Biochemistry

Article

Cite This: Biochemistry 2019, 58, 3069−3077

pubs.acs.org/biochemistry

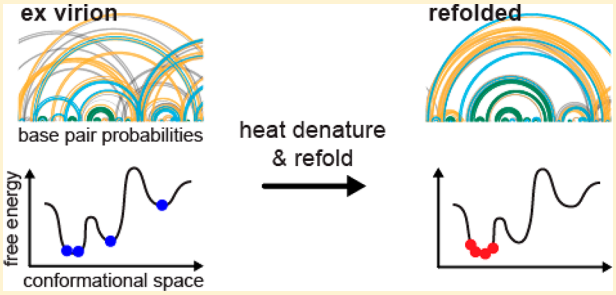
Effects of Refolding on Large-Scale RNA Structure

Elizabeth A. Dethoff and Kevin M. Weeks*

Department of Chemistry, The University of North Carolina, Chapel Hill, North Carolina 27599-3290, United States

Supporting Information

ABSTRACT: Understanding RNA-mediated functions requires a detailed characterization of the underlying RNA structure. In many cases, structure probing experiments are performed on RNA that has been "refolded" in some way, which may cause the conformation to differ from that of the native RNA. We used SHAPE−MaP (selective 2′-hydroxyl acylation analyzed by primer extension, read out by mutational profiling), to probe the structure of the Dengue virus (DENV) RNA genome after gentle extraction of the native RNA from intact virions (ex virion) and after heat denaturation and refolding. Comparison of mutiple SHAPE-informed structural features revealed that refolded RNA is more highly structured and samples fewer conformations than the ex virion RNA. Regions with similar structural features are generally those with low SHAPE reactivity and low Shannon entropy (lowSS regions), which correspond to elements with high levels of well-determined structure. This high-structure and low-entropy analysis framework, previously shown to make possible discovery of functional RNA structures, is thus now shown to allow de novo identification of structural elements in a refolded RNA that are likely to recapitulate RNA structures in the ex virion RNA state. Regions with less well-defined structures, which occurred more frequently in the more native-like ex virion RNA and may contain RNA switches, are challenging to recapitulate using refolded RNA.



The ability of RNA to fold back on itself to form three-dimensional structures is crucial for its ability to regulate transcription, translation, gene silencing, protein trafficking, and RNA splicing.[1,2] To understand the functions of RNA, the underlying RNA secondary and tertiary structures must be well understood. Ideally, RNA structure should be probed at comprehensive nucleotide resolution in healthy cells;[3−7] however, experimental requirements often necessitate studying RNA that has been purified under denaturing conditions or transcribed in vitro. These samples are often heat denatured and refolded to promote the formation of secondary and tertiary structure.

Evidence suggests that RNA structure formation begins during transcription.[8,9] Co-transcriptional folding processes can be affected by many factors, including the rate of transcription and the cellular environment. These factors differ greatly between in-cell and simplified in vitro environments. In-cell RNA folding can also be influenced by RNA chaperones that facilitate folding into the native structure.[10] Finally, full-length cellular RNAs may sample various distinct, functionally important conformations as they encounter specific ligands and proteins.[2,11] Every in-cell RNA molecule potentially has a unique folding history, and vestiges of these former structures may be evident in the native RNA structure. Thus, RNA structures in cells and in virus particles could differ substantially from that of refolded RNAs. A number of studies have examined the effects of refolding on small RNAs or small regions of large RNAs,[4,12−14] but the effects of heat denaturation and refolding on large RNAs, with complex and heterogeneous structures, are poorly understood.

We investigated the effects of heat denaturation and refolding on the structure of a Dengue virus serotype 2 (DENV2) RNA genome. We measured the local nucleotide flexibility for each position in the 10723-nucleotide genome using selective 2′-hydroxyl acylation analyzed by primer extension, read out by mutational profiling (SHAPE-MaP).[15−17] These data were used to model the secondary structure of protein-free DENV2 RNA under two conditions: gently extracted from virions (termed ex virion) and heat-denatured and refolded RNA (termed refolded). We identified 25 structural elements in the refolded RNA with well-determined secondary structures as defined by the experimentally determined level of RNA structure and conformational entropy. The structures of these specific elements agree well with the ex virion RNA secondary structure. We also identified extensive differences between the refolded and ex virion RNAs. We found that refolded DENV2 RNA generally has a conformation that is more well-defined than that of the more native-like ex virion RNA. Ultimately, many regions of the refolded RNA sample a single structure, whereas these same regions in the more native-like ex virion RNA sample multiple conformations.

Received: November 24, 2018
Revised: June 19, 2019
Published: June 24, 2019

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

METHODS

SHAPE-MaP RNA Structure Probing. Details about DENV2 RNA sample preparation and data processing have been described previously.[17] Briefly, RNA was gently extracted from purified DENV2 viral particles (strain S16803, GenBank GU289914; Microbix Biosystems) under nondenaturing conditions. For the ex virion RNA experiment, extracted RNA was exchanged into 1.11× SHAPE-MaP buffer [50 mM HEPES (pH 8.0), 200 mM potassium acetate (pH 8.0), and 3 mM $MgCl_2$, at a 1× concentration]. For the refolded RNA experiments, extracted RNA was exchanged into 0.5× TE [5 mM Tris (pH 7.5) and 0.5 mM EDTA] (RNeasy MinElute, Qiagen), heated at 85 °C for 5 min, and cooled on ice for 5 min. SHAPE-MaP buffer was added to a final concentration of 1.11× buffer. Both ex virion and refolded RNA samples were incubated under identical final conditions at 37 °C for 30 min. RNA was treated with 0.1 volume of 1-methyl-7-nitroisatoic anhydride (1M7), 1-methyl-6-nitroisatoic anhydride (1M6), or *N*-methyl-isatoic anhydride (NMIA). No-reagent controls were prepared in parallel by addition of neat dimethyl sulfoxide. A single set of denaturing controls were prepared using 1M7, 1M6, and NMIA. DNA libraries were prepared for massively parallel sequencing as described previously.[15] Ex virion and refolded libraries were sequenced on HiSeq 2500 and MiSeq instruments (Illumina). Sequencing data were analyzed using ShapeMapper software,[15,16] which has been extensively validated relative to reference structures.[7,15,18] Raw sequencing FASTQ files were aligned to a reference DENV2 sequence, and SHAPE reactivities were generated using the following parameters: minPhred = 10, minLength = 25, mapping quality = 30, and minPhredToCount = 20. SHAPE reactivities for nucleotides with a read depth of <2500 or with a high no-reagent mutation rate (>0.03) were excluded from analysis. We obtained 1M7 SHAPE reactivities for 99% of the nucleotides in the DENV2 genome in the ex virion and refolded states. Full biological replicate experiments for both ex virion and refolded RNAs yielded Pearson correlation coefficients of ≥0.9.

Kolmogrov–Smirnov Test. A Kolmogorov–Smirnov (KS) permutation test was used to evaluate whether the distributions of ex virion and refolded 1M7 SHAPE reactivities were significantly different. The null hypothesis of the KS test is that two data sets are drawn from the same distribution. First, 10% of the reactivities (∼1072) of each data set were randomly sampled. A KS test was performed on these two sample sets, and the p value was calculated. This process was repeated for a total of 1000 iterations. Control KS permutation tests were performed by comparing ex virion sample sets (or refolded sample sets) to each other. To account for multiple testing, a Bonferroni-corrected significance threshold value was used, α/n, where α equals the significance threshold (0.01) and n equals the number of hypotheses tested (1000) to give a corrected significance threshold of 0.00001. In the ex virion and refolded KS tests, all p values were <0.00001 (Figure S1). Thus, the ex virion and refolded 1M7 SHAPE reactivity distributions are significantly different. As expected, the p values of the control KS tests were all greater than the Bonferroni-corrected α value. The same KS permutation test procedure was also used to compare two refolded RNA 1M7 SHAPE reactivity replicates, collected months apart in full biologically distinct experiments (Figure S1). These two replicates had a Pearson correlation coefficient (R) of 0.88. Of the 1000 KS tests, 999 failed to reject the null hypothesis, indicating that the two replicates of refolded RNA SHAPE reactivities were drawn from the same distribution. Thus, the differences observed between ex virion and refolded SHAPE reactivities were not due to variability in the SHAPE-MaP approach but were a result of the refolding treatment.

Secondary Structure Modeling of Refolded RNA. The SHAPE-derived secondary structure model for the ex virion RNA was reported previously,[17] and the same methods were used here to characterize the secondary structure of refolded RNA. Briefly, pseudoknots were predicted using 1M7 SHAPE reactivities as constraints in the program ShapeKnots,[19] implemented in RNAStructure version 5.6.[20,21] The maximum base-pairing distance was set to 500 nucleotides; thus, in our analysis, we could examine nucleotide interactions that are within this pairing distance. All of the pseudoknots previously predicted in the ex virion RNA were also predicted to occur in the refolded RNA. Two additional pseudoknots were predicted in the refolded RNA: a small pseudoknot with the 5′-most nucleotide at position 3863 and a pseudoknot in the 3′-dumbbell structure in the 3′-untranslated region (3′-UTR). Pseudoknotted nucleotides were forced to be single-stranded in subsequent structure calculations.

Both 1M7 and differential SHAPE reactivities were input as constraints using *SuperFold*,[15,16] which uses RNAStructure *Partition* (version 5.6)[20,21] to generate base-pairing probabilities for all possible canonical base pairs. Base-pair probabilities were used to calculate a Shannon entropy value for every nucleotide.[22] We used *SuperFold*, which calls RNAStructure *Fold* (version 5.6),[20,21] to generate a minimum free energy structure.[15,16] Pseudoknotted base pairs were added to the structure. The sensitivity and positive predictive value (ppv) of the refolded model compared to the previously reported ex virion RNA model[17] were calculated, treating the ex virion structure as the reference state. These values are termed refolded-sensitivity (RF-sens) and refolded-ppv (RF-ppv) throughout the text. For whole-genome calculations, RF-sens is the percentage of base pairs in the ex virion structural model that are present in the refolded RNA model. RF-ppv is the percentage of base pairs in the refolded structural model that are also present in the ex virion model.

Identification of Well-Structured Elements. Structural elements with both 1M7 SHAPE reactivities and Shannon entropies below their respective medians were identified. The minimum length of selected regions was required to be at least 40 nucleotides. Some regions were combined or expanded to include all nucleotides of bisecting helices.[15,16] An exception was made for element 6. This element was not expanded to include the intersecting pseudoknotted helix, because the pseudoknot contains numerous low-probability, long-range base pairs that would have required expanding the element by >750 nucleotides. This single pseudoknot helix was ignored in the RF-sens and RF-ppv calculations of individual structural elements (discussed below).

The RF-ppv values for individual structural elements were calculated as described above for the entire genome, again treating the ex virion structure model as the reference state. The RF-sens was calculated differently, because the region of interest in the ex virion model often contained nucleotides for only one-half of a helix, with base-pairing partners located outside of the region of interest. Ignoring these base pairs would have artificially increased the calculated RF-sens values. Instead, for the individual element RF-sens calculations, we defined RF-sens as the percentage of base-paired nucleotides in the ex virion

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069–3077

model that have the same base-pairing pattern in the refolded RNA model.

Comparison of SHAPE-Directed Models to an Accepted DENV2 Structure. An accepted DENV2 secondary structure was created by unifying structural elements proposed in homology-based studies and as supported by subsequent structural studies of DENV2 RNA fragments, mini-genomes, and full-length genomes. Detailed descriptions of how the accepted structure was defined and comparisons of the SHAPE-predicted models to the accepted DENV2 structure are provided in the Supporting Information. Sensitivity (sens) and ppv values for the ex virion model were calculated relative to those elements for which well-determined accepted models exist, which together span 7% of the Dengue genome; these are elements 1 (nucleotides 1−105), 2 (nucleotides 148−308), and 24 (nucleotides 10247−10723). The sens value was defined as the percentage of accepted base pairs accurately predicted by the ex virion RNA model, and ppv was defined as the percentage of base pairs in the ex virion structural model that are also present in the accepted model.

RESULTS

Effects of Refolding Treatment on SHAPE Reactivities and Shannon Entropies. To understand the effects of heat denaturing and refolding on large-scale RNA structure, we obtained nucleotide-resolution SHAPE-MaP structural data for an entire DENV2 genome (strain S16803[23]) under two folding conditions. Under the first condition (ex virion), DENV2 RNA from intact virions was extracted under gentle, nondenaturing conditions (Figure 1). Under the second condition (refolded), isolated DENV2 RNA was denatured by heating to 85 °C in the absence of monovalent and divalent ions and cooling on ice. Both RNAs were allowed to fold at 37 °C in 50 mM HEPES (pH 8.0), 200 mM potassium acetate (pH 8.0), and 3 mM $MgCl_2$ before modification with 1M7[24] or the differential SHAPE reagents 1M6 and NMIA.[25,26] DNA libraries were prepared for massively parallel sequencing, and SHAPE reactivity profiles were generated.

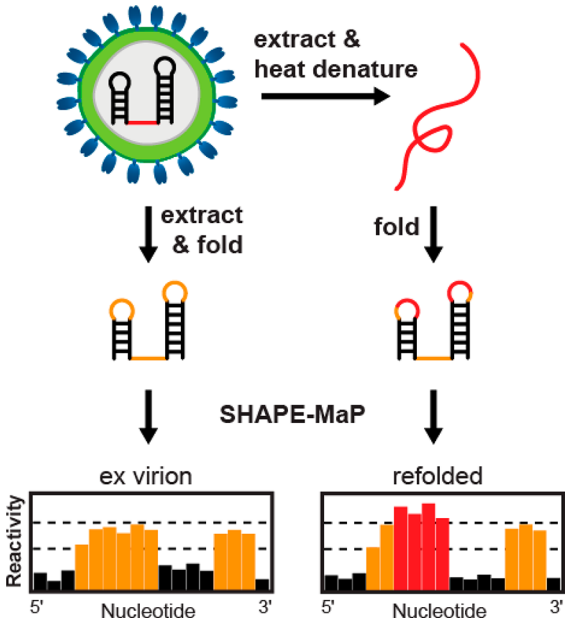


Figure 1. Native-like ex virion and refolded DENV2 genomic RNAs. The ex virion RNA sample was prepared by gentle phenol/chloroform extraction from purified DENV2 viral particles using an aqueous phase pre-equilibrated in folding buffer. Refolded RNA was heated briefly to 85 °C and snap-cooled on ice. Both ex virion and refolded RNA samples were incubated at 37 °C for 30 min in 200 mM potassium acetate and 3 mM $MgCl_2$ at pH 8.0 before SHAPE structure probing.

Generally, nucleotides with high SHAPE reactivities (>0.85) are conformationally flexible and likely to be single-stranded. Conversely, nucleotides with low SHAPE reactivities (<0.4) are structurally constrained and likely base paired.[19,27] A nucleotide with an intermediate reactivity may be only partially structured (for example, base-stacked but not base-paired) or may sample multiple conformations or participate in a structural switch. In the latter cases, two or more stable secondary structures may exist in equilibrium, and the measured SHAPE reactivity is the average flexibility of a nucleotide across all conformations. The median 1M7 SHAPE reactivity profiles of refolded and ex virion RNA are very similar across the genome (Figure 2A). This similarity indicates that DENV2 RNA likely adopts broadly similar architectures in the ex virion and refolded states.

Differential and 1M7 SHAPE reactivities were used in a partition function calculation to determine base-pairing probabilities for each base pair in the predicted ensemble of structures for both RNA states. The probabilities were summed to calculate a Shannon entropy for each nucleotide.[22] We emphasize that even though SHAPE data are an input into the Shannon entropy calculation, Shannon entropy provides additional information beyond that provided by raw SHAPE reactivities. SHAPE reactivity is a measure of the degree of RNA structure. Shannon entropy is a measure of our certainty or confidence in a predicted secondary structure. A region of RNA with low SHAPE reactivity indicates that the region is highly structured but does not report on our ability to model its structure using available structure probing tools. If multiple, different secondary structures are consistent with the SHAPE reactivity pattern, then this region will have a high Shannon entropy even though it might be highly structured. As expected, because these two metrics report on distinct features of RNA structure, they are not strongly correlated with each other. Pearson's R values between Shannon entropy and SHAPE reactivity for the ex virion and refolded RNAs are −0.25 and −0.19, respectively (Figure S2). The Shannon entropy profiles for refolded and ex virion RNAs are similar across the genome (Figure 2B), indicating that, overall, the two RNAs are structurally similar.

The Structure of Refolded RNA Is More Well-Determined Than That of Ex Virion RNA. Despite the broad similarities between the median SHAPE reactivities and Shannon entropies across the ex virion and refolded RNAs, there are nonetheless clear differences between the two RNA states. We first examined the frequency distribution of SHAPE reactivities across the two states, which is correlated with the overall level of structure of an RNA. A large number of low and high reactivities suggest an RNA samples a single, well-defined structure. Conversely, an RNA with many intermediate reactivities is less structured or samples multiple conformations.[26] The distributions of 1M7 SHAPE reactivities for refolded and ex virion DENV2 RNA are very different (Figure 2C). A greater number of nucleotides with low and high SHAPE reactivities are observed in refolded RNA, whereas a greater proportion of nucleotides with intermediate reactivities are detected in the ex virion RNA. These differences were significant at the $p < 0.01$ level [Kolmogorov−Smirnov permutation test (Figure S1A)]. A similar analysis of two replicate refolded RNA

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

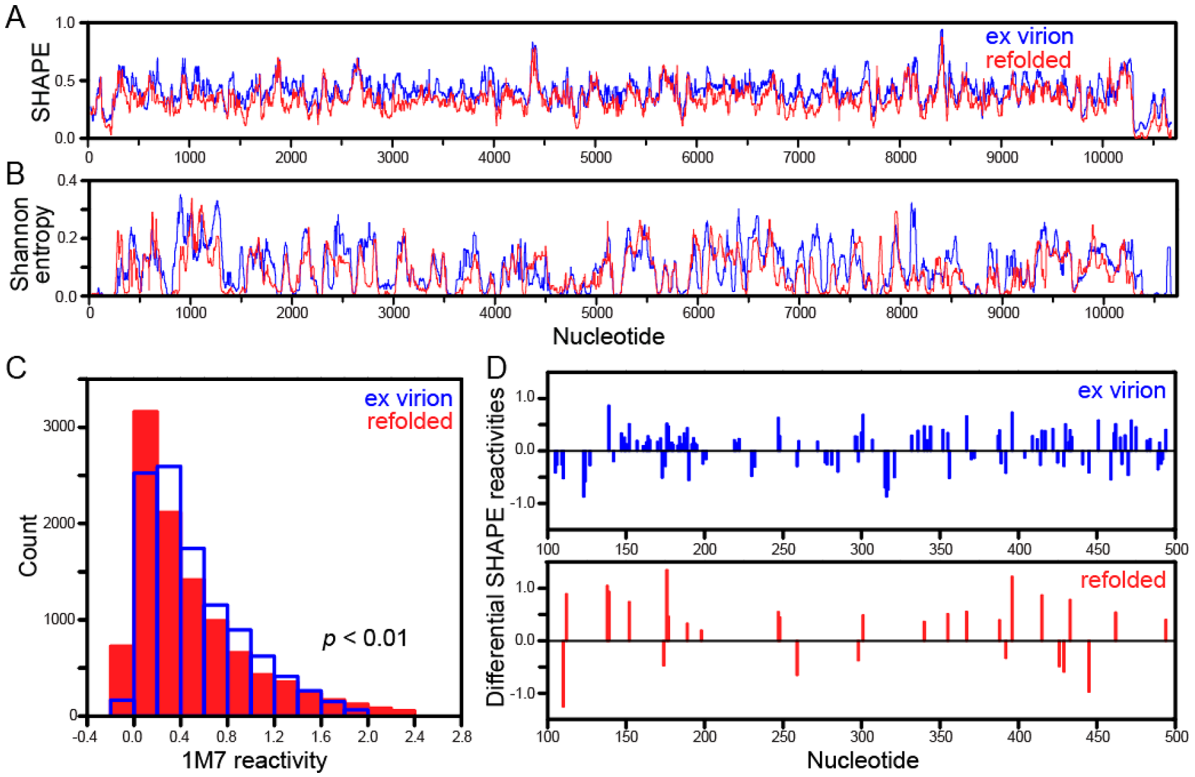


Figure 2. SHAPE reactivities and Shannon entropies for ex virion and refolded RNAs. (A) Median ex virion (blue) and refolded (red) RNA 1M7 SHAPE reactivities for the 10723-nucleotide DENV2 RNA genome. Reactivities are shown as centered 51-nucleotide windows. (B) Median ex virion (blue) and refolded (red) RNA Shannon entropies, calculated using differential SHAPE reactivities, are shown as centered 51-nucleotide windows. (C) Histogram of ex virion (blue) and refolded (red) RNA 1M7 SHAPE reactivities. (D) Representative differential SHAPE reactivities[25] for ex virion (blue) and refolded (red) RNAs.

data sets showed that the two replicates were statistically the same (Figure S1B). Thus, the differences between ex virion and refolded RNA SHAPE reactivities suggest that ex virion RNA is less structured or samples multiple conformations relative to the refolded RNA. In addition, the Shannon entropies of nucleotides in the refolded RNA were almost always lower than those of the ex virion RNA; the median Shannon entropies of the refolded and ex virion RNA are 0.062 and 0.085, respectively (Figure 2B). The smaller Shannon entropy of the refolded RNA suggests that the secondary structure of refolded RNA is more well-determined and likely composed of fewer heterogeneous structural variants.

The differential SHAPE experiment uses reagents 1M6 and NMIA to specifically detect nucleotides involved in non-canonical and local tertiary interactions.[25,26] For example, in prior work, differential SHAPE reactivities were measured for the thiamine pyrophosphate (TPP) riboswitch in the presence and absence of a TPP ligand.[26] The ligand-bound RNA forms stable tertiary interactions and is characterized by a small number of relatively strong differential SHAPE reactivities. In contrast, the ligand-free TPP riboswitch does not form significant tertiary structure and its differential reactivity profile is distinct in two ways: there are more nucleotides with differential reactivities, and the magnitudes of their reactivities are smaller. This pattern is consistent with the view that the ligand-free state of the TPP riboswitch contains less stable or only transient tertiary structure.[26] A similar trend was observed between refolded and ex virion DENV2 RNA. Refolded RNA had fewer (n = 583) differential reactivities than ex virion RNA (n = 2770), but the reactivities of the refolded RNA were larger in magnitude on average (Figure 2D and Figure S1C). Thus, the differential SHAPE reactivity profiles qualitatively suggest that refolded RNA has more nucleotides that participate in stable tertiary and noncanonical structural interactions. In summary, although the two DENV RNA states are broadly similar, heat denaturation and refolding result in a more structured state overall, containing more well-defined secondary and tertiary structure interactions than does the ex virion state.

Effects of Refolding Treatment on Secondary Structure. The complete three-reagent SHAPE approach[25] was used to model the secondary structures of the entire ex virion and refolded DENV RNAs (Figure 3). Extensive benchmarking supports the view that the three-reagent approach yields RNA structure models in which a large fraction of correct base pairs are recovered accurately in regions with a stable, well-determined structure.[25] We compared the entire refolded SHAPE-derived minimum free energy (MFE) structure to the ex virion MFE structure in terms of its sensitivity (RF-sens) and positive predictive value (RF-ppv), which are 66% and 59%, respectively (Figure 3C). We define RF-sens and RF-ppv by treating the ex virion structure (the more direct biological state) as the reference. The RF-sens then reveals the percentage of the ex virion base pairs that the refolded structure accurately predicts. The RF-ppv gives the percentage of the refolded structure base pairs that are common between the two structures. The RF-ppv is smaller than the RF-sens because the refolded RNA model contains more base pairs (2766 bp) than the ex virion model (2447 bp). The differences in base-pairing patterns could represent real changes in secondary structure upon refolding treatment. Alternatively, base-pairing

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

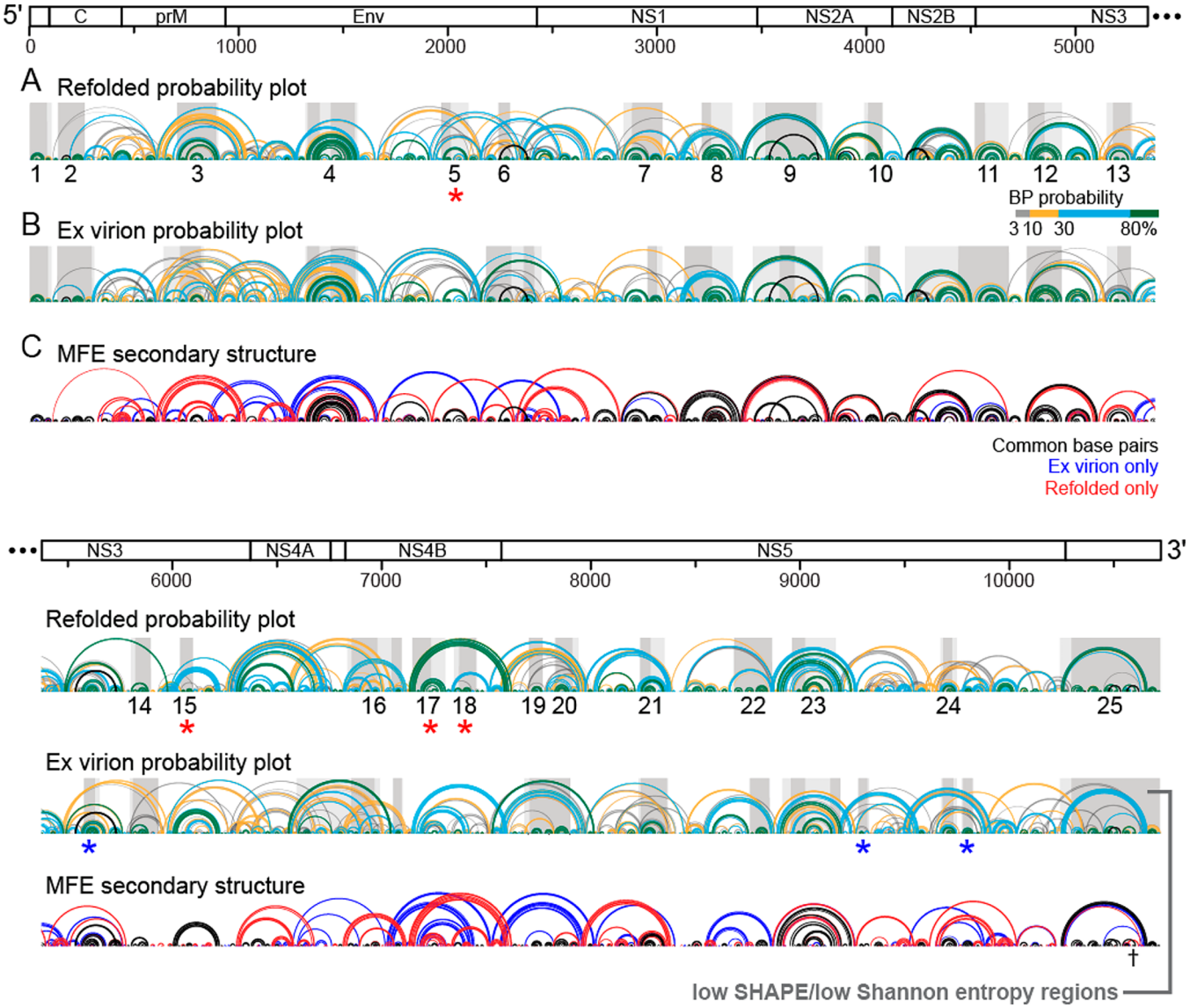


Figure 3. SHAPE-directed structure models for refolded and ex virion RNA states. (A) Refolded RNA base-pair probability plot. Base pairs are shown as arcs and are colored by their probability (see the key), with green arcs indicating the most probable helices. Black arcs indicate pseudoknots. Structural elements with both low SHAPE reactivity and low Shannon entropy are indicated by dark gray boxes. Regions that were expanded to include entire helices that intersect each low-SHAPE and low-entropy region are indicated by light gray boxes and are numbered. Red asterisks indicate structural elements that were identified as low SHAPE and low entropy in the (more highly structured) refolded RNA but not in the ex virion RNA. (B) Ex virion RNA base-pair probability plot. Arcs and elements of low SHAPE and low entropy are colored as described above. Blue asterisks indicate structural elements that were identified in the ex virion RNA but not in the refolded RNA. (C) Minimum free energy (MFE) secondary structure models. Base pairs common to both ex virion and refolded RNA are colored black. Base pairs unique to the ex virion and refolded models are colored blue and red, respectively. The dagger indicates a dumbbell pseudoknot structure that forms in the refolded RNA but not in ex virion RNA.

differences may not be significant if they occur in regions with high Shannon entropies, which are not structurally well-defined by our MFE models.

Stable, well-defined structural elements in large RNAs are those with both low SHAPE reactivity (corresponding to a high level of structure) and low Shannon entropy (reflecting a single highly probable structure). In prior analyses of SHAPE-MaP data for HIV-1, hepatitis C virus, and Dengue virus, regions of low SHAPE reactivity and low Shannon entropy (termed lowSS elements) overlapped nearly all motifs with known RNA-based functions and also identified regions with previously undetected function.[15,17,28] The lowSS framework identified 24 well-determined, highly structured elements in ex virion DENV2 RNA. Of the 22 elements located in the protein coding region, more than half were shown to be under clear evolutionary pressure.[17] The same lowSS analysis for the refolded DENV2 RNA yielded 25 lowSS elements (Figure 3A, numbered gray boxes). The 25 structural elements of the refolded RNA overlap well with the 24 previously identified lowSS elements in ex virion RNA (Figure 3B, gray boxes).[17] Four of the 24 identified elements are unique to the refolded RNA analysis [elements 5, 15, 17, and 18 (Figure 3A, red asterisks)].

We compared the structures of the ex virion and refolded RNAs across all 25 refolded RNA structural elements by calculating their refolded RNA model sensitivity (RF-sens) and RF-ppv. We also calculated the RF-sens and RF-ppv over the remaining or intervening 22 regions (with lengths of >50 nucleotides) that did not meet the low-SHAPE and low-entropy

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

criteria. The median RF-sens and median RF-ppv of the 25 low-SHAPE, low-entropy elements are 96% and 89%, respectively (Figure 4A); these values are much higher than the RF-sens and RF-ppv of the entire RNA genome (discussed above). Moreover, the median RF-sens and median RF-ppv for the 22 intervening (non-element, non-lowSS) regions are 54% and 47%, respectively (Figure 4B). The RF-sens and RF-ppv values of the 25 structural elements versus those of the 22 intervening regions are significantly different ($p < 0.001$; Kolmogorov–Smirnov test). Thus, whereas previous studies have shown that the lowSS metric is useful in identifying functional RNA structures in a large RNA, this study newly demonstrates that lowSS elements also show much stronger conservation of base-paired secondary structure after refolding treatment than does the RNA as a whole.

Comparison between refolded and ex virion RNAs

A

	lowSS elements		
	RF-sens	RF-ppv	length
1	84	97	105
2	97	97	110
3	57	49	195
4	84	75	244
5	100	98	119
6	100	100	57
7	100	98	183
8	98	89	147
9	90	87	328
10	100	100	68
11	90	93	151
12	98	96	155
13	100	93	118
14	100	100	85
15	100	82	54
16	67	66	254
17	28	28	141
18	26	22	94
19	100	95	52
20	88	76	115
21	98	85	113
22	71	64	161
23	85	84	205
24	96	96	75
25	87	85	477
median	96	89	119

B

	All other regions with lengths >50 nts		
	RF-sens	RF-ppv	length
	53	45	451
	27	20	419
	55	52	409
	64	54	146
	49	41	545
	67	61	189
	63	45	101
	67	57	204
	78	71	460
	89	62	107
	78	61	216
	51	52	532
	84	54	156
	39	36	743
	0	0	74
	43	31	274
	86	80	59
	40	33	296
	49	44	335
	43	50	116
	56	38	508
	42	36	496

Comparison to accepted structures

C

	Ex virion model		Refolded model	
	sens	ppv	sens	ppv
5'-UTR	97	82	91	88
5' capsid	99	98	100	100
3'-UTR*	86	84	94	90
3'-UTR	76	68	87	76

Figure 4. Comparison of DENV2 RNA structure models as a function of refolding and relative to regions with accepted structures. (A) RF-sens and RF-ppv values for the 25 elements of low SHAPE reactivity and low Shannon entropy (lowSS) identified in refolded DENV2 RNA, relative to the ex virion structure model. (B) RF-sens and RF-ppv values for all remaining regions of the DENV2 genome with lengths of at least 50 nucleotides. (C) RNA structure modeling accuracies relative to accepted DENV2 structures. The sens and ppv values of ex virion and refolded RNA models were calculated for elements 1 (5′-UTR), 2 (5′ portion of the capsid-coding region), and 24 or 25. In the ex virion and refolded RNAs, the 3′-UTR corresponds to elements 24 and 25, respectively (because there is one additional element in the refolded RNA). For the 3′-UTR* calculation, nucleotides proposed to be involved in a medium-range base-pairing interaction, which are not in the accepted model but for which there is direct experimental evidence,[17] were omitted from the sens and ppv calculations. For reference, the medium-range interaction is included in the 3′-UTR calculation (no asterisk). See the Supporting Information for a discussion.

The refolded and ex virion DENV structures are very similar in all elements containing known functional structures that were identified prior to our whole-RNA-genome SHAPE analysis [elements 1 (5′-UTR), 2 (5′ portion of the capsid-coding region), and 25 (3′-UTR)],[29−35] with RF-ppv values of 0.97, 0.97, and 0.85, respectively (Figures 3 and 4A). The main difference between the two models in these three regions is that a "dumbbell" pseudoknot in the 3′-UTR is predicted to form in the refolded RNA but is reactive by SHAPE and appears to be unstable in ex virion RNA (Figure 3, dagger).

There are four structural elements in which the RF-sens is relatively low (<70%): elements 3 and 16−18 (Figure 4A). In these regions, the ex virion RNA structure contains many base pairs with low probability (Figure S3). The ex virion base-pair probability plot contains many of the base pairs from the refolded MFE structure, but they are predicted with lower probabilities. Thus, in these four elements, the data support a model in which ex virion RNA samples multiple conformations; one of these is the MFE structure adopted by the refolded RNA, but this structure does not dominate the ex virion conformational ensemble.

Comparison of SHAPE-Predicted Models to Accepted DENV2 Structures. There are three regions in the Dengue RNA genome for which there are phylogenetically supported accepted structures.[29−31] Many of these structures have been modeled in previous SHAPE studies of the UTRs of full-length DENV or DENV minigenomes.[17,32−38] These accepted structures are in the 5′-UTR, the 5′ portion of the capsid-coding element, and the 3′-UTR. Strikingly, we find that all of the accepted structures fall in lowSS regions. Both the ex virion and refolded models accurately recover every accepted structural feature in the 5′-UTR and 5′ portion of the capsid-coding region (Figure S4), recover most of the major structural features in the 3′-UTR (Figure S5), and correctly model two pseudoknots located in the 5′-UTR and 3′-UTR. Both SHAPE-directed models miss several short pseudoknots in the 3′-UTR and predict a medium-range helix (involving positions 10247−10270 and 10640−10661) that is not included among accepted DENV2 models but that nonetheless has strong experimental support for the specific DENV2 strain (S16803) studied here.[17] Sens and ppv values for each SHAPE model were calculated for each element with an accepted secondary structure (Figure 4C); the sensitivities of the two models for these three elements range from 86 to 100%. Values for ppv are smaller than for sens, because the SHAPE-predicted models contain more base pairs than the accepted DENV2 structure. In summary, SHAPE-directed models, based on the ex virion or refolded RNAs, recover most accepted base pairs, as expected, as all accepted structures reside in lowSS regions.

DISCUSSION

By either necessity or convenience, researchers often study protein-free RNAs that have been refolded after denaturation. There is little information about how commonly used refolding approaches affect the structure of long RNAs, especially relative to more native-like cellular or virus-based states. Analyses of 1M7 SHAPE reactivities, differential SHAPE reactivities, and Shannon entropies all suggest that refolded DENV2 RNA has many features that are similar to those of a viral RNA genome that is likely in a native-like conformation (ex virion RNA) but

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

that there are also notable differences. In general, heat denaturation and refolding treatment results in a more structured and well-determined conformation. This difference suggests that the more native-like ex virion RNA samples a wider variety of states and may not uniquely form the most thermodynamically stable structure (Figure 5). Instead, some regions are only partially folded or sample multiple structures.

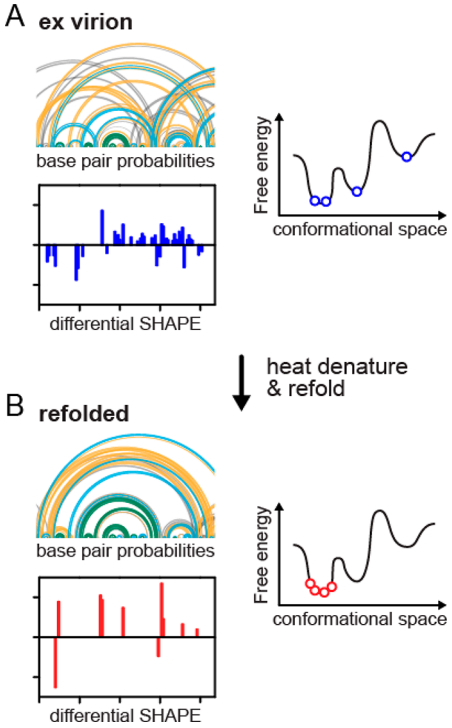


Figure 5. Model for the global effects of refolding on large-scale RNA structure. (A) Ex virion RNA contains multiple low-probability base pairs and a large number of low-magnitude differential SHAPE reactivities, suggesting that significant regions in the more native-like ex virion RNA are not structurally well-defined and instead sample multiple conformations. (B) Refolded RNA contains more highly probable base pairs and a smaller number of large differential SHAPE reactivities than the ex virion RNA, suggesting that heat denaturation and refolding result in a more structured and well-defined RNA conformation.

SHAPE reactivities were used to create secondary structure models for the entire DENV2 genome for the two RNA states. We were able to examine internucleotide interactions up to 500 residues apart. Most of the structural differences between the two models are located in regions that have high Shannon entropies and are thus intrinsically not well-defined by a single structural model. We identified 25 elements in the refolded RNA that are highly structured and predicted to favor a single, highly probable conformation. When only these elements were considered, the median RF-sens (percentage of the ex virion base pairs that the refolded structure accurately predicts) between the two RNA states was ~96% (Figure 4A). This high RF-sens suggests that analysis of refolded RNA is predictive of native RNA structure in regions in which both SHAPE reactivities and Shannon entropies are low.

Studying refolded RNA is likely justifiable for small RNAs, such as those optimized and studied by X-ray crystallography and nuclear magnetic resonance, assuming that the full RNA motif is included. Even with small RNAs, it is clear that the native cellular environment can modulate their structure.[4,13,14] A much greater degree of care must be taken when studying larger, more complex systems. This work shows that the native DENV2 genomic RNA is structurally more heterogeneous and contains fewer well-determined structural elements than its artificially refolded analogue. There are clear limitations to studying RNA motifs that are conformationally dynamic or involved in structural switches in refolded RNA, given that refolding of the long DENV2 RNA resulted in a more homogeneous and structured RNA architecture (Figure 5).

The low-SHAPE and low-Shannon entropy hypothesis has been evaluated previously, which revealed that regions with these characteristics overlap significantly with known functionally important regions in the HIV-1, HCV, and Dengue RNA genomes.[15,17,28] The work described here illustrates a new and additional useful feature of the lowSS criterion: lowSS regions also and simultaneously identify structural elements in a refolded RNA that are likely to correctly recapitulate RNA structures in a native-like RNA state. Furthermore, these structural elements in the refolded RNA overlap well with the previously identified low-SHAPE and low-Shannon entropy elements in ex virion RNA, of which many were shown to be under evolutionary pressure and likely necessary for viral replication. A subset of these elements likely form local folds with higher-order and tertiary structures.[15,17,28] The three regions in the DENV RNA genome with previously defined accepted structures also fall within lowSS regions for both ex virion and refolded models (Figure 4C). Thus, the lowSS structure and entropy metric appears to be highly useful for the discovery of important RNA structures, even in artificially refolded genomic RNA.

The lowSS framework emphasizes that not all predicted base pairs in a structural model should be given equal consideration and that structural well-determinedness can be established de novo. We strongly recommend reporting Shannon entropy values or related metrics alongside modeled RNA secondary structures to inform on the confidence (or lack thereof) in the predicted structural features across RNA genomes or transcriptomes. Moreover, the lowSS metric should find broad application in identifying regions that represent strong candidates for further functional evaluation[7,15,17,39] and for which in vitro transcripts can be used as good facsimiles of native RNAs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b01219.

Methods describing generation of an "accepted" DENV2 structure and detailed comparison to ex virion and refolded SHAPE-directed models, two figures showing statistical analysis of ex virion and refolded 1M7 and differential SHAPE reactivities and comparing the base-pair probability arcs for low-RF-sens elements, and two figures comparing the ex virion and refolded models to the accepted DENV2 structure (PDF)

Complete RNA SHAPE data sets for both ex virion and refolded states, predicted secondary structure models, and tables of lowSS regions for ex virion and refolded DENV2 RNA (ZIP)

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

AUTHOR INFORMATION

Corresponding Author

*E-mail: weeks@unc.edu.

ORCID

Kevin M. Weeks: 0000-0002-6748-9985

Funding

This work was supported by a grant from the National Institutes of Health (R35 GM122532 to K.M.W.). E.A.D. was a Lineberger Postdoctoral Fellow in the Basic Sciences (T32 CA009156) and a Ruth L. Kirschstein NRSA Fellow (F32 GM103180).

Notes

The authors declare the following competing financial interest(s): K.M.W. is an advisor to and holds equity in Ribometrix, to which mutational profiling (MaP) technologies have been licensed.

ACKNOWLEDGMENTS

Custom software used in this work is freely available from the corresponding author's Web site. Massively parallel sequencing data are available from the Sequence Read Archive, accession number SRP066015.

REFERENCES

(1) Mauger, D. M., Siegfried, N. A., and Weeks, K. M. (2013) The genetic code as expressed through relationships between mRNA structure and protein function. *FEBS Lett. 587*, 1180−1188.

(2) Dethoff, E. A., Chugh, J., Mustoe, A. M., and Al-Hashimi, H. M. (2012) Functional complexity and regulation through RNA dynamics. *Nature 482*, 322−330.

(3) Tyrrell, J., Weeks, K. M., and Pielak, G. J. (2015) Challenge of mimicking the influences of the cellular environment on RNA structure by PEG-induced macromolecular crowding. *Biochemistry 54*, 6447−6453.

(4) Tyrrell, J., McGinnis, J. L., Weeks, K. M., and Pielak, G. J. (2013) The cellular environment stabilizes adenine riboswitch RNA structure. *Biochemistry 52*, 8777−8785.

(5) McGinnis, J. L., and Weeks, K. M. (2014) Ribosome RNA assembly intermediates visualized in living cells. *Biochemistry 53*, 3237−3247.

(6) McGinnis, J. L., Liu, Q., Lavender, C. A., Devaraj, A., McClory, S. P., Fredrick, K., and Weeks, K. M. (2015) In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. *Proc. Natl. Acad. Sci. U. S. A. 112*, 2425−2430.

(7) Mustoe, A. M., Busan, S., Rice, G. M., Hajdin, C. E., Peterson, B. K., Ruda, V. M., Kubica, N., Nutiu, R., Baryza, J. L., and Weeks, K. M. (2018) Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell 173*, 181−195.e18.

(8) Lai, D., Proctor, J. R., and Meyer, I. M. (2013) On the importance of cotranscriptional RNA structure formation. *RNA 19*, 1461−1473.

(9) Zhang, J., and Landick, R. (2016) A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem. Sci. 41*, 293−310.

(10) Woodson, S. A. (2010) Taming free energy landscapes with RNA chaperones. *RNA Biol. 7*, 677−686.

(11) Serganov, A., and Nudler, E. (2013) A decade of riboswitches. *Cell 152*, 17−24.

(12) Uhlenbeck, O. C. (1995) Keeping RNA Happy. *RNA 1*, 4−6.

(13) 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.

(14) Leamy, K. A., Assmann, S. M., Mathews, D. H., and Bevilacqua, P. C. (2016) Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biophys. 49*, No. e10.

(15) Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E., and Weeks, K. M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods 11*, 959−965.

(16) Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., and Weeks, K. M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc. 10*, 1643−1669.

(17) Dethoff, E. A., Boerneke, M. A., Gokhale, N. S., Muhire, B. M., Martin, D. P., Sacco, M. T., McFadden, M. J., Weinstein, J. B., Messer, W. B., Horner, S. M., and Weeks, K. M. (2018) Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U. S. A. 115*, 11513−11518.

(18) Busan, S., and Weeks, K. M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with Shape Mapper 2. *RNA 24*, 143−148.

(19) 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.

(20) Mathews, D. H. (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA 10*, 1178−1190.

(21) Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinf. 11*, 129.

(22) Huynen, M., Gutell, R., and Konings, D. (1997) Assessing the reliability of RNA folding using statistical mechanics. *J. Mol. Biol. 267*, 1104−1112.

(23) Kelly, E. P., Polo, S., Sun, W., and Falgout, B. (2011) Evolution of attenuating mutations in dengue-2 strain S16803 PDK50 vaccine and comparison of growth kinetics with parent virus. *Virus Genes 43*, 18−26.

(24) Mortimer, S. A., and Weeks, K. M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc. 129*, 4144−4145.

(25) Rice, G. M., Leonard, C. W., and Weeks, K. M. (2014) RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA 20*, 846−854.

(26) Steen, K.-A., Rice, G. M., and Weeks, K. M. (2012) Fingerprinting Noncanonical and Tertiary RNA Structures by Differential SHAPE Reactivity. *J. Am. Chem. Soc. 134*, 13160−13163.

(27) Weeks, K. M., and Mauger, D. M. (2011) Exploring RNA Structural Codes with SHAPE Chemistry. *Acc. Chem. Res. 44*, 1280−1291.

(28) 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.

(29) Selisko, B., Wang, C., Harris, E., and Canard, B. (2014) Regulation of Flavivirus RNA synthesis and replication. *Curr. Opin. Virol. 9*, 74−83.

(30) Gebhard, L. G., Filomatori, C. V., and Gamarnik, A. V. (2011) Functional RNA elements in the dengue virus genome. *Viruses 3*, 1739−1756.

(31) Villordo, S. M., Carballeda, J. M., Filomatori, C. V., and Gamarnik, A. V. (2016) RNA Structure Duplications and Flavivirus Host Adaptation. *Trends Microbiol. 24*, 270−283.

(32) Sztuba-Solinska, J., Teramoto, T., Rausch, J. W., Shapiro, B. A., Padmanabhan, R., and Le Grice, S. F. J. (2013) Structural complexity of Dengue virus untranslated regions: cis-acting RNA motifs and pseudoknot interactions modulating functionality of the viral genome. *Nucleic Acids Res. 41*, 5075−5089.

(33) Liu, Z.-Y., Li, X.-F., Jiang, T., Deng, Y.-Q., Zhao, H., Wang, H.-J., Ye, Q., Zhu, S.-Y., Qiu, Y., Zhou, X., Qin, E.-D., and Qin, C.-F. (2013) Novel cis-acting element within the capsid-coding region enhances flavivirus viral-RNA replication by regulating genome cyclization. *J. Virol. 87*, 6804−6818.

(34) Villordo, S. M., Filomatori, C. V., Sánchez-Vargas, I., Blair, C. D., and Gamarnik, A. V. (2015) Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathog. 11*, No. e1004604.

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077

(35) de Borba, L., Villordo, S. M., Iglesias, N. G., Filomatori, C. V., Gebhard, L. G., and Gamarnik, A. V. (2015) Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J. Virol. 89*, 3430−3437.
(36) Chapman, E. G., Moon, S. L., Wilusz, J., and Kieft, J. S. (2014) RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *eLife 3*, No. e01892.
(37) Liu, Z.-Y., Li, X.-F., Jiang, T., Deng, Y.-Q., Ye, Q., Zhao, H., Yu, J.-Y., and Qin, C.-F. (2016) Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. *eLife 5*, 39926.
(38) Huber, R. G., Lim, X. N., Ng, W. C., Sim, A. Y. L., Poh, H. X., Shen, Y., Lim, S. Y., Sundstrom, K. B., Sun, X., Aw, J. G., Too, H. K., Boey, P. H., Wilm, A., Chawla, T., Choy, M. M., Jiang, L., de Sessions, P. F., Loh, X. J., Alonso, S., Hibberd, M., Nagarajan, N., Ooi, E. E., Bond, P. J., Sessions, O. M., and Wan, Y. (2019) Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun. 10*, 1408.
(39) Boerneke, M. A., Ehrhardt, J. E., and Weeks, K. M. (2019) Physical and Functional Analysis of Viral RNA Genomes by SHAPE. *Annu. Rev. Virol.*, DOI: 10.1146/annurev-virology-092917-043315.

DOI: 10.1021/acs.biochem.8b01219
Biochemistry 2019, 58, 3069−3077