Biology Faculty Publications   Department of Biology

11-13-2019

Structure of the 5' Untranslated Region of Enteroviral Genomic RNA

Bejan Mahmud
University of Nebraska Omaha, bmahmud@unomaha.edu

Christopher M. Horn
University of Nebraska at Omaha

William E. Tapprich
University of Nebraska at Omaha, wtapprich@unomaha.edu

UNO LIBRARIES
LIBRARY.UNOMAHA.EDU

Structure of the 5′ Untranslated Region of Enteroviral Genomic RNA

Bejan Mahmud,[a] Christopher M. Horn,[a] William E. Tapprich[a]

[a]Biology Department, University of Nebraska at Omaha, Omaha, Nebraska, USA

ABSTRACT Enteroviral RNA genomes share a long, highly structured 5′ untranslated region (5′ UTR) containing a type I internal ribosome entry site (IRES). The 5′ UTR is composed of stably folded RNA domains connected by unstructured RNA regions. Proper folding and functioning of the 5′ UTR underlies the efficiency of viral replication and also determines viral virulence. We have characterized the structure of 5′ UTR genomic RNA from coxsackievirus B3 using selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) and base-specific chemical probes in solution. Our results revealed novel structural features, including realignment of major domains, newly identified long-range interactions, and an intrinsically disordered connecting region. Together, these newly identified features contribute to a model for enteroviral 5′ UTRs with type I IRES elements that links structure to function during the hierarchical processes directed by genomic RNA during viral infection.

IMPORTANCE Enterovirus infections are responsible for human diseases, including myocarditis, pancreatitis, acute flaccid paralysis, and poliomyelitis. The virulence of these viruses depends on efficient recognition of the RNA genome by a large family of host proteins and protein synthesis factors, which in turn relies on the three-dimensional folding of the first 750 nucleotides of the molecule. Structural information about this region of the genome, called the 5′ untranslated region (5′ UTR), is needed to assist in the process of vaccine and antiviral development. This work presents a model for the structure of the enteroviral 5′ UTR. The model includes an RNA element called an intrinsically disordered RNA region (IDRR). Intrinsically disordered proteins (IDPs) are well known, but correlates in RNA have not been proposed. The proposed IDRR is a 20-nucleotide region, long known for its functional importance, where structural flexibility helps explain recognition by factors controlling multiple functional states.

KEYWORDS 5′ untranslated region, coxsackievirus B3, RNA structure, internal ribosome entry site

The genus *Enterovirus* includes 15 species of animal viruses that belong to the family *Picornaviridae* (1). Each species contains many subtypes and variants, with over 80,000 enteroviral genome sequences available in the nucleotide database. The diversity of the viruses in the genus is reflected by the diversity of disease processes that result from infection. In humans, seven species of enteroviruses are responsible for disease states that range from mild conditions like the common cold to very severe diseases, such as poliomyelitis, myocarditis, and acute flaccid paralysis (2). Despite the broad diversity of virus types and diseases, enteroviruses are remarkably similar in structural organization and infection mechanisms. They are small, nonenveloped, icosahedral viruses with a positive (+)-polarity single-stranded RNA genome typical of the *Picornaviridae* (3, 4). Enteroviruses also share a highly structured 5′ untranslated region (5′ UTR) of approximately 750 nucleotides with a conserved domain organization and a type I internal ribosome entry site (IRES) (5–7). The enteroviral 5′ UTR is

responsible for recruiting ribosomes, canonical initiation factors, and noncanonical host factors, collectively known as IRES *trans*-acting factors (ITAFs), to initiate translation of the viral polyprotein via the IRES element (8) and also provides an RNA domain called the cloverleaf on the 5′ end that is recognized by ITAFs and the viral polymerase for genome replication (9). We have characterized the RNA structure in the 5′ UTR of coxsackievirus B3 (CV-B3) as a model of enteroviral genomic RNA.

CV-B3 is an important human pathogen. While most infections are inapparent, they can lead to serious and even fatal diseases, such as pancreatitis, myocarditis, and dilated cardiomyopathy (10–15). The 5′ UTR of CV-B3 and the closely related poliovirus (PV) has been studied in detail because it participates in all parts of the infection cycle and contains critical virulence determinants (16–20). During infection, the 5′ UTR orchestrates a hierarchical series of activities that ensure efficient gene expression and replication. In general, translation and replication are controlled by different 5′ UTR RNA domains. The highly conserved cloverleaf structure (domain I) directs replication of both positive and negative strands (9, 21–25). Domains II to V work together to form the viral IRES that controls translation. In enteroviruses, the IRES is classified as type I and is distinct from the type II and III, HCV-like, and IGR IRES elements present in other picornavirus genera (26).

As genome sequences became available and solution techniques for determining RNA structure were developed, models of the enteroviral 5′ UTR were proposed (27–29). These models have been refined as experimental systems have advanced and as more sequences have become available for detailed comparative analyses (6, 20). Significant advances in methods that interrogate RNA structure in solution, most notably selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) (30, 31), together with highly advanced RNA-folding algorithms that incorporate SHAPE constraints into thermodynamic and comparative structure predictions (32), make it possible to propose significant changes to the existing models. We have applied these new approaches and present a model for the 5′ UTR of CV-B3 that adds structural information in three major ways. First, although the domain structure of previous models is recognizable and domains I and IV have only minor adjustments, there is significant reorganization of domains II, III, V, and VI. Second, our analysis has identified two new long-range interactions, one that has base pairing that is conserved in enteroviral 5′ UTRs and another that has base pairs that are CV-B3 specific. Third, we propose that the conserved pyrimidine-rich region downstream of domain V with the motif Y_n-X_m-AUG (33, 34), where Y_n is 8 to 10 pyrimidines, X_m is 18 to 20 nucleotides, and AUG is a cryptic start codon, is an intrinsically disordered RNA region (IDRR). The Y_n-X_m-AUG region is known to be important for binding multiple factors, including the ribosome (35). Similar to the role of intrinsically disordered regions in proteins (36), this disordered RNA region may facilitate multiple hierarchical interactions.

RESULTS

5′ UTR secondary-structure model. We combined data from three complementary approaches to propose a new model for the secondary structure of the 5′ UTR of enteroviral genomic RNA: chemical modification in solution, multiple-sequence analysis, and free-energy analysis using maximum expected accuracy (MEA) calculations. These three approaches were applied to the 5′ UTR of CV-B3 as a model for enteroviruses. The results from the three types of analysis provided inputs for constraining base-pairing interactions in the TurboFold II folding algorithm (32).

Data from chemical modification in solution was generated by SHAPE and was supplemented by the base-specific chemical probes dimethyl sulfate (DMS) and 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT). SHAPE chemicals, such as *N*-methylisatoic anhydride (NMIA), target and react with the 2′-OH group of the RNA sugar backbone. The reactivity of an individual base depends on the flexibility of its ribose backbone, which is determined by base pairing and other conformational constraints. More flexible positions are modified to a greater extent. For further structural analysis, we utilized two base-specific chemical probes that helped to

distinguish between positions involved in Watson-Crick pairs and positions that are unpaired. For both SHAPE and base-specific probing, modified sites were identified by reverse-transcriptase (RT) extension of fluorescently labeled cDNA primers. The cDNA fragments were separated using capillary electrophoresis, and the resulting fluorescence intensity files were analyzed using ShapeFinder (37). Modification intensities from ShapeFinder were inputs for TurboFold II, a structural prediction algorithm that relies on experimentally generated probing data and sequence conservation among functionally homologous sequences. For the sequence data set, we selected all of the 5′ UTR sequences available in the NCBI database where published characterization indicated that the genomes were derived from virulent, wild-type strains of CV-B3 (see Table S1 in the supplemental material).

SHAPE data (Fig. 1; see Table S2 in the supplemental material) were used to generate a secondary structure of the CV-B3 5′ UTR (Fig. 2). Base-specific probes, DMS and CMCT, were used in parallel to confirm the findings of SHAPE probing. When the base-specific data set was used independently in the TurboFold II analysis, the overall resultant structure prediction was over 92% similar to the one generated with SHAPE alone (Fig. 3). Comparison of the two predictions showed that the overall structure of each domain was preserved, including the newly proposed long-range interactions. The differences were primarily small extensions or reductions of base-paired stem regions. These differences included small changes in domains I, III, and IV. In domain I, the base-specific data set removed 2 base pairs that increase the size of a loop region. In domain III, the basal stem was extended, which shortened the basal stem of domain IV. The interconnecting region between domains IV and V was also shortened.

Two models of the CV-B3 5′ UTR structure proposed previously (6, 38) identified 11 (A to K) and 7 (I to VII) domains defined by base-paired basal stems. Our new model maintains some domains from both models but makes substantial changes to the structural organization of the majority of the molecule. Our model confirms the structure of domain I, also called the cloverleaf. Extensive characterization of the cloverleaf showed that the domain plays a critical role in initiation of replication by forming ribonucleoprotein (RNP) complexes with both viral and host factors (24, 25, 39). A single G-C pair (positions 20 and 24) is proposed to close the apical part of stem-loop b in domain I. The high SHAPE reactivity of 20G indicates that this pair may not be stably formed, in which case the apical part of stem-loop b would consist of a single larger loop that includes the nucleotides within the base pair and the bulge (19U, 20G, and 24C). This matches the proposed model of stem-loop b of the human rhinovirus 14 cloverleaf as determined by nuclear magnetic resonance (NMR) (40).

The interconnecting region between domains I and II has been shown to interact with poly(rC) binding protein 2 (PCBP2) with high affinity (41). Position 91C of the C-rich region near the 3′ end of the cloverleaf (positions 90 to 94) is moderately reactive; however, only 92C and 93C are highly conserved (97%). Overall, the pyrimidine-rich region between domains I and II stretches between positions 88 and 104 and is mostly protected from modification.

Domain II has been suggested to be crucial for viral virulence (17, 18, 20, 42, 43). In our model, domain II occupies the region between positions 105 and 181, which is larger than the model previously reported by Bailey and Tapprich for CV-B3 (6), as well as a recently proposed model for poliovirus (44), but identical to the model proposed by Liu et al. (38) and the model proposed by Zell and Stelzner for bovine enterovirus (28).

The predicted secondary structure of domain III consists of a single stem-loop occupying the region between bases 189 and 209, much shorter than the domain previously reported by Bailey and Tapprich, which included multiple stems and internal loops (6), but consistent with the small domain proposed by Liu et al. (38). Interestingly, deletion of the nucleotides included in the shorter domain does not inhibit poliovirus replication and translation (45–47).

As consistently documented in previous models, the >200-nucleotide domain IV assumes a very complex structure. The majority of our proposed model for domain IV
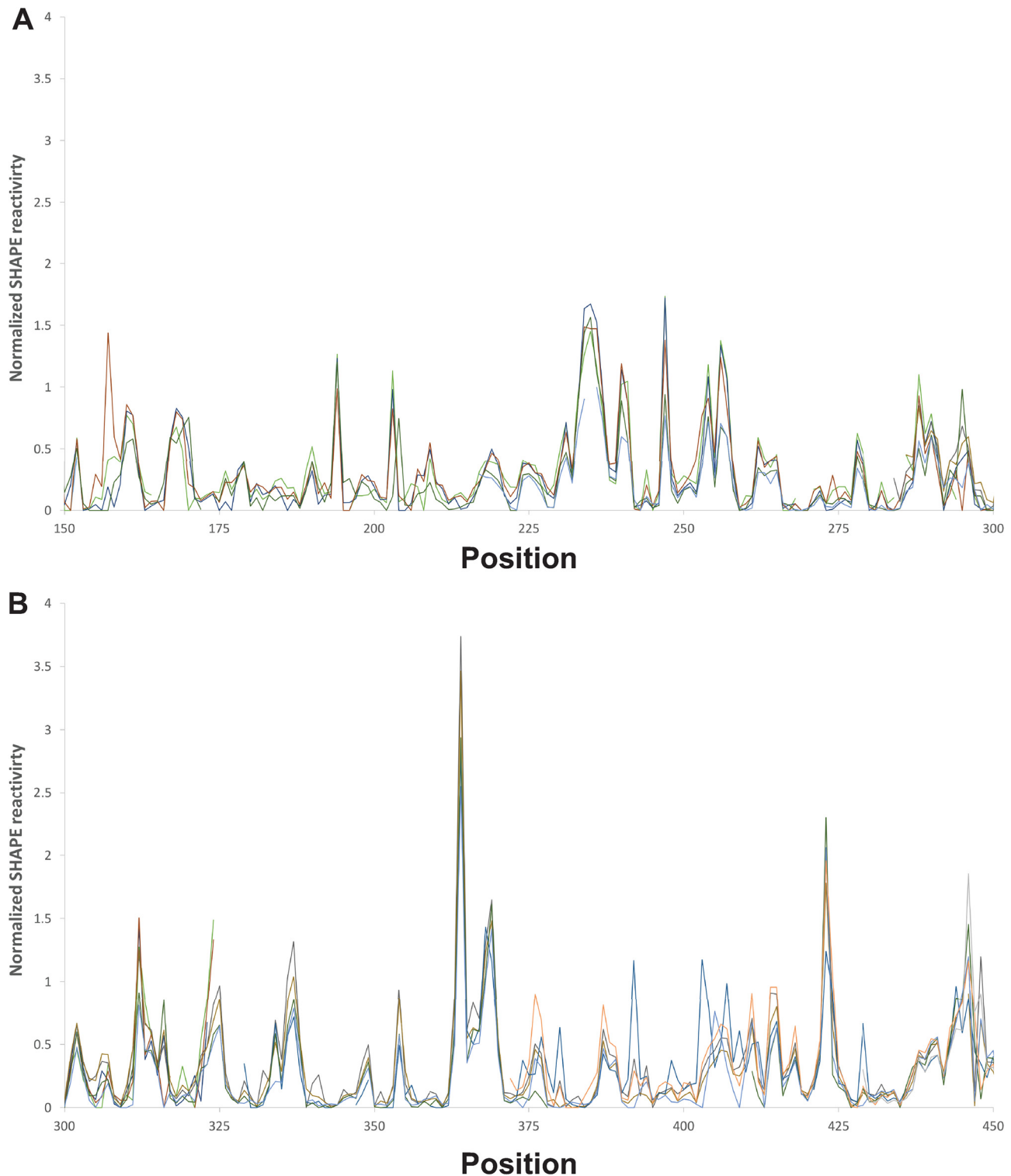
FIG 1 SHAPE reactivity traces by position. Reactivity traces for independent SHAPE experiments are overlaid on the graphs. Raw data for all SHAPE reactivity values for all experiments and all positions, together with means and standard deviations for each position, are listed in Table S3. (A) Reactivity traces for positions 150 to 300. (B) Reactivity traces for positions 301 to 450.

is consistent with prior models, particularly near the apex of the complex stem-loop domain. The junction loop atop domain IV branches out into a cruciform structure consisting of stem loops 4a, 4b, and 4c. Both stem-loops 4a and 4b have single-stranded poly(rC) regions, which may be involved in binding to poly(rC) binding

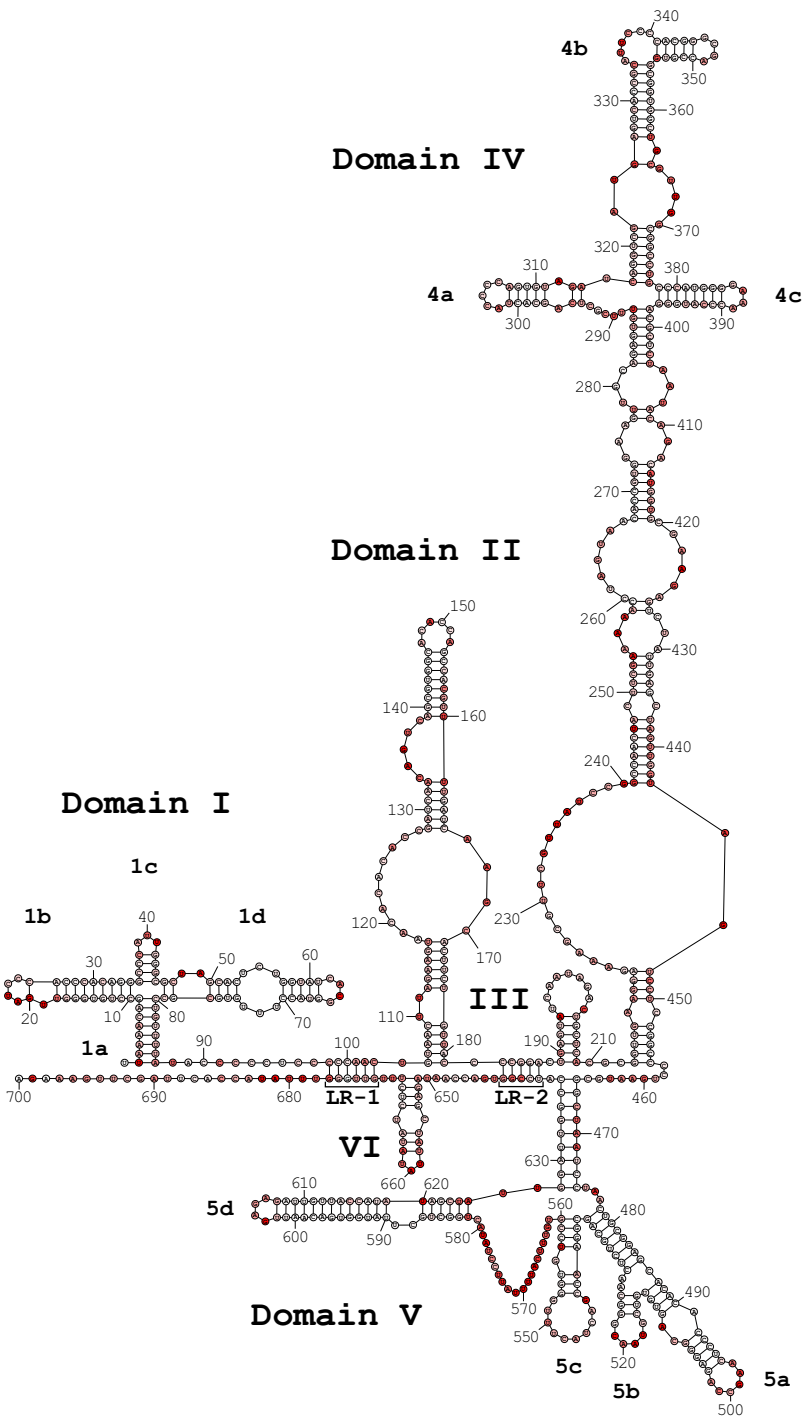**FIG 2** Secondary-structure model for the CV-B3 5′ UTR. SHAPE reactivity at each position is indicated by red shading. No shading shows no/low reactivity, with a normalized reactivity value below 0.3. Light-red shading shows moderate reactivity, with a normalized reactivity value between 0.3 and 0.7. Dark-red shading shows high reactivity, with a normalized reactivity value above 0.7. Domains are labeled I to VI. Stem-loops in domains with junction loops are labeled with letters. Two long-range interactions are labeled as LR-1 and LR-2.

proteins. The poly(rC) region within the apical loop of stem-loop 4a (positions 303 to 306) is highly conserved among enteroviruses but was not reactive when tested with SHAPE; it did, however, exhibit moderate reactivity when probed with base-specific agents. The poly(rC) region within stem-loop 4b (positions 338 to 341) is also highly
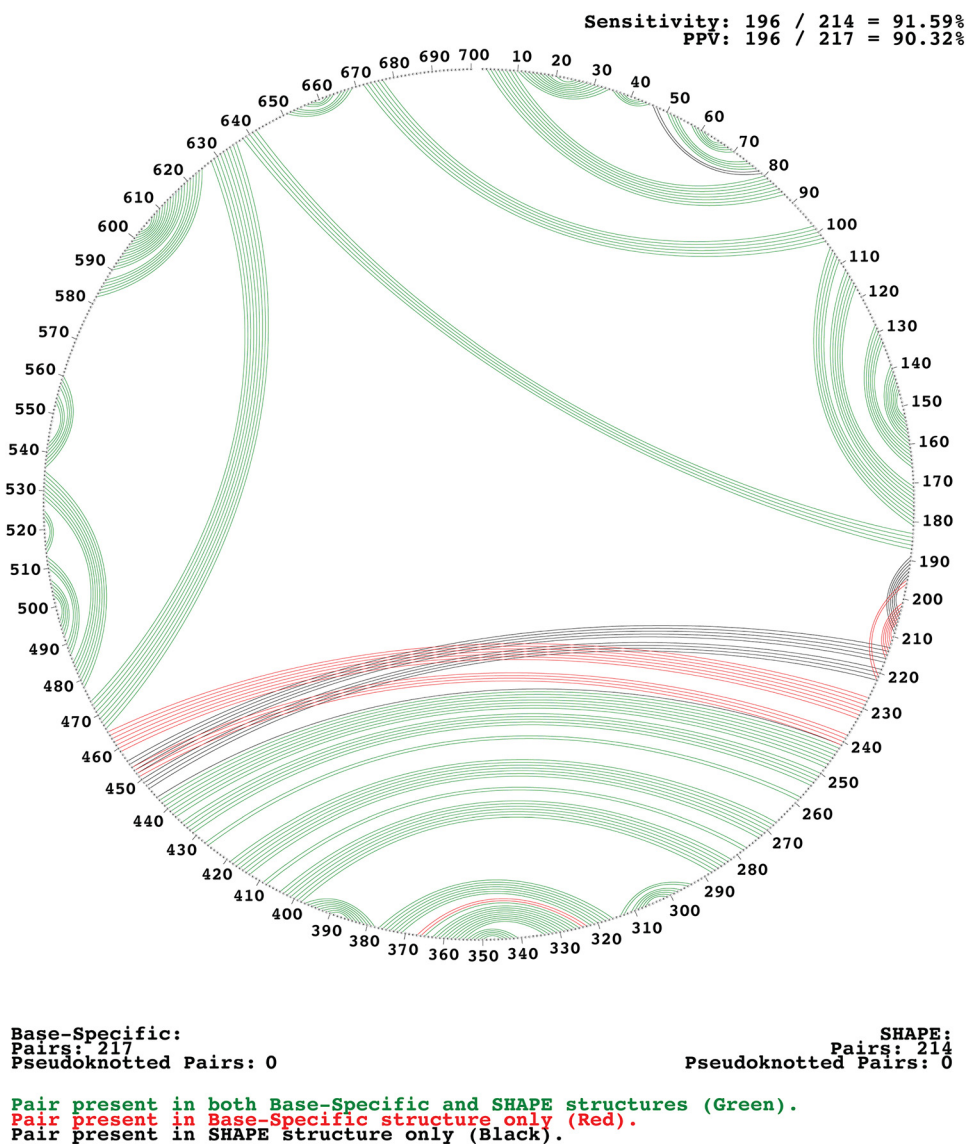
```
                                    Sensitivity: 196 / 214 = 91.59%
                                            PPV: 196 / 217 = 90.32%
```

Base-Specific:
Pairs: 217
Pseudoknotted Pairs: 0

SHAPE:
Pairs: 214
Pseudoknotted Pairs: 0

Pair present in both Base-Specific and SHAPE structures (Green).
Pair present in Base-Specific structure only (Red).
Pair present in SHAPE structure only (Black).

FIG 3 Circle diagram comparing secondary-structure models generated by SHAPE reactivity and base-specific reactivity. Nucleotide positions are displayed outside the circle. Base pairs are connected by lines inside the circle. Differences in the models generated are shown in red and black. PPV, positive predictive value.

conserved across the enterovirus A, B, and C sequences and is largely inaccessible to SHAPE modification. While position 338C is moderately accessible, the other cytidine bases are not accessible.

Our new model adds roughly 60 residues to the length and substantially rearranges domain V into a complex junction loop. This rearrangement brings a number of functionally important elements into close structural proximity. For example, the pyrimidine-rich region from position 561 to position 581, which plays a role in ribosome recognition (35, 48), was previously thought to bridge domains V and VI but is now fully encompassed within the newly updated domain V. The predicted structure of domain V is highly complex and can be broken into four stem-loop regions that are connected by two multiloops. Sabin vaccine strains of poliovirus have functionally deleterious mutations within this domain. Sabin-like mutations of the CV-B3 5′ UTR have debilitating effects on viral replication and translational efficiency (49). Positions 484A, which is mutated in Sabin 1-like strains, and 485G, which is mutated in Sabin 2-like strains, are both protected from any sort of modification and are involved in Watson-Crick base

pairing within a stem near a three-way junction that also branches out into stem-loops 5a and 5b. These two positions exhibit high levels of conservation across our data set of enterovirus A, B, and C species. 484A is conserved in 99% of the sequences, while 485G is present in 97%. The most deleterious is the Sabin 3-like mutation of 473C. In our model, 473C is protected from modification by both SHAPE and base-specific methods and is located near the junction within the basal stem of domain V. Position 473C is conserved across 99% of the enteroviral A, B, and C sequences in our data set. Stem-loop 5d was previously thought to be a discrete domain (domain VI) outside domain V (6). The structure of stem-loop 5d is identical to the domain VI structure of previous models. In our model, stem-loop 5d and the upstream stem-loops share a basal stem (nucleotides 466 to 474 and 628 to 636). This basal stem, which is a newly proposed structural feature of our model, has support from SHAPE analysis and is also supported by base-specific probing. TurboFold II generates this basal stem when constrained by either data set. The sequence of the basal stem is extremely conserved. Of the 18 nucleotides that constitute the basal stem, 14 are protected from SHAPE modification and the other 4 (468C, 469U, 470A, and 471A) are moderately reactive. The 5′ side of the basal stem is highly conserved (Fig. 4) (Stockholm Alignment File [https://unomaha.box.com/s/2osxk3l8xizj3x6usbqd6sqym9qdt8vi]). Positions 466G, 467G, and 472U are perfectly conserved across all of the 1,653 enterovirus A, B, and C sequences analyzed, while 470A and 471A are changed in only one sequence. 469U and 474C are present in 99% of the sequences, and 98% conservation was observed for 468C. Similarly, the 3′ side of the basal stem is highly conserved, with every nucleotide within the region (positions 628 to 636) present in at least 97% of the analyzed sequences. The 469-633 pair has 27 instances of compensatory changes among the 1,653 enterovirus A, B, and C sequences, but there are also 15 examples where a mutation in position 633 does not result in a Watson-Crick pairing. For the other pairs in the basal stem, an average of 0.9% of aligned sequences have mutations that create non-Watson Crick pairs (Stockholm Alignment File [https://unomaha.box.com/s/2osxk3l8xizj3x6usbqd6sqym9qdt8vi]).

The cryptic AUG (positions 591 to 593) within stem-loop 5d is essential for translation, and mutations of this region have been reported to inhibit viral translation (50). Our analysis indicates that this AUG is highly conserved (99%) across enteroviruses A, B, and C species and is mostly inaccessible, except for 591A, which showed moderate reactivity to base-specific probes.

The beginning of domain VI in our proposed model is shifted by about 65 residues from the previously proposed structure due to the increase in the size of domain V in the new model. Our data indicate that domain VI forms a single stem-loop with a single internal loop.

Long-range interactions. The secondary-structure model for CV-B3 has two newly identified long-range interactions (Fig. 2, LR-1 and LR-2). LR-1 involves six Watson-Crick base pairs spanning positions 98 to 103 and 672 to 677. Bases 99C, 100C, 673U, 675G, and 676G are not reactive, while other nucleotides constituting the long-range interaction displayed moderate reactivity values. Sequence comparison does not support generalization of this long-range interaction when analyzing enterovirus A, B, and C species (Fig. 4). The 98-to-100 C-rich region is conserved across the aligned enterovirus strains; however, the region also has a large incidence of insertions and deletions. The 672-to-677 region is more conserved. Bases 675G and 676G are present in over 95% of the enterovirus A, B, and C sequences, and over 90% conservation was observed for 673U. 672G and 674U are moderately conserved at 66% and 84%, respectively. Only 7% of the sequences have a guanine at the position corresponding to 677G of the CV-B3 strain 28 5′ UTR. Moreover, among the set of 28 virulent CV-B3 sequences used in TurboFold II to generate the model, the 98-677 pair was conserved in only in 5 of the sequences. Taken together, the data support LR-1 for CV-B3 strain 28, used in our analysis, but the interaction cannot be generalized. We found that LR-1 had sufficient support to be included in our model of CV-B3 because SHAPE probing constraints and
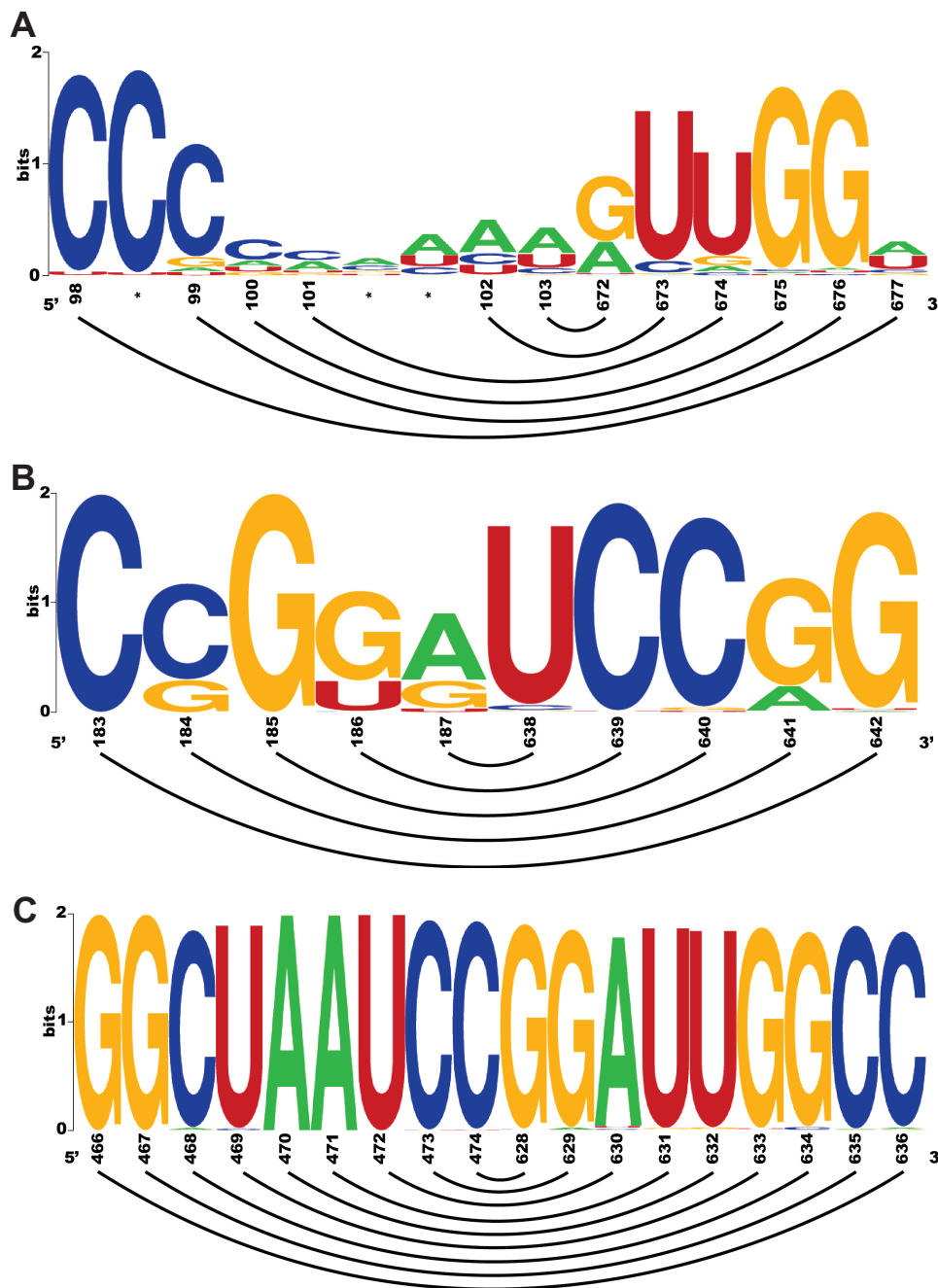
FIG 4 WebLogos showing sequence conservation of positions in proposed long-range interactions LR-1 (A) and LR-2 (B) and in the domain V stem (C). Proposed base-pairing interactions are shown by connecting lines.

base-specific probing constraints independently generated the proposed interaction when analyzed by TurboFold II. However, we cannot say that LR-1 will be a common feature when the structures of other enteroviral 5′ UTRs are determined.

Five Watson-Crick pairs are involved in the long-range binding indicated in box LR-2 (Fig. 2). This interaction is well supported by SHAPE reactivity, and sequence comparison showed a high degree of conservation, together with an example of a pair with compensatory changes (Fig. 4). Nucleotides 640C, 641G, and 642G are moderately reactive to NMIA, while other bases constituting this interaction are not reactive. Bases 183C and 185G were present in over 99% of the enterovirus A, B, and C sequences. 184C and 186G were each found in 74% of the sequences. 187A was 68% conserved
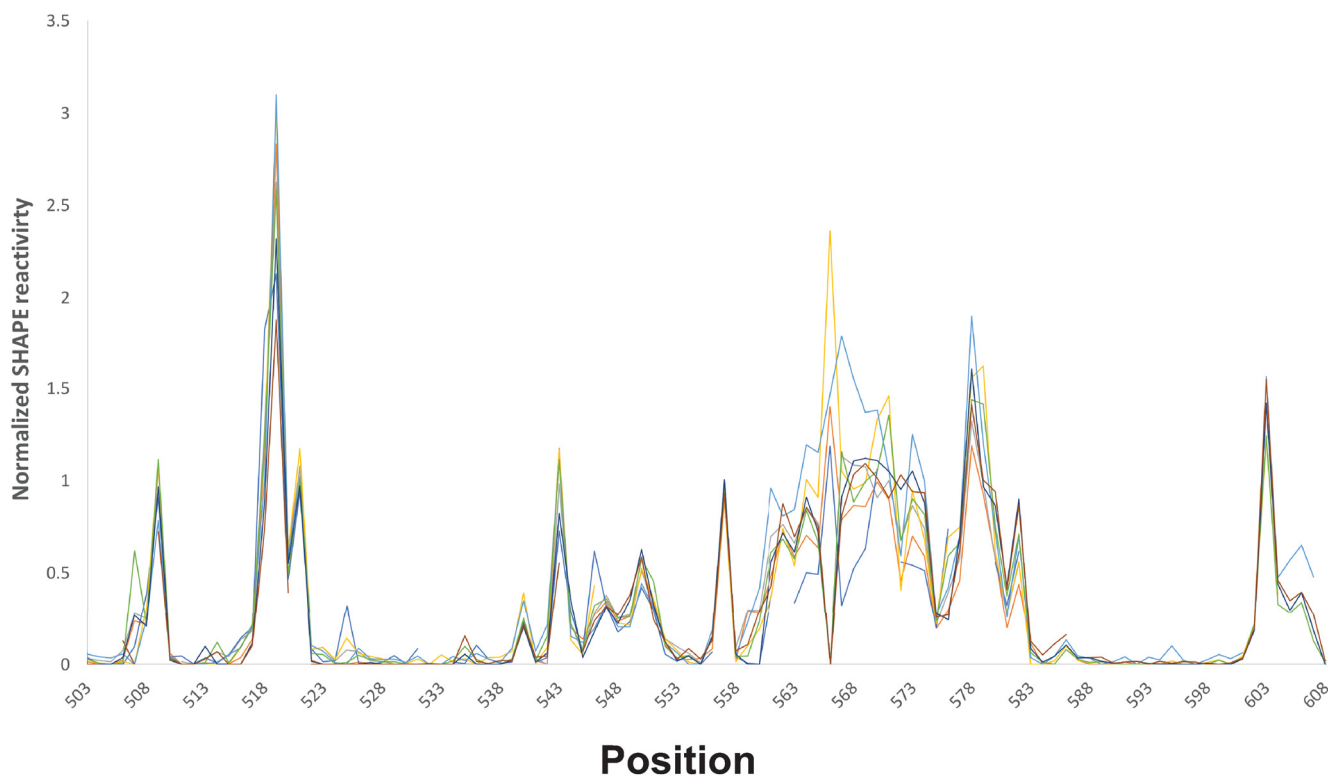
FIG 5 SHAPE reactivity traces by position for positions 503 to 608. Reactivity traces for independent SHAPE experiments are overlaid on the graph. The traces show consistent reactivity for regions upstream and downstream of the proposed intrinsically disordered RNA region (positions 561 to 581).

and was switched to a guanine in 28% of the analyzed sequences, which maintained the pairing interaction. This was the only pair in either proposed long-range interaction that showed a clear compensatory change. The 638-to-642 region is highly conserved: 638U, 639C, 640C, and 642G were found in over 95% of the sequences. 641G was present in 80% of analyzed enteroviruses. The 184C-641G pair was found in 26 of the 28 virulent CV-B3 sequences, while the other four base pairs were preserved in all of the utilized sequences. Among the enterovirus A, B, and C sequences, 183C-642G and 186G-639C pairs are conserved in over 99% of the strains, while the 185G-640C pair was found in 98% of the sequences.

An intrinsically disordered RNA region. The pyrimidine-rich single-stranded region leading into stem-loop 5d has been reported to be the ribosomal landing site for the IRES and thus plays an essential role in translation of the viral polyprotein (51). The nucleotides in this single-stranded region (561 to 581) are either highly reactive or moderately reactive, with the exception of 575C, which was not chemically accessible when probed with SHAPE but was moderately accessible to base-specific probes. This pyrimidine-rich region displayed extremely high variability among individual runs when tested with both SHAPE (Fig. 5) and base-specific probes. The modification pattern was highly inconsistent, with bases displaying variable reactivity. As also shown in Fig. 5, the variability in reactivity of the 561-to-581 region was not observed in the regions immediately at the 5′ or 3′ side of this region. Indeed, no other region in the molecule displayed this variability. Positions 561 to 581 represent the longest single-stranded stretch within the CV-B3 5′ UTR. A similar-size single-stranded stretch between positions 223 and 240 was found in an asymmetric internal loop within domain IV. The modification pattern within this single-stranded region was consistent across distinct runs ($n = 5$), suggesting that variability in reactivity to the chemical is not inherent to long single-stranded regions of RNA. Position 566C showed the most variable reactivity in the 5′ UTR but, interestingly, is almost perfectly conserved across enterovirus A, B, and C sequences. Overall, the pyrimidine-rich nature of the entire single-stranded region is highly

conserved. However, high rates of insertions and deletions were observed within the 573-to-577 window, and lower conservation levels were observed for positions 570 through 581. The unusual chemical modification behavior of the 561-to-581 region suggests that the positions do not adopt a single, or even small, set of structural states. We propose that this region is intrinsically disordered and represents a feature that functions similarly to intrinsically disordered regions in proteins (36).

DISCUSSION

The CV-B3 RNA genome is organized into a 742-nucleotide 5′ UTR, a 6,555-nucleotide coding region that produces a 2,185-amino-acid polyprotein, a 98-nucleotide 3′ UTR, and a poly(A) tail (52). This general genome organization is common among enteroviruses and also among all picornaviruses. For all picornaviruses, the highly structured 5′ UTR plays a critical role in gene expression and genome replication. As modeled by the 5′ UTR of CV-B3, replication of enteroviral RNA by the viral RNA-dependent RNA polymerase to produce both negative and positive strands requires a cloverleaf structure on the 5′ end of the genome (9). Beyond the cloverleaf, a large portion of the 5′ UTR functions as an IRES element that is responsible for recruiting ribosomes for translation (26). In enteroviruses, this IRES element is classified as type I. Given the critical functional activities of the 5′ UTR, it is not surprising that many virulence determinants have been localized there (17, 18). Mutagenic, chimeric, and deletion studies have confirmed the close structure-function relationship in the 5′ UTR (17, 49, 53, 54). The prevalence and health consequences of enterovirus infection make improved understanding of the 5′ UTR structure fundamentally important. Recent advances in methods for determining the structure of RNA molecules enable just such improved understanding.

Modern RNA structural analysis represents an amalgamation of biochemistry and computational biology. This is particularly true for chemical probing approaches, where recently introduced chemical agents have joined powerful computational analyses to generate reliable secondary-structure models for large RNAs. High-throughput data-gathering methods, such as capillary electrophoresis and next-generation sequencing, are coupled with custom computational software for quantitative analysis of chemical modification results (55, 56). Folding algorithms have also seen a staggering amount of improvement in both how they predict structures and how they interpret the data from such experiments (57). Minimum-free-energy (MFE) analysis is quite accurate for short RNA regions, but for large RNAs, prediction is much more difficult, even with chemical probing constraints. While MFE analysis remains the most intuitive and popular method of prediction, a new paradigm for structure prediction has been developed. Using base pair probabilities derived from a partition function calculation, a structure with the MEA can be predicted. While MEA structures may not necessarily have the lowest free energy in the ensemble of structures, the pairs in these structures have been shown to be more likely to be correctly predicted than those in their MFE counterparts (58). The accuracy of this approach is further supplemented by the application of probing data in RSample, an algorithm that focuses on the agreement between experimental probing data and estimated probing data by the stochastic sampling of RNA structure models rather than the interpretation of data in the absence of models (59). RSample is the current standard by which SHAPE reactivities are integrated into structure prediction. All of this is combined with comparative sequence analysis in TurboFold II (32) to produce an accurate secondary-structure model.

The structural model we propose for the 5′ UTR of CV-B3 relies upon the results of RSample and TurboFold II, as constrained by our chemical modification values. In some of the proposed double-stranded regions of our model, including some positions in LR-2 and in the basal stem of domain V, there are examples of nucleotides that display moderate or high SHAPE reactivities. It is possible to see reactivity in unexpected places because the SHAPE reactivity value is only one part of the pseudo-free-energy calculation during the structural prediction. As more data accumulate, small adjustments may be necessary to better align the model with experimental results.

A remarkable feature of picornaviral 5′ UTRs is their ability to direct their functional activities independently. Importantly however, despite a well-organized and modular domain structure, these activities require the intact element. Single domains cannot be uniquely ascribed to a function; rather, the domains work in concert (60). Consequently, the structures of the domains and their relationship to each other become important precursors to understanding their functions. In this study, we have defined a new domain organization that directly informs CV-B3 and enteroviral infection efficiency and replication. Among the most influential new features are the long-range interactions, the rearrangement of domain V into a junction loop, and an intrinsically disordered RNA region. Our new model presents a number of targets where mutagenesis is predicted to have deleterious effects on replication.

RNA folding follows a hierarchical pathway in which the primary sequence first forms a base-paired secondary structure that subsequently folds into tertiary modules, which are based primarily on non-Watson-Crick pairing interactions (57). Consequently, a reliable secondary structure is the foundation upon which the RNA structure is built. Our current results provide solid support for a number of previously proposed structural elements and also establish important new elements. The 5′ cloverleaf (domain I), which is so important for replication of the viral genome (24), is confirmed by our probing results. Just downstream of the cloverleaf is a 10-nucleotide C-rich sequence that is largely protected from chemical modification but is not proposed to be part of a base-pairing interaction. This region is recognized by PCBP2 as part of the complex regulating genome replication (41, 61). Protection from modification suggests that these bases adopt a stable structure that restricts backbone flexibility and base accessibility, but the nature of that structure is not currently known. Also, in the connecting region between domains I and II, our model proposes a long-range interaction (LR-1) of 6 bp that involves positions 98 to 103 and 672 to 677. While this interaction is supported by experimental results from virulent CV-B3 strains, it is not supported by a more comprehensive comparison of enteroviral sequences.

In our model, domain II, also called SLII, spans positions 105 to 181. This complex stem-loop contains two bulge loops, a large internal loop, and a capping hairpin loop with a stable stem. Various models for the size and pairing arrangement of domain II have been proposed (6, 20, 28, 38). In our current model, the domain is identical to that in early proposals. A number of studies show that domain II modulates virulence (17, 18, 20, 42, 43). Interestingly, the portion of the domain between positions 118 and 127, where sequences differ the most between virulent and nonvirulent strains, is a long single-stranded region that is protected from modification. As described above for the bases connecting domains I and II, this region must be involved in a higher-order structure that has yet to be identified.

A major rearrangement is proposed for the connecting region between domains II and III, as well as domain III itself. The connecting region contains a long-range interaction (LR-2) involving positions 183 to 187 and 638 to 642. The proposed LR-2 interaction brings the functionally active elements in the 3′ half of the UTR into close proximity with those in the 5′ half. This interaction has the support of one pair with compensatory changes, but most of the bases are highly conserved when enterovirus A, B, and C genomes are compared. Evidence for base pairs is strongest when compensatory changes are present, and conservation of the bases on both sides of a stem is not, by itself, sufficient to prove a secondary structure (62). However, our analysis from TurboFold II combines sequence comparison with SHAPE and thermodynamic constraints. Our model accepted some pairs where SHAPE reactivity was moderately high on one side of the stem and some pairs where sequence comparison did not reveal compensatory base changes. However, the combination of SHAPE constraints, sequence comparison, and MEA values leads to a model consistent with all three lines of evidence. Domain III itself is reduced to a single stem-loop spanning positions 189 to 209. On the 3′ side of the stem-loop, positions 214 to 240, which were previously part of domain III, have been added to a new basal region of domain IV.

Many functional activities are ascribed to domain IV, where host ITAFs, such as PCBP2, and canonical translation initiation factors, such as eIF-4G and eIF-3, have specific interaction sites (48, 63, 64). As in previous proposals, our results show that domain IV has a very long and complex basal stem topped by a junction loop that radiates into three stem loops. Interestingly, nearly all of the functional activities conducted by domain IV map to the junction loop, particularly the C-rich 4a hairpin loop, 4b bulge loop, 4c GNRA tetraloop, and underlying helices (48, 64, 65). Despite its length and complexity, the long stem leading up to the junction loop has not been linked to specific functions. Structurally, the basal extension of the domain 4 stem proposed in our model is intriguing. A long single-stranded region between positions 223 and 240 separates the closing stem from the longer stem beyond. It is unusual for RNA to have such a lengthy unstructured region. However, our SHAPE results support the absence of structure in these positions.

The most significant new insights into enteroviral 5′ UTR structure are found in domain V. These findings are particularly important given the role of domain V in ITAF binding, canonical initiation factor binding, ribosome binding, and virulence (48, 49, 64). Our model for domain V establishes a stable basal helix of nine pairs that defines a large junction loop. The domain encompasses 170 nucleotides and contains three stem-loops that radiate from the central junction. The stem-loop on the 3′ side of the junction includes the bases assigned to domain VI of previous 5′ UTR models (6). The structure of these bases is identical to that of the previously proposed domain VI.

Within large RNA molecules, junction loops provide an architecture that enables conformational and functional flexibility (66). The radiating helices stack coaxially, and the bases in the central core often form tertiary interactions with distant regions (67). As such, they are fundamentally important in functionally active RNAs. The domain V junction loop joins the cloverleaf junction loop as an RNA module with well-established functional importance.

SHAPE analysis of domain V revealed intriguing behavior of the pyrimidine-rich tract between positions 561 and 581. This region has the Y_n-X_m-AUG motif, where Y_n is 8 to 10 pyrimidines, X_m is 18 to 20 nucleotides, and AUG is a cryptic start codon (33, 34). The Y_n-X_m-AUG region is known to be important for binding multiple factors, including the ribosome, and is present in type I, type II, and type III IRES elements (5, 35). The functional importance of the polypyrimidine region as part of the Y_n-X_m-AUG motif in the proposed IDRR was confirmed by point mutations and deletion mutations shortly after identification of the poliovirus IRES element (47). The same study proposed that the motif represented the ribosome landing pad. Since that time, the Y_n-X_m-AUG motif has been found to be a common feature of type I, II, and III picornavirus IRES elements (60) and has been defined as the binding site for the ribosome, canonical translation factors, and ITAFs (26). The polypyrimidine region has also been modeled as single-stranded region in previous models. Our results show that region 561 to 581 exists as a long single-stranded region in the central core of the domain V junction loop that displays high SHAPE reactivity on average. However, when reactivity is analyzed base by base, there is wild variability between individual runs at every position between 561 and 575, including one position (566) where reactivity fluctuates between highly positive and highly negative SHAPE values. We propose that this region, which includes the Y_n segment of the Y_n-X_m-AUG motif, is an IDRR. Intrinsically disordered regions in proteins have been known for some time (68), and they are particularly frequent in proteins that occupy central positions in protein interaction networks, where they bind to multiple partners (69). They are also found in linkers that separate well-structured domains (36). An IDRR in the Y_n-X_m-AUG motif of domain V of the enteroviral 5′ UTR may serve a parallel purpose.

MATERIALS AND METHODS

Viral plasmid construction. A full-length CV-B3 strain 28 (accession no. AY752944.2) genomic construct was engineered and kindly provided by Nora Chapman of the Enterovirus Research Laboratory,

TABLE 1 Primers for primer extension of CV-B3 strain 28 5′ UTR RNA

Primer extension starting position	Sequence (5′–3′)
168	GGTAACAGAAGTGCTTGATC
368	CCCCATGGGCAGGCCGCC
627	AGTCACCGGATGGCCAATCC
722	TTGCTGTATTCAACTTAAC

University of Nebraska Medical Center, Omaha, NE, USA. The construct was modified to contain a T7 promoter sequence (TAATACGACTCACTATAGGG) upstream of the 5′ UTR and a 38-nucleotide ribozyme sequence (ATGAGGCCGAAAGGCCGAAAACCCGGTATCCCGGGTTC). Upon in vitro transcription, the ribozyme sequence self-cleaves to yield the authentic 5′ uridine on the 5′ end of the transcript.

The CV-B3 strain 28 plasmid construct was isolated using a Qiagen (Valencia, CA, USA) QIAprep Spin Miniprep kit in accordance with the manufacturer's protocol. The isolated plasmid was digested using restriction enzyme Ecl136II to produce the template for in vitro transcription that yielded an RNA of 756 bases following ribozyme cleavage.

***In vitro* viral RNA transcription and purification.** The 5′ UTR was transcribed *in vitro* using a MEGAscript T7 transcription kit (Invitrogen, Carlsbad, CA, USA) according to the protocol provided by the manufacturer. The transcription reaction mixture included 4 μg of the digested template DNA in a total reaction volume of 80 μl. The reaction mixture was incubated at 37°C for 8 h, after which 4 μl of DNase I provided in the kit was added and incubated at 37°C for an additional 30 min to remove the template DNA. The reaction was stopped by adding 460 μl of nuclease-free water and 60 μl of 0.5 M ammonium acetate. The RNA was then purified with phenol-chloroform-isoamyl alcohol (25:24:1), precipitated by incubation with 600 μl of isopropyl alcohol overnight at −20°C, pelleted by centrifugation for 20 min at 4°C, washed with 400 μl of 70% ethanol, and resuspended in 80 μl of 1× TE buffer (5 mM Tris [pH 8.0] and 0.5 mM EDTA), pH 7.6. The RNA was further purified using a MEGAclear transcription clean-up kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. The purified RNA was stored in small aliquots at −80°C.

NMIA modification. NMIA modification reaction mixtures contained 2.4 μg of purified RNA in 60 μl of 0.5× TE buffer, pH 8.0. The RNA was denatured by incubation at 95°C for 3 min and immediately placed on ice. While it was on ice, 30 μl of 3.3× folding buffer (333 mM HEPES, pH 8.0, 20 mM MgCl$_2$, 333 mM NaCl) was added to the tube and incubated at 37°C for 20 min to allow the RNA to fold into a stable structure. These buffer conditions and folding conditions match those established and validated during the development of the SHAPE approach (56). Ten microliters of 32.5 mM NMIA solution in dimethyl sulfoxide (DMSO) was added to the RNA and incubated at 37°C for an additional 45 min. The modified RNA was precipitated by incubation with 4 μl of 50 mM EDTA, pH 8.0, 1 μl of 20-μg/μl glycogen, 10 μl of 3 M NaCl, and 300 μl of 100% ethanol at −80°C for 20 min and then pelleted by centrifugation for 20 min at 4°C. The pellet was washed twice with 150 μl of 70% ethanol, dried, and resuspended in 90 μl of 0.5× TE buffer. The control RNA was subjected to all the steps in the protocol except that 10 μl of DMSO was added instead of the NMIA solution.

DMS modification. DMS modification reaction mixtures contained 2.4 μg of purified RNA in 100 μl of DMS buffer (40 mM K-cacodylate, pH 7.2, 10 mM MgCl$_2$, 50 mM NH$_4$Cl). The RNA was denatured by incubation at 80°C for 2 min and slowly cooled to 45°C to allow the RNA to fold into a stable structure. These buffer and folding conditions were chosen to match those established and validated during previous base-specific chemical probing studies (6, 20). Two microliters of 0.4% DMS in 100% ethanol was added to the experimental samples, while 2 μl of DMS buffer was added to the control sample. The samples were then allowed to incubate for 10 min at 37°C. The reaction mixtures were then adjusted to 0.2 μg/μl glycogen, 4 mM EDTA (pH 8.0), and 300 mM NaCl and precipitated with 75% (vol/vol) ethanol. The RNA pellets were dissolved in 90 μl of 0.5× TE and divided into 9-μl (1-pmol) aliquots.

CMCT modification. CMCT modification reaction mixtures contained 2.4 μg of purified RNA in 50 μl of CMCT buffer (40 mM K-borate, pH 8.0, 10 mM MgCl$_2$, 50 mM NH$_4$Cl). The RNA was denatured by incubation at 80°C for 2 min and slowly cooled to 45°C to allow the RNA to fold into a stable structure. These buffer and folding conditions were chosen to match those established and validated during previous base-specific chemical probing studies (6, 20). Fifty microliters of CMCT buffer with or without 6.7 mg/ml CMCT was added and allowed to incubate for 10 min at 37°C. The reaction mixtures were then adjusted to 0.2 μg/μl glycogen, 4 mM EDTA (pH 8.0), and 300 mM NaCl and precipitated with 75% (vol/vol) ethanol. The RNA pellets were dissolved in 90 μl of 0.5× TE and divided into 9-μl (1-pmol) aliquots.

Primer extension. The modification sites were analyzed by reverse-transcriptase primer extension. Four primers were utilized to cover the entire 5′ UTR. The sequences and priming positions of the primers are provided in Table 1.

Two sets of the four oligonucleotide primers were designed: one set of primers was labeled on the 5′ end using VIC fluorescent dye, while the other set of primers was labeled with 6-carboxyfluorescein (6-FAM) fluorescent dye. The primer extension protocol was adopted from that of Burrill and Andino (70). Two fluorescently 5′-labeled versions (6-FAM and VIC) of each primer were used in primer extension. The extension reaction mixtures included experimental (plus) and control (minus) RNAs, as well as dideoxy-sequencing reaction mixtures needed for ShapeFinder analysis (see below). For each experiment, eight primer extension reactions were performed: 1 plus, 2 minus, and 3 to 8 for sequencing. For each run, 9 μl (1 pmol) of RNA was denatured at 95°C for 3 min and placed on ice. Two microliters of 1 μM VIC-labeled

primer was added to reaction mixtures 1 through 4, and the same amount of 6-FAM-labeled primer was added to reaction mixtures 5 through 8. The reaction mixtures were incubated at 65°C for 5 min and at 35°C for 10 min and placed on ice. Nine microliters of the appropriate RT mixture was added. The basic RT mixture (1 μl of SuperScript III [Invitrogen], 4 μl of 5× first-strand buffer [Invitrogen], 1 μl of 0.1 M dithiothreitol, 1 μl of the 10 mM mixture of deoxynucleoside triphosphates, and 2 μl of water) was added to reaction mixtures 1 and 2. The basic mixture was supplemented with 1 μl of 10 mM ddATP for reaction mixtures 3 and 5 through 8 and 1 μl of 10 mM ddCTP for reaction mixture 4. The reaction mixtures were incubated at 52°C for 15 min and placed on ice. To degrade the RNA template, 2.5 μl of 1 M NaOH was added to each reaction mixture, and the mixtures were incubated at 98°C for 15 min and then placed on ice. To neutralize the reaction, 2.5 μl of 1 M HCl was added. The VIC and 6-FAM reaction mixtures were pooled in a pairwise manner (1 and 5, 2 and 6, 3 and 7, and 4 and 8) and diluted 2-fold with water. The four pooled reaction mixtures were adjusted to 0.2 μg/μl glycogen and 300 mM sodium acetate (NaOAc) (pH 5.2) and precipitated by adding ethanol to 75% (vol/vol). The cDNA pellets were washed twice with 75% ethanol and dissolved in 10 μl of HiDi formamide (Life Technologies). Capillary electrophoresis was performed by the University of Nebraska Medical Center (UNMC) sequencing core facility according to their fragment analysis protocol. Each of the four pooled reaction mixtures was analyzed in a separate capillary during capillary electrophoresis.

Quantitative analysis of reactivities. Analysis of raw electropherograms, which contained data for fluorescence intensity versus elution time, was carried out as described previously (70). The data files were analyzed in ShapeFinder (37). Data from the 6-FAM and VIC channels for each of the four capillaries were first combined into a single file. A fitted baseline adjustment with a window width of 200 was applied to all eight channels. Cubic mobility shifts were performed manually on the four 6-FAM channels to align these identical ddATP sequencing reactions. The shift values from these alignments were then applied to the appropriate VIC channels, aligning the experimental (plus), control (minus), and sequencing channels with one another. The data from the four 6-FAM channels were then removed before continuing the analysis. Signal decay correction was applied to each of the four remaining VIC channels. The region of interest for the signal decay correction and further analysis was determined by visually inspecting the quality of peaks within the plus channel, and the same window was applied to all four channels. The rescale factor and equation parameters (A, q, and C) were kept at their default values of 10,000, 1,000,000, 0.999, and 10,000, respectively. A scale factor was then applied to the minus channel so that most peaks were of a height equal to or less than that of the corresponding plus peak. This factor was typically between 0.3 and 0.7. Alignment and integration were then performed over a range equal to or narrower than that used for signal decay correction. After manual inspection and correction of peaks, the ShapeFinder software performed a whole-channel Gaussian integration to quantify all individual peak areas, and the minus peak areas were subtracted from the plus areas. The raw reactivity data for each primer were normalized using a 2% to 8% normalization scheme (56). Accordingly, the top 2% of data were taken as outliers. The average of the next 8% of the data was found, and each datum, including the top 2%, was then divided by this average reactivity. At least three independent runs were completed for each primer. The normalized reactivity values for each 5′ UTR position were tested for outliers using Dixon's Q test at a 95% confidence interval. The outliers were excluded from further analysis.

RNA structural prediction. The RNA secondary structure was modeled with TurboFold II, a computational tool that incorporates experimental probing data with sequence comparison to inform predictions of RNA secondary structure (32). For the sequence comparison data set, genomes from 28 strains of coxsackievirus experimentally shown to be virulent and also present as full-length genomes in the database were selected. CV-B3 strain 28 was included in this data set. The first 700 nucleotides of these genomes were input into the TurboFold II analysis. SHAPE data were applied to CV-B3 strain 28 during the prediction, with RSample enabled using its default parameters (59). The SHAPE slope and intercept were changed to 2.6 and −0.8, respectively (56). The prediction was performed in MEA mode with default settings (58). The resultant connectivity table (CT) files were converted into image files with the draw algorithm from RNAstructure (71).

Multiple-sequence alignment. Analysis of sequence conservation for enterovirus 5′ UTRs was performed by multiple-sequence alignment in Clustal Omega using the default parameters (72). Complete full-length genomes of 1,653 viral sequences from the enterovirus A, B, and C species were obtained from the NCBI database (see Table S3 in the supplemental material). The CV-B3 strain 28 sequence (accession no. AY752944.2) was set as the reference. Sequences selected for the data set had to be greater than 5,000 nucleotides and had to be full length to the 5′ end. These sequences were trimmed to display the 5′-most 700 nucleotides and were loaded into Clustal Omega as a single FASTA file. The sequence alignment is provided as a Stockholm Alignment File (https://unomaha.box.com/s/2osxk3l8xizj3x6usbqd6sqym9qdt8vi).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JVI.01288-19.

SUPPLEMENTAL FILE 1, XLSX file, 0.01 MB.
SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.
SUPPLEMENTAL FILE 3, XLSX file, 0.1 MB.

REFERENCES

1. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res 46:D708–D717. https://doi.org/10.1093/nar/gkx932.
2. Baggen J, Thibaut HJ, Strating J, van Kuppeveld F. 2018. The life cycle of non-polio enteroviruses and how to target it. Nat Rev Microbiol 16:368–381. https://doi.org/10.1038/s41579-018-0005-4.
3. Pallansch MA, Oberste MS, Whitton JL. 2013. Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses, p 490–530. In Knipe DM, Howley PM (ed), Fields virology. Lippincott Williams & Wilkins, Philadelphia, PA.
4. Racaniello VR. 2013. Picornaviridae: the viruses and their replication, p 454–489. In Knipe DM, Howley PM (ed), Fields virology, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
5. Lozano G, Martinez-Salas E. 2015. Structural insights into viral IRES-dependent translation mechanisms. Curr Opin Virol 12:113–120. https://doi.org/10.1016/j.coviro.2015.04.008.
6. Bailey JM, Tapprich WE. 2007. Structure of the 5′ nontranslated region of the coxsackievirus b3 genome: chemical modification and comparative sequence analysis. J Virol 81:650–668. https://doi.org/10.1128/JVI.01327-06.
7. Lee KM, Chen CJ, Shih SR. 2017. Regulation mechanisms of viral IRES-driven translation. Trends Microbiol 25:546–561. https://doi.org/10.1016/j.tim.2017.01.010.
8. Martinez-Salas E. 2008. The impact of RNA structure on picornavirus IRES activity. Trends Microbiol 16:230–237. https://doi.org/10.1016/j.tim.2008.01.013.
9. Sharma N, Ogram SA, Morasco BJ, Spear A, Chapman NM, Flanegan JB. 2009. Functional role of the 5′ terminal cloverleaf in Coxsackievirus RNA replication. Virology 393:238–249. https://doi.org/10.1016/j.virol.2009.07.039.
10. Tracy S, Chapman NM, McManus BM, Pallansch MA, Beck MA, Carstens J. 1990. A molecular and serologic evaluation of enteroviral involvement in human myocarditis. J Mol Cell Cardiol 22:403–414. https://doi.org/10.1016/0022-2828(90)91476-n.
11. Tracy S, Höfling K, Pirruccello S, Lane PH, Reyna SM, Gauntt CJ. 2000. Group B coxsackievirus myocarditis and pancreatitis: connection between viral virulence phenotypes in mice. J Med Virol 62:70–81. https://doi.org/10.1002/1096-9071(200009)62:1<70::AID-JMV11>3.0.CO;2-R.
12. Reyes MP, Lerner AM. 1985. Coxsackievirus myocarditis–with special reference to acute and chronic effects. Prog Cardiovasc Dis 27:373–394. https://doi.org/10.1016/0033-0620(85)90001-5.
13. Maier R, Krebs P, Ludewig B. 2004. Immunopathological basis of virus-induced myocarditis. Clin Dev Immunol 11:1–5. https://doi.org/10.1080/10446670410001670427.
14. Harrison AK, Bauer SP, Murphy FA. 1972. Viral pancreatitis: ultrastructural pathological effects of Coxsackievirus B3 infection in newborn mouse pancreas. Exp Mol Pathol 17:206–219. https://doi.org/10.1016/0014-4800(72)90070-6.
15. Vuorinen T, Kallajoki M, Hyypia T, Vainionpaa R. 1989. Coxsackievirus B3-induced acute pancreatitis: analysis of histopathological and viral parameters in a mouse model. Br J Exp Pathol 70:395–403.
16. Rinehart JE, Gomez RM, Roos RP. 1997. Molecular determinants for virulence in coxsackievirus B1 infection. J Virol 71:3986–3991.
17. Dunn JJ, Bradrick SS, Chapman NM, Tracy SM, Romero JR. 2003. The stem loop II within the 5′ nontranslated region of clinical coxsackievirus B3 genomes determines cardiovirulence phenotype in a murine model. J Infect Dis 187:1552–1561. https://doi.org/10.1086/374877.
18. Dunn JJ, Chapman NM, Tracy S, Romero JR. 2000. Genomic determinants of cardiovirulence in coxsackievirus B3 clinical isolates: localization to the 5′ nontranslated region. J Virol 74:4787–4794. https://doi.org/10.1128/jvi.74.10.4787-4794.2000.
19. Yeh MT, Wang SW, Yu CK, Lin KH, Lei HY, Su IJ, Wang JR. 2011. A single nucleotide in stem loop II of 5′-untranslated region contributes to virulence of enterovirus 71 in mice. PLoS One 6:e27082. https://doi.org/10.1371/journal.pone.0027082.
20. Prusa J, Missak J, Kittrell J, Evans JJ, Tapprich WE. 2014. Major alteration in coxsackievirus B3 genomic RNA structure distinguishes a virulent strain from an avirulent strain. Nucleic Acids Res 42:10112–10121. https://doi.org/10.1093/nar/gku706.
21. Rohll JB, Percy N, Ley R, Evans DJ, Almond JW, Barclay WS. 1994. The 5′-untranslated regions of picornavirus RNAs contain independent functional domains essential for RNA replication and translation. J Virol 68:4384–4391.
22. Murray KE, Steil BP, Roberts AW, Barton DJ. 2004. Replication of poliovirus RNA with complete internal ribosome entry site deletions. J Virol 78:1393–1402. https://doi.org/10.1128/jvi.78.3.1393-1402.2004.
23. Lyons T, Murray KE, Roberts AW, Barton DJ. 2001. Poliovirus 5′-terminal cloverleaf RNA is required in cis for VPg uridylylation and the initiation of negative-strand RNA synthesis. J Virol 75:10696–10708. https://doi.org/10.1128/JVI.75.22.10696-10708.2001.
24. Barton DJ, O'Donnell BJ, Flanegan JB. 2001. 5′ Cloverleaf in poliovirus RNA is a cis-acting replication element required for negative-strand synthesis. EMBO J 20:1439–1448. https://doi.org/10.1093/emboj/20.6.1439.
25. Vogt DA, Andino R. 2010. An RNA element at the 5′-end of the poliovirus genome functions as a general promoter for RNA synthesis. PLoS Pathog 6:e1000936. https://doi.org/10.1371/journal.ppat.1000936.
26. Martinez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Lozano G, Diaz-Toledano R. 2015. Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res 206:62–73. https://doi.org/10.1016/j.virusres.2015.01.012.
27. Skinner MA, Racaniello VR, Dunn G, Cooper J, Minor PD, Almond JW. 1989. New model for the secondary structure of the 5′ non-coding RNA of poliovirus is supported by biochemical and genetic data that also show that RNA secondary structure is important in neurovirulence. J Mol Biol 207:379–392. https://doi.org/10.1016/0022-2836(89)90261-1.
28. Zell R, Stelzner A. 1997. Application of genome sequence information to the classification of bovine enteroviruses: the importance of 5′- and 3′-nontranslated regions. Virus Res 51:213–229. https://doi.org/10.1016/s0168-1702(97)00096-8.
29. Rivera VM, Welsh JD, Maizel JV, Jr. 1988. Comparative sequence analysis of the 5′ noncoding region of the enteroviruses and rhinoviruses. Virology 165:42–50. https://doi.org/10.1016/0042-6822(88)90656-3.
30. Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc 1:1610–1616. https://doi.org/10.1038/nprot.2006.249.
31. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127:4223–4231. https://doi.org/10.1021/ja043822v.
32. Tan Z, Fu Y, Sharma G, Mathews DH. 2017. TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. Nucleic Acids Res 45:11570–11581. https://doi.org/10.1093/nar/gkx815.
33. Jang SK, Pestova TV, Hellen CU, Witherell GW, Wimmer E. 1990. Cap-independent translation of picornavirus RNAs: structure and function of the internal ribosomal entry site. Enzyme 44:292–309. https://doi.org/10.1159/000468766.
34. Pestova TV, Hellen CU, Wimmer E. 1991. Translation of poliovirus RNA: role of an essential cis-acting oligopyrimidine element within the 5′ nontranslated region and involvement of a cellular 57-kilodalton protein. J Virol 65:6194–6204.

35. Pestova TV, Hellen CU, Wimmer E. 1994. A conserved AUG triplet in the 5′ nontranslated region of poliovirus can function as an initiation codon in vitro and in vivo. Virology 204:729–737. https://doi.org/10.1006/viro.1994.1588.

36. Dyson HJ. 2016. Making sense of intrinsically disordered proteins. Biophys J 110:1013–1016. https://doi.org/10.1016/j.bpj.2016.01.030.

37. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA 14:1979–1990. https://doi.org/10.1261/rna.1166808.

38. Liu Z, Carthy CM, Cheung P, Bohunek L, Wilson JE, McManus BM, Yang D. 1999. Structural and functional analysis of the 5′ untranslated region of coxsackievirus B3 RNA: In vivo translational and infectivity studies of full-length mutants. Virology 265:206–217. https://doi.org/10.1006/viro.1999.0048.

39. Xiang W, Harris KS, Alexander L, Wimmer E. 1995. Interaction between the 5′-terminal cloverleaf and 3AB/3CDpro of poliovirus is essential for RNA replication. J Virol 69:3658–3667.

40. Warden MS, Tonelli M, Cornilescu G, Liu D, Hopersberger LJ, Ponniah K, Pascal SM. 2017. Structure of RNA stem loop B from the picornavirus replication platform. Biochemistry 56:2549–2557. https://doi.org/10.1021/acs.biochem.7b00141.

41. Zell R, Ihle Y, Seitz S, Gundel U, Wutzler P, Gorlach M. 2008. Poly(rC)-binding protein 2 interacts with the oligo(rC) tract of coxsackievirus B3. Biochem Biophys Res Commun 366:917–921. https://doi.org/10.1016/j.bbrc.2007.12.038.

42. Bradrick SS, Lieben EA, Carden BM, Romero JR. 2001. A predicted secondary structural domain within the internal ribosome entry site of echovirus 12 mediates a cell-type-specific block to viral replication. J Virol 75:6472–6481. https://doi.org/10.1128/JVI.75.14.6472-6481.2001.

43. Lee CK, Kono K, Haas E, Kim KS, Drescher KM, Chapman NM, Tracy S. 2005. Characterization of an infectious cDNA copy of the genome of a naturally occurring, avirulent coxsackievirus B3 clinical isolate. J Gen Virol 86:197–210. https://doi.org/10.1099/vir.0.80424-0.

44. Burrill CP, Westesson O, Schulte MB, Strings VR, Segal M, Andino R. 2013. Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity. J Virol 87:11670–11683. https://doi.org/10.1128/JVI.01560-13.

45. Dildine SL, Semler BL. 1989. The deletion of 41 proximal nucleotides reverts a poliovirus mutant containing a temperature-sensitive lesion in the 5′ noncoding region of genomic RNA. J Virol 63:847–862.

46. Haller AA, Nguyen JH, Semler BL. 1993. Minimum internal ribosome entry site required for poliovirus infectivity. J Virol 67:7461–7471.

47. Nicholson R, Pelletier J, Le SY, Sonenberg N. 1991. Structural and functional analysis of the ribosome landing pad of poliovirus type 2: in vivo translation studies. J Virol 65:5886–5894.

48. de Breyne S, Yu Y, Unbehaun A, Pestova TV, Hellen CU. 2009. Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. Proc Natl Acad Sci U S A 106:9197–9202. https://doi.org/10.1073/pnas.0900153106.

49. Ben M'hadheb-Gharbi M, Gharbi J, Paulous S, Brocard M, Komaromva A, Aouni M, Kean KM. 2006. Effects of the Sabin-like mutations in domain V of the internal ribosome entry segment on translational efficiency of the Coxsackievirus B3. Mol Genet Genomics 276:402–412. https://doi.org/10.1007/s00438-006-0155-3.

50. Pelletier J, Flynn ME, Kaplan G, Racaniello V, Sonenberg N. 1988. Mutational analysis of upstream AUG codons of poliovirus RNA. J Virol 62:4486–4492.

51. Yang D, Cheung P, Sun Y, Yuan J, Zhang H, Carthy CM, Anderson DR, Bohunek L, Wilson JE, McManus BM. 2003. A Shine-Dalgarno-like sequence mediates in vitro ribosomal internal entry and subsequent scanning for translation initiation of coxsackievirus B3 RNA. Virology 305:31–43. https://doi.org/10.1006/viro.2002.1770.

52. Lindberg AM, Stalhandske PO, Pettersson U. 1987. Genome of coxsackievirus B3. Virology 156:50–63. https://doi.org/10.1016/0042-6822(87)90435-1.

53. De Jesus N, Franco D, Paul A, Wimmer E, Cello J. 2005. Mutation of a single conserved nucleotide between the cloverleaf and internal ribosome entry site attenuates poliovirus neurovirulence. J Virol 79:14235–14243. https://doi.org/10.1128/JVI.79.22.14235-14243.2005.

54. Haller AA, Stewart SR, Semler BL. 1996. Attenuation stem-loop lesions in the 5′ noncoding region of poliovirus RNA: neuronal cell-specific translation defects. J Virol 70:1467–1474.

55. Busan S, Weeks KM. 2018. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. RNA 24:143–148. https://doi.org/10.1261/rna.061945.117.

56. Low JT, Weeks KM. 2010. SHAPE-directed RNA secondary structure prediction. Methods 52:150–158. https://doi.org/10.1016/j.ymeth.2010.06.007.

57. Miao Z, Westhof E. 2017. RNA structure: advances and assessment of 3D structure prediction. Annu Rev Biophys 46:483–503. https://doi.org/10.1146/annurev-biophys-070816-034125.

58. Lu ZJ, Gloor JW, Mathews DH. 2009. Improved RNA secondary structure prediction by maximizing expected pair accuracy. RNA 15:1805–1813. https://doi.org/10.1261/rna.1643609.

59. Spasic A, Assmann SM, Bevilacqua PC, Mathews DH. 2018. Modeling RNA secondary structure folding ensembles using SHAPE mapping data. Nucleic Acids Res 46:314–323. https://doi.org/10.1093/nar/gkx1057.

60. Martinez-Salas E, Francisco-Velilla R, Fernandez-Chamorro J, Embarek AM. 2018. Insights into structural and mechanistic features of viral IRES elements. Front Microbiol 8:2629. https://doi.org/10.3389/fmicb.2017.02629.

61. Toyoda H, Franco D, Fujita K, Paul AV, Wimmer E. 2007. Replication of poliovirus requires binding of the poly(rC) binding protein to the cloverleaf as well as to the adjacent C-rich spacer sequence between the cloverleaf and the internal ribosomal entry site. J Virol 81:10017–10028. https://doi.org/10.1128/JVI.00516-07.

62. Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods 14:45–48. https://doi.org/10.1038/nmeth.4066.

63. Gamarnik AV, Andino R. 2000. Interactions of viral protein 3CD and poly(rC) binding protein with the 5′ untranslated region of the poliovirus genome. J Virol 74:2219–2226. https://doi.org/10.1128/jvi.74.5.2219-2226.2000.

64. Sweeney TR, Abaeva IS, Pestova TV, Hellen CU. 2014. The mechanism of translation initiation on type 1 picornavirus IRESs. EMBO J 33:76–92. https://doi.org/10.1002/embj.201386124.

65. Sean P, Nguyen JH, Semler BL. 2009. Altered interactions between stem-loop IV within the 5′ noncoding region of coxsackievirus RNA and poly(rC) binding protein 2: effects on IRES-mediated translation and viral infectivity. Virology 389:45–58. https://doi.org/10.1016/j.virol.2009.03.012.

66. Cruz JA, Westhof E. 2009. The dynamic landscapes of RNA architecture. Cell 136:604–609. https://doi.org/10.1016/j.cell.2009.02.003.

67. Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 126:309–320. https://doi.org/10.1016/j.cell.2006.06.036.

68. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci U S A 93:11504–11509. https://doi.org/10.1073/pnas.93.21.11504.

69. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272:5129–5148. https://doi.org/10.1111/j.1742-4658.2005.04948.x.

70. Burrill CP, Andino R. 2013. RNA structure analysis of viruses using SHAPE. Curr Protoc Microbiol Chapter 30:Unit 15H.3. https://doi.org/10.1002/9780471729259.mc15h03s30.

71. Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11:129. https://doi.org/10.1186/1471-2105-11-129.

72. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75.