The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures

Highlights
- The genome of HCV is folded into specific RNA structures
- HCV genomic structures are conserved across multiple genotypes
- Elaborate RNA structures are present within protein-coding sequences
- Genetic manipulation of these structures affects replication and infectivity

Authors
Nathan Pirakitikulr, Andrew Kohlway, Brett D. Lindenbach, Anna M. Pyle

Correspondence
anna.pyle@yale.edu

In Brief
Using biochemistry and viral genetics, the RNA genome of HCV was shown to contain numerous well-conserved RNA structures within protein-coding regions. Genetic manipulation of these structures alters the ability of HCV to replicate and infect, thereby demonstrating that RNA sequences can store multiple layers of information.
The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures

Nathan Pirakitikul,1 Andrew Kohlway,1 Brett D. Lindenbach,2 and Anna M. Pyle1,3,4,*
1Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06511, USA
2Department of Microbial Pathogenesis, Yale University, New Haven, CT 06519, USA
3Department of Chemistry, Yale University, New Haven, CT 06511, USA
4Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
*Correspondence: anna.pyle@yale.edu
http://dx.doi.org/10.1016/j.molcel.2016.01.024

SUMMARY

RNA is a versatile macromolecule that accommodates functional information in primary sequence and secondary and tertiary structure. We use a combination of chemical probing, RNA structure modeling, comparative sequence analysis, and functional assays to examine the role of RNA structure in the hepatitis C virus (HCV) genome. We describe a set of conserved but functionally diverse structural RNA motifs that occur in multiple coding regions of the HCV genome, and we demonstrate that conformational changes in these motifs influence specific stages in the virus’ life cycle. Our study shows that these types of structures can pervade a genome, where they play specific mechanistic and regulatory roles, constituting a “code within the code” for controlling biological processes.

INTRODUCTION

Over 170 million people are infected with hepatitis C virus (HCV) worldwide. In the United States, Europe, and Japan, HCV-related cirrhosis is the leading indication for liver transplant. New targeted therapies developed over the past two decades now offer hope of curing HCV for many. Yet in the shadow of these clinical advances, questions about the fundamental biology of the virus remain.

HCV is a positive-sense RNA virus. The viral genome (9.6 kb) encodes a single large open reading frame (ORF), producing an ~3,000 aa polypeptide that is co- and posttranslationally cleaved into ten proteins. The first three, core, E1, and E2, are structural; the last five, NS3 (helicase/protease), NS4A, NS4B, NS5A, and NS5B (RNA-dependent RNA polymerase), form the replication complex (Bartenschlager et al., 2004).

RNA structures are critical to the entire viral life cycle, and the full genome is known to support a high degree of internal base pairing (Davis et al., 2008; Simmonds et al., 2004). The search for unique RNA elements within the HCV genome has spanned the better part of two decades. By using bioinformatics tools and RNase mapping, several groups have contributed to the identification of RNA structures in untranslated regions (UTRs) as well as the core- and NS5B-encoding regions (Fricke et al., 2015; Tuplin et al., 2004). Many of these structures have since been validated by mutational analysis in cell-culture models of HCV replication and infectivity, and they have helped assign functional roles for specific RNA structural elements (Diviney et al., 2008; Friebie and Bartenschlager, 2002; Friebie et al., 2001, 2005; Kolykhalov et al., 2000; Lee et al., 2004; McMullan et al., 2007; Oakland et al., 2013; Vassilaki et al., 2008; You and Rice, 2008; You et al., 2004). However, these methodologies have been less successful in identifying and characterizing RNA structures that occur elsewhere in the HCV genome (Chu et al., 2013).

Recently, Mauger et al. performed a comprehensive chemical probing of RNA structure on three HCV genomes (Mauger et al., 2015). This study confirmed that a high degree of folding occurs across the HCV genome, including coding regions. Further mutational analysis of highly conserved regions revealed that disrupting select structures affected viral fitness in cell culture.

In this study, we report the identification of functional RNA structures within the HCV genome by using sequential application of techniques that include selective ²-hydroxyl acylation analyzed by primer extension (SHAPE) chemical probing (Merino et al., 2005), comparative analysis that encompasses both primary sequence and secondary structure (Nawrocki and Eddy, 2013), and viral genetics (Figure 1A). By using this modified approach, we identified a large complementary set of biologically functional, conserved RNA elements that occur within all strains of HCV. Many of these RNA structures, and the dynamic interplay among them, modulate key stages in the viral life cycle.

RESULTS

An Approach for Discovering RNA Structures in Protein-Coding RNA

To predict functional RNA structure within the HCV genome, we developed a strategy that incorporates SHAPE, comparative sequence analysis and functional characterization by mutagenesis, and viral genetics.

In vitro RNA transcripts of functional, full-length HCV genomic RNA were folded in buffers that approximate physiological salt and pH—150 mM KCl and 5 mM MgCl₂ (pH 7.4) (Deigan et al., 2009; Kieft et al., 1999)—and determined to be monodisperse by size-exclusion chromatography (Figure S1A, available online).
Figure 1. Genome-wide Analysis of HCV RNA Structure

(A) Experimental overview. HCV RNA was in vitro transcribed and folded, then modified with NAI. Reverse transcription (RT) of modified and unmodified RNA generated fluorescent cDNA fragments that were analyzed by capillary electrophoresis (CE) to provide a SHAPE reactivity map. Reactivities were used to calculate candidate secondary structures. We performed covariation analyses to identify conserved structures, then examined functional relevance of structures by reverse genetics.

(B) Summary of SHAPE reactivity and covariation analysis. Regions of low reactivity (< 0.7) are colored blue, high reactivity (> 0.7) are red, and regions missing SHAPE data are gray. Each bar represents two nucleotides. Yellow boxes denote regions of the genome containing RNA motifs with covarying mutations present across all HCV genotypes.

(C) Positions of peptidase/protease cleavage sites in translated polyprotein (triangles).

(D) SHAPE reactivity profiles of regions containing highly conserved RNA motifs (average of two datasets). For reference, a dotted line is drawn through a reactivity of 0.7.
SHAPE reactivity was incorporated into the calculation of candidate secondary structures for overlapping 1 kb segments that span the genome (for a representative map, see Figure S1B).

The RNA secondary structures modeled exclusively from SHAPE reactivity were used as a starting point for comparative sequence analysis of complete HCV genomes in the NCBI database. To perform this analysis, we built preliminary covariance models based on a sequence-based alignment of genotype 2 viruses and the secondary-structure models generated from our analysis of the Jc1 strain. Previous analyses of other large RNAs also employed an approach involving alignment by primary sequence conservation, but we expanded this approach to include information about sequence covariance, enabling us to build an alignment composed of more divergent sequences. Borrowing a strategy developed to identify highly structured RNA motifs in shorter RNAs (El Korbi et al., 2014), sequences of 1,125 HCV genomes (Table S1) were incorporated into the covariation model based on structural similarity. By aligning divergent sequences based on RNA structure, this approach identifies conserved RNA structures even if they occur at different positions in the genome. We next focused on identifying conserved substructures by searching for positions with multiple consecutive covarying nucleotides. To this end, we employed the software package R2R (Weinberg and Breaker, 2011), which is commonly used to annotate riboswitches and other locally well-defined RNA structures. We found that within a given structured region, certain structural features (e.g., specific base pairs, nucleotide sequences, and bulges) were more conserved than others. Therefore, we targeted these sites for mutagenesis and functional characterization.

To preserve the amino acid sequence, only synonymous codons were incorporated in the mutant constructs. Additionally, we avoided rare codons to minimize effects on protein expression. This marked limited the number of mutable positions and available nucleotide changes. Notably, the constraints faced in nature are less stringent: at many positions the amino acid is poorly conserved, which permits multiple nucleotide combinations that covary between paired positions.

The Coding Region of HCV RNA Is Extensively Folded

We found that the coding region of the HCV RNA is extensively folded (Figure 1B), containing specific substructures that appear in all genotypes of the virus. Regions of low SHAPE reactivity (≤ 0.7 normalized reactivity, constrained) (Low et al., 2014) are distributed across the entire genome (Figure 1B and Table S4). Specifically, we identified 17 well-conserved stem loops spread across 9 protein-coding regions (numbered boxes; Figures 1B and 1D).

Six of the motifs (SL833, SL1412, SL2531, SL2549, SL6038, and SL8001) occur in regions not previously described to contain RNA structures (Figures 1B, 1D, and S2). Elsewhere, our structures are in agreement with descriptions found in the literature (Chu et al., 2013; Diviney et al., 2008; Mauger et al., 2013; McMullan et al., 2007; Tuplin et al., 2004), but our analysis enables us to describe features of these elements in far greater detail. These structures include SL427, SL588, SL669, SL761, SL783, SL8670, SL9198, SL9294, SL9326, and SL9389 (Figure S2).

We chose five well-conserved substructures (SL588, SL783, SL1412, SL6038, and SL8001) from different regions of the genome to study in depth. We further subdivided SL588 and SL783 into sections (A, B, and C) based on the extent of covariation observed. SL588A and SL783A (Figure 2A) constitute the most highly conserved segments, and we specifically targeted these sections to pinpoint functional structural features from within the highly folded architecture of the genome. To validate the function of individual substructures, we introduced mutations that changed RNA structure without changing protein sequence and measured their effect on viral replication and infectivity. We used a full-length Jc1 chimera (Table S3) with a Gaussia princeps luciferase (GLuc) reporter that is highly infectious in cell culture (Phan et al., 2009). The reporter is inserted between the p7 and NS2 genes, which are distant from the RNA structures being studied, and has minimal effects on viral replication and virus assembly.

At least two sets of mutations were designed for each structure. The need to maintain the conserved amino acid sequence prevented us from generating directly compensatory mutations to restore broken base pairs. Instead, we confirmed RNA structures by introducing mutations that preserved RNA structural features but changed the primary sequence, thereby demonstrating that the effect is functionally independent of primary sequence. Therefore, “unzip” mutations replaced Watson-Crick paired bases with nonpairing bases; “lock” mutations stabilized putative structures by replacing nonpaired bases with pairing bases. At SL783A and SL8001, in lieu of lock mutations, we “recoded” the sequence to maintain the same base-pair configuration found in the original structure. Where multiple conformations were possible, we introduced mutations that thermodynamically favor one form over others. We further avoided codons under-represented in the wild-type (WT) virus (Table S2). We confirmed that mutations induce the predicted structural changes by performing SHAPE analysis on each of the mutant constructs (Figures S3 and Table S4). Apart from SL783A, each substructure resulted in some degree of change to replication or infectivity when either unzipped or locked.

A Network of Functional RNA Structures within the Core-Encoding Region

The region that encodes the core protein is notable because it contains a high density of RNA structures. Our in vitro SHAPE data indicate that ~90% of the 573 nucleotides within this region are base paired. To confirm that these structures were also present in vivo, we performed DMS probing on actively replicating viral RNA in infected cells. We specifically examined the highly structured region spanning SL588, SL669, SL761, and SL783 and found the in vivo DMS reactivity to be in good agreement with predicted structures (Figures 2A and S4A). To functionally characterize these structures, we next focused on areas with the largest number of covarying base pairs: SL588A and SL783A (Figure 2A). In previous studies of the core region, mutations were introduced in segments containing SL588 and SL783 but were not targeted to individual substructures such as SL588A and SL783A (Mauger et al., 2015; Vassilaki et al., 2008).

In our study, unzip mutations of SL588A abolished viral replication, and lock mutations restored replication, modestly
Figure 2. Conserved RNA Motifs in the Core-Encoding Region

(A) Covariation analysis of core-encoding region RNA structures. Kissing interaction highlighted in red. Nucleotides are colored by degree of conservation. Base pairs are colored based on the presence of covarying mutations. ΔGs (kcal/mol) are reported for the tested structures calculated by thermodynamic modeling of each full stem loop. Dashes mark every tenth nucleotide.

(B) SL588A
- unzip
- lock
- unzip
- recode

SL783A
- ΔG -30.6 kcal/mol
- -50.1
- -3.80
- -22.1

ΔG -27.2

SL427

ΔG -47.1

(D) Replication

(E) Replication

(log(RU) vs wt, Δcore, GNN, mock, unzip, lock, unzip, recode)

Infectivity

(log(RU) vs wt, Δcore, GNN, mock, ΔG +, ΔG -)

Infectivity

(legend continued on next page)
increasing the rate above WT (p = 0.13) (Figures 2B and 2D). Mutations to SL783A (Figure 2B) had no effect on replication, infectivity (Figure 2D), or early replication (Figure S4B). These results indicate that SL588A behaves as a cis-acting replication element (CRE) in the context of the full-length genome. But it is important to note that virus constructs lacking the coding sequences for core protein (Δcore strains, which lack nucleotides 525–788) are able to replicate, which suggests that SL588 exerts its influence over viral replication in the context of an intact system, potentially by acting in concert with surrounding RNA structures.

To understand how SL588A might interface with neighboring RNA structures, we evaluated the propensity of the core-encoding region to form alternate base-pair configurations and long-range interactions. Close examination of the SHAPE reactivity map (Figure S1B) revealed loop regions of low reactivity (e.g., nucleotides 459–461 and 638–639), suggesting that the nucleotides in these regions are structurally constrained. We therefore searched the neighboring regions for complementary sequences, identifying a match between the terminal loop of SL427A and the 3' end of SL588C (red; Figure 2A). Further covariation analysis revealed that this putative “kissing loop” interaction was conserved in all genotypes.

To test this interaction, we designed the following mutants: (1) the “(−)-kiss” mutant that disrupts the kissing interaction but maintains intact stems SL427 and SL588, (2) the “(+)-kiss” mutant that strengthens the kissing interaction, and (3) the “(−)-stem” mutant that weakens stems in SL427 and SL588 (Figure 2C). The “(−)-kiss” mutant modestly slowed replication and drastically lowered infectious virus production. The “(+)-kiss” mutant abolished replication entirely. The “(−)-stem” mutant had no effect on either replication or infectivity (Figure 2E). Although the “(−)-kiss” mutant did not reach WT levels of replication, it behaved similarly to the noninfectious Δcore mutant, as expected (Figure S4C). The data indicate that there are two functional states of SL427 and SL588. In the first state, where the two stems do not interact (tested with the “(−)-kiss” mutant), replication proceeds, but infectious virus was severely compromised. In the second state, where the two stems are capable of forming a tertiary interaction, replication halts (tested with the “(+)-kiss” mutant). Taken together, these findings indicate that there are distinct and mutually exclusive RNA structures that facilitate replication and infectivity.

RNA Structures within the NS4B- and NS5B-Encoding Regions Tune the Efficiency of Replication

Based on chemical probing and covariation analysis of the NS4B-encoding region (Figure 3A), we observe that SL6038 can adopt two possible structural states: a stem (ΔG = −104 kcal/mol) and a cloverleaf (ΔG = −91.8 kcal/mol). To probe this, we introduced mutations that would favor one or the other of these conformations. Silent mutations that favored the SL6038 stem conformation (“stemlock”; ΔG(stem) = −118 kcal/mol, ΔG(cloverleaf) = −96.6 kcal/mol) abolished replication, whereas mutations that destabilized SL6038 stem formation (“unzip”; ΔG = −45 kcal/mol) had modest effects (Figures 3B and 3C). Indeed, one silent U-to-G substitution at position 6181 (“C-G,” ΔG(stem) = −109 kcal/mol, ΔG(cloverleaf) = −97.3 kcal/mol), which replaces a conserved C-U mismatch with a C-G pair, was sufficient to inhibit replication (Figures 3B and 3D). Moreover, this inhibition was observed only when the stem conformation is formed, and it was not observed in the presence of mutations (“cloverleaf;” ΔG(stem) = −92.7 kcal/mol, ΔG(cloverleaf) = −98.9 kcal/mol) that favored the cloverleaf conformation (Figures 3B and 3E). This suggests that the RNA element behaves as a toggle whereby the stem conformation inhibits replication and the cloverleaf conformation functions as a conserved device for activating the replication-competent state (Figure 3A).

Although SL8001 was not required for overall replication (Figure 4A), conformational changes in this motif tune the initial rate of replication. For example, unzip mutations in SL8001 increased the rate of replication, particularly at early time points (Figures 4B and 4C). By 32 hr, the SL8001 unzip mutant expressed significantly more luciferase than WT (p = 0.005). By contrast, recode mutations in SL8001 slowed replication to WT levels (Figure 4C). Thus, SL8001 may act as a governor to slow down the initial steps in replication, and its disruption may have adverse effects on viral fitness not obvious in a cell-culture assay.

RNA Structures Influence the Rate of Infectious Virus Production

Of the structures studied, only SL1412 in the E1-encoding region (Figure 4E) was found to specifically affect infectivity by suppressing the production of infectious virus. Unzip mutations to SL1412 increased infectivity, while lock mutations decreased infectivity relative to WT (Figures 4F and 4G). Although both unzip and lock mutants replicated slightly faster than WT (p = 0.08 and 0.65), it is notable that the mutants had drastically different effects on infectivity (p = 0.025). This provides evidence that individual RNA motifs do not simply exert general effects on viral fitness. Rather, RNA conformers are capable of influencing specific steps of the viral program independently. Thus, the location of SL1412 in the genome and its unique structural attributes help to mediate its specific effects on infectivity.

DISCUSSION

In this study, we show that HCV RNA contains extensive functional information within its primary sequence and its secondary
Figure 3. Conformational Sampling of RNA Structure in NS4B-Encoding Region

(A) Covariation analysis of SL6038 in NS4B-encoding region. Structures are annotated as in Figure 2.

(B) SL6038 mutant designs. The SL6038 unzip construct contains mutations that extend beyond the predicted structure in order to prevent alternate base pairs from forming within the region.

(C and D) Replication and infectivity time courses for SL6038 mutants (C) and SL6038 mismatch mutants in stem versus cloverleaf context (D). Data are represented as the mean of three experiments ± SEM.

Please cite this article in press as: Pirakitikul et al., The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures, Molecular Cell (2016), http://dx.doi.org/10.1016/j.molcel.2016.01.024
Figure 4. RNA Motifs within the NS5B- and E1-Encoding Regions

(A) Covariation analysis of SL8001 in NS5B-encoding region. Structures are annotated as in Figure 2.

(B) SL8001 mutant designs.

(E) SL1412 structures.

(F) SL1412 mutant designs.

(C and D) Replication assays for wild-type (wt), Δcore, GNN, and mock samples with unzip and lock conditions.

(G) Infectivity assays for wild-type (wt), Δcore, GNN, and mock samples with unzip and lock conditions.

Please cite this article in press as: Pirakitikul et al., The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures, Molecular Cell (2016), http://dx.doi.org/10.1016/j.molcel.2016.01.024

(legend continued on next page)
and tertiary structure. We demonstrate that individual structures, such as SL427 and SL588, interact to direct the viral genome through steps of the viral program. We also show how HCV can use conformational changes of an RNA structure (SL6038) to regulate replication. Most remarkably, we demonstrate that these phenomena occur in an RNA that simultaneously codes for protein. Owing to redundancy in the genetic code, we can then tune viral replication and infectivity by selectively stabilizing and destabilizing RNA substructures without affecting the coding of viral proteins.

These discoveries were the direct result of a search strategy that combined single-nucleotide chemical probing of RNA with detailed comparative genomic analysis of RNA structure. Incorporating a large set of diverse sequences based on structural similarity enabled us to pinpoint important structural features from the overall RNA fold and predict biologically relevant structures with high probability. Of the five RNA structures we characterized genetically, four had quantifiable biological effects on viral fitness.

Although deciphering precise mechanisms for the function of these RNA structures will require further analysis, we provide evidence that multiple pathways are involved. For example, formation of one substructure increases replication (SL588A), whereas another RNA substructure suppresses replication (SL6038). For SL6038, we observed a high tolerance for sequence and major structural changes, which argues against a process that requires site-specific protein recognition. Elsewhere, most notably in the core-encoding region, the mechanism of action likely involves the selective recruitment and exclusion of specific protein factors. Incorporating a large set of diverse sequences based on structural similarity enabled us to pinpoint important structural features from the overall RNA fold and predict biologically relevant structures with high probability. Of the five RNA structures we characterized genetically, four had quantifiable biological effects on viral fitness.

The effects of conformational changes in RNA structure can be absolute, as seen with SL588A and SL6038, or incremental, as seen with SL1412 and SL8001. This latter observation suggests that key decisions, such as whether to package or replicate, may ultimately hinge upon an accumulation of signals, some of which are present as RNA substructures embedded in protein-coding sequence. SL1412 can modulate the infectivity of the virus, potentially by increasing the rate of viral packaging or promoting efficient viral entry when unraveled. So far it is the only single RNA structure in HCV that has been shown to affect infectivity independently of replication, and further characterization of this element may reveal additional RNA features that determine viral infectivity.

In addition to direct roles in RNA replication and virus production, the conservation of observed RNA structures indicated that they contribute to overall viral fitness. A high degree of selective advantage must be present to maintain the structures through all HCV strains; however, not all facets of viral fitness can be readily measured. RNA structures may be involved in evading the immune system, determining tissue tropism, or controlling genetic drift. Incorporation of RNA structures into protein-coding sequence could serve to streamline genome size, camouflage immunogenic RNA signals, control translation of the large ORF, or assist folding of the genomic RNA. Intriguingly, work by Simmonds and others has described an association between viruses with a high degree of genome-scale ordered RNA structure (GORS), such as HCV, and viral persistence (Simmonds et al., 2004). SL783A could play a role in any one of these processes or just as likely work in concert with neighboring structures to help establish persistent infection. These could also be important roles for the structures (SL669, SL761, SL833, SL2531, SL2549, and SL8670) that were not characterized in our functional assay (Figure S2).

The location of RNA elements in the genome can offer important clues to their function. For example, many RNA structures within the core-encoding region perform functions that are only necessary when the core protein gene is present. SL588 is only critical in the context of an intact viral packaging pathway. It is known that the efficiency of replication and of virus production are often at odds, and it is thought that replication-enhancing mutations retain viral RNA in the replication complex at the expense of transferring the RNA to be packaged (Pietschmann et al., 2009). SL588 may therefore represent an important component for controlling this transfer process. By interacting with SL427, SL588 may shunt the genome away from the replication machinery and allow the viral RNA to interact with core and other factors involved in the assembly process. When the tertiary interaction is absent, SL588 could maintain replication in the presence of competing packaging machinery. SL588 is contained entirely within the core-encoding region, and it is notable that deletion of proteins necessary for viral packaging (core-NS2) therefore simultaneously removes the key RNA regulatory elements that are needed to control this pathway. In this way, the genes of HCV can be seen as modular units that function in coordinated fashion at both RNA and protein levels. This elegant organizational strategy offers multiple evolutionary advantages, minimizing genome size and increasing the likelihood that, in the rare event that HCV strains recombine, they form functional chimeras.

The organizational scheme seen in the core-encoding region also appears to be operative in other sections of the genome. In the NS5B-encoding region, which expresses the viral polymerase, numerous structures have been demonstrated to be involved in RNA replication (Cheng et al., 1999; Diviney et al., 2008; Tuplin et al., 2004). In this study, we demonstrate that SL8001 formation negatively regulates the initial replication rate. This could represent a feedback mechanism to limit NS5B expression or to limit the immunogenic signals that result from active expression of NS5B (Yu et al., 2012).
Conservation of these structures and their distribution throughout the viral genome demonstrates that nature uses redundancy in the genetic code to form RNA structures that control biological processes. In this study we have taken advantage of this redundancy to genetically sample different RNA conformations and observe their unique contributions to viral function. HCV may adopt a similar strategy in natural infections by using a high mutation rate to generate quasispecies with distinct RNA conformations that fill different regions of the fitness landscape. Our findings introduce a new layer of complexity in understanding how the genetic code is optimized (Shen et al., 2015), and they suggest a diversity of ways that gene expression can be regulated. The search strategy outlined in this study can be adopted for identifying RNA structures occurring in any protein-coding RNA. In future studies, it will be interesting to see how other biological systems harness the capacity to form RNA structures in protein-coding sequence as a means to control metabolic processes.

**EXPERIMENTAL PROCEDURES**

**RNA Purification and Folding**

Plasmids were linearized with XbaI and treated with mung bean nuclease to form blunt ends. Jc1 RNA (Pietschmann et al., 2006) was made by runoff transcription with T7 RNA polymerase, then treated with DNase (20 U/ml, Ambion). Transcription products were purified by using quasidenuclearizing renaturation methods (RNasey) and native, non-renaturation methods as follows: for the quasidenuclearizing renaturation method, DNase-treated RNA transcripts were loaded onto RNasey columns (QIAGEN), washed according to manufacturer’s protocol, and eluted in RNase-free water. For folding, RNA was buffered in 50 mM K-HEPES (pH 7.4), 0.1 mM EDTA, 150 mM KCl, and 5 mM MgCl₂. Samples were heated to 65°C and cooled to 37°C over 45 min on a PCR block. For the native, non-renaturation method, following DNase treatment, the reaction mixture was treated with proteinase K (0.3 mg/ml, Ambion). Digested products were diluted 15X in 100 mM KCl, 8 mM K-MOPS (pH 6.5), and 0.1 mM NaEDTA and loaded onto an Amicon Ultra-500 (Millipore) centrifugal filter (100 kDa molecular weight cutoff). Three consecutive filtrations were performed at 4,000 g and cooled to 37°C over 45 min on a PCR block. For the native, non-renaturation method, following DNase treatment, the reaction mixture was treated with proteinase K (0.3 mg/ml, Ambion). Digested products were diluted 15X in 100 mM KCl, 8 mM K-MOPS (pH 6.5), and 0.1 mM NaEDTA and loaded onto an Amicon Ultra-500 (Millipore) centrifugal filter (100 kDa molecular weight cutoff). Three consecutive filtrations were performed at 4,000 g and cooled to 37°C over 45 min on a PCR block.

**SHAPE**

NAI (2-methylisonic acid imidazolide) (Spitale et al., 2013) or an equivalent volume of DMSO was added to samples and incubated for 10 min at 37°C. Modified RNA was promptly extracted with phenol:chloroform, LCl precipitated, and reverse transcribed (RT) by using Superscript III (Invitrogen) and 6-JOE-labeled primers spaced 200–300 nt apart (Table S3A). Formamide, betaine, and trehalose to final concentrations of 5% (v/v), 1 M, and 600 mM, respectively, were added to the reaction to improve yield. cDNA products were ethanol precipitated, resuspended in formamide, coloaded with sequencing ladders, and resolved by capillary electrophoresis. Sequencing ladders were prepared by using 5-FAM-labeled primers, dyeoexy-NTPs (Roche), and cycle sequenase (Affymetrix) on template DNA.

**SHAPE Analysis**

Traces were processed in QuSHAPE, and signals were normalized by primer read as previously described (Karabiber et al., 2013). DMSO control peaks were subtracted from reagent peaks, and resulting normalizations were normalized such that the average intensity of the top 10% most reactive peaks was 1.0. We performed two or more replicates per primer from separately prepared and folded transcripts. Reads were assembled according to genome position, and data from replicate experiments were then averaged in Excel. Final figures were prepared in Graphpad Prism.

**Structure Modeling and Phylogenetic Analysis**

SHAPE reactivity was converted into pseudoenergy constraints in RNAstructure (Deigan et al., 2009) and used to fold the Jc1 RNA sequence in overlapping 1 kb segments spanning the genome. Calculations were performed at 37°C with parameters m = 1.8 kcal/mol and b = −0.6 kcal/mol. Candidate structures were appended to a Stockholm alignment of genotype 2 sequences. We used Infernal 1.1 (Navrotsky and Eddy, 2013) to build a covariance model from this initial alignment and incorporated more divergent sequences based on structural similarity to the model. To match distantly related strains (e.g., genotype 6 strains) to the conserved RNA structure, the program permitted mismatches and gaps to appear in the primary sequence alignment. With the positions of paired nucleotides properly matched, we used R29 (Weinberg and Breaker, 2011) to score and visualize covarying substructures. All HCV sequences were obtained from the NCBI taxonomy browser.

**Construct Design and Preparation**

Secondary structures and free energies of mutated motifs were calculated by using mfold (Zuker, 2003). Codon frequencies were calculated by using the Sequence Manipulation Suite (Stothard, 2000). Mutants were prepared by site-directed mutagenesis or were synthesized en bloc (Invitrogen), then subcloned back into the parental plasmid using available restriction sites (Table S3B). The pJc1/GLuc plasmid was previously described (Phan et al., 2009). All sequences were verified at the Yale Keck facility.

**Cell-Culture Replication and Infectivity**

Plasmids were linearized as above, and RNA was prepared by runoff transcription with T7 RNA polymerase, purified by QIAGEN RNasey columns, then stored at −20°C in ME buffer (10 mM MOPS (pH 6.5), 1 mM EDTA). Huh-7.5 cells were grown in DMEM (Invitrogen) supplemented with 10% fetal calf serum (Omega Scientific) and 1 mM nonessential amino acids (Invitrogen). Replication assays were performed in 24-well plate format. HCV RNA was transfected with MIR-2250 mRNA lipid transfection reagent (Mirus) according to the manufacturer’s protocol. Media were collected at the noted time points (12, 24, 48, 72, and 120 hr for full replication time course; every 8 hr for 48 hr for early replication time course), clarified in a microcentrifuge at 16,000 × g, and mixed with 5X luciferase lysis buffer (NEB). Luminescence was measured on either a Berthold Centro LB 960 luminometer or a Biotek Synergy H1 plate reader by using 10 μl of media, 50 μl of Gaussia luciferase reagent (NEB), and a 2–10 s integration time. Infectivity assays were performed in 96-well format. Naïve cells were infected with 100 μl of media collected from the replication assay. After 16 hr of infection, cells were washed with PBS, and fresh DMEM was added to each well. Media were collected 72 hr after washing, and luminescence was measured as described above. Data were normalized relative to the lowest value of “mock.” Statistical significance was calculated in Graphpad Prism by performing an analysis of covariance (ANCOVA) on the logarithmic transformation of normalized luciferase data measured at each time point beginning at hour 24.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.01.024.

**AUTHOR CONTRIBUTIONS**


**ACKNOWLEDGMENTS**

We thank S. Somarowthu and Z. Weinberg for guidance in bioinformatics analysis. We thank O. Fedorova for the synthesis of NAI. We thank K. Sanbonmatsu, R. Breaker, A. Iwasaki, D. Rawling, and all members of the A.M.P. lab for discussions. This work was supported by the NIH RO1 AI089826 (to A.M.P. and B.D.L.) and T32 GM07205 (to N.P.) A.M.P. is an investigator.
of the Howard Hughes Medical Institute. Funding for open access charge is from NIH.

Received: July 8, 2015
Revised: December 22, 2015
Accepted: January 21, 2016
Published: February 25, 2016

REFERENCES


