Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx1273

Probing of RNA structures in a positive sense RNA virus reveals selection pressures for structural elements

Kyle E. Watters[1], Krishna Choudhary[2], Sharon Aviran[2], Julius B. Lucks[3], Keith L. Perry[4] and Jeremy R. Thompson[4,*]

[1]Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, USA, [2]Department of Biomedical Engineering and Genome Center, University of California Davis, Davis, CA, USA, [3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60201, USA and [4]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

Received August 11, 2017; Revised December 07, 2017; Editorial Decision December 09, 2017; Accepted December 18, 2017

ABSTRACT

In single stranded (+)-sense RNA viruses, RNA structural elements (SEs) play essential roles in the infection process from replication to encapsidation. Using selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq) and covariation analysis, we explore the structural features of the third genome segment of cucumber mosaic virus (CMV), RNA3 (2216 nt), both *in vitro* and in plant cell lysates. Comparing SHAPE-Seq and covariation analysis results revealed multiple SEs in the coat protein open reading frame and 3′ untranslated region. Four of these SEs were mutated and serially passaged in *Nicotiana tabacum* plants to identify biologically selected changes to the original mutated sequences. After passaging, loop mutants showed partial reversion to their wild-type sequence and SEs that were structurally disrupted by mutations were restored to wild-type-like structures via synonymous mutations *in planta*. These results support the existence and selection of virus open reading frame SEs in the host organism and provide a framework for further studies on the role of RNA structure in viral infection. Additionally, this work demonstrates the applicability of high-throughput chemical probing in plant cell lysates and presents a new method for calculating SHAPE reactivities from overlapping reverse transcriptase priming sites.

INTRODUCTION

The structure of viral RNA (vRNA) is intimately involved in a number of essential viral functions such as replication, translation and encapsidation. Determination of vRNA structure has led to a broadening of our understanding of its involvement in the virus infection cycle (1,2), and research on plant viruses has played a central role in expanding our knowledge of RNA structure. For example, in the 1980s RNA structural predictions of the 3′ terminus of turnip yellow mosaic virus suggested a structure that was similar to that of a tRNA, but contained a new RNA structural motif later termed a pseudoknot (3,4). Similar tRNA-like structures were also confirmed for brome mosaic virus (BMV) (5,6) and tobacco mosaic virus (7,8) leading to the dissection of their function by mutational analyses (9–11).

Cucumber mosaic virus (CMV; species *Cucumber mosaic virus*, genus *Cucumovirus*, family *Bromoviridae*) has arguably the widest known host range of any virus, infecting over 1200 species in >100 plant families (12) including many agronomically important crops (13). As a result, it is one of the most intensively studied of plant viruses (14). It has a tripartite positive sense single-stranded genome that encodes five genes (Figure 1A). The longest genome segment, RNA1 (ca. 3200 nt), encodes the 1a protein which contains methyltransferase and helicase domains. RNA2 (ca. 3000 nt) encodes two proteins: 2a, containing an RNA-dependent RNA polymerase domain, and the overlapping 2b silencing suppressor. The bicistronic RNA3 encodes the movement protein (MP) and coat protein (CP) separated by the non-coding intergenic region (IGR). The 3′ positioned CP is translated through the subgenomic RNA4, the promoter of which lies in the IGR (15). Each RNA segment is capped at the 5′ terminus and has a highly conserved tRNA-like pseudoknot structure at the 3′ terminus (16–18).

RNA structural determination of CMV began almost simultaneously in conjunction with that of its relative BMV (5,19). The 135 3′ terminal nucleotides of all three genome segments are able to assume a tRNA-like conformation that facilitates negative strand synthesis (20,21). Within the tRNA-like structure, one stem–loop structure (SLC) has

*To whom correspondence should be addressed. Tel: +1 607 255 0872; Email: jrt36@cornell.edu

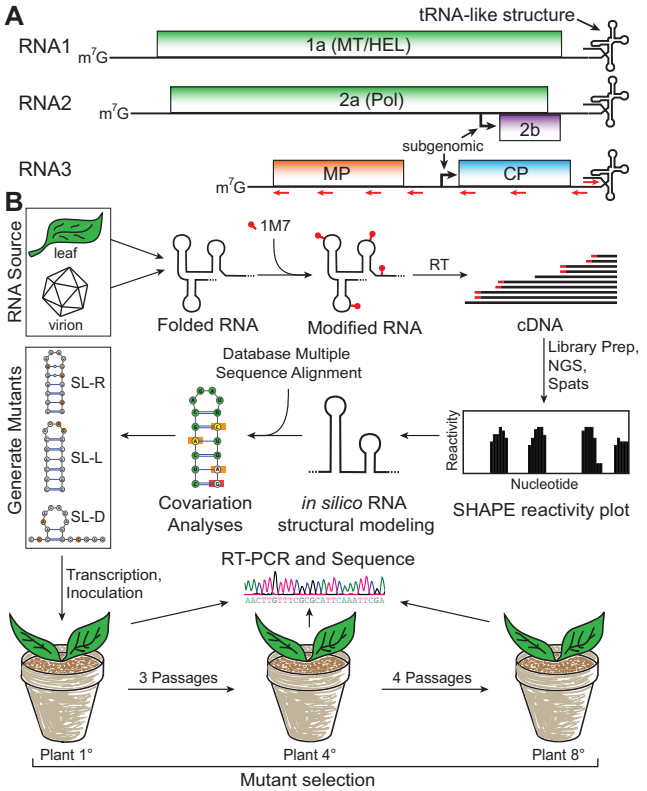

Figure 1. CMV genome organization and experimental overview. (**A**) Genome organization of cucumber mosaic virus (CMV) viral RNA (vRNA). The three genomic (+)-sense single stranded segments of CMV (RNA1, RNA2, and RNA3) are depicted with the ORFs they encode. Proteins CP and 2b are expressed via a subgenomic RNAs, with the approximate promoter positions marked with arrows. All genome segments have a 5′ m^7G cap and a tRNA-like structure at the 3′ end. RNA3 contains two ORFs, for the movement protein (MP) and the coat protein (CP), separated by a non-coding intergenic region (IGR). Reverse transcription priming sites used for SHAPE-Seq are denoted by red arrows. (**B**) Experimental overview. vRNA is first isolated from infected leaf or purified viral particles and analyzed with SHAPE-Seq. Restrained RNA folds from *in silico* structural modeling, using the obtained chemical probing data, and covariation analysis are performed to identify conserved RNA structural features across the entire CMV RNA3. Last, a subset of the identified conserved features are mutated and assayed for their stability *in planta*.

been shown to interact with the viral replicase and is required for RNA synthesis (22,23), with two nucleotides (5′-CA-3′) in the trinucleotide loop of SLC being essential for replicase interaction (10). Conserved structures for the complete 3′ UTR of CMV RNA3 have been identified using a combination of enzymatic probing and covariation analyses (6,24,25), but much of their function is unknown, although some are associated with high levels of recombination (24–26). Aside from the 3′ UTR, the only other region in which functional RNA structural elements (SEs) have been inferred is the IGR. The IGR of BMV contains a highly conserved box-B motif that, in yeast, forms a hairpin loop analogous to the TΨC-stem loop in tRNA (27). The same motif is present in the IGR and 5′ UTRs of RNA1 and RNA2 in CMV and other related viruses. The interaction of box-B with the 1a protein and its involvement in replication has been demonstrated within the 5′ UTR of RNA1 and RNA2 of BMV (28–30).

Much of the RNA structural work performed to date has focused on small, isolated regions of the viral genome. However, the recent development of many high-throughput probing techniques (31,32) is allowing for the complete coverage of viral genome structures in a single experiment. The parallel development of computational methods for comparative and integrative analysis of probing data has made it possible to recover biological insights from data generated in these experiments (32,33). To date, large-scale structural information exists for three ssRNA positive sense viral genomes: hepatitis C virus (HCV) (34), human immunodeficiency virus (35), and tomato bushy stunt virus (TBSV) (36) as well as the satellite tobacco mosaic virus (37,38). By structurally characterizing large regions of viral genomes at once, potentially interesting structures can be discovered much more rapidly, and in the proper context. Such information is enabling research into a more global understanding of the functionality of viral RNA genomes with the recent identification of RNA SEs within previously unexplored protein coding regions (34,36,39).

In this study, we use selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq) v2.1 (40) and covariation analysis to analyze the complete sequence of RNA3 from CMV strain Bn57 (41). Using SHAPE-Seq, we directly compare nucleotide flexibility across RNA3 in infected cell lysates and *in vitro* refolded purified viral RNA and present predicted secondary structure maps of RNA3 using SHAPE-Seq data to restrain secondary structure predictions. In agreement with previous studies, we observe structural features in the 3′ UTR consistent with the tRNA-like structure and viral replicase binding. Additional covariation analysis on RNA3 identified conserved RNA SEs in the 3′ UTR and CP ORF that were supported by the SHAPE-Seq analysis. Last, we test the biological relevance of the identified SEs by passaging mutants designed to disrupt these structures *in planta* (Figure 1B). The results of these experiments strongly support the existence of RNA SEs within the CP ORF.

MATERIALS AND METHODS

Plants, inoculations, virus passaging and virion purification

Plants of *Nicotiana tabacum* were maintained in either greenhouse conditions of 20 ± 3°C and 16 h:8 h of light:dark or growthroom conditions of 22 ± 2°C, 16 h:8 h light (100 μmol/m^2/s):dark and 50% relative humidity. Initial inoculations were done by rubbing a carborundum dusted leaf of a healthy plant at a 4–6 leaf stage of growth with 1 μl (∼4 μg) of a mix of Bn57-CMV RNA1, RNA2 and RNA3 transcripts generated from infectious clones pBn57-1, pBn57-2 and pBn57-3 (or its mutant derivatives), respectively. To passage the virus, once symptoms appeared in the inoculated plant, approximately 100 mg of leaf material was ground in 5 ml 0.1 M Na_2HPO_4 with a sterilized mortar and pestle and rubbed onto a carborundum dusted leaf of a healthy plant at a 4–6 leaf stage of growth. This process was carried out for eight passages in total with infected leaf material being collected and snap frozen for storage. All

passages were carried out in greenhouse conditions. Virions were extracted from 1 g of leaf material taken from the primary (transcript initiated) inoculated plants, resuspended in 300 μl 5 mM sodium borate (pH 9.0) and 0.5 mM EDTA, and stored in 50% glycerol at –20°C.

Viral RNA purification

Virion RNA was isolated by taking 100 μl of purified virions and adding 400 μl phenol, 100 μl chloroform and 100 μl disruption buffer (200 mM Tris pH 8.5, 1 M NaCl, 2 mM EDTA, 1% SDS). The mix was then vortexed for 30 s and centrifuged at 13 000 × g for 5 min. The supernatant was then transferred to a new tube along with 100 μl phenol and 50 μl chloroform vortexed again for 30 s and centrifuged at 13 000 × g for 5 min. Nucleic acids in the resulting supernatant were then ethanol precipitated and resuspended in 50 μl sterile distilled water.

Cell lysates and RNA modification

Cell lysates were prepared as described by Hacisuleyman et al (42), with some minor modifications, by grinding 20 mg of symptomatic *N. tabacum* leaf, inoculated, two weeks prior, with Bn57-CMV, in lysis buffer (150 mM KCl, 25 mM Tris pH 7.5, 5 mM EDTA, 5 mM $MgCl_2$, 0.5% NP-40, 1× HALT protease inhibitor cocktail (ThermoFisher Scientific), 0.5 mM DTT, 100 U/ml RNase OUT (ThermoFisher Scientific)), and rapidly snap freezing in liquid nitrogen before storing at –80°C. To perform SHAPE modification on cell lysates, the frozen lysates were first thawed slowly on ice, then quickly spun to pellet cell debris. The cleared lysates were then incubated at 20°C for 15 min. Then, 180 μl of the cleared lysates was added to either 20 μl dimethyl sulfoxide (DMSO; (–) control) or 65 mM 1-methyl-7-nitroisatoic anhydride (1M7; (+) sample) in DMSO and incubated for 5 min at 20°C. 600 μl of TRIzol reagent (Life Technologies) was added to the lysate samples and extracted according to the manufacturer's protocol using 20 μg of glycogen as a carrier. The resulting pellet was dissolved in 10 μl RNase-free H_2O. To modify *in vitro* refolded viral RNA samples, 3 μg of viral RNA was diluted to 12 μl in RNase-free H_2O then incubated for 95°C for 2 min and snap-cooled on ice for 1 min. The viral RNA was refolded by adding 6 μl of 3.3× folding buffer (333 mM HEPES, 333 mM NaCl, 33 mM $MgCl_2$) and incubated at 20°C for 20 min. Then, the 18 μl of folded viral RNA was split, adding 9 μl of RNA to 1 μl of either DMSO or 65 mM 1M7. The RNAs were modified for 3 min at 20°C then ethanol precipitated, using glycogen as a carrier, and dissolved in 10 μl RNase-free H_2O.

Reverse transcription

To the 10 μl of modified RNA (or unmodified control), 3 μl reverse transcription primer mix was added, containing 0.5 μM of each oligonucleotide A-H (Supplementary Table S1), spanning the length of RNA3. The resulting mix was heated to 95°C for 2 min, then incubated at 65°C for 5 min before placing on ice for ∼30 s. Next, 7 μl of SSIII master mix was added, containing: 0.5 μl of Superscript III (Life Technologies), 4 μl 5× First Strand Buffer (Life Technologies), 1 μl 100 mM DTT, 1 μl 10 mM dNTPs, and 0.5 μl RNase-free H_2O. The complete reaction mix was then incubated at 42°C for 1 min, followed by extension at 52°C for 25 min and deactivation at 65°C for 5 min. The RNA was hydrolyzed by addition of 1 μl of 4 M NaOH solution and heating to 95°C for 5 min. The basic solution containing the cDNA was partially neutralized with 2 μl of 1 M HCl and precipitated with 69 μl cold EtOH with thorough washing with 70% EtOH. The washed pellet was dissolved in 22.5 μl of nuclease-free H_2O.

Adapter ligation

To the cDNA, 3 μl 10× CircLigase Buffer (Epicentre), 1.5 μl 50 mM $MnCl_2$, 1.5 μl 1 mM ATP, 0.5 μl 100 μM DNA adapter (oligonucleotide I; Supplementary Table S1), and 1 μl CircLigase I (Epicentre) were added. The reaction was incubated at 60°C for 2 h, then 80°C for 10 min to inactivate the ligase. The ligated DNA was EtOH precipitated with 20 mg glycogen as a carrier and dissolved in 20 μl of nuclease-free H_2O. Then the cDNA was purified using 36 μl of Agencourt XP beads (Beckman Coulter), according to manufacturer's instructions and eluted with 20 μl TE buffer.

Quality analysis

For quality analysis, a separate PCR reaction for each (+) and (–) sample was mixed by combining: 13.75 μl nuclease-free H_2O, 5 μl 5× Phusion Buffer (New England Biolabs), 0.5 μl 10 mM dNTPs, 1.5 μl of 1 μM labeling primer (oligonucleotides J/K; Supplementary Table S1), 1.5 μl of 1 μM primer PE_F (oligonucleotide L; Supplementary Table S1), 1 μl of 0.1 μM selection primer mix (0.1 μM each of oligonucleotides M-T or U-AB; Supplementary Table S1), 1.5 μl ssDNA library (+ or –), and 0.25 μl Phusion DNA polymerase (New England Biolabs). Both fluorescent primers were purchased from Applied Biosystems and the selection primers were purchased from Integrated DNA Technologies. Phosphorothioate modifications were added to prevent the 3′→5′ exonuclease activity of Phusion polymerase (Supplementary Table S1). Amplification was performed for 15 cycles, then 50 μl nuclease-free H_2O was added, and the diluted reaction was ethanol precipitated. The resulting pellet was dissolved in formamide and analyzed with an ABI 3730xl capillary electrophoresis instrument.

Library preparation and next generation sequencing

To construct sequencing libraries, a separate PCR for each (+) and (–) sample was mixed by combining: 33.5 μl nuclease-free H_2O, 10 μl 5× Phusion Buffer (New England Biolabs), 0.5 μl 10 mM dNTPs, 0.25 μl of 100 μM TruSeq indexing primer (oligonucleotide AC (Supplementary Table S1)), 0.25 μl of 100 μM primer PE_F, 2 μl of 0.1 μM selection primer mix (+ or –, as noted above), 3 μl ssDNA library (+ or –), and 0.5 μl Phusion DNA polymerase (New England Biolabs). Amplification was performed as indicated in 'Quality analysis' above. Completed reactions were chilled at 4°C for 2 min before addition of 5 U exonuclease I (New

England Biolabs) to remove unextended primer. The reactions were then incubated at 37°C for 30 min. After incubation, the libraries were purified using 90 μl of Agencourt XP beads (Beckman Coulter) according to manufacturer's instructions. The complete libraries were eluted with 20 μl TE buffer and quantified with the Qubit 2.0 Fluorometer (Life Technologies). To prepare the libraries for sequencing, the average length of each sample was determined using the results from the quality analysis in order to calculate the molarity of each (+) or (–) reaction separately. Sequencing pools were mixed to be equimolar, such that all of the sequencing libraries were present in the solution at the same level. Sequencing was performed on the Illumina MiSeq using 2 × 35 bp paired end reads.

Data analysis with Spats

Analysis of sequencing reads was performed with Spats v1.0.1 (https://github.com/LucksLab/spats/releases/), using cutadapt v1.5 (43) and Bowtie 0.12.8 (44) to remove sequencing adapters and subsequently map the trimmed reads. Details of the statistical model that was used to obtain reactivity estimates for overlapping reverse transcriptase priming sites from the mapped reads output by Spats are described in detail in Supplementary Data Analysis. Briefly, reads were summarized as *stop counts*, X_k and *pass counts*, C_k for every nucleotide k separately for each of the DMSO and 1M7 channels for each replicate. X_k is the number of reads that resulted from termination of reverse transcription at *k-1*, where *l* is the length of the primer. These reads map starting one nucleotide downstream of k. C_k is the number of reads that extended to nucleotide k and hence, indicate absence of termination induced by noise/modification at k. These reads map starting at or upstream of k. Consequently, $(X_k + C_k)$ is the local coverage at nucleotide k, i.e. the number of reads that have information about the state of modification or noise at k (Supplementary Figure S1). Then, for each replicate, we calculated the stop frequency as $X_k/(X_k + C_k)$ in corresponding DMSO and 1M7 channels. Using superscripts '–' and '+' to denote DMSO and 1M7 channels respectively, reactivity β_k was estimated as

$$\hat{\beta}_k = \max\left(\frac{X_k^+/(X_k^+ + C_k^+) - X_k^-/(X_k^- + C_k^-)}{1 - X_k^-/(X_k^- + C_k^-)}, 0\right).$$

This yielded a reactivity profile for each replicate. Notably, when noise levels are very low, as in this study, reactivity calculation can be simplified and well-approximated by setting the reactivity estimate as the numerator in the equation above (45), that is,

$$\hat{\beta}_k^* = \max\left(X_k^+/(X_k^+ + C_k^+) - X_k^-/(X_k^- + C_k^-), 0\right).$$

Reactivities (β) were subsequently normalized such that the average reactivity for each profile was one. This was done by dividing each nucleotide's reactivity by sum of all reactivities for that profile and multiplying by the number of nucleotides for which reactivity data exists (2191 nts). Reactivity data was deposited in the RNA Mapping Database (https://rmdb.stanford.edu/) (46) with accession numbers: CMVBN3_1M7_0001 (lyRNA), CMVBN3_1M7_0002 (viRNA).

Mutagenesis

Infectious transcripts from clones of the CMV isolate Bn57 were generated as described previously using the mMessage mMachine® T7 transcription kit (ThermoFisher Scientific, Waltham, MA, USA). PCR-mediated mutagenesis of pBn57-3, containing RNA3 of Bn57-CMV, was done using PfuUltra II Fusion HS DNA Polymerase (Agilent, Santa Clara, CA, USA) according to the manufacturer's instructions. For the primary PCR step external primers ReFor, ReRev were used in combination with the corresponding *rev* and *for* primers for each viral structural domain, respectively (Supplementary Table S2). The fused PCR product was then digested with SacII and XbaI and ligated into the corresponding sites in the vector pBn57-3. All resulting mutant clones were sequenced to confirm the presence of the introduced mutations and the absence of any spurious nucleotide changes within the amplified fragment using primers Refor and ReRev (Supplementary Table S2).

In silico structural analyses and selection pressure determination

CMV CP sequences from members of subgroup I ($n = 71$) were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/) (Supplementary Table S4) and aligned using R-Coffee (47). The generated alignment was then analyzed by PPfold (48), assigning confidence levels to conserved RNA structures. Amino acid coding consistency in the mutated sequences was confirmed in Vector NTI (ThermoFisher Scientific). For covariation comparisons of specific structural regions global (subgroup I and II) CMV sequences ($n \leq 431$) were obtained from the NCBI database, aligned using R-Coffee, and manually filtered to remove non-informative sequences.

RT-PCR and virus sequencing

400 ng RNA, extracted from infected plant material using the RNeasy Plant Mini kit (Qiagen), was used as a template for reverse transcription with a CMV specific reverse primer (p1033.CMV-Bn57RNA3; Supplementary Table S2) using M-MLV Reverse Transcriptase (ThermoFisher Scientific) according to the manufacturer's instructions. The cDNA generated was then amplified by PCR using PfuUltra II Fusion HS DNA Polymerase (Agilent) with primers p766 and p1033.CMV-Bn57RNA3 (Supplementary Table S2) to yield a 1071 bp product encompassing the CP and 3′ UTR that were sequenced at the Cornell Genomics Facility (http://www.biotech.cornell.edu/brc/genomics-facility) in both directions. The rest of the CMV genome was not sequenced. Samples containing atypical (non-engineered) sequences were re-amplified and sequenced. Sequencing of purified virions was done using RNA extracted prior to RT-PCR as described earlier.

RESULTS

CMV RNA3 has a highly branched structure *in vivo* and *in vitro*

The secondary structure of CMV RNA3 was mapped with SHAPE-Seq from two different sources of CMV RNA: (i) *in vitro* refolded, purified viral RNA extracted from virions (viRNA) and (ii) infected cell lysates (lyRNA). We compared both of these sources because while obtaining data from virions is more tractable, plant cell lysates most closely approximate *in vivo* conditions; its use here offering the most practically feasible approach to studying the RNA structure of the viral population *in planta*. Eight target-specific priming locations were selected to obtain SHAPE reactivity information for the entire RNA3 sequence (Figures 1A and 2A). SHAPE reactivity values represent the flexibility of each nucleotide such that large values suggest a higher likelihood of single-strandedness (and vice versa) (49) and can be used to restrain folding algorithms to improve RNA structural predictions (45). In SHAPE experiments, a SHAPE chemical probe (such as 1-methyl-7-nitroisatoic anhydride (1M7), used in this study) reacts with an RNA molecule's accessible 2′-OH groups in a structure-dependent fashion. The SHAPE adduct causes reverse transcription (RT) to stop in a primer extension reaction. In SHAPE-Seq, these RT stops are detected using Illumina sequencing (40,50–52). Last, RT stops are mapped and used to calculate nucleotide reactivity values across the RNA.

To calculate reactivity values across RNA3, we used a stratified priming approach with eight priming locations spaced so as to ensure consistently high sequencing coverage over the entire stretch of 2216 nt (Figure 1A and Supplementary Figure S1) for three and five replicates of the viRNA or lyRNA SHAPE-Seq measurements, respectively. Local coverages at each nucleotide in all replicates at both conditions were greater than $\sim 10^3$, and reaction conditions were optimized to ensure a high level of confidence in chemical reactivity estimates (53). To estimate chemical reactivities from sequencing reads, we extended previous work on modeling SHAPE-Seq data obtained with a single targeted primer (54,55) to our stratified priming approach. The extended model and adjoined maximum-likelihood estimation method rigorously fuse reads obtained from multiple priming sites into a single reactivity profile by explicitly accounting for discrepancies in primer binding efficiencies and for terminations due to roadblocking by neighboring primers (see Supplementary Information for details). To inform comparative analysis between probing conditions, replicate data was inspected with the SEQualyzer platform for data exploration and quality control at both nucleotide and regional levels (56,57). Last, calculated reactivities were normalized such that the average reactivity value (ρ) was one across the entire RNA3 (Figure 2A and Supplementary Figure S2).

Using the complete RNA3 reactivity map, we obtained a secondary structure of RNA3 using the *ShapeKnots* algorithm with SHAPE restraints (Figure 3) (58). The predicted structures contain a number of previously identified elements (5,19,24) within the 3′ UTR, most notably two stem–loops: SLC (nts 2140–2175) and SLF (nts 2069–2089). These stem–loops display a typical low-high-low reactivity pattern that is frequently observed for RNA hairpins (Figure 2B). Other previously identified (5,19) 3′ UTR structures with moderate or highly reactive ($\rho > 0.5$) bases within predicted loops include SLB (nts 2182–2199), SLE (nts 2100–2019), and SLG (nts 2052–2067). Within the 3′ UTR, the observed reactivities agree well with both the previously recognized cruciform (5) and pseudoknotted structures (19). We also observe a TΨC stem–loop and box-B motif in the IGR at nts 1116–1130 analogous to those in the BMV replication enhancer (27). Additionally, both structures contain a predicted pseudoknot outside of the 3′ UTR (only one pseudoknot is allowed with *ShapeKnots*, preventing the observation of the pseudoknot in the tRNA-like structure). In the structure derived from the viRNA reactivities, a pseudoknot is predicted between nts 942–948 just upstream of the MP stop codon and nts 1484–1490 in the CP ORF, however, the structure restrained with the lyRNA-derived reactivities suggests a different pseudoknot in the CP ORF between nts 1368–1372 in SL1362 and nts 1218–1222 (Figure 3). It should be noted however that possible differences in free Mg^{2+} between the cell lysates and viRNA refolding buffer may contribute to the observed differences in the predicted pseudoknots.

Because we observe two predicted pseudoknots other than the known pseudoknot in the tRNA-like region of the 3′ UTR, we asked whether the known pseudoknot would be observed if either of the newly predicted pseudoknots were suppressed during computational folding. Therefore, we refolded RNA3 using the lyRNA-derived reactivities after forcing either nts 1218–1222 or nts 1368–1372 to be single-stranded. After effectively 'breaking' this pseudoknot, *ShapeKnots* predicted the formation of the tRNA-like pseudoknot with a few small rearrangements in the main structure of RNA3 (Supplementary Table S3). A similar analysis using the viRNA-derived reactivities demonstrated that the pseudoknot in the tRNA-like region is only predicted when both sides of the viRNA-derived pseudoknot are forced to be single-stranded, with an alternative pseudoknot forming when only nts 1484–1490 are forced single-stranded (Supplementary Table S3). Our ability to recover the pseudoknot in the tRNA-like region by preventing the formation of the newly predicted pseudoknots suggests that multiple pseudoknots may co-exist in the RNA3 structure, including the one frequently observed in the 3′ UTR. Further study will be necessary to determine if any of these predicted long-range interactions occur during the viral life cycle.

We observe a number of short stem–loops common between the lyRNA and viRNA restrained structures, but much of the internal base pairing in the ORFs and IGR differs for the two RNA3 structures. Because the predicted structures only differ in their reactivity maps (Figure 2A and Supplementary Figure S2), the predicted structural variations must come from changes in RNA accessibility between the two experimental conditions. While we observe some significant reactivity differences ($P < 0.05$) that did not create differences in the predicted structures, others, such as the differences within the 83–89 and 318–338 nt ranges resulted in different local structural contexts in the 5′ UTR and MP ORF, respectively (Figure 3). There are

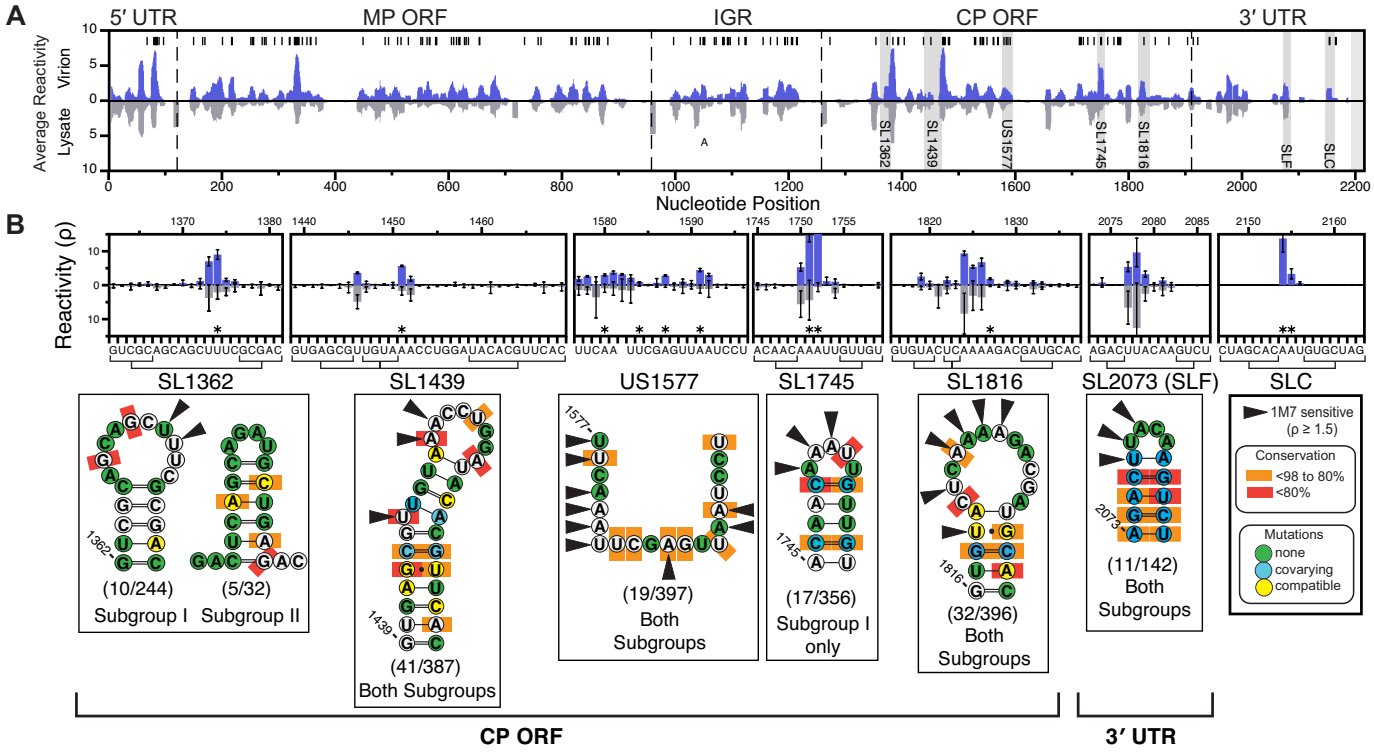

Figure 2. SHAPE-Seq reactivities across the complete RNA3 of Bn57-CMV. (**A**) Histogram of a 10 nt moving average of the SHAPE-Seq reactivities measured for the viRNA (blue, top) and lyRNA (gray, bottom). Significant differences (*t*-test; $P < 0.05$) between individual nucleotide viRNA and lyRNA reactivity values are indicated with a black vertical bar. Individual nucleotide reactivities are shown in Supplementary Figure S2 and can be downloaded from the RNA Mapping Database (https://rmdb.stanford.edu/) (46) with accession numbers CMVBN3_1M7_0001 (lyRNA), CMVBN3_1M7_0002 (viRNA). The high average lyRNA reactivities at the UTR boundaries are anomalous and are due to one replicate exhibiting uncharacteristically high reactivity at three positions. (**B**) SHAPE-Seq reactivity values (top) and covariation analysis results (bottom) for six conserved RNA SEs (SL – stem loop, US – unstructured) in RNA3. Brackets indicate base pairing in structural elements for strain Bn57, which differs from the consensus in SL1816 and SLF. Reactivity error bars represent the standard deviation of three (viRNA) or five (lyRNA) replicates. Significant differences between the viRNA and lyRNA reactivities (*t*-test, $P < 0.05$) are marked with asterisks. In the covariation analysis, SL1362 was observed as two distinct structures for each subgroup (I and II), and the fractions in parentheses denote the number of unique sequences represented within the structure relative to the total number of database sequences aligned. Bases that are not as strongly conserved are denoted with orange (80–98% conserved) or red (<80% conserved) rectangles.

three most likely causes for these types of observed differences: (i) protein/cofactor binding to RNA3 in the lysates changing the reactivity map (directly observable for SLC), making it less representative for folding prediction, (ii) refolding the viRNA *in vitro* may result in a different folded structure, yielding different SHAPE reactivities or (iii) the lyRNA reactivity maps' higher level of experimental noise may be affecting some of the folding motifs in the predicted structure.

SHAPE-Seq and covariation analysis identify conserved stem–loops in the CP ORF

Given the appearance of several stem–loops in common between the structures obtained from the lyRNA and viRNA reactivity maps, and a general lack of structural analysis in vRNA ORFs, we decided to examine the CP ORF for potential uncharacterized RNA SEs. We chose the CP ORF over the MP ORF for reasons of both experimental continuity (CP is proximal to the already well-characterized 3′ UTR) and anticipated biological impact, as the presence of subgenomic RNA4 results in higher levels of the CP ORF compared to MP ORF during infection. To look for evidence of SEs in the CP ORF, we used the program PPfold (48) that predicts a consensus secondary structure from RNA alignments. Analysis of an alignment (Supplementary Table S4 and Figure S3) of subgroup I sequences ($n = 71$; identities 89.9–99.7%), to which Bn57 belongs, generated similar structures to those obtained with *ShapeKnots* and the SHAPE-Seq data with varying levels of statistical support. Specifically, we identified four structures in the CP ORF (SL1362, SL1439, SL1745, and SL1816) that agreed with the SHAPE-restrained structural predictions and contained 1M7 sensitive ($\rho \geq 1.5$) nucleotides predicted to occur within loops and bulges (Figures 2B and 3 and Supplementary Figures S3 and S4). We then extended the alignments to include all CMV GenBank sequences ($n \leq 431$), irrespective of subgroup, focusing specifically on each of the previously predicted structures. We also included two additional regions in the analysis: (i) the previously described 3′ UTR stem–loop SL2073 and (ii) a 20 nt region in the CP ORF predicted to be unstructured (US1577) from SHAPE-Seq analysis. The extent of the variation in all six of these regions, including covariation and compatible base changes, is shown in Figure 2B. The most striking result is the variation within SLF (SL2073). All five stem nts of SLF covary

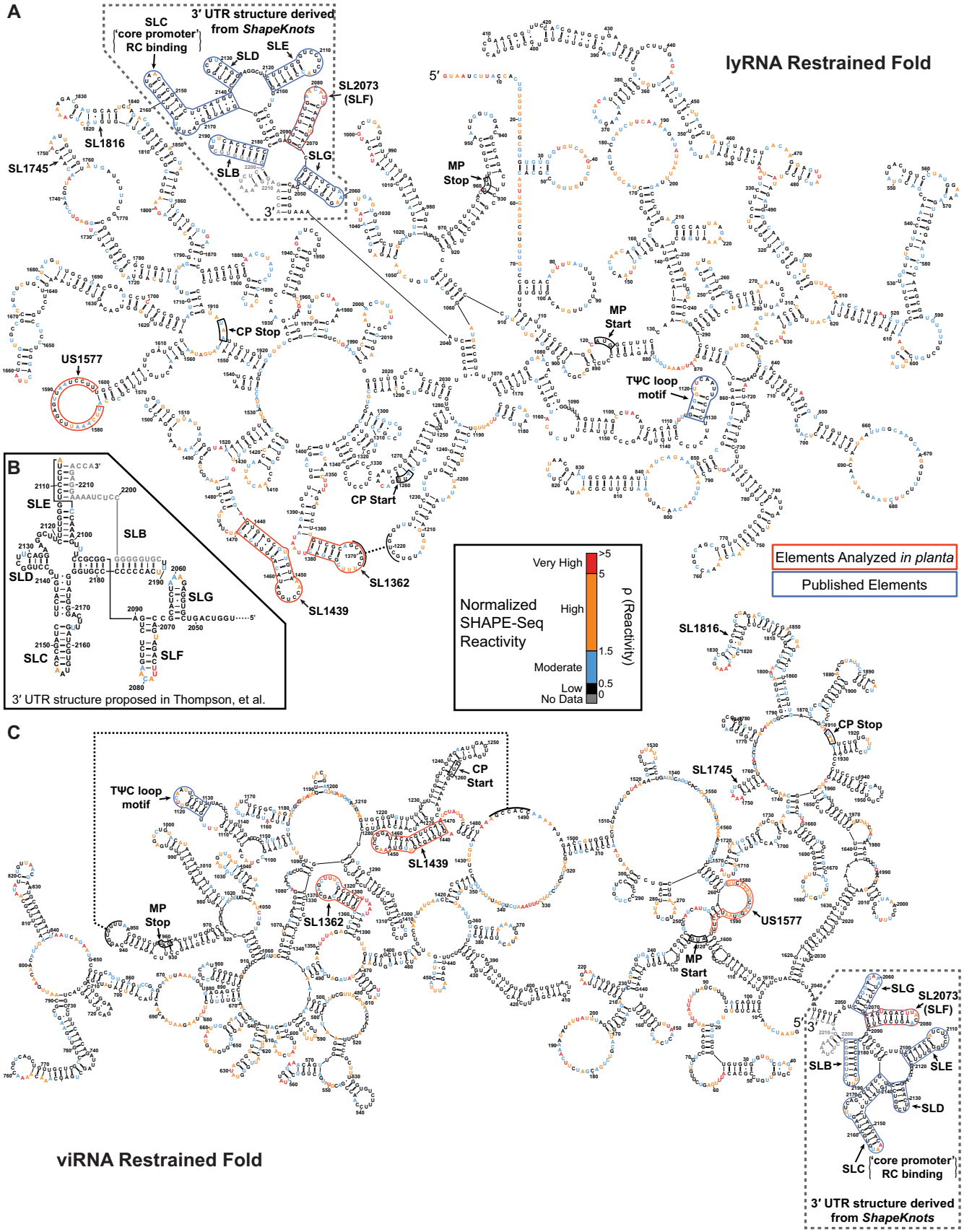

Figure 3. Predicted secondary structures of CMV RNA3 restrained with lyRNA and viRNA SHAPE-Seq data. (**A**) The secondary structure shown represents the minimum free energy structure prediction of the CMV RNA3 using the *ShapeKnots* algorithm of *RNAstructure* (51) restrained with the

while the tetraloop UACA is fully conserved. However, in stem–loops found within the CP ORF the level of covariation is much lower, and the level of conservation higher, although this is not necessarily unexpected give the additional restraint of maintaining the amino acid sequence (Figure 2B).

Disruption of SL2073 is strongly selected against *in vivo*

To examine the significance of these predicted structures we first mutated the well-characterized stem–loop SL2073, better known as SLF (5) or B3 (5,8) and examined how the mutants behave *in planta*. We created three mutants (Figure 4A)—one with a disrupted stem–loop (D-SL), another in which the sequence was changed but the predicted structure was maintained or restored (R-SL), and another in which only the loop sequence was altered (L-SL). These mutants were then individually inoculated onto *N. tabacum* and passaged to allow for the establishment of the fittest genotype. Viral RNA was extracted at the fourth (4°) and eighth (8°) passages and the sequence was compared with the original input RNA mutant. In three independent experiments, there was no instance of R-SL2073 or L-SL2073 diverging from the input sequence. However, the disrupted stem–loop in D-SL2073 consistently showed partial reversion of the predicted stem–loop to a structure that contained a hexaloop (GUACAA) instead of the wild-type tetraloop (UACA). Two types of changes occurred: A2074 reverted to the wild-type G or the wild-type C2085 mutated to U to allow pairing with the mutant A2074. Both structural reversions were subsequently found to be the dominant sequence in the initial (1°) recipient plant upon the appearance of systemic symptoms.

Mutation of a predicted unstructured region in the CP ORF remains unchanged during passaging

We next examined what effect mutating a predicted unstructured region might have on the virus. We introduced synonymous mutations at various nucleotide positions within US1577 (nts 1577–1597), a region predicted to be unstructured, to generate two independent mutants (m1 and m2) and analyzed them using the same method as for SL2073 (Figure 4B). In three independent experiments neither mutant sequence was observed to change from the originally mutated inoculated transcript, despite having introduced a number of codons under-represented in the CMV genome that could potentially be selected against due to codon usage bias (Supplementary Table S5).

Stem stability of SL1362 and SL1439 is selected for *in vivo*

Next, we wanted to test how SEs within the CP ORF would behave using the same methodologies as applied for SL2073 and US1577. Stem–loop SL1362 was mutated in three ways analogously to SL2073 (D-SL, R-SL and L-SL), but with the added criterion of only introducing synonymous mutations (Figure 4C). Experiments using D-SL1362 showed that at the 4° or 8° passage the mutant's predicted structure was partially reverted to a bulge-containing stem–loop that was achieved by a single nucleotide change on either side of the stem as similarly observed for SL2073. Also, as with SL2073, the altered base is immediately adjacent to the closing base pair of the stem. Interestingly, both changes are synonymous; C1365 changes to U to maintain a serine and A1380 mutates to G to maintain an arginine. The resulting stem consists of five base pairs like the wild-type, but also includes a bulge at A1367.

The disrupted stem–loop mutant of SL1439 (D-SL1439) is unstable in the initial (1°) recipient plant. In all three passages with D-SL1439, the engineered base C1440 reverts to the wild-type U, favoring the formation of a larger loop and a single, short stem that is closer to the wild-type structure (Figure 4D). Importantly, when the SL1362 and SL1439 SEs were mutated to the restored structures (R-SL1362 and R-SL1439), maintaining the predicted wild-type structures, the introduced mutations remained stable throughout passaging (Figure 4C, D).

Unlike SL2073, mutations to the loop in the SL1362 predicted structure results in a gradual reversion, with a single base, A1371, reverting to the wild-type G. Similarly, one of the loop mutations in L-SL1439 results in a A1455U change, restoring the wild-type. Both of these positions are centrally located within their respective loops. Concurrent with the A1371G reversion in L-SL1362 and the C1365U covariation in D-SL1362, there was a consistent distal change at C1487 to U, rendering an amino acid shift from a proline to a leucine. This change occurred gradually at the 4° and 8° passages (Supplementary Figure S4).

DISCUSSION

The structure-function relationship of RNA structures contained within the genome of RNA viruses is still largely unexplored, partly due to the size and complexity of viral genomes. With the advent of new molecular probing methods, global folding predictions of entire molecules has become a tractable endeavor. Recently, bioinformatics analyses of (+)-sense ssRNA viruses have found evidence for genome-scale ordered RNA structure. Genome-scale struc-

←

averaged lyRNA reactivities ($m = 1.1$ and $b = -0.3$). Nucleotides are colored by average reactivity intensity as indicated in the legend. The six identified structural elements are marked, as well as the CP and MP start and stop codons. The stem–loops analyzed *in planta* (Figure 4) are outlined in red and other notable, published elements that have been observed in CMV or BMV are marked in blue (5,6). A potential pseudoknot was identified by *ShapeKnots* between nts 1368–1372 and 1218–1222. The predicted 3′ UTR structure is outlined in with gray dashes for comparison to (B). (**B**) Structure of the 3′ tRNA-like region as proposed by Thompson *et al.* (25) shown for comparison to the predicted structure using the lyRNA (A) and viRNA (C) reactivities. Due to the limit of one pseudoknot in the fold prediction, the pseudoknot in the 3′ tRNA-like region could not be observed in the complete RNA3 structure predictions using *ShapeKnots*. (**C**) Predicted secondary structure of CMV RNA3 restrained with viRNA SHAPE-Seq data. The folding method and annotations are the same as described in (A). A potential pseudoknot was identified by *ShapeKnots* between nts 942–948 and 1484–1490, which differs from the pseudoknot predicted for the lyRNA restrained structure in (A). Other notable differences between the viRNA and lyRNA structures include much lower reactivities in the loop of the core promoter SLC in the lyRNA structure, indicative of replicase binding, and a few bases of US1577 that are predicted to be paired in the viRNA structure and unpaired in the lyRNA structure.

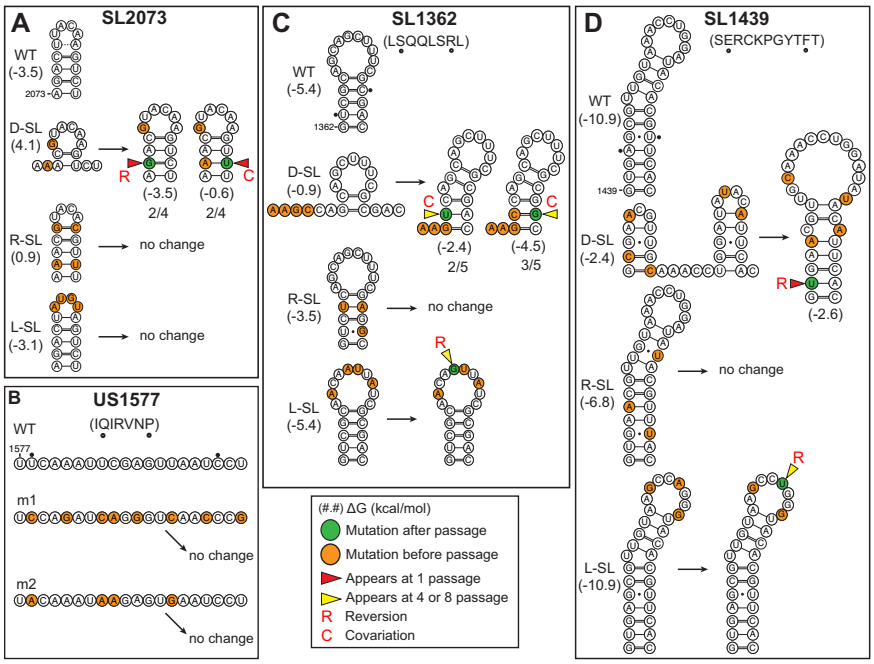

Figure 4. *In planta* selection of the SL1362, SL1439, SL2073, and US1577 mutants. (**A**) Four versions of SL2073 are depicted: wild-type (wt), disrupted (D-SL), restored (R-SL) and loop (L-SL). Dashed line in SL2073 indicates difference between consensus and SHAPE restrained predictions. (**B**) Two versions of US1577 mutants containing synonymous nucleotide changes are shown: m1 and m2. Three mutants (D-SL, R-SL, L-SL) and the wild-type (wt) structure are shown for SL1362 (**C**) and SL1439 (**D**). In all mutants, black arrows point to the predicted structures of the sequences observed after passaging. Orange circles indicate engineered changes; green circles are positions mutated during *in planta* passaging. Red wedges show changes that occurred at 1° passage. Where applicable, letters in parentheses show the amino acid sequence encoded with black dots serving as positional references for the first nucleotide of the codon. Structures are based on predictions made by MFold (64) using the sequences shown with the calculated minimum free energy (kcal/mol) in parentheses. A minimum of three biological replicates were passaged for all constructs; fractions below paired structures indicate number of experiments yielding depicted structure over total number of experiments carried out.

ture has been shown to be particularly important in plant viruses with icosahedral particles, like CMV, and is predicted to be more extensive compared with most animal viruses (59). Studies on higher-order RNA structure and its potential function, however, have only focused on a few structures identified within a few viruses. At the moment, large-scale RNA structural maps exist for three ssRNA (+)-sense viral genomes: hepatitis C virus (34), human immunodeficiency virus (35), and tomato bushy stunt virus (36). In addition, the structure of satellite tobacco mosaic virus has also recently been reported (37,38). The identification of SEs within ORFs has also recently been explored in HCV using SHAPE reactivity data and covariation analyses. Mutated variants of those SEs were shown to vary drastically in their effect on replication and infectivity in cell cultures, providing evidence that there are multiple SEs across the HCV genome critical to the infection cycle (34,39). For plants, SHAPE analysis has been used to predict SEs within ORFs of TBSV, and structurally destabilized mutants of these SEs reduced accumulation of TBSV in protoplast competition assays (36).

The rationale for analyzing both viRNA and lyRNA was that the latter source of RNA would more closely represent *in vivo* conditions. A modified version of the lysate extraction method used here was successfully used for human cell-lines in endogenous RNA pull-downs (42). Its use here offered the most practically feasible approach to studying the RNA structure of the viral population *in planta*. In our SHAPE results, the viRNA and lyRNA reactivity data shared similar reactivity maps, although a number of differences existed that in at least one instance could be explained by the putative binding of the viral replicase to the loop and bulge of SLC (10,23,60). It contains an analog to the clamped adenine motif (CAM) identified in BMV at position nt 2154, which forms part of the triloop (AUA in BMV; AAU in CMV). This nucleotide is highly reactive in the viRNA data, but significantly reduced in the lyRNA data, along with other bases predicted to occur in the triloop and bulge.

Based on previous work where we demonstrated the capacity of the CMV genome to rapidly adapt to novel environments and hosts (61), we used the adaptability of CMV to select for the fittest genotype derivative *in planta* from engineered mutants (Figure 1B). We hypothesized that if there was any selective advantage to the SEs predicted then any changes in their structural integrity would be selected against and would be quickly observed in passaged virus populations given the high mutation rate of RNA viruses. We first tested mutants in SL2073 where we suspected a high selection pressure, then we examined a predicted unstructured region in the CP ORF (US1577) that we did not suspect to be under any structural selection pressure. A structurally disrupted mutant of SL2073 showed rapid reversion to form a new structure close to wild-type, but other mutants with altered sequences in the stem (R-SL2073) or loop (L-SL2073) that maintained the original stem–loop struc-

ture remained stable (Figure 4A). In contrast to these results is the apparent stability of the predicted unstructured region US1577 for the two mutants passaged in this study (Figure 4B). Parallel to these passages we also monitored wild-type Bn57 infection and found no nucleotide changes occurred in the mutant or wild-type US1577, a result that appears to rule out any stochastic effects due to random genetic drift.

We then examined two of four newly identified potential SEs in the CP ORF (Figure 2B): SL1362 and SL1439. Using the approach as described above (Figure 1), we demonstrated that in SL1362 and SL1439 nucleotide changes occur in structurally disrupted mutants to return its predicted structure to some semblance of the original wild-type, but engineered changes in the stem that are predicted to maintain stem integrity are stable *in planta*. Like SL2073, neither mutant (D-SL1362 and D-SL1439) was restored to structures identical to the wild-type, but in D-SL1362 the changes that occurred were compensatory and were not reversions to the wild-type sequence. Additionally, the D-SL1362 changes were synonymous, and although they appear to be weakly selected for (appearing at the 4° or 8° passage), they support the progressive selection for the presence of a SE. In D-SL1439, the C1440U reversion to wild-type occurred immediately in the first passage to partially restore the wild-type stem.

Unlike the non-coding SL2073, the CP ORF SEs showed a consistent reversion of a centrally positioned loop nucleotide to the wild-type, both of which were at the third position of a codon triplet. The reason for this is unclear, as both nucleotides are less conserved than others in their respective loops (Figure 2B). One possible reason for the observed changes *in planta* might be a negative selection against less commonly codon triplets in these loops (62,63), however, some of the least common CMV codons were introduced into the US1577 mutants with no clear negative effects (Supplementary Table S5). It is also possible that these loops are involved in long-range interactions, as was predicted for SL1362 when restrained with the lyRNA reactivity data (Figure 3A). Another curiosity of the *in planta* L-SL1362 derivatives was the partial alteration of a distal nucleotide (C1487U) that causes a nonsynonymous change, P77L, in the CP. The reason for the distal mutation is not obvious, potentially pointing to selection pressures operating at a non-codon level. Interestingly, this mutation falls within the pseudoknot predicted in the viRNA-derived structure (Figure 3C) and would not introduce a mismatch. The change P77L was also reported as a spontaneous compensatory mutation in an engineered movement deficient Fny-CMV mutant that restored systemic infection (57). The biological significance of this mutation remains unclear.

In conclusion, we have used a novel combination of methods to explore the RNA secondary structure of CMV RNA3 to find evidence for the existence of previously undescribed SEs. While the function of these SEs is currently unknown, they may serve important roles protecting the viral RNA from the host RNAi machinery, interacting with viral and host proteins, or taking part in long-range interactions (1). Identification of these SEs, however, is an important first step in understanding the importance of the RNA structure-function relationship in the life cycle of an RNA virus. The work presented here provides a framework for identifying new SEs in RNA plant virus genomes and investigating the role of RNA structure in viral infections.

AVAILABILITY

Reactivity data was deposited in the RNA Mapping Database (https://rmdb.stanford.edu/) (46) with accession numbers: CMVBN3_1M7_0001 (lyRNA), CMVBN3_1M7_0002 (viRNA). A scripts for calculating the reactivities via the stratified priming approach can be found at either https://github.com/LucksLab/LucksLab_Publications/tree/master/Watters_CMV_RNA_Structures or as part of SEQualyzer (57).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Heather McLane and Allison Schloop for their technical expertise and Nick van Eck and crew for growth-room facilities support. We also thank Peter Palukaitis for critically reviewing the manuscript, and Ezgi Hacisuleyman for advice on the lysate preparations. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

FUNDING

National Institutes of Health (NIH) [R00HG006860 to S.A.]; New Innovator Award through the National Institute of General Medical Sciences of the National Institutes of Health [1DP2GM110838 to J.B.L.]; Gates Foundation Grand Challenges [OPP 1068482 to J.R.T.]; K.E.W. was a Fleming Scholar in the Robert F. Smith School of Chemical and Biomolecular Engineering at Cornell University during part of this work. Funding for open access charge: Gates Foundation.
Conflict of interest statement. None declared.

REFERENCES

1. Nicholson,B.L. and White,K.A. (2014) Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.*, **12**, 493–504.
2. Nicholson,B.L. and White,K.A. (2015) Exploring the architecture of viral RNA genomes. *Curr. Opin. Virol.*, **12**, 66–74.
3. Pleij,C.W.A., Rietveld,K. and Bosch,L. (1985) A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res.*, **13**, 1717–1731.
4. Rietveld,K., Vanpoelgeest,R., Pleij,C.W.A., Vanboom,J.H. and Bosch,L. (1982) The transfer RNA-like structure at the 3′ terminus of turnip yellow mosaic-virus RNA - differences and similarities with canonical transfer-RNA. *Nucleic Acids Res.*, **10**, 1929–1946.
5. Ahlquist,P., Dasgupta,R. and Kaesberg,P. (1981) Near identity of 3′RNA secondary structure in bromoviruses and cucumber mosaic virus. *Cell*, **23**, 183–189.
6. Felden,B., Florentz,C., Giege,R. and Westhof,E. (1994) Solution structure of the 3′-end of brome mosaic virus genomic RNAs. Conformational mimicry with canonical tRNAs. *J. Mol. Biol.*, **235**, 508–531.

7. Rietveld,K., Linschooten,K., Pleij,C.W. and Bosch,L. (1984) The three-dimensional folding of the tRNA-like structure of tobacco mosaic virus RNA. A new building principle applied twice. *EMBO J.*, **3**, 2613–2619.
8. Felden,B., Florentz,C., Giege,R. and Westhof,E. (1996) A central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5′ upstream pseudoknots in the 3′ terminus of tobacco mosaic virus RNA. *RNA*, **2**, 201–212.
9. Takamatsu,N., Watanabe,Y., Meshi,T. and Okada,Y. (1990) Mutational analysis of the pseudoknot region in the 3′ noncoding region of tobacco mosaic virus RNA. *J. Virol.*, **64**, 3686–3693.
10. Sivakumaran,K., Bao,Y., Roossinck,M.J. and Kao,C.C. (2000) Recognition of the core RNA promoter for minus-strand RNA synthesis by the replicases of Brome mosaic virus and Cucumber mosaic virus. *J. Virol.*, **74**, 10323–10331.
11. Osman,T.A., Hemenway,C.L. and Buck,K.W. (2000) Role of the 3′ tRNA-like structure in tobacco mosaic virus minus-strand RNA synthesis by the viral RNA-dependent RNA polymerase In vitro. *J. Virol.*, **74**, 11671–11680.
12. Edwardson,J.R. and Christie,R.G. (1991) Cucumoviruses. *CRC Handbook of Viruses Infecting Legumes*, CRC Press, Boca Raton, pp. 293–319.
13. Rybicki,E.P. (2015) A Top Ten list for economically important plant viruses. *Arch. Virol.*, **160**, 17–20.
14. Scholthof,K.B.G., Adkins,S., Czosnek,H., Palukaitis,P., Jacquot,E., Hohn,T., Hohn,B., Saunders,K., Candresse,T., Ahlquist,P. *et al.* (2011) Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.*, **12**, 938–954.
15. Chen,M.H., Roossinck,M.J. and Kao,C.C. (2000) Efficient and specific initiation of subgenomic RNA synthesis by cucumber mosaic virus replicase in vitro requires an upstream RNA stem–loop. *J. Virol.*, **74**, 11201–11209.
16. Jacquemond,M. (2012) Cucumber mosaic virus. *Adv. Virus Res.*, **84**, 439–504.
17. Palukaitis,P., Roossinck,M.J., Dietzgen,R.G. and Francki,R.I. (1992) Cucumber mosaic virus. *Adv. Virus Res.*, **41**, 281–348.
18. Palukaitis,P. and Garcia-Arenal,F. (2003) Cucumoviruses. *Adv. Virus Res.*, **62**, 241–323.
19. Joshi,R.L., Joshi,S., Chapeville,F. and Haenni,A.L. (1983) tRNA-like structures of plant viral RNAs: conformational requirements for adenylation and aminoacylation. *EMBO J.*, **2**, 1123–1127.
20. Chapman,M.R., Rao,A.L.N. and Kao,C.C. (1998) Sequences 5′ of the conserved tRNA-like promoter modulate the initiation of minus-strand synthesis by the brome mosaic virus RNA-dependent RNA polymerase. *Virology*, **252**, 458–467.
21. Miller,W.A., Bujarski,J.J., Dreher,T.W. and Hall,T.C. (1986) Minus-strand initiation by brome mosaic virus replicase within the 3′ transfer RNA-like structure of native and modified RNA templates. *J. Mol. Biol.*, **187**, 537–546.
22. Dreher,T.W., Bujarski,J.J. and Hall,T.C. (1984) Mutant viral RNAs synthesized in vitro show altered aminoacylation and replicase template activities. *Nature*, **311**, 171–175.
23. Chapman,M.R. and Kao,C.C. (1999) A minimal RNA promoter for minus-strand RNA synthesis by the brome mosaic virus polymerase complex. *J. Mol. Biol.*, **286**, 709–720.
24. Thompson,J.R., Buratti,E., de Wispelaere,M. and Tepfer,M. (2008) Structural and functional characterization of the 5′ region of subgenomic RNA5 of cucumber mosaic virus. *J. Gen. Virol.*, **89**, 1729–1738.
25. Thompson,J.R. and Tepfer,M. (2009) The 3′ untranslated region of cucumber mosaic virus (CMV) subgroup II RNA3 arose by interspecific recombination between CMV and tomato aspermy virus. *J. Gen. Virol.*, **90**, 2293–2298.
26. Morroni,M., Thompson,J.R. and Tepfer,M. (2009) Analysis of recombination between viral RNAs and transgene mRNA under conditions of high selection pressure in favour of recombinants. *J. Gen. Virol.*, **90**, 2798–2807.
27. Baumstark,T. and Ahlquist,P. (2001) The brome mosaic virus RNA3 intergenic replication enhancer folds to mimic a tRNA TpsiC-stem loop and is modified in vivo. *RNA*, **7**, 1652–1670.
28. Noueiry,Chen.J. and Ahlquist,P. (2001) Brome mosaic virus Protein 1a recruits viral RNA2 to RNA replication through a 5′ proximal RNA2 signal. *J. Virol.*, **75**, 3207–3219.
29. Yi,G. and Kao,C. (2008) cis- and trans-acting functions of brome mosaic virus protein 1a in genomic RNA1 replication. *J. Virol.*, **82**, 3045–3053.
30. Sullivan,M.L. and Ahlquist,P. (1999) A brome mosaic virus intergenic RNA3 replication signal functions with viral replication protein 1a to dramatically stabilize RNA in vivo. *J. Virol.*, **73**, 2622–2632.
31. Lu,Z.P. and Chang,H.Y. (2016) Decoding the RNA structurome. *Curr. Opin. Struc. Biol.*, **36**, 142–148.
32. Strobel,E.J., Watters,K.E., Loughrey,D. and Lucks,J.B. (2016). RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Curr. Opin. Biotechnol*. **39**, 182–191.
33. Choudhary,K., Deng,F. and Aviran,S. (2017) Comparative and integrative analysis of RNA structural profiling data: current practices and emerging questions. *Quant. Biol.*, **5**, 3–24.
34. Mauger,D.M., Golden,M., Yamane,D., Williford,S., Lemon,S.M., Martin,D.P. and Weeks,K.M. (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3692–3697.
35. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–U787.
36. Wu,B.D., Grigull,J., Ore,M.O., Morin,S. and White,K.A. (2013) Global organization of a positive-strand RNA virus genome. *PLoS Pathog.*, **9**, e1003363.
37. Athavale,S.S., Gossett,J.J., Bowman,J.C., Hud,N.V., Williams,L.D. and Harvey,S.C. (2013) In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS One*, **8**, e54384.
38. Larman,B.C., Dethoff,E.A. and Weeks,K.M. (2017) Packaged and free STMV RNA genomes adopt distinct conformational states. *Biochemistry*, **56**, 2175–2183.
39. Pirakitikulr,N., Kohlway,A., Lindenbach,B.D. and Pyle,A.M. (2016) The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol. Cell*, **62**, 111–120.
40. Watters,K.E., Yu,A.M., Strobel,E.J., Settle,A.H. and Lucks,J.B. (2016) Characterizing RNA structures in vitro and in vivo with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods*, **103**, 34–48.
41. Thompson,J.R., Langenhan,J.L., Fuchs,M. and Perry,K.L. (2015) Genotyping of cucumber mosaic virus isolates in western New York State during epidemic years: characterization of an emergent plant virus population. *Virus Res.*, **210**, 169–177.
42. Hacisuleyman,E., Goff,L.A., Trapnell,C., Williams,A., Henao-Mejia,J., Sun,L., McClanahan,P., Hendrickson,D.G., Sauvageau,M., Kelley,D.R. *et al.* (2014) Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.*, **21**, 198–206.
43. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**, 10–12.
44. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10**, R25.
45. Ge,P. and Zhang,S.J. (2015) Computational analysis of RNA structures with chemical probing data. *Methods*, **79–80**, 60–66.
46. Yesselman,J. D., Tian,S., Liu,X., Shi,L., Li,J.B. and Das,R. (2017) Updates to the RNA mapping database (RMDB), version 2. *Nucleic Acids Res.*, doi:10.1093/nar/gkx873.
47. Wilm,A., Higgins,D.G. and Notredame,C. (2008) R-Coffee: a method for multiple alignment of non-coding RNA. *Nucleic Acids Res.*, **36**, e52.
48. Sukosd,Z., Knudsen,B., Kjems,J. and Pedersen,C.N. (2012) PPfold 3.0: fast RNA secondary structure prediction using phylogeny and auxiliary data. *Bioinformatics*, **28**, 2691–2692.
49. McGinnis,J.L., Dunkle,J.A., Cate,J.H.D. and Weeks,K.M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **134**, 6617–6624.
50. Loughrey,D., Watters,K.E., Settle,A.H. and Lucks,J.B. (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.*, **42**, e165.
51. Lucks,J.B., Mortimer,S.A., Trapnell,C., Luo,S.J., Aviran,S., Schroth,G.P., Pachter,L., Doudna,J.A. and Arkin,A.P. (2011) Multiplexed RNA structure characterization with selective

2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11063–11068.
52. Watters,K.E. and Lucks,J.B. (2016) Mapping RNA structure in vitro with SHAPE chemistry and next-generation sequencing (SHAPE-Seq). (Eds)Turner,D.H. and Mathews,D.H., In:*Methods in Molecular Biolog - RNA Structure Determination: Methods and Protocols*, Vol. **1490**, pp. 135–162.
53. Aviran,S. and Pachter,L. (2014) Rational experiment design for sequencing-based RNA structure mapping. *RNA*, **20**, 1864–1877.
54. Aviran,S., Trapnell,C., Lucks,J.B., Mortimer,S.A., Luo,S.J., Schroth,G.P., Doudna,J.A., Arkin,A.P. and Pachter,L. (2011) Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11069–11074.
55. Aviran,S., Lucks,J.B. and Pachter,L. (2011) RNA structure characterization from chemical mapping experiments. In: *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing*. Monticello, pp. 1743–1750.
56. Choudhary,K., Shih,N.P., Deng,F., Ledda,M., Li,B. and Aviran,S. (2016) Metrics for rapid quality control in RNA structure probing experiments. *Bioinformatics*, **32**, 3575–3583.
57. Choudhary,K., Ruan,L.Y., Deng,F., Shih,N. and Aviran,S. (2017) SEQualyzer: interactive tool for quality control and exploratory analysis of high-throughput RNA structural profiling data. *Bioinformatics*, **33**, 441–443.
58. Hajdin,C.E., Bellaousov,S., Huggins,W., Leonard,C.W., Mathews,D.H. and Weeks,K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5498–5503.
59. Simmonds,P., Tuplin,A. and Evans,D.J. (2004) Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. *RNA*, **10**, 1337–1351.
60. Kim,C.H., Kao,C.C. and Tinoco,I. (2000) RNA motifs that determine specificity between a viral replicase and its promoter. *Nat. Struct. Biol.*, **7**, 415–423.
61. Thompson,J.R., Doun,S. and Perry,K.L. (2006) Compensatory capsid protein mutations in cucumber mosaic virus confer systemic infectivity in squash (Cucurbita pepo). *J. Virol.*, **80**, 7740–7743.
62. Adams,M.J. and Antoniw,J.F. (2004) Codon usage bias amongst plant viruses. *Arch. Virol.*, **149**, 113–135.
63. Lauring,A.S., Acevedo,A., Cooper,S.B. and Andino,R. (2012) Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe*, **12**, 623–632.
64. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.