White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/237977/

Version: Published Version

Article:
Twarock, Reidun and Stockley, Peter G (2026) Packaging Signal-mediated Assembly:How Viruses Outsmart Their Hosts. Journal of Molecular Biology. 169646. ISSN: 0022-2836

https://doi.org/10.1016/j.jmb.2026.169646

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.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Packaging Signal-mediated Assembly – How Viruses Outsmart Their Hosts

Reidun Twarock [1], **and Peter G. Stockley** [2,*]

*1 - **Departments of Mathematics and Biology,** University of York, York YO10 5GH, UK*
*2 - **Astbury Centre for Structural Molecular Biology,** University of Leeds, Leeds LS2 9JT, UK*

***Correspondence to Peter G. Stockley:** p.g.stockley@leeds.ac.uk (P.G. Stockley)*
https://doi.org/10.1016/j.jmb.2026.169646
Edited by Eric O. Freed

Abstract

The importance of the genetic code in virology is universally acknowledged. However, it is less known that viral genomes can harbour a second code, embedded within the genetic code, that orchestrates the efficient assembly and genome packaging in many viral systems. Since its discovery in a bacterial virus, the molecular details and function of this mechanism have been characterised in a broad range of viral families, including major human pathogens. This Perspective article reports on the hallmarks of this "assembly/packaging code", the journey of its discovery, and the enticing opportunities it brings both in antiviral therapy and in virus nanotechnology.

© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

We believe that all important concepts in Molecular Biology should advance our understanding of biological function. Hence the power of the double-helical structure of DNA in explaining heredity and genetics [1].

The same is true of simple, single-stranded RNA viral pathogens. They are ubiquitous in our environment – think of Human Rhinoviruses, the most common infections of humans [2], poliomyelitis [3] or Foot and Mouth Disease Virus [4], many of which are known to cause potentially lethal diseases in people or animals. However, at present we have few tools for their control other than vaccination for poliomyelitis and foot and mouth disease viruses [5–7]. Understanding their assembly mechanism at a molecular level reveals multiple essential steps that could be targeted for anti-viral therapy.

To alter this situation, we need a deeper understanding of the molecular mechanisms underpinning the behaviour of ssRNA virions. We started this process by addressing the questions of how these viruses assemble their virions efficiently in the arms race against their hosts' defence mechanisms. We have discovered that they achieve this feat via a common assembly mechanism, we term "RNA Packaging Signal-mediated Assembly" [8]. This mechanism relies on *multiple*, sequence-variable, sites with a shared core motif (we term these as Packaging Signals, PSs) that are distributed across positive-sense viral genomes in the form of stem-loops (SLs) [8–11]. Such sites act collectively in promoting efficient virion assembly and ensure packaging of the cognate gRNA against a backdrop of competitor cellular RNAs. This outcome is achieved via a distribution of different PS-CP affinities, that are tuned via a variation of the nucleotide sequences surrounding a consensus recognition motif [11]. Whilst the coat proteins (CPs) of these viruses can typically also assemble into protein shells in the absence of the gRNA, the assembly kinetics for such reactions are much slower, and the assembly outcomes are, in many cases, less defined. For example, CPs recognising, and interacting with, their cognate

0022-2836/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

gRNAs result in multiple effects: they create a defined nucleus for assembly, e.g. via a region within the genome encompassing multiple PSs, as in the gRNA of STNV-1 (Satellite Necrosis Virus-1, a $T = 1$ virion); they also prevent the formation of local kinetic traps, and they bias assembly to the cognate virion morphology. An example for the latter occurs in Hepatitis B Virus (HBV), where the PS:CP contacts alter the ratio of nuclear capsids to favour the infectious $T = 4$ from the smaller $T = 3$ shells [12,13]. Some of these functions can also, at least partially, be explained by a single *pac* site within viral genomes, but this does not account for the full spectrum of gains in assembly efficiency seen in both *in vitro* and *in vivo* experiments, and in the models of assembly cited above for the *multiple* packaging signal mediated assembly mechanism.

These mechanistic insights have far-reaching consequences for our understanding of simple viruses. For example, they explain why virion particles form only around the positive-sense version of their genomes, discriminating against both their negative-sense gRNA strands, and their obligatory dsRNA replicate intermediates. Negative-sense strands can form SLs at similar sites along the complements to their viral genomes, but they are distinct with respect to their folding free energies, as well as in the sequences of their loop motifs. The latter are a big driver of CP recognition. Recently, we have shown that multiple PS sites within the gRNA of HBV also prevent assembly competition by the transcriptome (*E. coli*), whereas non-viral RNAs are successfully packaged when we delete the major PS sites with the viral genome [13,14].

This assembly mechanism appears ubiquitous in all the positive-sense, single-stranded virions we have examined to date [8], including those of the entero- [15], and most recently the picorna-viral families [16]. The asymmetric cryo-EM structure of human rhinovirus, for instance, reveals a network of interactions between the capsid and duplex RNA fragments [16]. This provides insights into virus assembly and genome uncoating [17], attesting to the importance of this mechanism for our understanding of viral infections.

This hypothesis started out directed simply at understanding how a protein “recognises” a specific RNA sequence [18]. The RNA chosen for this task was the translational repression (TR) site from bacteriophage MS2, itself a multiple groundbreaking virus originally isolated from the Metropolitan Sewers of New York City and the first virion completely sequenced [19,20]. Using combinations of RNA (and chemical) synthesis, multiple variants of TR, a simple 19 nucleotide SL, were synthesised and purified *in vitro* [21]. In order to understand in detail what the molecular consequence(s) of complex formation between a phage coat protein (CP) dimer and TR was, we determined three-dimensional structures of those complexes using X-ray crystallography of virus-like particles (VLPs) [18,22], and also examined the interaction of TR RNA SLs and cognate CPs using single-molecule fluorescence spectroscopy, mass spectroscopy and modelling [23,24].

These studies confirmed the conclusions from genetic studies that CP complex formation at TR blocks ribosomal access to the replicase gene (cistron) on the phage genome. They also reveal that a conformational change in the CP dimer is imposed by this sequence-specific RNA-CP interaction, favouring formation of an asymmetric coat protein dimer (termed A/B), from what in solution in the absence of the cognate RNA oligonucleotide is a symmetrical CP dimer (C/C). The sequence-specific asymmetry of this complex explains why the bacteriophage CP does only very slowly assemble a VLP lacking its gRNA (on the order of days), or package non-phage RNAs. Only RNAs containing a stably-folded TR SL, or TR-like SLs, trigger this RNA-induced conformational switch sufficiently efficiently [25–27]. Since a specific ratio of C/C and A/B dimers is required for phage particle formation (30:60), this raises the question of whether other RNA SLs within the same gRNA are also likely to promote this reaction, however inefficiently. Using insights from the geometric organisation of the viral capsid in combination with bioinformatics and modelling, we were able to answer this question in the affirmative [10,28]. Not only are variants of TR functional, but more importantly, *multiple* distinct sites with a common core motif act *collectively* in promoting virion assembly. Since our initial discovery of multiple PSs in the bacteriophage MS2 gRNA over a decade ago, we have characterised similar RNA-CP interactions in both simple plant and animal (including human) viruses. These outcomes confirm the biological importance of the PS-mediated virion assembly mechanism in Virology. It is possible that this mechanism is present in even wider groups of ssRNA viruses, for which at present only a single PS is known, perhaps even in viruses such as HIV [29].

One might also then think about the process of infection more broadly, rather than simply the process of assembly of the virion CP shell (capsid). What motivates an inert virus particle to disassemble its protein shell to release its gRNA, and thus cause disease? Even though disassembly is not always a mere reversal of the assembly process, for example in more complex scenarios involving maturation events, it is important to understand how the CP-PS contacts contribute to both processes (see the Graphical Abstract, illustrating the roles of PSs at different

stages of a viral life cycle). CP-PS interactions vary dramatically in their strengths due to alterations in the underlying gRNA sequences at the PS sites. Thus, weaker PSs on viral gRNAs are very likely to dissociate first from the overlying CP layer *in virio* [30]. This creates a phenomenon known as "molecular frustration" in those parts of the CP shell that are no longer bound to RNA [28]. CP dimers lacking their cognate bound PS RNA ligands, like those in RNA bacteriophage MS2 [18], would be expected to alter their conformational preferences. However, they are kept in their defined, non-preferred, asymmetric conformational states because of their positions in the protein lattice of the virus-like particle. The result is a particle having "weaknesses" at defined positions that can be exploited by gRNAs to promote their exits into the "infected" cell from the viral particle at well-defined capsid sites. We have characterised this phenomenon using X-ray footprinting and asymmetric cryo-EM structure determination [15]. Within each virus there appears to be a gradient of PS-CP affinities allowing the stronger interactions to outcompete the weaker ones, explaining why part of the gRNA will tend to lose its affinity for CP after assembly into virions (where CP concentrations are high).

The multiply dispersed nature of the PSs, and their tolerance to sequence variation around a core motif allows virions to overcome the error-prone genome replication that is typical for ssRNA viruses [31]. However, the latter also complicates the development of anti-viral strategies, explaining why pharmacists' shops are not full of simple anti-viral remedies. It is very difficult to target such variable, multiply-dispersed phenomena with small molecular weight ligands that are more suited to inhibiting ligand-binding to a single defined protein target, like the active site of a protein enzyme. However, major diseases are caused by this class of virus, and to date the only hope of (pre-)treatment is vaccination with the viral protein shell. Our deeper understanding of the PS-mediated mechanism, together with recent advances in the development of small molecular weight compounds against RNA targets, is starting to turn the tables on this situation [12].

Modern medicine must also envisage the day when gene therapy, driven by understanding the vectors that currently move their own genetic material(s) into preferred hosts [32], are routinely adapted to customise delivery of oligonucleotides encoding preferred genetic characteristics. However, ssRNA viruses also face competition from the host transcriptome. We have evidence in HBV that multiple PS sites allow the virus to preferentially package its pre-genomic RNA over such competitor RNAs, but this is not the case for current applications. Once such packaging issues are resolved, the costs for the use of such technologies on the current human population at large are likely to become much more manageable [33]. Our deeper mechanistic understanding of *Nature*'s simple virions might be one way to achieve this outcome.

CRediT authorship contribution statement

Reidun Twarock: Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Peter G. Stockley:** Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

DATA AVAILABILITY

No data was used for the research described in the article.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received 8 October 2025;
Accepted 13 January 2026;
Available online 19 January 2026

Keywords:
virus;
assembly;
RNA;
packaging signals

References

[1]. Watson, J.D., Crick, F.H., (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171** (4356), 737–738.

[2]. Puenpa, J. et al, (2025). Seasonal dynamics and genetic diversity of human rhinoviruses in patients with acute respiratory infection in Bangkok in 2024. *Arch. Virol.* **170** (10), 208.

[3]. Tuells, J., Hurtado-Sánchez, A.P., (2025). The paradigm of an eliminated disease: sequelae and challenges to eradication. *J. Rev. Esp. Anestesiol. Reanim.*. https://doi.org/10.1016/j.redare.2025.500484.

[4]. Zha, J. et al, (2025). Foot-and-mouth disease virus vaccine with VP1 G-H loop substitution of the Cathay strain broadens antigen spectrum. *Appl. Microbiol. Biotechnol.* **109** (1), 199.

[5]. Salk, D., Salk, J., (1984). Vaccinology of poliomyelitis. *Vaccine* **2**, 59–74.

[6]. Sabin, A.B., (1984). Oral poliovirus vaccine: history of its development and use and current. *J. Infect. Dis.* **151** (3), 420–436.

[7]. Belsham, G.J., (2020). Towards improvements in foot-and-mouth disease vaccine performance. *Acta Vet. Scand.* **62** (1), 20.

[8]. Twarock, R., Stockley, P.G., (2019). RNA-mediated virus assembly: mechanisms and consequences for viral evolution and therapy. *Ann. Rev. Biophys.* **48**, 495–514.
[9]. Borodavka, A. et al, (2012). Evidence that viral RNAs have evolved for efficient, two-stage packaging. *PNAS* **109**, 15769–15774.
[10]. Dykeman, E.C., Stockley, P.G., Twarock, R., (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.
[11]. Dykeman, E.C., Stockley, P.G., Twarock, R., (2014). Solving a Levinthal's paradox for virus assembly suggests a novel anti-viral therapy. *PNAS* **111**, 5361–5366.
[12]. Patel, N. et al, (2024). Dysregulation of hepatitis B virus nucleocapsid assembly in vitro by RNA-binding small ligands. *J. Mol. Biol.* **434**, (10)167557
[13]. Patel, N., Clark, S., Weiß, E.U., Mata, C.P., Bohon, J., Farquhar, E.R., Maskell, D.P., Ranson, N.A., Twarock, R., Stockley, P.G., (2021). *In vitro* functional analysis of gRNA sites regulating assembly of hepatitis B virus. *Commun. Biol.* **4** (1), 1407.
[14]. Routh, A., Domitrovic, T., Johnson, J.E., (2012). Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *PNAS* **109**, 1907–1912.
[15]. Chandler-Bostock, R. et al, (2020). Assembly of infectious enteroviruses depends on multiple, conserved genomic RNA-coat protein contacts. *PLoS Pathog.* **16**, (12) e1009146
[16]. Gil-Cantero, D. et al, (2024). Cryo-EM of human rhinovirus reveals a network of interactions between the capsid and duplex RNA regions that provides insights into virus assembly and genome uncoating. *Commun. Biol.* **7**, 1501–1511.
[17]. Valiente et al, (2022). Molecular determinants of human rhinovirus infection, assembly, and conformational stability at capsid protein interfaces. *J. Virol.* **96**, (23) e0084022
[18]. Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J., Liljas, L., (1994). Crystal-structure of a bacteriophage-RNA coat protein-operator complex. *Nature* **371**, 623–626.
[19]. Loeb, T., Zinder, N.D., (1961). A bacteriophage containing RNA. *PNAS* **47**, 282.
[20]. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min, J.W., Molemans, F., Raeymaekers, A., Van den Berghe, A., (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260** (5551), 500–507.
[21]. Adams, C.J., Murray, J.B., Arnold, J.R.P., Stockley, P.G., (1994). Incorporation of a fluorescent nucleotide into oligoribonucleotides. *Tetra Let.* **35**, 1597–1600.
[22]. Grahn, E., Stonehouse, N.J., Murray, J.B., Van den Worm, S., Valegård, K., Fridborg, K., Stockley, P.G., Liljas, L., (1999). Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements. *RNA* **5**, 131–138.
[23]. Stockley, P.G., Ashcroft, A.E., Francese, S., Thompson, G.S., Ranson, N.A., Smith, A.M., Homans, S.W., Stonehouse, N.J., (2005). Dissecting the fine details of assembly of a $T = 3$ phage capsid. *Comput. Math. Methods Med.* **6**, 119–125.
[24]. 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.
[25]. Rolfsson, O., Toropova, K., Morton, V., Francese, S., Basnak, G., Thompson, G.S., Homans, S.W., Ashcroft, A. E., Stonehouse, N.J., Ranson, N.A., Stockley, P.G., (2008). RNA packing specificity and folding during assembly of the bacteriophage MS2. *Comput. Math. Methods Med.* **9** (3–4), 339–349.
[26]. Morton, V.L., Dykeman, E.C., Stonehouse, N.J., Ashcroft, A.E., Twarock, R., Stockley, P.G., (2010). The impact of viral RNA on assembly pathway selection. *J. Mol. Biol.* **401**, 298–308.
[27]. Dykeman, E.C., Grayson, N.E., Toropova, K., Ranson, N. A., Stockley, P.G., Twarock, R., (2011). Simple rules for efficient assembly predict the layout of a packaged viral RNA. *J. Mol. Biol.* **408**, 399–407.
[28]. Twarock, R., Leonov, G., Stockley, P.G., (2018). Hamiltonian path analysis of viral genomes. *Nature Commun.* **9**, 2021.
[29]. Bergland, J.A., Charpentier, B., Rosbash, M., (1997). A high affinity binding site for the HIV-1 nucleocapsid protein. *Nucleic Acids Res.* **25**, 1042–1049.
[30]. Wroblewski et al, (2024). Visualizing viral RNA packaging signals in action. *J. Mol. Biol.* **435**, 168765.
[31]. Mizuuchi, R., Furubayashi, T., Ichihashi, N., (2022). Evolutionary transition from a single "RNA replicator" to a multiple "replicator" network. *Nature Commun.* **13** (1), 1460.
[32]. Herchenröder, O., Pützer, B.M., (2024). Novel viral vectors for gene therapy. *Viruses* **16** (3), 387.
[33]. Gote, V., Bolla, P.K., Kommineni, N., Butreddy, A., Nukala, P.K., Palakurthi, S.S., Khan, W., (2023). A comprehensive review of mRNA vaccines. *Int. J. Mol. Sci.* **24** (3), 2700.