometric properties of these paths can be classified for different types of capsid geometries. These local properties of the paths (as illustrated in Figure 4b for MS2) can then be used, in combination with a bioinformatics search for potential PS
candidates, to identify the likely PS distribution [32, 46, 47**]. Note that it is not necessary for all binding sites to be occupied, and that the Hamiltonian path constraints can be more restrictive in some regions of the genome than in others. For example, our Hamiltonian Paths Analysis predicted PSs for bacteriophage MS2, that are in excellent agreement with the RNA:CP binding sites identified via cross-linking immunoprecipitation (CLIP) experiments [48**]. Our analysis shows that PSs are more constrained in one half of the MS2 capsid.
(see red rhombs in Figure 4c based on Ref. [32]), which agrees well with an asymmetric EM reconstruction of MS2 at 8.7 Å resolution [49**]. Moreover, all PSs identified in a subsequent EM reconstruction at 3.6 Å resolution [50**] had previously been identified via our Hamiltonian Path Analysis method (Figure 4d). This demonstrates the utility of mathematical tools in identifying salient features in the organization of a packaged viral genome.

Conclusions

Modeling of PS-mediated assembly demonstrates the distinct advantages of PSs for efficient capsid formation. As PS-mediated assembly confers fitness advantages to viral particles assembling via this mechanism, it is likely that it is widespread in nature. The discovery of PSs in a number of viral families infecting different hosts including humans supports this hypothesis. Even Hepatitis B virus, a DNA virus, has been shown to reveal packaging signals in its pregenomic RNA, that impact on capsid geometry by biasing assembly towards formation of $T = 4$ shells [36*]. It is likely that multiple dispersed PSs will be discovered in many more viral systems over the next decade, for example, in the alphaviruses [51]. Similar assembly mechanisms may even occur more widely in nature, for example in the assembly of repurposed Gag-like proteins [52*] with roles in intercellular RNA transfer across synaptic boutons [53*].

The models of PS-mediated assembly have provided mechanistic insights that could not have been obtained via experiment alone. They revealed that hallmarks of PS-mediated assembly can only be observed in the context of scenarios reflecting in vivo infections, and demonstrated the importance of the PS affinity distribution for efficient capsid formation. The Hamiltonian path approach has moreover served as a tool for the identification of PSs [32]. The discovery of PS-mediated assembly has opened up novel opportunities for antiviral therapy, for example, via small molecular weight compounds blocking either the PS or CP sites of the PS:CP interactions. The modelling paradigm reviewed here provides a basis for the study of viral infections and viral evolution, and such models have been constructed in order to study the merits of different anti-viral strategies [54*] and the resilience of PS-mediated assembly under mutational pressures [55*]. The detailed understanding of the characteristics and functional roles of the PS distribution has moreover enabled novel applications in biomanufacturing. The PS assembly code can be isolated and repurposed for the construction of stable virus-like particles with improved assembly efficiency compared with their viral counterparts, as demonstrated for Satellite Tobacco Necrosis Virus [56**]. Such particles might be used as decoys, gene delivery vectors, or for vaccination purposes.

Acknowledgements

Funding from a Joint Wellcome Trust Investigator Award to RT and PG (110145 & 110146) is gratefully acknowledged, as is assistance with the figures from Prof Neil Ranson (University of Leeds).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- of special interest
- of outstanding interest

12. The authors study the impact of synonymous recoding of viral sequences on the ability of the genome to compact, which is a prerequisite for packaging into the limited volume of a viral capsid.
15. The authors study the assembly of cowpea chlorotic mottle virus establishing that at low pH capsid assembly is a homogeneous nucleation process, while at neutral pH heterogeneous nucleation dominates, guided by RNA:CP interactions.


The authors model RNA as a confined polymer chain, to understand the optimal degree of base-pairing for viral encapsidation given the strength of attraction between the capsid and the RNA.


Representing the RNA as a polyelectrolyte chain, the authors incorporate the interaction of the chain segments, capsid protein and ions at the level of Poisson–Boltzmann theory. Considering the degree of branching and the stiffness of the RNA (using the Kuhn length), a optimal RNA length for encapsidation is found.


The authors study the assembly of cowpea chlorotic mottle virus virions while varying the U content of the packaged nucleotide content. The polyU sequences are less likely to contain secondary structures but can outcompete viral RNA for CP, however the particles formed are more likely to kinetically trapped and structurally heterogeneous.


This is an excellent recent review of coarse-grained modelling approaches in the study of virus assembly.


Using a combination of SELEX, bioinformatics techniques and mutational analysis, the molecular details of the PS-CP interactions in human parecho virus are identified and their role in assembly established via mutagenesis. A 10-log drop in viral titre is found for a replication & translation competent mutant with changes to a critical PS.


PSs are identified in Hepatitis B virus using SELEX, binding assays and bioinformatics methods. HBV is a DNA virus, which assemblies around its pregenomic RNA, meaning that the PS-mediated assembly is not restricted to ssRNA viruses, in which it had been first discovered.


This paper demonstrates how multiple dispersed packaging signals with different affinities for coat protein promote efficient capsid assembly and selective genome packaging.


Using mutagenesis experiments, the authors demonstrate that the PSs in STNV are evolutionarily constrained by position and sequence motif. This demonstrates the existence of an assembly code, in terms of the PSs, that can be targeted by antiviral therapy, or be repurposed for the design of virus-like particles.


This paper demonstrates the predictive power of the Hamiltonian path approach based on a recent cryo-EM map for MS2 (see Ref 50 below).


The authors use CLIP-Seq and desorption/ionization mass spectrometry to determine the multiple dispersed, sequence-specific PSs in MS2.

This asymmetric EM reconstruction of bacteriophage MS2 at medium resolution (8.7 Å) reveals for the first time the structures of both the protein shell, the asymmetric genomic RNA and the unique maturation protein.


This asymmetric cryo-EM reconstruction of MS2 at 3.6 Å resolution identifies the nature of 15 PS:CP contacts, each of which had been predicted with the Hamiltonian Path Analysis reviewed here.


See annotation to Ref. [53].


The two references above describe the neuronal gene Arc, that exhibits homologies with the retroviral Gag protein. The gene encodes a protein that forms viral capsid-like structures, which specifically transfer Arc mRNAs. The analysis shows that assembly of these containers is more efficient in the presence of the mRNA, suggesting that there could be secondary structure elements in the mRNA that function similarly to PSs during container assembly, implying that the mechanism of PS-mediated assembly may be occurring more widely in biological systems beyond virology.


The virus assembly model reviewed here has been coupled with a model of viral replication in order to describe a viral infection at the scale of an individual infected cell. This has been coupled with a quasispecies model of viral evolution in the context of a viral infection. Application of this model to a chronic Hepatitis C infection demonstrates that drug strategies targeting PSs are less likely triggering drug resistance than conventional forms of therapy.


The dodecahedral model is expanded to include an explicit nucleotide sequence, where packaging efficiency depends on the ability of the RNA to fold locally into PSs. This enables the mutation of the RNA primary structure, thus allowing exploration of the effects of such mutations on viral assembly. The study suggests that viruses rely on degenerate PSs to ensure mutational resilience.


The knowledge of the PSs of STNV is used to create a synthetic nucleotide fragment which assembles more efficiently than WT, and even outcompetes the WT in a head-to-head competition. This demonstrates that the PS-encoded virus assembly instruction manual can be optimised and repurposed for the synthesis of virus-like particles, with potential applications as gene vectors or in vaccinology.