BIOCHEMISTRY
including biophysical chemistry & molecular biology

New Concept

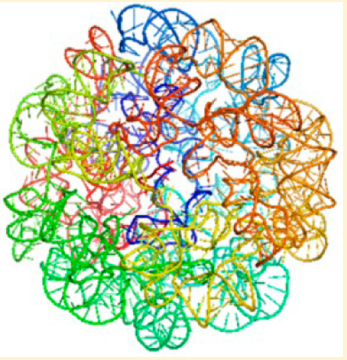
pubs.acs.org/biochemistry

Probing Viral Genomic Structure: Alternative Viewpoints and Alternative Structures for Satellite Tobacco Mosaic Virus RNA

Susan J. Schroeder*

Department of Chemistry and Biochemistry and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma 73019, United States

ABSTRACT: Viral RNA structure prediction is a valuable tool for development of drugs against viral disease. This work discusses different approaches to predicting encapsidated viral RNA and highlights satellite tobacco mosaic virus (STMV) RNA as a model system with excellent crystallography data. Fundamentally important issues for debate include thermodynamic versus kinetic control of virus assembly and the possible consequences of quasi-species in the primary structure on RNA secondary structure prediction of a single structure or an ensemble of structures. Multiple computational tools and chemical reagents are now available for improved viral RNA structure prediction. Two different predicted structures for encapsidated STMV RNA result from differences in three main areas: a different approach and philosophy to studying encapsidated viral RNA, an emphasis on different RNA motifs, and technical differences in computational methods and chemical reagents. The experiments with traditional chemical probing and SHAPE reagents are compared in terms of chemistry, results, and interpretation for STMV RNA as well as other RNA protein assemblies, such as the 5′UTR of HIV and the ribosome. This discussion of the challenges of viral RNA structure prediction will lead to new experiments and improved future predictions for viral RNA.



The prediction of RNA structure, function, and drug targets from sequence is currently a grand challenge in RNA biology. The need for accurate computational and chemical tools to predict RNA structures from sequence is increasing as a deluge of RNA sequence information is generated from transcriptome projects and next-generation sequencing technology. The application of such predictions holds promise for treating diseases caused by viruses with RNA genomes. Critical assessment and discussion of the strengths and weaknesses of different tools and approaches can move the field forward in developing the prediction tools that will help realize the full potential of new sequencing technology.

This concept paper uses satellite tobacco mosaic virus (STMV) as an example to contrast different approaches to viral RNA structure prediction and to compare traditional chemical probing reagents and SHAPE reagents. RNA structure prediction involves interpretation of chemical probing data. STMV is an excellent model because STMV has been experimentally studied by sequencing,[1−3] chemical probing,[4−6] and crystallography.[7,8] The crystal structure of STMV provides the first evidence of well-ordered RNA inside a virus particle and an amazing 1.8 Å resolution view of the RNA and RNA−protein interactions[7,8] (Figure 1). At each of the 30 2-fold axes of the $T = 1$ virus particle, an RNA helix of nine pairs interacts with two coat protein subunits. The icosahedral averaging used to determine the structure, however, obscures the details of which nucleotides in the STMV RNA genome form helices and the connections between helices. McPherson et al. proposed a cotranscriptional folding and virus assembly model based on the crystallography structure.[9] An ensemble model of STMV RNA secondary structure proposed by Schroeder et al. uses crystallographic data, traditional chemical probing data, and the cotranscriptional folding and virus assembly model to define a region of RNA conformational space that satisfies the experimental data on encapsidated STMV RNA.[6] The best scoring secondary structure from this ensemble provides the basis for the first all-atom model of a virus particle.[10]

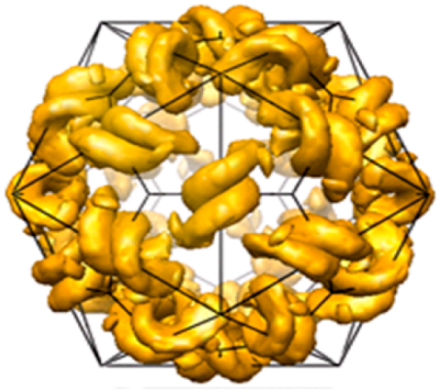


Figure 1. Electron density at 1.8 Å resolution for 30 helices of nine pairs in STMV RNA projected onto a $T = 1$ icosahedron.[7,77] The identities of the nucleotides in the helices and the density for nucleotides connecting the helices are obscured by icosahedral averaging.

Received: August 20, 2014
Revised: October 6, 2014
Published: October 16, 2014

dx.doi.org/10.1021/bi501051k | *Biochemistry* 2014, 53, 6728−6737

Recently, two alternative secondary structures were proposed on the basis of SHAPE reactivities in the *in vitro* transcribed[5] and encapsidated STMV RNA.[4] The structures for a viral RNA genome *in capsa* and *in vitro* are expected to be different. Different RNA structures are likely to occur from *in vivo* or *in vitro* assembly mechanisms and in the presence or absence of capsid proteins. Chemical probing for cucumber satellite mosaic virus clearly shows differences in the structure of the RNA genome *in planta*, *in capsa*, and *in vitro*.[11] Similarly, poliovirus RNA changes patterns of SHAPE reactivities *in capsa* and *in vivo*.[12] Differences across the transcriptome RNA structures in *Arabidopsis thaliana* were observed *in vitro* and *in vivo* for both DMS and SHAPE probing.[13] On the other hand, the substantial differences in the predicted structures for encapsidated STMV RNA presented by Schroeder et al. and Archer et al. lie in three main areas: a fundamentally different approach and philosophy to studying encapsidated viral RNA, an emphasis on different RNA motifs, and technical differences in computational methods, chemical reagents, and handling of viral particles. This work explores the data and different assumptions used in these models.

PHILOSOPHICALLY DIFFERENT APPROACHES TO PREDICTING ENCAPSIDATED VIRAL RNA STRUCTURE

Two fundamentally different approaches have been employed for interpreting chemical probing results in STMV. In one case, a single, minimum free energy structure is assumed. In the other case, an ensemble that is not based on free energy is assumed. These assumptions are considered below. Thermodynamic parameters and free energy minimization have been an integral part of RNA structure prediction and form the core of many widely used Web server tools.[14−16] However, a growing body of experimental evidence demonstrates the importance of kinetics and cotranscriptional folding in the determination of RNA structure.[17−22] Figure 2 shows a cartoon comparing

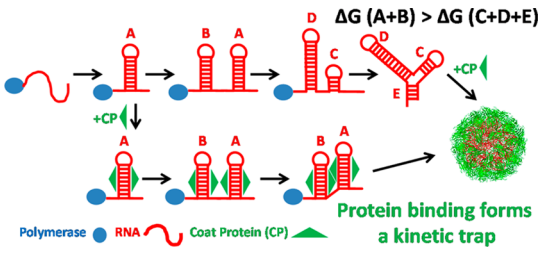


Figure 2. Cotranscriptional folding and virus assembly model. A thermodynamically driven path is shown above. A kinetically determined path in which coat proteins bind cotranscriptionally and stabilize helices in a kinetic trap is shown below. The final state is the three-dimensional model of STMV with the RNA colored red and the coat proteins colored green.[10]

possible thermodynamic and kinetic folding pathways for STMV. Free energy minimization has been a very successful strategy and predicts approximately 73% of base pairs correctly on average for RNAs known to have essentially single secondary structures.[23] The Contrafold program estimated nearest-neighbor parameters for Watson−Crick pairs from known three-dimensional RNA structures in the Protein Data Bank.[24] The rank order of nearest-neighbor parameters in Contrafold matched the order of stabilities for experimentally determined thermodynamic parameters for Watson−Crick pairs.[25] This result suggests that many known structures do fold to the lowest-free energy structure and validates one of the fundamental assumptions in free energy minimization approaches, i.e., that the lowest-free energy structure will be the functional *in vivo* structure. Protein chaperones may also facilitate RNA folding to the lowest-energy state by overcoming kinetic barriers.[26,27]

On the other hand, thermodynamic equilibrium rarely occurs *in vivo*,[28] and the assumptions of free energy minimization may not hold true for all RNA. In the case of viral genomic RNA, the RNA sequence encodes multiple functional structures that fold and unfold during the virus life cycle. If a viral RNA folded into a very stable minimum free energy structure, then another source of energy or a protein chaperone would be required to unfold the RNA before moving onto the next stage of the viral life cycle. The absence of covariation patterns in viral genomic phylogenies may be a consequence of selection for multiple functional structures during the stages of the virus life cycle, a kinetically driven and irreversible process to which the assumptions of free energy minimization may not apply. Thus, new tools, new approaches, and new ideas are necessary to evaluate the contributions of thermodynamics and kinetics in the determination of viral RNA structure.

Viral RNA May Be Different from Other RNA Types in Sequence, Structure, and Folding Pathway. Viral RNA may be very different at many levels of sequence, structure, and function from other RNA such as rRNA or tRNA that have been studied well by structural biologists. rRNA and tRNA have predominantly one function, one structure, and one sequence within an organism, while viral RNAs have multiple sequences, structures, and functions. Although sequences vary for tRNAs carrying different amino acids and vary for tRNA in different species, these sequence variations are consistent within a species and have a functional role in molecular recognition and specificities of tRNA aminoacylation.[29] In the case of viral RNA, the sequence variation is very high even within a single infection due to low-accuracy polymerases. The consensus sequence is really an ensemble average of sequences, or a quasi-species.[30] New sequencing technology is just beginning to probe the wide range of sequence variation in viral RNA such as poliovirus,[31] and the impact of the quasi-species character of viral RNA sequences on secondary and tertiary structures is only beginning to be understood. Whether multiple sequence variants fold into similar structures or into different structures with similar functional features is an exciting area of future research in viral RNA.

Viral RNA differs from tRNA or rRNA in several key aspects, such as the number of functional folds, the regulation of ribonucleoprotein complex assembly, and the lifetime of a folded functional structure. These differences may create different evolutionary selective criteria for favorable RNA conformations and different dependencies on thermodynamic stability. The single sequence for tRNA or rRNA folds into predominantly one structure with a primary function in protein translation. The recent discovery of an alternate functional fold of tRNA[32] and the ability of tRNA to unfold and act as a primer for some viral replication strategies[33] are exceptions to the predominant cloverleaf tRNA structure and its primary roles in reading the genetic code and catalyzing peptide bond formation.[34] The complex multicomponent folding and assembly pathways for rRNA are highly regulated and result in one ribonucleoprotein complex. tRNA and rRNA are among the most abundant cellular RNAs and have long-lived stable

dx.doi.org/10.1021/bi501051k | *Biochemistry* 2014, 53, 6728−6737

structures. After folding into a functional structure, tRNA and rRNA do not unfold to perform another function. In this case, thermodynamic stability may be a favorable characteristic in natural selection. In contrast, the consensus viral RNA sequence encodes multiple structures and functions throughout the viral infectious cycle. The structures of viral RNA cannot remain the same as the RNA genome replicates, translates, encapsidates, and escapes host defense strategies. Viral RNA must repeatedly fold and unfold into different transient structures during the virus life cycle. Thus, in this case, thermodynamic stability may be a disadvantage in natural selection. An ensemble of metastable structures that rapidly change conformations when different protein binding partners are present may be more advantageous for viral RNA genomes.

One consequence of the different characteristics of tRNA or rRNA versus viral RNA sequence variation is the existence of covariation patterns. tRNA and rRNA have clear patterns of covariation when sequences from different species are compared.[35] This covariation is used to determine the single conserved functional structure that the tRNA or rRNA sequence encodes. In contrast, viral RNA shows little or no nucleotide covariation, even in noncoding regions. The lack of covariation patterns in viral RNA may be a consequence of several factors, such as a high degree of sequence variation, conservation of the protein sequence code, and multiple metastable, functional RNA structures during the virus life cycle. An ensemble model may better accommodate the viral sequence diversity.[6] A lack of covariation may also be an indicator of the existence of multiple functional structures and signal that free energy minimization assumptions may not apply.

In the case of STMV RNA, nucleotides 53−642 encode the coat protein, and there is no covariation among the approximately 20 different natural consensus sequences.[1,2] Although the 3′ end of STMV RNA has been proposed to fold into a tRNA-like structure,[36] this proposed secondary structure is not consistent with either DMS or SHAPE chemical probing in the encapsidated state.[4,6] There is not yet any experimental evidence to support the assumption that STMV folds into a single structure in a way that is similar to that of tRNA or rRNA. The assumption that viral RNA folds into domains of ≤600 nucleotides like rRNA domains[4,37] also has not yet been supported experimentally. Thus, the motifs observed and assumptions for tRNA or rRNA folding may or may not apply to the way in which STMV and viral genomes fold.

Different Prediction Tools for a Single Minimum Energy Structure, an Ensemble, or Kinetically Determined Structures. Different computational tools for RNA structure prediction will generate an ensemble, a subset of suboptimal structures, or a single RNA structure. The characteristics of some common prediction approaches are discussed first (Table 1), and then the different approaches used to predict structures of encapsidated STMV RNA are compared. Thermodynamic parameters and other kinds of experimental data can be used in the scoring function of different prediction programs. Because there are only four different nucleotides, many RNA sequences have more than one possible secondary structure if only Watson−Crick pairing rules are applied with a minimum number of nucleotides for a hairpin turn and the exclusion of pseudoknots. The number of possible structures increases exponentially with the length of the sequence. The Zuker algorithm generates the single lowest-free energy structure and a subset of suboptimal structures in addition to the lowest-energy structure.[38,39] Considering suboptimal structures helps identify other functional structures that may not be the lowest-energy structure because protein binding and RNA tertiary structure are not accounted for in the thermodynamic parameters used for free energy minimization. The McCaskill algorithm considers an ensemble of structures and calculates the probabilities of base pairing, which facilitates interpretation of the likliehood of different regions of a structure.[40] For longer RNA, many possible structures exist within a narrow range of free energy. To address this challenge, the sfold algorithm by Ding and Lawrence calculates a Boltzman centroid, which best represents a cluster of similar structures.[41] Very recently, a spectral clustering approach has been developed to define RNA structural clusters with DMS chemical probing.[42]

Table 1

algorithm	refs	key features
Free Energy Minimization Approaches		
Zuker	38, 39	calculates the minimum free energy structure (MFE), samples suboptimal structures
McCaskill	40	computes base pairing probabilities
Ding and Lawrence	76	generates a centroid structure and cluster analysis
spectral clustering	42	identifies clusters in an ensemble
Contrafold	24	incorporates data from a structural database
Combinatorially Complete Enumeration Approaches		
Pipas and McMahon	44	first method for complete enumeration of structures
Wuchty	43	computes all structures within a free energy window
Crumple	45	enumerates all structures independently of free energy
Kinetic Approaches		
Zhao et al.	49	predictions use a cotranscriptional folding model that iteratively folds a growing sequence
Kinwalker	50	computes a folding pathway to the MFE
CoFold	51	considers cotranscriptional folding using pairing distance parameters

The Wuchty algorithm,[43] the algorithm by Pipas and McMahon,[44] and the Crumple algorithm[45] calculate all possible non-pseduoknotted structures for a given RNA sequence. The Wuchty algorithm calculates all possible structures within a given free energy window, the size of which depends on the length of the sequence. The Crumple algorithm can use thermodynamics as a filter to reduce the possible number of structures but does not require thermodynamic parameters. The advantage of the Crumple algorithm is the ability to modulate the importance of thermodynamic parameters in RNA structure predictions.[45] Chemical probing and other experimental constraints can reduce the possible number of structures and improve the accuracy of free energy minimization predictions.[23] However, chemical probing constraints do not necessarily define a single structure,[45] and consideration of suboptimal structures or ensembles of structures provides additional insight for biological RNA structure and function. The abundance of computational tools to address the RNA folding problem highlights the problem that a single minimum free energy structure is often insufficient to describe or predict biologically relevant RNA secondary structures.

In the case of STMV, the sequence space is not a folding funnel with a sharp point at a single lowest-free energy structure. Rather, the sequence space for STMV resembles a

dx.doi.org/10.1021/bi501051k | Biochemistry 2014, 53, 6728−6737

low broad basin with many diverse low-energy secondary structures. For example, the STMV sequence can form 42768 different structures with energies within 1 kcal/mol as predicted by using the Wuchty algorithm as implemented in RNAStructure,[46] and two different STMV structures have 560 nucleotides paired differently and yet differ by only 0.2 kcal/mol.[47] The McCaskill algorithm as implemented in both RNAStructure and the Vienna suite[48] reveals that the majority of the base pairs in these STMV structures have less than 50% probability. A modified Wuchty algorithm implemented in the context of the Vienna suite computes more than 4 billion structures without multibranch loops and more than 1 billion structures with multibranch loops that all satisfy the DMS, CMCT, and kethoxal chemical probing data and crystallography constraints for encapsidated STMV RNA but have up to 306 of 1058 nucleotides paired differently (J. W. Stone, S. Bleckley, S. Lavelle, and S, J, Schroeder, unpublished observations). Thus, there are many different possible folds with similar predicted free energies for the STMV sequence.

Archer et al.[4] and Schroeder et al.[6] used chemical probing in different ways to identify a single structure and an ensemble of structures, respectively. Archer et al. expanded the RNAStructure program to incorporate SHAPE constraints into RNA structure prediction and thus used SHAPE constraints and the Zuker algorithm to identify a single lowest-energy structure. In contrast, Schroeder et al. developed the Crumple algorithm and the Sliding windows and Assembly programs to include global constraints such as the length and minimum number of helices as a constraint in RNA structure prediction. Schroeder et al. used these tools to generate a set of hairpins that satisfy the crystallographic data; DMS, kethoxal, and CMCT chemical probing data; and the hypothesis of cotranscriptional folding and assembly. Any nonoverlapping set of 30 of these hairpins would be a reasonable solution to the STMV RNA folding problem. Schroeder et al. also used both the sfold algorithm and the Zuker algorithm in RNAStructure with constraints from traditional chemical probes to generate alternative structures that had lower predicted free energies but satisfied the crystallographic data less well.[6] Schroeder et al. demonstrated multiple possible structures using several different computational tools. Both groups developed new computational tools to incorporate new kinds of experimental constraints into RNA structure prediction.

Recently, deterministic computational methods for predicting RNA structures with consideration of a cotranscriptional folding process such as the kinetic model by Zhao et al.,[49] Kinwalker,[50] and CoFold[51] have been developed. The program based on the kinetic model of Zhao et al. can compute a cotranscriptionallly folded RNA secondary structure only for sequences of approximately ≤100 nucleotides and thus requires further development to tackle challenges such as the STMV folding problem. Kinwalker in the Vienna package generates a folding pathway to the minimum free energy structure. The minimum free energy structure for STMV RNA as computed in the Vienna package does not generate a structure with 30 helices of nine pairs that is consistent with the crystallographic data. Folding the STMV RNA sequence with the CoFold program, which also calls the Vienna package tools, generates a structure of 30 helices of nine pairs but is not consistent with either the traditional chemical probing data or SHAPE data and does not yet include an option to allow chemical probing data as constraints. Developing methods for computing or simulating RNA cotranscriptional folding accurately is an important area of active research that will have many applications for viral RNA genomes.

Experimentally Testing Computational Predictions with Sequence Mutations. Secondary structure predictions are often tested with site-directed mutagenesis and recovery of compensating Watson−Crick base pair changes.[52,53] In the case of STMV RNA, initial predictions from RNAStructure were tested by site-directed mutagenesis of helices with high probabilities. Care was taken to make silent mutations or minimal changes to the capsid protein sequence and to avoid any mutations in the polymerase binding site. Although the mutations, such as G436C, C466G, C206U, and G219A, predicted a significant change in the minimum free energy secondary structure, all mutated STMV RNA sequences replicated *in planta* without any noticeable phenotype.[54] This result is consistent with an ensemble model of hairpins that use imperfect helices. When STMV RNA sequences with the site-directed mutations or naturally occurring sequence variations are computationally folded,[1,2] a single-nucleotide change can completely change the predicted minimum free energy structure in Vienna's RNAfold or RNAStructure. However, the same mutations cause at most one or two helices to change in the predictions from Crumple, Sliding Windows, and Assembly. The mutations cause a change in score for only the hairpin that contains the mutation; thus, the ensemble view of STMV RNA changes only in the score of one helix (J.-W. Liu and S. J. Schroeder, unpublished observations). Thus, an ensemble view is more robust with respect to single-nucleotide changes that occur at relatively high rates in viral RNA. These *in planta* and *in silico* analyses of STMV mutations support the ensemble view of encapsidated STMV RNA.

■ DIFFERENT EMPHASIS ON RNA MOTIFS

The two different secondary structure models for encapsidated STMV RNA emphasize different structural motifs, either consecutive terminal mismatches or multibranch loops. The consecutive terminal mismatches in the model by Schroeder et al. have a score that is higher and more favorable than the scores of internal mismatches in the scoring function for STMV RNA in the Crumple program because the *B* factors in the crystal structure are higher for nucleotide bases at the end of the helix.[7] Higher *B* factors indicate less well-defined electron density that could be explained by the effects of icosahedral averaging electron densities for Watson−Crick pairs and a few noncanonical pairs. Consecutive terminal mismatches are thermodynamically stable[55,56] and can form A-form RNA helices (X. Gu, J. Malone, L. Thomas, S. Harris, B. H. M. Mooers, and S. J. Schroeder, unpublished observations) that would fit well into the electron density observed in the crystal structure. The thermodynamic parameters for consecutive terminal mismatches are not yet incorporated into the database for RNAstructure, which was used to generate the model of Archer et al.

The Crumple, Sliding Windows, and Assembly approach[6] explicitly focuses on a series of 30 hairpin loops and makes no attempt to predict additional base pairs that might occur in multibranch loops or provide additional thermodynamic stability. The model by Schroeder et al. was intentionally underpredicted to facilitate three-dimensional modeling. As stated in the article, "no predictions beyond the best 30 helices are attempted at this point because such interactions are likely convoluted with the tertiary and quaternary structure of the STMV RNA and capsid proteins."[6] The three-dimensional

dx.doi.org/10.1021/bi501051k | *Biochemistry* 2014, 53, 6728−6737

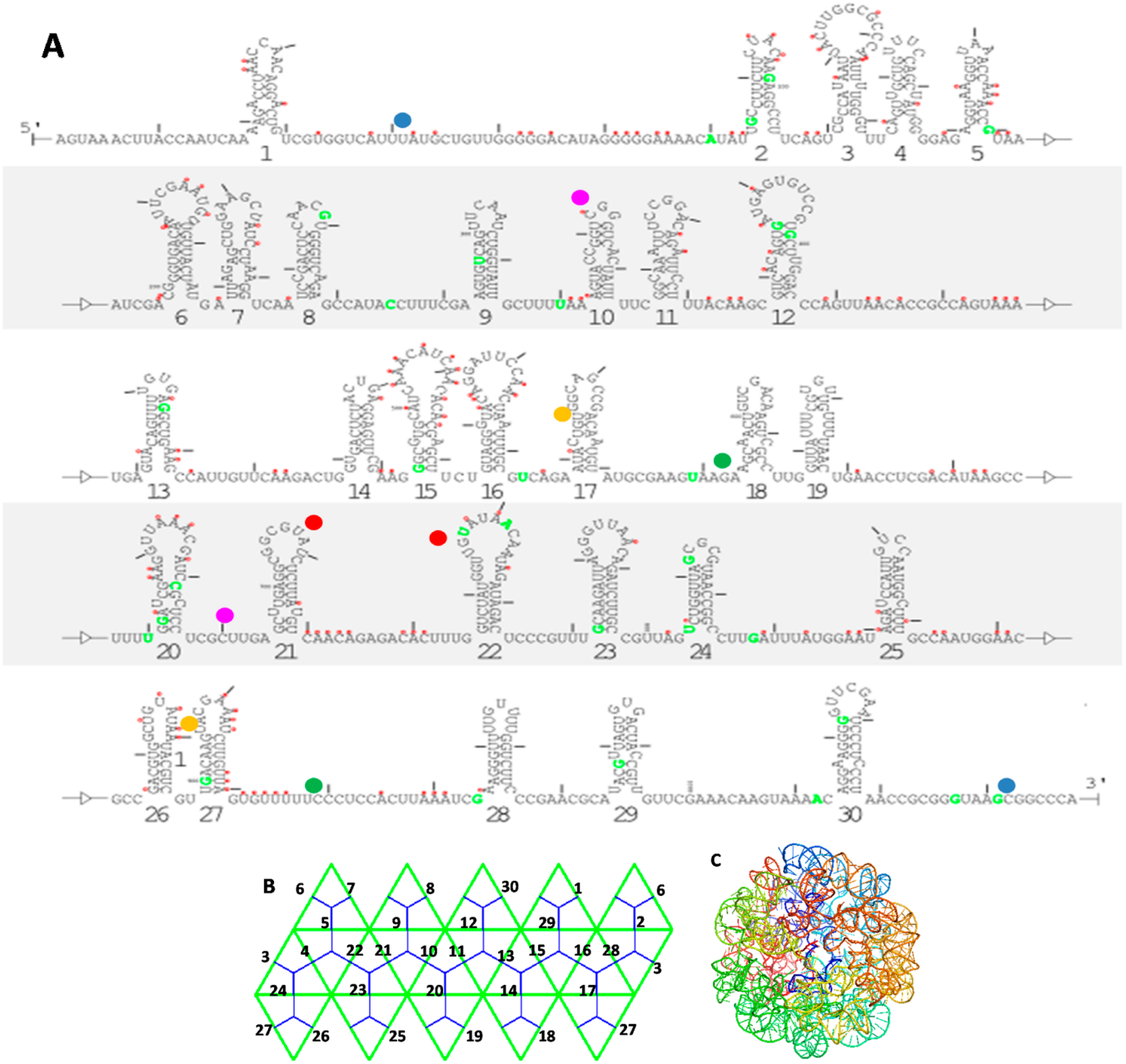


Figure 3. Models of STMV RNA secondary and tertiary structure. (A) Secondary structure model with the best score from the ensemble[6] labeled with points of tertiary base pairs in the three-dimensional model of STMV.[10] Nucleotides 54−53:1052−1053 are depicted as blue dots, nucleotides 332−334:691−693 as purple dots, nucleotides 567−568:909−908 as gold dots, nucleotides 592−594:929−927 as green dots, and nucleotides 707−708:747−746 as red dots. (B) Two-dimensional projection of the symmetry for a $T = 1$ virus particle.[77] The green lines are axes of 2-fold symmetry and are labeled with the RNA helix number from panel A as modeled in ref 10. Each green triangle is a face of the icosahedron, and the blue lines show points of 3-fold symmetry. (C) Three-dimensional structure of STMV RNA from the model proposed in ref 10. The RNA is colored from red to blue from the 5′ to 3′ end, respectively. The two ends of the RNA do come close in three-dimensional space in this model. All proteins were deleted from the structure for the sake of clarity.

model of STMV[10] reveals several potential pairing interactions, such as kissing hairpins and pseudoknots, as well as multibranch loops. For example, nucleotides 592−594:929−927 can form pairs that would create a multibranch loop, nucleotides 54−53:1052−1053 and 332−334:693−691 could form pseudoknots, and nucleotides 707−708:747−746 and 567−568:909−908 could form kissing hairpins in the model by Yingying et al. (Figure 3). These long-range stacked pairs could easily form to stabilize the encapsidated state but also would not require a large amount of energy to unfold. There are many possible additional pairing interactions that depend on how the 30 helices are placed in the 3D model. Currently, predicting such long-range pseudoknot interactions directly from a sequence of 1058 nucleotides is beyond the abilities of current computational methods that consider non-nested pairing.[57−59]

In contrast, the model by Archer et al. focuses on very stable, long-range multibranch loop motifs in the initial secondary structure prediction of the 30 helices observed in the electron density. The thermodynamic parameters for multibranch loops are difficult to measure experimentally, and different RNA

dx.doi.org/10.1021/bi501051k | *Biochemistry* 2014, 53, 6728−6737

folding programs implement the prediction rules for multibranch loops in different ways. For example, the 3′ end of the model by Archer et al. has a junction of seven helices, a very complicated motif for which no experimental observations have been previously made and no measurements could be made by optical melting. Thus, it is difficult to accurately calculate the additional thermodynamic stability for forming a multibranch structure. The helices that create multibranch loops in the Archer et al. model are much longer (up to 20 bp) than the two or three pair stacks that potentially form tertiary interactions in the model of Yingying et al. Thus, the long-range tertiary interactions in the model of Yingying et al. could possibly facilitate folding and unfolding transitions with lower energy barriers. The approach of Schroeder et al. predicted only 30 helices of nine pairs and left the remaining nucleotides unassigned to provide flexibility in modeling the three-dimensional structure and assembly pathways. In contrast, the inclusion of many stable long-range multibranch loops will restrict the number of possible ways to form a three-dimensional structure.

TECHNICAL DIFFERENCES IN VIRUS PARTICLE PREPARATION, REAGENTS, ANALYSIS, AND CHEMICAL PROBING RESULTS

Differences in Biochemical Protocols. The differences between the two proposed structures for encapsidated STMV RNA also result from distinctly different protocols in virus preparation, chemical reagents, and computational tools. Both studies began with STMV viral particles purified from infected *Nicotiana tabacum* leaves following procedures developed in the Dodd laboratory.[3,60] The particles used in the studies by Schroeder et al. were stored at 4 °C and never frozen or partially deproteinated prior to chemical analysis. In contrast, the STMV particles in the study by Archer et al. were frozen, and some were partially deproteinated. Freezing can damage virus particles and reduce infectivity. The consistency of the SHAPE results of Archer et al. between frozen and partially protein-degraded particles suggests that the freezing did indeed damage the virus particles rather than providing support for the partial protein degradation protocols. The atomic force microscopy studies presented by Archer et al.[4] do not provide sufficient resolution on the nucleotide level to assess differences in structure caused by freezing conditions or partial protein degradation. The intimate RNA−protein interactions observed in the high-resolution crystal structures suggest that any partial protein degradation would also disrupt the local RNA structure.

Traditional chemical probes and SHAPE reagents have different chemistries, different correlations with solvent accessibility, and different quenching requirements. Traditional chemical probing reagents, such as dimethyl sulfate (DMS), kethoxal (KX), and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluene sulfonate (CMCT), react at the Watson−Crick face of nucleotide bases (Figure 4). SHAPE reagents, such as *N*-methylisotoic anhydride (NMIA) and 1-methyl-7-nitrosatoic anhydride (1M7), react with the 2′-hydroxyl group on the ribose (Figure 4). Unlike traditional chemical probes, the SHAPE reagents require some dimethyl sulfoxide in the solution. DMS, CMCT, and KX require quenching by another reagent. In contrast, SHAPE reagents are quenched by water over time and do not require additional quenching reagents, which can be an advantage of this reagent in some experiments. Thus, the different reagents utilize different chemistries and probe different parts of the RNA. The

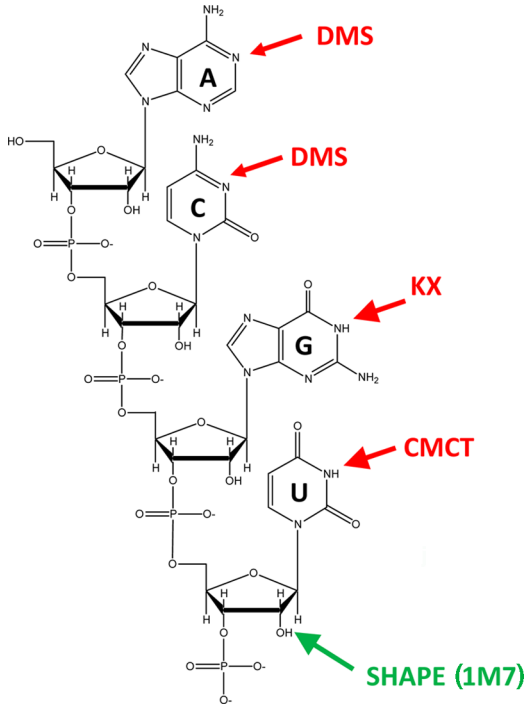


Figure 4. Site of the chemical modification for traditional chemical probes, including DMS, kethoxal, CMCT, and SHAPE reagents, such as NMIA and 1M7.

chemical accessibility of traditional chemical probes correlates well with solvent accessibility in rRNA and ribosome crystal structures with only a few exceptions explained by proteins quenching the reagent.[61] In contrast, hits from SHAPE reagents do not correlate with solvent accessibility in ribosome crystal structures,[62] leaving the mechanism of action for SHAPE probes not yet fully understood. One possible explanation for the low correlation with solvent accessibility is that SHAPE reagents may be probing partially unfolded or partially degraded RNA−protein interactions. Partially unfolded RNA would presumably have flexible ribose groups susceptible to SHAPE reactivity. If any unquenched SHAPE reagent remains active during the purification of RNA from the RNA−protein complex (including proteinase K digestion, phenol extraction, and ethanol precipitation), then SHAPE hits that do not correlate with solvent accessibility in crystal structures could occur. This is one possible explanation for the low signal-to-noise ratio observed in SHAPE reactions compared to DMS reactions in STMV RNA in the experiments of Schroeder et al.[6] (Figure 2 of the Supporting Information of ref 6).

The two approaches also employ different control reactions for reading chemical probing data on capillary electrophoresis. Schroeder et al. performed additional controls with dideoxy sequencing reactions on encapsidated STMV RNA that had been treated with buffers and no probing reagent and then purified with the same protocols. With increasing proteinase K digestion times, no additional peaks were observed in dideoxy sequencing reactions of unmodified encapsidated STMV RNA, thus demonstrating that RNA degradation did not occur during the purification process and the capsid protein was completely degraded before reverse transcription. No additional peaks occurred during increasing protinease K digestion times in chemically modified STMV RNA reverse transcription reactions, which suggests that the chemical probes were fully

dx.doi.org/10.1021/bi501051k | *Biochemistry* 2014, 53, 6728−6737

Table 2. Comparison of Traditional Chemical Probes and SHAPE Reagents

		SHAPE		traditional probes				
RNA	nucleotide	strong hits	medium hits	strong hits	medium hits	same strong hits	same strong and medium hits	ref
STMV RNA in virus particles	1058	127	251	160	–	30	94	4, 6
16S rRNA in 30S subunits	1542	255	269	38	82	13	71	67, 68
HIV 5′UTR *in vivo*[a]	340	27	24	62	27	27	49	37, 69
RNase P transcribed[b]	155	30	16	46	14	27	40	73
influenza A M1 mRNA transcribed	97	5	15	15	8	3	18	72

[a]The HIV 5′UTR was probed *in vivo* with only DMS in ref 69, and thus, only SHAPE hits for adenine and cytosine nucleotides are counted. [b]The RNase P hits were compared for the buffer condition of 135 mM KCl, 25 mM NaCl, and 50 mM HEPES without magnesium. This condition showed the most hits, and other buffer conditions showed changes in chemical reactivities at some nucleotides.

quenched (Supporting Information of ref 6). In contrast, Archer et al. performed different controls that focused on control reactions with *in vitro*-transcribed STMV RNA in dideoxy sequencing reactions as a reference ladder and did not perform additional control reactions to test for RNA degradation, incomplete capsid degradation, or unquenched chemical probes during purification.

Differences in Computational Analyses. The two approaches also used very different computational tools for analysis of chemical probing data, inclusion of chemical probing results, and RNA secondary structure prediction. Archer et al. used sophisticated SHAPEFinder software[63] to normalize signals and correct for signal decay in individual capillary electrophoresis traces. Signal levels are adjusted manually, and traces are subtracted to correct for background noise. Primers with different dye labels are used to generate traces for dideoxy sequencing, reagent probing, and no-reagent controls, and the signal processing and background corrections are done independently for each trace. The no-reagent control trace is subtracted after signal processing, and data are then averaged from multiple replicates. In contrast, Schroeder et al. used minimal computational processing in the Beckman Fragment Analysis equipment software to overlay results from capillary electrophoresis with no normalization, correction for signal decay, or background subtraction. Traces were aligned with Beckman size standards and dideoxy sequencing traces. All primers for reverse transcription used the same Cy5 dye label. A qualitative, simplistic, very conservative, manual process identified only the strongest, reproducible hits from chemical probing. Strong hits were observed at the lowest chemical probe concentration, could be reproduced at several chemical probe concentrations, and were consistently higher in intensity than other local signals. Any ambiguous hits were simply not used as constraints. The DMS experiments were analyzed by both traditional ^{32}P-labeled gel electrophoresis and capillary gel electrophoresis, which gave matching results from both independent analyses of "strong" hits. Thus, the data processing in the two studies was completely different and reflects different philosophies of signal processing, which is still a topic of debate in the literature.[64−66]

Differences in Chemical Probing Data. The chemical probing data and the way chemical probing data were used differ significantly in the two approaches. Schroeder et al. used only the strongest hits from chemical probing because, in an ensemble of structures, it is not possible to distinguish a weak hit in the majority of structures from a strong hit in a minority of structures. The strongest hits were then used as constraints in the Crumple, Vienna, RNAStructure, and Sfold programs. An ensemble was generated using the Crumple, Sliding Windows, and Assembly programs, and any nonoverlapping combination of 30 helices would be an experimentally valid structure. All hairpins in the ensemble and the best scoring structure are completely consistent with the defined rules for chemical probing hits. Chemical probing hits are allowed to occur in unpaired nucleotides and also in nucleotides in Watson−Crick pairs at the end of a helix, adjacent to a loop, or adjacent to GU pairs.[23]

In contrast, Archer et al. used all strong and weak hits for all nucleotides and thus used more total constraints for structure prediction. Only the RNAStructure software program specifically incorporates SHAPE data, so this was the program selected to generate a single minimum free energy secondary structure. As a result of the way RNAStructure incorporates SHAPE data as a pseudoenergy term, the lowest-energy structure may not be entirely consistent with the chemical probing rules that RNAStructure uses for other chemical probing reagents. The rules for DMS, CMCT, and kethoxal probing in RNAStructure allow chemical probing hits to occur in single strands, in pairs at the end of a helix, and in pairs adjacent to GU pairs but not in Watson−Crick pairs between two other Watson−Crick pairs.[23] For example, in the structure of Archer et al., three nucleotides, 217, 650, and 999, show strong SHAPE reactivities in the range of 0.7−1 and also occur in a Watson−Crick pair between two other Watson−Crick pairs in stable RNA helices. An additional 28 nucleotides show SHAPE reactivities in the range of 0.3−0.7 and also occur in a Watson−Crick pair between two other Watson−Crick pairs. Archer et al. also computed a minimum free energy structure with a maximum distance for pairing to generate a series of hairpins. This structure with a series of hairpins is more consistent with the SHAPE data, with only 15 violations of SHAPE reactivities in the range of 0.3−0.7. Archer et al. argued in favor of the multibranch loop model over the hairpin model because the multibranch loop model had a more favorable free energy. From an alternative viewpoint, perhaps structures that satisfy the SHAPE data very well can be generated if less emphasis is placed on minimizing the free energy.

The different approaches to STMV RNA structure prediction generated structures that have different levels of consistency with experimental chemical probing data. The free energy minimized structures proposed by Archer et al. are not fully self-consistent with the SHAPE data. The multibranched loop model[4] is not consistent with 31 strong and medium SHAPE hits[4] or 26 strong hits from chemical probing previously reported.[6] In contrast, the best scoring structure from the ensemble model[6] is entirely consistent with the traditional DMS, CMCT, and kethoxal chemical probing data and violates only five of the strong hits observed in subsequent SHAPE

dx.doi.org/10.1021/bi501051k | *Biochemistry* 2014, 53, 6728−6737

studies.[4] There are multiple possible combinations of 30 helices from the possible hairpins in the ensemble model[6] that would satisfy all the traditional chemical probing and SHAPE data. Thus, the STMV structures from the ensemble model[6] are more consistent with the reported data[4,6] than the multibranch STMV RNA model.[4]

The nucleotides that react with traditional chemical probes form a set very different from that of the nucleotides that react with SHAPE reagents in STMV RNA and also other RNA (Table 2). In the case of STMV RNA, only 30 nucleotides share the same strong hits in both traditional chemical probes and SHAPE reagents. Similar differences in chemical probing results occur in 16S rRNA in 30S subunits[67,68] and the 5′UTR in HIV *in vivo*.[37,69] The differences between traditional probes and SHAPE reagents would affect not only structure predictions but also identification of functional sites. For example, in a comparison of traditional chemical probing of 23S rRNA in 50S subunits and SHAPE data, 4 of 12 nucleotides that change intensities of reactivities and thus report on antibiotic binding, 14 of 26 nucleotides that report on A and P site tRNA binding,[70] 10 of 16 nucleotides that report on E site tRNA binding,[70] and 8 of 27 nucleotides that report on 30S subunit binding[71] were not reactive at all to SHAPE reagents.[67] Surprisingly, even A2451 at the *Escherichia coli* peptidyl transferase center was not reactive to SHAPE reagents.[67] Many of the differences in the ribosome probing involve flexible nucleotides on the surface whose conformation and dynamics change upon substrate binding. Thus, these are both cases in which the nucleotide base is reactive to traditional probes while the ribose is not reactive to SHAPE reagents and vice versa.

There are several possible explanations for the different results in traditional chemical probes and SHAPE reagents. Differences in sample preparation and methods of partial protein digestion as discussed above may account for some of the differences. In support of this source of differences, the match between traditional chemical probes and SHAPE reagents is better for two *in vitro* transcribed RNAs probed with both traditional and SHAPE reagents in the same laboratory[72,73] (Table 1). Additionally, a recent comparison of SHAPE and traditional chemical probes in the same laboratory in three riboswitches, tRNA Phe, and the P4−P6 group I intron reveals many sites of different reactivities for different reagents, some of which also correlate with specific three-dimensional structures, such as certain conformations of sheared GA pairs.[64]

The physical basis of SHAPE reactivities remains to be fully explained. SHAPE reactivity does not correlate with solvent accessibility[62,74] but does show some correlation to NMR order parameters for the C1′ atom with r values ranging from 0.73 to 0.89.[74] SHAPE data do not report directly on base pairing but can be parametrized to improve RNA structure prediction. Of course, structure predictions with different sets of constraints will generate different structures using any computational approach. Thus, understanding the physical basis for SHAPE reactivities will be important for integrating SHAPE data with other experimental constraints to predict structures of unknown RNA.

■ CONCLUSIONS AND FUTURE DIRECTIONS

The structures of viral RNA genomes will continue to be a challenging area of active research and debate. Improvements in experimental chemical probes, next-generation sequencing for chemical probing readout,[13,75] and computational tools will undoubtedly yield more insight into these potential drug targets. In future transcriptome projects and viral RNA studies, using all available chemical probes that give good signal-to-noise ratios will provide the most information about RNA conformations. Information from multiple chemical probes may also provide constraints for three-dimensional structural motifs.[64] Using multiple computational tools provides more confidence in the common features from different predictions. Computational tools for modeling cotranscriptional folding are a key focus for future development. Ensemble models may further develop to better describe mixed populations of RNA conformations and RNA shapeshifters. As more RNAs are probed under cellular conditions and single-cell and single-virion sequencing develops, we will gain even more detailed insight into the vast and varied landscape of RNA conformations. Future research on the folding pathways of viral RNA could test these assumptions about viral RNA and its similarity to other well-structured RNA. The goals of RNA structure prediction from sequence may shift from calculating a single lowest-energy structure to identifying the common functional structural features of an ensemble of RNA conformations.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: susan.schroeder@ou.edu. Fax: (405) 325-6111. Phone: (405) 325-3092.

Funding

This work was supported by grants from the National Science Foundation (CAREER Grant 0844913) and the Oklahoma Center for the Advancement of Science and Technology (HR13-206).

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

1M7, 1-methyl-7-nitrosatoic anhydride; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate; DMS, dimethyl sulfate; KX, kethoxal; NMIA, *N*-methylisotoic anhydride and 1-methyl-7-nitrosatoic anhydride; SHAPE, selective 2′-hydroxyl acylation analyzed by primer extension; STMV, satellite tobacco mosaic virus; UTR, untranslated region.

■ REFERENCES

(1) Kurath, G., Rey, M. E., and Dodds, J. A. (1992) Analysis of Genetic Heterogeneity Within the Type Strain of Satellite Tobacco Mosaic Virus Reveals Variants and a Strong Bias for G to A Substitution Mutants. *Virology 189*, 233−244.

(2) Kurath, G., Heick, J., and Dodds, J. A. (1993) RNAse Protection Analyses Show High Genetic Diversity among Field Isolates of Satellite Tobacco Mosaic Virus. *Virology 194*, 414−418.

(3) Mirkov, T. E., Mathews, D. M., Du Plessis, D. H., and Dodds, J. A. (1989) Nucleotide Sequence and Translation of Satellite Tobacco Mosaic Virus RNA. *Virology 170*, 139−146.

(4) Archer, E. J., Simpson, M. A., Watts, N. J., O'Kane, R., Wang, B., Erie, D. A., McPherson, A., and Weeks, K. M. (2013) Long-range architecture in a viral RNA genome. *Biochemistry 52*, 3182−3190.

(5) Athavale, S. S., Gossett, J. J., Bowman, J. C., Hud, N. V., Williams, L. D., and Harvey, S. C. (2013) In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. *PLoS One 8*, e54384.

(6) Schroeder, S. J., Stone, J. W., Bleckley, S., Gibbons, T., and Mathews, D. M. (2011) Ensemble of Secondary Structures for Encapsidated Satellite Tobacco Mosaic Virus RNA Consistent with

Chemical Probing and Crystallography Constraints. *Biophys. J. 101*, 167−175.
(7) Larson, N. B., Day, J., Greenwood, A., and McPherson, A. (1998) Refined Structure of Satellite Tobacco Mosaic Virus at 1.8 Å Resolution. *J. Mol. Biol. 277*, 37−59.
(8) Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A., and McPherson, A. (1993) Double Helical RNA in Stallite Tobacco Mosaic Virus. *Nature 361*, 179−182.
(9) Larson, S. B., and McPherson, A. (2001) Satellite Tobacco Mosaic Virus RNA: Structure and Implications for Assembly. *Curr. Opin. Struct. Biol. 11*, 59−65.
(10) Zeng, Y., Larson, S. B., Heitsch, C. E., McPherson, A., and Harvey, S. C. (2012) A model for the structure of satellite tobacco mosaic virus. *J. Struct. Biol. 180*, 110−116.
(11) Rodrigues-Alvarado, G., and Roossinck, M. J. (1997) Structural Analysis of a Necrogenic Strain of Cucumber Mosaic Cucumovirus satellite RNA in planta. *Virology 236*, 155−166.
(12) Burrill, C. P., Westesson, O., Schulte, M. B., Strings, V. R., Segal, M., and Andino, R. (2013) Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity. *J. Virol. 87*, 11670−11683.
(13) Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., and Assmann, S. M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature 505*, 696−700.
(14) Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinf. 11*, 129.
(15) Lorenz, R., Bernhart, S. H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol. AMB 6*, 26.
(16) Markham, N. R., and Zuker, M. (2008) UNAFold: Software for nucleic acid folding and hybridization. *Methods Mol. Biol. 453*, 3−31.
(17) Greenleaf, W., Frieda, K., Foster, D., Woodside, M., and Block, S. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science 319*, 630−633.
(18) Mahen, E. M., Harger, J. W., Calderon, E. M., and Fedor, M. J. (2005) Kinetics and thermodynamics make different contributions to RNA folding in vitro and in yeast. *Mol. Cell 19*, 27−37.
(19) Mahen, E. M., Watson, P. Y., Cottrell, J. W., and Fedor, M. J. (2010) mRNA Secondary Structures Fold Sequentially but Exchange Rapdily In Vivo. *PLoS Biol. 8*, e1000307.
(20) Lai, D., Proctor, J. R., and Meyer, I. M. (2013) On the importance of cotranscriptional RNA structure formation. *RNA 19*, 1461−1473.
(21) Frieda, K. L., and Block, S. M. (2012) Direct observation of cotranscriptional folding in an adenine riboswitch. *Science 338*, 397−400.
(22) Grohman, J. K., Gorelick, R. J., Kottegoda, S., Allbritton, N. L., Rein, A., and Weeks, K. M. (2014) An immature retroviral RNA genome resembles a kinetically trapped intermediate state. *J. Virol. 88*, 6061−6068.
(23) Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A. 101*, 7287−7292.
(24) Do, C., Woods, D., and Batzglou, S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics 22*, e90−e98.
(25) Xia, T., Santa Lucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic Parameters for an Expanded Nearest-neighbor Model for Formation of RNA Duplexes with Watson-Crick Base Pairs. *Biochemistry 37*, 14719−14735.
(26) Jarmoskaite, I., Bhaskaran, H., Seifert, S., and Russell, R. (2014) DEAD-box protein CYT-19 is activated by exposed helices in a group I intron RNA. *Proc. Natl. Acad. Sci. U.S.A. 111*, E2928−E2936.
(27) Jarmoskaite, I., and Russell, R. (2014) RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem. 83*, 697−725.
(28) Mohr, H., and Schopfer, P. (1995) The Cell as an Open System, Dynamic Equilibrium. In *Plant Physiology*, p 41, Springer-Verlag, Berlin.
(29) Shepotinovskaya, I., and Uhlenbeck, O. C. (2013) tRNA residues evolved to promote translational accuracy. *RNA 19*, 510−516.
(30) Biebricher, C. K., and Eigen, M. (2006) What is a quasispecies? *Curr. Top. Microbiol. Immunol. 299*, 1−31.
(31) Acevedo, A., Brodsky, L., and Andino, R. (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature 505*, 686−690.
(32) Rudinger-Thirion, J., Lescure, A., Paulus, C., and Frugier, M. (2011) Misfolded human tRNA isodecoder binds and neutralizes a 3′ UTR-embedded Alu element. *Proc. Natl. Acad. Sci. U.S.A. 108*, E794−E802.
(33) Mak, J., and Kleiman, L. (1997) Primer tRNAs for reverse transcription. *J. Virol. 71*, 8087−8095.
(34) Zaher, H. S., and Green, R. (2009) Fidelity at the molecular level: Lessons from protein synthesis. *Cell 136*, 746−762.
(35) Shang, L., Gardner, D. P., Xu, W., Cannone, J. J., Miranker, D. P., Ozer, S., and Gutell, R. R. (2013) Two accurate sequence, structure, and phylogenetic template-based RNA alignment systems. *BMC Syst. Biol. 7* (Suppl. 4), S13.
(36) Felden, B., Florentz, C., McPherson, A., and Giege, R. (1994) A histidine accepting tRNA-like fold at the 3′end of satellite tobacco mosic virus RNA. *Nucleic Acids Res. 22*, 2882−2886.
(37) Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature 460*, 711−716.
(38) Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science 244*, 48−52.
(39) Zuker, M., and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res. 9*, 133−148.
(40) McCaskill, J. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers 29*, 1105−1119.
(41) Ding, Y., Chan, C., and Lawrence, C. (2004) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res. 31*, 7280−7301.
(42) Homan, P. J., Favorov, O. V., Lavender, C. A., Kursun, O., Ge, X., Busan, S., Dokholyan, N. V., and Weeks, K. M. (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U.S.A. 111*, 13858−13863.
(43) Wuchty, S., Fontana, W., Hofacker, I. L., and Schuster, P. (1999) Complete Suboptimal Folding of RNA and the Stability of Secondary Structures. *Biopolymers 49*, 145−165.
(44) Pipas, J., and McMahon, J. (1975) Methods for Predicting RNA Secondary Structure. *Proc. Natl. Acad. Sci. U.S.A. 72*, 2017−2021.
(45) Bleckley, S., Stone, J. W., and Schroeder, S. J. (2012) Crumple: A Method for Complete Enumeration of All Possible Pseudoknot-Free RNA Secondary Structures. *PLoS One 7*, e52414.
(46) Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: Software for RNA structure prediction and analysis. *BMC Bioinf. 11*, 129.
(47) Schroeder, S. J. (2009) Advances in RNA structure prediction from sequence: New tools for generating hypotheses about viral RNA structure-function relationships. *J. Virol. 83*, 6326−6334.
(48) Gruber, A., Lorenz, R., Bernhart, S., Neubock, R., and Hofacker, I. (2008) The Vienna RNA Websuite. *Nucleic Acids Res. 36*, W70−W74.
(49) Zhao, P., Zhang, W., and Chen, S. J. (2011) Cotranscriptional folding kinetics of ribonucleic acid secondary structures. *J. Chem. Phys. 135*, 245101.
(50) Geis, M., Flamm, C., Wolfinger, M. T., Tanzer, A., Hofacker, I. L., Middendorf, M., Mandl, C., Stadler, P. F., and Thurner, C. (2008) Folding Kinetics of Large RNAs. *J. Mol. Biol. 379*, 160−173.

(51) Proctor, J. R., and Meyer, I. M. (2013) COFOLD: An RNA secondary structure prediction method that takes co-transcriptional folding into account. *Nucleic Acids Res. 41*, e102.
(52) Andino, R., Rieckhof, G. E., and Baltimore, D. (1990) A Functional Ribonucleoprotein Complex Forms around the 5′ End of Poliovirus RNA. *Cell 63*, 369−380.
(53) Tian, S., Cordero, P., Kladwang, W., and Das, R. (2014) High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA 20*, 1815−1826.
(54) Clanton-Arrowood, K. (2010) *Determining the Secondary Structure of Satellite Tobacco Mosaic Virus RNA: A Mutagenesis Approach*, Univerrsity of Oklahoma, Norman, OK.
(55) Clanton-Arrowood, K., McGurk, J., and Schroeder, S. J. (2008) 3′ Terminal Nucleotides Determine Thermodynamic Stabilities of Mismatches at the Ends of RNA Helices. *Biochemistry 47*, 13418−13427.
(56) Nguyen, M.-T., and Schroeder, S. J. (2010) Consecutive Terminal GU Pairs Stabilize RNA Helices. *Biochemistry 49*, 10574−10581.
(57) Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., and Weeks, K. M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A. 110*, 5498−5503.
(58) Dirks, R., and Pierce, N. (2004) An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *J. Comput. Biol. 25*, 1295−1304.
(59) Rivas, E., and Eddy, S. (2000) The language of RNA: A formal grammar that includes pseudoknots. *Bioinformatics 16*, 334−340.
(60) Mirkov, T. E., Kurath, G., Mathews, D. M., Elliott, K., Dodds, J. A., and Fitzmaurice, L. (1990) Factors Affecting Efficient Infection of Tobacco with in vitro RNA Transcripts from Cloned cDNAs of Satellite Tobacco Mosaic Virus. *Virology 179*, 395−402.
(61) Nguyenle, T., Laurberg, M., Brenowitz, M., and Noller, H. F. (2006) Following the dynamics of changes in solvent accessibility of 16 and 23 S rRNA during ribosomal subunit association using synchrotron-generated hydroxyl radicals. *J. Mol. Biol. 359*, 1235−1248.
(62) McGinnis, J. L., Dunkle, J. A., Cate, J. H., and Weeks, K. M. (2012) The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc. 134*, 6617−6624.
(63) Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M., and Giddings, M. C. (2008) ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA 14*, 1979−1990.
(64) Kladwang, W., Mann, T. H., Becka, A., Tian, S., Kim, H., Yoon, S., and Das, R. (2014) Standardization of RNA chemical mapping experiments. *Biochemistry 53*, 3063−3065.
(65) Kladwang, W., VanLang, C. C., Cordero, P., and Das, R. (2011) Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry 50*, 8049−8056.
(66) Leonard, C. W., Hajdin, C. E., Karabiber, F., Mathews, D. H., Favorov, O. V., Dokholyan, N. V., and Weeks, K. M. (2013) Principles for understanding the accuracy of SHAPE-directed RNA structure modeling. *Biochemistry 52*, 588−595.
(67) Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A. 106*, 97−102.
(68) Moazed, D., Stern, S., and Noller, H. F. (1986) Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J. Mol. Biol. 187*, 399−416.
(69) Paillart, J.-C., Dettenhofer, M., Yu, X., Ehresmann, C., Ehresmann, B., and Marquet, R. (2004) First Snapshots of the HIV-1 RNA Structure in Infected Cells and in Virions. *J. Biol. Chem. 279*, 48397−48403.
(70) Moazed, D., and Noller, H. F. (1989) Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell 57*, 585−597.
(71) Merryman, C., Moazed, D., Daubresse, G., and Noller, H. F. (1999) Nucleotides in 23S rRNA protected by the association of 30S and 50S ribosomal subunits. *J. Mol. Biol. 285*, 107−113.
(72) Jiang, T., Kennedy, S. D., Moss, W. N., Kierzek, E., and Turner, D. H. (2014) Secondary Structure of a Conserved Domain in an Intron of Influenza A M1 mRNA. *Biochemistry 53*, 5236−5248.
(73) Liang, R., Kierzek, E., Kierzek, R., and Turner, D. H. (2010) Comparisons between chemical mapping and binding to isoenergetic oligonucleotide microarrays reveal unexpected patterns of binding to the *Bacillus subtilis* RNase P RNA specificity domain. *Biochemistry 49*, 8155−8168.
(74) Gherghe, C. M., Shajani, Z., Wilkinson, K. A., Varani, G., and Weeks, K. M. (2008) Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J. Am. Chem. Soc. 130*, 12244−12245.
(75) Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M., and Bevilacqua, P. C. (2013) Determination of in vivo RNA structure in low-abundance transcripts. *Nat. Commun. 4*, 2971.
(76) Ding, Y., and Lawrence, C. (2001) Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res. 29*, 1034−1046.
(77) Shepherd, C. M., Borelli, I. A., Lander, G., Natarajan, P., Siddavanahalli, V., Bajaj, C., Johnson, J. E., Brooks, C. L., and Reddy, V. S. (2006) VIPERdb: A Relational Database for Structural Virology. *Nucleic Acids Res. 34*, D386−D389.