Published online 19 January 2018

Nucleic Acids Research, 2018, Vol. 46, No. 7 ***3657–3670***
doi: 10.1093/nar/gky012

Structural divergence creates new functional features in alphavirus genomes

Katrina M. Kutchko[1,2,†], Emily A. Madden[3,†], Clayton Morrison[4], Kenneth S. Plante[4], Wes Sanders[3,5], Heather A. Vincent[3], Marta C. Cruz Cisneros[4], Kristin M. Long[4], Nathaniel J. Moorman[3,5,*], Mark T. Heise[3,4,*] and Alain Laederach[1,2,5,*]

[1]Department of Biology, UNC-Chapel Hill, USA, [2]Curriculum in Bioinformatics and Computational Biology, UNC-Chapel Hill, USA, [3]Department of Microbiology and Immunology, UNC-Chapel Hill, USA, [4]Department of Genetics, UNC-Chapel Hill, USA and [5]Lineberger Comprehensive Cancer Center, UNC-Chapel Hill, USA

Received July 21, 2017; Revised December 10, 2017; Editorial Decision December 24, 2017; Accepted January 05, 2018

ABSTRACT

Alphaviruses are mosquito-borne pathogens that cause human diseases ranging from debilitating arthritis to lethal encephalitis. Studies with Sindbis virus (SINV), which causes fever, rash, and arthralgia in humans, and Venezuelan equine encephalitis virus (VEEV), which causes encephalitis, have identified RNA structural elements that play key roles in replication and pathogenesis. However, a complete genomic structural profile has not been established for these viruses. We used the structural probing technique SHAPE-MaP to identify structured elements within the SINV and VEEV genomes. Our SHAPE-directed structural models recapitulate known RNA structures, while also identifying novel structural elements, including a new functional element in the nsP1 region of SINV whose disruption causes a defect in infectivity. Although RNA structural elements are important for multiple aspects of alphavirus biology, we found the majority of RNA structures were not conserved between SINV and VEEV. Our data suggest that alphavirus RNA genomes are highly divergent structurally despite similar genomic architecture and sequence conservation; still, RNA structural elements are critical to the viral life cycle. These findings reframe traditional assumptions about RNA structure and evolution: rather than structures being conserved, alphaviruses frequently evolve new structures that may shape interactions with host immune systems or co-evolve with viral proteins.

INTRODUCTION

Alphaviruses are mosquito-borne positive sense RNA viruses that infect a wide range of vertebrate and invertebrate hosts. Sindbis virus (SINV) is the prototype virus of the alphavirus genus and generally infects avian species (1,2). However, SINV can spill over into humans, where it causes symptoms such as fever, rash, myalgia, and arthralgia (3). In contrast, Venezuelan equine encephalitis virus (VEEV), which is normally spread in an enzootic cycle between rodents and mosquito vectors, can emerge in an epizootic form leading to large-scale epidemics associated with high mortality in equine species, and symptoms ranging from flu-like symptoms to severe encephalitis in humans (4). As such, these viruses can survive and replicate in a variety of vertebrate hosts and arthropod vectors, and both protein sequences and the structure of the RNA genome itself are important for the virus life cycle.

After alphavirus entry, the positive sense genome acts as a messenger RNA, leading to the translation of the viral nonstructural proteins, which mediate viral RNA synthesis and are encoded by the first two thirds of the viral genome (1). The positive sense genome is then transcribed to produce the viral negative strand RNA, which in turn serves as a template for the synthesis of both the positive sense viral genome, as well as a shorter 26S RNA encompassing the last third of the genome that encodes the viral structural proteins (1). Viral genomes contain a large amount of information in a limited amount of space; consequently, al-

*To whom correspondence should be addressed. Tel: +1 919 962 4565; Fax: +1 919 962 1625; Email: alain@unc.edu
Correspondence may also be addressed to Mark T. Heise. Email: mark_heisem@med.unc.edu
Correspondence may also be addressed to Nathaniel J. Moorman. Email: nathaniel_moorman@med.unc.edu
†These authors contributed equally to this work as first authors.
Present addresses:
Clayton Morrison, University of Mount Olive, Mt. Olive, NC, USA.
Kenneth S. Plante, World Reference Center for Emerging Viruses and Arboviruses, University of Texas Medical Branch, Galveston, TX, USA.
Kristin M. Long, North Carolina State Laboratory of Public Health, Raleigh, NC, USA.

© The Author(s) 2018. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

phavirus RNAs contain important regulatory structures in addition to the protein-coding sequence. These structures occur both in non-coding portions of the genome, such as the 5′ and 3′ UTRs (5–7), and in coding regions of the genome, such as the 51-nt conserved sequence element (5′ CSE) in nsP1, the packaging signal, and a frameshift signal in the structural polyprotein (8–11).

Despite the importance of these RNA structures to the virus life cycle, structures from one alphavirus are not necessarily compatible with other alphaviruses (7,9,12). This phenomenon raises that possibility that differences in viral RNA structure may contribute to the variation in host range, cell tropism, or disease pathogenesis between alphaviruses. However, to this point, the level of RNA structural conservation among alphaviruses has not been systematically characterized and in particular not at the whole genome level, with the most extensive analysis of structural conservation focused on the 5′ UTR and the 51-nt 5′ CSE (7,8,11).

Historically, covariation in a multiple sequence alignment of related RNAs served as the 'gold standard' method of identifying conserved (and, by extension, functional) RNA secondary structures (13–17). Base pairs that *covary*, or evolve together to preserve base pairing but not sequence identity, reveal the conservation of secondary structure within an RNA element. Covariation is most evident for RNAs having strong and highly conserved structures, such as transfer RNAs and ribosomal RNAs (14,17,18). For less-conserved RNAs, including those that are specific to multicellular organisms, the structural covariation signal is much weaker or non-existent (19,20). Many cellular RNAs including long non-coding RNAs do not exhibit any covariation (21). This finding raises the question of whether non-conserved RNA structural elements are functional, or whether their functions are derived solely from their sequence and not their structure.

In order to understand whether novel RNA elements have a functional role in the alphavirus life cycle, and to determine whether RNA structural elements are conserved within the alphavirus family, we used SHAPE-MaP (selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling) to determine the structural profile of the SINV and VEEV genomes. SHAPE-MaP has previously been used to identify structures in the HIV and hepatitis C virus genomes (22–24). Here, we used SHAPE-MaP to obtain high quality, high-resolution structural data for the complete SINV and VEEV genomic RNAs. This analysis found that aside from a few previously defined functional structures in SINV—in particular, the 5′ conserved sequence element and the frameshift element in the 6K coding region (7,10,25,26)—there seems to be little overall conservation of structured elements throughout the alphavirus family. Instead, the SINV and VEEV genomes contain a large number of highly structured regions that are unique to each virus and not shared between different alphavirus family members. Further analysis of one of these novel structures found that it affects SINV replication, indicating that these novel RNA structures are likely to play an important role in the viral lifecycle. These results suggest that alphaviruses utilize mutational space to evolve novel RNA structural elements specific to their individual biology, rather than preserving structures throughout virus evolution. Consequently, lack of conservation of structure does not indicate that a specific conformation lacks a function; instead, lack of structural conservation suggests that the role of RNA structure is highly context-dependent.

MATERIALS AND METHODS

SHAPE data collection

Sindbis virus (Girdwood strain; accession #MF459683) and VEEV (ZPC783 strain; accession #MF459684) virions were concentrated by ultracentrifugation over a 20% sucrose cushion. Concentrated virions were lysed with TRIzol (Ambion) and full-length genomic RNA was purified following the manufacturer's protocol.

Modified RNA was obtained by incubation of 2 μg of total RNA at 37°C for 15 min in the presence of 10 mM $MgCl_2$ and 111 mM KCl, then treated with 100 nM of 1-methyl-7-nitroisatoicanhydride (1M7) for 5 min at 37°C. Negative control RNA was obtained by incubation of 2 μg total RNA at 37°C for 15 min, then incubated with 5 μl DMSO for 5 min at 37°C. Denatured control RNA was obtained by incubation of 2 μg total RNA at 95°C for 2 min, then treated with 100 nM 1M7 for 2 min at 95°C. Following treatment, RNA was purified using illustra MicroSpin G-50 columns (GE Healthcare). Total purified RNAs were then incubated with 500 ng Random Primer 9 (NEB) at 65°C for 5 min, cooled to 0°C, and mixed with 10 mM dNTPs, 0.1 M DTT, 500 mM Tris pH 8.0, 750 KCl and 500 mM $MnCl_2$. The mix was then incubated at 42°C for 2 min, followed by the addition of 200 units of SuperScript II (Thermo Fisher Scientific) and a further incubation at 42°C for 180 min, heat inactivated at 70°C for 15 min, and then purified using illustra MicroSpin G-50 columns. Double stranded cDNA was created by NEBNext mRNA Second Strand Synthesis Module (NEB) using the manufacturer's protocol. The double stranded cDNA was then fragmented, tagged, amplified and barcoded using Nextera XT DNA Library Preparation Kit (Illumina) following the manufacturer's directions. Libraries were cleaned with Agencourt AMPure XP beads (BeckmanCoulter) at a DNA to bead ratio of 0.6:1, library size was determined by 2100 Bioanalyzer (Agilent Technologies) and quantified with a Qubit Flourometer using Qubit dsDNA HS Assay Kit (ThermoFisher Scientific). Sequencing was performed on a MiSeq Desktop Sequencer (Illumina).

Sequencing reads from the background SHAPE-MaP condition were used to assemble the correct virus sequences. SHAPE reactivities for the CHIKV, SINV and VEEV viruses were derived using the ShapeMapper pipeline (22). Because the background mutation rate for VEEV was higher than the CHIKV background mutation rate, SHAPE reactivities for VEEV were re-calculated using a scaled background mutation rate and 2%-8% normalization (27). The SHAPE data for SINV and VEEV are available as supplementary data in SNRNASM format (Supplementary Data 1), as well as online at https://docs.google.com/spreadsheets/d/1BHORlaXPbC0npQK-zy4YEE938qzZOrsp9L6FmouWN3U/edit?usp=sharing (28). Windowed SHAPE values were calculated by finding the median SHAPE reactivity over a rolling 55-nt window

and comparing those values to the global median SHAPE (22,29,30).

Multiple sequence alignment

The alignment was built on the conserved protein-coding sequence alignment from (31). The full nucleotide sequence for each virus in the alignment was downloaded from GenBank. Only viruses with complete genome sequences were included in the final alignment, and the assembled sequences for SINV, CHIKV and VEEV were added to the alignment. Non-conserved portions of the genome (5′ UTR, 3′ UTR, C-terminus of nsP3, and non-coding junction and N-terminus of the capsid sequence) were aligned with MAFFT (32,33) (v7.221) and manually refined. The non-conserved and non-coding alignments were concatenated to create the final multiple sequence alignment. The phylogenetic tree was created using PhyML (34) (v3.0) with default parameters and midpoint-rooted.

Sequence conservation

The sequence conservation score $C(x)$, ranging from 0 to 1, at each alignment position x was computed using the following equation, adapted from (35):

$$C(x) = (1 - t(x))^{0.5} \cdot (1 - g(x))^{1} \quad (1)$$

where $g(x)$ is the frequency of gaps at position x, and $t(x)$ is the Shannon entropy (36) at position x. The sequences were weighted using the algorithm in (37), and the weights were incorporated as in (35).

Correlations in SHAPE data

To compare the correlation between two sets of SHAPE data, thereby enabling comparisons across the entire genome while avoiding distortions by outliers, all gaps and missing and negative values were set to zero before calculating the Pearson correlation coefficient over a 55-nt rolling window. For the correlation between SINV and VEEV, the SHAPE reactivities were aligned according to the multiple sequence alignment (with all-gap positions removed). To generate the background distribution, the SINV and VEEV reactivities from the previous analysis were each scrambled prior to the rolling correlation coefficient calculation. The SHAPE correlation distribution for biological replicates was generated using the SHAPE data for the first 11 400 nt of the SHAPE-MaP data for two biological replicates of CHIKV. Regions in SINV and VEEV with correlation coefficients in the top 99th percentile were expanded 27 nt on each side to incorporate all nucleotides within the window, for a total of nine highly correlated regions.

Identification of structured regions

RNASurface (38) (v1.0) was used to find regions of significant structure in the SINV genome, with a minimum z-score threshold of –2.5. Overlapping regions were merged, leading to 20 distinct predicted structured regions. In 17 of those regions, the majority of positions had below average windowed median SHAPE reactivities. These 17 regions are the final set of structured regions in the SINV genome, supported by both prediction and experimental data.

Structure modeling

Minimum free energy models for each structured region were generated using RNAstructure's *Fold* program (39) (v5.8.1), incorporating SHAPE reactivities as a pseudo-free energy term (40), with the maximum base pairing distance set at 200 nt and standard parameters (intercept = −0.6 kcal/mol, slope = 1.8 kcal/mol, temperature = 310.15 K) otherwise.

To model the whole genome structure, the Superfold program was used with SHAPE reactivities incorporated as a pseudo-free energy term (22), with a maximum base pairing distance of 500 nt and standard parameters otherwise. This whole-genome structural model was used to obtain the Shannon entropies at each position from the base pairing probabilities (41,42). Structures within regions with highly correlated SHAPE data were extracted from this whole-genome structural model, with long-range base pairs removed.

Creation of mutant clones

To generate mutations for each region while keeping the amino acid sequence unchanged, the program CodonShuffle (43) was used with the dn231 algorithm, which scrambles sets of trinucleotides while ensuring that the first and third bases of each trinucleotide set are preserved. This method also preserves sequence composition and dinucleotide frequency. For each region, 1000 shuffled sequences were randomly generated, in most cases representing hundreds of unique sequences.

Out of these shuffled sequences, mutant sequences were selected to maximize structural disruption while also avoiding large changes in codon usage. Because the virus must survive in multiple hosts, organism-specific measures such as the codon adaptive index (44) are not useful to quantify change in codon usage. Instead, codon usage change was calculated using the sum of square differences of codon frequencies within the virus transcript. Structural disruption was determined by calculating the percentage of base pairs in the SHAPE-directed structural model that could no longer form Watson–Crick or wobble base pairs in the mutant sequence, as well as confirming that the predicted structure of the mutant was not similar to the structural model for wildtype.

Structure mutants were designed from the Girdwood S.A. cDNA clone (pg100) of SINV and created by Gibson assembly (New England BioLabs). Fragments of the Girdwood genome containing structure disrupting mutations were purchased from Integrated DNA Technologies (IDT, Iowa, USA) with ∼23 bp of overhang beyond restriction endonuclease cut sites. Clones were confirmed by Sanger sequencing through UNC sequencing core.

Infectious RNA was transcribed from the cDNA clones after linearization by NotI using mMESSAGE mMACHINE SP6 Transcription Kit (Invitrogen). RNA was introduced to BHK-21 cells by electroporation. Supernatants were collected 24–48 h after electroporation based on observed cytopathic effects and aliquoted into single use

aliquots stored at –80°C. Virus titer was quantified by plaque assay on Vero81 cells.

Cell culture

BHK-21 cells were maintained in 1× aMEM (Gibco) supplemented with 10% heat inactivated fetal bovine serum (FBS) (Sigma) and 0.2 mM L-glutamine (Gibco). Vero81 cells were maintained in 1× DMEM (Gibco) supplemented with 10% FBS and 0.2 mM L-glutamine. NIH-3T3 cells were maintained in 1× DMEM supplemented with 10% bovine calf serum (Colorado Serum Co., Denver, USA). Mosquito C6/36 cells were maintained in 1× Leibovitz L-15 (Corning/Cellgro) supplemented with 10% FBS, 10% tryptose phosphate broth (Sigma) and 0.2 mM L-glutamine.

Viral growth and plaque assays

Multistep growth curves were conducted by infecting cells at a multiplicity of infection (MOI) equal to 0.01 in biological triplicate. Sample of cell culture supernatants were taken at indicated times after infection and stored at –80°C. Viral titer was quantified by plaque assay. During plaque assays, Vero81 monolayers were infected with virus samples titrated in 1× PBS (Gibco) with 1% FBS and Ca^{2+}/Mg^{2+} and overlaid with 1× aMEM with 10% FBS, 0.2 mM L-glutamine, 1 mM HEPES (Corning), 1% penicillin streptomycin (Gibco) and 1.25% carboxymethylcellulose sodium (CMC) (Sigma). Virus was allowed to plaque for 40 h before cells were fixed with 4% paraformaldehyde (PFA) (Sigma), washed, and stained with 0.25% crystal violet (Fisher Chemical).

Specific infectivity assays

Wildtype and mutant RNA were electroporated into BHK-21 cells in parallel. An aliquot of electroporated cells were titrated and plated overtop subconfluent Vero81 cell monolayer. BHK-21 cells were allowed to attach for 1.5 h, at which point the monolayers were overlaid with CMC overlay detailed previously and incubated for 40 h. After incubation, cells were fixed, washed and stained as detailed for plaque assays.

RNA stability and transcription assays

Full length genomic RNA from the wildtype and mutant viruses was produced using the SP6 DNA dependent RNA polymerase (Ambion) as described above. Wildtype and mutant RNA were electroporated into BHK-21 cells in biological duplicate or quadruplicate as described above, at which point the cells were washed 1× with media to remove remaining extracellular RNAs. To test genomic RNA stability, half the cells were treated with cyclohexamide to prevent translation of the viral nonstructural proteins, which mediate viral RNA synthesis, thereby preventing replication of virus RNA. Cells used to analyze viral genome transcription kinetics were not treated with cyclohexamide. For both experiments, cells were plated and RNA harvested at the indicated times. Monolayers were washed 1× with PBS before cells were lysed in TRIzol reagent. Total RNA was purified following manufacturer's protocol, and treated with DNAse (Promega) to remove any input plasmid from the initial transcription reaction. Virus genome and 18S copy number was quantified by qRT-PCR using iTaq Universal Probes One-Step Kit (Bio-Rad) alongside a standard for absolute quantitation. SINV primer-probe sets were used as described previously, and 18S primer probe was purchased ThermoFisher (Catalog #4331182) (45). We also performed PCR amplification on samples without reverse transcriptase to test for carryover plasmid contamination within the electroporated RNA stocks. After quantitation of genome and 18S copies, SINV genomes were normalized to 1×10^6 18S copies and log transformed.

Western blots

Wildtype and mutant RNA were electroporated into BHK-21 cells in biological duplicate as described above. Cells were washed 1× with media to remove remaining extracellular RNAs and plated. Cell lysates were harvested at indicated time points with RIPA buffer containing protease inhibitor at indicated times. Lysates were prepared by incubation on ice for 30 min followed by centrifugation at 12 000 rpm for 15 min. Total protein was quantified by BCA Pierce assay (ThermoFisher) using a BSA standard curve. Equal amounts of protein were boiled in SDS loading buffer for 5 min. Samples were run on a 4–15% gradient Mini-PROTEAN TGX Precast Protein gel (BioRad) and transferred to PVDF membrane (Li-Cor) with the BioRad Wet Transfer system. Membranes were blocked in 1× TBS–0.1% tween-20 (TBST) and 5% milk overnight at 4°C. Primary antibodies were diluted in TBST + 5% milk (1:2000 rabbit polyclonal anti-nsP3; 1:1000 goat anti-actin, Santa Cruz) (46). Membranes were washed 3× in 1× TBST while rocking at room temperature for 10 min before incubation with secondary antibody (1:10 000) IRDye conjugated mouse anti-rabbit for nsP3 and IRDye conjugated anti-goat for actin (Li-Cor). Secondary antibodies were incubated for 1 h at room temperature in 5% milk with 0.01% SDS in 1× TBST on a rocker. Membranes were washed 3× with 1× TBST for ten minutes each wash. The membranes were then washed 3× in 1× TBS for 10 min each time. The membranes were visualized with the Odyssey infrared Imaging system (Li-Cor).

Structural conservation

The structure compatibility (*SC*) score of a structure within a given, related sequence (x) was defined as follows:

$$SC = \frac{bp_x}{bp_{orig}} \quad (2)$$

specifically, the fraction of base pairs (*bp*) in the SINV structure (*orig*) that can form in the related sequence, using the multiple sequence alignment to identify the locations of homologous base pairs.

To search for homologous structures, we used the Infernal software suite (v1.1.1) (47,48). For each SINV structured region, the sequence and the minimum free energy structure were used to create an alignment in Stockholm format (with long-range base pairs in the packaging signal removed). A covariance model was built and calibrated

using *cmbuild* and *cmcalibrate* for each region. Hits in homologous alphaviruses were found using *cmsearch* with the model on the sequences in the multiple sequence alignment, and those hits were assembled into a new alignment with *cmalign*, to create a structure-informed alignment for each region of interest.

The R-scape program (v0.3.2) was used to identify base pairs with significant covariance in each structure-informed alignment >50 nucleotides (21), and applied to conserved RNA alignments from the same report. Average percent sequence identities were calculated with the *alistat* program, part of the HMMER package (49).

RESULTS

Structure conservation and divergence identified by high-resolution SHAPE profiling

RNA structural elements play essential roles in many aspects of the alphavirus lifecycle, including regulation of viral RNA synthesis, viral translation, and evasion of the host innate immune system. However, most of this analysis has focused on a fairly limited number of RNA structures, such as the 51-nt conserved sequence element (5′ CSE) and the 414-nt packaging signal located in the nsP1 coding sequence (9,25). To locate additional functional structures within the SINV genome, we used SHAPE-MaP (22) to obtain a high-resolution structural profile for the entire SINV genome (Figure 1A). We performed these experiments on refolded viral RNA in the absence of viral and cellular proteins. Highly structured regions that are likely to fold into a single, unique conformation have below-average median SHAPE reactivities (22,50).

We also determined the sequence conservation score at each position using a multiple sequence alignment containing 37 alphaviruses representing the entire alphavirus phylogeny (31). The sequence conservation score uses sequence identity and gappiness to calculate the conservation at each position (35) (Equation 1). Although most of the protein-coding portion of the genome is highly conserved, highly divergent regions occur in both protein-coding and non-coding sections (Figure 1A; light gray at the 5′ end, end of nsP3, non-coding junction, and 3′ end).

We also used the intersection of sequence features and SHAPE reactivities to find regions of highly stable structure within the SINV genome, as well as average structural entropy across the genome (Figure 1B). We identified 17 'structured regions' with low median SHAPE reactivities and high structural significance based on the z-score using RNAsurface (38), which compares the free energy for a region to what would be expected if the sequence were shuffled at random (51). For each of these structured regions, we used SHAPE reactivities to guide secondary structure prediction to derive a structural model (40,52) (Figure 1C). Our method successfully recapitulates known or previously predicted structured regions, including regions overlapping the 5′ CSE, the packaging signal, the non-coding junction, and the frameshift signal (10,53), confirming the utility of the SHAPE-MaP method for accurately predicting RNA secondary structures within viral RNA genomes. In addition to identifying several previously characterized areas of RNA secondary structure, SHAPE-MaP analysis also identified several regions of the genome that contain previously uncharacterized stable RNA structures (Figure 1C). Therefore, novel areas of RNA secondary structure are broadly distributed throughout the SINV genome, which raises the possibility that these structures play as yet undefined roles in the alphavirus lifecycle.

Previous work has found that structures such as the 5′ CSE and the RNA packaging signal are conserved between two or more alphavirus family members. Given the large number of stable RNA structures present in the SINV genome, we wanted to determine whether any of these novel structured regions were conserved between alphaviruses. A subset of RNA structural elements, such as the 5′ CSE and the RNA packaging signal, are known to have structural conservation between SINV and Venezuelan equine encephalitis virus (VEEV) (8,9). Therefore, we also performed SHAPE-MaP on the ZPC738 strain of VEEV (Supplementary Figure S1A).

To identify stable structures within the VEEV genome, we applied the same analysis used for SINV (Supplementary Figure S1B). Similar to SINV, this analysis identified several highly structured regions within the VEEV genome that overlap with the 5′ CSE, packaging signal, and frameshift signal, as well as several regions with high structural stability that have not previously been characterized in VEEV (Supplementary Figure S1C). Therefore, similar to SINV, this data suggests that the VEEV genome contains previously uncharacterized RNA structural motifs that may play an important role in VEEV replication and pathogenesis, while also providing us with an opportunity to directly assess whether stable structures are conserved between two different alphaviruses.

We used the SINV and VEEV SHAPE data to look at the correlation of SHAPE reactivities between the related genomes to assess conservation of RNA structure (29) (Figure 1D). This analysis identified nine regions in the genome with correlation coefficients that surpassed the 99th percentile, including the region covering the 51 nt 5′ CSE. However, when compared with correlations between biological replicates, the correlation distribution between SINV and VEEV SHAPE data overlaps minimally. The distribution is more similar, but not identical, to a random distribution, representing a limited amount of structural conservation (Figure 1E). Therefore, while a small number of regions are correlated in their SHAPE signal, which could indicate structural conservation, the vast majority of highly structured regions in both the SINV and VEEV genome are unique and not shared between the two viruses. We also compared the structures of SINV and VEEV that overlap these highest-correlated regions in the 99th percentile (Supplementary Figure S2A). While the patterns of SHAPE data are similar between SINV and VEEV in these regions, that effect is the result of similar but not conserved base-pairing patterns; only the 5′ CSE and the frameshift region adopt similar structures between the viruses (Supplementary Figure S2B). Therefore, even within regions of relatively high SHAPE correlation, structures generally diverge between SINV and VEEV.

Despite the general lack of structure conservation between SINV and VEEV, one of the most correlated regions

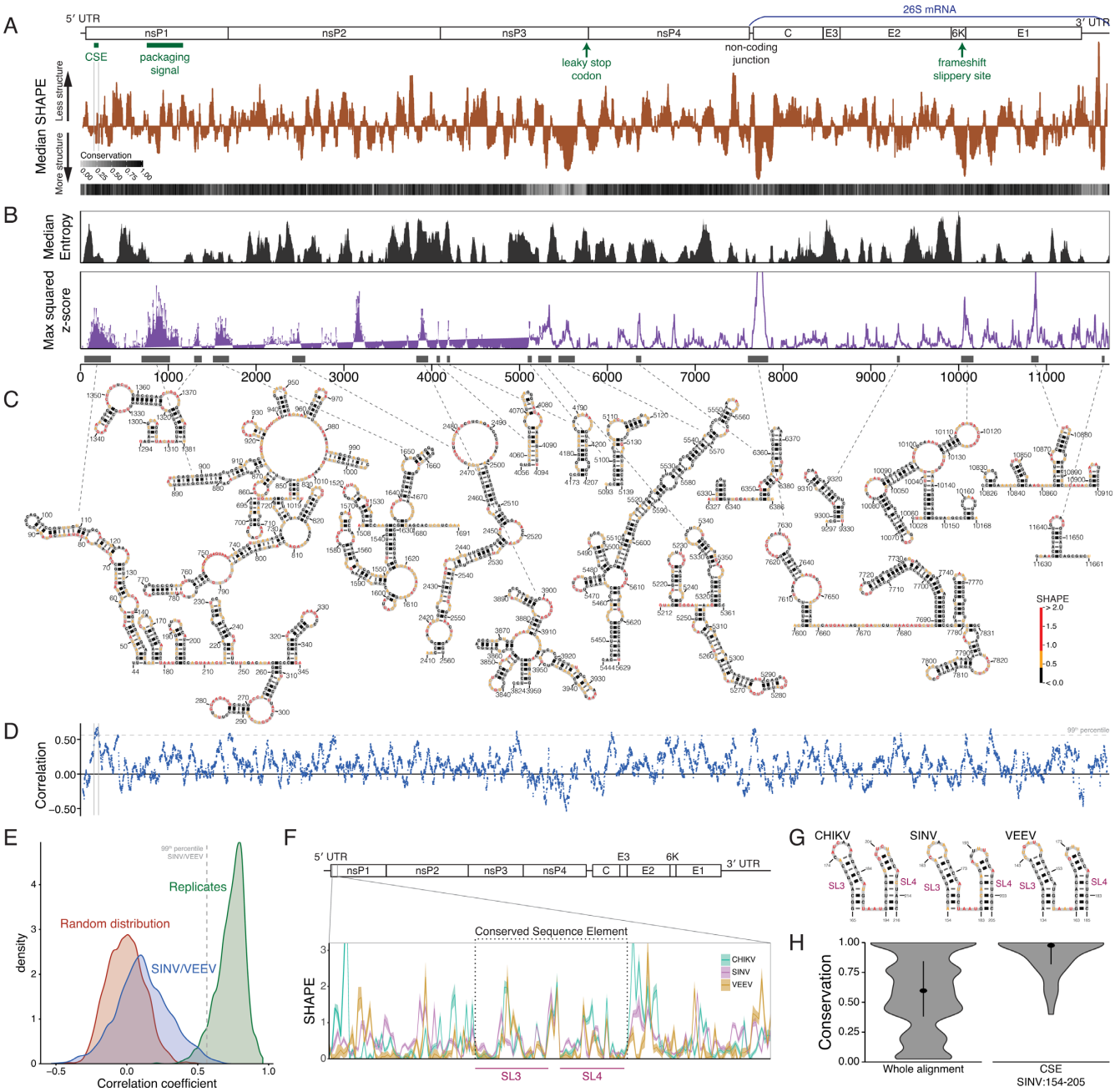

Figure 1. The Sindbis virus genome contains a multitude of diverse RNA structures. (**A**) Top: schematic of the virus genome organization, with annotated elements. Middle: SHAPE data for the Sindbis virus genome, represented by the local median (55-nt window) compared with the global median. Reactivities below the x-axis indicate a region more structured than average. Gray lines denote the conserved sequence element (5′ CSE), which has low SHAPE reactivities and is highly structured. Bottom: sequence conservation at each position, based on sequence identity and gappiness, from a multiple sequence alignment of 37 alphaviruses. The protein-coding sequence contains both well-conserved (black and dark gray) and less-conserved (light gray) regions. (**B**) Top: median (55-nt window) Shannon entropies of base pairing across the SINV genome. Middle: Maximum squared *z*-score at each position in the genome, with higher values corresponding to greater structural significance. Bottom: structured regions in the SINV genome, based on the intersection of regions with low SHAPE and low *z*-scores. (**C**) SHAPE-directed structural models of SINV structured regions. Nucleotide color indicates low, medium, or high SHAPE reactivity. (**D**) Windowed correlation coefficients of SHAPE data between the SINV and VEEV genomes. The dashed line indicates the top 1% of correlation coefficients. SHAPE data within the 5′ CSE are among the most correlated within the genome, indicating high structural conservation within that region. (**E**) Distribution of windowed correlation coefficients of SHAPE data. Red: a background distribution, blue: correlation coefficients between SINV and VEEV, green: correlation coefficients of two biological replicates of a virus. Although SINV and VEEV are more correlated than expected at random, there is little overlap with the correlations of the same virus, indicating little widespread correlation. Dashed line indicates top 1% of SHAPE correlations between SINV and VEEV. (**F**) SHAPE data of the 5′ CSE in CHIKV, SINV, and VEEV. Within the 5′ CSE, the SHAPE profiles are very similar, representing conservation of structure, but the correlation immediately disappears outside of the 5′ CSE. (**G**) SHAPE-directed structural models of the CSE in CHIKV, SINV and VEEV. The 5′ CSE structure is compatible with the SHAPE data and conserved in all three viruses. (**H**) Distribution of alignment-derived sequence conservation scores in the entire alignment (left) and 5′ CSE only (right). Dot indicates the median, with the line extending from the 25th to 75th percentile.

is the CSE, which shows sequence conservation across multiple alphavirus family members (8). We therefore compared the SHAPE reactivities in that region between SINV, VEEV, and Chikungunya virus (CHIKV) (Figure 1F). The 5′ CSE is highly correlated within the two conserved stem loops 3 and 4 (8), but the SHAPE reactivities are not similar outside of the conserved region. The similarity in SHAPE derives from the identical structures of stem loops 3 and 4 (Figure 1G). In addition, sequence conservation within the 5′ CSE is remarkably high, especially when compared with the alphavirus family as a whole, based on sequence conservation scores using the multiple sequence alignment (Figure 1H). Outside the 5′ CSE, however, slight divergence in sequence results in little conservation of structure. Thus, only very local structural motifs like the two hairpins in the 5′ CSE are structurally conserved.

RNA structure plays a critical role in SINV replication

Given the highly structured nature of both the SINV and VEEV RNA genomes, we set out to test the functional impact of a subset of these structures. To determine whether an RNA structure is functional, it is necessary to disrupt the structure without changing other aspects of the sequence, such as the encoded amino acid sequence. We used the program CodonShuffle to create mutant RNA sequences that preserve the amino acid sequence (43). The algorithm we used shuffles sets of trinucleotides (not in reading frame) in which the first and third bases remain identical, and the second base only changes when it would not affect the protein sequence (Figure 2A). This method also preserves sequence composition and dinucleotide frequency. For most RNA sequences, this method generates hundreds of possible sequence mutants, so we chose mutation strategies designed to maximize disruption of the structural model by changing base pairing in a particular region with only very minor changes in codon usage (Supplementary Table S1). The frequencies of each codon in the mutant viruses remain nearly identical to WT, differing in usage by one or two instances at most for each codon. In addition, although the effects of small changes in codon usage have not been studied in alphaviruses, slight changes in codon usage have not been shown to affect viral growth in polioviruses (54,55).

To validate our structure-disrupting method, we mutated two known SINV RNA structures, the 5′ CSE and the packaging signal (Figure 2B). To disrupt the 5′ CSE, we created mutations both within the element and within the long hairpin immediately 5′ of the element, creating twenty mutations throughout the region. Prior studies have suggested that the 5′ CSE structure has a mild impact on growth in mammalian cells but necessary in mosquito cells (8). Consistent with these prior results, we found that disruption of the 5′ CSE with the 5′ hairpin resulted in decreased viral growth in C6/36 mosquito cells compared to the wildtype virus (Supplementary Figure S2). We also found that disruption of the 5′ CSE and the 5′ hairpin resulted in a significant growth defect in mammalian cells compared to the wildtype virus (Figure 2C). Therefore, these results suggest that the 5′ hairpin and 5′ CSE broadly impact viral replication in both mammalian and mosquito cells.

The viral RNA packaging signal, which falls within a region spanning nucleotides 613 to 1019, is less sequence conserved and longer than the 5′ CSE. Therefore, we introduced a total of sixty-nine mutations into the packaging signal that were designed to maximize disruption of the RNA secondary structure, while avoiding impact on the protein sequence (Figure 2B). Consistent with prior results, disrupting the RNA packaging signal resulted in a significant decrease in virus yield (Figure 2C), thereby validating both the SHAPE-MaP-derived structural predictions, while demonstrating our ability to directly test the functional importance of stable RNA secondary structures within the SINV genome.

Novel RNA structures in the SINV genome

Next, we applied our method to two structured regions with low SHAPE reactivity: one downstream of the packaging signal in nsP1, and one extremely low-SHAPE region in the non-conserved domain of nsP3 (Figure 3A). The nsP1 structured region (nsP1 SR) is conserved, but less so than the packaging signal. The novel nsP3 structured region (nsP3 SR) has little sequence conservation, and is located 133 nucleotides upstream of the leaky stop codon. We disrupted the nsP1 SR and nsP3 SR with 6 and 36 point mutations, respectively.

As with the previously tested regions, we infected Vero cells, a mammalian cell line, with the structural mutants (Figure 3B, left). In Vero cells, the mutants grew at the same rate as the wildtype virus. We also infected NIH/3T3 cells which, unlike Vero cells, have a competent interferon system (56) (Figure 3B, right). Neither structural mutant had a change in phenotype in NIH/3T3 cells, indicating that these structures are not necessary for viral growth in mammalian cells. Additionally, however, we measured the specific infectivity of *in vitro* transcribed genomic RNA of the packaging signal, nsP1 SR and nsP3 SR mutants by electroporating the RNA into BHK-21 cells and plating serial dilutions of cells over a Vero cell monolayer (Figure 3C, Supplementary Figure S3A), an assay which measures the gross ability of naked viral RNA to produce infectious virus and detects early defects in viral fitness. We used equal amounts of RNA for the specific infectivity assay (Supplementary Figure S3B). The packaging signal and nsP3 SR mutants had the same specific infectivity as wildtype virus, whereas the specific infectivity of nsP1 SR was reduced by three to four orders of magnitude, indicating that nsP1 SR is critical for the virus. The absence of phenotypic change between the nsP3 SR mutant, which spans almost 200 nt of coding sequence, and WT virus confirms that merely changing the RNA sequence alone is not enough to disrupt the virus life cycle.

Curiously, mutated nsP1 SR overcame its infectivity defect in the viral growth assays. Sequencing of the rescued virus did not reveal any compensatory mutations to restore structure or otherwise regain function.

We considered the possibility that disruption of nsP1 SR results in destabilization of the viral RNA, so we assessed transcription and stability of the input RNA after electroporation (Supplementary Figure S3C). We saw no differences in the stability of nsP1 SR mutant RNA compared to

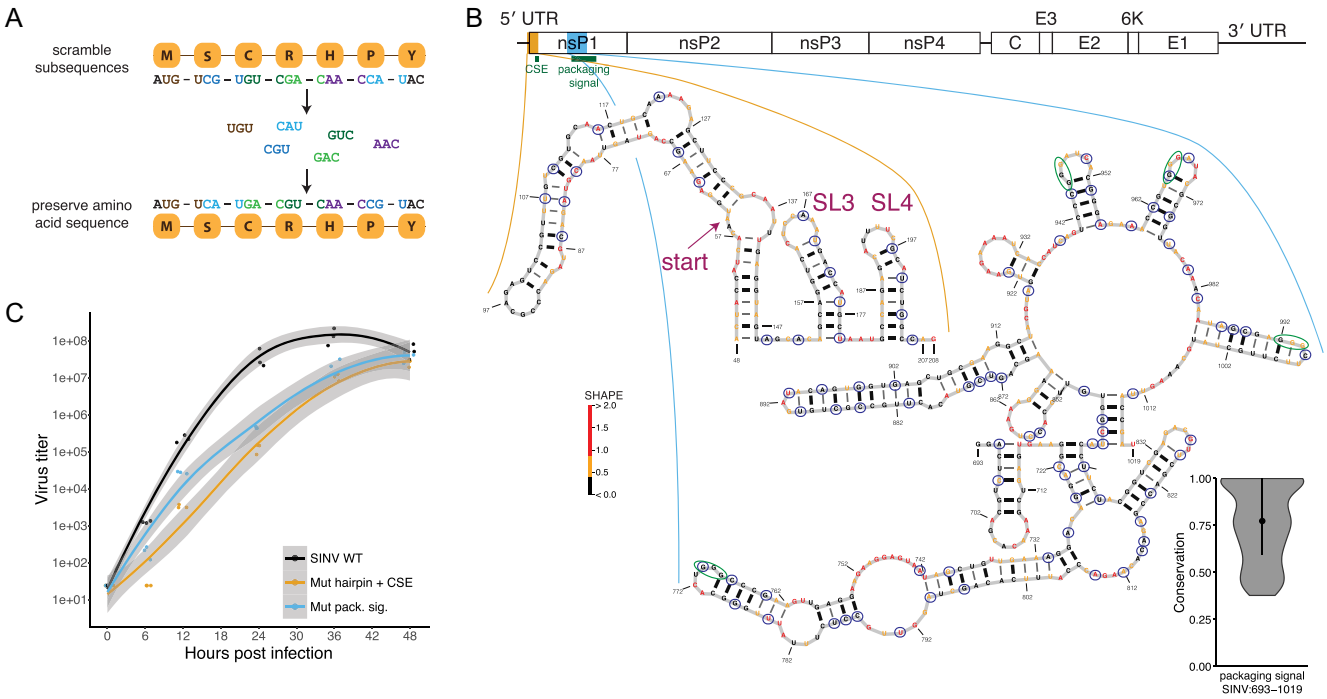

Figure 2. Structure-disrupting mutations successfully confirm function by impeding virus growth. (**A**) Method used to disrupt RNA structure. Trinucleotide sets are shuffled, changing the nucleotide sequence while the amino acid sequence is preserved. (**B**) Left: structure of the hairpin + CSE element. Start codon and stem loops 3 and 4 are indicated. Right: structure of region overlapping packaging signal. Blue circles indicate positions that are mutated to disrupt the structure. Green circles indicate previously observed GGG motifs within packaging signal structure (9). Nucleotide color represents SHAPE reactivity. Violin plot displays conservation scores within the packaging signal region. (**C**) Growth curves for SINV WT (black), mutated hairpin + CSE (gold), and packaging signal (blue) in Vero81 cells at a MOI of 0.01. Shading indicates standard error. Both structures are necessary for optimal virus growth.

WT RNA. This finding, along with the nsP1 SR mutant's ability to recover from its initial defect, suggests that the infectivity defect of the nsP1 SR mutant is caused by disruptions of events associated with genome replication early during infection. We found that levels of viral RNA synthesis (Figure 3D) and nonstructural protein expression (Figure 3E) were reduced over time compared to the WT virus. These results could be due to either defects in early nonstructural protein synthesis or transcription of the genome itself. Overall, our findings suggest that nsP1 SR plays an important role in regulating early stages of the viral replication process.

Functional RNA structures are not conserved

Having confirmed that at least two known RNA structures and one novel RNA structure in the SINV genome are functional, we wanted to assess their level of conservation in related alphaviruses. For each structure, we compared the SINV model to the sequences in the thirty-six related alphaviruses (Figure 4A). The structure compatibility score represents the fraction of structural model base pairs that can form at homologous positions in each virus (Equation 2).

The 5′ CSE has a high structural compatibility score in nearly every other alphavirus, indicating that the two 5′ CSE stem–loops are conserved throughout the alphavirus family. The packaging signal has less structural conservation compared with the 5′ CSE, with strong conservation only within a small branch of the alphavirus phylogeny. The nsP1 SR follows a similar pattern, but it is even less conserved than the packaging signal. The nsP3 SR, in a non-conserved region of the SINV genome, has no structure conservation outside of immediate relatives. These results indicate that functional structures are not necessarily conserved, and, in fact, they are in this case unique to an individual virus.

We also considered the possibility that functional structures may have shifted locations in other alphaviruses and would therefore not appear to be conserved based on the multiple sequence alignment. For each of the structured regions we identified (Figure 1C), we constructed a stochastic context free grammar (SCFG) model from the predicted SHAPE-informed secondary structure and corresponding sequence (15,47). We used this model to search through the related alphavirus sequences and refined our alignments in an attempt to build a covariance models for the SINV structures (57,58). Only three regions, including the hairpin + 5′ CSE and the packaging signal, exist in almost all related alphaviruses (Supplementary Table S2).

We quantified structural covariation of these structure-informed alignments with the new program R-scape (21), which identifies base pairs having significant covariation. Because R-scape applies a significance threshold to an alignment, it filters base pairs that may appear to be conserved but do not covary more than expected by chance. We calculated the percentage of base pairs in each covariance model that R-scape found to be significant, both in the structure-informed alignments and in known conserved

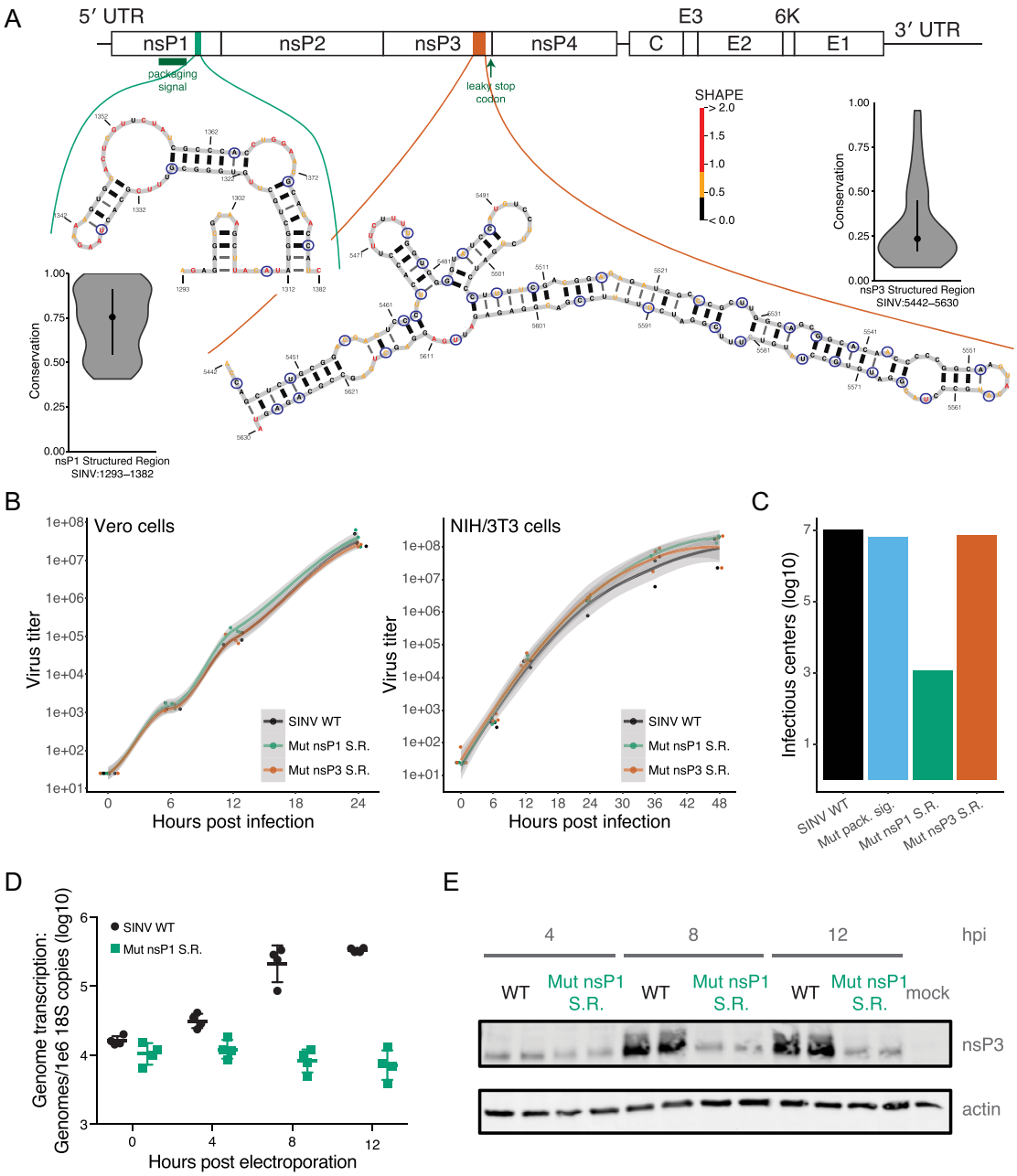

Figure 3. Novel virus structures tune SINV growth. (**A**) Structures for the new nsP1 structured region (nsP1 SR; left) and the new nsP3 structured region (nsP3 SR; right). Nucleotide color represents SHAPE reactivity. Violin plots display sequence conservation scores within each region. Blue circles indicate positions that are mutated. (**B**) Growth curves for SINV WT (black), the nsP1 SR (green), and the nsP3 SR (red). Mutant growth is nearly identical to WT in both Vero cells (left) and NIH/3T3 cells (right). (**C**) Specific infectivity of mutant viruses. The nsP1 SR mutant has a large defect in infectivity. Graph is a representative experiment of three or more replicates. (**D**) Genome transcription levels of WT and nsP1 SR mutants, measured by qRT-PCR. The nsP1 SR mutant has a defect for genome transcription. (**E**) RNA translation of WT and nsP1 SR. Expression of the nonstructural proteins is impaired in nsP1 SR compared to WT as measured by probing for nsP3.

RNA structures (Figure 4B; Supplementary Table S3). Although the structure-informed alignment of the long hairpin + 5′ CSE in nsP1 contains the highest number of significantly covarying base pairs among SINV regions, the sensitivity of these covariation models is well below classic structured RNAs such as riboswitches, tRNA, and Rnase P (21) (Figure 4B; Supplementary Table S3). Importantly, the nsP1 SR, whose disruption dramatically decreases specific infectivity (Figure 3C) has no covarying base-pairs.

The only other regions in SINV where R-scape found more than one significantly covarying base pair are the regions overlapping the 26S promoter (SINV:7600–7831) and the region containing the frameshift element (SINV:10028–10168). The two covarying base pairs in the frameshift element are contiguous and part of a stem loop found in

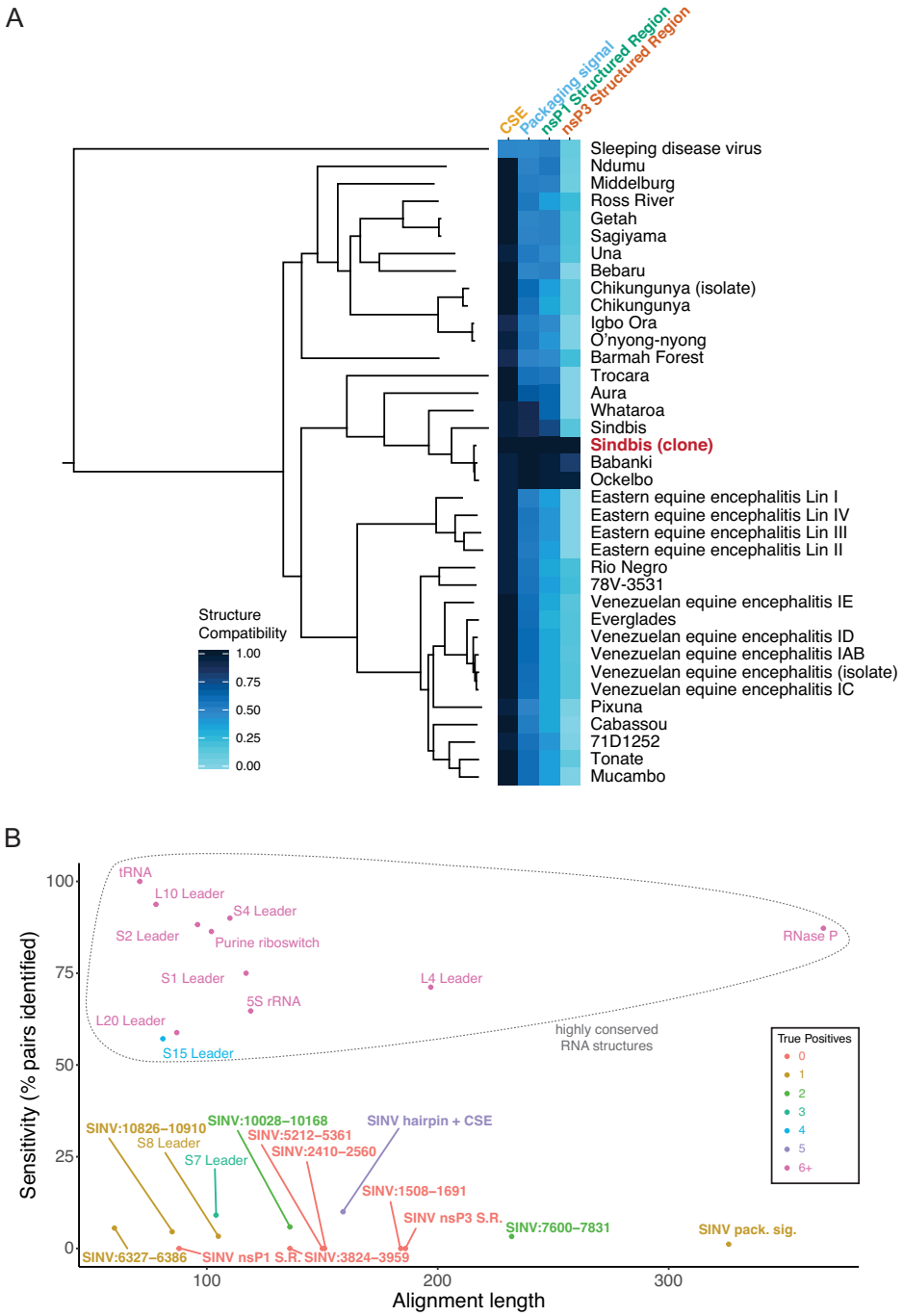

Figure 4. Outside of the conserved sequence element, SINV functional structures are not conserved. (**A**) Structure compatibility scores for the four tested sequences, with the phylogenetic tree on the left. The SINV strain used for the reference sequence and structure is bolded and in red. The structural model for the 5′ CSE is highly compatible with other alphavirus sequences, but the other functional structures are less conserved. For the new region in nsP3, the structure essentially does not exist outside of closely related strains. (**B**) R-scape results for structure-informed alignments. X-axis: alignment length; y-axis: percentage of base pairs in structure found by R-scape; color: number of true positives found by R-scape. Bold typeface indicates R-scape results from SINV structure-informed alignments, whereas regular typeface indicates R-scape results for alignments from (21). While no SINV region has anywhere near the amount of covariation of highly-conserved RNA structures, the region containing the 5′ CSE is notable for the highest number of covarying base pairs within SINV.

the New World clade of alphaviruses, including VEEV (10) (Supplementary Figure S5). Elements similar to this hairpin exist in 30 out of 37 alphavirus sequences in our alignment, so although it is highly conserved, it lacks the complete structural conservation of the 51-nt 5′ CSE (Supplementary Table S2).

Consequently, although there is slight covariation in SINV within the 5′ CSE, most other regions including functional RNA structural elements have little to no covariation. The lack of structural covariation indicates that sequence, not structure, is the primary driver of similarity in our covariance models. Despite RNA structural elements being important for the growth of SINV, there is no covariation evidence to indicate that in general these structures are conserved among alphaviruses. Thus, these results suggest that alphaviruses evolve idiosyncratic, functional RNA structures specific to their individual biology. Furthermore, these results demonstrate the difficulty in using covariation as a signal to identify functional motifs in these viruses.

DISCUSSION

Through a combination of sequence analysis and experimental probing data, we generated whole-genome structural models for the previously uncharacterized SINV and VEEV RNA genomes. Using these models, we identified previously known and novel structures in each genome, and we found that both non-coding and coding regions of the genome contain highly structured RNA elements. We applied a systematic mutational method to disrupt RNA structures while preserving amino acid sequence, nucleotide composition, and dinucleotide frequencies. With this method, we confirmed that disrupting two known functional structures—the 5′ CSE and the packaging signal—decreases virus growth. Also, we identified a new functional RNA element in nsP1 whose disruption greatly diminishes viral RNA specific infectivity. The mutant viruses have distinct phenotypes: 5′ SL/5′ CSE and packaging signal mutants have a sharp decrease in growth in Vero cells (and the former also has greatly decreased growth in C6/36 cells), mutated nsP3 SR has no change in phenotype compared with wildtype virus, and mutated nsP1 SR has drastically impaired specific infectivity. These phenotypic differences indicate different mechanisms by which RNA structure regulates the infectivity and growth of SINV.

The data presented in Figure 2 recapitulates known aspects of SINV biology, that the 5′ CSE and packaging signal are critical to viral replication (8,9,25). These data are nonetheless important as they serve as a positive control for our automatic codon usage optimized shuffling strategy (Figure 2A), suggesting that we can disrupt known RNA structures using thus approach. Furthermore, both the 5′ CSE and packaging signal have low-median SHAPE (Figure 1A), confirming that our structural data is predictive of likely important RNA structures in the coding region of the virus.

We also examined the conservation of these structured regions among related alphaviruses and found that most structured regions, aside from the 5′ CSE, are highly divergent. Despite high sequence conservation within most coding regions, there is little evidence of structural conservation. Instead, alphaviruses seem to quickly discard existing structures and evolve new ones, likely a result of their own particular environmental requirements. These viruses must survive in at least two organisms, the arthropod vector and the vertebrate host, and among the alphavirus family there is great diversity in which organisms these viruses infect (1,3,59). The diversity of these viruses is underscored by the discovery of Eilat virus, an alphavirus that cannot survive in vertebrates (60). The environmental diversity of these viruses is mirrored in the diversity of their RNA structures: common elements but individual uniqueness.

Structured regions in the SINV genome

RNA elements around the 5′ end of the SINV genome are critical for virus growth (7,8,11,12,25). We specifically investigated the structured region at the beginning of nsP1 including both the 51-nt 5′ CSE and its preceding 5′ hairpin. It is important to note that the two most highly conserved hairpins we report here in the 5′ CSE are 3′ of the start codon, indicating that specific structures can be conserved in a coding region. In contrast with (but not in contradiction to) previous research that found the 5′ CSE to be functional in mosquito cells but not in vertebrate cells (8), when we disrupted the 5′ CSE and hairpin, the virus grew poorly in both vertebrate and arthropod cells (Figure 2C, Supplementary Figure S4). Although we do not yet know the mechanism by which these structures function, we conclude that the combination of the 5′ hairpin and the 5′ CSE is important for the SINV life cycle in both vector and vertebrate host.

We also confirmed that disrupting the packaging signal, which is also located in a coding region of the virus, disturbs the virus growth cycle in Vero cells, confirming the importance of the structure. Our SHAPE-directed structural model for the packaging signal region found repeated GGG-motifs in stem-loops, as previously suggested (9) (Figure 2B). Because disrupting these stem-loops interferes with growth, the RNA structure throughout this region is critical for viral proteins to recognize genomic RNA. Although we have made every effort to minimize the impact of our coding mutations on codon optimality (Supplementary Table S1), we cannot exclude that our coding mutations may also have an effect on translation by altering codon usage.

Other structured regions include the non-coding junction, which overlaps with the subgenomic promoter, the frameshift element, and a highly structured hairpin ∼100 nucleotides upstream of the leaky stop codon. Although that hairpin was not found to be functional in Vero cells, another study suggested that structure formation downstream of the leaky stop codon plays a role in stop codon read-through (61). It is possible that structural elements, in conjunction with virus or host proteins, regulate the read-through frequency and translation of nsP4.

The frameshift element in the 6K coding region is particularly interesting because its two covarying base pairs are next to each other, indicating that the hairpin is conserved. This hairpin also exists in equine encephalitis viruses, but even among those related equine encephalitis viruses the

structure diverges outside of that hairpin (10). Here, we find the hairpin also exists in SINV. Although this hairpin in the frameshift element is conserved in most alphaviruses that we examined, homologs were not found in every alphavirus, again indicating some degree of structural divergence within the virus family.

Lack of structural conservation among alphaviruses

Covariation, in which base pairs evolve together, is a useful indicator to identify conserved RNA structural elements (15,18,21,62). We used the new program R-scape (21) to quantify the number of significantly covarying base pairs as a metric for conservation of structure. The region with the hairpin and 5′ CSE had the most conservation, with five covarying base pairs. Compared with highly structured, highly conserved RNAs, however, this number of covarying base pairs is quite limited (15,18,21).

In Figure 4B and Supplementary Table S3, we report the percentage of pairs identified by R-scape, or sensitivity of the covariation model, for the SINV structured regions we identified experimentally, compared to classic conserved RNA structures such as tRNA and the L4 leader (21). Sensitivity, as measured by the percentage of pairs identified (Figure 4B), enables us to evaluate the covariation and conservation support for RNA structures in SINV. It is striking that not a single region within the alphavirus genomes, including the highly conserved and functional 5′ CSE, has as much covariation as known conserved elements. The R-scape analysis used to compute this sensitivity is tailored to detect very high levels of conservation, mostly in non-coding RNA, such as those observed in tRNAs and ribosomal RNA (21). We are applying it here to coding regions of the genome, where protein coding sequence is likely the main driver of sequence conservation. As such, the result that all of our structures have far lower sensitivity than known structured RNAs is not necessarily surprising. However, it is essential to note that none of the structures here are supported by covariation evidence, yet we were still able to identify one novel structure with a clear viral phenotype (Figure 3).

Experimental structure probing combined with detailed functional characterizations is likely necessary for identifying novel structured regions in alphaviruses, and it remains to be seen whether this is true for other single stranded RNA viruses. Fortunately, recent technological advances leveraging next-generation sequencing to obtain SHAPE and other forms of chemical and enzymatic probing data will facilitate this approach (22,63–67). Furthermore, these data sets will enable further development of hybrid sequence/experiment approaches for reconciling conservation and experimental data.

Comparing the SHAPE profiles between related sequences is a useful, model-free approach to finding similarities in RNA structure (19,29,30). In this instance, the divergence in SHAPE data between SINV and VEEV supports the conclusion that, outside of highly conserved elements, these viruses are mostly structurally divergent (Figure 1D and E, Supplementary Figure S1). We also measured the structural compatibility of related alphavirus sequences with the SHAPE-derived structures for the 5′ CSE, packaging signal, and the nsP1 and nsP3 structured regions (Figure 4A). It is evident from this analysis that only the 5′ CSE has broad structural conservation across the family. What these data suggest is that specific structural elements are generally not conserved; nonetheless overall patterns of structure vs. unstructured, as evidenced by median SHAPE fluctuations, appear more conserved. These findings are consistent with the idea that in many cases, specific structures are not as important to function as the presence, or absence, of RNA structure in a particular region.

New considerations for RNA structure and evolution

It is often dogmatically suggested that functional structural elements are conserved in related organisms, with the converse being that non-conserved elements are not functional (21). However, we found a new functional, albeit non-conserved, SINV RNA element and a large amount of structural divergence within the SINV packaging signal. Consequently, traditional methods for identifying structure in certain RNAs do not apply adequately to alphaviruses, and may also be problematic with other RNA viruses. Viruses are highly divergent structurally, yet they preserve particular elements such as a single hairpin. Therefore, to adequately study RNA structure in the context of RNA viruses, new computational methods are necessary to integrate high-throughput experimental techniques such as SHAPE-MaP, and to allow for flexibility of structure outside of the most conserved elements.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

The authors thank Naomi Forrester for her assistance with the alphavirus multiple sequence alignment, Elena Rivas for her helpful comments, and Diane Griffin for kindly providing the SINV nsP3 antibody. The authors would also like to thank the Weeks lab for assistance with plotting RNAs and analyzing large RNAs.

FUNDING

National Institutes of Health [HL111527 to A.L., GM101237 to A.L., HG008133 to A.L., AI03311 to N.J.M., AI123811 to N.J.M., AI109680 to M.T.H., AI109761 to M.T.H., AI107810 to M.T.H., AI137887 to M.T.H., A.L. and N.J.M., AI007151-36A1 to K.S.P., AI126730 to K.S.P.]; National Science Foundation [DGE-1144081 to K.M.K.]; North Carolina University Cancer Research Fund (to N.J.M.). The open access publication charge for this paper has been waived by Oxford University Press – *NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
Conflict of interest statement. None declared.

REFERENCES

1. Strauss,J. and Strauss,E. (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.*, **58**, 491–562.

2. Laine,M., Luukkainen,R. and Toivanen,A. (2004) Sindbis viruses and other alphaviruses as cause of human arthritic disease. *J. Intern. Med.*, **256**, 457–471.
3. Suhrbier,A., Jaffar-Bandjee,M.-C. and Gasque,P. (2012) Arthritogenic alphaviruses—an overview. *Nat. Rev. Rheumatol.*, **8**, 420–429.
4. Go,Y.Y., Balasuriya,U.B.R. and Lee,C. (2014) Zoonotic encephalitides caused by arboviruses: transmission and epidemiology of alphaviruses and flaviviruses. *Clin. Exp. Vaccine Res.*, **3**, 58.
5. Hyde,J., Gardner,C., Kimura,T., White,J., Liu,G., Trobaugh,D., Huang,C., Tonelli,M., Paessler,S., Takeda,K. *et al.* (2014) A viral RNA structural element alters host recognition of nonself RNA. *Science*, **343**, 783–788.
6. Hyde,J.L., Chen,R., Trobaugh,D.W., Diamond,M.S., Weaver,S.C., Klimstra,W.B. and Wilusz,J. (2015) The 5' and 3' ends of alphavirus RNAs - Non-coding is not non-functional. *Virus Res.*, **206**, 99–107.
7. Frolov,I., Hardy,R. and Rice,C. (2001) Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA*, **7**, 1638–1651.
8. Fayzulin,R. and Frolov,I. (2004) Changes of the secondary structure of the 5′ end of the Sindbis virus genome inhibit virus growth in mosquito cells and lead to accumulation of adaptive mutations. *J. Virol.*, **78**, 4953–4964.
9. Kim,D.Y., Firth,A.E., Atasheva,S., Frolova,E.I. and Frolov,I. (2011) Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. *J. Virol.*, **85**, 8022–8036.
10. Kendra,J.A., de la Fuente,C., Brahms,A., Woodson,C., Bell,T.M., Chen,B., Khan,Y.A., Jacobs,J.L., Kehn-Hall,K. and Dinman,J.D. (2016) Ablation of programmed -1 ribosomal frameshifting in Venezuelan equine encephalitis virus results in attenuated neuropathogenicity. *J. Virol.*, **91**, e01766-16.
11. Michel,G., Petrakova,O., Atasheva,S. and Frolov,I. (2007) Adaptation of Venezuelan equine encephalitis virus lacking 51-nt conserved sequence element to replication in mammalian and mosquito cells. *Virology*, **362**, 475–487.
12. Gorchakov,R., Hardy,R., Rice,C.M. and Frolov,I. (2004) Selection of functional 5' cis-acting elements promoting efficient sindbis virus genome replication. *J. Virol.*, **78**, 61–75.
13. Noller,H.F., Kop,J., Wheaton,V., Brosius,J., Gutell,R.R., Kopylov,A.M., Dohme,F., Herr,W., Stahl,D.A., Gupta,R. *et al.* (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.*, **9**, 6167–6189.
14. Gutell,R.R., Power,A., Hertz,G.Z., Putz,E.J. and Stormo,G.D. (2005) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.*, **20**, 11.
15. Eddy,S.R. and Durbin,R. (1994) RNA sequence analysis using covariance models. *Nucleic Acids Res.*, **22**, 2079–2088.
16. Hofacker,I.L., Fekete,M. and Stadler,P.F. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
17. Holley,R.W., Apgar,J., Everett,G.A., Madison,J.T., Marquisee,M., Merrill,S.H., Penswick,J.R. and Zamir,A. (1965) Structure of a ribonucleic acid. *Science*, **147**, 1462–1465.
18. Gutell,R.R., Lee,J.C. and Cannone,J.J. (2002) The accuracy of ribosomal RNA comparative structure models. *Curr. Opin. Struct. Biol.*, **12**, 301–310.
19. Kutchko,K.M., Sanders,W., Ziehr,B., Phillips,G., Solem,A., Halvorsen,M., Weeks,K.M., Moorman,N. and Laederach,A. (2015) Multiple conformations are a conserved and regulatory feature of the RB1 5′ UTR. *RNA*, **21**, 1–12.
20. Somarowthu,S., Legiewicz,M., Chillón,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell*, **58**, 353–361.
21. Rivas,E., Clements,J. and Eddy,S.R. (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods*, **14**, 1–8.
22. 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.
23. 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.
24. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W., Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
25. Niesters,H. and Strauss,J. (1990) Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. *J. Virol.*, **64**, 1639–1647.
26. Firth,A.E., Chung,B.Y., Fleeton,M.N. and Atkins,J.F. (2008) Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virol. J.*, **5**, 108.
27. Wilkinson,K.A., Gorelick,R.J., Vasa,S.M., Guex,N., Rein,A., Mathews,D.H., Giddings,M.C. and Weeks,K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
28. Rocca-Serra,P., Bellaousov,S., Birmingham,A., Chen,C., Cordero,P., Das,R., Davis-Neulander,L., Duncan,C.D.S., Halvorsen,M., Knight,R. *et al.* (2011) Sharing and archiving nucleic acid structure mapping data. *RNA*, **17**, 1204–1212.
29. Pollom,E., Dang,K.K., Potter,E.L., Gorelick,R.J., Burch,C.L., Weeks,K.M. and Swanstrom,R. (2013) 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.
30. Kutchko,K.M. and Laederach,A. (2017) Transcending the prediction paradigm: Novel applications of SHAPE to RNA function and evolution. *Wiley Interdiscip. Rev. RNA*, **8**, e1374.
31. Forrester,N.L., Palacios,G., Tesh,R.B., Savji,N., Guzman,H., Sherman,M., Weaver,S.C. and Lipkin,W.I. (2012) Genome-scale phylogeny of the alphavirus genus suggests a marine origin. *J. Virol.*, **86**, 2729–2738.
32. Katoh,K., Misawa,K., Kuma,K. and Miyata,T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059–3066.
33. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
34. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
35. Valdar,W.S.J. (2002) Scoring residue conservation. *Proteins*, **48**, 227–241.
36. Shannon,C. (1948) A mathematical theory of communiction. *Bell Syst. Tech. J.*, **27**, 379–423.
37. Henikoff,S. and Henikoff,J. (1994) Position-based sequence weights. *J. Mol. Biol.*, **243**, 574–578.
38. Soldatov,R.A., Vinogradova,S.V. and Mironov,A.A. (2014) RNASurface: fast and accurate detection of locally optimal potentially structured RNA segments. *Bioinformatics*, **30**, 457–463.
39. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 7287–7292.
40. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
41. Huynen,M., Gutell,R. and Konings,D. (1997) Assessing the reliability of RNA folding using statistical mechanics. *J. Mol. Biol.*, **267**, 1104–1112.
42. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
43. Jorge,D.M.D.M., Mills,R.E. and Lauring,A.S. (2015) CodonShuffle: a tool for generating and analyzing synonymously mutated sequences. *Virus Evol.*, **1**, vev012.
44. Sharp,P.M. and Li,W.-H. (1987) The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.
45. Wollish,A.C., Ferris,M.T., Blevins,L.K., Loo,Y., Gale,M. and Heise,M.T. (2013) An attenuating mutation in a neurovirulent Sindbis virus strain interacts with the IPS-1 signaling pathway in vivo. *Virology*, **435**, 269–280.

46. Park,E. and Griffin,D.E. (2009) The nsP3 macro domain is important for Sindbis virus replication in neurons and neurovirulence in mice. *Virology*, **388**, 305–314.
47. Nawrocki,E.P., Kolbe,D.L. and Eddy,S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics*, **25**, 1335–1337.
48. Nawrocki,E.P. and Eddy,S.R. (2013) Computational identification of functional RNA homologs in metagenomic data. *RNA Biol.*, **10**, 1170–1179.
49. Eddy,S. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
50. Smola,M.J., Calabrese,J.M. and Weeks,K.M. (2015) Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry*, **54**, 6867–6875.
51. Le,S. and Maizel,J. (1989) A method for assessing the statistical significance of RNA folding. *J. Theor. Biol.*, **138**, 495–510.
52. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
53. Snyder,J.E., Kulcsar,K.A., Schultz,K.L.W., Riley,C.P., Neary,J.T., Marr,S., Jose,J., Griffin,D.E. and Kuhn,R.J. (2013) Functional characterization of the alphavirus TF protein. *J. Virol.*, **87**, 8511–8523.
54. Mueller,S., Papamichail,D., Coleman,J.R., Skiena,S. and Wimmer,E. (2006) Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.*, **80**, 9687–9696.
55. Burns,C.C., Shaw,J., Campagnoli,R., Jorba,J., Vincent,A., Quay,J. and Kew,O. (2006) Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J. Virol.*, **80**, 3259–3272.
56. Desmyter,J., Melnick,J.L. and Rawls,W.E. (1968) Defectiveness of interferon production and of rubella virus interference in a line of African green monkey kidney cells (Vero). *J. Virol.*, **2**, 955–961.
57. Nawrocki,E.P., Burge,S.W., Bateman,A., Daub,J., Eberhardt,R.Y., Eddy,S.R., Floden,E.W., Gardner,P.P., Jones,T.A., Tate,J. *et al.* (2015) Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.
58. Gardner,P.P. (2009) The use of covariance models to annotate RNAs in whole genomes. *Brief. Funct. Genomic. Proteomic.*, **8**, 444–450.
59. Weaver,S.C., Winegar,R., Manger,I.D. and Forrester,N.L. (2012) Alphaviruses: population genetics and determinants of emergence. *Antiviral Res.*, **94**, 242–257.
60. Nasar,F., Palacios,G., Gorchakov,R. V, Guzman,H., Da Rosa,A.P.T., Savji,N., Popov,V.L., Sherman,M.B., Lipkin,W.I., Tesh,R.B. *et al.* (2012) Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14622–14627.
61. Firth,A.E., Wills,N.M., Gesteland,R.F. and Atkins,J.F. (2011) Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic Acids Res.*, **39**, 6679–6691.
62. Ritz,J., Martin,J.S. and Laederach,A. (2013) Evolutionary evidence for alternative structure in RNA sequence co-variation. *PLoS Comput. Biol.*, **9**, e1003152.
63. Loughrey,D., Watters,K.E., Settle,A.H. and Lucks,J.B. (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.*, **42**, e165.
64. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
65. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., Bevilacqua,P.C. and Assmann,S.M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.
66. Wan,Y., Qu,K., Zhang,Q.C., Flynn,R.A., Manor,O., Ouyang,Z., Zhang,J., Spitale,R.C., Snyder,M.P., Segal,E. *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, **505**, 706–709.
67. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.