

Rotavirus RNA chaperone mediates global transcriptome-wide increase in RNA backbone flexibility

Aaztli Coria^{a,b}, Michael L Knight^c, Anastacia Wienecke^{b,d}, Daniel Desirò^c, Alain Laederach^{b,d*}, Alexander Borodavka^{c*}.

*To whom correspondence should be addressed.

^a Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599

^b Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599

^c Department of Biochemistry, University of Cambridge, Cambridge, UK

^d Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Abstract

Due to genome segmentation, rotaviruses must co-package a set of eleven distinct genomic RNAs. The packaging is mediated by virus-encoded RNA chaperones, such as the rotavirus NSP2 protein. While the activities of distinct viral RNA chaperones are well studied on synthetic RNA substrates, little is known about their global effect on the entire viral transcriptome. Here, we used Selective 2'-hydroxyl Acylation Analyzed by Primer Extension and Mutational Profiling (SHAPE-MaP) to systematically examine the secondary structure of the rotavirus transcriptome composed of eleven distinct transcripts in the absence and presence of increasing concentrations of NSP2. Surprisingly, SHAPE-MaP data reveals that despite the well-documented helix-unwinding activity of NSP2 *in vitro*, its incubation with cognate rotavirus transcripts does not induce a significant change in the SHAPE reactivities. However, a quantitative analysis of the per nucleotide mutation rate measured by mutational profiling, from which SHAPE reactivities are derived, reveals a global five-fold rate increase in the presence of molar excess of NSP2. We demonstrate that the standard normalization procedure used in deriving SHAPE reactivities from mutation rates can mask an important global effect of an RNA chaperone activity. Further analysis of the mutation rate in the context of structural classification reveals a larger effect on stems rather than loop elements. Together, these data provide the first experimentally derived secondary structure model of the rotavirus transcriptome and reveal that NSP2 acts by globally increasing RNA backbone flexibility in a concentration-dependent manner, consistent with its promiscuous RNA-binding nature.

Introduction

Virions of rotaviruses (RV), a large group of important human and animal pathogens, contain 11 distinct dsRNA molecules, or genomic segments, all of which are essential for rotavirus infectivity (1–3). Each dsRNA segment serves as a template for transcribing differently sized (0.67–3.4 kb) protein-coding +ssRNA transcripts (mRNAs) that can fold into regulatory structures, such as those predicted to be found in the 5′ and 3′ untranslated regions (UTRs) (4, 5). In addition, consensus sequences found in 5′ and 3′ terminal regions of these segments also have important regulatory roles in ensuring proper viral infection, such as recruiting viral proteins to enhance viral translation (6, 7). Li *et al* provided one of the most comprehensive studies of rotavirus RNAs based on the minimum free energy consensus structures deduced from multiple sequence alignments and covariation analysis of group A rotavirus (RVA) strains (5). This study revealed multiple conserved RNA regions present in all RNA segments that were proposed to fold into regulatory elements. However, in this study, structure probing validation of generated models was only carried out for the smallest segment 11 transcript. While purified infectious rotavirus particles contain eleven dsRNA molecules thus precluding its chemical structure probing *in virio*, structural analysis of a complete rotavirus transcriptome with single-nucleotide precision would enable identification of RNA structures for further functional validation through recently established reverse genetics studies (8). Structural characterization of multiple RNA viruses have provided insights into previously unknown functional regulatory motifs (9, 10). Thus, the first step of identifying functional elements in an RNA viral genome is to model experimentally validated RNA secondary structures.

Importantly, rotaviruses encode an RNA chaperone NSP2 that is believed to be involved in RNA selection and stoichiometric segmented genome packaging (3, 11). Single-molecule fluorescence studies of NSP2-mediated RNA remodeling with short (< 50 nts) RNA substrates indicate that

NSP2 binds single-stranded RNAs with nanomolar affinity, causing transient helix destabilization and melting of its secondary structure (12, 13). These studies also revealed that NSP2 octamers can bind to RNA stem-loops in both open and closed conformations, raising further questions about the mode of action of NSP2 as an RNA chaperone that promotes inter-segment RNA interactions in RVs (13). While these studies were important for our understanding of the RNA chaperone-mediated RNA annealing, there is no direct evidence of structural reorganization of the RVA RNA upon NSP2 binding. Moreover, NSP2 binding itself poses another major challenge for modelling RV RNA structures, as there is very little understanding of how multivalent RNA chaperones can alter RNA folding on a transcriptome level (14). NSP2 is an RNA chaperone known to melt RNA helices, its helix-destabilizing properties have been characterized using single-molecule FRET and gel electrophoresis with synthetic RNA and DNA substrates (12, 13, 15). However, this helix-destabilizing activity has yet to be investigated using RV RNA at transcriptome scale.

RNA folding is mediated by base-pairing that forms secondary structures which play a myriad of fundamental regulatory roles including gene expression (9, 16, 17). To experimentally interrogate RNA structure, chemical structure probing coupled with thermodynamic modeling provides single nucleotide resolution information with transcriptome scale throughput (18, 19). This is particularly pertinent for viruses, whose RNA genomes, or ssRNA precursors of dsRNA genomes, can be directly translated into viral proteins, or participate in the assembly of a nascent virus (20, 21). Thus, experimentally investigating RNA structure with chemical probing allows us to identify potentially new and important roles for RNA structure in rotaviruses.

Here, we provide experimentally validated structure models of a complete rotavirus RNA pre-genome probed by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) and

mutational profiling (MaP) analysis (19, 22). We identified multiple conserved structured regions with low SHAPE reactivity located in the 5' terminal regions of RV RNAs. We show that the RNA chaperone NSP2 binding to RV transcripts dramatically increases mutational rates induced by the SHAPE reagent, but has no effect on the computed SHAPE reactivity, suggesting a global effect on backbone flexibility with minimal structural rearrangement. Analysis of SHAPE-informed secondary structure models in the presence of NSP2 reveals that despite its promiscuous RNA binding mode, its binding preferentially affects base-paired elements, consistent with the RNA chaperone activity. These findings support a model in which promiscuous binding of NSP2 to the rotavirus transcriptome allows this chaperone to uniformly increase the flexibility of the RNA backbone, thus leading to the increased accessibility of RNA towards the establishment of inter-molecular contacts.

Materials and Methods

Rotavirus A transcripts

pUC19 clones encoding for segments 1-11 of bovine rotavirus A (strain RF, G6P6[1]) (13, 23) were used for *in vitro* transcription reactions. Sequences for the 11 segments are listed in Supplementary Table 4 in SNRNASM format. The plasmids coding for segments 2 and 8 were digested using BbsI, the plasmid coding for segment 10 was linearized with BsaI, and the plasmids coding for the remaining segments were linearized using BsmBI restriction enzymes, following the manufacturer's protocol (NEB). Following linearization, the digestion product was purified via gel extraction (QiaQuick Gel Extraction Kit). Each RV segment was then transcribed using 1 µg of linearized DNA with HiScribe T7 from NEB at 37° for 3 hours. The transcription reaction was subsequently treated with Turbo DNase (ThermoFisher) for 30 minutes at 37. The RNA was purified using Purelink RNA columns (ThermoFisher) and eluted in 30 µL of nuclease free water. Each RNA segment quality and size were assessed on a 2% denaturing MOPS formaldehyde agarose gel, prior to loading on the gel RNA was incubated at 95° in formaldehyde and denaturing buffer (ThermoFisher).

NSP2 Expression and Purification

A sequence-verified pET-28b-NSP2 construct (13) was used for protein expression, as described in (24). Ni-affinity-purified NSP2 fractions were further purified over a HiTrap SP cation-exchange column (Cytiva), and the concentrated peak fractions were resolved on a Superdex 200 10x300 GL column and pre-equilibrated with RNase-free SEC buffer (25 mM HEPES-Na, pH 7.5, 150 mM NaCl) to ensure high purity and homogeneity of the preparation. The peak fraction was at least 99% pure by SDS-PAGE, with a characteristic 260 nm/280 nm absorbance ratio of < 0.57 suggesting that the purified protein was pure from contaminating nucleic acids.

Structure Probing using SHAPE-MaP

To prepare RV RNA for structure probing – for single RNA experiments the single RNA segments were diluted to a final concentration of 260 nM and incubated at 65°C for 5 minutes then slow cooled back to room temperature for 30 minutes then kept at 4. For transcriptome wide experiments each RV segment was diluted to a concentration of 25 nM, incubated at 65°C for 5 minutes then slow cooled back to room temperature were then pooled, resulting in an equimolar mix of all 11 RV segments. Stock NSP2 was diluted using 1 mM HEPES at pH 7 to concentrations required for the titrations described; 40 μ M, 20 μ M and 10 μ M.

The RNA was then buffered in the following conditions: 1 mM MgCl₂, 200 mM NaCl, and 200 mM HEPES (pH 7). NSP2 was added at equal volume to the RNA, for the no protein condition, 1mM HEPES was added at equal volume to the RNA. The RNA:NSP2 samples were then incubated at 37° for 30 minutes prior to probing. For probing, RNA:protein mix were incubated with 10% volume of 125 mM 5-nitroisatoic-anhydride (5NIA) dissolved in dimethyl sulfoxide (DMSO) at a final concentration of 12.5 mM 5NIA. For 1M7 (1-Methyl-7-nitroisatoic anhydride) treated samples, RNA was treated with 10% volume of 100 mM 1M7 for a final concentration of 10 mM 1M7. Untreated samples were incubated with 10% DMSO. Both modified and untreated samples were vortexed then incubated at 37° for 5 minutes. RNA was subsequently purified using the Qiagen MinElute columns and eluted in 30 μ L of nuclease-free water.

DMS Probing of transcripts for segment-specific amplification

Rotavirus segment 11 RNA was refolded as described above. The RNA was diluted to a final concentration of 250 nM in the RNA folding buffer (5 mM HEPES-Na pH 7.5, 0.5 mM MgCl₂, and 100 mM NaCl). The diluted RNA was then mixed with an equal volume of either 20 μ M NSP2 and samples were incubated at 37°C for 30 minutes. For DMS probing, 1 M bicine pH 8.0 was

added to each sample give a final concentration of 200 mM. Samples were incubated at 37°C for 1 minute with 10% DMS in ethanol to give a final DMS concentration of either 0.5% or 0.1% v/v. Control samples were incubated with an equivalent volume of ethanol. Reactions were quenched by addition of 1 M DTT to a final concentration of 167 mM. RNA was purified using the Zymo RNA Clean and Concentrator 5 kit. Samples were then prepared using our segment specific library preparation protocol as described below.

In-Cell Structure Probing using SHAPE-MaP

MA104 cells were grown in T25 flasks in DMEM-high glucose (Sigma), supplemented with 2 mM GlutaMAX (Gibco), 100 U/ml of Penicillin-Streptomycin (Gibco), and 10% Fetal Calf Serum (FCS) until confluent. Cells were washed by incubating with FCS free media at 37°C for 30 minutes. Group A rotavirus (RVA), strain RF (bovine) was activated by incubating with 1.5 µg/ml of trypsin for 30 minutes at 37°C, as previously described (25). Infections were performed in 3 ml of FCS-free media for 2 hours, at 37°C, at an MOI of 1.4. Following the initial incubation, an additional 7.5 ml of FCS-free media was added and cells were incubated for a further 6 hours at 37°C.

Cells were washed twice with PBS and detached by incubation with trypsin. Detached cells were pelleted at 500 x g for 5 minutes and resuspended in 1.8 ml of 200 mM HEPES-Na pH 7.5 in PBS. The sample was then split in two and incubated with a 10% (v/v) of 250 mM 5NIA (25 mM final concentration) or DMSO at 37°C for 10 minutes. The reaction was quenched by the addition of 1 M DTT to a final concentration of 125 mM, and RNA was purified using the Monarch Plasmid Miniprep Kit (NEB).

Tagmentation-Based Library Preparation and Next Generation Sequencing

First strand cDNA synthesis of treated RNA was carried out as previously described (22). First, 29 μ l of purified modified RNA was added to 100 ng Random Primer 9 (NEB) and 0.2 mM of each dNTP. This mix then incubated at 65°C for 5 min then incubated on ice. This reaction was buffered for a final concentration of 50 mM Tris (pH 8.0), 75 mM KCl, 10 mM DTT, and 6 mM $MnCl_2$ and 2 μ l of SuperScript II Reverse Transcriptase (Invitrogen). The reaction was supplemented with 20 U of Murine RNase inhibitor (NEB) and the reaction was incubated at 25°C for 10 min, then 42°C for 3 h, then 70°C for 15 min. First strand cDNA was desalted using G50 columns (GE). The desalted product of first strand synthesis was immediately introduced into a second strand synthesis reaction. The cDNA second strand was generated using mRNA Non-Directional Second Strand Synthesis module (NEB). Double stranded DNA was then cleaned up using Ampure XP purification beads and eluted in 30 μ L of water and quantified using a Qubit dsDNA HS Assay (Invitrogen). Double-stranded DNA was fragmented, repaired, and adaptor ligated for sequencing library preparation using the NEBNext[®] Ultra[™] II DNA Library Prep Kit for Illumina[®]. Library quality was assessed using the High Sensitivity Bioanalyzer protocol (Agilent). Libraries were then loaded on an Illumina MiSeq at 10 pM and run at 300 cycles (150x2) or 600 cycles (300x2).

Segment-Specific Library Preparation and Next Generation Sequencing

In order to reverse transcribe samples 500 ng of total RNA was combined with 2 pMol of reverse transcription primer (Supplementary Table 3) in a volume of 8 μ l. Following this, 2 μ l of 10 mM dNTPs were added and samples were incubated for 5 minutes at 37°C, before transferring them immediately to ice for 2 minutes. A reverse transcription master mix was created by mixing 2 μ l of 10x NTP minus buffer (0.5 M Tris-Cl pH 8.0, 750 mM KCl, and 100 mM DTT), 4 μ l of 5 M betaine, and 3 μ l of 40 mM $MnCl_2$ per sample. A total of 9 μ l of the reverse transcription master

mix was added to each sample. The samples were then incubated at room temperature for 2 minutes before the addition of 1 µl of 200 U/µl SuperScript II Reverse Transcriptase (Invitrogen). Reverse transcription was performed using the following program in a thermocycler: 10 minutes at 25°C, 90 minutes at 42°C, 10 cycles of 50°C for 2 minutes followed by 42°C for 2 minutes, and a final incubation of 70°C for 10 minutes. The resulting cDNA was purified using the Zymo DNA Clean and Concentrator 5 kit. A total of 30-50 ng of cDNA was combined with a final concentration of 0.2 mM dNTPs, 0.25 µM of gene segment 11-specific primer (Supplementary Table 3) and 20 U/µl of Q5 Hot Start High-Fidelity DNA Polymerase in Q5 DNA Polymerase buffer (NEB). Amplification was performed using the following program in a thermocycler: 98°C for 30 seconds, 20 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 20 seconds, followed by a final incubation at 72°C for 2 minutes. DNA was purified by addition of a 1x volume of AMPure XP beads (Beckman-Coulter). Beads were separated on a magnet and washed twice with 80% ethanol before eluting in 10 µl of water. A total of 2 ng of DNA from the first amplification was combined with a final concentration of 0.2 mM dNTPs, 0.5 µM of a universal forward primer to add Illumina adaptor (Supplementary Table 3), 0.5 µM of a sample specific reverse primer to add a sample specific index and Illumina adaptor (Supplementary Table 3), and 20 U/µl of Q5 Hot Start High-Fidelity DNA Polymerase (NEB) in Q5 DNA Polymerase buffer (NEB). Amplification was performed using the following program in a thermocycler: 98°C for 5 minutes, 10 cycles of 98°C for 10 seconds, 66°C for 30 seconds, and 72°C for 20 seconds, followed by a final incubation at 72°C for 2 minutes. DNA was purified by addition of a 0.8x volume of AMPure XP beads (Beckman-Coulter). Beads were separated on a magnet and washed twice with 80% ethanol before eluting in 10 µl of water. Libraries were then loaded on an Illumina MiSeq at 10 pM and run for 600 cycles (300x2).

SHAPE-MaP Analysis

Sequencing reads produced from our mutational profiling experiments were analyzed using the ShapeMapper2 pipeline (26), version 2.1.4 using Bowtie2 version 2.3.4.3 as the aligner. ShapeMapper2 computes mutation rates for both treated and untreated samples by counting the mutations at each individual nucleotide along the transcript in comparison to the reference sequence (Supplementary Table 4). Reactivity is computed by subtracting the mutation rate of the untreated sample from the mutation rate of the treated sample. Statistical outliers are then excluded from further analysis and reactivities are normalized as previously described (27).

RNA Secondary Structure Modeling and Forgi Classification

All minimum free energy diagrams were generated using RNAStructure (28) by incorporating SHAPE reactivities as a pseudo free energy term and using a maximum pairing distance of 150 nucleotides unless otherwise stated. Using connectivity tables (.CT) generated from RNAStructure we then used the Forgi python library to classify each nucleotide into a specific RNA secondary structure motif (29). These nucleotide classifications were generated from our SHAPE informed secondary structures derived from structure probing experiments of the RV transcriptome incubated together at an equimolar ratio.

Shannon entropy modelling

Base-pairing probabilities for each transcript were predicted using RNAfold from the ViennaRNA package with the -p option (30). This option enables the calculation of the partition function and base pairing probability matrix (31). We predicted the base-pairing probability matrix for each RNA sequence R . Shannon entropy values (32), H_i for all nucleotides $i \in R$ were then calculated using all base pair probabilities p_{ij} of a paired nucleotide i with j across all potential paired nucleotides R as follows:

$$H_i = - \sum_{j=1}^R p_{ij} \log_{10} p_{ij}$$

A median of a 50-nucleotide sliding window was used to create the final Shannon entropy profiles for each transcript.

RNA conservation analysis

All nucleotide sequences of rotavirus A (taxid: 28875) for NSP5, NSP4 and VP6 were extracted from the NCBI virus database (07.27.2021). Pairwise alignments of all sequences of each segment with the respective segment of the bovine RF strain used in our experiments (KF729678.1, KF729677.1, KF729658.1) were carried out using MAFFT with default parameters, and the Levenshtein distance for each pair was calculated (33, 34). Sequences that had less than 90% sequence identity with our reference sequence, and/or introduced gaps in the reference sequence with missing conserved terminal sequences were removed. This resulted in a total of 79 sequences for NSP5, 77 sequences for NSP4 and 328 sequences for VP6. The 90% sequence cut-off for NSP5, VP6 and NSP4 was chosen as appropriate following the recommendations for full-genome classification of rotaviruses by Matthijnssens *et al* (35). Multiple sequence alignments for each segment were generated with MAFFT as described above, and these alignments were used to calculate the nucleotide conservation of the most abundant nucleotide at each position. All sequences, and the corresponding GenBank accession numbers are provided in the Supplementary Table 5. Since the nucleotides were generally conserved, we also computed the mean of a 10-nucleotide window to generate conservation heatmap plots to depict the variation in the nucleotide conservation between 0.75 and 1.00. RNA structures were created using RNAfold from the ViennaRNA package with the -p option to calculate the partition function and the --shape option to incorporate our SHAPE-MaP values in the prediction (30). These structures were then

visualized and coloured using VARNA by reassigning the conservation values of each segment between 0 and 1 (36).

Statistical Analyses

Mutation rate distribution analysis and establishment of linear relationships between NSP2 titration points (Figure 3, 4, 5) was analyzed using R studio (version 2021.09.1+372) and visualized with the ggplot2 package. Box plots and histograms (Figures 5 and 6) comparing the log2FoldChange of RNA structural motifs was done using the python package Matplotlib.

Results

SHAPE-MaP derived structural analysis of rotavirus A transcriptome

Recently established reverse genetics approaches for rescuing viable rotaviruses utilize RVA transcripts produced by T7 polymerase-driven transcription (8). Therefore, we utilized a set of plasmids used for the virus rescue to produce eleven RVA transcripts in vitro (23). The produced RNA was intact and homogenous to be used for subsequent analyses by high-throughput selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) to interrogate local nucleotide flexibilities (Supplementary Figure 1). To develop an accurate model of RNA structure of the RVA transcriptome, we co-incubated an equimolar mix of eleven RV transcripts (0.25 μ M total RNA concentration) in the RNA folding buffer (100 mM NaCl, 100 mM HEPES-Na, 0.5 mM MgCl₂, pH 7) at 37°C for 30 min. Under these conditions, eleven RVA transcripts did not form stable inter-molecular contacts in our previous single-molecule fluorescence studies (13). The RNA mix was then incubated with the SHAPE reagent 5-nitroisatoic anhydride (5NIA), as described in Materials and Methods (37). Conformationally flexible nucleotides exhibit high reactivities towards the electrophile 5NIA (37), thus allowing us to assess the RV transcriptome reactivity with single-nucleotide resolution. We performed at least two replicates of all SHAPE experiments

and SHAPE reactivities for each one were reproducible and highly correlated with correlation coefficient in the range **between 0.73 – 0.98 (Supplementary table 1)**. We quantified relative flexibility of each transcript using median SHAPE analysis as previously described (37–41) (Figure 1A). Regions of low median SHAPE are generally more likely to fold into a single, well-defined conformation, while above median SHAPE generally adopt multiple conformations (29). We modeled the secondary structure and calculated base-pairing probabilities for the entire RV transcriptome using our experimentally determined SHAPE reactivities averaged over both replicates as a pseudo-free energy term as previously described (39, 42, 43). The resulting modelled base-paired nucleotides are shown as arcs, while the color of the arc represents the likelihood of the base-pair forming (Figure 1A). We show selected minimum free energy structures for regions of low SHAPE in Figure 1B.

To further validate our SHAPE-MaP probing data, we compared our SHAPE-informed secondary structure model of the segment 11 transcript to a predicted secondary structure model derived from our previously published RNA-RNA SELEX data (13). The analysis revealed that both models were broadly similar (PPV of 66.8% and sensitivity of 84.52%, Supplementary Figure 1). **In addition, we compared our SHAPE-derived models with the secondary structure models generated using sequence conservation analysis alone. Our conservation analysis identified highly conserved regions present in segments 6, 10 and 11 (Figure 2). For group A rotavirus segments 6 and 10, the analysis of 328 unique VP6-coding and 77 NSP4-coding sequences, respectively (sequence selection criteria are described in Materials and Methods) has revealed multiple discrete regions of high conservation, most notably encompassing terminal nucleotides (Figure 2). In segment 6 RNA, the conserved 3' terminal sequences are predicted to fold into highly structured stem-loops (Figure 2A), fully supported by our SHAPE-MaP data. We further identified two previously**

predicted stem-loops located in the 3' terminal region of segment 6 known to enhance translation (44). Similarly, segment 10 RNA was predicted to have a highly conserved terminal structure comprising stable stem-loops in the 3' UTR (Figure 2B), fully in agreement with previous covariation analysis carried out by Li *et al.* (5). Moreover, analysis of the shortest segment 11 transcript reveals extended conserved sequences, broadly correlating with lower median SHAPE reactivity (Figure 2C), while our SHAPE-MaP data supports the presence of long-range interactions (LRIs, including conserved helices H1 to H3 shown in Figure 2C), in agreement with the prediction of Li *et al.* Finally, our probing data enabled us to identify the predominant LRI structure in segment 9 amongst the alternative models that had been previously proposed by Li *et al.* (Supplementary Figure 2). Collectively, these data reveal that all RV transcripts have structured 5' terminal regions, expanding into coding regions, except for segments 2 and 5. Previously, it was found that the first 78 nucleotides in segment 1 and segment 7 are highly conserved (5). Additionally, the first 82 nucleotides and the first 94 nucleotides of segments 9 and 3, respectively, are conserved. All these regions of high conservation appear to be structured, i.e., with low median SHAPE reactivity, as seen in Figure 1A. In contrast, segments 2 and 5 both have unstructured 5' terminal regions, with neither of these segments exhibiting sequence conservation in the 5' terminal regions (5). Thus, our data support the notion that RNA structural pervasiveness is present within areas of high sequence conservation in RV genomes, and demonstrate that our experimental data are in good agreement with previously determined co-variation models.

To validate our findings *in cell*, we also carried out SHAPE-MaP by treating RV-infected cells at 6 hours post infection, i.e., prior to dsRNA synthesis when the majority of the viral RNA is single-stranded. Overall, SHAPE-MaP-guided structures of both segment 11 RNAs probed *in vitro* and *in cell* were similar (Figure 3), revealing that the RV RNA conformational ensembles in cells are

remarkably comparable to the ones obtained *in vitro*. We also noted that the highly conserved sequence immediately before the start codon of NSP6 encoded by segment 11, had high SHAPE reactivity both *in vitro* and in cells (Figure 3A, green box), strongly supporting the notion that this region may be highly accessible for translational initiation, as proposed by (5).

Mutation rate analysis reveals that NSP2 increases segment 11 RNA backbone flexibility

To measure the effects of NSP2 on RVA transcripts, we began by incubating *in vitro* transcribed segment 11 RNA (0.67 kb) with increasing amounts of NSP2 and carried out SHAPE chemical probing (Figure 4A). Individual SHAPE reactivities of nucleotides across the segment 11 transcript are represented in a bar chart, where high SHAPE reactivities above 0.7, red bars, indicate a flexible nucleotide and nucleotides with low SHAPE reactivities below 0.5, black bars, represent constrained nucleotides. Visual inspection of these data does not reveal any obvious large differences in the SHAPE profiles. This is confirmed in Figure 3B where we plot the SHAPE reactivities of segment 11 alone (x axis) and segment 11 SHAPE reactivities when incubated with increasing amounts of NSP2 (y axis). We found that the slopes for all the titration points were close to 1, and the data remained highly correlated with $R^2 = 0.78$ for the 20 μM NSP2 end point (Figure 4B, purple line), which is within expected replicate to replicate variability (Supplementary Table 1). Further analysis of SHAPE reactivities of segment 11 RNA in the presence of increasing concentrations of NSP2 reveals only a subtle change in the global median of SHAPE reactivities (Figure 4C). Thus, SHAPE-MaP analysis detects only minimal structural rearrangements without a significant apparent global change in SHAPE reactivity.

SHAPE reactivities are calculated from raw mutation rates (Figures 4D and Figure 4E). The mutation rate of the untreated sample (Figure 4D, DMSO control) is subtracted from the mutation rate of the treated sample (Figure 4D, 5NIA) for each individual nucleotide. Statistical outliers are

removed, and the remaining mutation rate intensities are normalized so all reactivities fall between 0 (highly constrained) to 2 (highly flexible) (26, 27). Thus, if NSP2 binding uniformly increases reactivities of all bases, the apparent SHAPE reactivities would remain unchanged, while the mutation rates would uniformly increase for all bases. Thus, this prompted us to investigate the raw mutation rates. As expected, raw mutational rate analysis confirmed that nucleotides in the 5NIA-treated samples are more reactive compared to the DMSO-treated control. Furthermore, both mutation rate and SHAPE reactivity analyses replicates are highly correlated and have a slope near one (Figure 4F and Figure 4G). We therefore hypothesized that if NSP2 has a global effect on backbone flexibility, it would effectively be normalized out and only an analysis of the raw mutation rates could reveal such effects.

Remarkably, the addition of NSP2 resulted in a concentration-dependent increase in the distribution of nucleotide mutation rates (Figure 5A), in agreement with our hypothesis. The pattern of the mutation rates for all bases however remained similar (Figure 5A), in agreement with a very similar pattern observed in the normalized SHAPE reactivities (Figure 4A). Remarkably, when we plotted the mutation rate upon addition of NSP2, we observed a strong linear correlation for increasing concentrations of the RNA chaperone, with the slope increase up to 5.0 ± 0.1 for the highest 20 μM NSP2 concentration tested (Figure 5B). We did not attempt higher NSP2 concentrations as those have been previously shown to cause severe RNA aggregation rather than promote strand-annealing reactions (13). The mutation rate increase in the presence of NSP2 is further confirmed by our analysis of distribution of mutation rates in NSP2-concentration dependent manner (Figure 5C). Thus, NSP2 has a global effect on segment 11 backbone reactivity as measured by mutation rate, that is masked upon normalization when SHAPE reactivities are computed using the standard normalization procedure. Specifically, the

quartile normalization step used when deriving SHAPE reactivities from mutational profiling data (27, 38) has the unintended effect of masking these global changes in reactivity in the data.

Finally, to ensure that the changes in mutation rate are mediated by NSP2 and not perturbations caused by the addition of protein, we also incubated segment 11 RNA with 20 μ M BSA and found no changes in mutation rate or SHAPE reactivity (Supplementary Figure 3), suggesting the observed effect is specific to NSP2.

NSP2 uniformly increases backbone flexibility in RV transcripts

To assess if NSP2 increases backbone flexibility across different RV segments, or if the global increase in backbone flexibility is specific to segment 11 RNA, we examined *in vitro* transcribed segment 5, 6, and 10 transcripts with increasing concentrations of NSP2. We found a similar increase in mutation rate in segments 5, 6, and 10 as segment 11. As can be seen in Figures 6A-C the mutation rate 20 μ M endpoint of the titration yields slopes >4 for segments 5, 6, 10, respectively. Collectively, these data suggest NSP2 increases mutation rate across RV transcripts indiscriminately.

NSP2 preferentially increases mutation rate in base-paired structural motifs

The overall pattern of mutation rates remains highly correlated upon addition of up to 20 μ M NSP2 (Figures 6A-C) for each segment, as does the SHAPE data. Therefore, we applied mutation rate analysis to determine if NSP2 has any structural specificity at transcriptome scale, by categorizing all nucleotides using the Forgi library classifications (interior bulge, multiloop, stem, and hairpin loop) as described in the Methods (Figure 6D)(29). These classifications are based on the SHAPE-directed minimum free energy secondary structures reported in Figure 1.

To compare nucleotides in the four different structural contexts, we computed the $\log_2(\text{Fold Change})$ of the mutation rate at 20 and 0 μM NSP2 respectively. As can be seen in Figure 5E, a vast majority of nucleotides have $\log_2(\text{FC}) > 0$, consistent with the positive slopes we observed in the scatter plots in Figures 6A-C. Remarkably, we also observed a statistically significant increase in the $\log_2(\text{FC})$ for paired nucleotides (stem) compared with unpaired (hairpin loop, multi-loop, and interior bulge, Figure 6E). The advantage of computing $\log_2(\text{FC})$ in this way is that it allows quantitative comparison of nucleotides with differing mutation rates. **Thus, our data show that although the effect of NSP2 on mutation rates is uniformly positive, NSP2 has a larger relative effect on the flexibility of stems vs. unpaired nucleotides, consistent with its helix-destabilizing activity.**

Structural consequences of rotaviral transcriptome on $\log_2(\text{FC})$

As we have previously shown, at sub-nanomolar concentrations, rotaviral segments do not appear to interact intermolecularly when NSP2 is absent (13). Indeed, when we compare segments probed alone vs. those probed in the presence of all 11 segments (Supplementary Figure 4) we observe highly correlated mutation rates and slopes close to 1. We were therefore interested to see if our $\log_2(\text{FC})$ analysis can detect any statistical differences at transcriptome scale when all eleven segments are incubated with NSP2. We therefore probed the entire transcriptome in the presence of increasing concentrations of NSP2. As can be seen in Figures 6A-D, the 20 μM NSP2 mutation rates are highly correlated in both the absence (x-axis, alone) and presence (y axis, transcriptome) of the other segments, with slopes near 1 in all cases. We also computed slope and R^2 for increasing concentrations of NSP2 and report these in Table 1, and all transcripts exhibit a similar concentration dependent trend. This suggests very little difference in the overall structure of these segments in the presence or absence of the transcriptome.

However, when we compared the $\log_2(\text{FC})$ distributions for segments 5, 6, 10 and 11 alone and in the presence of the rest of the viral transcriptome (Figure 6E), we observed a structure-specific difference. The analysis shows that the $\log_2(\text{FC})$ is systematically higher for paired nucleotides (as observed in Figure 5E) for both individual transcripts alone and in the presence of transcriptome. However, there are subtle but statistically significant differences in the quartile distributions for paired and unpaired nucleotides. Notably, the quartile distributions of mutation rates are significantly lower for paired nucleotides (green) in the presence of the transcriptome, while for hairpin and multi-loops they are significantly higher (orange, yellow and red). Similarly, for interior bulges, the increase is positive, but not significant, likely due to the small number of nucleotides in this conformation relative to loops and stems.

In the context of our previous results, this analysis reveals a subtle but important result that could be explained by transient intermolecular interactions between unpaired nucleotides in the context of the transcriptome. First, neither the addition of NSP2 nor the transcriptome has any detectable effect on the pattern of the SHAPE reactivities (Figure 3A). Thus, it can be concluded that NSP2 is not systematically refolding the RNA in a way that is detectable by chemical probing. Despite fairly subtle changes, our analysis suggests that NSP2 exerts a larger effect on the $\log_2(\text{FC})$ for paired nucleotides. Remarkably, in the presence of NSP2, paired bases are significantly more reactive towards 5NIA compared to unpaired nucleotides. However, their median reactivity decreases in the presence of the remaining rotaviral transcriptome (Figure 6E, green), consistent with the proposed model that the RNA chaperone NSP2 facilitates inter-molecular RNA-RNA

interactions by exposing complementary sequences that are normally sequestered by the local structure (13).

Taken together, these data show that quantitative analysis of mutation rates can detect subtle but important structural preferences in promiscuous RNA-binding proteins like NSP2 at transcriptome scale, even if they do not appear to have strong structural or sequence preference (45) or when analyzing the normalized SHAPE-MaP data.

Discussion

Our transcriptome-wide chemical structure probing of eleven rotavirus A transcripts data validates previous computational and co-variance models of these RNAs, providing first experimental support to enable modelling of these structures. In addition, we have identified multiple well-defined structured regions of low SHAPE reactivity across the entire transcriptome (Figure 1). Thus, our data suggest that RNA structure is pervasive in rotaviral segments and not limited to untranslated terminal regions. Pervasive RNA secondary structures is a feature of all RNA viruses studied to date using chemical probing including flaviviruses (46), H1N1 (41), HIV (9), alphaviruses (10, 39), and coronaviruses (47, 48) Remarkably, *in cell* probing of one of the segments, gene segment 11 transcript, revealed a very similar chemical reactivity profile to that seen *in vitro*. Interestingly, this result contrasts with the observations that transcripts of influenza A viruses exhibit distinct reactivities when refolded *in cell* vs *in vitro* transcribed mRNAs that appeared to be less structured in cells (49). It should be noted that the structures we have identified here only exist prior to genome packaging, i.e., before they are replicated into fully double-stranded RNA form inside the virion (1, 50, 51), thus precluding the analysis of the RNA structure *in virio*. Nonetheless,

the analysis of rotavirus transcriptome has generated insights into the pervasiveness and dynamic nature of RNA structures for all RV transcripts that contain similarly sized regions of low SHAPE reactivity ranging from 50-200 nucleotides, similar to other RNA viruses. Moreover, we used SHAPE-MaP to evaluate how NSP2, a key RNA chaperone involved in RNA assortment and packaging in rotaviruses, may interact with these pervasive structures.

Despite multiple biophysical and biochemical studies that have demonstrated helix-destabilizing activity of NSP2 with shorter RNA substrates (12, 13, 15), surprisingly, incubation of any of the rotaviral segments with even molar excesses of NSP2 had no significant effect on apparent SHAPE reactivity. At first, these results appeared to be at odds with our previous biophysical studies of the RNA chaperoning effect of NSP2 on small RNA stem-loops that revealed a melting of the structure (13, 15). However, direct analysis of adduct-induced mutations via mutational profiling allowed us to observe the previously unknown global effect of NSP2 on the rotaviral transcripts.

Indeed, the quartile normalization used in the SHAPE-MaP data processing effectively masks any large global changes in mutation rates. Thus, although next generation sequencing enables unprecedented throughput, detailed understanding of the experimental system (i.e., specific vs non-specific RNA binding, RNA:protein binding stoichiometry, complementary biophysical data supporting RNA conformational changes) should always be considered for careful data interpretation. Furthermore, visualization of the raw mutational data during analysis should always be carried out to reveal any changes in the RNA backbone dynamics that might be normalized out in subsequent data processing steps. Additionally, we carried RNA probing using the established SHAPE reagent 1M7 in the presence and absence of NSP2. Our results with 1M7 reagent confirm the 5NIA data, suggesting that upon addition of NSP2, RNA backbone of segment 11 transcripts increases its reactivity as quantified by the mutation rate (Supplementary Figure 5). 1M7 has much

faster reaction kinetics than 5NIA [Cite Costin et al.] and in previous studies a small subset of specific nucleotides in highly structured RNAs were found to have differential reactivity.

Having established that we observe a very significant, global, and NSP2 concentration-dependent increase in mutation rate upon addition of the RNA chaperone, we decided to evaluate whether we observe any structural specificity. Previously, it has been suggested that with small RNAs, NSP2 preferentially binds unpaired nucleotides with sub-nanomolar affinity (13). Our classification analysis of bases as paired or unpaired nucleotides allowed us to observe that the $\log_2(\text{Fold Change})$ in mutation rate upon addition of NSP2 was more positive for paired nucleotides across the entire RV transcriptome (Figure 6). Our previous studies suggest that NSP2 preferentially binds unpaired RNA regions (45), therefore the higher $\log_2(\text{Fold Change})$ for paired regions does not reflect any structural preference in binding of NSP2. Rather, our mutational profiling analysis indicates that NSP2 has a larger relative effect on paired nucleotide's flexibility, consistent with its role as an RNA chaperone that promotes RNA-RNA interactions by making sequestered complementary sequences more accessible for inter-molecular base-pairing. It is important to note that the increase in mutation rate indicates a global change in backbone flexibility, but does not change the overall SHAPE pattern. To directly interrogate any base pairing changes mediated by NSP2, we performed an analogous NSP2 titration using dimethyl sulfate (DMS), which methylates unpaired adenines and cytosines, thus directly reporting on the propensity for base-pairing. NSP2-mediated increase in mutation rate is consistent between DMS- and 5-NIA-treated samples (Supplementary Figure 6). Indeed, when treated with DMS, we observed that stem-forming paired nucleotides experience a more significant increase in $\log_2(\text{FC})$ in the presence of NSP2 (Supplementary Figure 6), further confirming helix-destabilizing effect of NSP2 as the result of increased RNA backbone flexibility. It should be noted that further in-depth analyses followed by

an extensive functional validation is required to identify the nucleotides involved in the assembly of eleven distinct RV transcripts. Recent studies of the assortment and packaging of 8 RNA segments in influenza A viruses suggest that in IAVs such interactions are likely to be highly redundant, and potentially transient in nature (41, 52), thus making their identification and validation a particularly challenging task. Nevertheless, understanding the mechanistic role of promiscuous RNA-binding proteins with RNA chaperone activity will help our endeavors to elucidate the RNA assembly pathways leading to a stoichiometric segment packaging in these viruses.

These data further support a model in which NSP2 is able to capture folded RNA stem-loop structures within a single RNA-binding groove, resulting in general relaxation of RNA structure without fraying of the RNA stem termini (15). This allows the RNA structure to be sufficiently flexible that it can adopt a conformation compatible with stabilization of sequence-specific, potentially transient RNA-RNA contacts. By binding to multiple RNAs concurrently via surface-exposed positively charged grooves, NSP2 octamers act as matchmakers of complementary sequences, promoting intermolecular RNA–RNA interactions (15). While NSP2 promiscuously binds any ssRNA with sub-nanomolar affinity, the proposed mechanism would require NSP2-RNA interactions to be transient to allow sampling of multiple interacting RNA partners. We propose that on a transcriptome level, NSP2 achieves its RNA chaperone activity by globally increasing RNA backbone flexibility, while its RNA-RNA matchmaking is governed by specific, yet to be revealed, complementary sequences.

Data Availability

Raw sequencing reads are available on the NCBI SRA database under BioProject ID PRJNA791398. All SHAPE data used in this study are available in Supplementary Table 4 in SNRRNASM format.

Funding and Acknowledgements

This work was supported by the Wellcome Trust [103068/Z/13/Z and 213437/Z/18/Z to A.B.], National Institutes of Health grants R35 GM140844 and R01 HL111527 to A.L. A.C. was supported by F31 GM130040-02 through the National Institute of General Medical Sciences.

Funding for open access charge: Wellcome Trust.

References

1. Gallegos, C.O. and Patton, J.T. (1989) Characterization of rotavirus replication intermediates: A model for the assembly of single-shelled particles. *Virology*, **172**, 616–627.
2. Jayaram, H., Estes, M.K. and Prasad, B.V.V. (2004) Emerging themes in rotavirus cell entry, genome organization, transcription and replication. *Virus Res.*, **101**, 67–81.
3. McDonald, S.M. and Patton, J.T. (2011) Assortment and packaging of the segmented rotavirus genome. *Trends Microbiol.*, **19**, 136–144.
4. Silvestri, L.S., Taraporewala, Z.F. and Patton, J.T. (2004) Rotavirus Replication: Plus-Sense Templates for Double-Stranded RNA Synthesis Are Made in Viroplasms. *J. Virol.*, **78**, 7763–7774.
5. Li, W., Manktelow, E., von Kirchbach, J.C., Gog, J.R., Desselberger, U. and Lever, A.M. (2010) Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential cis-acting functions. *Nucleic Acids Res.*, **38**, 7718–7735.
6. Poncet, D., Aponte, C. and Cohen, J. (1993) Rotavirus protein NSP3 (NS34) is bound to the 3' end consensus sequence of viral mRNAs in infected cells. *J. Virol.*, **67**, 3159–3165.
7. Vende, P., Piron, M., Castagné, N. and Poncet, D. (2000) Efficient Translation of Rotavirus mRNA Requires Simultaneous Interaction of NSP3 with the Eukaryotic Translation Initiation Factor eIF4G and the mRNA 3' End. *J. Virol.*, **74**, 7064–7071.
8. Desselberger, U. (2017) Reverse genetics of rotavirus. *Proc. Natl. Acad. Sci.*, **114**, 2106–2108.

9. 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.
10. Madden,E.A., Plante,K.S., Morrison,C.R., Kutchko,K.M., Sanders,W., Long,K.M., Taft-Benz,S., Cruz Cisneros,M.C., White,A.M., Sarkar,S., *et al.* Using SHAPE-MaP To Model RNA Secondary Structure and Identify 3'UTR Variation in Chikungunya Virus. *J. Virol.*, **94**, e00701-20.
11. Alexander Borodavka, Ulrich Desselberger, Ulrich Desselberger, and John T. Patton (2018) Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. *Curr. Opin. Virol.*, 10.1016/j.coviro.2018.08.001.
12. Taraporewala,Z.F. and Patton,J.T. (2001) Identification and Characterization of the Helix-Destabilizing Activity of Rotavirus Nonstructural Protein NSP2. *J. Virol.*, **75**, 4519–4527.
13. Borodavka,A., Dykeman,E.C., Schrimpf,W. and Lamb,D.C. (2017) Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *eLife*, **6**, e27453.
14. Woodson,S.A. (2010) Taming free energy landscapes with RNA chaperones. *RNA Biol.*, **7**, 677–686.
15. Bravo,J.P.K., Bartnik,K., Venditti,L., Acker,J., Gail,E.H., Colyer,A., Davidovich,C., Lamb,D.C., Tuma,R., Calabrese,A.N., *et al.* (2021) Structural basis of rotavirus RNA chaperone displacement and RNA annealing. *Proc. Natl. Acad. Sci.*, **118**.
16. Warf,M.B. and Berglund,J.A. (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.*, **35**, 169–178.
17. Mortimer,S.A., Kidwell,M.A. and Doudna,J.A. (2014) Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.*, **15**, 469–479.
18. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci.*, **106**, 97–102.
19. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
20. Nicholson,B.L. and White,K.A. (2014) Functional long-range RNA–RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.*, **12**, 493–504.
21. Tuplin,A., Struthers,M., Simmonds,P. and Evans,D.J. (2012) A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication. *Nucleic Acids Res.*, **40**, 6908–6921.
22. Smola,M.J., Rice,G.M., Busan,S., Siegfried,N.A. and Weeks,K.M. (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.
23. Richards,J.E., Desselberger,U. and Lever,A.M. (2013) Experimental Pathways towards Developing a Rotavirus Reverse Genetics System: Synthetic Full Length Rotavirus ssRNAs Are Neither Infectious nor Translated in Permissive Cells. *PLoS ONE*, **8**, e74328.

24. Diebold,O., Gonzalez,V., Venditti,L., Sharp,C., Blake,R.A., Tan,W.S., Stevens,J., Caddy,S., Digard,P., Borodavka,A., *et al.* (2022) Using Species a Rotavirus Reverse Genetics to Engineer Chimeric Viruses Expressing SARS-CoV-2 Spike Epitopes. *J. Virol.*, **0**, e00488-22.
25. Busan,S. and Weeks,K.M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA N. Y. N.*, **24**, 143–148.
26. Low,J.T. and Weeks,K.M. (2010) SHAPE-Directed RNA Secondary Structure Prediction. *Methods San Diego Calif*, **52**, 150–158.
27. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
28. Thiel,B.C., Beckmann,I.K., Kerpedjiev,P. and Hofacker,I.L. (2019) 3D based on 2D: Calculating helix angles and stacking patterns using forgi 2.0, an RNA Python library centered on secondary structure elements. *Fl000Research*, **8**, ISCB Comm J-287.
29. Lorenz,R., Bernhart,S.H., Höner zu Siederdisen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
30. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
31. Shannon,C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.*, **27**, 623–656.
32. 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.
33. Levenshtein,V.I. (1965) Binary codes capable of correcting deletions, insertions, and reversals. *Sov. Phys. Dokl.*, **10**, 707–710.
34. Matthijnssens,J., Ciarlet,M., Heiman,E., Arijs,I., Delbeke,T., McDonald,S.M., Palombo,E.A., Iturriza-Gómara,M., Maes,P., Patton,J.T., *et al.* (2008) Full Genome-Based Classification of Rotaviruses Reveals a Common Origin between Human Wa-Like and Porcine Rotavirus Strains and Human DS-1-Like and Bovine Rotavirus Strains. *J. Virol.*, **82**, 3204–3219.
35. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
36. Busan,S., Weidmann,C.A., Sengupta,A. and Weeks,K.M. (2019) Guidelines for SHAPE Reagent Choice and Detection Strategy for RNA Structure Probing Studies. *Biochemistry*, **58**, 2655–2664.
37. 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.
38. Kutchko,K.M., Madden,E.A., Morrison,C., Plante,K.S., Sanders,W., Vincent,H.A., Cruz Cisneros,M.C., Long,K.M., Moorman,N.J., Heise,M.T., *et al.* (2018) Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Res.*, **46**, 3657–3670.
39. Smola,M.J. and Weeks,K.M. (2018) In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.*, **13**, 1181–1195.
40. Dadonaite,B., Gilbertson,B., Knight,M.L., Trifkovic,S., Rockman,S., Laederach,A., Brown,L.E., Fodor,E. and Bauer,D.L.V. (2019) The structure of the influenza A virus genome.

Nat. Microbiol., **4**, 1781–1789.

41. Woods,C.T., Lackey,L., Williams,B., Dokholyan,N.V., Gotz,D. and Laederach,A. (2017) Comparative Visualization of the RNA Suboptimal Conformational Ensemble In Vivo. *Biophys. J.*, **113**, 290–301.
42. Lackey,L., Coria,A., Woods,C., McArthur,E. and Laederach,A. (2018) Allele-specific SHAPE-MaP assessment of the effects of somatic variation and protein binding on mRNA structure. *RNA N. Y. N*, **24**, 513–528.
43. Yang,A.-D., Barro,M., Gorziglia,M.I. and Patton,J.T. (2004) Translation enhancer in the 3'-untranslated region of rotavirus gene 6 mRNA promotes expression of the major capsid protein VP6. *Arch. Virol.*, **149**, 303–321.
44. Bravo,J.P.K., Borodavka,A., Barth,A., Calabrese,A.N., Mojzes,P., Cockburn,J.J.B., Lamb,D.C. and Tuma,R. (2018) Stability of local secondary structure determines selectivity of viral RNA chaperones. *Nucleic Acids Res.*, **46**, 7924–7937.
45. Dethoff,E.A., Boerneke,M.A., Gokhale,N.S., Muhire,B.M., Martin,D.P., Sacco,M.T., McFadden,M.J., Weinstein,J.B., Messer,W.B., Horner,S.M., *et al.* (2018) Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci.*, **115**, 11513–11518.
46. Schlick,T., Zhu,Q., Dey,A., Jain,S., Yan,S. and Laederach,A. (2021) To Knot or Not to Knot: Multiple Conformations of the SARS-CoV-2 Frameshifting RNA Element. *J. Am. Chem. Soc.*, **143**, 11404–11422.
47. Huston,N.C., Wan,H., Strine,M.S., de Cesaris Araujo Tavares,R., Wilen,C.B. and Pyle,A.M. (2021) Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell*, **81**, 584-598.e5.
48. Simon,L.M., Morandi,E., Luganini,A., Gribaudo,G., Martinez-Sobrido,L., Turner,D.H., Oliviero,S. and Incarnato,D. (2019) In vivo analysis of influenza A mRNA secondary structures identifies critical regulatory motifs. *Nucleic Acids Res.*, **47**, 7003–7017.
49. Acs,G., Klett,H., Schonberg,M., Christman,J., Levin,D.H. and Silverstein,S.C. (1971) Mechanism of Reovirus Double-Stranded Ribonucleic Acid Synthesis In Vivo and In Vitro. *J. Virol.*, **8**, 684–689.
50. Desselberger,U., Richards,J., Tchertanov,L., Lepault,J., Lever,A., Burrone,O. and Cohen,J. (2013) Further characterisation of rotavirus cores: Ss(+)RNAs can be packaged in vitro but packaging lacks sequence specificity. *Virus Res.*, **178**, 252–263.
51. Haralampiev,I., Prisner,S., Nitzan,M., Schade,M., Jolmes,F., Schreiber,M., Loidolt-Krüger,M., Jongen,K., Chamiolo,J., Nilson,N., *et al.* (2020) Selective flexible packaging pathways of the segmented genome of influenza A virus. *Nat. Commun.*, **11**, 4355.
52. Gog,J.R., Afonso,E.D.S., Dalton,R.M., Leclercq,I., Tiley,L., Elton,D., von Kirchbach,J.C., Naffakh,N., Escriou,N. and Digard,P. (2007) Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Res.*, **35**, 1897–1907.

Figure Legends

Figure 1. The Rotavirus transcriptome exhibits a diverse ensemble of RNA structures.

(A) Median SHAPE reactivity of each of the 11 transcripts representing the rotavirus transcriptome averaged over two highly correlated ($R^2 \sim 0.73 - 0.98$) replicates. SHAPE-MaP experiments were carried out by incubating equimolar amounts of each RV segment transcript and probed with 5NIA, as described in Materials and Methods. A 50-nucleotide window was used to calculate median SHAPE values along each transcript and normalized to the overall median of the individual transcript. Areas of high median SHAPE (i.e., above zero, shown in green) represent conformationally flexible RNA regions. Areas of low median SHAPE (i.e., below zero, shown in blue) represent regions with high propensity to base-pair. Arc plots show pseudo-free energy derived base-pairing probabilities computed using nearest-neighbor RNAstructure parameters constrained by incorporating SHAPE reactivities (18, 28). Probabilities of base-pairing are color-coded, with green arcs corresponding to the most probable base-pairs, and grey being the least probable ones. Boxed regions indicate conserved nucleotides. Conformational entropy plots (median Shannon entropy calculated with a 50-nucleotide sliding window) computed for RV transcripts are plotted underneath median SHAPE values. High Shannon entropy regions typically correspond to positions with alternative conformations, while low entropy regions identify sequences that are more prone to adopting a dominant MFE configuration.

Figure 2. SHAPE-MaP derived models support covariation-based secondary structure models.

(A) Median SHAPE reactivity of segment 6 in the presence of the RV transcriptome. Median SHAPE profile and base-pairing probabilities were calculated as described in Figure 1. Grey box indicates two terminal stem-loops that were predicted in segment 6 by Li *et al* (5), and are further confirmed by our SHAPE-informed secondary structure model. Beneath the median SHAPE profile is a conservation score heatmap plot calculated using a 10-nt rolling window as described in Materials and Methods. (B) Median SHAPE-MaP profile and base-pairing probabilities of segment 10 in the presence of the RV transcriptome. Median SHAPE profile and base-pairing probabilities were calculated as described in Figure 1. Grey box highlights three stem-loops modelled in the 3' terminal region of segment 10, as previously predicted (5). Additionally, these three loops are located in a highly-conserved region of segment 10 as indicated by the high conservation score. (C) SHAPE-MaP profile of segment 11 RNA shown along with the conservation heatmap plot (*left*), and the corresponding models of conserved RNA helices H2 and H3 (*right*) supported by the SHAPE-MaP data. Highly conserved nucleotides were previously predicted to participate in LRIs between terminal regions and our SHAPE-informed MFE model validates these predictions (5, 53).

Figure 3. SHAPE-MaP analysis of *in vitro* transcribed segment 11 RNA *versus* in-cell transcript probing during RV infection. (A) Top panel: SHAPE reactivities of segment 11 transcript probed with 5NIA along with RV transcriptome. Bottom panel: in-cell SHAPE reactivity profile of segment 11 RNA probed with 5NIA at 6 hours post infection, as described in Materials and Methods. Individual bars represent SHAPE reactivities at single nucleotide resolution, black bars indicate nucleotides with the lowest reactivities (< 0.4) and red bars denote nucleotides with the highest reactivities (> 0.8). (B) Overlaid arc diagrams show similarities and differences between base-pairing in segment 11 RNA in the context of the complete RV transcriptome *in vitro* (Top) *vs* in the context of RV-infected cells (Bottom). Minimum free energy arc diagrams were generated by incorporating SHAPE-MaP data as a pseudo-free energy term in RNAstructure, as described in Figure 1. (C) Left: Scatterplot correlation analysis reveals that segment 11 transcript SHAPE reactivities between the *in cell* and *in vitro transcribed* conditions are highly correlated ($R^2 = 0.8$). Right: Violin plots showing distribution of SHAPE reactivities of segment 11 RNA when probed *in cell* or *in vitro transcribed*. Boxes represent the 25th/75th interquartile range, and medians are shown as central bands. SHAPE reactivities are not significantly different as assessed by Kruskal-Wallis test ($p < 0.05$).

Figure 4. NSP2 addition does not result in detectable changes in the median SHAPE reactivities of segment 11 RNA. (A) Segment 11 was incubated with increasing amounts of NSP2 (0 μ M, 5 μ M, 10 μ M, and 20 μ M) then probed with 5NIA, as described in the Methods section. SHAPE reactivities for each of the listed titration points show no obvious difference due to the normalization protocol. (B) Scatterplots showing the linear relationship between segment 11 SHAPE reactivity (x axis) and segment 11 SHAPE reactivity when incubated with increasing amounts of NSP2 (y axis). Note high correlation between the data ($R^2 = 0.96$, $R^2 = 0.78$), and the linear relationship exhibits no significant change between the titration points. (C) Violin plots comparing distribution of SHAPE reactivities of segment 11 RNA upon incubation with NSP2. Boxes represent the 25th/75th interquartile range, and medians are shown as central bands. Addition of 20 μ M NSP2 changes the distribution of SHAPE reactivities, while the median SHAPE reactivity does not significantly change, as assessed by the Kruskal-Wallis test ($p < 0.05$). Note there is no significant difference between the conditions despite a significant global increase in mutation rate. (D) Raw mutation rates of segment 11 RNA structure probing experiment. The red line is representative of segment 11 RNA mutation rates induced by the incubation with 12.5 mM of the modifying reagent, 5-nitroisatoic anhydride (5NIA), and the blue is the measured mutation rate across the segment 11 transcript treated with dimethyl sulfoxide (DMSO) as a control. Normalization of these mutation rates results in the SHAPE profiles shown in panel A. (E) Histogram representation of distribution of nucleotide mutation rates in treated and untreated samples of segment 11 RNA shown in panel D. (F) Scatterplot showing linear relationship of mutation rate data between reproducible replicate experiments. The slope when comparing both replicates is near 1, showing that mutation rate data are highly correlated ($R^2 = 0.74$, slope = 1.08).

± 0.03). (G) Scatterplot showing linear relationship of SHAPE-MaP data between reproducible replicate experiments. SHAPE-MaP data are highly correlated ($R^2=0.93$, slope = 0.94 ± 0.02).

Figure 5. NSP2 uniformly increases segment 11 RNA mutation rate in a concentration- dependent manner. (A) Segment 11 was incubated with increasing amounts of NSP2 (0 μ M, 5 μ M 10 μ M, and 20 μ M), then probed with 5NIA, as described in Materials and Methods. Higher mutation rates reveal increased flexibility in the RNA backbone, while lower mutation rates indicate less conformationally flexible nucleotides. Note a uniform increase in mutation rates proportional to NSP2 concentration. Zoomed-in: representative regions of segment 11 RNA. (B) Violin plots comparing distribution of mutation rates of segment 11 RNA upon incubation with NSP2. Boxes represent the 25th/75th interquartile range, and medians are shown as central bands. Note, both change in mutation rate distribution and significant change in median mutation rate. Significance values were calculated using Kruskal-Wallis test ($p < 0.05$). (C) Mutation rates of segment 11 RNA (x-axis) plotted against mutation rates of segment 11 RNA in the presence of increasing amounts of NSP2 to reveal linear relationship with distinct slopes. The slope increase as NSP2 is titrated in, while the linear dependence persists, suggesting a global change in the RNA flexibility in response to saturation with NSP2.

Figure 6. NSP2 uniformly increases backbone flexibility in RV RNAs. (A-C) Mutation rates of individual RNAs (segment 5, segment 6, and segment 10, respectively) alone (x axis) plotted against the mutation rates of the RNAs incubated with 20 μ M NSP2 (y axis). (D) Schematic of structural motifs found within an RNA. Using SHAPE-MaP-informed secondary structure models generated as shown in Figure 1, individual nucleotides (color-coded) were categorized into each of the different structural modalities described: interior bulge (red), multiloop (yellow), stem (green), and hairpin loop (orange) using the Forgi library (Materials and Methods). (E) NSP2-mediated mutation rate change analysis for individual structural motifs as described in panel D. $\text{Log}_2(\text{FC})$ denotes $\log_2(\text{mutation rates induced by 20 } \mu\text{M NSP2 divided by the mutation rates without NSP2})$. A vast majority of $\log_2(\text{FC})$ values are positive. Statistical significance was assessed using Kruskal-Wallis test. Stem-forming nucleotides involved in base-pairing have a higher average $\log_2(\text{FC})$ than those located within predicted forming hairpins, multiloops and interior bulges.

Figure 7. Changes in global backbone flexibility in RV transcriptome mediated by NSP2. Addition of NSP2 to an equimolar mix of 11 distinct RV transcripts results in global increase of mutation rates across all segments. (A-D): Mutation rates of individual RNAs incubated with an RNA mix representing the RV transcriptome in the presence of 20 μ M NSP2(y-axis) plotted against mutation rates of the RNA alone in the presence of 20 μ M of NSP2 (x-axis) to reveal highly correlated data. (E): Analysis of NSP2-mediated mutation rate changes in the presence or absence of the rotavirus transcriptome. Comparison of distribution of NSP2 mediated mutation rate changes (i.e., the log2 of the quotient of the mutation rates in the presence of 20 μ M NSP2 over the mutation rates for RNA without NSP2) for each the structural motif (stem, hairpin loops, multiloops, interior bulge, as described in Figure 5D) for RNAs alone and RNAs in the context of the transcriptome. Q1 and Q3 represent the 25th and 75th interquartile range, respectively, and the median, Q2, is labeled shown as solid lines. Statistical significance was assessed using Kruskal-Wallis test.

Tables and Table Legends

Segment	5μM NSP2/0μM NSP2		10μM NSP2/0μM NSP2		20μM NSP2/0μM NSP2	
	Slope	R ²	Slope	R ²	Slope	R ²
<i>Segments probed in the context of RV Transcriptome</i>						
Segment 1	2.1 ± 0.01	0.84	3.3 ± 0.03	0.67	5.4 ± 0.06	0.54
Segment 2	2.3 ± 0.02	0.81	3.3 ± 0.04	0.70	5.4 ± 0.07	0.55
Segment 3	2.3 ± 0.02	0.90	3.2 ± 0.03	0.79	4.5 ± 0.06	0.69
Segment 4	2.2 ± 0.02	0.92	3.3 ± 0.03	0.81	5.5 ± 0.07	0.63
Segment 5	2.4 ± 0.02	0.90	3.0 ± 0.03	0.88	4.2 ± 0.06	0.83
Segment 6	2.1 ± 0.02	0.84	2.9 ± 0.05	0.73	5.9 ± 0.1	0.44
Segment 7	2.0 ± 0.03	0.88	2.9 ± 0.05	0.79	5.7 ± 0.1	0.74
Segment 8	2.1 ± 0.03	0.97	2.9 ± 0.04	0.96	4.9 ± 0.1	0.86
Segment 9	2.1 ± 0.04	0.84	3.0 ± 0.04	0.79	4.2 ± 0.07	0.73
Segment 10	2.2 ± 0.05	0.80	2.8 ± 0.08	0.65	5.9 ± 0.2	0.59
Segment 11	1.9 ± 0.04	0.88	2.8 ± 0.08	0.73	6.6 ± 0.2	0.53
<i>Segments probed alone</i>						
Segment 5	2.9 ± 0.03	0.87	2.8 ± 0.03	0.88	4.8 ± 0.05	0.69
Segment 6	2.4 ± 0.04	0.72	3.7 ± 0.06	0.56	4.8 ± 0.1	0.51
Segment 10	2.1 ± 0.04	0.70	3.3 ± 0.08	0.42	3.9 ± 0.1	0.54
Segment 11	2.6 ± 0.07	0.84	4.0 ± 0.1	0.67	4.6 ± 0.1	0.65

Table 1. NSP2 increases mutation rates indiscriminately across all RV RNAs.

Slopes indicate NSP2 induced increases in mutation rate for each segment calculated by dividing the mutation rate at each point in the titration experiment by the mutation rate of the segment without NSP2 as measured by 5NIA chemical probing. As NPS2 concentration increases, as does the slope. We calculated slope values for the entire RV transcriptome at increasing concentrations of NSP2 and for Segments 5, 6, 10, and 11 alone.