



Structure-first identification of RNA elements that regulate dengue virus genome architecture and replication

Mark A. Boerneke^a, Nandan S. Gokhale^b, Stacy M. Horner^{b,c}, and Kevin M. Weeks^{a,1}

Edited by Juli Feigon, University of California, Los Angeles, CA; received October 5, 2022; accepted March 2, 2023

The genomes of RNA viruses encode the information required for replication in host cells both in their linear sequence and in complex higher-order structures. A subset of these RNA genome structures show clear sequence conservation, and have been extensively described for well-characterized viruses. However, the extent to which viral RNA genomes contain functional structural elements—unable to be detected by sequence alone—that nonetheless are critical to viral fitness is largely unknown. Here, we devise a structure-first experimental strategy and use it to identify 22 structure-similar motifs across the coding sequences of the RNA genomes for the four dengue virus serotypes. At least 10 of these motifs modulate viral fitness, revealing a significant unnoticed extent of RNA structure-mediated regulation within viral coding sequences. These viral RNA structures promote a compact global genome architecture, interact with proteins, and regulate the viral replication cycle. These motifs are also thus constrained at the levels of both RNA structure and protein sequence and are potential resistance-refractory targets for antivirals and live-attenuated vaccines. Structure-first identification of conserved RNA structure enables efficient discovery of pervasive RNA-mediated regulation in viral genomes and, likely, other cellular RNAs.

SHAPE-MaP | viral RNA structure | evolutionary conservation | functional motif discovery | RNA genome architecture

RNA viruses—including dengue, influenza, Ebola, Zika, and coronaviruses—represent serious threats to human health. Complex internal RNA structures in the genomes of these viruses usurp cellular metabolism and create gene regulation machineries that enable viral replication. RNA viruses often contain highly structured and obviously conserved RNA elements in their 5'- and 3'-untranslated regions (UTRs) (1–4), and less frequently in their coding regions (3, 5), with functions critical to viral replication and fitness. These highly conserved elements can often be identified using comparative sequence (or covariation) analysis, and it is likely that most genome regions with highly conserved structures and sequences have already been discovered in well-studied RNA viruses. We hypothesized that RNA viruses might contain additional, currently undetected, functional structural elements across their genomes. We therefore sought to assess if incorporating experimentally determined RNA structure information into a search strategy might make it possible to discover new functional elements in related viruses.

Dengue virus (DENV) is a single-stranded, positive-sense, enveloped RNA virus in the *Flaviviridae* family. DENV infection is the leading cause of mosquito-borne viral disease in humans (6, 7). The four major DENV serotypes share about 70% nucleotide identity but are antigenically distinct (8, 9). First-time DENV infections can cause mild to severe dengue (6), and subsequent heterotypic infections are associated with a higher risk for severe forms of the disease, resulting in significant mortality (7). DENV threatens more than one-third of the human population, and effective vaccines and therapeutics remain elusive (10). Discovery of functional elements co-occurring across the genomes of all DENV serotypes would facilitate efforts to design antidengue therapies.

The DENV RNA genome is 10.7 kb in length and encodes a single polypeptide that is processed into three structural (capsid, membrane and envelope) and seven nonstructural (or enzymatic) proteins (11). The 5'- and 3'-UTRs and the first 300 nucleotides of the coding region (encoding capsid) contain clearly conserved RNA structures with functions critical to the viral replication cycle (1, 5, 9, 12). Well-determined RNA tertiary structures are prevalent across a single serotype (DENV2) of dengue, several of which are important for viral fitness (13). Long-range RNA–RNA interactions, mapped by crosslinking, occur broadly in all four DENV serotypes and show partial conservation (14). Here we interrogated the secondary (base pairing) structure of the RNA genomes for all four DENV serotypes.

We find that all DENV genomes are highly structured overall as measured by SHAPE-MaP (selective 2'-hydroxyl acylation analyzed by primer extension and mutational

Significance

RNA viruses hijack cellular metabolism and promote their own replication using compact genomes that encode information in both their primary sequences and in higher-order structures. Most functional structures identified to date have been found by identifying specific sequences conserved across a group of related viruses. By inverting the conventional approach, first defining RNA structures and then looking for conservation of these motifs, we efficiently discover multiple previously unannotated motifs important for viral fitness. This work reveals that RNA structure is a powerful tool to find functional elements in viruses, identifies potential motifs useful in antiviral and vaccine development, and sets the stage for further discovery of functional elements in large viral, messenger, and noncoding RNAs.

Author contributions: M.A.B. and K.M.W. conceived the project; M.A.B., N.S.G., S.M.H., and K.M.W. designed research; M.A.B. performed research; M.A.B. contributed new reagents/analytic tools; M.A.B., S.M.H., and K.M.W. analyzed data; and M.A.B. and K.M.W. wrote the paper.

Competing interest statement: The authors have organizational affiliations to disclose, K.M.W. is an advisor to and holds equity in Ribometrix.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: weeks@unc.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2217053120/-/DCSupplemental>.

Published April 3, 2023.

profiling) chemical probing, but observed structures are not broadly conserved among serotypes. However, by taking a structure-first approach, we identified numerous compact genome regions with evidence of structural conservation. We found that a significant subset of these, newly identified, structure-similar RNA motifs affect viral fitness, promote a compact global genome architecture, interact with proteins, and regulate viral replication. Identifying experimentally defined structural conservation across the genomes of related RNA viruses, such as DENV serotypes, thus leads directly to discovery of new RNA-mediated functions and new opportunities for defining contributors to viral replication and for interfering with viral fitness.

Results

RNA Structural Conservation and Divergence in the Genomes of the Four DENV Serotypes. We used SHAPE-MaP (15, 16) chemical probing to obtain comprehensive single-nucleotide resolution measurements of RNA structure across full-length DENV1, DENV2, DENV3, and DENV4 genomes, gently extracted from viral particles (Fig. 1*A*). SHAPE reactivities were used to create high-quality (17) genome-wide secondary structure models. Large RNA molecules contain regions both that adopt well-determined stable structures and that are structurally dynamic and sample multiple conformations (16, 18, 19). We therefore calculated base-pairing probabilities across all possible structures in the Boltzmann ensembles of structures, consistent with SHAPE data for each DENV serotype. These pairing probabilities, visualized as arcs, were compared across DENV serotypes (Fig. 1*B*). Extensive prior work has shown that well-determined and highly structured RNA elements [termed low SHAPE-low Shannon entropy (lowSS) motifs (4, 16); in green in Fig. 1*B*] are overrepresented with functional elements.

As expected, SHAPE-directed models for regions with well-determined structures readily identified structures in the 5'- and 3'-UTRs and in the 5' end of the capsid-coding region known to be conserved across and functional for the four DENV serotype genomes (*SI Appendix, Fig. S1*) (1, 5, 9, 12). Thus, an experimentally (SHAPE)-informed, structure-first approach recapitulates conserved, functionally important motifs. Despite success in recapitulating known functional elements and the notable number of well-determined structures observed in each DENV genome, most well-determined RNA structures we defined across the four DENV genomes are not conserved (Fig. 1*B*, green arcs), even though these genomes share 70% sequence identity.

We therefore took a more focused approach and looked for compact, well-determined elements with notable structural similarity. Comparative analysis of SHAPE-informed structure models revealed 22, previously unnoted, RNA motifs with similar structures in two or more serotypes (Fig. 1*C* and *Materials and Methods*). The sequences of these elements were not as highly conserved as structures in and near the 5'- and 3'-UTRs, and their levels of structural similarity varied. Nonetheless, individual motifs could be identified that clearly showed similar overall RNA architectures, similar or identical base-paired stems with covarying base pairs, and similar or identical hairpin and internal loop sequences (Fig. 2 and *SI Appendix, Fig. S2*). None of these structures had been previously identified by conventional covariation analyses.

Structure-Similar RNA Motifs Regulate DENV Fitness. We assessed the contribution of structure-similar motifs to viral fitness by introducing mutations that individually disrupted 17 of the 22 elements identified by comparative structural analysis and examined these in functional assays (Fig. 3). Mutated

sequences for five elements could not be recovered without off-target mutations, despite multiple independent attempts using methods optimized for cloning of difficult viral repeat sequences (*Materials and Methods*). Structures are named by their location in the RNA genome sequence. We introduced synonymous mutations designed to disrupt the structure of these RNA motifs (and preserve protein sequence) into a full-length DENV2 RNA construct (13) (Fig. 3*A* and *SI Appendix, Fig. S3*). Synonymous mutations avoided rare codons and minimized changes in overall codon usage, nucleotide composition, dinucleotide content, and predicted formation of alternate structures to minimize indirect effects. Capped wild-type (WT) and mutant DENV2 RNAs were transfected into BHK-21 cells, and viral replication and infectivity were assessed by measuring intracellular viral RNA and infectious viral particles in the supernatant (DENV titer), respectively. As a control, we confirmed that known structure-disrupting mutations (13) in the NS2A coding-region reduced viral RNA levels and infectious particles. Strikingly, 10 of the 17 mutants moderately or severely attenuated one or both measures of viral fitness relative to WT virus (Fig. 3*B* and *C*), reducing viral RNA or DENV titer by at least 50% at 72 h post-transfection. Five mutants reduced levels of viral RNA or viral titer by greater than 75% (motifs 2200, 5600, 7080, 8500, 10000; Fig. 3*B* and *C*). By taking a structure-first approach, we thus identified numerous RNA elements in DENV genomes important for viral replication.

Multiple Structures Promote a Compact Global Genome Architecture. We next examined whether structure-similar RNA motifs in the DENV genomes are important for higher-order genome architecture and organization. We used dynamic light scattering to evaluate the hydrodynamic radii of WT and RNA structure-disrupting mutant DENV2 genomes. In the absence of protein, WT DENV2 RNA genomes fold into a state with a hydrodynamic radius of ~45 nm (Fig. 4*A*); an intact DENV virion, by comparison, has a radius of roughly 25 nm (20), indicating that interactions with viral structural proteins in the assembled virion further compact genome structure. RNA structure-disrupting mutations in five of the 17 elements significantly disrupted compaction of the protein-free RNA genome (Fig. 4*A* and *B*), reflected as radial size increases of 8 to 15 nm (15 to 25%) and 30 to 50% increases in mean global genome volume. Of these five structural elements, mutations in four also attenuated viral fitness in DENV2 functional assays (Fig. 3*D*). Thus, mutations in compact (30 to 70 nucleotide) local RNA structures caused large-scale changes in global folding of the DENV genome concomitant with, in most cases, functional consequences for viral fitness. These and other local structures across the DENV genome show low SHAPE reactivities in regions modeled to be single-stranded, consistent with additional higher-order RNA interactions that contribute to genome compaction. These results suggest one function of structure-similar DENV RNA motifs is to promote a compact global genome architecture.

A Structure-Similar RNA Motif in the Region Encoding NS3 Plays an Important Role in Replication. We selected two structure-similar RNA elements, whose disruption led to severe attenuation of viral fitness, for further investigation (the 5600 and 8500 elements, Fig. 3*B* and *C*). The 5600 element is located in the region that encodes NS3 and occurs in serotypes 1, 2, and 4. Mutation of this RNA structure [Fig. 5*A*, 5600^{mut} (the original screening mutant)] severely attenuated viral replication, reducing viral RNA by 60% and viral titer by 95% at 72 h post-transfection (Fig. 5*B* and *C*, 5600^{mut}). We designed two additional mutants to further investigate the functional

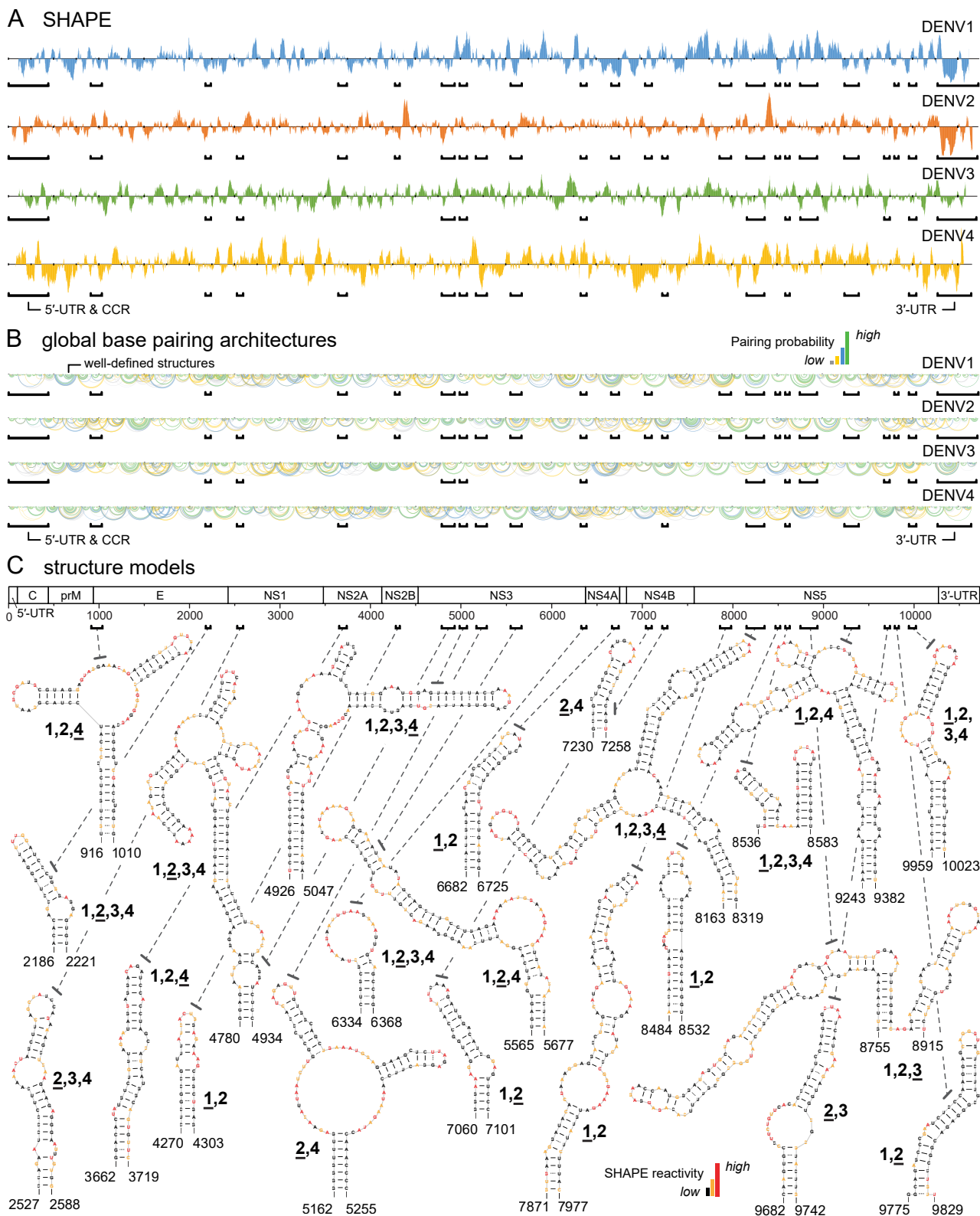


Fig. 1. Well-determined RNA structures across the four DENV serotype genomes. (A) Median SHAPE reactivities plotted over centered 55-nt windows. Regions of low SHAPE reactivity correspond to high levels of RNA structure. Motifs with similar structures in two or more serotypes are emphasized with brackets. (B) Structure models for the four DENV genomes displayed as base pair probability arcs. Green arcs indicate highly probable base pairs (see scale). Base pairing models for representative conserved structures in the 5'- and 3'-UTRs and CCR (capsid-coding region) are shown in *SI Appendix, Fig. S1*. (C) Twenty-two newly defined structure-similar RNA motifs in the coding regions of multiple serotypes. Numbers (1 to 4) indicate the serotypes containing each RNA structure; number is underlined for the representative serotype structure shown. Structures for other serotypes are shown in *SI Appendix, Fig. S2*. Secondary structures are colored by SHAPE reactivity (see scale).

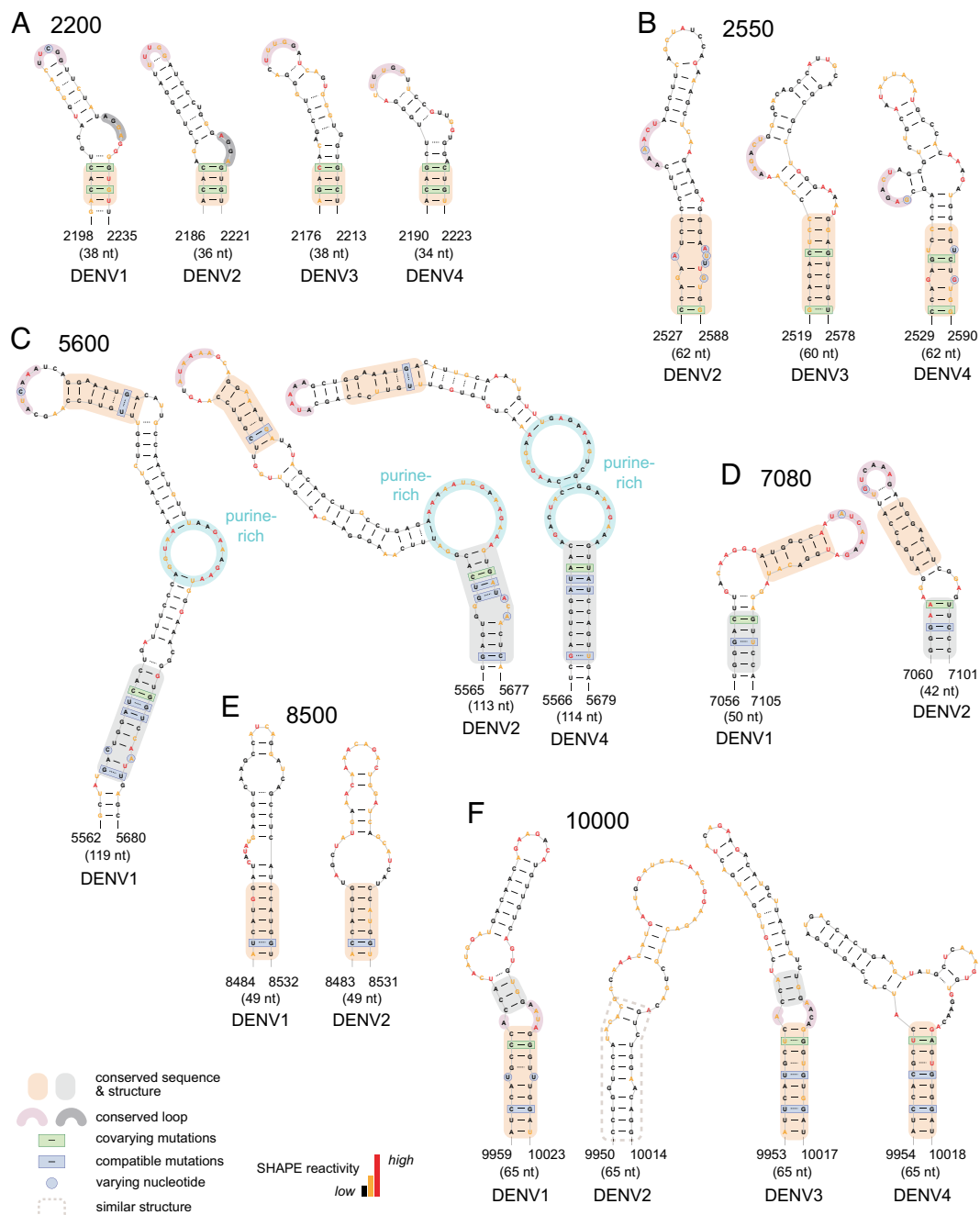


Fig. 2. Structure-similar RNA motifs in DENV genomes that affect viral fitness. Representative structure-similar motifs at genome locations (A) 2200, (B) 2550, (C) 5600, (D) 7080, (E) 8500, and (F) 10000 for individual serotypes. Mutants are named by their genome location. Structures are highlighted to show conserved structures and sequences, and covarying and compatible mutations across serotypes. If multiple serotype RNAs form a Watson-Crick or G-U base pair at a given location and both nucleotides differ, the base pair is classified as having covarying mutations. If a given base pair only varies between serotype RNAs at one nucleotide, the base pair is classified as having compatible mutations. Secondary structures are colored by SHAPE reactivity (see scale). Additional structure-similar motifs not shown here are illustrated in *SI Appendix, Fig. S2*.

importance of this RNA structure. The first mutant (*unzip2*) disrupted or unzipped this structure by introducing synonymous mutations that break five base pairs in the stem (Fig. 5A). These five base pairs were chosen as they could be recoded to restore base pair formation with additional synonymous compensatory mutations in a second mutant (*recode2*, Fig. 5A). The first mutant showed an attenuating functional effect similar to the original screening mutant (5600^{mut}), even though it disrupted fewer base pairs. The second mutant (*recode 2*), containing compensatory mutations, showed a modest rescue of fitness compared to *unzip2* in DENV2 infectivity and replicon (discussed below) assays (Fig. 5 B–D).

We next examined the role of the RNA structure at position 5600 using a DENV2 replicon construct, which reports on viral translation and replication, but not viral entry or viral packaging and assembly steps. In this construct, viral structural protein genes are replaced with a *Renilla luciferase* gene (21, 22). The 5600 mutations (Fig. 5A) were introduced into the replicon construct. The mutations had no effect on replicon reporter expression at 8 h post-transfection, indicating that translation (which occurs relatively rapidly in this system) is unaffected (*SI Appendix, Fig. S4A*). At 72 h post-transfection, levels of replicon reporter expression from 5600^{mut} , *unzip2*, and *recode2* mutant constructs were attenuated relative to levels of the WT construct (Fig. 5D).

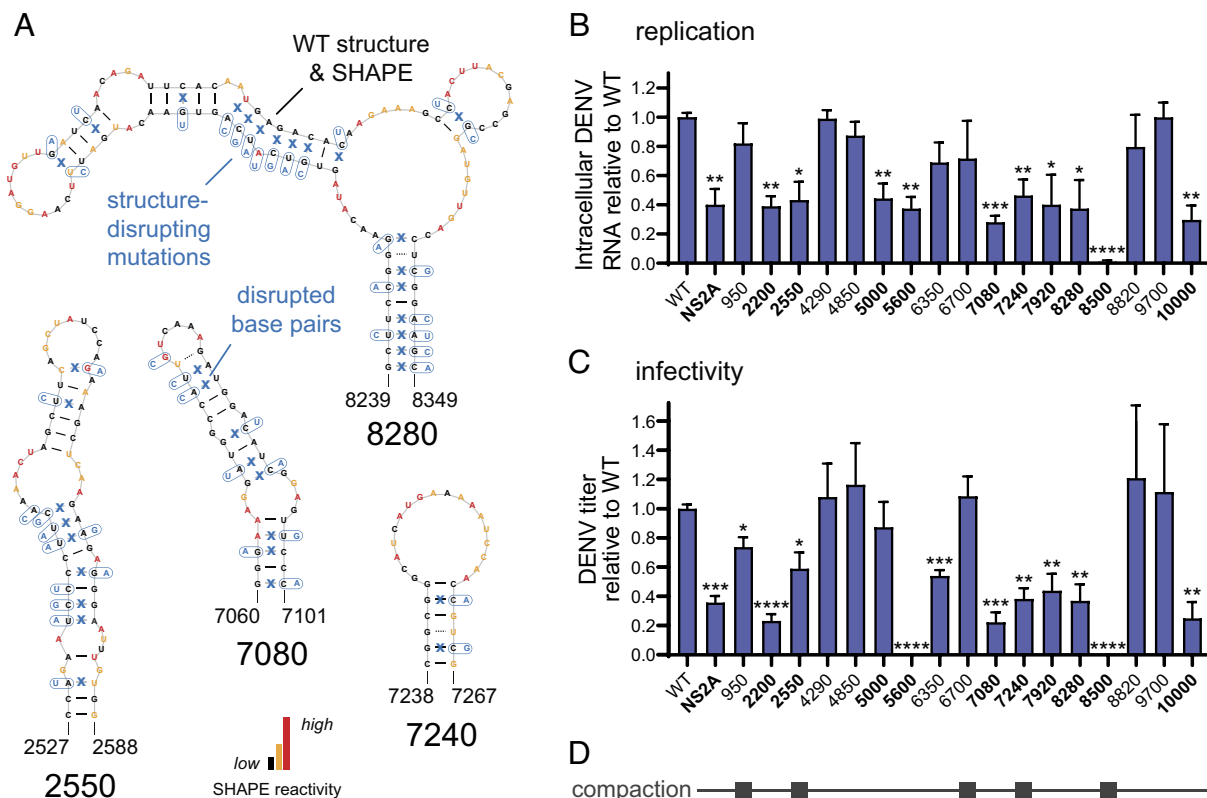


Fig. 3. Mutation of structure-similar DENV RNA motifs and effect on fitness. (A) Examples of synonymous mutations (blue) designed to disrupt the structures of DENV RNA elements (others shown in *SI Appendix*, Fig. S3). WT secondary structures are colored by SHAPE reactivities. (B) Effects of structure-disrupting mutations on replication, quantified by measuring intracellular DENV RNA relative to WT by RT-qPCR at 72 h post-transfection. (C) Effects of structure-disrupting mutations on infectivity quantified by measuring infectious viral particles in the supernatant (DENV titer) relative to WT at 72 h post-transfection. For panels B and C, values are plotted as mean \pm SEM of three biological replicates; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; two-tailed unpaired t test. An attenuating structure-disrupting mutation in the NS2A coding-region (13) is shown for comparison (NS2A). Mutants that attenuated viral replication and infectivity, and reduced viral RNA by $>50\%$ are emphasized in boldface text. (D) Summary of structure-similar DENV RNA motif mutants that significantly affect genome compaction (compare with Fig. 4).

The reductions were similar in magnitude as those observed for replication and infectivity assays using the full-length DENV infectious construct (Fig. 5 B and C). Combined, these results suggest that the RNA structure at position 5600 of the DENV2 genome is important for RNA genome replication but not viral entry, translation, or packaging.

A Structure-Similar RNA Motif in the Region Encoding NS5 Regulates Viral Replication and Packaging and Is Bound by a Protein Partner. The 8500 element is located in the NS5-encoding

region and is present in serotypes 1 and 2. The mutation of this RNA structure [Fig. 5E, 8500^{mutt} (the original screening mutant)] led to the most severely attenuating phenotype in our large functional screen (Fig. 5 F and G, 8500^{mutt}). We designed two additional mutations to individually destabilize the lower or upper stem structures in this motif (Fig. 5E, unzip-lower and unzip-upper). These two mutants showed intermediate attenuating functional effects in DENV2 replication and infectivity assays (Fig. 5 F and G) relative to the original screening mutant (8500^{mutt}), emphasizing the importance of this entire RNA structure for virus function. It was

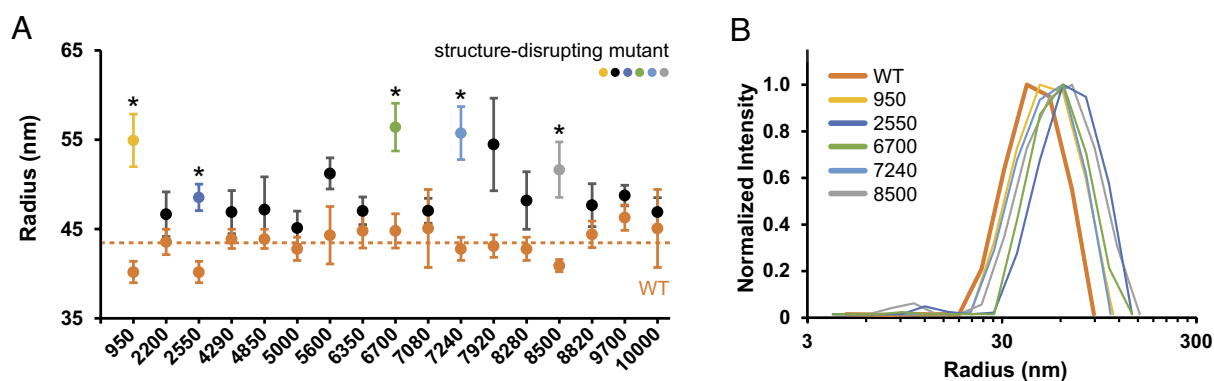


Fig. 4. Global genome sizes of WT and mutant DENV RNAs. (A) Radii of WT (orange) and mutant RNAs (black, colored). Mutant RNAs whose radii show a significant size difference relative to WT are emphasized with asterisks. Mutant RNA sizes were measured in parallel with WT for each independent measurement, and illustrate variability in dynamic light scattering measurements. Median WT RNA size across all independent measurements is indicated with dashed orange line. Values are mean \pm SEM calculated from three experiments. * $P < 0.025$, two-tailed unpaired t test. (B) Experimental measurements of the hydrodynamic radii of DENV RNAs by dynamic light scattering for representative genomic RNAs.

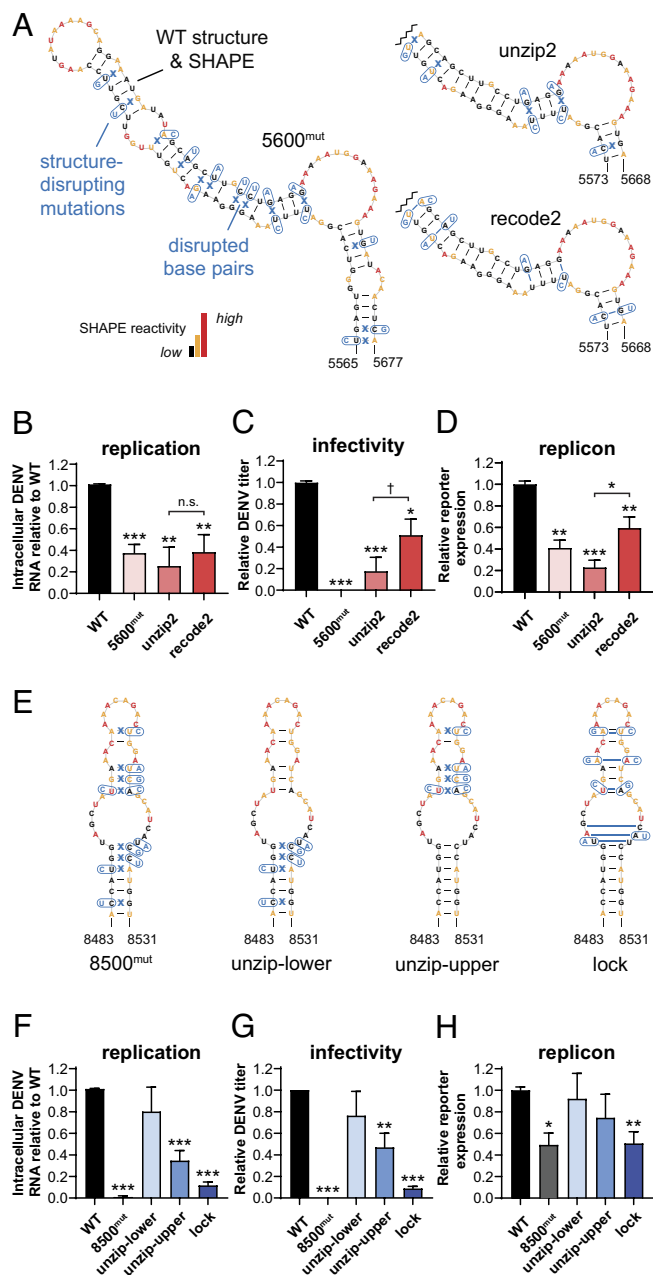


Fig. 5. Regulation of viral fitness by RNA elements 5600 and 8500. (A) Structure-disrupting [5600^{mut} (original screening mutant) and unzip2] and structure-restoring (recode2) mutations (blue) in the 5600 RNA element. (B–D) Effects of mutation in the 5600 element on (B) replication, quantified by measuring intracellular DENV RNA relative to WT, (C) infectivity, quantified by measuring infectious viral particles in the supernatant (DENV titer) relative to WT, and (D) replicon reporter expression relative to WT, post-transfection of full-length or replicon RNA. (E) Structure-disrupting [8500^{mut} (original screening mutant), unzip-lower, and unzip-upper] and structure-strengthening (lock) mutations (blue) in the 8500 RNA element. (F–H) Effects of mutation on (F) replication, (G) infectivity, or (H) replicon reporter expression. WT secondary structures are colored by SHAPE reactivity (see scale in A). The 5600^{mut} and 8500^{mut} mutants are the same as those shown in Fig. 3. Values plotted as mean \pm SEM of three biological replicates; † $P < 0.15$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; n.s., not significant; two-tailed unpaired t test.

not possible to introduce synonymous compensatory mutations to restore the original stem structure. Instead, we designed a structure-strengthening mutant by introducing synonymous mutations that created additional base pairs (Fig. 5E, lock). This locked structure mutant displays a severe attenuating phenotype in DENV replication and infectivity assays (Fig. 5F and G, lock), similar to that of the original screening mutant (8500^{mut}). The

functional role of the 8500 RNA structure thus is finely tuned, as structure strengthening and weakening both lead to severe attenuation of viral fitness.

In the context of the replicon construct, the 8500 mutations had no effect on replicon reporter expression at early time points post-transfection, consistent with normal translation (SI Appendix, Fig. S4B). At 72 h post-transfection, however, the mutations had consistently smaller effects on replicon reporter expression relative to the attenuating functional effects observed in replication and infectivity assays using the full-length construct (Fig. 5H). This difference suggests that the 8500 element functions in packaging or entry stages of the viral replication cycle as well as in the replication stage. Unbiased RNA–protein interaction crosslinking experiments (23, 24) in infected cells revealed dense RNA–protein crosslinking sites to the 8500 region of the genomic RNA (SI Appendix, Fig. S5). Crosslinking was substantially reduced when the structure was mutated to form the “locked” structure. These observations suggest that protein binding at this RNA structure is important for its regulatory function.

In sum, follow-up investigations of the structure-similar RNA motifs at 5600 and 8500 (Fig. 5) provide evidence that the structures these RNA elements form, rather than the encoded sequence, are important for their functions. Further investigation of the 2200 RNA element, present in the Env-encoding region of all four serotypes, also supports the importance of RNA structure for its function (SI Appendix, Fig. S6).

Discussion

RNA viruses densely encode RNA-based information in their small genomes. Most viral RNA structures characterized to date are located in 5′- and 3′-UTRs, where covariation analyses can identify highly conserved, functional structures. Identifying functional elements within coding sequences or that show more modest levels of conservation is much more challenging. We hypothesized that RNA structure might provide a powerful guide for discovering new functional elements in RNA viruses, including elements for which sequence conservation is not readily detectable. This strategy proved remarkably efficient and successful: we characterized structures across the entire genomic RNAs of the four DENV serotypes, identified 22 previously unnoted structure-similar motifs, were physically able to test 17 of these (in one of the largest screens of viral RNA structure and function to date), and identified 10 RNA structures that affect viral fitness. Our work significantly expands the list of known functional RNA elements in the DENV genome and reveals that DENV (and likely other RNA viruses) extensively exploits RNA structure-based mechanisms, based on motifs in the protein-coding regions of their genomes. These RNA motifs function to promote a compact global genome architecture, interact with proteins, and regulate the DENV replication cycle (Fig. 6A).

Using synonymous codon substitutions, it was remarkably straightforward to disrupt RNA motifs with similar structures across DENV serotypes and thereby create attenuated viruses. This insight directly motivates a strategy for designing live-attenuated vaccines with a reduced likelihood of reversion to virulence. Four RNA elements whose disruption led to attenuated viruses (2200, 5600, 7920, and 10000) simultaneously encode amino acid sequences overlapping key functional motifs in DENV proteins (Fig. 6B and SI Appendix, Fig. S7 and Table S1). 2200 encodes a part of a domain in Env (termed the stem) that mediates a conformational change involved in viral fusion (25). 5600 encodes a domain in the NS3 helicase involved in sequence-specific recognition of viral RNA (26). Two elements, 7920 and 10000 fall in NS5.

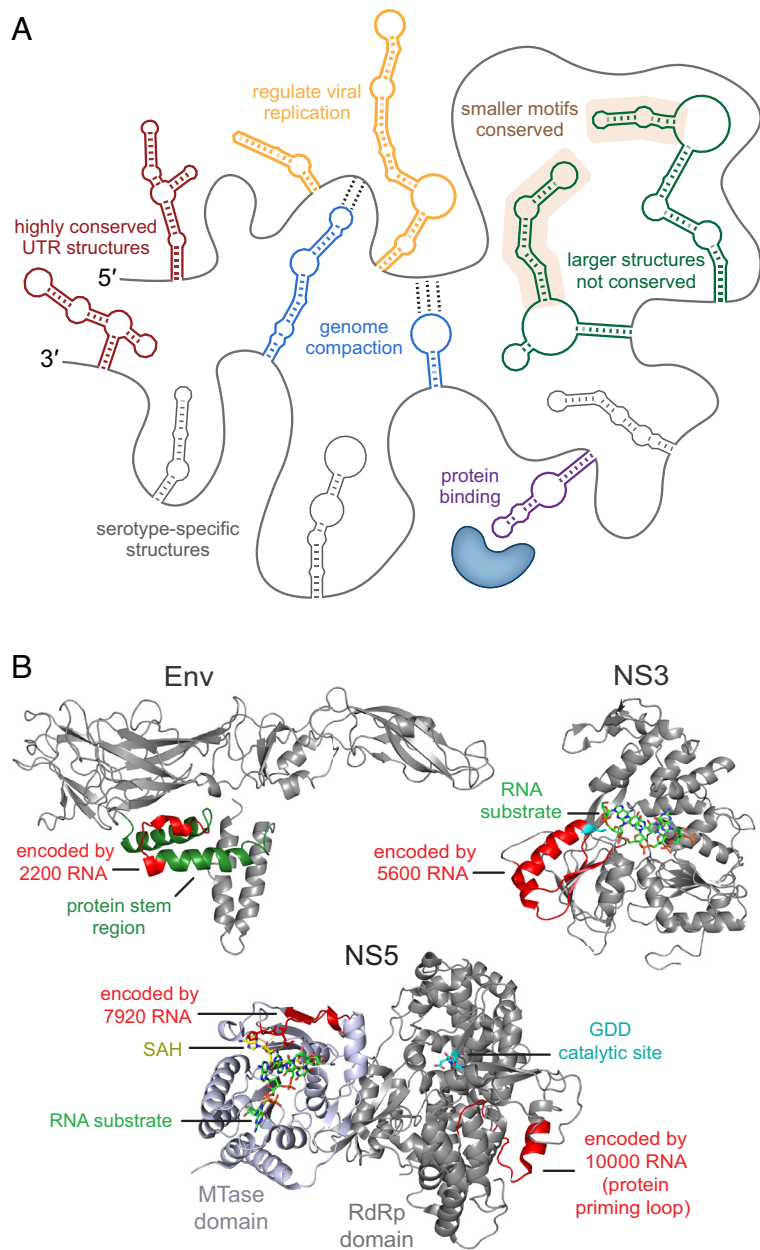


Fig. 6. Classes of functional viral RNA genome structures. (A) RNA genomes of related viruses contain serotype-specific RNA structures, highly conserved RNA structures (primarily in the 5'- and 3'-UTRs), and internal coding-region RNA structures with varying degrees of conservation. Structure-similar coding-region motifs play diverse functions including promoting genome compaction, interacting with protein binding partners, and regulating specific stages of the viral replication cycle. (B) Visualization of RNA sequences with functional roles at both protein-coding and RNA-structure levels. Functional regions in DENV Env (PDB: 3J27), NS3 (5XC6), and MTase and RdRp domains of NS5 (5D7O) proteins are encoded by functional RNA structure motifs 2200, 5600, 7920, and 10000, respectively. RNA and protein sequences, with protein structures and functions, are shown in greater detail in *SI Appendix, Fig. S7*.

7920 encodes the pocket in the MTase (methyl transferase) domain that recognizes conserved nucleobases at the 5' start of the genome and the SAM (S-adenosyl methionine) cofactor (27). 10000 overlaps the RdRp (RNA dependent RNA polymerase) domain and encodes a portion of the polymerase priming loop that enables RNA polymerization in the absence of a primer strand (27). These four regions are thus under dual evolutionary constraints, at both RNA structure and protein coding sequence levels, and face heightened barriers to reversion. In addition, small molecule inhibitors targeting doubly-constrained regions at the RNA and protein levels might encounter greater impediments to the development of viral resistance, perhaps especially for DENV NS5 RdRp inhibitors (*SI Appendix, Fig. S7*) (28, 29).

Well-defined (highly structured and low entropy) motifs are common and ubiquitous throughout each of the four DENV genomic RNAs (Figs. 1 and 6A). Widespread formation of internal secondary structure in large RNAs is now clearly an unremarkable and expected result (16). The observation of extensive internal structure is consistent with our previous findings in the

DENV2 genome (13), with independent studies of positive-sense RNA viruses (4, 14, 15, 30–36), and for essentially all large RNAs (16, 37, 38). Structure-first analysis further revealed that, although well-defined structures are common, most RNA structures in coding regions are not strongly conserved across serotypes. Extensive structural divergence between variants or serotypes has also been observed for the genome coding regions of RNA viruses from lentivirus, hepacivirus, and alphavirus genera (31, 34, 39). Lack of conservation does not preclude functional importance. For example, a subset of genome structures important for function in HIV-1 and Sindbis virus are not conserved in the related simian immunodeficiency and Venezuelan equine encephalitis viruses, respectively (34, 39). In addition, RNA viruses may require a general architecture in a specific genome region, rather than requiring specific individual structures to function. For example, multiple studies find that related RNA viruses often contain highly or lowly structured regions, respectively, at similar genome locations (14, 31, 34, 35, 40). Well-defined RNA secondary structures also often create

accessibility-switches, either occluding or presenting key functional motifs, where a specific structure is not required but the simple formation of some kind of structure is (41).

Each of the structure-similar RNA motifs we identified across DENV coding sequences (Fig. 6A) has a precedent among functional elements found previously in non-coding regions. Five RNA structures (950, 2550, 6700, 7240, and 8500) promote a compact global genome architecture. We hypothesize, first, that genome compaction facilitates assembly of viral replication complexes and of RNA packaging into the limited space within the virion and, second, that there are many RNA-structure-directed ways to accomplish these functions. Analogously, the more compact circularized conformation of the DENV RNA genome (as compared to the linear conformation) is required for packaging into a virion, consistent with a link between genome compaction and viral packaging (13). Two structure-similar motifs (5600, 8500) are important specifically for viral replication, and one of these (8500) also functions in packaging or entry. Similarly, RNA structures regulating replication, packaging, or both, viral stages have been identified in the 5'- and 3'-UTRs of the DENV genome (1, 5, 9). The 8500 element specifically also appears to bind a protein partner; again similarly, RNA viruses use structures in their 5'- and 3'-UTRs to mediate interactions with proteins (3, 42, 43). This work ultimately significantly extends our understanding of the extent to which information for function is encoded in viral RNA genomes beyond sequence alone, at the level of higher-order RNA structure, and sets the stage for future investigations in other RNA viruses.

In sum, structure-first comparison of DENV RNA genomes led to the efficient discovery of multiple novel functional RNA elements across DENV serotypes, specifically across coding regions. Most coding-region structures display weaker signatures of conservation than UTR region structures, but many still play important roles in viral fitness. Identification of structure-similar RNA motifs will likely be broadly useful for the discovery of RNA-mediated functions that are not immediately detectable at the sequence level in diverse other viral RNA genomes, and potentially in messenger RNAs and long noncoding RNAs. Efficient discovery of conserved functional elements in related RNA viruses outlines a pathway for development of antiviral therapies and live-attenuated vaccines that inclusively target multiple serotypes, genotypes, or strains of a virus. Broadly, functional RNA motifs, like those found to have similar structures in multiple DENV genomes, are harbingers for new biological roles of higher-order RNA structure and are potential targets for RNA-directed therapeutics.

Materials and Methods

Complete, detailed descriptions of all methods, including stepwise descriptions of the data analysis pipelines, are provided in the *SI Appendix*.

RNA Structure Probing. Structure probing experiments were performed on RNA gently extracted from purified virions [DENV1 (strain: West Pacific 74), DENV2 (strain 16681), DENV3 (CH53489), and DENV4 (TVP-360)] avoiding heating, metal ion chelation, ethanol precipitation, and other potentially denaturing steps (13). RNA was modified with the SHAPE reagent 1-methyl-7-nitroisatoic anhydride (1M7) (44). SHAPE-MaP and RNP-MaP experiments, library preparation, and sequencing were performed as described (23, 44).

Pipeline for Identifying Structure-Similar Motifs. We devised a strategy for specifically identifying structure-similar RNA motifs, colocalized in defined local regions across the four DENV RNA genomes. The pipeline, with details given below, involved: 1) Sequence alignment and mutation parsing, 2) secondary structure modeling, 3) initial analysis of structure-similar RNA motifs among DENV serotypes, and 4) three-model strategy for defining structure-similar RNA motifs.

Sequence Alignment and Mutation Parsing. FASTQ files from sequencing runs were inputted into *ShapeMapper 2* (45) for read alignment, mutation counting, and SHAPE reactivity profile generation. Median read depths of all SHAPE-MaP and RNP-MaP samples and controls were greater than 50,000; nucleotides with read depths of less than 5,000 were excluded from analysis.

Secondary Structure Modeling. *Superfold* (44) was used with SHAPE reactivity data to inform RNA structure modeling. Default parameters (except that the maximum pairing distance was set to 300 nt) were used to generate base-pairing probabilities for all nucleotides and minimum free energy structure models. These parameters yielded good models for previously characterized structures in the 5'- and 3'-UTRs (*SI Appendix, Fig. S1*). Local median SHAPE reactivities were calculated over centered sliding 55-nt windows to identify structured RNA regions.

Initial Analysis of Structure-Similar RNA Motifs among DENV Serotypes. DENV1, DENV2, DENV3, and DENV4 genome structure models were displayed as base pairing probability arcs (max pairing distance of 300 nt) and viewed as four tracks in the Integrative Genomics Viewer (IGV) (46). These comparisons revealed well-determined RNA structures [with high probability (>80%) base pairs] in the 5'- and 3'-UTRs and the 5' end of the capsid-coding region (including cHP and DCS-PK structures; capsid-coding region hairpin and downstream of 5' cyclization sequence pseudoknot, respectively), as expected, reflecting known conservation across the four DENV serotype genomes. Each serotype further contained numerous additional well-determined structures, only a small fraction of which showed potential conservation across serotypes.

Three-Model Strategy for Defining Structure-Similar RNA Motifs. *Superfold* was used to generate two additional structure models for each serotype with maximum base-pairing distances set to 100 and 600 nts. For each individual serotype, base pairing probability arcs for each of the three structure models were visualized as three tracks in a single IGV session file. RNA structure motifs with pairing probabilities >30% that were modelled identically in all three structure models for an individual serotype were selected as search motifs to query the other serotypes for similar structures; base pairs with pairing probabilities >30% in any of the three structure models of the other serotypes were accepted as potential support for similar structures. To be selected as similar structures for functional analyses, entire structures or substructure motifs (dashed boxes in *SI Appendix, Fig. S2*) were required to be identically sized and located at similar genome locations, or meet defined similarity thresholds in structure, size, or sequence (Fig. 2 and *SI Appendix, Fig. S2*). Structure-similar RNA elements were required to occur in DENV2 to leverage established DENV2 functional assays (13).

Design of RNA Structure-Disrupting DENV Mutants. Sequences were generated by introducing synonymous mutations designed to disrupt base pairing in the SHAPE-directed structural model. Mutant sequences were chosen to: 1) avoid rare codons, 2) minimize change in overall codon usage, minimize change in 3) nucleotide composition and 4) dinucleotide content, and 5) minimize formation of alternate structures. Structures for mutant sequences were modeled using *Superfold* to confirm disruption of the WT structure and to avoid sequences that altered neighboring structural elements. Mutants were generated in the context of p16681-T7G, which is based on the full-length DENV2 infectious clone p16681 (13).

Physical Size Evaluation and Phenotypic Analysis of DENV Mutants. RNA was synthesized and transfected as described (13). Dynamic light scattering experiments were performed on uncapped, in vitro transcribed, refolded DENV2 RNA (13). Replication assays were performed using capped in vitro transcribed RNA (13, 23). Viral translation and replication was measured for RNA structure-disrupting mutations in the 5600 and 8500 elements using a *Renilla* luciferase-expressing DENV2 replicon assay (22). RNP-MaP treatment was performed on infected cells and RNP-MaP site densities (23) were calculated over centered sliding 15-nt windows to compare WT and mutant RNAs.

Data, Materials, and Software Availability. All software employed in this paper is published. *ShapeMapper 2* (45) and *Superfold* (44) are available at <https://weekslab.com/software> (under "ShapeMapper 2" and "SHAPE-MaP" headings, respectively) and at <https://github.com/weeks-unc> (in the "shapemapper2" and "Superfold"

repositories, respectively). VARNA (47), IGV (46), and Sequence Manipulation Suite 2 (48) are third-party, open-source software. Raw and processed sequencing datasets analyzed in this study have been deposited in the Gene Expression Omnibus (GEO) database, <https://www.ncbi.nlm.nih.gov/geo/> (accession number GEO: GSE226865). All other data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by NIH R35 GM122532 (to K.M.W.). M.A.B. was supported by an NIH Ruth L. Kirschstein Postdoctoral Fellowship (F32 GM128330) and an NIH Pathway to Independence Award (K99 AI156640). N.S.G. and S.M.H. were supported by the Burroughs Wellcome Fund

and NIH grant R01AI125416. The DENV2 replicon sequence-containing plasmid was the generous gift of the J. Carette laboratory (Stanford U.), and we thank W. Qiao (Stanford U.) for her initial support in its application. Dynamic light scattering experiments were performed at the University of North Carolina Macromolecular Interactions Facility (NIH grant P30CA016086).

Author affiliations: ^aDepartment of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290; ^bDepartment of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710; and ^cDepartment of Medicine, Duke University Medical Center, Durham, NC 27710

1. L. G. Gebhard, C. V. Filomatori, A. V. Gamarnik, Functional RNA elements in the dengue virus genome. *Viruses* **3**, 1739–1756 (2011).
2. R. P. Smyth, M. Negroni, A. M. Lever, J. Mak, J. C. Kenyon, RNA structure—A neglected puppet master for the evolution of virus and host immunity. *Front. Immunol.* **9**, 2097 (2018).
3. Z. A. Jaafar, J. S. Kieft, Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol.* **17**, 110–123 (2019).
4. M. A. Boerneke, J. E. Ehrhardt, K. M. Weeks, Physical and functional analysis of viral RNA genomes by SHAPE. *Annu. Rev. Virol.* **6**, 93–117 (2019).
5. L. de Borja *et al.*, Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J. Virol.* **89**, 3430–3437 (2015).
6. S. Bhatt *et al.*, The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
7. T. C. Pierson, M. S. Diamond, The continued emerging threat of flaviviruses. *Nat. Microbiol.* **5**, 796–812 (2020).
8. M. G. Guzman, E. Harris, Dengue. *Lancet* **385**, 453–465 (2015).
9. S. M. Villordo, J. M. Carballeda, C. V. Filomatori, A. V. Gamarnik, RNA structure duplications and flavivirus host adaptation. *Trends Microbiol.* **24**, 270–283 (2016).
10. D. R. Martinez, S. W. Metz, R. S. Baric, Dengue vaccines: The promise and pitfalls of antibody-mediated protection. *Cell Host Microbe* **29**, 13–22 (2021).
11. D. J. Gubler, Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* **11**, 480–496 (1998).
12. K. Clyde, J. Barrera, E. Harris, The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. *Virology* **379**, 314–323 (2008).
13. E. A. Dethoff *et al.*, Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11513–11518 (2018).
14. R. G. Huber *et al.*, Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.* **10**, 1408 (2019).
15. N. A. Siegfried, S. Busan, G. M. Rice, J. A. E. Nelson, K. M. Weeks, RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–65 (2014).
16. K. M. Weeks, SHAPE directed discovery of new functions in large RNAs. *Acc. Chem. Res.* **54**, 2502–2517 (2021).
17. C. E. Hajdin *et al.*, Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5498–503 (2013).
18. E. A. Dethoff, K. M. Weeks, Effects of refolding on large-scale RNA structure. *Biochemistry* **58**, 3069–3077 (2019).
19. C. A. Giannetti, S. Busan, C. A. Weidmann, K. M. Weeks, SHAPE probing reveals human rRNAs are largely unfolded in solution. *Biochemistry* **58**, 3377–3385 (2019).
20. R. J. Kuhn *et al.*, Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717–725 (2002).
21. C. D. Marceau *et al.*, Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* **535**, 159–163 (2016).
22. Y. S. Ooi *et al.*, An RNA-centric dissection of host complexes controlling flavivirus infection. *Nat. Microbiol.* **4**, 2369–2382 (2019).
23. C. A. Weidmann, A. M. Mustoe, P. B. Jariwala, J. M. Calabrese, K. M. Weeks, Analysis of RNA-protein networks with RNP-MaP defines functional hubs on RNA. *Nat. Biotechnol.* **39**, 347–356 (2021).
24. C. Iserman *et al.*, Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell* **80**, 1078–1091.e6 (2020).
25. S.-R. Lin *et al.*, The helical domains of the stem region of dengue virus envelope protein are involved in both virus assembly and entry. *J. Virol.* **85**, 5159–5171 (2011).
26. C. M. D. Swarbrick *et al.*, NS3 helicase from dengue virus specifically recognizes viral RNA sequence to ensure optimal replication. *Nucleic Acids Res.* **45**, 12904–12920 (2017).
27. A. E. Sahili, J. Lescar, Dengue virus non-structural protein 5. *Viruses* **9**, 91 (2017).
28. H. Shimizu *et al.*, Discovery of a small molecule inhibitor targeting dengue virus NS5 RNA-dependent RNA polymerase. *PLoS Negl. Trop. Dis.* **13**, e0007894 (2019).
29. R. Arora *et al.*, Two RNA tunnel inhibitors bind in highly conserved sites in dengue virus NS5 polymerase: Structural and functional studies. *J. Virol.* **94**, e01130–20 (2020).
30. C. P. Burrill *et al.*, Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity. *J. Virol.* **87**, 11670–116783 (2013).
31. D. M. Mauger *et al.*, Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3692–3697 (2015).
32. N. Pirakitikulr, A. Kohlway, B. D. Lindenbach, A. M. Pyle, The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol. Cell* **62**, 111–120 (2016).
33. B. Dadonaite *et al.*, The structure of the influenza A virus genome. *Nat. Microbiol.* **4**, 1781–1789 (2019).
34. K. M. Kutchko *et al.*, Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Res.* **46**, 3657–3670 (2018).
35. P. Li *et al.*, Integrative analysis of Zika Virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe* **24**, 875–886.e5 (2018).
36. I. Manfredonia, D. Incarnato, Structure and regulation of coronavirus genomes: State-of-the-art and novel insights from SARS-CoV-2 studies. *Biochem. Soc. Trans.* **49**, 341–352 (2020).
37. A. M. Mustoe *et al.*, Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* **173**, 181–195 (2018).
38. L. Sun *et al.*, RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **26**, 322–330 (2019).
39. E. Pollom *et al.*, Comparison of SIV and HIV-1 Genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. *PLoS Pathog.* **9**, e1003294 (2013).
40. C. A. Lavender, R. J. Gorelick, K. M. Weeks, Structure-based alignment and consensus secondary structures for three HIV-related RNA genomes. *PLoS Comput. Biol.* **11**, e1004230 (2015).
41. A. M. Mustoe, M. Corley, A. Laederach, K. M. Weeks, Messenger RNA structure regulates translation initiation: A mechanism exploited from bacteria to humans. *Biochemistry* **57**, 3537–3539 (2018).
42. Y. Liu *et al.*, Structures and functions of the 3' untranslated regions of positive-sense single-stranded RNA viruses infecting humans and animals. *Front. Cell. Infect. Microbiol.* **10**, 453 (2020).
43. E. Lee *et al.*, Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. *Nat. Commun.* **12**, 2530 (2021).
44. M. J. Smola, G. M. Rice, S. Busan, N. A. Siegfried, K. M. Weeks, 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 (2015).
45. S. Busan, K. M. Weeks, Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**, 143–148 (2018).
46. S. Busan, K. M. Weeks, Visualization of RNA structure models within the Integrative Genomics Viewer. *RNA* **23**, 1012–1018 (2017).
47. K. Darty, A. Denise, Y. Ponty, VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
48. P. Stothard, The sequence manipulation suite: Javascript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**, 1102–1104 (2000).