Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2

Óttar Rolfsson[1,†], Stefani Middleton[2,3,†], Iain W. Manfield[1], Simon J. White[1], Baochang Fan[2], Robert Vaughan[2], Neil A. Ranson[1], Eric Dykeman[4], Reidun Twarock[4], James Ford[3], C. Cheng Kao[2] and Peter G. Stockley[1]

1 - Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom
2 - Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN 47405, USA
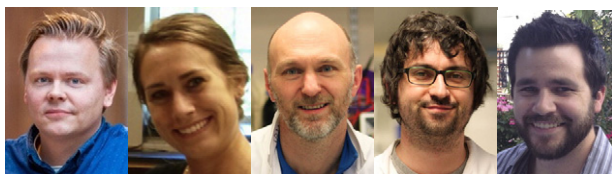3 - The Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN 47405, USA
4 - Department of Biology and Mathematics and York Centre for Complex Systems Analysis, University of York, York YO10 5DD, United Kingdom

Correspondence to :
http://dx.doi.org/10.1016/j.jmb.2015.11.014
Edited by E. O. Freed

Óttar Rolfsson, Stefani Middleton, Iain W. Manfield, Simon J. White and Robert Vaughan

Abstract

Using cross-linking coupled to matrix-assisted laser desorption/ionization mass spectrometry and CLIP-Seq sequencing, we determined the peptide and oligonucleotide sequences at the interfaces between the capsid proteins and the genomic RNA of bacteriophage MS2. The results suggest that the same coat protein (CP)–RNA and maturation protein (MP)–RNA interfaces are used in every viral particle. The portions of the viral RNA in contact with CP subunits span the genome, consistent with a large number of discrete and similar contacts within each particle. Many of these sites match previous predictions of the locations of multiple, dispersed and degenerate RNA sites with cognate CP affinity termed packaging signals (PSs). Chemical RNA footprinting was used to compare the secondary structures of protein-free genomic fragments and the RNA in the virion. Some PSs are partially present in protein-free RNA but others would need to refold from their dominant solution conformations to form the contacts identified in the virion. The RNA-binding peptides within the MP map to two sections of the N-terminal half of the protein. Comparison of MP sequences from related phages suggests a similar arrangement of RNA-binding sites, although these N-terminal regions have only limited sequence conservation. In contrast, the sequences of the C-termini are highly conserved, consistent with them encompassing pilin-binding domains required for initial contact with host cells. These results provide independent and unambiguous support for the assembly of MS2 virions via a PS-mediated mechanism involving a series of induced-fit viral protein interactions with RNA.

Introduction

Spherical viral capsids assemble by two distinct mechanisms [1]. Many double-stranded DNA

phages and Herpesviruses form a protein-only shell of coat and scaffold proteins, a pro-capsid, into which the genome is actively pumped via specialzsed energy-requiring enzymes as the scaffold proteins are removed. Spherical, positive-sense, single-stranded RNA (ssRNA) viruses, which include many major pathogens in all kingdoms of life, assemble by the co-assembly of coat protein (CP) subunits around the genome [2]. The free energy for this process has been thought to come mostly from favorable electrostatic interactions between the negatively charged RNA and positively charged residues and domains in the CPs [3–12]. Adding to this view are the observations from *in vitro* reassembly experiments showing that many viral CPs can assemble into virus-like particles in the absence of RNA, in the presence of non-cognate RNAs, or even anionic polymers. High-resolution X-ray structures of ssRNA viruses mostly reveal only a limited ordering of the encapsidated genome, again implying that RNA has no, or very little, overt roles in virion assembly [13–17]. However, the specificity of genome packaging is incompletely explained by such an assembly mechanism. Virions from natural infections predominantly encapsidate their cognate genomes, with only minor misincorporation of cellular or degraded RNAs [18,19]. These outcomes hint at more complex regulation of the assembly process.

Most families of these viruses encode single-copy, high-affinity CP-binding sites in their genomes that are thought to act as assembly initiation sites [20,21]. These RNA sequences form defined secondary structure elements that are specifically recognized by the viral CPs. The best characterized of these interactions is the 19-nt-long stem–loop named TR in the MS2 genome (Fig. 1) that acts as a translational operator for the replicase cistron [22]. This also functions as the assembly initiation site, both *in vitro* and *in vivo* [23,24]. Mutational disruption of this packaging signal (PS), however, does not result in discernable defects in viral infectivity or virion formation, although similar mutation of the TR recognition site in the CP ablates both these events [25]. This result led to the suggestion that RNA sites other than TR act as the site of assembly initiation in its absence, providing a robust assembly mechanism. Previously, we showed (Fig. 1) that TR–CP interaction has another consequence, namely, that binding of the stem–loop biases the conformational equilibrium of the CP dimer away from the symmetric, RNA-free C/C quasi-conformer and favors the asymmetric A/B dimer [26,29,32]. The mechanism underpinning this allosteric conformer switch was investigated using all-atom normal mode analysis [33,34]. The results suggest that many RNA stem–loops with similar structural features to TR could trigger similar changes; that is, although translational repression is highly sequence specific, the allosteric effect is much less so. This both explains the mutagenesis data and suggests that up to 60 such

stem–loops might lie within the MS2 genome, this being the number of A/B quasi-conformer dimers required to construct the *T* = 3 capsid. Electron microscopy (EM) reconstruction of the virion is also consistent with this concept revealing extensively ordered genome segments in contact with the inner surface of the CPs (Fig. 1b, right half). Since the X-ray structures [27,28,35] of MS2 and other RNA phages do not show much density for the RNA, the previous inferences based on the apparent absence of such contacts can be also be ignored.

These ideas led to the introduction of a novel model for virion assembly based on the presence of multiple, degenerate and dispersed RNA sequence motifs/structures called PSs that bind their cognate CPs. This concept explains results from single-molecule fluorescence assays of *in vitro* assembly of the model viruses bacteriophage MS2 and STNV [36–39]. Highly cooperative, selective RNA packaging was observed to occur *in vitro* at nanomolar concentrations, consistent with the existence of multiple dispersed stem–loops in these genomes with recognition motifs similar to the known high-affinity PSs [40,41]. Mathematical modeling of this mechanism suggests that it provides significant selective advantages for the viruses that use it, including solving the viral equivalent of Levinthal's paradox [42,43]. Therefore, PS-mediated assembly could be common for viruses that infect many different types of host, although the precise molecular effects arising from PS–CP contacts could be distinct in each case. An essential feature of the PS model is that the dispersed PSs have widely differing CP affinities in order to regulate the number of initiation events occurring and avoid kinetic traps. This implies that PS sequences and structures will be variable and difficult to detect by sequence searches alone. For MS2 (Fig. 1), we have determined the molecular details of sequence-specific recognition of preferred RNA sequences/structures bound by its capsomere [35,44–54], a CP dimer (CP_2), using X-ray crystallography, and we have investigated the PS-mediated assembly mechanism using mass spectrometry [55], analytical ultracentrifugation and smFCS (*single-molecule fluorescence correlation spectroscopy*) [26,36]. Combined with affinity data [56] and the experimentally determined solution structure of the genome [57,58], we predicted the sequences and locations of its PSs [41], identifying a large percentage of the putative 60 copies required to regulate assembly of the *T* = 3 shell (Fig. 1).

In order to confirm the presence of multiple PSs within the MS2 genome, we have now determined the contact points between the proteins and RNA components of intact virions. The results show that the previous predictions of PS sites in contact with the phage CP were remarkably accurate, confirming unambiguously the PS-mediated assembly of this virion. The data also allow us to assign the functions of the single-copy/virion maturation protein (MP) to
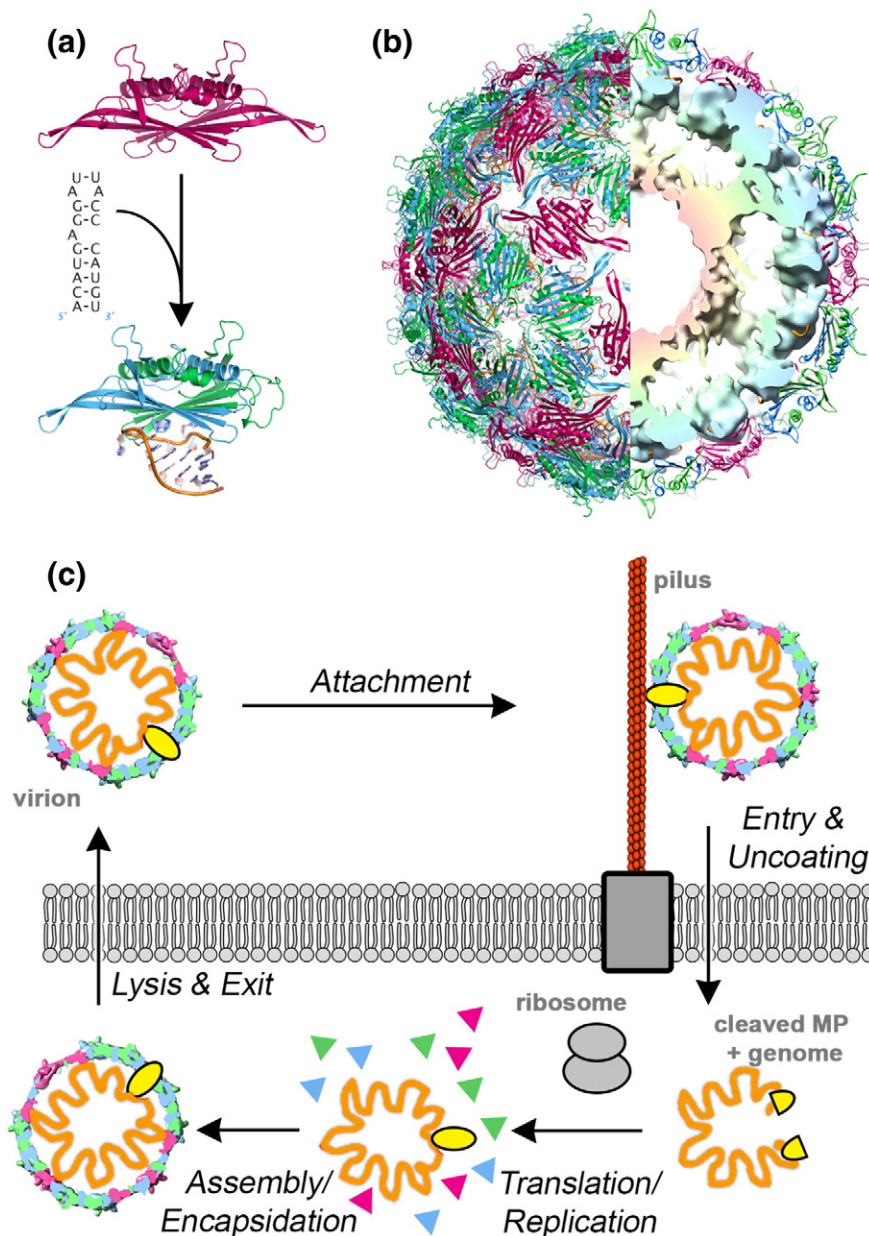
Fig. 1. Components of the MS2 virion and its lifecycle. (a) Ribbon representations of the two quasi-equivalent MS2 CP dimers: The symmetric dimer (C/C) is colored pink, while the asymmetric (A/B) is blue/green. Binding the TR high-affinity PS, sequence and secondary structure triggers conformer switching of C/C to A/B [26]. In principle, 60 such conformer switching events are required to create the $T = 3$ capsid shown in (b). (b) Structure of the MS2 virion. Left, surface view of the icosahedrally averaged X-ray structure of the $T = 3$ MS2 virion [27,28]. Right, cutaway schematic of the equivalent EM reconstitution at 9 Å resolution [29], showing the extensive contacts between the genome, density radially colored pink to blue as the radius increases and the overlying protein shell. X-ray coordinates for these images were taken from PDB ID 1ZDH. (c) Schematic of the phage lifecycle. Virions initially bind to the sides of the bacterial pilus via the MP. MPs are an essential single-copy structural component of RNA phages. By a mechanism that remains largely obscure, the RNA–MP complex but not the remaining capsid shell enters the cell. MP also binds to its own PSs located toward either end of the MS2 genome [30]. Recent asymmetric reconstruction of the MS2–pilus complex suggests that MP replaces a CP dimer of the C/C conformer in an otherwise entirely icosahedral protein shell [31], from which position it is ideally placed to contact the cellular receptor and escort the RNA into the cytoplasm. Note that the presence of the asymmetric MP component could not be detected in averaged X-ray and EM density maps. Once internalized, the MP is cleaved into two separate fragments by protease, allowing translation and replication to start. Temporal control of phage gene expression then regulates the production of progeny genomes and structural proteins that assemble prior to the action of the phage lysis protein.

separate domains (Fig. 1), the N-terminal domain seemingly encompassing the RNA-binding function while the pilus-binding function lies within the C-terminal portion. Sequence comparisons with other phage MPs suggest that this is a conserved arrangement. In addition, RNA structure probing has been used to monitor the presence of PS sites in CP-free genome transcripts or the virion, revealing that PSs fall into at least two classes, those that are present at least some of the time in the protein-free RNA and those that must refold as assembly proceeds. An ordered series of induced-fit interactions, similar to those occurring during ribosome assembly [59–62], likely accounts for MS2 virion formation and presumably for assembly of other ssRNA viruses using the PS-mediated mechanism.

Results

Mapping the RNA contacts on both CP and MP subunits

The genomic RNA footprint on the capsid proteins within the virion was determined by an RNA cross-linking and peptide mapping assay (RCAP), used previously to map the interaction between the capsid and the nucleic acid of brome mosaic virus [63]. Briefly, the RNA was cross-linked to the virion proteins in amine-free buffer by treatment with formaldehyde, followed by digestion with proteases. RNA and RNA–peptide complexes were then precipitated with lithium chloride, the cross-links were reversed and the released peptides were identified by MALDI (*matrix-assisted laser desorption/ionization*)-ToF (*time of flight*) mass spectrometry (Fig. 2a). Control reactions that lacked formaldehyde did not yield significant amounts of recovered peptides. In contrast, several peptides were detected in the cross-linked samples. Peptides within 0.5 Da of their theoretical molecular mass were assigned to those generated from theoretical digests of the MS2 CP or MP proteins (Fig. 2b and/or c, Fig. 3a, Supplementary Fig. 1 and Supplementary Table 1). There is an extensive overlap in the peptides identified by digestion either with trypsin alone or with trypsin and Glu-C (Fig. 2b). Arginines or lysines that remain uncleaved in these peptides are assumed to be the sites of cross-linking to the RNA.

MS2 virion peptides identified using RCAP are shown mapped onto the amino acid sequences of both the CP (Fig. 2c) and the MP (Fig. 3a). The CP peptides in contact with the genomic RNA encompass all the amino acid residues that contact TR in VLP crystal structures (Supplementary Fig. 1), including the four β-strands (C–F) that form the interior face of the CP subunit. Importantly, there are no RNA contacts that map outside the known
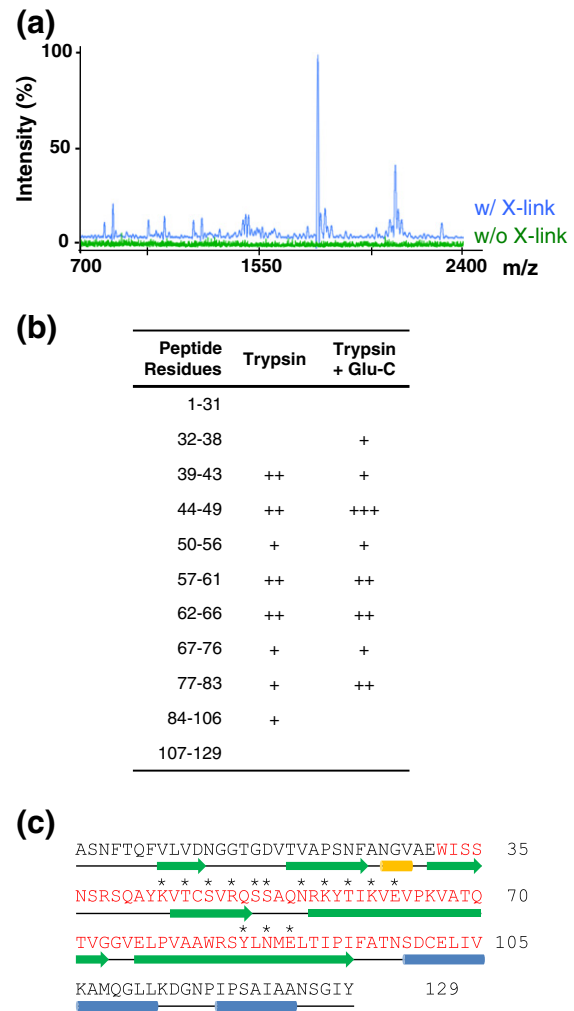

Fig. 2. RCAP of the MS2 virion. (a) MALDI-ToF spectra of trypsin-generated peptides co-purified with the MS2 RNA (Materials and Methods) from samples mock-treated or cross-linked with formaldehyde. (b) Residue numbers of CP peptides recovered in experiments similar to those shown in (a) for the different protease treatments. "+" indicates the intensity of each peptide segment. (c) MS2 CP peptides in contact with the genome in the virion. Peptides assigned following RCAP are shown in red and comprise the entire region from amino acid 32 to amino acid 105, which form the four β-strands, C, D, E and F, facing the interior of the virion. The secondary structure elements (PDB ID 2MS2; Fig. 1a) are represented by green arrows, blue bars and a gold bar for β-strands, α-helices and a 3_{10} helix, respectively. Residues contacting the TR PS are shown with an asterisk above (Fig. 1b).

RNA-binding site. This is consistent with the idea that CP binding to the genomic RNA occurs along the same RNA-binding face of each CP, as expected for a PS-mediated assembly mechanism.

The RNA–MP complex is the only part of the virion to enter the host cell during infection where the MP undergoes proteolytic cleavage into two fragments of 15 and 24 kDa that are presumed to be separate

globular domains [64]. The MP RCAP data (Fig. 3a) are consistent with this idea. All the RNA-contacting peptides are localized to the first 233 amino acid residues of the 393-residue protein, implying that it is orientated in the virion in the C- to N-terminal direction from out to in. This is also consistent with a prediction of RNA-binding functional domains [65], which identifies two motifs located between amino acids 52–60 and amino acids 162–170 (Fig. 3a) that are within peptides detected via RCAP. Sequence comparison with the MPs from other phages reveals
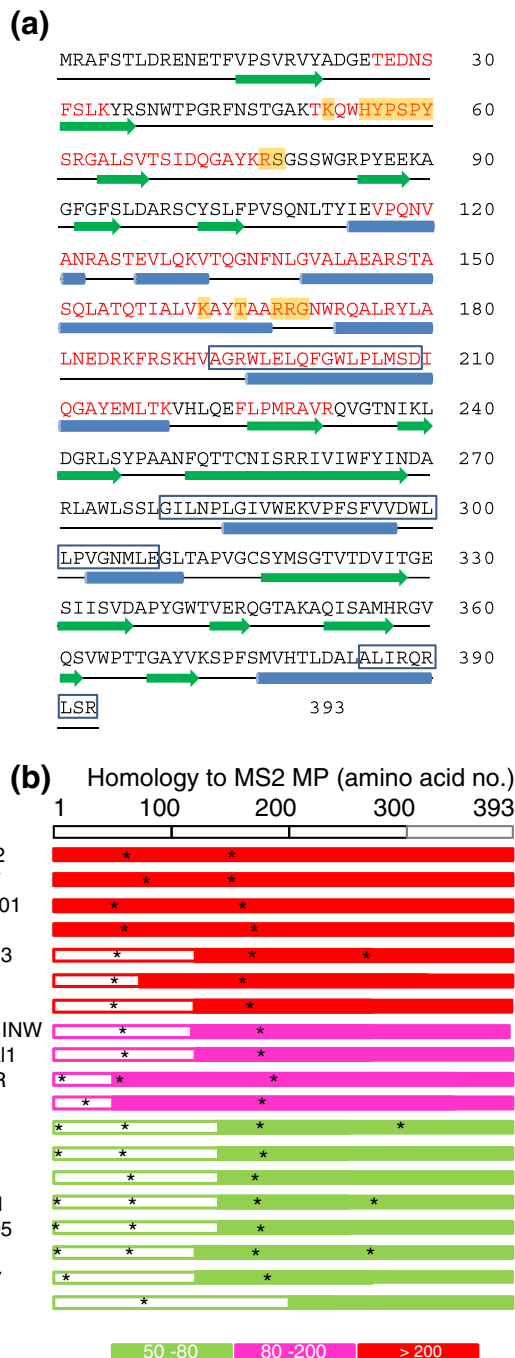
a conserved pattern consistent with this organization (Fig. 3b). The MPs of phages in the Levivirus or Allolevivirus families are, respectively, predicted to encode RNA-binding motifs around amino acids 50–70 and 160–180 despite considerable sequence divergence over the N-terminal ~140 residues. In contrast, the C-terminal fragments are highly conserved, perhaps reflecting the requirement to bind the same or very similar pilin targets.

Mapping the CP contacts on the viral genome

RNA sequences in contact with the MS2 CP subunits were identified using CLIP-Seq [63]. The genomic RNA was cross-linked to virion proteins using UV irradiation (Fig. 4a). The virion was then dissociated and the protein-bound RNA fragments were minimized by hydrolysis in the presence of zinc ions to produce fragments in the ~25- to 40-nt range, based on conditions established for a polyAU control (Fig. 4b). RNA fragments cross-linked to the CP were then concentrated by immunoprecipitation with an anti-CP-specific serum. In the absence of MP-specific antibodies, it is not possible to do the equivalent experiment for RNA fragments in contact with MP. The extent of cross-linking was assessed by SDS-PAGE. UV irradiation produces a small amount of cross-linking between CP dimers, compared to the non-irradiated control (Fig. 4c); however, quantification of radiolabeled RNA revealed that it co-migrated with the CP and that there is an ~60-fold higher signal for CP-associated RNA in the irradiated sample (Fig. 4d). The material in these regions of the gel for both irradiated and control samples was then released from the membrane by treatment with proteinase K, TRIzol™ extraction, was ethanol precipitated and was subjected to NextGen sequencing (Illumina MiSeq).

(a)

```
MRAFSTLDRENETFVPSVRVYADGETEDNS      30

FSLKYRSNWTPGRFNSTGAKTKQWHYPSPY      60

SRGALSVTSIDQGAYKRSGSSWGRPYEEKA      90

GFGFSLDARSCYSLFPVSQNLTYIEVPQNV     120

ANRASTEVLQKVTQGNFNLGVALAEARSTA     150

SQLATQTIALVKAYTAARRGNWRQALRYLA     180

LNEDRKFRSKHVAGRWLELQFGWLPLMSDI     210

QGAYEMLTKVHLQEFLPMRAVRQVGTNIKL     240

DGRLSYPAANFQTTCNISRRIVIWFYINDA     270

RLAWLSSLGILNPLGIVWEKVPFSFVVDWL     300

LPVGNMLEGLTAPVGCSYMSGTVTDVITGE     330

SIISVDAPYGWTVERQGTAKAQISAMHRGV     360

QSVWPTTGAYVKSPFSMVHTLDALALIRQR     390

LSR                                393
```

(b) Homology to MS2 MP (amino acid no.)

Fig. 3. Roles of the MP. (a) Sequence of the MP with the peptides identified via RCAP highlighted in red. The Jpred3-predicted secondary structure elements of the MP are shown using the same format as for the CP (Fig. 2c). Predicted RNA-binding residues (RNABindR-Plus) are highlighted in orange. Peptides that are highly conserved across both Leviviridae and Alloleviviridae phage MPs are boxed (motif 1, amino acids 193–210, 51% average identity over 18 amino acids; motif 2, amino acids 279–308, 56% average identity over 30 amino acids; motif 3, amino acids 385–393, average 52% identity over 9 amino acids). These regions may therefore be part of the pilin-binding site. (b) Sequence homology of Leviviridae MP sequences. Filled bars represent regions homologous to MS2 MP and open boxes represent non-homologous regions. These bars cover the full length of each protein, without indicating the short gaps required to accommodate the alignment. Most proteins in this comparison are ~400 amino acids long. Colors represent the homology score using the BLOSUM62 matrix. Potential RNA-binding sites predicted by RNABindRPlus are represented by asterisks.
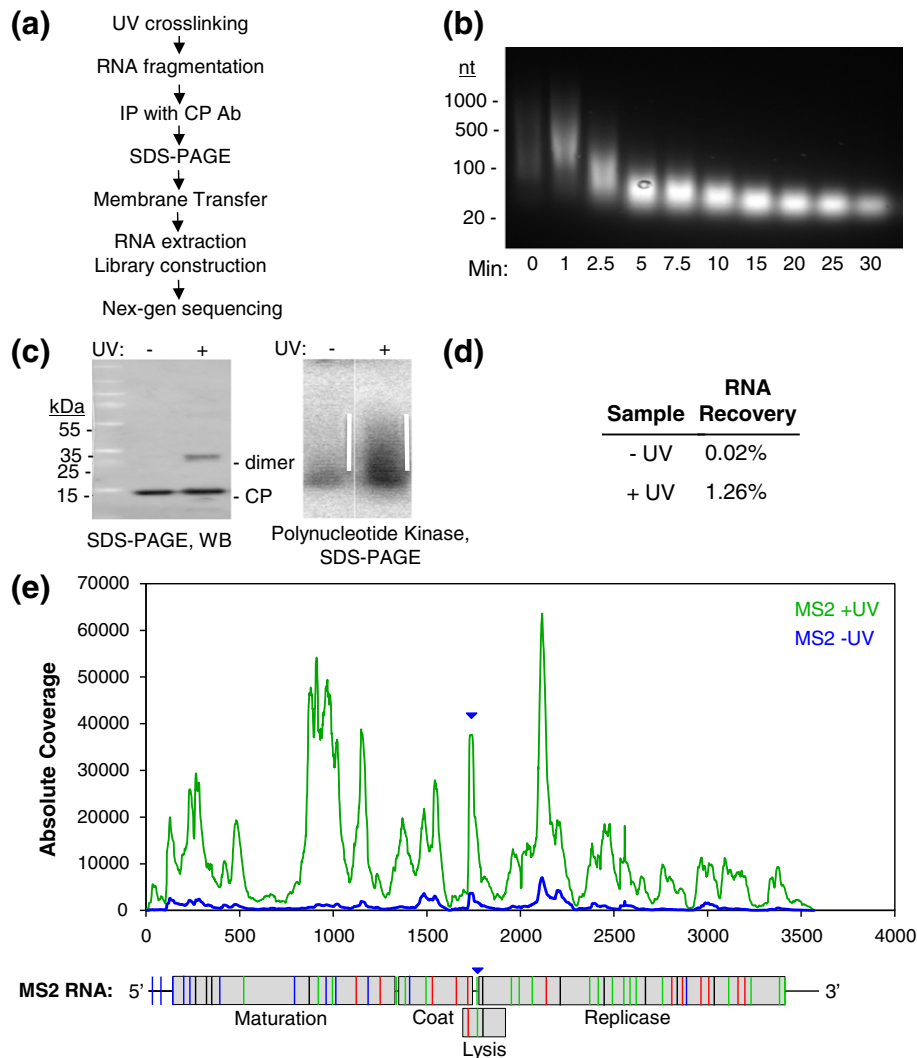
(a) UV crosslinking
↓
RNA fragmentation
↓
IP with CP Ab
↓
SDS-PAGE
↓
Membrane Transfer
↓
RNA extraction
Library construction
↓
Nex-gen sequencing

(b)

(c)

SDS-PAGE, WB   Polynucleotide Kinase,
                      SDS-PAGE

(d)

Sample	RNA Recovery
- UV	0.02%
+ UV	1.26%

(e)

Fig. 4. RNA sequences in the MS2 virion that contact the capsid proteins. (a) Outline of the CLIP-Seq protocol used to map RNA residues that contact the MS2 capsid proteins. (b) PAGE analysis of a time course of RNA cleavage using 10 mM $ZnCl_2$ with poly(A:U) RNA. (c) SDS-PAGE gel of immune-precipitated MS2 CP containing covalently linked RNA treated with zinc ions. Irradiated and control samples were transferred to a nitrocellulose membrane and subsequently visualized by a Western blot with an anti-MS2 CP primary antibody (left-hand panel). The regions of the same gel containing cross-linked RNA were identified in parallel samples by prior 5′ end labeling of the RNA fragments with polynucleotide kinase (right-hand panel). The smear likely represents cross-linked CP–RNA. The white lines identify the regions excised from the membrane and processed for deep sequencing (see Materials and Methods). (d) Amount of input RNA recovered with CP after purification. (e) Abundances of cDNA fragments that co-purified with the MS2 CP. The sequences for the cDNAs identified by Illumina DNA sequencing were aligned with the MS2 genome, and histograms were produced denoting the frequency of particular sequences within the datasets for irradiated (green) and control (blue) samples. The peak that corresponds to the MS2 operator hairpin is identified by an inverted blue triangle. A schematic of the MS2 genome is shown below the graph to allow identification the approximate locations of the most significant peaks. Lines in blue highlight those sites matching previous PS predictions by Dykeman *et al.* [41]; those in orange, similarly for Bleckley *et al.* [66]; and green, for matches to both predictions. Black lines indicate the 10 peaks not predicted previously.

The individual MS2 sequence reads were converted into sense-strand RNA and mapped onto the genomic sequence of the phage creating a histogram of hits at each nucleotide for both irradiated (green) and control (blue) samples, yielding Fig. 4e. The frequency of matched nucleotides varies by factors of up to 6000 across the genome and, at every position, is higher in the cross-linked sample than at the same position in the control. In the sequences from the non-cross-linked sample, some enrichment of the RNAs is apparent. In some places, a small peak in the non-irradiated sample appears to

match a peak in the irradiated sample but at a much lower frequency. However, this is clearly not the case for the majority of the peaks in the irradiated sample, confirming that specific enrichment had occurred. The background may be due to some RNA sequences or conformations being recognized by the capsid with higher affinity or that some RNA fragments may precipitate better than others. This is one reason why the peak heights cannot be interpreted simply in terms of RNA–CP affinity.

The pattern of peaks, some of which are clearly symmetrical implying discrete sites of contact, is consistent with the presence of the same series of discrete RNA oligonucleotides in contact with the capsid in every virus particle, that is, that the highest matching nucleotide in each peak should be within or adjacent to a PS. A random arrangement of the genomic RNA within each capsid would not be expected to yield this distribution. Importantly, the level of sequencing multiplicity in the MiSeq data (Supplementary Table 2) suggests that they provide an accurate sampling of the virion population; thus, the peaks represent the most frequently cross-linked fragments. Unfortunately, most peaks are not well resolved from their neighbors, making such direct identification more difficult, but the TR site is clearly within one of these higher peaks (see arrow in Fig. 4e) as expected. In order to analyze these peaks further, we identified a significance threshold (3658; see Materials and Methods) for the number of hits required to generate a significant peak. This results in 54 peaks (Table 1), and their locations are mapped onto the MS2 genome below the histogram of Fig. 4e.

Since cross-linking can occur at any point of contact between a PS fragment and a CP subunit and the rate of zinc ion cleavage is locally RNA sequence/structure dependent, the peak nucleotides do not necessarily coincide with the central nucleotides of a PS. Another problem is the nature of the pile-up analysis could produce a misleading single peak due to overlapping sequence reads from adjoining fragments. In order to take account of these issues, Table 1 lists both the peak nucleotide of each of the 54 peaks and the 20 nucleotides of the genomic flanking sequences on each side of it, thus creating a fragment of about the average probe size that was sequenced. If these fragments encompass PSs, they would be expected to form a stem–loop. Mfold was therefore used to explore the potential secondary structures for each fragment and any stem–loops that can form compared to previous predictions of the PS sites [41]. Sequences capable of forming stem–loops are shown in red in Table 1, with those nucleotides predicted to base pair being underlined. This analysis rapidly identified a large number of potential PSs within the CLIP-Seq fragments. In addition, it revealed several fragments that encompass two potential PSs, presumably for

the reasons set out above. It also suggested that the PSs may extend beyond the boundaries of the 41 nucleotides listed, and the regions of these sites within such CLIP-Seq fragments are shown in Table 1. This misalignment could be a consequence of mis-identification of peaks due to their overlaps (Fig. 4e).

There have been two previous attempts to predict the locations of the PSs in the MS2 genome. We used information on the preferred RNA-binding sites for CP from X-ray crystal structures and SELEX together with the reported solution structure of CP-free RNA to create a minimal RNA motif. This yielded 54 of the expected 60 PSs required in this $T = 3$ virus. The CLIP-Seq fragments encompass 33/54 sites predicted in this manner. The Schroeder laboratory has also used their RNA folding algorithm CRUMPLE [66] with a slightly more relaxed probe motif to predict PS sites. The CLIP-Seq dataset has 32 matches to their 60 stem–loop predictions. The matches to the different predictions only partially overlap, meaning that, together, they account for 44/54 experimentally determined sites. Mfold of these remaining 10 orphan sites fails to identify stable folds that are likely PS sites; thus, their presence in the dataset is confusing but could be due to artifacts in aligning CLIP-Seq peaks within the histogram. The multiple matches between the observed CP–RNA contacts in the virion and the predicted PS sites are, however, excellent evidence in support of PS-mediated assembly. It is interesting to examine why neither prediction yielded a complete match to the experimental data, but this requires a more detailed discussion that will be described separately in Dykeman *et al.* (unpublished results).

The CLIP-Seq data also contain a sequence previously reported as one of two RNA-binding sites of the MP, Peak 8 (Table 1) corresponding to the 5′ site. The previous identification of these sites relied on direct sequencing of the RNA fragments resulting from RNase A digestion of the acid-insoluble MP–RNA complex that can be isolated from the virion [30]. This identified two sites, one toward each end of the genome, consistent with a role for the MP in circularizing the RNA. Since the RNA–protein complexes here were recovered by anti-CP antibodies, this suggests that there is a CP-binding site adjacent to the MP-binding site, consistent with the known location of the MP in the virion, where it replaces one of the C/C CP dimers [31].

Identification of classes of PSs

The experiments described above provide direct evidence that MS2 assembles via a PS-mediated mechanism, but they do not address how the CPs locate their respective PS sites during virion assembly. The latter cannot always be present in the active form of a stem–loop due to the many roles the

Table 1. Flanking sequences of the MiSeq peaks and their relationship to previous predicted PSs. Matches to the predictions of Dykeman *et al.* [41] are highlighted in blue in the peak nucleotide. Column, those to Bleckley *et al.* [66], is in orange and with matches to both in green. There are 10 CLIP-Seq sites that do not match any prediction (no color). Predicted hairpins are underlined. Predicted MS2 CP-binding motifs are in red. The peak column relates to the peaks identified in Fig. 4e with peak 1 nearest the 5′ end of the genome.

Peak	Peak nt.	Peak Sequence
1	35	UCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUUU
2	51	CUUCCUGUCGAGCUAAUGCCAUUUUUAAUGUCUUUAGCGAG
3	128	CCAUUCCUAGGAGGUUUGACCUGUGCGAGCUUUUAGUACCC
4	230	AGAUAACUCAUUCUCUUUAAAAUAUCGUUCGAACUGGACUC
5	266	GACUCCCGGUCGUUUUAACUCGACUGGGGCCAAAACGAAAC
6	281	UAACUCGACUGGGGCCAAAACGAAACAGUGGCACUACCCCU
7	352	CAUCGAUAGAUCAAGGUGCCUACAAGCGAAGUGGGUCAUCG
8	419	CGGUUUCGGCUUCUCCCUCGACGCACGCUCCUGCUACAGCC
9	474	CAAAACUUGACUUACAUCGAAGUGCCGCAGAACGUUGCGAA
10	770	GAGTGATATCCAGGGUGCAUAUGAGAUGCUUACGAAGGUUC
11	803	GAAGGUUCACCUUCAAGAGUUUCUUCCUAUGAGAGCCGUAC
12	872	UGGCCGUCUGUCGUAUCCAGCUGCAAACUUCCAGACAACGU
13	910	CGUGCAACAUAUCGCGACGUAUCGUGAUAUGGUUUUACAUA
14	951	AACGAUGCACGUUUGGCAUGGUUGUCGUCUCUAGGUAUCUU
15	983	AGGUAUCUUGAACCCACUAGGUAUAGUGUGGGAAAAGGUGC
16	1020	GUGCCUUUCUCAUUCGUUGUCGACUGGCUCCUACCUGUAGG
17	1112	AACAGUUACUGACGUAAUAACGGGUGAGUCCAUCAUAAGCG
18	1149	AGCGUUGACGCUCCCUACGGGUGGACUGUGGAGAGACAGGG
19	1242	CCAACAACUGGCGCGUACGUAAAGUCUCCUUUCUCGAUGGU
20	1325	UAGAUAGAGCCCUCAACCGGAGUUUGAAGCAUGGCUUCUAA
21	1372	UCAGUUCGUUCUCGUCGACAAUGGCGGAACUGGCGACGUGA
22	1404	GCGACGUGACUGUCGCCCCAAGCAACUUCGCUAACGGGGUC
23	1485	AAGUAACCUGUAGCGUUCGUCAGAGCUCUGCGCAGAAUCGC
24	1552	UAAAGUGGCAACCCAGACUGUUGGUGGUGUAGAGCUUCCUG
25	1634	CCAAUUUUCGCUACGAAUUCCGACUGCGAGCUUAUUGUUAA
26	1696	AGAUGGAAACCCGAUUCCCUCAGCAAUCGCAGCAAACUCCG
27	1741	CUACUAAUAGACGCCGGCCAUUCAAACAUGAGGAUUACCCA
28	1807	CUCUUUAUGUAUUGAUCUUCCUCGCGAUCUUUCUCUCGAAA
29	1938	UCACAAAGCAUCCGACCUUAGGUUCUGGUAAUGACGAGGCG
30	1960	UUCUGGUAAUGACGAGGCGACCCGUCGUACCUUAGCUAUCG
31	2056	ACAUGACAAAUCCUUGUCAUGGGAUCCGGAUGUUUUACAAA
32	2116	UGGCAACCUCCUCUCUGGCUACCGAUCGUCGUUGUUUGGGC
33	2211	CUUACAAGAAGUUCGCUGAACAAGCAACCGUUACCCCCCGC
34	2332	AUUUAGGCUCGUUGUAGGGAACGGAGUGUUUACAGUUCCGA
35	2383	AAUAGAUCGGGCUGCCUGUAAGGAGCCUGAUAUGAAUAUGU
36	2451	UCAGACGCCGGCUCAAAUCCGUUGGUAUAGACCUGAAUGAU
37	2481	ACCUGAAUGAUCAAUCGAUCAACCAGCGUCUGGCUCAGCAG
38	2509	UCUGGCUCAGCAGGGCAGCGUAGAUGGUUCGCUUGCGACGA
39	2557	AUCGUCUGCAUCCGAUUCCAUCUCCGAUCGCCUGGGUGUGGA
40	2606	CCACCAGAGCUAUAUUCAUAUCUCGAUCGUAUCCGCUCACA
41	2646	ACUACGGAAUCGUAGAUGGCGAGACGAUACGAUGGGAACUA
42	2731	GAUAUUCUGGGCAAUAGUCAAAGCGACCCAAAUCCAUUUUG
43	2759	CAAAUCCAUUUUGGUAACGCCGGAACCAUAGGCAUCUACGG
44	2800	GGACGAUAUUAUAUGUCCCAGUGAGAUUGCACCCCGUGUGC
45	2854	CUACUACGGUUUUAAACCGAAUCUUCGUAAAACGUUCGUGU
46	2951	CCGUUUUACAUCAAGAAACCUGUUGACAAUCUCUUCGCCCU
47	2982	UCUUCGCCCUGAUGCUGAUAUUAAAUCGGCUACGGGGUUGG
48	3007	UCGGCUACGGGGUUGGGGGAGUUGUCGGAGGUAUGUCAGAUC
49	3091	UUCGAUGUUCUUCGGUGGGACGGACCUCGCUGCCGACUACU
50	3131	UACGUAGUCAGCCCGCCUACGGCAGUCUCGGUAUACACCAA
51	3161	GUAUACACCAAGACUCCGUACGGGCGGCUGCUCGCGGAUAC
52	3202	CCGUACCUCGGGUUUCCGUCUUGCUCGUAUCGCUCGAGAAC
53	3345	UACGCACUUCGGAGUGGCUAACGCCGGUUCCCACAUUCCCU
54	3374	CCACAUUCCCUCAGGAGUGUGGGGCCAGCGAGCUCUCCUCGG

genome plays in the phage lifecycle. Previous smFCS assays of assembly *in vitro* with the full genome and large sub-fragments, all containing the TR site, show that assembly is divided into two separate kinetic phases. There is an initially rapid phase that occurs above a minimum threshold CP concentration during which the RNA conformational ensemble collapses to facilitate encapsidation [38,67]. This CP concentration-independent process is followed by a kinetically slower phase, dependent on CP concentration, during which additional CPs are recruited to complete formation of $T = 3$ particles. Controlling the partitioning of these two stages is presumably an essential feature of assembly

efficiency. One way that this could be achieved is via PSs with a distinct hierarchy of affinities, with higher-affinity interactions forming in the first phase and lower-affinity ones forming in the later phase. Alternatively, the RNA could undergo conformational change revealing functional PSs after the binding of the initial CP subunits.

In order to discriminate between these mechanisms, we carried out secondary structure probing of transcripts encompassing large sections of the MS2 RNA, as well as the intact virion (Fig. 5a). This was performed using Pb^{2+} ions that hydrolyse RNA via the deprotonation of the 2′ OH followed by cleavage of the phosphodiester bond at the 3′ position [68–72]. Lead ions are ideal for this purpose because their small size allows complete penetration into the MS2 capsids, which have pores at the 3-fold and 5-fold axes, and their reactivity is rapidly controllable
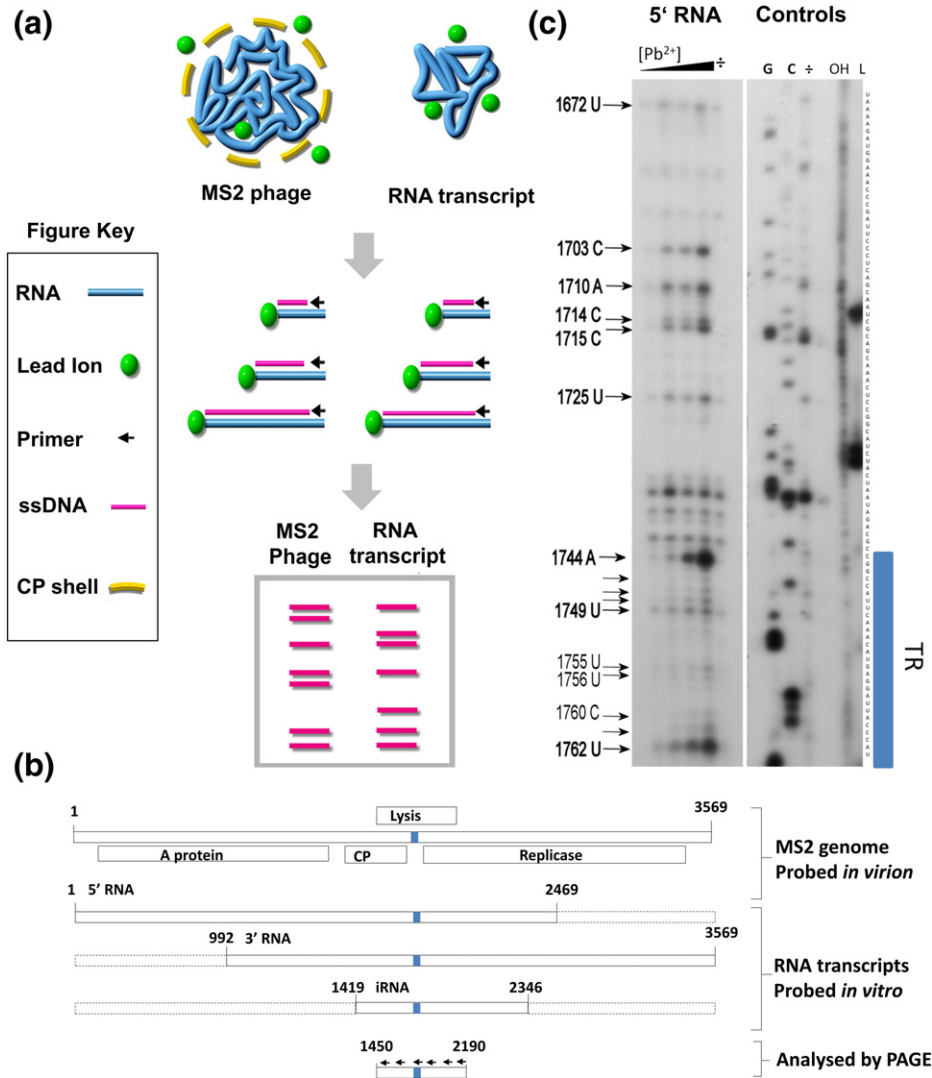

Fig. 5. Lead acetate probing of the MS2 ssRNA genome. (a) Outline of the lead probing experiment. The time-dependent cleavage of the genome within MS2 virions or sub-genomic fragments was performed in separate reactions at room temperature. Following Pb^{2+} incubation, we quenched the reactions with EDTA and we precipitated the RNAs at high salt concentration. Following addition of EDTA, we phenol/chloroform extracted the MS2 phage sample to remove virion proteins. (b) Schematic showing the RNA fragments footprinted and the region from 1419 to 2190 probed. The arrows indicate the sites of hybridisation of the primers; blue bar, the location of the TR site. (c) An autoradiograph of a 6% (w/v) PAGE assay of primer extension products obtained by reverse transcription of the 5′ RNA treated with Pb^{2+}. The TR operator is highlighted by the blue bar. Samples were treated with 0.4 mM Pb^{2+} in assembly buffer for 5, 10, 30 and 60 min. A control reaction lacking Pb^{2+} is indicated by "÷". Arrows highlight sites cleaved specifically by lead-ion-induced hydrolysis positions. Dideoxy sequencing ladders (G and C), a hydrolysis ladder (OH) and size standards (L) analyzed in adjacent lanes allowed identification of the nucleotide sequence and hence the sites of cleavage.
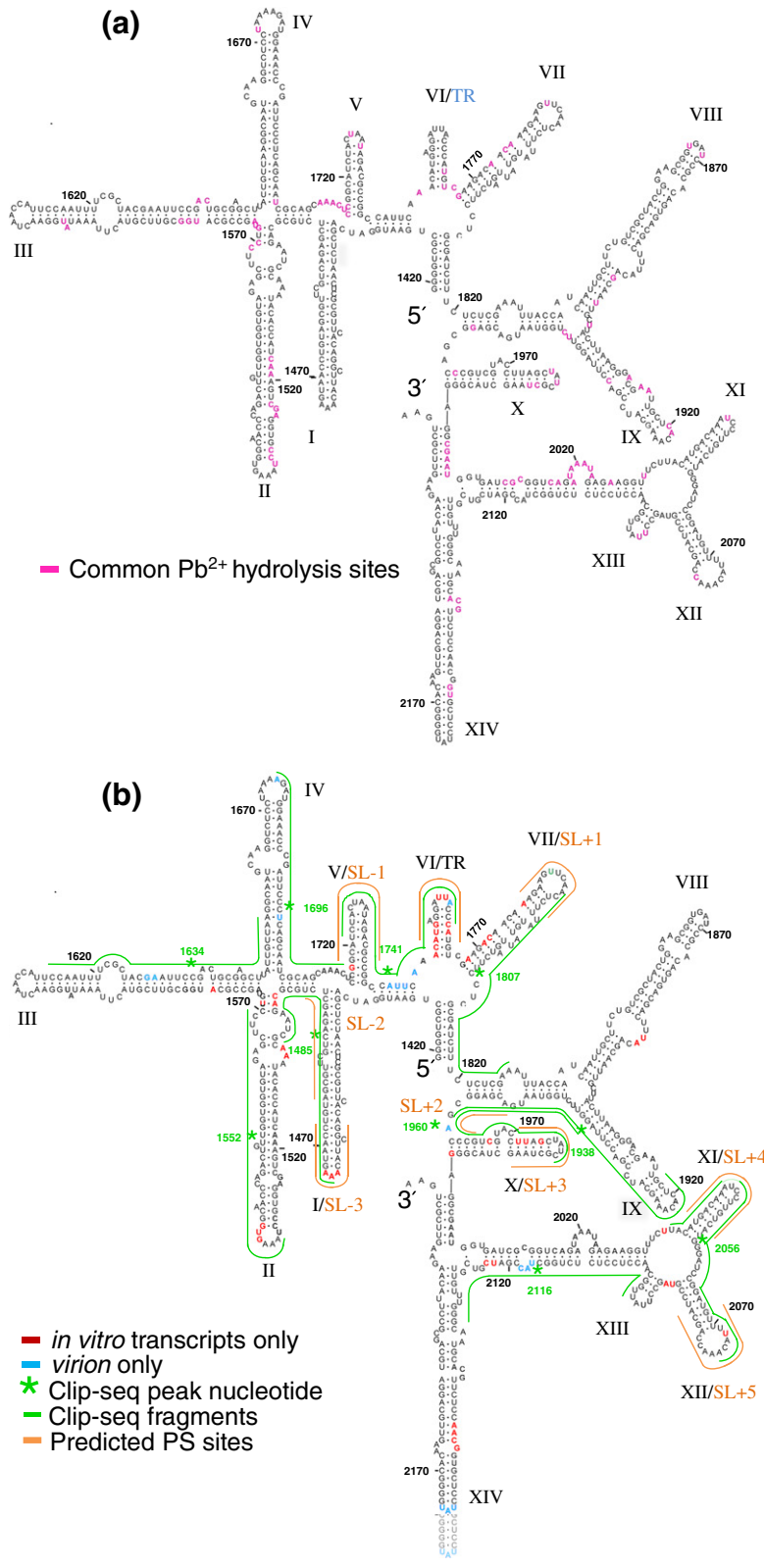
Fig. 6 (*legend on next page*)

by addition of excess ethylenediaminetetraacetic acid (EDTA). In addition, lead ions are expected to have minimal interactions with capsid proteins. The rate of cleavage is affected by the accessibility of the 2′ OH and the flexibility of the RNA backbone. Therefore, efficient cleavage occurs mainly in single-stranded regions and is inhibited in base-paired, stacked or higher-order structures in the RNA or due to interaction with proteins [71].

The lead-ion-induced hydrolysis sites on three MS2 RNA transcripts (referred to as the 5′ RNA, the 3′ RNA and the internal iRNA) were compared to that of the genome within virions (Fig. 5b). We assayed the genome between nucleotide 1450 and 2190, using primers spaced at roughly 130-nt intervals in the MS2 genome, resulting in partial overlap of the primer extension products when resolved on 6–8% (w/v) polyacrylamide sequencing gels. This region is present in all the RNAs, encompasses the TR PS and is well away from MP-binding sites. By analyzing the kinetics of lead-ion-induced hydrolysis, it was possible to assign nucleotide cleavage positions from autoradiographs as signals whose intensities increased with incubation time (see Materials and Methods). An example of the lead-ion-induced cleavage patterns observed on autoradiograms is shown in Fig. 5c. The pattern of cleavages observed within the TR operator in the 5′ RNA is similar in both the 3′ RNA and the iRNA (Supplementary Fig. 2). Three distinct patterns of lead ion cleavage are seen across the different fragments. These are sites that are present in the virion only, in the protein-free RNA fragments only or in both. Note that lead ion cleavage is relatively slow and the signals seen on the gels are from the averaging of many individual molecules. Therefore, it is not possible to be definitive about whether any cleavage sites detected are universally present. However, the qualitative distinction between sites that cleave within the assay period and those that do not provides a clear indication about the propensity for a particular structure to form.

RNA structure is partially conserved *in vitro* and within the virion

Conserved lead ion hydrolysis sites were observed across all the RNA fragments that together span the MS2 genome (Supplementary Fig. 2). For example, ~75% of the cleavage sites in iRNA are also observed in both the 5′ and 3′ RNAs, with the similarity between the latter two being ~90%. These results suggest that the solution secondary structures of the overlapping portions of all three transcripts are very similar. Within the probed region in the virion, there are 121 cleavages, 78 of which (64%) are also detected in each of the protein-free transcripts (Fig. 6a). Most of these are within or adjacent to RNA regions proposed, on the basis of previous selective chemical and enzymatic structure probing, to be single stranded [58,73]. Thus, some conservation of secondary structures extends to all four RNAs. This is consistent with these features being the result of local interactions, that is, that short-range contacts drive the formation of similar stem–loops [72,74] that are indifferent to the flanking sequences in each case. For clarity, the stem–loops of the solution secondary structure are labeled 1 to 24 in Roman numerals and, where appropriate, also with our previously predicted PS labels, based on the TR site, that is, SL ± 1 and so on (Fig. 6). Importantly, the clarity of these cleavage patterns suggests that, within each ensemble of probed RNAs, there is a consensus structure and that, within each virion, the genome conformation is largely the same. Such an arrangement is a consequence of PS-mediated assembly and consistent with the structure obtained by asymmetric cryo-tomographic reconstruction [31].

Identification of the CP–RNA genome interactions in the virion

The conserved solution secondary structure elements in the virion imply that the tertiary RNA conformation is a compacted version of the RNA solution structure [67]. The susceptibility of several nucleotides to lead ion hydrolysis, however, differs between the virion and the protein-free RNA. Some of these differences must be a consequence of the more compact structure, and others likely reflect protection from hydrolysis due to CP binding or a protein-induced remodeling of RNA secondary structure. To identify and differentiate between these scenarios, we compared the oligonucleotides isolated by CLIP-Seq (Fig. 4d and Table 1), the previously predicted PS sites [41], the proposed 2° genome structure [57,58,75] and lead ion cleavage positions on protein-free RNA (Fig. 6b).

Fig. 6. Identification of different classes of PS sites. (a) Shared lead ion cleavage positions detected in both the transcripts and the MS2 virion are highlighted in magenta on the RNA secondary structure [73]. Proposed hairpins are indicated as I–XIV. (b) Altered lead-ion-induced cleavage sites are indicative of altered RNA structure following genome packaging. The lead-ion-induced cleavage sites detected only in transcripts, that is, which are protected in the virion, are highlighted in red. Those that cleave only in the virion are in blue. Comparison of these positions with MiSeq reads, shown here with a green line spanning 40 nt around the peak nucleotide (*), and previously predicted PSs [41] within the MS2 genome, shown here as orange lines and annotated as SL − 3 to SL + 5, allows the identification of PSs that have a propensity to fold prior to capsid assembly or that must refold during assembly.

TR is the highest-affinity PS within the MS2 genome and is thus the site [56] where lead cleavage patterns would be expected to differ the most between solution and the virion. Consistent with this idea, the cleavage patterns of stem–loop VI/TR differ significantly (Fig. 6b) with a large number of nucleotide positions becoming less accessible (red) to the lead ions in the virion, while one nucleotide involved in sequence-specific recognition (Fig. 1b) becomes more accessible (blue). This result suggests that marked alterations in cleavage site accessibility occur as a result of CP binding. A peak centered at nucleotide 1741 in the CLIP-Seq data lies between VI/TR and the 5′ V stem–loop, which we predict is also a PS, SL − 1, confirming the inference from the multiple potential PSs in the CLIP-Seq fragments (Fig. 4d and Table 1) and allowing us to interpret the wider RNA cleavage patterns with confidence.

Starting from the 3′ end of the probed region, there is a CLIP-Seq peak at 2116 that is adjacent to nucleotides with altered accessibility to lead hydrolysis in the virion, implying that this is a PS although it was not previously predicted to be one. The next peak at 2056 lies between two stem–loops predicted to be PSs, XII/SL + 5 and XI/SL + 4, the former having one nucleotide with a detectable change in accessibility. The results with TR and SL − 1 suggest that these sites are indeed PSs. Peak 1960 lies between X/SL + 3 and SL + 2, with both sequences showing some altered lead accessibility, implying two PSs, one present in solution and one that must refold during assembly. Peak 1938 overlaps with the region of SL + 2 and with IX. The latter is not predicted as a PS and shows no changes in lead accessibility; thus, its status cannot be confirmed. Peak 1807 overlaps with VII/SL + 1, which shows some altered accessibility and therefore is likely a PS. Peak 1696 overlaps with V/SL − 1 and IV, the pattern showing increased cleavage at a single nucleotide but seemingly not at a PS. Peak 1634 overlaps an area that undergoes slight changes in accessibility but does not form stem–loops and hence cannot be a PS. Peak 1485 is centered on SL − 2 but is not in the form of a stem–loop and therefore must refold extensively into its PS conformation. It also overlaps with I/SL − 3 where the decreased lead accessibility to the loop region confirms its role as a PS. These correlations with solution secondary structures, CLIP-Seq and PS predictions confirm the ideas of a PS-mediated assembly reaction without the need for extensive mutagenesis that could have pleiotropic effects on MS2 infection.

Discussion

The virions of ssRNA viruses spontaneously self-assemble *in vitro* and, it is assumed, *in vivo*.

This reaction has been proposed to be largely driven by electrostatic interactions. The data presented above show clearly that the end product of such assembly has a highly organized relationship between its virion proteins and genomic RNA, which is most unlikely to arise by a simple charge compaction process. Instead, we have shown that the process is highly regulated and involves multiple RNA–protein contacts at sites distributed throughout the genome, PSs. Such interactions occur in viruses from bacteria (MS2 [2,23,31,37,41]), plants (TCV [76], BMV [63] and STNV [38]) and humans (HPEV-1 [77]). For BMV, it appears that these interactions can also be altered by post-translational modification of the CP subunits [63] and that virion assembly is controlled by more than electrostatics [78].

PS-mediated assembly sets up a unique relationship between the CP shell and the genomic sequence, allowing the virus to control the spatial positioning of those sequences. This may be an essential feature during the early uncoating events that occur during the infection of the new host since extrusion of the RNA occurs at a unique site in the protein shell [79,80]. These properties explain why a similar mechanism is found in viruses from such wide-ranging hosts. Since capsid assembly has no cellular homologue that we know of, disrupting PS-mediated assembly can be exploited for therapeutic benefit. Given the potential importance of this mechanism, it is important to validate it as widely as possible. The data presented here contribute significantly to that process. By observing the CP–PS contacts in virions, we have confirmed the roles of multiple PS sites without having to make multiple mutant genomes.

The results are not exhaustive because they do not identify every possible site of contact between virion proteins and the enclosed genome. RNA sequences that lack obvious PS motifs are also present in the CLIP-Seq data, possibly because they contain features that favor precipitation in the CLIP-Seq assay. Nonetheless, the overwhelming majority of contacts are consistent with PS-mediated assembly. The CP peptides in contact with the genomic RNA (RCAP) are similar to those seen in high-resolution X-ray structures of the TR bound to every position. In principle, each PS contact within a genome would be to an A/B dimer (Fig. 1). The VLPs used for X-ray structure determination of bound TR and its variants were soaked with excess oligonucleotide that binds to all dimer sites. At C/C sites, due to the 2-fold symmetry, they make degenerate contacts but bind in a unique orientation at the same RNA–CP interface seen for TR at A/B [27]. The peptide data therefore confirm that every virion has a similar arrangement of RNA–protein contacts for the first time. In addition, CLIP-Seq has been used to identify the RNA sequences in close contact with

these peptides, revealing extensive overlap with previous predictions of PSs within this virus. This is the most compelling evidence to date confirming that MS2 assembly occurs via multiple PS–CP contacts.

We have also investigated the architecture of the MP. Its contacts to the genome are localized to the N-terminal half of the molecule, consistent with predictions of two RNA-binding sites in this region. Although these RNA-binding motifs seem to occur across a large range of the known MPs, the sequences of these N-termini are quite variable. In contrast, there is extensive conservation at the C-termini, perhaps consistent with it encompassing the pilus-binding domain.

Combining CLIP-Seq information with RNA chemical structure probing allows us to interrogate the presence/appearance of PSs before and after virion assembly. It is clear that there are two classes of PS sites, those that are detectable in the protein-free RNA and those that are only detected after CP-induced refolding and capsid assembly. Presumably, the PS folds are flickering in and out of existence and the lead ions are sensitive to equilibrium. This is consistent with the two phases of assembly previously detected using smFCS *in vitro*. This scenario is somewhat analogous to that seen for assembly of the ribosome [59–62]. However, ssRNA viruses face a more complex assembly task than the ribosome. They must co-assemble around RNA genomes, often many times longer than any ribosomal RNA, using multiple copies, often numbering in the hundreds, of the same or very similar CP subunits, to form a highly symmetric container of restricted volume. Confining a complex folded polymer, such as genomic RNA, into the limited volume of the viral capsid requires proper regulation. Furthermore, the assembly of the virion must allow proper extrusion of the RNA and proper viral gene expression in the subsequent round of infection. Thus, while electrostatic interactions are useful in collapsing the conformational ensemble of protein-free RNA and for lowering the overall free energy of the assembled particle, rapid and efficient assembly requires that they occur in a regulated fashion. This control is precisely what PS-mediated assembly provides. The tendency of viral genomes to form branched structures can also contribute to this assembly efficiency [81,82]. Since PS binding by MS2 CP leads to conformational switching in the CP dimer and CP binding to PSs results in local refolding of the genome, assembly occurs via a series of events that each relies on induced fit. Given that this is a highly cooperative event involving multiple RNA–protein and protein–protein contacts, the free energy involved in this induced fit is likely to be considerable. This idea is not new [83], but here we have been able to probe the details of how the process is controlled by the genomic PSs for the first time.

Materials and Methods

Defining the RNA–protein contacts in the virion

RCAP

MS2 virions were from the American Type Culture Collection and propagated in male strains of *Escherichia coli*. The virions were purified [84] and suspended in SMH buffer [100 mM NaCl, 8 mM MgSO$_4$ and 50 mM Hepes (pH 7.5)] to identify regions in the MS2 CP and MP that contact the RNA. The RCAP protocol was as described in Ni *et al.* [63]. Aliquots of the trypsin-treated samples were subsequently treated with sequencing-grade Glu-C (Promega) prior to purification, concentration of the sample using a c18 ZipTip and mass spectrometric analysis using a Bruker Autoflex III mass spectrometer (Agilent Technologies) in positive-ion mode. Some samples were also examined using the reflectron mode. Database searches used the program Mascot (Matrix Science), with the search directed against the MS2 protein sequences [53]. Peptide assignments were only made when they were within 0.5 Da of the expected mass.

CLIP-Seq

MS2 virions in SM buffer [100 mM NaCl, 8 mM MgSO$_4$ and 50 mM Tris–HCl (pH 7.5)] were irradiated at 400 mJ/cm^2 by 254-nm light, on ice in a CL-1000 ultraviolet cross-linker (UVP). The samples were mixed with gentle pipetting and incubated on ice for 3 min. Irradiation was repeated twice more. The virions were then pelleted by ultracentrifugation at 60,000 RPM for 4 min at 4 °C and resuspended in RNA fragmentation buffer [100 mM Tris–HCl (pH 7.0) and 10 mM ZnCl$_2$] at 70 °C for 20 min. The reaction was quenched by adding EDTA, pH 8, to a final concentration of 42 mM. Immunoprecipitation used an anti-MS2 CP serum (EMD Millipore) bound to protein A/G magnetic beads (Pierce) according to the manufacturer's protocol. The eluted CP–RNA complex was electrophoresed on a NuPAGE 4–12% (w/v) Bis-Tris SDS-PAGE gel in 1× 4-morpholineethanesulfonic acid buffer on ice and then transferred to a nitrocellulose membrane using a wet transfer apparatus (BioRad) at 100 V for 1 h. The membrane was stained with 0.1% (w/v) Ponceau-S in 5% (v/v) acetic acid for ~5 min, and the region 6–100 kDa above the CP band was excised from the membrane. Material on the membrane was digested with 0.8 mg/mL proteinase K at 37 °C for 1 h to release the cross-linked RNA fragments. The RNA was extracted using Trizol and precipitated with 3 volumes of ethanol containing 0.3 M ammonium acetate.

In preparation for cDNA library construction, the RNA was treated with T4 polynucleotide kinase (NEB) to remove the 3′ phosphate and add a 5′ phosphate. The pre-treated RNA was then prepared using an Illumina Truseq Small RNA preparation kit. After PCR amplification of the cDNA using Pfu DNA polymerase, the reaction was incubated with 1 µL Taq DNA polymerase at 72 °C for 10 min and the amplified cDNA was purified with a Qiaquick Gel Extraction Kit (Qiagen) from a 2% (w/v) agarose-TAE gel. The libraries from different samples were pooled at equimolar ratios for Illumina MiSeq sequencing using MiSeq reagent kit V3 (150 cycles). All samples from the UV cross-linked and un-cross-linked samples were processed in parallel and sequenced within the same run.

Labeling of CP-bound RNA

To visualize and quantify the RNAs covalently attached to MS2 CP, we incubated CP–RNA complexes treated with $ZnCl_2$ with 2 U of T4 polynucleotide kinase and [γ-^{32}P]ATP for 30 min at 37 °C according to the manufacturer's instructions (NEB). The material was then electrophoresed in NuPAGE 4–12% (w/v) Bis-Tris SDS-PAGE gel in 1× 4-morpholineethanesulfonic acid buffer. The gel was wrapped in plastic film and the signal was detected using a Phosphorimager. Western blots of the CP were performed as described in Hema *et al.* [85].

DNA analysis

Analysis of the MiSeq reads used the Galaxy user interface [86]. The reads were processed with the FASTX toolkits to trim sequences with a quality score lower than 20 using *FASTQ Quality Trimmer,* removing the 3′ adaptor and discarding any reads shorter than 15 nt using *Clip.* The cleaned reads were aligned to the MS2 reference genome using Bowtie2 [87] under the default "local" alignment settings and the actual coverage of each nucleotide position on MS2 RNA was obtained using Mpileup in SAMtools [88]. The Mpileup output from the forward and reverse MiSeq reads were combined. All peaks with nucleotide coverage higher than the threshold of 3685 reads were assigned as putative capsid-interacting sites. The threshold was established with results from a "no UV" control library, using the number of reads at the apex of the translational operator (TR) hairpin as the cutoff.

MP sequence analysis

Leviviridae MP sequences were identified using BLASTP and TBLASTN searches at National Center for Biotechnology Information [89] using MS2 MP as the query (sequence accession P03610). Excluding MS2 search results and using an Expect value of 10 allowed recovery of low homology sequences. Matches were scored using default parameters, with the BLOSUM62 matrix.

MS2 MP secondary structure composition was predicted using the Jpred3 algorithm [90] with sequence accession P03610 query and homologous UniRef sequences identified by the server. CP secondary structures were taken from the annotation of PDB file 2MS2.

RNA-binding sites in MPs were predicted using the RNABindRPlus server [65] with sequence accessions MS2 (P03610), R17 (EF108465), JP501 (AAF67668), phage fr (CAA33135), BZ13 (FJ483837), KU1 (AAF67673), GA (CAA27496), C-1-INW (AFN37812), Hgal1 (AFN37816), PRR (ABH03626), phage M (JX625144), phage Fl (ACT66757), M11 (AAC06249), phage SP (X07489), MX1 (AF059242), NL95 (AF059243), Qβ (AY099114), PP7 (A80191), VK (EU372698) and AP205 (AF334111).

Probing RNA conformation before and after encapsidation

Lead ion probing

The production and purification of the MS2 RNA fragments was as previously described [67]. All structural probing reactions were performed in 10 µL final volume. The 5′ and 3′ sub-genomic fragments were at a 0.1 µM final RNA concentration. The iRNA was kept at 0.27 µM final concentration. RNA concentrations were measured by absorbance at 260 nm assuming that one asymmetric unit is equivalent to 40 µg/µL. Wild-type MS2 phage was used at a final concentration of 0.285 µg/µL. Prior to cleavage with lead ions, the MS2 sub-genomic RNAs were heated to 65 °C for 10 min and then cooled to room temperature at 2 °C/min on a thermal cycler in 6 µL final volume. At this point, 2 µL of 5× assembly buffer was added to the sample resulting in a final concentration of 40 mM NH_4OAc and 1 mM $Mg(OAc)_2$ (pH 7.2). We then added 2 µL of freshly prepared 2 mM lead acetate to each reaction to a final concentration of 0.4 mM lead ions. Reactions were incubated at room temperature for 5, 10, 30 and 60 min followed by addition of 5 µL of 0.1 M EDTA, 1.5 µL of 3 M NaOAc and 35 µL of EtOH and were stored at −20 °C for $\geq$2 h. RNA was recovered as a pellet by centrifugation at 13,000 RPM and washed once with 70% EtOH prior to resuspension in 5 µL DEPC-treated H_2O. Wild-type MS2 capsids were treated similarly but were phenol/chloroform extracted prior to precipitation with 3 M NaOAc. Prior to reverse transcription, the appropriate ^{32}P-labeled primer was annealed to the cleaved RNA by addition of 1 µL of 1 µM labeled primer stock solution and incubated at 65 °C for 10 min followed by snap cooling on ice. The primers used for reverse transcription were as follows:

```
1568_R:AAGCTCTACACCACCAACAGTCT
1666_R:CCTTGCATTGCCTTAACA
1812_R:CGCGAGGAAGATCAATACATA
1951_R:TCATTACCAGAACCTAAGGTCGGA
2061_R:GATCCCATGACAAGGATTTG
2223_R:TAACGGTTGCTTGTTCAGC
```

Reverse transcription was performed with Transcriptor® reverse transcriptase (Roche Diagnostics) at 52 °C for 30 min according to the manufacturer's description. Single-stranded DNA products were then precipitated by addition of 3 µL of 3 M NaOAc and 60 µL of EtOH, recovered by centrifugation and washed once with 70% EtOH and allowed to dry at room temperature for 15 min. Single-stranded DNA was dissolved in 95% formamide RNA loading buffer and run on 6–8% (w/v) denaturing polyacrylamide slab gels [91]. The gels were then exposed to film at −80 °C overnight and digitized using an image scanner.

Data analysis

Lead ion cleavage positions were identified from exposed film first and foremost as signals whose intensity increased with increasing lead ion incubation time. An absence of a signal in the control lane was taken as an indicator but not a requirement of lead-ion-induced cleavage. Because control reactions were incubated in assembly buffer for 60 min at room temperature, non-specific hydrolysis sites were always detected. If intensity at these positions was observed to increase in the presence of lead acetate over time and the intensity with lead ions *versus* control after 60 min was higher in sample containing lead ions, this was defined as a hydrolysis position. The nucleotide position of induced hydrolysis within the MS2 genome was deduced from dideoxy sequencing ladders (G and C) and a size standard.

Identified positions were not quantified, and thus, a comparison of cleavage intensities between RNAs was not performed. All lead-ion-induced hydrolysis was therefore identified in a Boolean manner with regard to the above-mentioned criteria. Lead ion hydrolysis positions were mapped onto the MS2 genome and their nucleotide positions were compared with respect to whether cleavage occurred only in the RNA sub-genomic fragments *in vitro*, only within the virion or in both the virion and the sub-genomic fragments *in vitro*. The numbers of cleavages occurring in these three groups were taken as an indicator of structural similarity of these RNAs in the different environments.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2015.11.014.

Present addresses: Ó. Rolfsson, Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland, and University of Iceland Biomedical Center, Laeknagardur, 101 Reykjavik, Iceland.

†Ó.R. and S.M. contributed equally to this work.

Abbreviations used:
CP, coat protein; MP, maturation protein; PS, packaging signal; ssRNA, single-stranded RNA; EM, electron microscopy; EDTA, ethylenediaminetetraacetic acid.

References

[1] Viral molecular machines, 2012, http://dx.doi.org/10.1007/978-1-4614-0980-9.

[2] P.G. Stockley, N.A. Ranson, R. Twarock, A new paradigm for the roles of the genome in ssRNA viruses, Futur. Virol. 8 (2013) 531–543, http://dx.doi.org/10.2217/fvl.12.84.

[3] R.D. Cadena-Nava, M. Comas-Garcia, R.F. Garmann, A.L.N. Rao, C.M. Knobler, W.M. Gelbart, Self-assembly of viral capsid protein and RNA molecules of different sizes: Requirement for a specific high protein/RNA mass ratio, J. Virol. 86 (2012) 3318–3326, http://dx.doi.org/10.1128/JVI.06566-11.

[4] J.B. Bancroft, G.J. Hills, R. Markham, A study of the self-assembly process in a small spherical virus. Formation of organized structures from protein subunits *in vitro*, Virology 31 (1967) 354–379.

[5] J.B. Bancroft, E. Hiebert, C.E. Bracker, The effects of various polyanions on shell formation of some spherical viruses, Virology 39 (1969) 924–930.

[6] O.M. Elrad, M.F. Hagan, Encapsulation of a polymer by an icosahedral virus, Phys. Biol. 7 (2010) 045003, http://dx.doi.org/10.1088/1478-3975/7/4/045003.

[7] P. van der Schoot, R. Bruinsma, Electrostatics and the assembly of an RNA virus, Phys. Rev. E Stat. Nonlinear Soft Matter Phys. 71 (2005) 061928, http://dx.doi.org/10.1103/physreve.71.061928.

[8] A. Kivenson, M.F. Hagan, Mechanisms of capsid assembly around a polymer, Biophys. J. 99 (2010) 619–628, http://dx.doi.org/10.1016/j.bpj.2010.04.035.

[9] R. Balint, S.S. Cohen, The incorporation of radiolabeled polyamines and methionine into turnip yellow mosaic virus in protoplasts from infected plants, Virology 144 (1985) 181–193, http://dx.doi.org/10.1016/0042-6822(85)90316-2.

[10] R.F. Bruinsma, Physics of RNA and viral assembly, Eur. Phys. J. E Soft Matter 19 (2006) 303–310, http://dx.doi.org/10.1140/epje/i2005-10071-1.

[11] R.F. Garmann, M. Comas-Garcia, M.S.T. Koay, J.J.L.M. Cornelissen, C.M. Knobler, W.M. Gelbart, Role of electrostatics in the assembly pathway of a single-stranded RNA virus, J. Virol. 88 (2014) 10472–10479, http://dx.doi.org/10.1128/JVI.01044-14.

[12] V.A. Belyi, M. Muthukumar, Electrostatic origin of the genome packing in viruses, Proc. Natl. Acad. Sci. 103 (2006) 17174–17178, http://dx.doi.org/10.1073/pnas.0608311103.

[13] S.C. Harrison, A.J. Olson, C.E. Schutt, F.K. Winkler, G. Bricogne, Tomato bushy stunt virus at 2.9 Å resolution, Nature 276 (1978) 368–373, http://dx.doi.org/10.1038/276368a0.

[14] M.G. Rossmann, E. Arnold, J.W. Erickson, E.A. Frankenberger, J.P. Griffith, H.J. Hecht, et al., Structure of a human common cold virus and functional relationship to other picornaviruses, Nature 317 (1985) 145–153, http://dx.doi.org/10.1038/317145a0.

[15] T.A. Jones, L. Liljas, Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution, J. Mol. Biol. 177 (1984) 735–767, http://dx.doi.org/10.1016/0022-2836(84)90047-0.

[16] J.M. Hogle, M. Chow, D.J. Filman, Three-dimensional structure of poliovirus at 2.9 Å resolution, Science 229 (1985) 1358–1365.

[17] J.M. Hogle, A. Maeda, S.C. Harrison, Structure and assembly of turnip crinkle virus. I. X-ray crystallographic structure analysis at 3.2 Å resolution, J. Mol. Biol. 191 (1986) 625–638, http://dx.doi.org/10.1016/0022-2836(86)90450-X.

[18] A. Routh, T. Domitrovic, J.E. Johnson, Packaging host RNAs in small RNA viruses: An inevitable consequence of an error-prone polymerase? Cell Cycle 11 (2012) 3713–3714, http://dx.doi.org/10.4161/cc.22112.

[19] A. Routh, T. Domitrovic, J.E. Johnson, Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 1907–1912, http://dx.doi.org/10.1073/pnas.1116168109.

[20] D.Y. Kim, A.E. Firth, S. Atasheva, E.I. Frolova, I. Frolov, Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution, J. Virol. 85 (2011) 8022–8036, http://dx.doi.org/10.1128/JVI.00644-11.

[21] J. Sasaki, K. Taniguchi, The 5′-end sequence of the genome of Aichi virus, a picornavirus, contains an element critical for viral RNA encapsidation, J. Virol. 77 (2003) 3542–3548, http://dx.doi.org/10.1128/jvi.77.6.3542-3548.2003.

[22] H.F. Lodish, K. Horiuchi, N.D. Zinder, Mutants of the bacteriophage f2. V. On the production of noninfectious phage particles, Virology 27 (1965) 139–155, http://dx.doi.org/10.1016/0042-6822(65)90064-4.

[23] C.M. Ling, P.P. Hung, L.R. Overby, Independent assembly of Qbeta and MS2 phages in doubly infected *Escherichia coli*, Virology 40 (1970) 920–929.

[24] D. Beckett, H.N. Wu, O.C. Uhlenbeck, Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly, J. Mol. Biol. 204 (1988) 939–947, http://dx.doi.org/10.1016/0022-2836(88)90053-8.

[25] D.S. Peabody, Role of the coat protein–RNA interaction in the life cycle of bacteriophage MS2, Mol. Gen. Genet. 254 (1997) 358–364.

[26] G. Basnak, V.L. Morton, O. Rolfsson, N.J. Stonehouse, A.E. Ashcroft, P.G. Stockley, Viral genomic single-stranded RNA directs the pathway toward a T = 3 capsid, J. Mol. Biol. 395 (2010) 924–936, http://dx.doi.org/10.1016/j.jmb.2009.11.018.

[27] K. Valegård, J.B. Murray, N.J. Stonehouse, S. van den Worm, P.G. Stockley, L. Liljas, The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein–RNA interactions, J. Mol. Biol. 270 (1997) 724–738, http://dx.doi.org/10.1006/jmbi.1997.1144.

[28] R. Golmohammadi, K. Valegård, K. Fridborg, L. Liljas, The refined structure of bacteriophage MS2 at 2.8 Å resolution, J. Mol. Biol. 234 (1993) 620–639, http://dx.doi.org/10.1006/jmbi.1993.1616.

[29] K. Toropova, G. Basnak, R. Twarock, P.G. Stockley, N.A. Ranson, The three-dimensional structure of genomic RNA in bacteriophage MS2: Implications for assembly, J. Mol. Biol. 375 (2008) 824–836, http://dx.doi.org/10.1016/j.jmb.2007.08.067.

[30] T. Shiba, Y. Suzuki, Localization of A protein in the RNA-A protein complex of RNA phage MS2, Biochim. Biophys. Acta 654 (1981) 249–255.

[31] K.C. Dent, R. Thompson, A.M. Barker, J.A. Hiscox, J.N. Barr, P.G. Stockley, et al., The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release, Structure 21 (2013) 1225–1234, http://dx.doi.org/10.1016/j.str.2013.05.012.

[32] P.G. Stockley, O. Rolfsson, G.S. Thompson, G. Basnak, S. Francese, N.J. Stonehouse, et al., A simple, RNA-mediated allosteric switch controls the pathway to formation of a T = 3 viral capsid, J. Mol. Biol. 369 (2007) 541–552, http://dx.doi.org/10.1016/j.jmb.2007.03.020.

[33] E.C. Dykeman, R. Twarock, All-atom normal-mode analysis reveals an RNA-induced allostery in a bacteriophage coat protein, Phys. Rev. E Stat. Nonlinear Soft Matter Phys. (2010), http://dx.doi.org/10.1103/physreve.81.031908.

[34] E.C. Dykeman, P.G. Stockley, R. Twarock, Dynamic allostery controls coat protein conformer switching during MS2 phage assembly, J. Mol. Biol. 395 (2010) 916–923, http://dx.doi.org/10.1016/j.jmb.2009.11.016.

[35] K. Valegård, J.B. Murray, P.G. Stockley, N.J. Stonehouse, L. Liljas, Crystal structure of an RNA bacteriophage coat protein-operator complex, Nature 371 (1994) 623–626, http://dx.doi.org/10.1038/371623a0.

[36] A. Borodavka, R. Tuma, P.G. Stockley, A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence, RNA Biol. 10 (2013) 481–489, http://dx.doi.org/10.4161/rna.23838.

[37] A. Borodavka, R. Tuma, P.G. Stockley, Evidence that viral RNAs have evolved for efficient, two-stage packaging, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 15769–15774, http://dx.doi.org/10.1073/pnas.1204357109.

[38] N. Patel, E.C. Dykeman, R.H.A. Coutts, G.P. Lomonossoff, D.J. Rowlands, S.E.V. Phillips, et al., Revealing the density of encoded functions in a viral RNA, Proc. Natl. Acad. Sci. 112 (2015) 2227–2232, http://dx.doi.org/10.1073/pnas.1420812112.

[39] P.G. Stockley, R. Twarock, S.E. Bakker, A.M. Barker, A. Borodavka, E. Dykeman, et al., Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism, J. Biol. Phys. 39 (2013) 277–287, http://dx.doi.org/10.1007/s10867-013-9313-0.

[40] D.H.J. Bunka, S.W. Lane, C.L. Lane, E.C. Dykeman, R.J. Ford, A.M. Barker, et al., 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 (2011) 51–65, http://dx.doi.org/10.1016/j.jmb.2011.07.063.

[41] E.C. Dykeman, P.G. Stockley, R. Twarock, Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome, J. Mol. Biol. 425 (2013) 3235–3249, http://dx.doi.org/10.1016/j.jmb.2013.06.005.

[42] C. Levinthal, How to Fold Graciously, Mossbauer Spectroscopy in Biological Systems Proceedings 67 (41) (1969) 22–26.

[43] E.C. Dykeman, P.G. Stockley, R. Twarock, Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 5361–5366, http://dx.doi.org/10.1073/pnas.1319479111.

[44] J. Carey, P.T. Lowary, O.C. Uhlenbeck, Interaction of R17 coat protein with synthetic variants of its ribonucleic acid binding site, Biochemistry 22 (1983) 4723–4730, http://dx.doi.org/10.1021/bi00289a017.

[45] H.-N. Wu, O.C. Uhlenbeck, Role of a bulged A residue in a specific RNA–protein interaction, Biochemistry 26 (1987) 8221–8227, http://dx.doi.org/10.1021/bi00399a030.

[46] S.J. Talbot, S. Goodman, S.R. Bates, C.W. Fishwick, P.G. Stockley, Use of synthetic oligoribonucleotides to probe RNA–protein interactions in the MS2 translational operator complex, Nucleic Acids Res. 18 (1990) 3521–3528.

[47] P.G. Stockley, N.J. Stonehouse, J.B. Murray, S.T. Goodman, S.J. Talbot, C.J. Adams, et al., Probing sequence-specific RNA recognition by the bacteriophage MS2 coat protein, Nucleic Acids Res. 23 (1995) 2512–2518.

[48] I. Hirao, M. Spingola, D. Peabody, A. Ellington, The limits of specificity: An experimental analysis with RNA aptamers to MS2 coat protein variants, Mol. Divers. 4 (1998) 75–89, http://dx.doi.org/10.1023/A:1026401917416.

[49] S. Rowsell, N.J. Stonehouse, M.A. Convery, C.J. Adams, A.D. Ellington, I. Hirao, et al., Crystal structures of a series of RNA aptamers complexed to the same protein target, Nat. Struct. Biol. 5 (1998) 970–975, http://dx.doi.org/10.1038/2946.

[50] M.A. Convery, S. Rowsell, N.J. Stonehouse, A.D. Ellington, I. Hirao, J.B. Murray, et al., Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution, Nat. Struct. Biol. 5 (1998) 133–139, http://dx.doi.org/10.1038/nsb0298-133.

[51] E. Grahn, N.J. Stonehouse, J.B. Murray, S. van den Worm, K. Valegård, K. Fridborg, et al., Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: Implications for binding requirements, RNA 5 (1999) 131–138.

[52] E. Grahn, T. Moss, C. Helgstrand, K. Fridborg, M. Sundaram, K. Tars, et al., Structural basis of pyrimidine specificity in the MS2 RNA hairpin-coat-protein complex, RNA 7 (2001) 1616–1627.

[53] C. Helgstrand, E. Grahn, T. Moss, N.J. Stonehouse, K. Tars, P.G. Stockley, et al., Investigating the structural basis of purine specificity in the structures of MS2 coat protein RNA translational operator hairpins, Nucleic Acids Res. 30 (2002) 2678–2685, http://dx.doi.org/10.1093/nar/gkf371.

[54] W.T. Horn, K. Tars, E. Grahn, C. Helgstrand, A.J. Baron, H. Lago, et al., Structural basis of RNA binding discrimination between bacteriophages Qbeta and MS2, Structure 14 (2006) 487–495, http://dx.doi.org/10.1016/j.str.2005.12.006.

[55] V.L. Morton, E.C. Dykeman, N.J. Stonehouse, A.E. Ashcroft, R. Twarock, P.G. Stockley, The impact of viral RNA on assembly pathway selection, J. Mol. Biol. 401 (2010) 298–308, http://dx.doi.org/10.1016/j.jmb.2010.05.059.

[56] H. Lago, A.M. Parrott, T. Moss, N.J. Stonehouse, P.G. Stockley, Probing the kinetics of formation of the bacteriophage MS2 translational operator complex: Identification of a protein conformer unable to bind RNA11. Edited by J. Karn, J. Mol. Biol. 305 (2001) 1131–1144, http://dx.doi.org/10.1006/jmbi.2000.4355.

[57] H. Groeneveld, K. Thimon, J. van Duin, Translational control of maturation-protein synthesis in phage MS2: A role for the kinetics of RNA folding? RNA 1 (1995) 79–88.

[58] R.C.L. Olsthoorn, Structure and Evolution of RNA Phages, Ph.D. Thesis, Leiden University, 1996.

[59] M.W.T. Talkington, G. Siuzdak, J.R. Williamson, An assembly landscape for the 30S ribosomal subunit, Nature 438 (2005) 628–632, http://dx.doi.org/10.1038/nature04261.

[60] W.K. Ridgeway, D.P. Millar, J.R. Williamson, Quantitation of ten 30S ribosomal assembly intermediates using fluorescence triple correlation spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 13614–13619, http://dx.doi.org/10.1073/pnas.1204620109.

[61] T. Adilakshmi, D.L. Bellur, S.A. Woodson, Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly, Nature 455 (2008) 1268–1272, http://dx.doi.org/10.1038/nature07298.

[62] S.A. Woodson, RNA folding pathways and the self-assembly of ribosomes, Acc. Chem. Res. 44 (2011) 1312–1319, http://dx.doi.org/10.1021/ar2000474.

[63] P. Ni, R.C. Vaughan, B. Tragesser, H. Hoover, C.C. Kao, The plant host can affect the encapsidation of brome mosaic virus (BMV) RNA: BMV virions are surprisingly heterogeneous, J. Mol. Biol. 426 (2014) 1061–1076, http://dx.doi.org/10.1016/j.jmb.2013.09.007.

[64] P.M. Krahn, R.J. O'Callaghan, W. Paranchych, Stages in phage R17 infection. VI. Injection of A protein and RNA into the host cell, Virology 47 (1972) 628–637, http://dx.doi.org/10.1016/0042-6822(72)90552-1.

[65] R.R. Walia, L.C. Xue, K. Wilkins, Y. El-Manzalawy, D. Dobbs, V. Honavar, RNABindRPlus: A predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins, PLoS ONE 9 (2014) e97725, http://dx.doi.org/10.1371/journal.pone.0097725.

[66] S. Bleckley, S.J. Schroeder, Incorporating global features of RNA motifs in predictions for an ensemble of secondary structures for encapsidated MS2 bacteriophage RNA, RNA 18 (2012) 1309–1318, http://dx.doi.org/10.1261/rna.032326.112.

[67] O. Rolfsson, K. Toropova, N.A. Ranson, P.G. Stockley, Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid, J. Mol. Biol. 401 (2010) 309–322, http://dx.doi.org/10.1016/j.jmb.2010.05.058.

[68] R.S. Brown, B.E. Hingerty, J.C. Dewan, A. Klug, Pb(II)-catalysed cleavage of the sugar-phosphate backbone of yeast tRNAPhe—Implications for lead toxicity and self-splicing RNA, Nature 303 (1983) 543–546.

[69] L.A. Kirsebom, J. Ciesiolka, Pb^{2+}-Induced Cleavage of RNA, Wiley-VCH Verlag GmbH, Weinheim, Germany, 2005, http://dx.doi.org/10.1002/9783527619504.ch13.

[70] D. Winter, N. Polacek, I. Halama, B. Streicher, A. Barta, Lead-catalysed specific cleavage of ribosomal RNAs, Nucleic Acids Res. 25 (1997) 1817–1824, http://dx.doi.org/10.1093/nar/25.9.1817.

[71] J. Ciesiołka, D. Michałowski, J. Wrzesinski, J. Krajewski, W.J. Krzyzosiak, Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs, J. Mol. Biol. 275 (1998) 211–220, http://dx.doi.org/10.1006/jmbi.1997.1462.

[72] J.R. Fresco, B.M. Alberts, P. Doty, Some molecular details of the secondary structure of ribonucleic acid, Nature 188 (1960) 98–101, http://dx.doi.org/10.1038/188098a0.

[73] W. Fiers, Chemical structure and biological activity of bacteriophage MS2 RNA, in: N.D. Zinder (Ed.), RNA Phages, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1975, pp. 353–396.

[74] I. Tinoco, C. Bustamante, How RNA folds, J. Mol. Biol. 293 (1999) 271–281, http://dx.doi.org/10.1006/jmbi.1999.3001.

[75] E.A. Skripkin, M.R. Adhin, M.H. de Smit, J. van Duin, Secondary structure of the central region of bacteriophage MS2 RNA. Conservation and biological significance, J. Mol. Biol. 211 (1990) 447–463, http://dx.doi.org/10.1016/0022-2836(90)90364-R.

[76] F. Qu, T.J. Morris, Encapsidation of turnip crinkle virus is defined by a specific packaging signal and RNA size, J. Virol. 71 (1997) 1428–1435.

[77] S. Shakeel, E. Dykeman, S.J. White, et al., PLoS Pathog, 2015 (submitted for publication).

[78] P. Ni, Z. Wang, X. Ma, N.C. Das, P. Sokol, W. Chiu, et al., An examination of the electrostatic interactions between the N-terminal tail of the Brome Mosaic Virus coat protein and encapsidated RNAs, J. Mol. Biol. 419 (2012) 284–300, http://dx.doi.org/10.1016/j.jmb.2012.03.023.

[79] S.E. Bakker, R.J. Ford, A.M. Barker, J. Robottom, K. Saunders, A.R. Pearson, et al., Isolation of an asymmetric RNA uncoating intermediate for a single-stranded RNA plant virus, J. Mol. Biol. 417 (2012) 65–78, http://dx.doi.org/10.1016/j.jmb.2012.01.017.

[80] T.J. Tuthill, D. Bubeck, D.J. Rowlands, J.M. Hogle, Characterization of early steps in the poliovirus infection process: Receptor-decorated liposomes induce conversion of the virus to membrane-anchored entry-intermediate particles, J.

Virol. 80 (2006) 172–180, http://dx.doi.org/10.1128/JVI.80.1.172-180.2006.

[81] A. Gopal, Z.H. Zhou, C.M. Knobler, W.M. Gelbart, Visualizing large RNA molecules in solution, RNA 18 (2012) 284–299, http://dx.doi.org/10.1261/rna.027557.111.

[82] A.M. Yoffe, P. Prinsen, A. Gopal, C.M. Knobler, W.M. Gelbart, A. Ben-Shaul, Predicting the sizes of large RNA molecules, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 16153–16158, http://dx.doi.org/10.1073/pnas.0808089105.

[83] T. Lin, J. Cavarelli, J.E. Johnson, Evidence for assembly-dependent folding of protein and RNA in an icosahedral virus, Virology 314 (2003) 26–33.

[84] J.H. Strauss, R.L. Sinsheimer, Purification and properties of bacteriophage MS2 and of its ribonucleic acid, J. Mol. Biol. 7 (1963) 43–54, http://dx.doi.org/10.1016/S0022-2836(63)80017-0.

[85] M. Hema, A. Murali, P. Ni, R.C. Vaughan, K. Fujisaki, I. Tsvetkova, et al., Effects of amino-acid substitutions in the Brome mosaic virus capsid protein on RNA encapsidation, Mol. Plant Microbe Interact. 23 (2010) 1433–1447, http://dx.doi.org/10.1094/MPMI-05-10-0118.

[86] J. Goecks, A. Nekrutenko, J. Taylor, Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences, Genome Biol. 11 (2010) R86, http://dx.doi.org/10.1186/gb-2010-11-8-r86.

[87] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, http://dx.doi.org/10.1038/nmeth.1923.

[88] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, http://dx.doi.org/10.1093/bioinformatics/btp352.

[89] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

[90] C. Cole, J.D. Barber, G.J. Barton, The Jpred 3 secondary structure prediction server, Nucleic Acids Res. 36 (2008) W197–W201, http://dx.doi.org/10.1093/nar/gkn238.

[91] M.R. Green, J. Sambrook, Molecular cloning: A laboratory manual, 2012, http://dx.doi.org/10.1002/abio.370050118.