Trends in
Microbiology

CellPress
OPEN ACCESS

Opinion

Molecular frustration: a hypothesis for regulation of viral infections

Reidun Twarock,[1] Greg J. Towers,[2] and Peter G. Stockley[3,*]

The recent revolution in imaging techniques and results from RNA footprinting *in situ* reveal how the bacteriophage MS2 genome regulates both particle assembly and genome release. We have proposed a model in which multiple packaging signal (PS) RNA-coat protein (CP) contacts orchestrate different stages of a viral life cycle. Programmed formation and release of specific PS contacts with CP regulates viral particle assembly and genome uncoating during cell entry. We hypothesize that molecular frustration, a concept introduced to understand protein folding, can be used to better rationalize how PSs function in both particle assembly and genome release. More broadly this concept may explain the directionality of viral life cycles, for example, the roles of host cofactors in HIV infection. We propose that this is a universal principle in virology that explains mechanisms of host–virus interaction and suggests diverse therapeutic interventions.

Highlights

Viral infections are directional; the mechanisms of assembly and disassembly are distinct and coordinated.

Molecular frustration describes the coexistence of alternative states of a macromolecular system where all favourable interactions cannot be satisfied simultaneously.

We hypothesize that molecular frustration of the coat proteins and their complexes with the genome in bacteriophage MS2 accounts for regulation of assembly and disassembly during infection.

We therefore hypothesize that the same phenomenon may also underlie the directionality of other viral infections, for example, how cellular cofactors regulate infection by HIV.

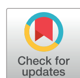
Introduction

Assembly of an infectious virion is a vital step in any viral life cycle. It transforms otherwise harmless molecules into an infection machine capable of detecting, transferring its genome into, and then subverting the gene expression of, a host cell. This directional process gives rise to the release of multiple progeny virions able to repeat this process of infection with additional susceptible cells. Once a target cell has been engaged, virions become sensitive to further molecular cues from the host cell, ultimately resulting in sequential genome release, gene expression, and the formation of progeny virions. The molecular basis of this infection directionality is not well understood, but recent advances hint at mechanistic explanations that may be applicable to all viral systems.

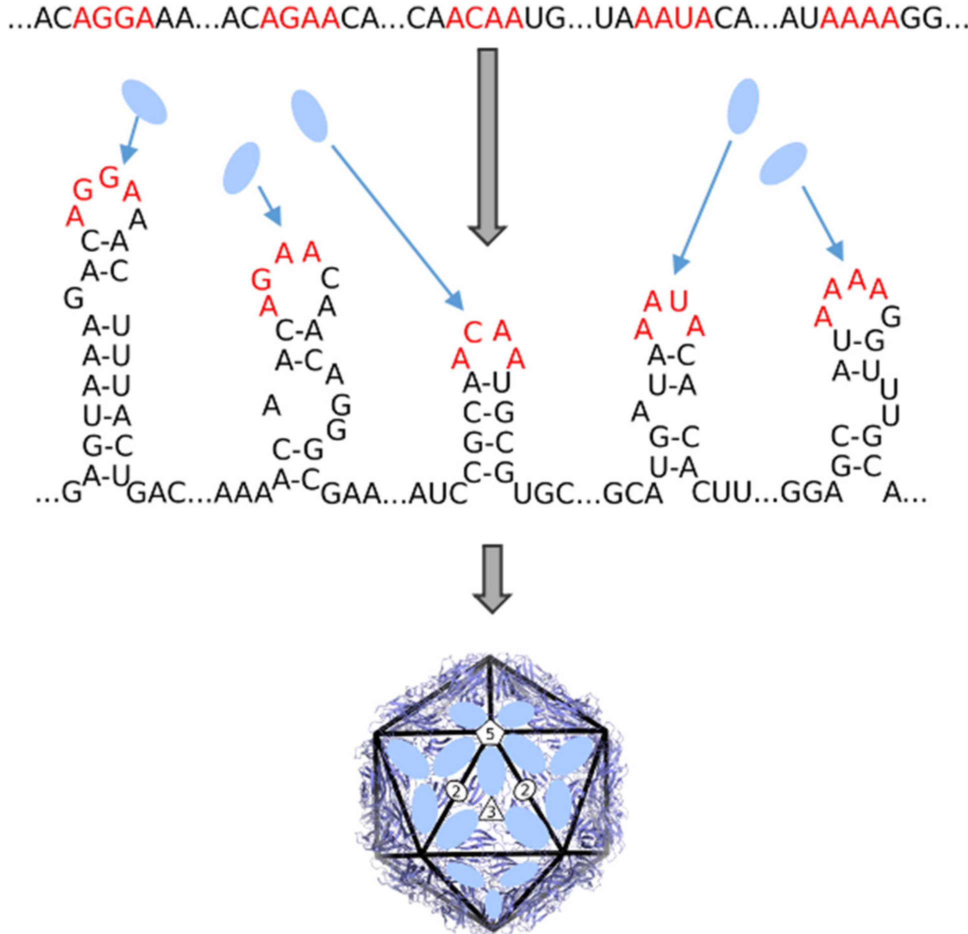
For single-stranded (ss)RNA viruses, which occur in many viral subfamilies and include major human viral pathogens, we proposed and validated an assembly mechanism (Figure 1) that ensures highly efficient, genetically robust, and precise assembly. The outcomes of RNA virus infections are progeny virions containing the cognate viral genomes encoding the **coat proteins (CPs)** (see Glossary) of their protective shell. We have shown that many viral RNA genomes (gRNA) contain an ensemble of PS motifs/sites dispersed throughout the gRNA. Each of these has a different affinity for the cognate CP, collectively defining a preferred assembly pathway that ensures the observed packaging specificity. For the RNA phage MS2, asymmetric cryo-electron microscopy (cryo-EM) reconstructions and gRNA X-ray footprinting (XRF) reveal that, after assembly, the majority of these RNA–CP contacts dissociate, creating local 'strain' in the capsid shell. This makes genome release during infection easier. RNA PS-mediated assembly regulation can be viewed as a type of molecular Velcro. Individually, PS–CP contacts are easily broken, but as an ensemble they are very stable. A viral capsid cannot be too stable since it would then not be able to uncoat. Thus, the multiply dispersed contacts that promote capsid assembly are, at least partially, reversed during or after assembly, sufficiently destabilizing the capsid that it is able to release the viral genome on the appropriate cellular cue. The molecular frustration created during these steps is advantageous evolutionarily, and we propose that this lies at the heart of regulating all viral life cycles.

[1]Departments of Mathematics and Biology & York Cross-Disciplinary Centre for Systems Analysis, University of York, York, UK
[2]Division of Infection and Immunity, University College London, Gower Street, London WC1E 6BT, UK
[3]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK

*Correspondence:
p.g.stockley@leeds.ac.uk
(P.G. Stockley).



© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



Trends in Microbiology

Figure 1. The **packaging-signal-mediated assembly** mechanism. Packaging-signal-mediated virus assembly [1] relies on multiple secondary structure elements (**packaging signals**, **PSs**) with a shared sequence motif (here AxxA, red) that is recognized by viral coat protein (CP, blue ovals). Individual stem-loops will drive sequence-specific assembly *in vitro*, whilst multiple sites, such as those at the 5' end of the Satellite Tobacco Necrosis Virus-1 gRNA shown, act co-operatively to promote assembly [3]. The cooperativity is sensitive to the folding propensity of the stem-loop, as well as the relative nucleotide separation of the stems.

Assembly via the PS-mediated mechanism relies on the presence of multiple dispersed sequences (motifs) scattered across the genome (gRNA) that each have affinity for their cognate CPs [1]. These RNA PSs vary in sequence around a consensus CP-binding motif that is typically presented in the context of RNA stem-loops, that is, the simplest gRNA secondary structure elements that fold by local self-base-pairing [2–6]. Base-pairing propensity and the degree to which each PS matches the consensus motif create a hierarchy of CP-affinities across the gRNA that drives assembly along a preferred assembly pathway [7]. Such a stepwise mechanism prevents multiple assembly initiation events occurring on the same gRNA that cannot easily be reconciled by formation of a single capsid. We [2,5,6,8,9], and now others [10,11], have shown that PS-mediated assembly regulates virion/nucleocapsid formation in a number of ssRNA viruses, for example, the picornaviruses and alphaviruses, as well as the para-retrovirus hepatitis B virus. Strikingly, a mechanism similar to PS-mediated regulation of assembly evolves spontaneously in a directed evolution experiment studying mRNA encapsidation by a cage-forming bacterial

Glossary

Coat/capsid proteins (CPs): the proteins constituting the viral (nucleo) capsid that provides protection for the viral genome.

FG motifs: characteristic sequence motif/structural folds of **phenylalanine-glycine-rich nucleoporins** (**FG-Nups), given by tandem repeats** of phenylalanine-glycine (FG repeats). These FG repeats are found in approximately one-third of Nups (also called FG-Nups).

Nuclear pore complex (NPC): a permeability barrier between the nucleus and the cytoplasm.

Packaging-signal-mediated assembly: a mechanism of virus assembly in which multiple dispersed packaging signals in a gRNA act collectively to promote formation of the viral (nucleo)capsid.

Packaging signals (PSs): sequence/structure motifs in genomic RNAs recognized by the respective cognate coat proteins. In the systems we have characterized to date, these consist of RNA stem-loops, that is, their formation is driven by short-range interactions.

Phenylalanine-glycine-rich nucleoporins (FG-Nups): a molecular filter regulating the selective NPC crossing of biomolecules.

enzyme [12]. This outcome suggests that this mechanism of assembly regulation confers significant evolutionary advantages [13]. It also opens novel opportunities for exploitation. Drug therapy directed against PS–CP contacts could be particularly effective because such multiple dispersed targets cannot simultaneously mutate to become resistant to drug binding [14,15].

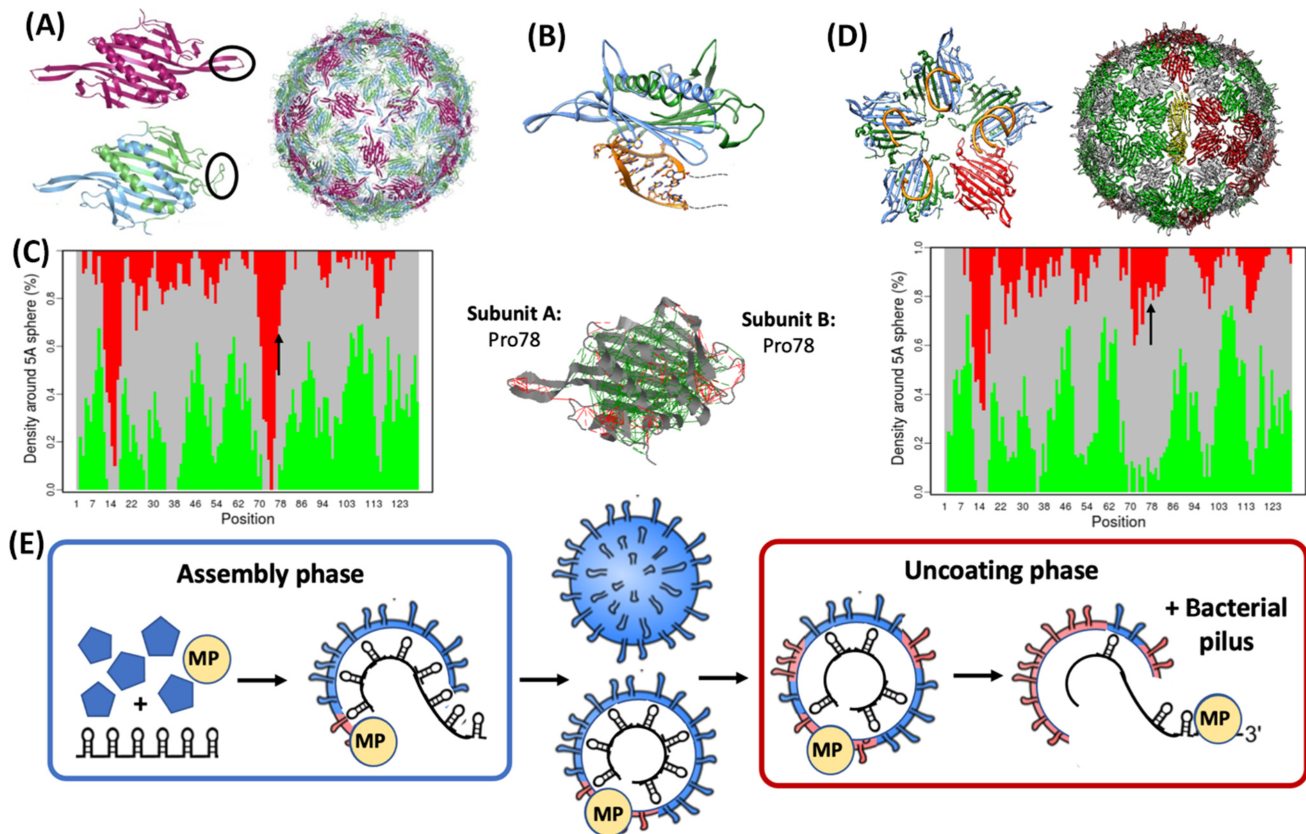
PS-mediated assembly accounts for the observed viral genome packaging efficiency seen in *in vivo* infections, but it can only be replicated in *in vitro* reassembly reactions by working at very low concentrations [16]. Genome encapsidation specificity is of course an immediate consequence of the mechanism, but a similar outcome has been ascribed to the formation of viral replication factories in cells that sequester CPs and gRNAs so that they inevitably interact [17]. However, such *in vivo* sequestration does not explain the widespread occurrence of PS sequences in many viral genomes that re

gulate assembly *in vitro*. Nor is it informative of the details of virion assembly, that is, the role(s) of gRNA in regulating the formation of defined CP complexes that act as the building blocks (capsomers) of the assembled virion that emerges. Indeed, the massive differences in self-assembly rates with and without PS-containing genomes or genomic fragments *in vitro* is consistent with this form of regulation. For most viruses, virion assembly is followed by disassembly-dependent infection, two apparently competing processes. Our understanding of this necessary directionality is currently very poor, for example, how does a virus react to where it is in the target cell – is this via host cofactor interactions that regulate the infectious process? Which roles do PSs play? We propose here that the directionality of a viral infection can be better understood in terms of the molecular frustration concept introduced by Wolynes and colleagues in the context of protein folding [18].

Molecular frustration and fuzziness – the keys to biological functions

According to the minimal frustration principle, proteins fold by minimizing their internal energetic conflicts. The resulting energy landscape has a strong bias towards the native state, as most interactions present on it are favourable, in contrast to random interactions that could occur during the folding process. However, remaining energetic conflicts lead to a state referred to as molecular frustration, or fuzziness, and an ensemble of these are present in the native state. Proteins interconvert between these different conformations that are in a dynamic equilibrium. This flexibility ultimately allows proteins to perform their molecular functions, for example via allosteric control.

We investigate here the hypothesis that molecular frustration plays a key role in viral life cycles. The bacteriophage MS2 coat protein dimer adopts two conformations (A/B and C/C) that coexist and are required in a ratio 2:1 in the fully formed capsid (Figure 2A and Box 1). In the B subunit, Pro78 adopts a *cis* peptide conformation. As a result, the loop linking the F and G β-strands bends towards the globular body of the dimer, allowing A/B dimers to fit around the fivefold axes of the phage particle without steric clashes. In the A or C subunits, the Pro78 residue is *trans,* confining the adjoining FG-loops to be in the A and C conformation. The switch between A/B and C/C must occur in a controlled way in order to guarantee the relative ratio required for the particle. Indeed, in solution, the C/C dimer is the dominant species. However, contact with a gRNA PS shifts the conformational equilibrium in favour of the A/B dimer (Figure 2B), resulting in rapid and efficient assembly in the presence of the gRNA [19]. We demonstrated previously that this conformational switch occurs via a redistribution of the kinetic energy of the dimer upon PS contact. Whilst the FG-loop of the A subunit becomes restricted in motion, its counterpart in the B subunit explores a larger conformational space, ultimately leading to the formation of a salt-bridge arresting the FG-loop in the flipped conformation [20]. We note that the dynamics of the MS2 dimer can also be affected by the insertion of an additional polypeptide sequence, such as those required to functionalize the



Trends in Microbiology

Figure 2. Molecular frustration in bacteriophage MS2. (A) The MS2 capsid is composed of 30 symmetric C/C dimers (magenta, top), one of which is replaced by a single copy of maturation protein (MP) in the phage shell, and 60 asymmetric A/B dimers (blue/green, bottom). The latter are organized in groups of five around the particle 5-fold axes. The flip of the FG-loop (circled) in the B monomer (green) is required to avoid steric clashes. (B) The packaging signal (PS)-free MS2 coat protein (CP) dimer is predominantly in the C/C conformation in solution. When a PS (such as the translational operator TR; orange) binds, it triggers a conformational change to the A/B conformation (blue/green as in (A)). (C) Configurational molecular frustration of the MS2 CP dimer computed with the FrustratometeR [22,24]; there are lower levels of molecular frustration in the flipped loop conformation in subunit B (right), compared with the flexible loop in subunit A (left) mostly N-terminal to Pro78 (indicated by arrows). (D) During or post-assembly, some PSs detach from the CP dimer (red). The PSs detach preferentially from dimers in the vicinity of MP (yellow), potentially facilitating is extrusion from the capsid shell. (E) Cartoon illustrating the roles of PSs in the MS2 life cycle that we rationalize using the concept of molecular frustration. Increased strain in the capsid shell is indicated in red.

capsid surface. This results in a shift of the C/C to A/B ratio, leading to polymorphic assembly of particle morphologies with more C/C dimers than wild-type [21].

Wolynes and colleagues pointed out the role of molecular frustration in the allosteric control of protein assemblies [22] and have provided an on-line 'FrustratometeR' for its calculation [23,24]. We used this to compute configurational molecular frustration for the MS2 dimer. Comparison of frustration values for residues in the FG-loop, in particular for the Pro78 residue, reveals much reduced molecular frustration in the B subunit compared with the A subunit (Figure 2C). This implies that PS-binding results not only in one of the subunits adopting the B conformation but also reduces the molecular frustration of the CP dimer. It had been suggested that binding partners, or changes in the cellular milieu, for example pH or ion concentrations, exploit fuzziness by reducing the number of distinct conformational substates, thus enabling context-dependent regulation within cells [18,25]. By analogy to this, PS-binding results in allosteric control of the equilibrium between the different coat protein conformers. It is possible that it also plays a role

Box 1. Viral structure determination

At the beginning of the molecular analysis of biological structures such as viruses, X-ray crystallography was the only technique capable of structure determination. The first virus structure at atomic resolution, for the spherical ssRNA Tomato Bushy Stunt Virus (TBSV), was determined in 1978. It was then at the technical cutting edge. Two factors compromised the outcome, however. Firstly, the necessity of crystallization results in alignment of objects dominated by the symmetry of the viral coat protein capsid. Secondly, to simplify the essential structure calculations, icosahedral symmetry for this capsid was assumed. Since these assumptions yielded a high-resolution structure they seemed reasonable, although all nonsymmetrical aspects of the virion, for example, parts of the CP N-terminal arms and the entire RNA genome, became invisible. Similar X-ray analysis of phage MS2 also yields a genome-free symmetric capsid. Its X-ray structure also makes the genome invisible, and it lacks density for the unique MP within the protein shell. Worse still, it replaces MP with a CP dimer in the correct conformation for its quasi-equivalent location, a consequence of symmetry-averaging, that is, not only does this technique remove vital information it can inadvertently also create structures that were never present.

The imaging revolution of cryo-electron microscopy takes advantage both of an improved electron beam but also of dramatically improved computational power. The result is the ability to determine asymmetric structures in which the computer aligns three-dimensional objects imaged individually. Such reconstructions can reveal features that lack symmetry. For RNA phages, these include both the MP protein and most of the ssRNA genome. In these cases, for the first time in any virus, it is possible to see most of the structure of the infectious machinery of a virion.

during the formation of higher order protein complexes on the pathway to the fully formed capsid as protein interfaces in higher order assemblies become less frustrated upon complex formation [26]. Molecular frustration thus impacts virus assembly both by enabling allosteric regulation of the equilibrium between the two coat protein dimers, and by promoting complex formation, thus contributing to virus assembly in different ways.

Does the role of molecular frustration extend to entire viral life cycles?

Bacteriophage MS2 assembly provides an example of functionally important molecular frustration and fuzziness in virology. Here we propose that the concept extends to the later stages of a viral life cycle. This hypothesis is rooted in our recent observation that individual PSs in the MS2 gRNA dissociate from the protein capsid shell post/during assembly. Asymmetric cryo-EM structures [27,28] and X-ray RNA footprinting data [29] suggest that over half of the PSs that drive assembly ultimately dissociate in this way. These dissociations occur preferentially in the gRNA domain located between the single copy of maturation protein (MP) and the CP dimer bound at TR (Figure 2D). MP binds both close to the 3'-end of the gRNA as well as to the receptor for the cellular target, the F pilus, As the FG-loops of the associated B subunits are unable to revert back to extended conformations without disrupting the CP lattice of the phage shell, this creates local strain in the capsid surface, impacting its stability. Loss of disproportionate numbers of PSs around MP differentially destabilizes this area of the capsid, priming the protein lattice for localized rupture around MP, followed by genome release at this site. The loss of specific PS–CP contacts towards the 3'-end of the gRNA facilitates the genome–MP complex entering the host cell in a 3'–5' direction [30]. This is important for replication, as the first cistron to be translated must be the replicase, leading to rapid expansion of the gRNA copy number. The incoming MP subunit gets proteolytically cleaved during cell entry, making the 3'-end of the nascent gRNA accessible to the phage replicase enzyme. The dissociation of a significant number of PS–CP contacts thus both prepares the virion for infection and confers directionality to the infection process.

We propose that molecular frustration, extended to encompass the entire virion comprising protein shell and packaged genome, accounts for this directionality. This requires broadening consideration of frustrated interactions to include also gRNA self-interactions in its secondary and tertiary structure. In this view, the assembled virion, composed of capsid and packaged gRNA, can be described as an ensemble of alternative states, which are characterized by differing numbers of PS–gRNA contacts as well as distinct conformations of the packaged gRNA. Conformations in which PSs are bound, or in which PSs have unbound from the CP dimers, allow the gRNA

tertiary structure inside the particle to reorganize into a more favourable configuration. Akin to the classical molecular frustration argument of Wolynes and colleagues in the context of protein folding, the virion, in its ensemble of alternative states, follows the principle of minimal frustration. This results in coordinated and programmed breaking of PS–CP contacts to ultimately ensure successful infection at the correct time and location.

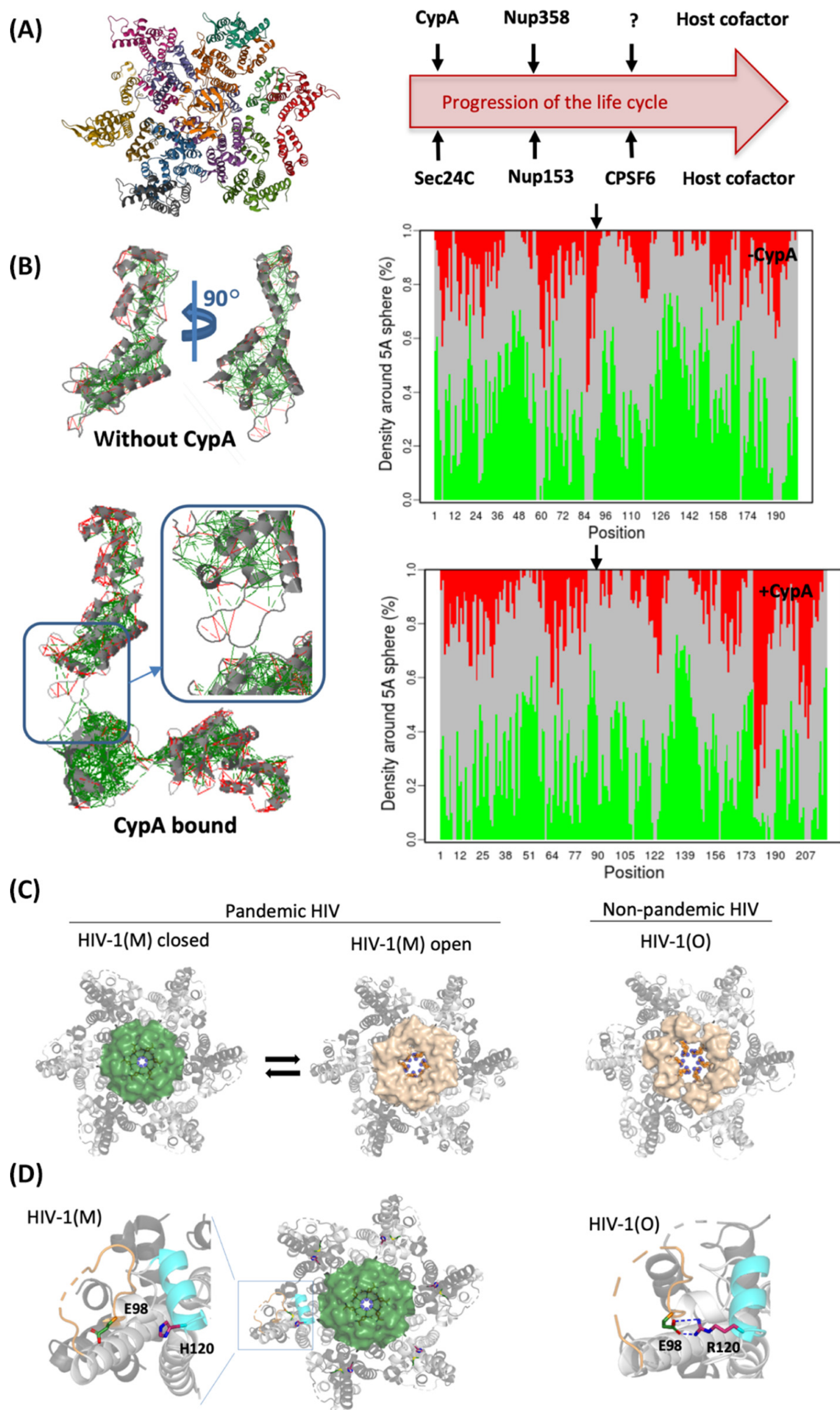
HIV as an example of viral infection and molecular frustration induced fuzziness

We infer that the principle of viral particle regulation by molecular frustration and fuzziness is widespread in virology and extends beyond regulation via PSs. As an example, we describe the regulation of HIV by host cofactors recruited to viral capsids. HIV capsids comprise 250 hexamers (Figure 3A) and 12 pentamers of the capsid protein. Intact capsids protect viral genome synthesis and travel across the cytoplasm, through a nuclear pore, and into the nucleus where they dissociate to release their genomes adjacent to chromatin [31–34]. Tight regulation of disassembly and genome release is essential to avoid genome degradation and nucleic acid sensing [31,35,36]. We propose that position-specific host cofactors act as waypoints, and confer directionality to infection, allowing capsids to respond to their location in target cells particularly by regulating uncoating [31,36,37] (Figure 3B). There are two distinct capsid cofactor binding sites. A flexible surface loop binds cytoplasmic cyclophilin A (CypA) and the Nup358 Cyp-like domain in the **nuclear pore complex (NPC)** [37,38]. A second site between capsid monomers sequentially recruits phenylalanine-glycine (FG) motifs provided by cytoplasmic protein Sec24C, NPC-associated proteins including Nup153, Nup98, and Nup35 and CPSF6 in the nucleus [39–45].

CypA recruitment reduces CypA binding loop mobility [46]. Configurational molecular frustration of a CA monomer (Figure 3B, top) is much reduced at the CypA binding site (P90, arrow), demonstrating that cofactor binding reduces molecular frustration. Strikingly, it also changes molecular frustration at multiple areas across the capsid, hinting at effects on downstream cofactor binding and uncoating. Notably, frustration appears to be increased by CypA in the C terminus of CA in the region where the downstream cofactor Nup153 binds CA via a Nup153 KKK motif [45]. Experimental evidence also supports our model. Prevention of CypA interaction by capsid mutation, or CypA inhibition, makes the virus insensitive to depletion of downstream NPC-associated and nuclear cofactors [37,39,44]. We hypothesize that failure to interact with CypA prevents the formation of the appropriate state to enable functional interactions with downstream FG-bearing cofactors in the NPC and nucleus. In this model, CypA-CA interactions affect molecular frustration, for example, at the FG binding site, potentially impacting the binding, or consequence of the FG–cofactor binding, maximizing infection and minimizing innate immune sensing.

We hypothesize that, in analogy to the CypA case, FG–cofactor binding reduces molecular frustration, conferring directionality to the sequence of cofactor binding events. Crystal structures of cofactor–capsid complexes do not reveal allosteric regulation or help understand cofactor activity [40–42]. By contrast, molecular frustration theory provides a dynamic model in which the existence of distinct cofactor binding modes in different environments, for examples in the cytoplasm, NPC, or nucleus, can be better understood. In this framework, progression through frustration states depends on cues from cofactor interactions, each reducing or recalibrating molecular frustration in the system, thus influencing the recruitment or consequences of downstream cofactor binding. For example, early interactors, for example, CypA, should bias the system to favour downstream interactions (e.g., NPC-associated proteins and CPSF6) and ensure uncoating at the appropriate time and place ready for immediate integration. As in the MS2 case, the ultimate outcome is genome release at the desired time and location in the life cycle. This is consistent with cofactor binding inducing capsid conformation changes to regulate viral trafficking, DNA synthesis, and uncoating.

CellPress
OPEN ACCESS



Trends in Microbiology

(See figure legend at the bottom of the next page.)

The importance of molecular frustration and fuzziness, and its regulation by cofactor recruitment, is also supported by recent comparison of pandemic and nonpandemic HIV. Remarkably, the single pandemic HIV-1(M) lineage has adapted its capsid to increase fuzziness, suggesting evolution of fuzziness as a way of regulating host virus interactions to maximize replication, transmission, and pandemic potential [47].

Discussion

Wolynes and colleagues introduced the concepts of molecular frustration and fuzziness to model protein function, including allosteric regulation. We provide examples of molecular frustration in virology, demonstrating that PS binding in a bacteriophage, and cofactor binding in HIV, reduce molecular frustration of the structural viral proteins in each case. Similar regulation is very likely to occur more widely in virology. We propose that this could provide an explanation of the behaviour of more complex viral systems and their mechanisms of regulating life cycle events at specific times and cellular locations through interaction with cellular components. Viruses require their virions to be meta-stable, assembling a container to protect their genomes and restricting their access to cellular systems. Yet at the same time they must retain the ability to uncoat and release genome at the appropriate time and place. For bacteriophages like MS2, this requires the priming of the virion for genome release upon contact with the bacterial receptor, the F-pilus; for viruses such as HIV, this includes sensing and responding to their cell location.

We argue that a detailed understanding of viral molecular frustration mechanisms will reveal therapeutic opportunities (see Outstanding questions). Such inhibitors could be potent, dysregulating multiple aspects of viral life cycles. A capsid-targeting, cofactor-mimicking drug has been developed by Gilead Sciences which inhibits multiple stages of the HIV life cycle [48,49] and represents such an inhibitor. Low-molecular-weight compounds targeting the conserved features of multiple, dispersed, genomic PSs are highly inhibitory to assembly [12]. Such 'drugs' are likely to target viruses independently of strain variation, and the multiple PS sites are unlikely to mutate simultaneously to resist treatment [14].

Molecular frustration likely manifests in distinct ways in different viral systems, as proposed here for MS2 and HIV. As we have demonstrated here, molecular frustration sheds new light on the mechanisms of genome release. In MS2, molecular frustration underpins a release mechanism that results in localized rupture of the capsid around MP. If frustration-linked conformational changes are greater at the narrow end of the conical HIV capsid, this could also explain the propensity of the capsid to rupture at these sites [50,51]. Identifying genome- and cofactor-driven changes in frustrated interactions over time is now tractable using XRF [29] on frozen samples, including both nucleic acid and protein interactions. Such experiments will reveal new knowledge of fundamental biological principles underpinning viral life cycles, perhaps universally. In HIV, the time resolution and ability to analyse frozen/trapped samples using XRF will enable characterization of how cofactors impact the overall system, revealing consequences for virion conformational complexity. As we have demonstrated here, such analyses must include all viral components,

Outstanding questions

Does molecular frustration theory explain how all viral infections are regulated?

Are viral particles individually distinct with respect to their degree of frustration and therefore differentially infectious?

Can insights into life cycle regulation be derived from molecular frustration theory and used to develop novel therapeutic approaches?

Can we develop drugs targeting PS–coat protein contacts, or decoys mimicking PS characteristics?

Can we exploit insights from PS-mediated regulation to improve viral vectors for gene therapy, for example, would recoding PSs into therapeutic nucleic acid cargoes improve vector production and infectivity?

Can we develop viral inhibitors that inappropriately mimic cofactor-driven molecular frustration effects on viral capsids and thus inactivate them?

Figure 3. Molecular frustration in HIV. (A) The CA hexamer (left) responds to host factors with Cyp-like (top) or **FG motifs** (bottom). (B) Configurational molecular frustration of a CA monomer (top) and a CA monomer in complex with CypA (bottom) computed with the FrustratometeR [22,24]. CypA binding reduces molecular frustration at the CypA binding site (P90, arrow). (C) The pandemic HIV-1(M) lineage has adapted to increase capsid dynamics in two ways. Pandemic HIV-1(M) has adapted the beta-hairpin hinge region to be able to form closed (green) and open (tan) BHP conformations. Nonpandemic HIV-1(O) retains the ancestral state and can only form the open BHP conformation (tan). (D) Pandemic HIV-1(M) has also lost the R120 salt bridge which is expected to reduce CypA binding loop (tan) dynamics and thus regulation of loop mobility by Cyclophilin A recruitment in nonpandemic HIV-1(O). Nonpandemic HIV-2 also resembles HIV-1(O). The parental simian immunodeficiency virus from chimpanzees (SIVcpz*ptt)* represents HIV-1(M), suggesting that these adaptations occurred during chimpanzee infection. Further experimental details and reconstructions of phylogenetic histories are reported in [47].

including gRNA (including its PS distributions as in MS2), and nonstructural proteins (such as integrase for HIV). XRF will reveal whether the loss of PS-type contacts between gRNA and capsid, akin to the molecular frustration-based mechanism in MS2, perhaps triggered by cofactor binding, can regulate reverse transcription. This would be a parallel to the situation in hepatitis B virus (HBV), which also reverse transcribes its gRNA inside the virion capsid [9,52]. The molecular frustration concept has proven useful in refining protein structure predictions [53,54] and in characterizing enzyme activity [55]. It applies at atomic resolution [56,57] and enhances our understanding of biological functions both at the protein level and in multiprotein systems. We argue that, by studying genomic RNA, and its allosteric impact on protein conformation and aggregation, molecular frustration and fuzziness, we can shed new light on the directionality of viral infections.

Acknowledgments

R.T. acknowledges funding via an EPSRC Established Career Fellowship (EP/R023204/1) and a Royal Society Wolfson Fellowship (RSWF/R1/180009). R.T. and P.G.S. acknowledge support from two Joint Wellcome Trust Investigator Awards (110145/110146/Z/15/Z & 125243/ 224509/A/21/Z); the UK BBSRC and a Roscoe Prize Award from the British Medical Association. G.J.T. is funded by a Wellcome Investigator Award, a Wellcome Collaborative Award and grants from the Rosetrees Foundation, the UK Medical Research Council and UKRI. HIV figures were prepared with input from Morten Govasli Larsen (UCL).

Declaration of interests

No interests are declared.

References

1. Twarock, R. and Stockley, P.G. (2019) RNA-mediated virus assembly: mechanisms and consequences for viral evolution and therapy. *Annu. Rev. Biophys.* 48, 495–514
2. Patel, N. *et al.* (2017) Rewriting nature's assembly manual for a ssRNA virus. *Proc. Natl. Acad. Sci. U. S. A.* 114, 12255–12260
3. Bunka, D.H. *et al.* (2011) Degenerate RNA packaging signals in the genome of satellite tobacco necrosis virus: implications for the assembly of a T=1 capsid. *J. Mol. Biol.* 413, 51–65
4. Dykeman, E.C. *et al.* (2013) Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J. Mol. Biol.* 425, 3235–3249
5. Shakeel, S. *et al.* (2017) Genomic RNA folding mediates assembly of human parechovirus. *Nat. Commun.* 8, 5
6. Chandler-Bostock, R. *et al.* (2020) Assembly of infectious enteroviruses depends on multiple, conserved genomic RNA-coat protein contacts. *PLoS Pathog.* 16, e1009146
7. Twarock, R. *et al.* (2018) Hamiltonian path analysis of viral genomes. *Nat. Commun.* 9, 2021
8. Rolfsson, O. *et al.* (2016) Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J. Mol. Biol.* 428, 431–448
9. Patel, N. *et al.* (2017) HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nat. Microbiol.* 2, 17098
10. Brown, R.S. *et al.* (2020) Multiple capsid protein binding sites mediate selective packaging of the alphavirus genomic RNA. *Nat. Commun.* 11, 4693
11. Brown, R.S. *et al.* (2021) Specific recognition of a stem-loop RNA structure by the alphavirus capsid protein. *Viruses* 13, 1517
12. Tetter, S. *et al.* (2021) Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* 372, 1220–1224
13. Dykeman, E.C. *et al.* (2014) Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5361–5366
14. Bingham, R.J. *et al.* (2017) RNA virus evolution via a quasispecies-based model reveals a drug target with a high barrier to resistance. *Viruses* 9
15. Patel, N. *et al.* (2022) Dysregulation of Hepatitis B Virus Nucleocapsid Assembly in vitro by RNA-binding Small Ligands. *J. Mol. Biol.* 434, 167557
16. Borodavka, A. *et al.* (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl. Acad. Sci. U. S. A.* 109, 15769–15774
17. Routh, A. *et al.* (2012) Packaging host RNAs in small RNA viruses: an inevitable consequence of an error-prone polymerase? *Cell Cycle* 11, 3713–3714
18. Gianni, S. *et al.* (2021) Fuzziness and frustration in the energy landscape of protein folding, function, and assembly. *Acc. Chem. Res.* 54, 1251–1259
19. Hill, H.R. *et al.* (1997) Analysis of phage MS2 coat protein mutants expressed from a reconstituted phagemid reveals that proline 78 is essential for viral infectivity. *J. Mol. Biol.* 266, 1–7
20. Dykeman, E.C. *et al.* (2010) Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J. Mol. Biol.* 395, 916–923
21. Biela, A.P. *et al.* (2022) Programmable polymorphism of a virus-like particle. *Commun. Mater.* 3, 7
22. Rausch, A.O. *et al.* (2021) FrustratometeR: an R-package to compute local frustration in protein structures, point mutants and MD simulations. *Bioinformatics* 37, 3038–3040
23. Ferreiro, D.U. *et al.* (2011) On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3499–3503
24. Parra, R.G. *et al.* (2016) Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* 44, W356–W360
25. Freiberger, M.I. *et al.* (2021) Frustration in fuzzy protein complexes leads to interaction versatility. *J. Phys. Chem. B* 125, 2513–2520
26. Ferreiro, D.U. *et al.* (2007) Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19819–19824
27. Koning, R.I. *et al.* (2016) Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure *in situ*. *Nat. Commun.* 7, 12524
28. Dai, X. *et al.* (2017) *In situ* structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* 541, 112–116
29. Chandler-Bostock, R. *et al.* (2022) Genome-regulated assembly of a ssRNA virus may also prepare it for infection. *J. Mol. Biol.* 434, 167797

30. Eigen, M. *et al.* (1991) The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry* 30, 11005–11018
31. Jacques, D.A. *et al.* (2016) HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* 536, 349–353
32. Muller, T.G. *et al.* (2021) HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. *eLife* 10, e64776
33. Zila, V. *et al.* (2021) Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* 184, 1032–1046 e18
34. Li, C. *et al.* (2021) HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2019467118
35. Lahaye, X. *et al.* (2013) The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. *Immunity* 39, 1132–1142
36. Rasaiyaah, J. *et al.* (2013) HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* 503, 402–405
37. Schaller, T. *et al.* (2011) HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.* 7, e1002439
38. Luban, J. *et al.* (1993) Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* 73, 1067–1078
39. Lee, K. *et al.* (2010) Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* 7, 221–233
40. Price, A.J. *et al.* (2012) CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathog.* 8, e1002896
41. Price, A.J. *et al.* (2014) Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLoS Pathog.* 10, e1004459
42. Rebensburg, S.V. *et al.* (2021) Sec24C is an HIV-1 host dependency factor crucial for virus replication. *Nat. Microbiol.* 6, 435–444
43. Dickson, C.F. *et al.* (2023) Karyopherin mimicry explains how the HIV capsid penetrates nuclear pores. *bioRxiv* 2023.03.23.534032
44. Xue, G. *et al.* (2023) The HIV-1 capsid core is an opportunistic nuclear import receptor. *Nat. Commun.* 14, 3782
45. Shen, Q. *et al.* (2023) The capsid lattice engages a bipartite NUP153 motif to mediate nuclear entry of HIV-1 cores. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2202815120
46. Lu, M. *et al.* (2015) Dynamic allostery governs cyclophilin A-HIV capsid interplay. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14617–14622
47. Zuliani-Alvarez, L. *et al.* (2022) Evasion of cGAS and TRIM5 defines pandemic HIV. *Nat. Microbiol.* 7, 1762–1776
48. Link, J.O. *et al.* (2020) Clinical targeting of HIV capsid protein with a long-acting small molecule. *Nature* 584, 614–618
49. Yant, S.R. *et al.* (2019) A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model. *Nat. Med.* 25, 1377–1384
50. Rankovic, S. *et al.* (2021) HIV-1 uncoating occurs via a series of rapid biomechanical changes in the core related to individual stages of reverse transcription. *J. Virol.* 95, e00166-21
51. Christensen , D.E. *et al.* (2020) Reconstitution and visualization of HIV-1 capsid-dependent replication and integration *in vitro*. *Science* 370, eabc8420
52. Wang, J.C. *et al.* (2014) Encapsidated hepatitis B virus reverse transcriptase is poised on an ordered RNA lattice. *Proc. Natl. Acad. Sci. U. S. A.* 111, 11329–11334
53. Chen, M. *et al.* (2020) Protein structure refinement guided by atomic packing frustration analysis. *J. Phys. Chem. B* 124, 10889–10898
54. Lin, X. *et al.* (2019) Forging tools for refining predicted protein structures. *Proc. Natl. Acad. Sci. U. S. A.* 116, 9400–9409
55. Freiberger, M.I. *et al.* (2019) Local frustration around enzyme active sites. *Proc. Natl. Acad. Sci. U. S. A.* 116, 4037–4043
56. Chen, M. *et al.* (2020) Surveying biomolecular frustration at atomic resolution. *Nat. Commun.* 11, 5944
57. Chen, J. *et al.* (2019) Localizing frustration in proteins using all-atom energy functions. *J. Phys. Chem. B* 123, 4497–4504