Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

In silico survey of the central conserved regions in viroids of the *Pospiviroidae* family for conserved asymmetric loop structures

PAUL FREIDHOFF and MICHAEL F. BRUIST

Department of Chemistry and Biochemistry, University of the Sciences, Philadelphia, Pennsylvania 19104, USA

ABSTRACT

Viroids are the smallest replicative pathogens, consisting of RNA circles (~300 nucleotides) that require host machinery to replicate. Structural RNA elements recruit these host factors. Currently, many of these structural elements and the nature of their interactions are unknown. All *Pospiviroidae* have homology in the central conserved region (CCR). The CCR of potato spindle tuber viroid (PSTVd) contains a sarcin/ricin domain (SRD), the only viroid structural element with an unequivocal replication role. We assumed that every member of this family uses this region to recruit host factors, and that each CCR has an SRD-like asymmetric loop within it. Potential SRD or SRD-like motifs were sought in the CCR of each *Pospiviroidae* member as follows. Motif location in each CCR was predicted with MUSCLE alignment and Vienna RNAfold. Viroid-specific models of SRD-like motifs were built by superimposing noncanonical base pairs and nucleotides on a model of an SRD. The RNA geometry search engine FR3D was then used to find nucleotide groups close to the geometry suggested by this superimposition. Atomic resolution structures were assembled using the molecular visualization program Chimera, and the stability of each motif was assessed with molecular dynamics (MD). Some models required a protonated cytosine. To be stable within a cell, the pK_a of that cytosine must be shifted up. Constant pH-replica exchange MD analysis showed such a shift in the proposed structures. These data show that every *Pospiviroidae* member could form a motif that resembles an SRD in its CCR, and imply there could be undiscovered mimics of other RNA domains.

Keywords: molecular dynamics; RNA domains; RNA dynamics; RNA mimicry; sarcin/ricin domain; constant pH replica exchange molecular dynamics; Loop E RNA; S-turn

INTRODUCTION

A critical conundrum in molecular biology today is the disparity between the central role of protein-encoding mRNA in the central dogma (Crick 1970) while mRNA only encompasses ~2% of the human genome. Meanwhile, it was discovered that up to 90% of that genome is transcribed (ENCODE Project Consortium 2007; Mattick 2009). We now know that many of these noncoding RNAs (ncRNAs) affect phenotype through regulatory, structural, catalytic, and trafficking roles in the cell. New classes and functions of RNA are being discovered as biochemical knowledge and techniques progress (Dozmorov et al. 2013; Cech and Steitz 2014).

The function of an RNA is linked to its structure. RNA structure can be dissected into motifs, which are conserved three-dimensional arrangements of nucleotides. These include standard A-form RNA but also embrace a vast catalog of nucleotide arrangements featuring noncanonical base pairs (ncbp), nucleotide interactions where the hydrogen bonding occurs between atoms in addition to those on the traditional "Watson–Crick" edges of the bases (Fig. 1A; Leontis and Westhof 2001; Leontis et al. 2002; Stombaugh et al. 2009). This work has shown that motifs share geometry rather than sequence. Just as any canonical base pair may fit in an A-form helix, many noncanonical pairings are isosteric and can substitute for one another (Moore 1999; Hendrix et al. 2005). Examples of such

Corresponding author: m.bruist@usciences.edu

Abbreviations: CCR, central conserved region; CProp, communication propensity; LCCR, lower central conserved region; LRCProp, long range communication propensity; mSRD, mimic of SRD; ncbp(s), noncanonical base pair(s); pH-REMD, pH-Dependent replica exchange molecular dynamics; RMSD, root mean square deviation; SRD, sarcin/ricin domain; UCCR, upper central conserved region; H, Hoogsteen edge; S, sugar edge; W, Watson-Crick edge. Viroid abbreviations may be found in the legend to Figure 2.

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.070409.119.

© 2019 Freidhoff and Bruist This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

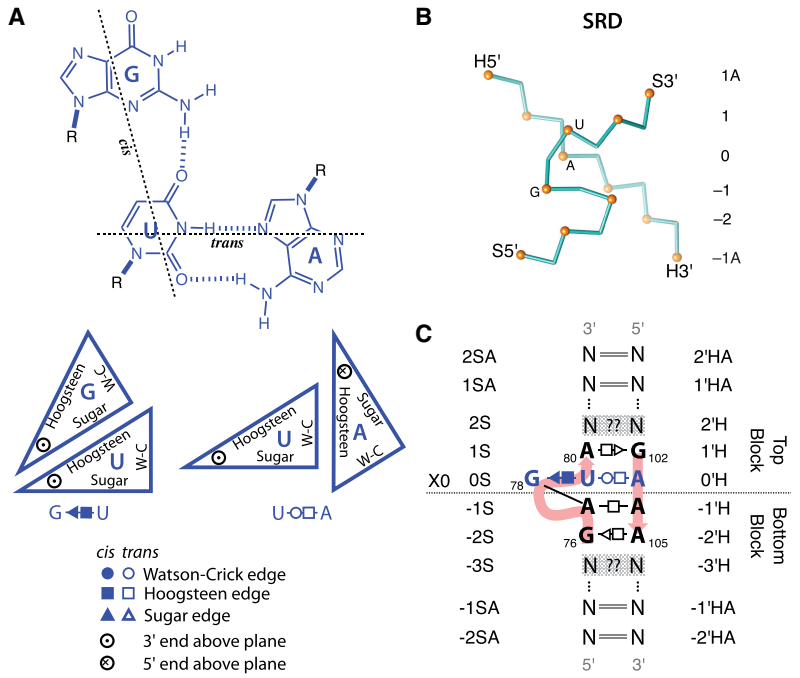

FIGURE 1. (*A*) The Westhof–Leontis (Leontis and Westhof 2001) nomenclature for ncbps in RNA is demonstrated using the triplet from a standard sarcin/ricin domain (SRD). (*B*) The S- and H-strands of the SRD in the 5S RNA of *H. marismortui* [PDBid 1S72, (Klein et al. 2004)]. The backbone is depicted using pseudobonds connecting the phosphorus (gold circle) and C4′ of each nucleotide. Depth cueing shows the H-strand is behind the S-strand. The bases in the triplet have been labeled. The levels described in part *C* are given to the *left*. (*C*) The 5S SRD of PDBid 1S72 has been labeled with the proposed SRD-like numbering system. The backbone paths of the S-strands meander and the H-strands are traced in red. N indicates any base; the double *horizontal* lines indicate any canonical or wobbled pairing. The nucleotides with the patterned background are optional extensions of the SRD-like domain. The double question marks indicate an unspecified noncanonical pairing.

geometric forms include kink-turns (Klein et al. 2001), A-minor loops (Nissen et al. 2001), and sarcin/ricin domains (SRD) (Branch et al. 1985; Correll et al. 1999). Such motifs act as recognition points and structural scaffolds for intermolecular interactions including RNA–protein, RNA–RNA, and RNA–small molecule interactions (Hermann and Patel 2000). They also serve as architectural and dynamic elements bending RNA or creating hinges (Klein et al. 2001; Mohan and Noller 2017).

One of the first ncRNAs studied was potato spindle tuber viroid (PSTVd; viroid abbreviations are listed separately in the legend to Fig. 2; Riesner et al. 1979). Viroids are small circular RNAs that infect plants, including economically important crops (Katsarou et al. 2015). There are 32 species of viroids classified as of 2014, four of them belong to the *Avsunviroidae* family, and 28 species, including PSTVd, belong in the *Pospiviroidae* family (Fig. 2; Di Serio et al. 2014). The 28 species in the *Pospiviroidae* family, along with their standard abbreviations are provided in Figure 2. Viroids are classified as infectious long noncoding RNAs (lncRNA), a class of RNAs tied to regulation of cell processes including gene modulation, development, and even viral infection and cancer (Mercer et al. 2009; Ponting et al. 2009; Chen et al. 2018). Some of these functions have been linked to conserved motifs; however, the exact mechanism of these lncRNAs are still being determined (Kung et al. 2013; Johnsson et al. 2014) while viroid structural–functional relationships are better understood. Thus, viroids serve as excellent models for the study of RNA structural elements and how they determine function inside of the cell.

Unlike viruses, viroids do not have a capsid, nor do they encode proteins (Diener and Raymer 1967). Therefore, viroids must hijack host cell mechanisms to replicate. Viroids belonging in the *Pospiviroidae* family replicate via an asymmetric rolling circle. The infecting (+)strand circle directs the synthesis of a multigenome linear (–) strand, which in turn directs the synthesis of multigenome linear (+) strands. These (+)strand multimers are cut by a host endonuclease to create monomers that are ligated into circles by a host ligase. Viroids perform this processing using multiple RNA structural elements to recruit and direct host proteins (Zhong et al. 2008). One of these motifs is the SRD (also known as Loop E in this context) of the central conserved region (CCR) in PSTVd, which plays a significant role in the ligation step as determined by mutational analysis and in vitro RNA processing assays (Semancik et al. 1993; Baumstark et al. 1997; Zhong et al. 2006).

The processing endonuclease does not cleave monomeric circles. It is thought that the viroid CCR is a context-driven RNA switch. In the case of a multigenomic strand, an RNA sequence longer than the monomer folds into a structure that the endonuclease recognizes (Baumstark et al. 1997; Gas et al. 2007). This may be a metastable structure that will rearrange to a different native state. That structure targeted by the endonuclease cannot form in a monomeric circle because the repeated sequences are not present. After the endonuclease acts, removing the duplicated sequences, the RNA folds into an alternative ligation-supporting structure.

The CCR of PSTVd includes the loop E SRD. In the postulated RNA switch, the two strands of the SRD are separated in the endonuclease-substrate fold. The rearrangement triggered by cleavage brings the SRD strands together as part of a structure that directs the ligase. The RNA of the

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

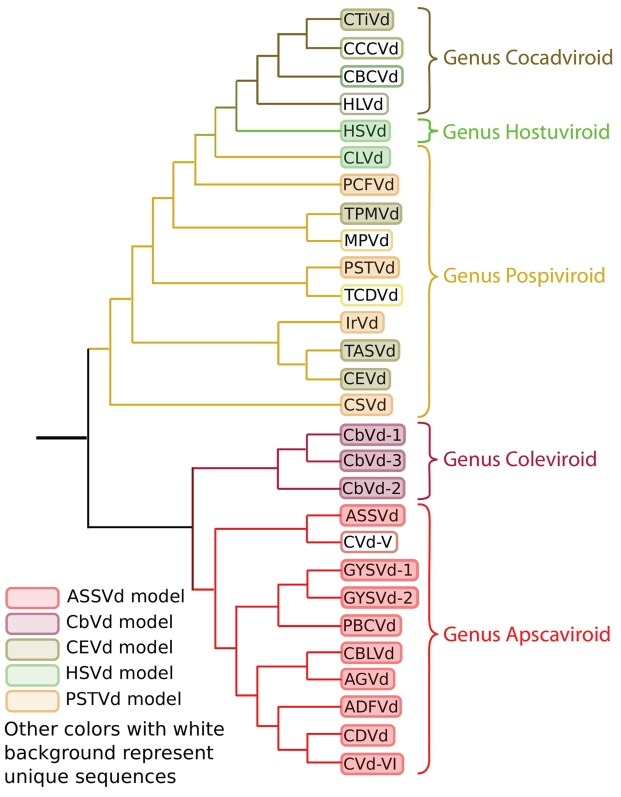

FIGURE 2. Taxonomy of *Pospiviroidae* (Di Serio et al. 2014) color coded by SRD-like internal loop sequence. *Horizontal* spacing does not correlate to precise evolutionary distance. ADFVd, Apple dimple fruit viroid; ASSVd, Apple scar skin viroid; AGVd, Australian grapevine viroid; CSVd, Chrysanthemum stunt viroid; CBCVd, Citrus bark cracking viroid; CBLVd, Citrus bent leaf viroid; CDVd, Citrus dwarfing viroid; CVd-V, Citrus viroid V; CVd-VI, Citrus viroid VI; CEVd, Citrus exocortis viroid; CCCVd, Coconut cadang-cadang viroid; CTiVd, Coconut tinangaja viroid; CbVd1, Coleus blumei viroid 1; CbVd2, Coleus blumei viroid 2; CbVd3, Coleus blumei viroid 3; CLVd, Columnea latent viroid; GYSVd-1, Grapevine yellow speckle viroid 1, GYSVd-2, Grapevine yellow speckle viroid 2; HLVd, Hop latent viroid; HSVd, Hop stunt viroid; IrVd Iresine viroid 1; MPVd, Mexican papita viroid; PBCVd, Pear blister canker viroid; PCFVd, Pepper chat fruit viroid; PSTVd, Potato spindle tuber viroid; TASVd, Tomato apical stunt viroid; TCDVd, Tomato chlorotic dwarf viroid; TPMVd, Tomato planta macho viroid.

CCR must have kinetic access to each of two possible folds, hence these folds are metastable, and they have similar energies. If the ligation conformation (SRD) is too stable, the RNA will switch before the endonuclease acts, causing aberrant circles. If the ligation conformation has insufficient stability, the RNA may be degraded at the new ends before they can be ligated (Friday et al. 2017).

While all members of the *Pospiviroidae* family use the asymmetric rolling circle pathway, not all members of the family have a recognized SRD near their ligation site, casting doubt on its role in processing (Ding 2010; Giguère and Perreault 2017). Figure 2 shows the currently accepted phylogenetic tree of viroids (Di Serio et al. 2014). Most of the members of *Pospiviroid*s and *Cocadviroids* have a recognized SRD. For members of the *Apscaviroids*, there is a recognizable upper block for the SRD, but the sequence where the lower block should reside has only a single ncbp. For some other viroids such as the *Coleviroids* (CbVd) and hop stunt viroid (HSVd), neither a top nor bottom block is discernable. In this paper, we show that upon more careful investigation of *Pospiviroidae* CCR sequences, we find that every viroid of this family has the potential to fold into an SRD-like domain. This supports models that propose an SRD motif is critical in replication, e.g. ligation.

Here we provide a brief description of the SRD domain and a numbering system that simplifies comparisons between species. More details may be found in the first section of Materials and Methods. We have adopted the 5S rRNA SRD (PDB 1S72) as standard for the domain (Fig. 1B,C). It has four ncbps and a base triplet. The overall structure has a strand whose backbone describes an S-turn (the S-strand, Fig. 1B,C), the other strand follows a more normal helical path (the H-strand). The base triplet sets the core. It has a "U~A handle" (Jaeger et al. 2009) that interacts with an extrahelical G. In our standard view, there is one ncbp above the triplet and three ncbps below it. The numbering system (Fig. 1C) is based on levels and strands and simplifies descriptions of base and geometry substitutions as well as insertions and deletions. The triplet is level 0, ncbps above it get positive numbers while those below get negative numbers. The S-turn strand gets an "S" and is numbered 5′–3′. The helical strand gets an "H" and is numbered 3′–5′. The numbers get a prime to emphasize this reversed numbering. The extrahelical nucleotide of the S-strand gets an "X" designation. The numbering resets at the boundary to A-form RNA. The S-strand nucleotide in the first canonical base pair above the domain is [1SA], while the H-strand nucleotide in the first canonical pair below the domain is [–1′HA].

Despite decades of research on viroids, knowledge about the three-dimensional structure of their RNA elements and the roles they have in viroid replication is incomplete. Some of the functional elements in PSTVd have been modeled, including the SRD [also called Loop E] (Zhong et al. 2006), loop 6 (Takeda et al. 2011), and the right terminal domain (Steger 2017). Progress is being made on several fronts in the elucidation of RNA structure and dynamics: combining chemical probe susceptibilities with structure comparison and prediction data (Tian et al. 2014; Giguère and Perreault 2017), cryo-electron microscopy (Haselbach et al. 2018), molecular dynamics (MD) (Smith et al. 2017; Šponer et al. 2018), and combinations of these approaches (Zhang et al. 2018), have all made important progress. MD is very useful in describing the dynamics of macromolecules. It has already been used to characterize the dynamics of SRD domains (Spackova and Šponer 2006; Spasic et al. 2018). Furthermore, the SRD has been used as a benchmark to test the quality

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

of force fields for MD (Havrila et al. 2013; Šponer et al. 2018).

We used "structure gazing" (Woods and Laederach 2017) and computer-assisted RNA structure searching FR3D (Sarver et al. 2008; Smith et al. 2017) to manually assemble models in which the nucleotides of candidate sequences adopt a structure with a backbone shape similar to an SRD. These models were refined and evaluated using MD. In searching for isosteres, we have tried to expand the candidate pool by considering nucleotides with *syn* glycosidic bonds (Sokoloski et al. 2011) and protonated nucleobases. Here we present candidate structures for each of the 28 members of the *Pospiviroidae* family, a strong indication that an SRD-like structure is universally required for viroid RNA processing.

The proposed structures are in a portion of the CCR that is postulated to be pleomorphic. Considering the fact that the same sequence of RNA, even when it is not a part of a switch, can assume different conformations (López-Carrasco and Flores 2017; Rangadurai et al. 2018), we acknowledge that these are only possible conformations that these sequences can adopt, either transiently or as their native state. We present considerable detail for our models in order to illustrate the chemical consistency of these structures. The actual structures, once determined, will certainly show many differences.

RESULTS AND DISCUSSION

It is known that SRDs occur in many members of the *Pospiviroidae* family, the SRD has demonstrated function in processing (Baumstark et al. 1997; Zhong et al. 2006; Eiras et al. 2007) of PSTVd, all *Pospiviroidae* have a CCR, and that all *Pospiviroidae* follow the same replication pathway. Therefore, we postulate that there is a conserved structural element resembling an SRD in the CCR of all viroids of this family. In order to test this hypothesis, sequences in the CCR of each of the *Pospiviroidae* were superimposed on an SRD, and the resulting models were tested for stability using MD. The standard minimal SRD we chose was the SRD 5S rRNA of the *H. marismortui*, nucleotides 76–80 and 102–105 (Fig. 1B; Klein et al. 2004).

Establishing the internal loops to model

The 28 *Pospiviroidae* viroids were aligned around their upper and lower CCR strands (UCCR and LCCR, respectively). Segments corresponding to PSTVd's UCCR nt 80–110 and LCCR nt 248–282 were used. These were paired by complementarity and joined to sequences to make left and right G-C-rich hairpins to fix the strand register, creating a virtual set of RNA minicircles (Schrader et al. 2003; Supplemental Fig. S1). In the case of HSVd-like minicircles, we chose probable locations for an SRD-like motif based on the known cleavage site of HSVd, apple scar skin viroid (ASSVd), and citrus exocortis viroid (CEVd) (Gas et al. 2007), the length of the internal loops of the PSTVd-like minicircles, sequence identity between HSVd and columnea latent viroid (CLVd), and partial sequence conservation with a sequence that is identical in all three *Coleviroids*. Two versions of the HSVd-like internal loop were seen, one for both HSVd and CLVd, and one for CbVd.

The secondary structure of each viroid minicircle was analyzed with RNAfold (Hofacker and Stadler 2006). Each folded minicircle had an internal loop in roughly the same location. The internal loop for PSTVd corresponded to the previously recognized loop E (Branch et al. 1985), which is an SRD variant. Ding's laboratory modeled this domain by making mutations with properties consistent with base pair isostericities (Zhong et al. 2006). In the case of the HSVd-like regions, RNAfold created two or four canonical base pairs in these regions. However, SHAPE mapping indicates that these viroids have either single-stranded RNA or ncbps at these positions (Giguère et al. 2014). Hence, we assume that these bases participate in ncbps rather than canonical pairs and designate the whole region as an internal loop.

In all, 11 different internal loops occurred in this section of the CCR in the 28 *Pospiviroidae* minicircles. The sequences are reported in Supplemental Figure S1 in Supplemental information and Table Construction of the Supplemental Information. Figure 2 color codes viroids according to internal loop sequence. These internal loops fall into two groups, PSTVd-like (nine sequences: 23 viroids) and HSVd-like (two sequences: HSVd and CLVd and the *Coleviroids*).

With the exception of CLVd, all of the *Pospiviroids* and *Cocadviroids* have a CCR consistent with the known sequence variants of the SRD. In the model for PSTVd proposed by Ding's laboratory, the top block has the standard capping A tHS G (PSTVd model A[1S] and G [1′H]) and a C~U~A base triplet (C[0X], U[0S], and A[0′H]) instead of a G~U~A triplet (Zhong et al. 2006). In the bottom block, the A tH A at the –1 level is conserved. U replaces G in the –2S position, a substitution recognized as near isosteric (Leontis et al. 2002), and an extra ncbp occurs at the bottom (Fig. 3A, left). Other viroids in this group have sequence variations that are also isosteric or near isosteric with the standard SRD. For example, CEVd, has the same sequence as PSTVd for the helical strand. In the S-strand, position –2S is G, the same as the standard SRD, and –S3 is A, which can make a wobble *c*WW pair with C.

For the *Apscaviroids*, their internal loop sequences support a top block of an SRD, but the ncbps that would form in the bottom block are not recognized as isosteric with those in PSTVd. The CCRs of CbVd, HSVd, and CLVd, do not have sequences discernable as either the top or bottom block of an SRD. We will show that all of these loops can, in theory, be folded into structural motifs that resemble the SRD.

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

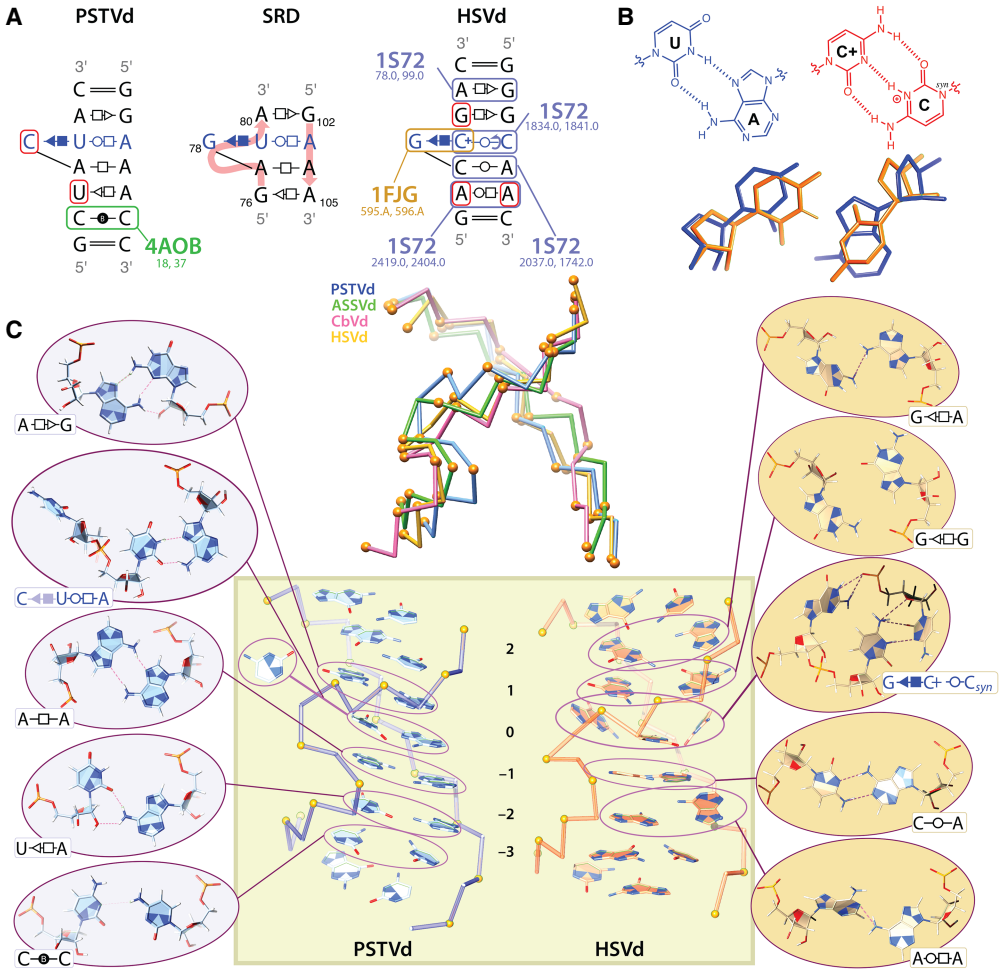

FIGURE 3. Proposed structures for the SRD in PSTVd and mSRD in HSVd. (*A*) Sequence of the internal loops from PSTVd and HSVd and the sources (PDBid) for the ncbp structures used in the production of the models. (*B*) A comparison of the U-A handles from the SRD of PSTVd and the C+ *t*WW C_{syn} ncbp of the triplet in the mSRD of HSVd. The *top* pair shows an idealized structure with explicit hydrogen bonding. *Below* that are the pairs from the triplets shown in part *C*. (*C*) Structures of the centroids from the major cluster for PSTVd and HSVd. The *top middle* shows a P-C4′ backbone of four groups: two SRD (PSTVd and ASSVd) and two mSRD (CbVd and HSVd). The phosphorus and C4′ atoms from the core of each domain fitted to the PSTVd backbone (H and S levels –2, –1, 0, and 1). The *bottom center* shows the PSTVd and HSVd SRD-like domains from the same point of view as the backbone structure *above*. Level is given between the two structures. Each noncanonical nucleotide pair or triplet is shown on the outside. The nucleotide labels follow the view in the *inset*, not the order in part *A*. Note that the triplet from PSTVd is in two ovals connected by a line. The symbol for the interaction of the C with the U-A handle is faded to indicate that this interaction is not maintained.

Proposed three-dimensional models for the internal loops

Possible three-dimensional structures for all 11 internal loops were produced by superimposing each loop sequence onto the standard SRD. The RNA structure search engine FR3D (Sarver et al. 2008) was used to find base pairs or triplets with the base identity of the viroid and the interaction geometry of the standard SRD. Where possible, base pairs with neighbors resembling the geometry of the neighbors in SRD were used. For example, in producing a model for the SRD-like loop in HSVd, an additional *t*HS ncbp is needed in the top block. This was modeled after pair G78 *t*HS A99 in the 28S rRNA of PDB 1S72, which transitions from a *t*HS ncbp to A-form RNA. In some cases, the desired pair was not found; in those instances, a related base pair was used and one or both bases replaced to match the required sequence. Figure 3A describes the source for modeling the SRD-like domains of PSTVd and HSVd. This information for all 11 models is given in Table Constructs. Some noteworthy changes are also described below in the description of each of the 11 SRD-like motifs.

The motifs were assembled using the molecular visualization and manipulation program Chimera (Pettersen et al. 2004). Bond angles of the phosphate and C5′ atoms were modified manually to approximate proper chemical dimensions and reduce clashes. The resulting models may be viewed at https://sites.google.com/usciences.edu/westcenter/databases described in the Supplemental Information.

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

MD results

These structures were simulated in the MD package AMBER14 (Case et al. 2005, 2014). They were surrounded by sufficient K^+ ions for neutralization and explicit water. Models were then minimized and equilibrated. During the initial equilibration restraints bound desired ncbps and base-phosphate interactions. These restraints were diminished and removed by the end of equilibration. Production runs were performed at constant volume at 300 K. The first 100 psec or more of the initial production run were not used in analyses.

In some cases, the models were not stable and required manual intervention. Because our RNA models use short chains with no loops, fraying of ends could allow the end residues to interact with the noncanonical bases. These interactions were minimized by changing the identity of the terminal bases, extending the stem by one base pair, or in the case of citrus bark cracking viroid (CBCVd) the terminal restraint was increased (see Table_for_SRD-Like_Constructs.xlsx in the Supplemental Material). Finding a stable state for the HSVd and CbVd internal loops required important substantive changes: The central cytosine was protonated (C+) and the base of its partner cytosine was rotated to a *syn* glycosidic bond, creating a pair similar to that seen in the i-motif or i-wire (Guéron and Leroy 2000). With these changes, we were able to find stable SRD or SRD-like structures in the CCR of every member of the *Pospiviroidae* family. None of the changes made altered the classification of these structures as PSTVd-like or HSVd-like, nor did any call for the introduction of another category. All of the PSTVd-like internal loops could make an SRD following the geometry proposed by Zhong et al. (2006). The HSVd-like internal loops could make an SRD-like domain with an S-turn strand and a helical strand, but with significantly different nucleoside interaction geometries (Fig. 3). We propose calling this an mSRD for mimic of an SRD. This domain has a central triplet based on a C+ *t*WW C_{syn} pair that has a glycosidic bond geometry resembling the U~A handle (Fig. 3B). The sugar (S) edge of guanine is adjacent to the Hoogsteen (H) edge of the protonated cytosine; however, they share no hydrogen bonds. The Watson–Crick (W) edge of this guanine does hydrogen bond to the phosphate of the unprotonated cytosine. The top block of the mSRD is completed with two A *t*HS G or G *t*HS G ncbps. The bottom block has a *t*WW pair under the triplet and one *t*WH ncbps transitioning to A-form RNA (Fig. 3A, right).

In each of our models, the top ncbp that makes the transition to A-form RNA is an A *t*HS G. In contrast, the triplet and the lower block are more variable, both in terms of the types of interactions creating the ncbps and in the number of ncbps. The triplet of the PSTVd-like domains is immediately below the top ncbp as in the 5S rRNA SRD. However, the extrahelical C does not interact strongly with the U~A handle; the domain adjusts to compensate for the loss of stacking with the extrahelical G of the standard SRD. The HSVd-triplet has an extra ncbp above it and is unrelated to the 5S rRNA triplet. It is discussed below. The lower block of both the PSTVd- and HSVd-like domains have a variable number of ncbps. Despite this, the extent of the S-turn is the same in every model. The bottom end of the S-turn [–2 and below] can accommodate canonical, wobbled *c*WW pairs, or other similar geometries.

Frames from 500 nsec MD trajectories for the each of the 11 models were analyzed by hierarchical agglomerative clustering using CPPTRAJ (Roe and Cheatham 2013) of AMBER (Case et al. 2014), and the outcome is presented in Table 1. The analysis divides a model's trajectory into clusters with similar conformations using a bottom-up approach. The 11 tables present data for all clusters representing 1% or more of the frames in a trajectory. The mean RMSD for each cluster (AvgDist) is given as well as the mean RMSD between clusters (AvgCDist). In each case, clusters have a mean RMSD of 0.8–1.3 Å, and clusters differed from one another by RMSD values of 1.2–2.3 Å. The number of clusters needed to include 90% of the frames is given above each viroid cluster table. All these measures indicate stable structures.

Constant pH MD supports a triplet with a protonated cytosine

N3 of cytosine is a weak base with an intrinsic pK_a of 4.2 in water (Thaplyal and Bevilacqua 2014), almost 3 units below neutral pH. Thus, normal cytosine is not detectably protonated in the cytosol of a plant cell. However, the pK_a of a residue within a biological macromolecule can shift considerably in response to its microenvironment. Charges, hydrogen bonding, hydration, and other factors in the surroundings that favor either the conjugate acid or base over the other will perturb the pK_a. Such perturbations have been documented experimentally (Isom et al. 2011) and reproduced computationally in both proteins and nucleic acids (Chen et al. 2017). Shifts greater than five pH units have been observed (Isom et al. 2011).

Constant pH Replica Exchange MD (pH-REMD) (Swails et al. 2014) was used to predict pK_a's of cytosines in our RNA models. In this method, the MD program generates a master trajectory as well as replicas that have the protonated and unprotonated states at different pH values. At regular intervals, the free energies of the master trajectory and one of the replicas are weighted by the chemical potential of H^+ (pH) and compared.

A Metropolis Monte Carlo criterion decides which state will be continued in the master trajectory. The pH weighting potential that makes the two protonation states equal in internal energy gives the pK_a. Similar results were obtained with constant pH MD (Mongan et al. 2004) which uses only two trajectories for exchange (data not shown).

TABLE 1. MD trajectory clusters

PSTVd-like internal loops

	PSTVd				CEVd				MPVd				TCDVd			
	90% clst #:		2		90% clst #:		4		90% clst #:		3		90% clst #:		2	
	Cluster	Frac	AvgDist	AvgCDist	Cluster	Frac	AvgDist	AvgCDist	Cluster	Frac	AvgDist	AvgCDist	Cluster	Frac	AvgDist	AvgCDist
Pospivioids	0	0.689	1.06	1.76	0	0.575	1.08	1.62	0	0.831	1.03	1.81	0	0.860	1.28	1.99
	1	0.257	1.05	1.86	1	0.219	1.08	1.60	1	0.068	1.12	1.89	1	0.049	0.97	2.37
	2	0.023	1.15	1.98	2	0.099	0.90	1.84	2	0.042	0.90	1.87	2	0.032	1.34	2.32
					3	0.040	1.05	1.93	3	0.016	1.06	2.24	3	0.022	0.90	2.29
					4	0.029	1.07	1.73	4	0.014	1.12	2.22	4	0.012	0.77	2.35
					5	0.011	1.04	1.73	5	0.012	1.08	2.08				
									6	0.012	1.02	2.11				

	CCCVd				CBCVd				HLVd			
	90% clst #:		4		90% clst #:		2		90% clst #:		2	
	Cluster	Frac	AvgDist	AvgCDist	Cluster	Frac	AvgDist	AvgCDist	Cluster	Frac	AvgDist	AvgCDist
Cocadviroids	0	0.790	0.99	1.44	0	0.841	1.01	1.50	0	0.881	1.01	1.74
	1	0.062	1.07	1.60	1	0.063	0.93	1.62	1	0.031	1.12	1.80
	2	0.046	1.11	1.67	2	0.037	0.96	1.59	2	0.023	0.99	1.70
	3	0.036	0.98	1.66	3	0.035	0.99	1.75	3	0.023	0.93	1.82
	4	0.023	0.91	1.67	4	0.010	1.00	1.98	4	0.014	0.78	1.81
	5	0.019	0.80	1.79					5	0.013	0.82	2.19
	6	0.015	0.94	1.75								

	ASSVd				CVdV			
	90% clst #:		2		90% clst #:		1	
	Cluster	Frac	AvgDist	AvgCDist	Cluster	Frac	AvgCDist	AvgCDist
Apscaviroids	0	0.521	0.84	1.28	0	0.941	0.72	1.22
	1	0.413	0.71	1.40	1	0.017	0.71	1.49
	2	0.039	0.86	1.28	2	0.015	0.90	1.36
	3	0.010	0.86	1.26	3	0.010	0.78	1.40

HSVd-like internal loops

	HSVd					CbVd1			
	90% clst #:		3			90% clst #:		4	
	Cluster	Frac	AvgDist	AvgCDist		Cluster	Frac	AvgDist	AvgCDist
Hostuviroids	0	0.502	1.16	2.17	Coleviroids	0	0.332	0.94	2.05
	1	0.262	1.24	2.01		1	0.324	1.01	1.79
	2	0.157	1.02	2.34		2	0.207	0.92	1.80
	3	0.051	1.23	2.14		3	0.070	0.93	1.94
	4	0.012	0.99	2.42		4	0.054	0.98	1.79
	5	0.011	0.92	2.23					

The sizes, average RMSD distance among frames in a cluster (AvgDist), and RMSD distance between centroids of each cluster (AvgCDist). Clusters are indexed from zero. Clusters with less than 1% of the frames are not shown. Percentages are based on all frames in the trajectory.

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

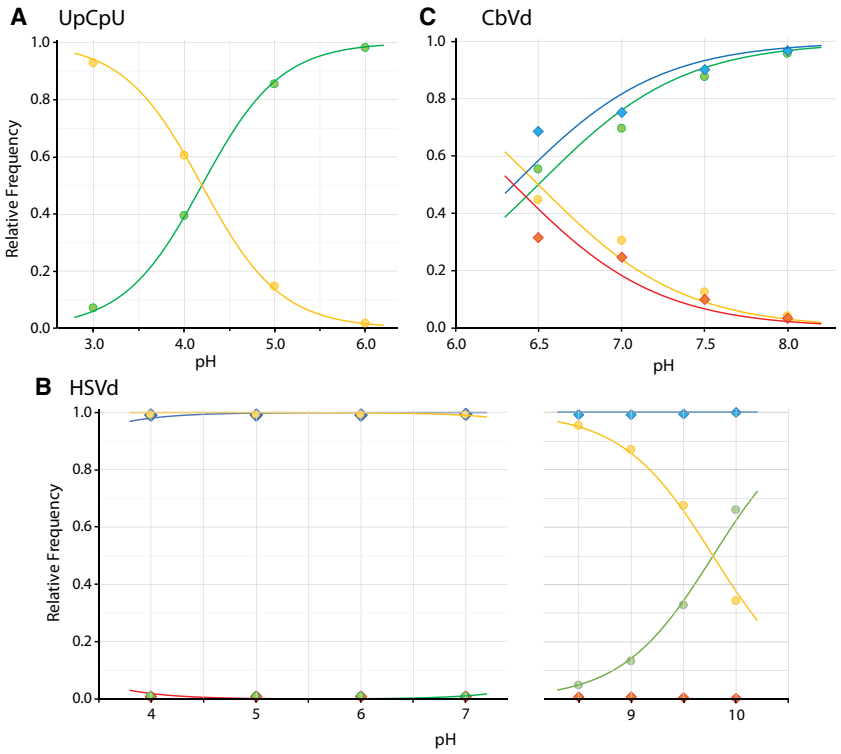

FIGURE 4. pH-REMD indicates that the mSRD structure perturbs the pK_a's of C[0S] and C[0′H]. (*A*) A control calculation on the trinucleotide UpCpU. The relative frequencies of acceptance of the protonated cytosine (yellow circles) and unprotonated form (green circle) are shown as a function of pH. The yellow and green curves give the pH dependence of the relative concentrations of the acid and base forms expected when the pK_a is 4.2. (*B*) Low and high pH runs for the HSVd mSRD. Both C[0S] (circles: yellow, protonated; green, unprotonated) and C[0′H] (diamonds: red, protonated; blue, unprotonated) were made dissociable. The yellow and green lines give the pH dependence of the acid and base forms for pK_a 9.8, while the red and blue lines give the pH dependence for pK_a 2.5. (*C*) The const pH-REMD exchange frequency versus pH for the mSRD of CbVd. The shape and color scheme are the same as part *B*. The red/blue and yellow/green pairs of lines are for pK_a's 6.3 and 6.4, respectively.

Figure 4 presents the frequency of acceptance of the protonated and unprotonated states at set pH values during a constant pH-REMD MD run. A control calculation on the trinucleotide UpCpU gave a pK_a of 4.2 (Fig. 4A), confirming the calibration of the system. When applied to the HSVd mSRD, pK_a's of 9.8 and < 3 were obtained for the [0S] and [0′H] cytosines respectively (Fig. 4B). These values support the MD run with a protonated C + [0S] and an unprotonated C[0′H]. A similar analysis of the mSRD from CbVd gave a rather different result: The pK_a's of the [0S] and [0′H] cytosines were 6.5 and 6.3, respectively (Fig. 4C). Indicating that in CbVd near neutrality, on average one cytosine on either the S or the H-strand is protonated, but not both. Thus, while the symmetry is different, our MD predictions imply that in the mSRD one of the two cytosines is protonated. A possible reason for the symmetry difference is presented below.

We probed the possibility of other protonated bases as well, but did not find evidence of shifted pK_a's. For example, pH-REMD did not switch to a protonated C in the pH range of six to seven for either the bifurcated C cW C pair of PSTVd or ASSVd's level –2 C cW G.

Communication propensity indicates stable SRDs and mSRDs have been found

MD trajectories were further assessed using communication propensity (CProp) (Chennubhotla and Bahar 2007). CProp is the variance in distance between pairs of atoms and it has been shown that these equilibrium fluctuations directly reflect the mean number of steps needed to send and receive information between residues. CProp reflects the stability of a nucleic acid structure, as stable domains should be tightly packed and, therefore, residues will move in unison, giving a low variance in the distance between any pairs of atoms in the domain. Thus, the lower the CProp value, the better the communication, and by inference stability. We use the glycosidic nitrogen as the representative atom of each nucleotide. Our modeling indicates A-form RNA stacked bases have a CProp of <0.2 $Å^2$, and canonically base paired residues have values <0.1 $Å^2$. CProps with other nucleotides greater than 1 $Å^2$ indicate considerable local motion and lack of a distinct local structure.

CProp values are presented as a two-dimensional heat map with the sequence along the diagonal and the domain position number repeated along each edge (Fig. 5). The S-strand is followed by the H-strand; a black line separates these. Each grid intersection above the diagonal gives the CProps as the variance in glycosidic nitrogen distances for that nucleotide couple. Such heat maps are dominated by the contact map. For helices, there are low CProp values immediately above the sequence diagonal, indicating stacking, and another line of low values perpendicular to the diagonal indicating base-pairing. The details of additional interactions are in the off-diagonal areas. The area of a heat map below the diagonal presents the difference map created by subtracting PSTVd CProps from the CProps of the model presented above the diagonal. PSTVd will be used for comparison analysis, as this motif has the most experimental evidence for its proposed structure (Zhong et al. 2008; Steger and Riesner 2018).

Figure 5 shows CProps for the PSTVd SRD and the HSVd mSRD. In the PSTVd map, the upper left to lower right

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

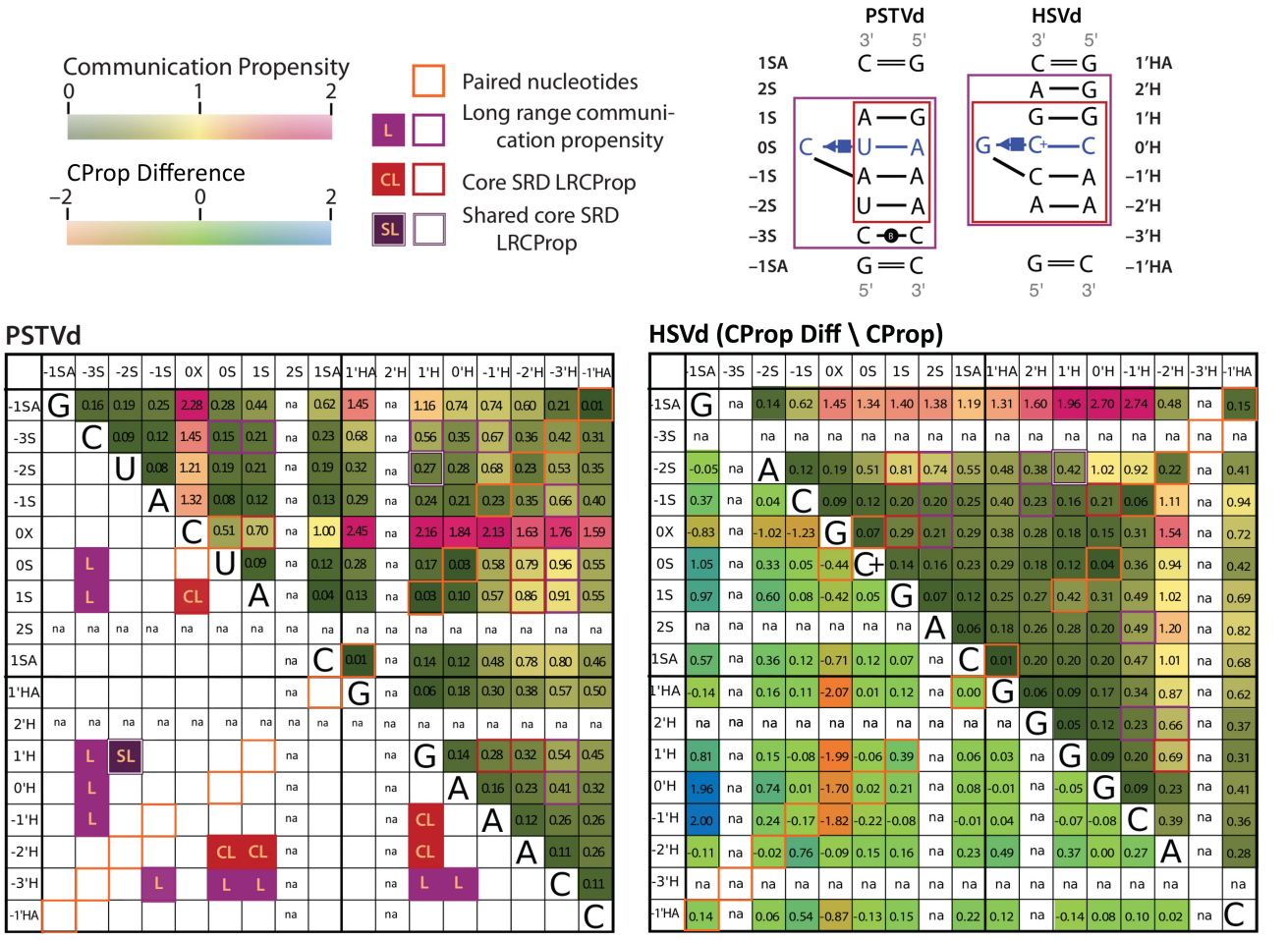

FIGURE 5. Communication propensities of PSTVd's SRD and HSVd's mSRD. The *upper right* triangle of each heat map gives the CProps of frames in the largest cluster from a 500 nsec MD trajectory for each domain. Squares are colored following the *upper* color key. Each grid has entries for all positions found in the two viroids. Since each has a nucleotide not present in the other, the squares of the row or column contain "na" for not applicable. Squares for paired bases are designated with an orange border. The *upper right* maps also present long-range communication propensities using colored borders as indicated by the colored rectangles in the key. These are given for LRCProps found in the domain core (small squares in sequence keys) and for the rest of the domain (large squares). The core LRCProp that is shared by all domains investigated has a navy-blue border. For clarity, the LRCProp data is also presented in the *lower left* triangle of the PSTVd map. The *lower left* triangle of the HSVd map presents a difference map, PSTVd subtracted from HSVd. If either PSTVd or HSVd has no nucleotide at a position, the square is marked "na."

diagonal, showing the stacked bases, is interrupted by the extrahelical C[0X] because this base does not stack or pair with any of the bases. The diagonal from the center to the upper right (boxed with orange) shows base pairs and ncbps, which have CProp values comparable to the stacked bases, with the exception of the cytosines paired at the –3 level. The base pair with the best communication overall is A *t*HS G [1S~1′H]. The map for HSVd shows that in general there is much more communication in this domain. There is no analog of the extrahelical C[0X]; the G at [0X] stacks and pairs as well as the other nucleosides of the motif. The area below the diagonal presents the difference map with PSTVd CProps.

An extension of CProp is long-range communication propensity (LRCProp) (Morra et al. 2009), which counts the number of atom pairs that have a CProp below a communication threshold and are separated by more than a distance threshold. Higher LRCProp values indicate more long-range communication. For glycosidic nitrogens in RNA, we use thresholds of 1.0 Å^2 for communication and 11.2 Å for distance (the separation of these nitrogens in a wide base pair). Using these thresholds, in A-form RNA almost all nucleotides have LRCProps with two or three other nucleotides. LRCProps are less common within the SRD; their instances are indicated as cells in the heat maps with red, purple or blue borders. In most cases, a path of low CProp values can be traced between residues showing LRCProp. However, it is possible for a path to have a step with a higher CProp value because the maps with only one atom for each residue are incomplete. The sugar,

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

phosphate, and base of a nucleotide may have different CProp values and the value for the glycosidic nitrogen is only one of multiple communication paths. Figure S-CProp and Table S-CProp in the Supplemental Material present maps and values for all 11 models.

Every motif modeled has a shared LRCProp between 1′H and -2S, which extends all the way across the motif. This may explain the stability of these motifs. Some LRCProps are unique to groups of models. For example, *Apscaviroids* show long-range communication between 0X cytosine and nucleotides at −2S and 0′H.

Plasticity of the SRD and mSRD RNA

All of the models except CVdV show considerable plasticity as reflected by higher CProp values for touching nucleotides compared to A-form RNA, multiple clusters observed for most models, and changes in base geometry as reported by RNAView (Yang et al. 2003) or DSSR (Hanson and Lu 2017). The U~A handle, while conserved in all *Pospiviroids*, is quite dynamic. Figure 6A shows this pair in CEVd. The center image is the centroid of the largest cluster with a standard U tWH A. In the right frame, the U has pulled away from the A and rotated its W edge away such that its S edge now faces the H edge of the A. In the left image, the tilt has decreased while the buckle has increased, but the S–H interaction is retained. HSVd displays an example of stacking taking priority over pairing. In Figure 6B, C[–1S] and A[–1′H] are paired and stacked on the bases below them. They can come apart and preserve this stacking (Fig. 6B, left), or they can separate and the A moves to stack with the base above it. The transition from one stack to another is rapid; this can be seen in the Supplemental Movie provided in the Supplemental Information.

Figure 6 also presents time courses of distances or angles indicative of the base rearrangements discussed. These show regular exchange between these conformations, indicative of different states in equilibrium.

These models also indicate that the bottom portion of the S-turn is tolerant of many different geometries for

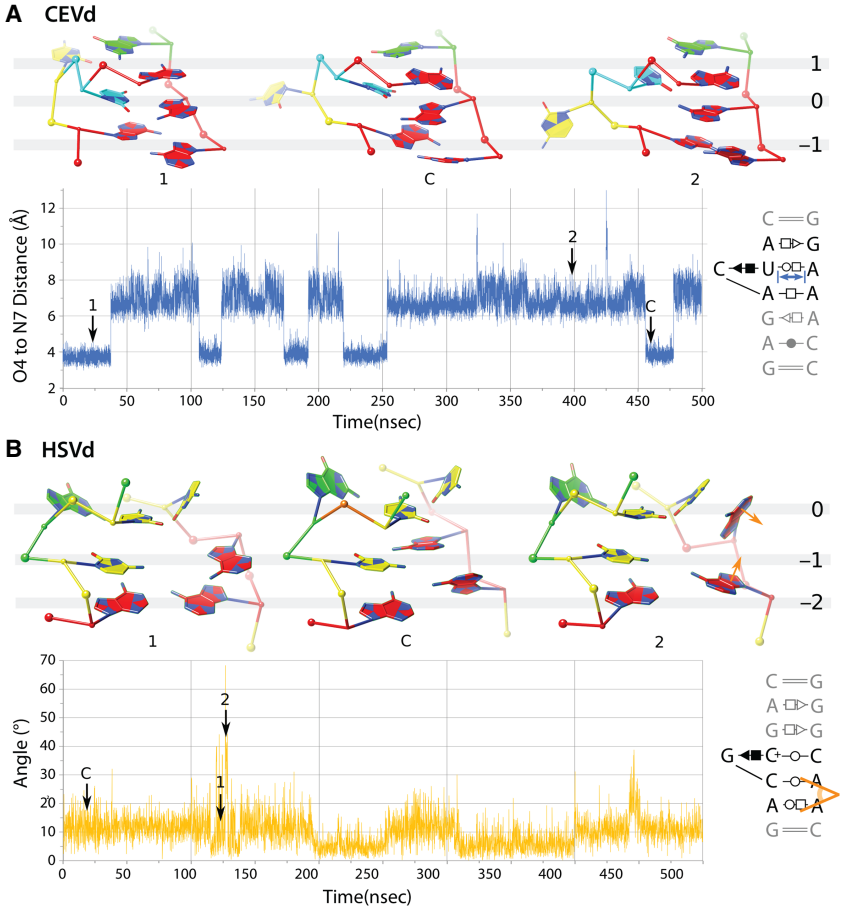

FIGURE 6. Examples of the plasticity of the SRD and mSRD domains of viroids. Conformation information from two MD trajectories is presented. Images show bases connected to a P-C4′ backbone with pseudobonds from each C4′ to its glycosidic nitrogen. Carbons and phosphates are NDB base color coded (A red; C yellow; G green; U cyan). Each view has the S-strand on the *left* and the H-strand on the *right*. Numbers to the *right* indicate the domain position level. Depth cueing fades the back of each structure. Each domain sequence is given to the *right* of the plots. The black bases are those depicted in the images. The central image shows the centroid from the largest cluster. Examples of different conformations are to the *left* and *right*. The plots show the variation of distances and angles over the course of the 500 nsec MD trajectory. The black arrows indicate the frames shown in the images. (*A*) The U~A handle of CEVd has different geometries. A[0′H] and A[1S] have extensive cross-stacking which remains constant throughout the trajectory. In contrast, the U moves. In the principle conformation, its W edge makes one or two hydrogen bonds to the H edge of A[0′H]. However, the U can rotate its S edge toward the A, requiring bridging waters (not shown) rather than direct hydrogen bonds to interact with the A. This separation and rotation can be monitored with the U[0S]O4 to A [0′H]N7 distance (purple dashed line). This distance is plotted *below* the images. The stacking also changes. In the centroid, A[0′H] stacks on both A[–1′H] and A[–1S]. In the *left-side* conformation, A[0′H] stacks predominantly on A[–1′H], while in the *right-side* conformation it stacks on A[–1S]. (*B*) The wobbled A-C pair at position –1 of HSVd has different geometries. C[–1S]~A [–1′H] of the centroid has a typical *cis* W interaction that makes two hydrogen bonds. The A does not stack well the bases above or below it. This pair often separates. The adenine will then stack with A[–2′H] (*left*) or with C[–1S] (*right*). This motion can be monitored via the angle A[–1′H] makes with A[–2′H] as determined from vectors normal to each base's six-membered ring (orange arrows). This angle is presented in the plot *below* the images.

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

base interaction. The bottom of the S-curve supports canonical and noncanonical cW pairs, as well as tSH pairs. This could be because stacking seems more important than pairing, as evidenced by partners separating and losing their direct hydrogen bonding, but by maintaining stacking with the bases above and below in tandem or sequentially, as described for HSVd above. This applies to both the SRD and mSRD structures and is supported by the consistent LRCProp between 1′H and –2S.

Descriptions of the models for each genus of the *Pospiviroidae* family

Pospiviroids

Five models cover the ten members of the *Pospiviroid* genus: PSTVd, CEVd, Mexican papita viroid (MPVd), and tomato chlorotic dwarf viroid (TCDVd) all have a PSTVd-like internal loop. CLVd has a HSVd-like internal loop and will be discussed with HSVd. Hop latent viroid (HLVd) is in the *Cocadviroid* genus, but because of the similarity of its internal loop to the *Pospiviroids*, it will be included here. As described above, the upper block is very similar to the standard SRD built upon a C∼U∼A triplet whose C rarely associates with any of the other residues in the domain. Each lower block has A tHH A at the –1 position and (U or A) tSH A at the –2 position. All four lower blocks are extended with an extra cWW ncbp at the –3 position, which is either wobbled or uses bifurcated hydrogen bonding. The distinction of the upper block from the lower one is shown in part by the generally poor CProp, for example, U of the U∼A handle of both PSTVd and TCDVd has a CProp greater than one with all nucleotides in the H-strand lower block.

In all four *Pospiviroid* models, the S- and H-strands have extensive stacking (Fig. 7, PSTVd). The most prominent stack extends from the –3′H or –2′H up into the A-form RNA starting at the 1SA position. The stacked bases switch from the H-strand to the S-strand where A[0′H] cross-stacks with the A[1S]. This is the hallmark cross-stack of an SRD. The other prominent stack is made by bases –1S through –1SA and can continue in the A-form RNA below the domain. A[1S] is part of both stacks; thus, it serves as a junction, creating the curve-shaped structure seen in Figure 6 for PSTVd. ASSVd has a similar λ stack (Fig. 7, ASSVd).

Among all of the viroids in this group, each of these stacks contains a base from the –3 position. Although designated cWW, these two bases often do not interact with each other. The stacking is strong enough to separate the pairing edges such that they hydrogen bond with water rather than each other. pH-REMD analysis indicated that in PSTVd, neither of these cytosines has a perturbed pK_a; thus, no protonation is expected for this C∼C pair.

The bifurcated C[–3S] cW C[–3′H] pair of PSTVd has numerous geometries. There is only direct hydrogen bonding between these bases in 70.6% of the trajectory frames. In the frames with no direct hydrogen bonding, water generally bridges interaction between the cytosines, which can be separated up to 10 Å. This separation is not affected by the hydrogen bonding status of the U tSH A immediately above these C's. When the C's separate, each base remains associated with its stacked base column. One hydrogen bond is observed in 33.6% of the frames, most often N3 of C[–3S] bound to N4 of C[–3′H]. The remaining are bifurcated hydrogen bonds in which both hydrogens of N4 of C[–3′H] interact with its partner's N3 or only one hydrogen on N4 of C[–3′H] interacts with both O2 and N3 of C[–3S].

The internal loops of the other *Pospiviroids*, CEVd and HLVd, each have a nucleotide change at both the –2 and –3 levels. These changes make these domains less dynamic, as indicated by lower CProps in both the upper and lower blocks for these two viroids and by the tenacity of hydrogen bonds between partners.

Cocadviroids

This genus is unusual in that each of its four members has a different internal loop sequence. These loops vary in size due to deletions or insertions in the lower block. Coconut tinangaja viroid (CTiVd) has the same internal loop as CEVd, which is the same size as PSTVd's loop and was described with the *Pospiviroids*. CBCVd lacks

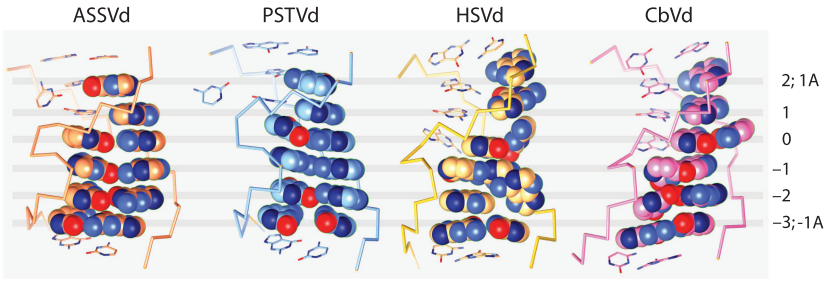

FIGURE 7. Base stacking in the SRD-like domains. The centroids from the largest cluster for ASSVd, PSTVd, HSVd, and CbVd are shown. Bases analogous to ten bases in the λ stack of PSTVd are shown as 90% van der Waals radii space-filling atoms [1′HA or 2′H, 1′H, 0S, 0′H, –1S,–1′H, –2 s,–2′H, –1′HA or –3S]. Partner bases not included in the λ stack and an additional A-form base pair *above* and *below* have been included to provide context. These are shown in stick mode. The models are aligned by the levels described in Figure 1 and labeled on the *right*. The backbone is indicated using P-C4′ pseudobonds. The base carbon colors match the backbone, which is specific to each viroid. Glycosidic nitrogens are navy blue; all other atoms have CPK coloring in all models.

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

one ncbp, while coconut cadang-cadang viroid (CCCVd) has an extra ncbp in the lower block. With the exception of CBCVd, all *Cocadviroids* have one or two bifurcated base pairs at the lower block transition to A-form RNA. There is no indication that the canonical base pairs of CBCVd have a nonstandard geometry. These changes in the lower block of the *Cocadviroids* influence the dynamics of the motif as indicated by their CProp values, but they do not alter the shape of the S-turn. The extent of that turn is the same for all of these sequences.

Apscaviroids

This genus has two internal loop sequences: the ASSVd loop, shared by nine viroids, and citrus viroid V (CVdV), the only member of its group. These loops are short with only four or five ncbps in three or four levels. As with the *Cocadviroids*, the extent of the S-turn is not reduced. The *Apscaviroid* SRD motifs have greater stabilities than the other models in the PSTVd group as indicated by only two clusters representing >94% of the conformations and the low CProp values. Figure 7 shows that this stability is most likely a consequence of extensive stacking. A similar long stack described for the *Pospiviroids* is seen, except that A[1S] does not contribute to both stacks, it only resides in the long stack. It is G[1SA] that participates in the two parallel stacks, raising the junction point one level. The G[-2′H] is an A in the *Pospiviroids* and *Cocadviroids*; this may be responsible for this change and the added stability.

Hostuviroids and coleviroids

Five viroids have internal loops with the HSVd-like sequence: CLVd has the same loop sequence as HSVd, while the three *Coleviroids* share a sequence that differs from HSVd at four residues. We have proposed the mSRD for these, which despite extensive sequence differences has a backbone trace remarkably similar to the SRD of the PSTVd-like group (Fig. 3C). The upper block of this mSRD has an additional *t*HS ncbp at the 2 level. We have observed other instances where the terminal *t*HS ncbp of a domain can repeat. [See RNA 3D Motif Atlas (Petrov et al. 2013) entries IL_64611.1 and IL_66584.1 and IL_06300.1 and IL_05062.1]. In HSVd these ncbps differ from those in the PSTVd-like group in that they can separate, losing one or more hydrogen bonds; however, as in the PSTVd-like case, they maintain strong stacking that continues into the A-form RNA. In CbVd the level 1 and 2 pairs both are A *t*HS G and follow the PSTVd group with good hydrogen bonding and stacking. In both cases, these stacking interactions produce a series of low CProp values, supporting a tight stack with correlated movements of its component bases.

Both genera use a G~C~C triplet for the base of the top block. For HSVd, the C[0S] is protonated, while for CbVd either C[0S] or C[0′H] is protonated. In both, the C at 0′H is *syn*; the χ values are around 40° and 60° for HSVd and CbVd. These values are on the boundary of intermediate *syn* and full *syn*, as defined by Sokoloski et al. (2011). This allows any of C[0′H] N3, O2, as well as O4′ to hydrogen bond or interact with the protonated N and N4 of C + [0S] (Figs. 3B, 8). The combination of protonation and a *syn* glycosidic bond allows for an extensive network of hydrogen bonds and dipole interactions described below (Fig. 8).

Clustering of the HSVd trajectory showed that C + [0S] can stack on either A[-1′H] (first and third cluster; Fig. 6B, centroid) or C[-1S] (second and fourth cluster; Fig. 6B, conformations 1 and 2). This resulted in the first and last half of the main production run having significantly different CProp values for C + . There were several transitions between these conformations, instances with C + [0S] on A [–1′H] clearly dominated, indicating this arrangement is slightly more stable.

The base triplet conformations are different in the HSVd and CbVd models. HSVd has a high positive buckle at both base-edge interactions, while the cytosines of the CbVd triplet are nearly coplanar, and the G[0X] is parallel but staggered down from the cytosine pair so that it can interact with the 2′ OH of G[1′H]. In HSVd, there is a dipole-dipole interaction or possible weak hydrogen bond between N2 of G[0X] and an H on N4 of C + [0S]. This interaction cannot form when the bases are coplanar as in CbVd (Fig. 8, bottom right). The buckled triplet of HSVd creates a basket in which G[1′H] resides, separating it from its partner, G[1S] (Fig. 8). Unlike the extrahelical C of the PSTVd-like group triplet, G[0X] of both HSVd and CbVd is an integral part of the stacked bases of the top block, making a cross-strand stack with G[1′H], but not stacking with any bases below it (Figs. 7, 8). This cross-stack involves the bases complementary to the cross-stacked bases in a standard SRD ([1S] stacking with [0′H], PSTVd in Fig. 3C). In the mSRD, it is C[0′H] of the triplet that has only marginal stacking with other bases of the domain.

We explored the properties of the HSVd mSRD triplet that stabilized the motif and favored protonation of C[0S] over C[0′H]. Protonation of cytosine's N3 makes three areas more positive or less negative: the N3-H3 group itself, the glycosidic nitrogen (N1) and the adjacent carbonyl (C2-O2), and the N4 amino group along with the adjacent C5-H5 (Fig. 9A). Starting with a representative frame of C+ HSVd, we altered the charges on C[0S] as follows: (i) a control with the full charge (C+; Fig. 9B); (ii) fully unprotonated (Fig. 9C); (iii) C+ with the glycosidic nitrogen and the ketone reverted to the unprotonated charges (Fig. 9D); (iv) C+ with N3 and H3 reverted to the unprotonated charges (Fig. 9E); and (v) C+ with the C4 and its amine reverted to their unprotonated charges (Fig. 9F). Each case was observed in 200 nsec of MD simulation. The fractional charge on the cytosine does not affect AMBER's production of

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

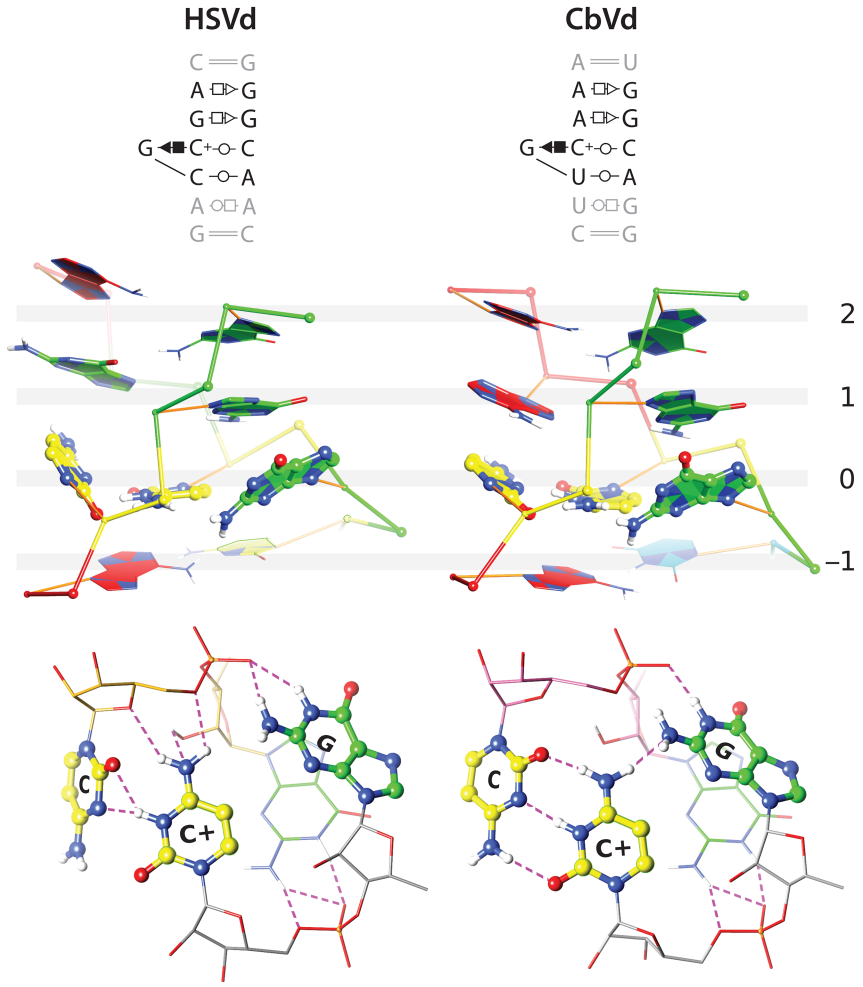

FIGURE 8. Interactions of the protonated cytosine in the triplets of HSVd and CbVd mSRD. (*Top*) Sequence schematics for the mSRDs. (*Middle*) Side view of the bases of the upper blocks of the mSRD domains. This view is from the back, the H-strand is in front, 5′–3′ is from the *upper right* to the *lower left*. The bases in the triplet are in ball and stick representation. Carbons have NDB colors; pseudobonds from C4′ to the glycosidic nitrogen are coral. (*Bottom*) View of the triplets showing interaction within the triplet and with G[1′H]. Bases of the triplet are ball and stick, the backbone and G[1′H] are wire frame. Potential hydrogen bonds are purple dashed lines. Not all of these will be present simultaneously. Note that because of the extreme buckling of C[0′H] in HSVd, the only hydrogen bonding it can have with C[0S] is a bifurcated interaction of the protonated H3 with N3 and O2 of C[0′H].

trajectories. CProps were calculated from these runs (Fig. 9C–F, upper right triangles) and compared to the control by subtracting those values to give the difference CProps in the lower left triangle.

The consequence of complete reversion to neutral cytosine is that C[0′H] loses communication with all other residues, as C[0′H] and C[0S] no longer form a stable ncbp. The trajectory shows that they lose interaction shortly after the trajectory starts, which allows C[0′H] to rotate into the *anti* conformation and fall out of the helix. Other nucleotides in the H-strand distort and become displaced as evidenced by the column of poor communication (blue difference CProps in the [1S] column and the [-1′H] and [-2′H] rows).

Reversion of the glycosidic N1 and ketone charges of C[0S] resulted in the greatest disruption. The CProps (Fig. 9E) show that the bases with which C[0S] stacks, [1S] and A[–1′H], lose communication with it. Also, unique to this scenario, G[0X] loses communication with the rest of the helix. The result was that both C[0S] and G[0X] fell out of the motif, destabilizing it. Note that the keto group does not participate in direct hydrogen bonds with other parts of the motif. There are more subtle interactions not revealed with a hydrogen bond or water that may mediate a key interaction. Alternatively, the changes in charge distribution disrupt the partial charge interaction that contributes to strong stacking (Calladine et al. 2004). Reversion of N3 to its unprotonated charge and giving H3 no charge, maintains the same charge on the Watson edge, but on different atoms. This change caused C[0′H] to fall out of the triplet allowing it to interact with the residues in the upper block. C[0′H] then flipped to the *anti* conformation allowing it to interact with G[1S], causing the motif to destabilize. The reversion of the charges around the amino group (N4/C5) was the smallest change (a loss of 0.173 charge). This resulted in a lack of interactions of C[0S] and G[1S] with most of the lower block. Thus, all three locations that pick up the positive charge from protonation make critical interactions that keep the motif stable. C[0′H] does not have such a network of interactions, and thus its protonation will not have the same effect. In CbVd, the C[0S]~C[0′H] is nearly planar and hence more symmetric. A combination of these effects may contribute to the preference of protonating C[0S] in HSVd, while either cytosine may be protonated in CbVd.

Cross-stacking in the SRD in PSTVd and the 5S rRNA have been assessed by UV cross-linking. Branch et al. (1985) showed that in PSTVd U[0S] cross-links to G[1H]. Interestingly, CDVd (formerly CVdIII) is resistant to cross-linking, and substitution of PSTVd' A[–1S] to G, in accordance with the CDVd sequence, abolishes its cross-linking (Owens and Baumstark 2007). Thus, nucleotides that are not cross-linked influence cross-linking. Our MD simulations for PSTVd, CDVd, and PSTVd with A[–1S] changed

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

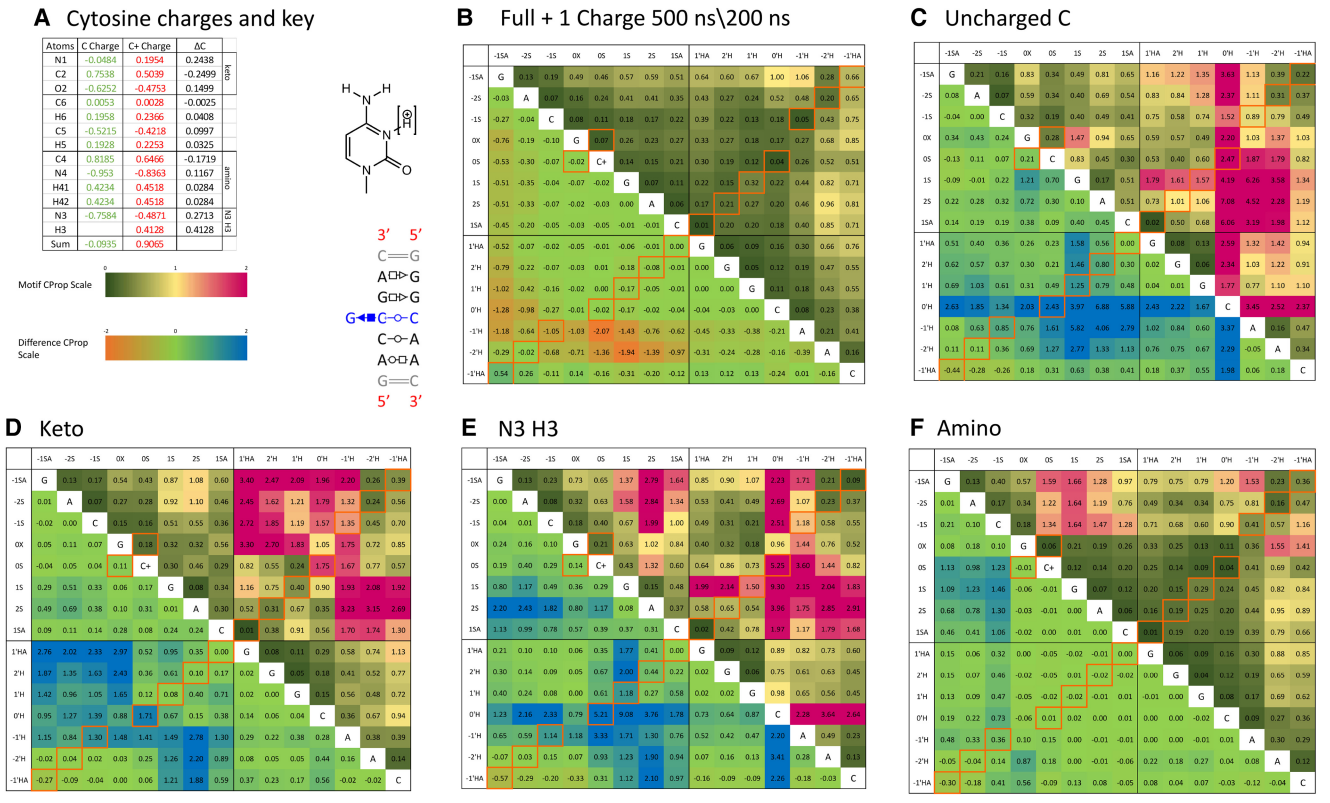

FIGURE 9. Dissection of the role of charged regions of cytosine in the HSVd mSRD. (*A*) Charge distribution on atoms of cytosine and protonated cytosine used in the AMBER force field for MD. ΔC presents the charge difference of protonated minus neutral cytosine. The color keys for the heat maps for CProps and CProp differences are presented as well as the secondary structure schematic for the HSVd mSRD. (*B*) Heat maps of the CProps for the HSVd mSRD with protonated cytosine from the 500 nsec run (Fig. 5) prior to clustering (*lower left*) and the 200-nsec control for this experiment (*upper right*). (*C–F, upper right*) Heat maps for CProps for the mSRD for various sets of charges. (*Bottom left*) The difference map in which the CProps 200-nsec run from the standard protonated cytosine (panel B, *upper right*) were subtracted from the values *above* the diagonal. (*C*) CProps for the mSRD with no cytosine protonated. (*D*) For this run, the charges of the glycosidic nitrogen (N1) and C2–O2 ketone were set to those of unprotonated cytosine. All other atoms had the charges for protonated cytosine. This produces a base with a fractional charge that is not compensated elsewhere. (*E*) The charge on N3 was that for unprotonated cytosine; H3 has 0 charge. (*F*) The charges on C4 and its amino group were set to unprotonated cytosine.

to G showed no prognostic such as a change in stacking overlap area of the bases that cross-link (data not shown). Consequently, we cannot predict which of these SRDs and mSRDs should show UV cross-linking.

Conclusion

The results from the in silico work show that every viroid in the *Pospiviroidae* family has a sequence in its CCR that is capable of forming an asymmetric internal loop that is geometrically similar to PSTVd's SRD. All share the distinctive S-curve and helical strands of the SRD. They fall into two general geometric families. One, which includes PSTVd, has variants closely resembling the standard SRD geometry as represented by the 5S rRNA SRD. The variations are limited to particular ncbps; they conserve the shape of the ncbp but may change its geometric class. The other structure, the mSRD, is significantly different in its internal structure, with different geometry at the triplet or 0 level and the ncbp below it (−1 level). The MD trajectories of members of each geometric family show arrays of interactions that keep the motif stable, including cross-stacking of bases in the triplet. These motifs share LRCProps, including one that is common to all 11 that demonstrate the stability of the full motif. However, there are unique interactions found within each motif that may give each species of viroids unique biochemical properties. Similar long-range communication involving a protonated cytosine has been proposed in the hepatitis D ribozyme (Veeraraghavan et al. 2010).

This conclusion rests strongly on the reliability of MD; many investigators have shown that there are still some shortcomings in the force field for RNA (Šponer et al. 2018). They saw unusual conformations accumulate in long runs. However, other cases have shown that MD clearly predicts unexpected structures, such as the unusual structure of the internal loop GAGU studied in the Mathews and Turner labs (Spasic et al. 2018). The long

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

time courses sought to explore a global minimum, which we are not doing. It is widely accepted that access of viroid RNA to the cellular endonuclease or ligase during processing is controlled by pleomorphic structures in the CCR. Thus, our structures may not represent a global minimum in the conditions used. Our models stayed in conformations that could all be related to the standard SRD. We did see conformational plasticity and many of these variant conformations are visited repeatedly during trajectories, supporting their designation as a family of conformations in a local minimum. Furthermore, we saw on numerous occasions that incorrect structures gave unstable trajectories in which one or more bases fell out of the helix, as explained in the presentation of varying the charges on protonated C[0S]. Each sequence was able to maintain a stable conformation for at least 500 nsec.

We present chemical consistency in these models. The mSRD required protonation of one cytosine in a C~C pair. This is extraordinary because cytosine normally has a pK_a of 4.2. Through the use of pH-REMD, we show that this structure perturbs the pK_a's of the cytosines in question. In the case of HSVd, one cytosine is strongly shifted and always protonated. For CbVd, the structure perturbs the pK_a's of both cytosine to near neutrality. Therefore, one of these will likely be protonated in neutral cytoplasm of plants. This pair also requited a *syn* glycosidic bond. This arrangement was stable in our structures. However, when these structures were perturbed, such as by charge changes or protonation of the cytosine, the glycosidic bond quickly reverted to *anti*.

The evidence for a structure with a conserved shape in the CCR has important implications for the asymmetric rolling circle model of replication of the *Pospiviroidae*. Where known, the CCR is the location of cleavage and ligation. It seems reasonable that related viroids would be using related host functions for these essential processes; hence, we must continue to learn more about potentially shared functions and learn how they are used to coerce the host into replicating viroid RNA. Some lncRNAs, like viroids, may not have high structural conservation, but do have large-scale shape conservation that maintains similar functions in related species. Therefore, care must be taken in dismissing the possibility that an RNA lacks a specific functionality.

PSTVd replication requires an alternative form of transcription factor TFIIIA, TFIIIA-7ZF, to complex RNA pol II so that it will transcribe PSTVd RNA. Ribosomal protein L5 (RPL5) regulates splicing of TFIIIA mRNA by inhibiting the pathway that produces of TFIIIA-7ZF. PSTVd binds RPL5 using the SRD of the CCR of PSTVd, as the mutation of the SRD eliminates binding (Jiang et al. 2018). PSTVd binding RPL5 prevents inhibition of TFIIIA-7ZF production (Dissanayaka Mudiyanselage et al. 2018; Jiang et al. 2018). HSVd has been shown to have a similar interaction of RPL5 (Wang and Ding 2010). With the current demonstration that HSVd is likely to have a similarly shaped motif, it is likely that RPL5 binds the mSRD of HSVd. Further studies are needed to show if other viroids have similar interactions. The discovery that viroids directly interact with ribosomal proteins introduces the potential of new biochemical pathways previously unseen. While both PSTVd and HSVd show interactions to exclusively RPL5, there is still a possibility that viroids interact with other ribosomal proteins (Zhou et al. 2015; Xie et al. 2018).

This in silico work provides approximations of these structures, other methods must be utilized to validate these structures. Further work into the structure of viroids and SRD mimic in mRNA is required to show that these structures are valid and functional. These predictions and other research previously explained imply that structural mimics are more common in RNAs than thought and may have a very broad array of functions in biochemical pathways.

MATERIALS AND METHODS

Description of the SRD and the numbering scheme used in this paper

We have adopted the 5S rRNA SRD (PDB 1S72 [Klein et al. 2004]) as standard for the domain (Fig. 1B,C). It has five ncbps, which we describe here using the nomenclature of Leontis and Westhof (Leontis and Westhof 2001). The overall structure has an S-turn (the S-strand, Fig. 1B,C), while its partner follows a more normal helical path (the H-strand). In the standard view, the S-strand goes up in the 5′–3′ direction and is presented on the left in the secondary structure depiction (Fig. 1C). At its core, an SRD has a base triplet containing a "U~A handle" (Jaeger et al. 2009), in which the W edge of U interacts with the H edge of A such that their glycosidic bonds are *trans*, a U *t*WH A interaction. The third base of this triplet is an extrahelical (bulged) G, whose S edge interacts with the H edge of U in a *cis* arrangement, G *c*SH U. This triplet is stacked on a *trans*-Hoogsteen-Hoogsteen ncbp (*t*HH), typically a pair of adenines. The local geometry of the U~A handle gives the backbone an antiparallel orientation, while the G *c*SH U and A *t*HH A pairs produce a parallel local backbone orientation. The base pairs above and below this core all have A-like antiparallel backbones. This combination creates the distinctive S-turn structure (Correll et al. 1999; Duarte et al. 2003).

The outer ncbps of the motif serve to transition the SRD's core to the surrounding A-form RNA. The top transition strictly uses a *t*SH ncbp. The bottom transition can use this pair as well. However, this boundary has more variability, often utilizing additional ncbps, which may end in an *c*WW pair to make the transition. The motif is characterized by extensive stacking, but not always with intra-strand neighbors. One such feature is the cross-strand stacking of the A's from the U~A handle and the upper *t*SH pair. Additional base-backbone interactions further stabilize this motif (Lu et al. 2010).

The nucleotides of an SRD create a geometric pattern that allows for many isosteric substitutions (Leontis et al. 2002). Thus, many different RNA sequences fold into this structure. In order

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

to guide comparisons of domains with different sequences, as well as with insertions and deletions, we present a structure-based numbering system (Fig. 1C). Nucleotides in the S-turn and helical strands are designated with an "S" or "H" suffix, respectively. S-strand nucleotides are numbered 5′–3′, while nucleotides in the H-strand use the same number as their S-strand partner and are given a prime to indicate the 3′–5′ numbering. The extrahelical base has an X suffix. The triplet is the 0 level. Nucleotides above it are positive, nucleotides below it are negative. Thus, the triplet is designated 0X~0S~0′H and the tHH pair below it –1S~–1′H. Noncanonical pairs above and below the triplet get larger positive and negative numbers. Canonical A-form bases get an A suffix. The first canonical pair above the motif is 1SA~1′HA; the first below is –1SA~–1′HA. The base pair step between the 0 and –1 levels splits the domain into upper and lower blocks. There are SRD-like junctions, in which an insertion or a switch to a different chain occurs at the 0X/–1S step of the S-like strand. For numbering these, bases not in the SRD helix have a X suffix. The numbering of bases connecting to 0S starts with 0X and continues positive; bases connecting to –1S get negative numbers.

Determining sequences in the CCR, predicting secondary structure of viroids, and predicting tertiary structure

Minicircles containing the CCR of each viroid were created as follows. Sequences of the wild-type viroids were obtained from NCBI Entrez Molecular Sequence Database (https://www.ncbi.nlm.nih.gov/Web/Search/entrezfs.html; sequences and ID numbers are presented in Supplemental Table S1). These sequences were aligned using the ClustalW output of MUSCLE (Edgar 2004). The most conserved sequences agreed with the CCRs recognized by others (Di Serio et al. 2014) and were divided into upper (UCCR) and lower CCR (LCCR) strands. These were roughly paired by complementarity and additional viroid sequence included to make the ends blunt. GC rich hairpins were added to these ends in order to create circles that fold into rods. Each sequence was submitted to RNAfold (Mathews et al. 2004; Hofacker and Stadler 2006) at ViennaRNA (Lorenz et al. 2011) to produce secondary structure predictions of the CCR for every classified viroid in the *Pospiviroidae* family.

The secondary structure predictions were analyzed for potential asymmetric loops in positions similar to PSTVd's SRD. Only unique motif sequences were used; thus, eleven different candidates for SRD-like motifs were created. Each internal RNA loop sequence was compared to the SRD in the 5S RNA of the *H. marismortui* large ribosomal subunit (PDB ID 1S72; nucleotides 76–80 and 102–105). The RNA structure search engine FR3D (Sarver et al. 2008) was used to find ncbps with the viroid sequence and a similar geometry to the SRD. The initial models were assembled in the molecular visualization and manipulation program Chimera (Pettersen et al. 2004). Structures were manipulated manually in Chimera to create phosphodiester bonds between residues and to minimize bad contacts. At least three canonical base pairs were added to each end of the SRD to hold the two strands of RNA together. The canonical base pairs used were designed to prevent undesired interactions with the SRD and do not need to follow the sequence of the viroid's canonical base pairs. In some cases, a fourth canonical pair was added when fraying ends created unwanted interactions with other nucleotides in the model.

In some cases, MD indicated unstable structures (see below), and variations were made. These included varying geometry, trying protonated bases, and rotating glycosidic bonds to the *syn* conformation.

Molecular dynamics

The models were submitted to the AmberMD 14 package (Case et al. 2014) using leap with the force field ff99bsc0_chiOL3. For models that included a protonated cytosine (C+) parameters from all_prot_nucleic10.lib and frcmod.protonated_nucleic were used.

The models were solvated in a shell of TIP3PBOX explicit water of ~8.0 Å (Jorgensen et al. 1983). Each RNA solute was covered with a shell 8 Å thick, and then potassium ions were added to neutralize the charge (20–24 ions depending on the number of phosphates and protonated residues). K^+ was used as it is the most common cation in the cytoplasm and nucleoplasm; no divalent ions were present. The model was solvated in a cuboid box to produce models with 11,000 H_20 residues, which ensured a minimum of 8 Å from the box surface to the RNA. The model was then minimized for 2000 steps and then equilibrated under constant pressure with 1 fsec time steps, using 30 kcal $mol^{-1}A^{-2}$ harmonic potential on the ions and RNA in addition to NMR restraints on heavy atoms that participate in the modeled base pairs. SHAKE (Ryckaert et al. 1977) was used to constrain all covalent bonds to hydrogens. The restraints were lowered over a course of 200 picosec to 10, 3, 1, 0.3, 0.1, 0.03, and 0.01 kcal $mol^{-1}A^2$ and then restraints were completely removed. After equilibrating, production simulations were run over 500 nsec with steps of 2 fsec, at constant volume. Temperature was kept at 300 K with a Langevin thermostat. During production runs, end fraying was reduced using an NMR restraint of 0.3 kcal $mol^{-1}A^2$ on the hydrogen-bonded heavy atoms of each terminal base pair. This weak restraint is analogous to an extension of the A-form stems as would occur in a larger biological RNA. For CBCVd fraying was problematic, so the restraints were raised to 1.0 kcal mol^{-1} A^2. No other restraints were active during production runs.

Communication propensity and clustering

Stability of models was assessed using communication propensity (Morra et al. 2009), a measure of the variance in the distance between two different atoms, which is a direct measure of signal propagation (communication) between the two residues. The lower the CProp, the higher the communication between the two residues and the more stable the interactions. Glycosidic nitrogens were used as indicators for each nucleotide. Distances between all pairs of glycosidic nitrogens were measured and averaged using the AMBER CPPTRAJ module (Roe and Cheatham 2013). A PERL script calculated communication propensities between atom pairs *i* and *j* as:

$$CProp_{ij} = \frac{1}{N_{\text{frames}}} \sum_{k=1}^{N_{\text{frames}}} (d_{ij,k} - \langle d_{ij} \rangle)^2,$$

where N_{frames} is the number of trajectory frames used, $d_{ij,k}$ is the

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

distance between glycosidic nitrogens (N1 of pyrimidines and N9 of purines) of residues *i* and *j* in frame *k*, and $<d_{ij}>$ is the average distance separating these nitrogens in residues *i* and *j*. When analyzing aggregates of CProp values, we report medians, as there is no upper bound and exact value of outliers has no significance. For the PSTVd-like group, the core is defined as levels –2 to 1, without the highly mobile extrahelical base [0X]. This leaves out the extended bottom of the domains in *Pospiviroids* and *Cocadviroids* and includes the bottom boundary canonical pair for *Apscaviroids*. For the HSVd-like group the same levels are used (excludes the 2 level), but the 0X nucleotide is included. Long-range CProps (LRCProps) (Morra et al. 2009) counts the number of pairs which have an average distance longer than any base pair (>11.2 Å), but has a CProp that signifies a stable base pair (<1.0 $Å^2$). Varying these thresholds does not change the qualitative information relayed by LRCProp; the values we use give a reasonable number of hits and relay the qualitative properties of the domains well. High LRCProp counts indicate stable motifs where motion is correlated among several residues that are not base-paired or stacked.

Average CProp values and number of LRCProps per nucleotide for A-form RNA were determined from a 250 nsec MD trajectory of a 42 nucleotide RNA with the top strand sequence CCGCGCUAUGAACCAGUAAAGGCGGGAGACACGAUUGCG GCG. This RNA was constructed in Chimera and entered in leap with 80 K^+ ions and 51,823 waters and equilibrated as described for the SRD RNAs, but only using end restraints.

Hierarchical centroid-linked clustering performed by cpptraj was used to analyze common conformations of each trajectory (Shao et al. 2007). The coordinates from the 500 nsec trajectory were stripped to only the ring skeletons, sugar, and phosphate without hydrogens. Ten clusters were produced with a sieve of twenty. 3DNA (Lu and Olson 2008) and DSSR (Hanson and Lu 2017) were utilized to determine the motif's nucleotide pairing and interaction for each cluster.

pH-dependent replica exchange MD

pH-REMD (Itoh et al. 2011; Swails et al. 2014) was used to determine the pK_a's of the protonated cytosines in ASSVd, CbVd1, HSVd, and PSTVd. pH-REMD uses explicit solvent MD and regularly allows exchanges between models with different protonation states. The Metropolis exchange criterion used energies from a Generalized Born implicit solvent calculation. If an exchange is made, the solute is immobilized briefly while the explicit water rearranges in response to the new charge arrangement. The pK_a was calculated from the populations of protonated and unprotonated cytosine using the Henderson-Hasselbach equation. pH-REMD was performed for 1 nsec per replica using the parameters of Swails et al. (2014). For the HSVd-like group, triplet C + [0S] and C + [0′H] were set as titratable. We also looked for other protonation by setting the following titratable: C[–2S] of ASSVd, as well as C[–3S] and C[–3′H] of PSTVd.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We would like to thank Dr. Zhiwei Liu, Associate Director West Center for Computational Chemistry and Drug Design for training and assistance for our computations. We thank Dr. Michael Yonkunas for critical reading of the manuscript. Funding was provided by the University of the Sciences Department of Chemistry and Biochemistry and the Robert D. Spiers Graduate Research Fellowship.

Received January 17, 2019; accepted May 14, 2019.

REFERENCES

Baumstark T, Schroder AR, Riesner D. 1997. Viroid processing: switch from cleavage to ligation is driven by a change from a tetraloop to a loop E conformation. *EMBO J* **16:** 599–610. doi:10.1093/emboj/16.3.599

Branch AD, Benenfeld BJ, Robertson HD. 1985. Ultraviolet light-induced crosslinking reveals a unique region of local tertiary structure in potato spindle tuber viroid and HeLa 5S RNA. *Proc Natl Acad Sci* **82:** 6590–6594. doi:10.1073/pnas.82.19.6590

Calladine C, Drew H, Luisi B, Travers A. 2004. *Understanding DNA*, 3rd ed. Academic Press, MA.

Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. *J Comput Chem* **26:** 1668–1688. doi:10.1002/jcc.20290

Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Gohlke H, et al. 2014. *AMBER 14*. University of California, San Francisco. http://ambermd.org/Manuals.php

Cech TR, Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157:** 77–94. doi:10.1016/j.cell.2014.03.008

Chen H, Giese TJ, Golden BL, York DM. 2017. Divalent metal ion activation of a guanine general base in the hammerhead ribozyme: insights from molecular simulations. *Biochemistry* **56:** 2985–2994. doi:10.1021/acs.biochem.6b01192

Chen X, Sun Y, Cai R, Wang G, Shu X, Pang W. 2018. Long noncoding RNA: multiple players in gene expression. *BMB Rep* **51:** 280–289. doi:10.5483/BMBRep.2018.51.6.025

Chennubhotla C, Bahar I. 2007. Signal propagation in proteins and relation to equilibrium fluctuations. *PLoS Comput Biol* **3:** 1716–1726. doi:10.1371/journal.pcbi.0030172

Correll CC, Wool IG, Munishkin A. 1999. The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J Mol Biol* **292:** 275–287. doi:10.1006/jmbi.1999.3072

Crick F. 1970. Central dogma of molecular biology. *Nature* **227:** 561–563. doi:10.1038/227561a0

Diener TO, Raymer WB. 1967. Potato spindle tuber virus: a plant virus with properties of a free nucleic acid. *Science* **158:** 378–381. doi:10.1126/science.158.3799.378

Ding B. 2010. Viroids: self-replicating, mobile, and fast-evolving noncoding regulatory RNAs. *Wiley Interdiscip Rev RNA* **1:** 362–375. doi:10.1002/wrna.22

Di Serio F, Flores R, Verhoeven JT, Li SF, Pallás V, Randles JW, Sano T, Vidalakis G, Owens RA. 2014. Current status of viroid taxonomy. *Arch Virol* **159:** 3467–3478. doi:10.1007/s00705-014-2200-6

Dissanayaka Mudiyanselage SD, Qu J, Tian N, Jiang J, Wang Y. 2018. Potato spindle tuber viroid RNA-templated transcription: factors and regulation. *Viruses* **10**. doi:10.3390/v10090503

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

Dozmorov MG, Giles CB, Koelsch KA, Wren JD. 2013. Systematic classification of non-coding RNAs by epigenomic similarity. *BMC Bioinformatics* **14:** S2. doi:10.1186/1471-2105-14-S14-S2

Duarte CM, Wadley LM, Pyle AM. 2003. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res* **31:** 4755–4761. doi:10.1093/nar/gkg682

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797. doi:10.1093/nar/gkh340

Eiras M, Kitajima EW, Flores R, Daròs JA. 2007. Existence in vivo of the loop E motif in potato spindle tuber viroid RNA. *Arch Virol* **152:** 1389–1393. doi:10.1007/s00705-007-0952-y

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816. doi:10.1038/nature05874

Friday D, Mukkara P, Owens RA, Baumstark T, Bruist MF. 2017. Processing of potato spindle tuber viroid RNAs in Yeast, a nonconventional host. *J Virol* **91:** e01078-17. doi:10.1128/JVI.01078-17

Gas ME, Hernández C, Flores R, Daròs JA. 2007. Processing of nuclear viroids in vivo: an interplay between RNA conformations. *PLoS Pathog* **3:** e182. doi:10.1371/journal.ppat.0030182

Giguère T, Perreault JP. 2017. Classification of the *Pospiviroidae* based on their structural hallmarks. *PLoS ONE* **12:** e0182536. doi:10.1371/journal.pone.0182536

Giguère T, Adkar-Purushothama CR, Perreault JP. 2014. Comprehensive secondary structure elucidation of four genera of the family *Pospiviroidae*. *PLoS ONE* **9:** e98655. doi:10.1371/journal.pone.0098655

Guéron M, Leroy JL. 2000. The i-motif in nucleic acids. *Curr Opin Struct Biol* **10:** 326–331. doi:10.1016/S0959-440X(00)00091-9

Hanson RM, Lu XJ. 2017. DSSR-enhanced visualization of nucleic acid structures in Jmol. *Nucleic Acids Res* **45:** W528–W533. doi:10.1093/nar/gkx365

Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, Urlaub H, Kastner B, Lührmann R, Stark H. 2018. Structure and conformational dynamics of the human spliceosomal B^{act} complex. *Cell* **172:** 454–464.e411. doi:10.1016/j.cell.2018.01.010

Havrila M, Réblová K, Zirbel CL, Leontis NB, Šponer J. 2013. Isosteric and nonisosteric base pairs in RNA motifs: molecular dynamics and bioinformatics study of the sarcin-ricin internal loop. *J Phys Chem B* **117:** 14302–14319. doi:10.1021/jp408530w

Hendrix DK, Brenner SE, Holbrook SR. 2005. RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys* **38:** 221–243. doi:10.1017/S0033583506004215

Hermann T, Patel DJ. 2000. RNA bulges as architectural and recognition motifs. *Structure* **8:** R47–R54. doi:10.1016/S0969-2126(00)00110-6

Hofacker IL, Stadler PF. 2006. Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics* **22:** 1172–1176. doi:10.1093/bioinformatics/btl023

Isom DG, Castaneda CA, Cannon BR, Garcia-Moreno B. 2011. Large shifts in pK_a values of lysine residues buried inside a protein. *Proc Natl Acad Sci* **108:** 5260–5265. doi:10.1073/pnas.1010750108

Itoh SG, Damjanović A, Brooks BR. 2011. pH replica-exchange method based on discrete protonation states. *Proteins* **79:** 3420–3436. doi:10.1002/prot.23176

Jaeger L, Verzemnieks EJ, Geary C. 2009. The UA_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res* **37:** 215–230. doi:10.1093/nar/gkn911

Jiang J, Smith HN, Ren D, Dissanayaka SM, Dawe AL, Wang L, Wang Y. 2018. Potato spindle tuber viroid modulates its replication through a direct interaction with a splicing regulator. *J Virol*. doi:10.1128/JVI.01004-18

Johnsson P, Lipovich L, Grandér D, Morris KV. 2014. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta* **1840:** 1063–1071. doi:10.1016/j.bbagen.2013.10.035

Jorgensen W, Chandrasekhar J, Madura J, Impey R MLK. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79:** 926–935. doi:10.1063/1.445869

Katsarou K, Rao AL, Tsagris M, Kalantidis K. 2015. Infectious long noncoding RNAs. *Biochimie* **117:** 37–47. doi:10.1016/j.biochi.2015.05.005

Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J* **20:** 4214–4221. doi:10.1093/emboj/20.15.4214

Klein DJ, Moore PB, Steitz TA. 2004. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* **340:** 141–177. doi:10.1016/j.jmb.2004.03.076

Kung JT, Colognori D, Lee JT. 2013. Long noncoding RNAs: past, present, and future. *Genetics* **193:** 651–669. doi:10.1534/genetics.112.146704

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499–512. doi:10.1017/S1355838201002515

Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30:** 3497–3531. doi:10.1093/nar/gkf481

López-Carrasco A, Flores R. 2017. Dissecting the secondary structure of the circular RNA of a nuclear viroid *in vivo*: a "naked" rod-like conformation similar but not identical to that observed *in vitro*. *RNA Biol* **14:** 1046–1054. doi:10.1080/15476286.2016.1223005

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6:** 26. doi:10.1186/1748-7188-6-26

Lu XJ, Olson WK. 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* **3:** 1213–1227. doi:10.1038/nprot.2008.104

Lu XJ, Olson WK, Bussemaker HJ. 2010. The RNA backbone plays a crucial role in mediating the intrinsic stability of the GpU dinucleotide platform and the GpUpA/GpA miniduplex. *Nucleic Acids Res* **38:** 4868–4876. doi:10.1093/nar/gkq155

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* **101:** 7287–7292. doi:10.1073/pnas.0401799101

Mattick JS. 2009. The genetic signatures of noncoding RNAs. *PLoS Genet* **5:** e1000459. doi:10.1371/journal.pgen.1000459

Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10:** 155–159. doi:10.1038/nrg2521

Mohan S, Noller HF. 2017. Recurring RNA structural motifs underlie the mechanics of L1 stalk movement. *Nat Commun* **8:** 14285. doi:10.1038/ncomms14285

Mongan J, Case DA, McCammon JA. 2004. Constant pH molecular dynamics in generalized Born implicit solvent. *J Comput Chem* **25:** 2038–2048. doi:10.1002/jcc.20139

Moore PB. 1999. Structural motifs in RNA. *Annu Rev Biochem* **68:** 287–300. doi:10.1146/annurev.biochem.68.1.287

Morra G, Verkhivker G, Colombo G. 2009. Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. *PLoS Comput Biol* **5:** e1000323. doi:10.1371/journal.pcbi.1000323

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci* **98:** 4899–4903. doi:10.1073/pnas.081082398

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

Owens RA, Baumstark T. 2007. Structural differences within the loop E motif imply alternative mechanisms of viroid processing. *RNA* **13:** 824–834. doi:10.1261/rna.452307

Petrov AI, Zirbel CL, Leontis NB. 2013. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* **19:** 1327–1340. doi:10.1261/rna.039438.113

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera? A visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605–1612. doi:10.1002/jcc.20084

Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* **136:** 629–641. doi:10.1016/j.cell.2009.02.006

Rangadurai A, Zhou H, Merriman DK, Meiser N, Liu B, Shi H, Szymanski ES, Al-Hashimi HM. 2018. Why are Hoogsteen base pairs energetically disfavored in A-RNA compared to B-DNA? *Nucleic Acids Res* **46:** 11099–11114. doi:10.1093/nar/gky885

Riesner D, Henco K, Rokohl U, Klotz G, Kleinschmidt AK, Domdey H, Jank P, Gross HJ, Sänger HL. 1979. Structure and structure formation of viroids. *J Mol Biol* **133:** 85–115. doi:10.1016/0022-2836(79)90252-3

Roe DR, Cheatham TE 3rd. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* **9:** 3084–3095. doi:10.1021/ct400341p

Ryckaert JP, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23:** 327–341. doi:10.1016/0021-9991(77)90098-5

Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. 2008. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol* **56:** 215–252. doi:10.1007/s00285-007-0110-x

Schrader O, Baumstark T, Riesner D. 2003. A mini-RNA containing the tetraloop, wobble-pair and loop E motifs of the central conserved region of potato spindle tuber viroid is processed into a minicircle. *Nucleic Acids Res* **31:** 988–998. doi:10.1093/nar/gkg193

Semancik JS, Szychowski JA, Rakowski AG, Symons RH. 1993. Isolates of citrus exocortis viroid recovered by host and tissue selection. *J Gen Virol* **74:** 2427–2436. doi:10.1099/0022-1317-74-11-2427

Shao J, Tanner SW, Thompson N, Cheatham TE. 2007. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J Chem Theory Comput* **3:** 2312–2334. doi:10.1021/ct700119m

Smith LG, Zhao J, Mathews DH, Turner DH. 2017. Physics-based all-atom modeling of RNA energetics and structure. *Wiley Interdiscip Rev RNA* **8:** e1422. doi:10.1002/wrna.1422

Sokoloski JE, Godfrey SA, Dombrowski SE, Bevilacqua PC. 2011. Prevalence of *syn* nucleobases in the active sites of functional RNAs. *RNA* **17:** 1775–1787. doi:10.1261/rna.2759911

Spackova N, Šponer J. 2006. Molecular dynamics simulations of sarcin-ricin rRNA motif. *Nucleic Acids Res* **34:** 697–708. doi:10.1093/nar/gkj470

Spasic A, Kennedy SD, Needham L, Manoharan M, Kierzek R, Turner DH, Mathews DH. 2018. Molecular dynamics correctly models the unusual major conformation of the GAGU RNA internal loop and with NMR reveals an unusual minor conformation. *RNA* **24:** 656–672. doi:10.1261/rna.064527.117

Šponer J, Bussi G, Krepl M, Banáš P, Bottaro S, Cunha RA, Gil-Ley A, Pinamonti G, Poblete S, Jurečka P, et al. 2018. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem Rev* **118:** 4177–4338. doi:10.1021/acs.chemrev.7b00427

Steger G. 2017. Modelling the three-dimensional structure of the right-terminal domain of pospiviroids. *Sci Rep* **7:** 711. doi:10.1038/s41598-017-00764-x

Steger G, Riesner D. 2018. Viroid research and its significance for RNA technology and basic biochemistry. *Nucleic Acids Res* **46:** 10563–10576. doi:10.1093/nar/gky903

Stombaugh J, Zirbel CL, Westhof E, Leontis NB. 2009. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* **37:** 2294–2312. doi:10.1093/nar/gkp011

Swails JM, York DM, Roitberg AE. 2014. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J Chem Theory Comput* **10:** 1341–1352. doi:10.1021/ct401042b

Takeda R, Petrov AI, Leontis NB, Ding B. 2011. A three-dimensional RNA motif in *Potato spindle tuber viroid* mediates trafficking from palisade mesophyll to spongy mesophyll in *Nicotiana benthamiana*. *Plant Cell* **23:** 258–272. doi:10.1105/tpc.110.081414

Thaplyal P, Bevilacqua PC. 2014. Experimental approaches for measuring pKa's in RNA and DNA. *Methods Enzymol* **549:** 189–219. doi:10.1016/B978-0-12-801122-5.00009-X

Tian S, Cordero P, Kladwang W, Das R. 2014. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **20:** 1815–1826. doi:10.1261/rna.044321.114

Veeraraghavan N, Bevilacqua PC, Hammes-Schiffer S. 2010. Long-distance communication in the HDV ribozyme: insights from molecular dynamics and experiments. *J Mol Biol* **402:** 278–291. doi:10.1016/j.jmb.2010.07.025

Wang Y, Ding B. 2010. Viroids: small probes for exploring the vast universe of RNA trafficking in plants. *J Integr Plant Biol* **52:** 28–39. doi:10.1111/j.1744-7909.2010.00900.x

Woods CT, Laederach A. 2017. Classification of RNA structure change by 'gazing' at experimental data. *Bioinformatics* **33:** 1647–1655. doi:10.1093/bioinformatics/btx041

Xie X, Guo P, Yu H, Wang Y, Chen G. 2018. Ribosomal proteins: insight into molecular roles and functions in hepatocellular carcinoma. *Oncogene* **37:** 277–285. doi:10.1038/onc.2017.343

Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. 2003. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* **31:** 3450–3460. doi:10.1093/nar/gkg529

Zhang K, Keane SC, Su Z, Irobalieva RN, Chen M, Van V, Sciandra CA, Marchant J, Heng X, Schmid MF, et al. 2018. Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid cryo-EM, NMR, and molecular dynamics approach. *Structure* **26:** 490–498.e3. doi:10.1016/j.str.2018.01.001

Zhong X, Leontis N, Qian S, Itaya A, Qi Y, Boris-Lawrie K, Ding B. 2006. Tertiary structural and functional analyses of a viroid RNA motif by isostericity matrix and mutagenesis reveal its essential role in replication. *J Virol* **80:** 8566–8581. doi:10.1128/JVI.00837-06

Zhong X, Archual AJ, Amin AA, Ding B. 2008. A genomic map of viroid RNA motifs critical for replication and systemic trafficking. *Plant Cell* **20:** 35–47. doi:10.1105/tpc.107.056606

Zhou X, Liao W, Liao J, Peng L, Hua L. 2015. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol* **7:** 92–104. doi:10.1093/jmcb/mjv014

Downloaded from rnajournal.cshlp.org on April 24, 2024 - Published by Cold Spring Harbor Laboratory Press

In silico survey of the central conserved regions in viroids of the *Pospiviroidae* family for conserved asymmetric loop structures

Paul Freidhoff and Michael F. Bruist

RNA 2019 25: 985-1003 originally published online May 23, 2019
Access the most recent version at doi:10.1261/rna.070409.119

Supplemental Material http://rnajournal.cshlp.org/content/suppl/2019/05/23/rna.070409.119.DC1

References This article cites 86 articles, 19 of which can be accessed free at: http://rnajournal.cshlp.org/content/25/8/985.full.html#ref-list-1

Creative Commons License This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

© 2019 Freidhoff and Bruist; Published by Cold Spring Harbor Laboratory Press for the RNA Society