BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 5989–5997 5989
DOI: 10.1021/bi100393r

Nuclear Magnetic Resonance Structure of the Prohead RNA E-Loop Hairpin[†,‡]

Steven Harris and Susan J. Schroeder*

Department of Chemistry and Biochemistry, University of Oklahoma, 620 Parrington Oval, Norman, Oklahoma 73019

Received March 14, 2010; Revised Manuscript Received June 12, 2010

ABSTRACT: The *Bacillus subtilis* phage ϕ29 packaging motor requires prohead RNA for genome encapsidation. The nuclear magnetic resonance structure of the prohead RNA E-loop hairpin, r(5′AUUGAGUU), is presented and compared to predictions from MC-SYM. The prohead RNA E-loop hairpins contain sequences similar to rRNA hairpins. Comparison of predicted and experimentally determined prohead and ribosomal hairpin structures reveals that sequence similarity is a stronger determinant of hairpin structural similarity than grouping similar types of RNA. All the hairpins contain a U-turn motif but differ in the first noncanonical pair and backbone orientation. These structures provide benchmarks for further improvements in RNA structure predictions from sequence.

Bacillus subtilis phage ϕ29 utilizes a nanomolecular motor to encapsidate the DNA genome into an empty viral capsid (*1*, *2*). The ϕ29 motor contains an essential prohead RNA (pRNA)[1] that works in concert with several proteins to complete the packaging process (*3*−*5*). Previous chemical modification and phylogenetic studies on pRNA show very little primary sequence conservation but a well-conserved secondary structure (*6*−*9*). The pRNA E-loop hairpin sequence is conserved in ϕ29, M2/NF, and SF5 pRNA sequences but varies in GA1 pRNA (Figure 1) (*7*). The pRNA E-loop hairpin interacts with gene product 10, the viral head−tail connector protein, a 290-amino acid monomer that forms a 422 kDa dodecameric complex (*10*−*14*). Mutations in the pRNA E-loop hairpin reduce the level of protein binding to less than 30%, and mutation of the A nucleotide in the hairpin loop to a C reduces packaging activity 95% (*15*), although substitution of this hairpin with a UUCG tetraloop retains some packaging activity in vitro (*16*). Nucleotides in the pRNA E-loop hairpin show protection from chemical modification when pRNA forms dimers and multimers (*17*). Thus, the pRNA E-loop hairpin is a conserved structural piece of the RNA and protein interactions in the ϕ29 packaging motor.

NMR studies of RNA hairpins provide benchmarks for methods for predicting three-dimensional RNA structure from sequence. MC-SYM, a computational program that predicts RNA three-dimensional structure from sequence using symbolic programming and cyclic motifs, is trained on high-resolution ribosome crystal structures (*18*). Isostericity matrices, which classify and predict base pairing interactions, are based on phylogenetic and structural studies of rRNA and structures in the rfam database (*19*, *20*). pRNA presents a good test of these prediction methods because the degree of sequence conservation is unusually low but the secondary structure is well-conserved.

Ribosomal hairpins S2 in the small subunit and L11 in the large subunit were determined by both NMR and crystallography and contain sequences similar to pRNA hairpins. The ribosomal hairpins have the same structure both in the small oligonucleotide model of the hairpin determined by NMR and in the context of the RNA−protein interactions determined by crystallography (*21*−*27*). Thus, in this case, the three-dimensional structure of the ribosomal hairpins is determined by the RNA sequence and is not altered by protein and RNA tertiary interactions. Comparisons of the experimental and predicted structures for the pRNA E-loop and ribosomal hairpins S2 and L11 test the applicability of MC-SYM and isostericity matrices in predicting RNA structure.

Both the ribosome and the ϕ29 packaging motor are large RNA−protein assemblies with complex architectures that support and direct biological catalysis. The S2, L11, and pRNA E-loop hairpins share a common structural function at junctions where RNA helices and proteins pack closely together through nonspecific electrostatic interactions between residues such as arginine, lysine, glutamine, asparagine, and the RNA phosphate backbone. The high-resolution ribosome crystal structures (*25*−*27*) provide an enormous database of RNA−protein structural interactions upon which structure prediction methods can be developed and tested. The small size of these six-nucleotide hairpins is advantageous for testing computational predictions and assessing the accuracy of local three-dimensional structure prediction in detail. Accurate structure prediction from sequence can provide models for stimulating hypotheses of function in RNA−protein complexes. In the absence of high-resolution structural information for the complete ϕ29 packaging motor, atomic-resolution structure determination of parts of the motor can be modeled into the low-resolution cryo-electron structures to design experiments for determining the mechanism of the packing motor and the role of the essential pRNA. Atomic-resolution crystal structures of proteins gp7, gp12, gp10, and gp13 have been determined and modeled into cryoEM maps of the prohead motor at 7.9 Å (*10*, *28*−*34*). Models of pRNA hexamers and dimers have been computationally predicted (*35*, *36*), and experimental validation

[†]This work was supported by Oklahoma Center for the Advancement of Science and Technology Grant HR09-160, National Science Foundation Bridges to Doctorate Program Grant 0902027, and an American Cancer Society Institutional Research Grant to the University of Oklahoma Cancer Center (Seed Grant 090039).

[‡]The structure has been deposited in the Protein Data Bank as entry 2KVN and in RCSB as entry 101629.

*To whom correspondence should be addressed. E-mail: susan.schroeder@ou.edu. Phone: (405) 325-3092. Fax: (405) 325-6111.

[1]Abbreviations: NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlation spectroscopy; HSQC, heteronuclear single-quantum coherence spectroscopy; HETCOR, heteronuclear correlation spectroscopy; pRNA, prohead ribonucleic acid; rRNA, ribosomal ribonucleic acid; MC-SYM, macromolecular conformations by symbolic programming.

Published on Web 06/15/2010

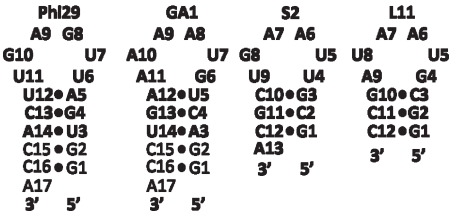

FIGURE 1: Sequence of the ϕ29 and GA1 E-loop and S2 and L11 rRNA hairpins. U3 is U50 in the ϕ29 RNA full-length sequence. Nucleotides not in bold were added for the stability of the NMR construct.

of prediction methods can increase our confidence in the ability of these types of models to inform structure−function hypotheses. This research presents the first experimental atomic-resolution structure of a piece of the pRNA component of the ϕ29 packaging motor. Thus, the NMR structure of the pRNA E-loop hairpin contributes to structure prediction and modeling efforts that build from small pieces following an *aufbau* approach to RNA structure prediction and structure−function relationships.

MATERIALS AND METHODS

RNA Preparation and Purification. The RNA was synthesized and gel purified by Dharmacon, Inc. Protecting groups were removed using the supplied deprotecting buffer, a 1% acetic acid solution at pH 3.8. The RNA oligonucleotide was lyophilized, resuspended in 10 mM $NaPO_4$, 10 mM NaCl, 0.5 mM Na_2EDTA pH 5.5 NMR buffer, dialyzed in 5 mM NaCl, 5 mM Na_3PO_4, 0.25 mM Na_2EDTA pH 6 buffer with a 1000 molecular weight cutoff membrane, and further purified using a Sephadex G-25 gel exclusion column. Oligomer purity was greater than 95%, proven by ^{32}P labeling and gel electrophoresis. The sample concentration was 1 mM on the basis of high-temperature UV absorbance measurements. For NMR experiments in D_2O, the oligomer was exchanged with 99.9% D_2O twice and once with 99.996% D_2O.

UV Experiments. A Beckman DU 800 UV−vis spectrometer was used to measure absorbance versus temperature melting curves at 260 nm from 20 to 90 °C with a heating rate of 1 °C/min. The sample was in NMR buffer, and the T_m was calculated with Meltwin version 3.5 (*37*) and the two-state model (*38*). Hairpin formation, rather than duplex formation, can be confirmed by a concentration-independent melting temperature.

NMR Spectroscopy. H_2O and D_2O NOESY, COSY, and aromatic $^1H-^{13}C$ HSQC NMR spectra were recorded on a Varian 500 MHz spectrometer. $^1H-^{31}P$ HETCOR and $^1H-^{13}C$ HSQC spectra were recorded on a Varian 600 MHz NMR spectrometer. NMR data were processed with Varian VNMR. The details of NMR data collection are included in the Supporting Information.

NMR Restraints and Peak Assignments. Distance restraints from the nonexchangeable NOESY spectra were based on NOE cross-peaks in the 50, 100, and 150 ms experiments. Restraints were classified as strong, medium, or weak if the NOE occurred in the 50, 100, or 150 ms spectra, respectively. The 400 ms spectra were used to confirm NOE assignments that were weak in the 150 ms spectra but were not used directly as distance restraints. Backbone, χ, and sugar pucker restraints for the stem nucleotides from the RNA backbone consortium (*39*) were applied during standard simulated annealing protocols (*40*). The backbone angles of the hairpin nucleotides were allowed to sample any conformation to fully explore conformational space with the exception of the syn G10 nucleotide. The backbone angles in the final structure ensemble are consistent with the observed cross-peaks in the COSY and ^{31}P HETCOR experiments. The SPARKY assignment program was used for peak assignment, labeling, and compilation of distance restraints and NOE lists (*41*). Table 1 contains the summary of NOE and dihedral restraints. A table of all the NOE restraints is part of the Supporting Information.

Table 1: Restraints and Ensemble Convergence for the ϕ29 E-Loop Hairpin[a]

base	intra	inter	dihedral	rmsd (Å)
G1	8	7	3	1.76
G2	8	9	5	1.38
U3	5	7	5	1.05
G4	6	8	5	1.00
A5	7	13	3	1.08
U6	7	12	0	1.23
U7	5	9	0	1.33
G8	11	4	0	3.49
A9	5	11	0	1.83
G10	10	9	2	2.34
U11	5	17	0	1.45
U12	7	7	3	1.35
C13	7	5	5	1.08
A14	6	7	5	0.92
C15	8	8	5	0.90
C16	10	7	5	1.02
A17	8	4	3	2.23
total	123	72	47	1.50

[a]Bases are numbered in the 5′ to 3′ direction. Intra refers to NOEs between protons on the same nucleotide. Inter refers to NOEs between protons on different nucleotides. Note that each NOE occurs between two protons, so the total internucleotide NOEs are the sum of the internucleotide NOEs for each nucleotide divided by 2. rmsd is an abbreviation for the all-atom root-mean-square deviation of the 31 structures in the final ensemble.

Structure Calculations. Structure calculations were performed with Crystallography and NMR Systems (CNS) version 1.2 (*42*) on a Dell Optiplex GX620 computer. An unrefined, initial MCSYM structure was heated for 10 ps at 500 K without any restraints to generate 30 randomized starting structures (rmsd of 6.1 Å). The RNA-DNA-all-atom parameter and topology files from CNS contain the force charges, atomic charges, atomic masses, and atom connectivity. The details of the simulation methods are presented in the Supporting Information. Pymol was used to analyze and visualize the structures (*43*).

Structure Predictions with MC-SYM. MC-SYM structure predictions were conducted using the web-hosted MC-Fold | MC-Sym pipeline service for RNA secondary and tertiary structure prediction at http://www.major.iric.ca/MC-Fold/. MC-Fold predicts secondary structures first with the Zuker algorithm (*44*) and Turner thermodynamics (*45*, *46*) and then selects structures to be submitted to MC-SYM, which generates predicted tertiary structures. These structures were then energy minimized and scored with the AMBER force field (*47*−*49*).

RESULTS

UV Absorbance. The plot of absorbance versus temperature of the hairpin gives a normal distinctive two-state melting profile. The predicted and measured melting temperatures and free energies at 37 °C are 70 °C and 7.5 kcal/mol in 1 M NaCl and 67 °C and 7.4 kcal/mol in 10 mM NaCl, respectively. These values are consistent within experimental error with estimates for the salt dependence of hairpin stabilities (*50*, *51*). The melting temperature was unperturbed with a varying oligomer concentration, which is consistent with the RNA sequence forming a hairpin and not a duplex (*38*).

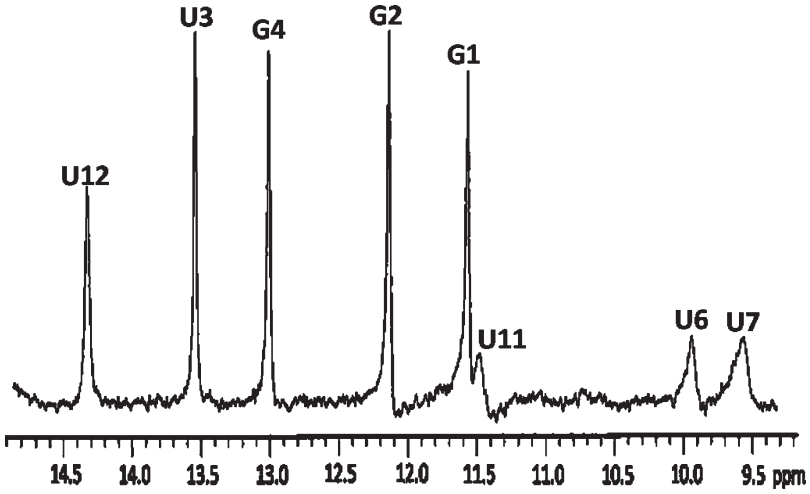

FIGURE 2: One-dimensional imino proton spectrum of ϕ29 E-loop hairpin r(5′GGUGAUUGAGUUCACCA) recorded at 1 °C and 500 MHz.

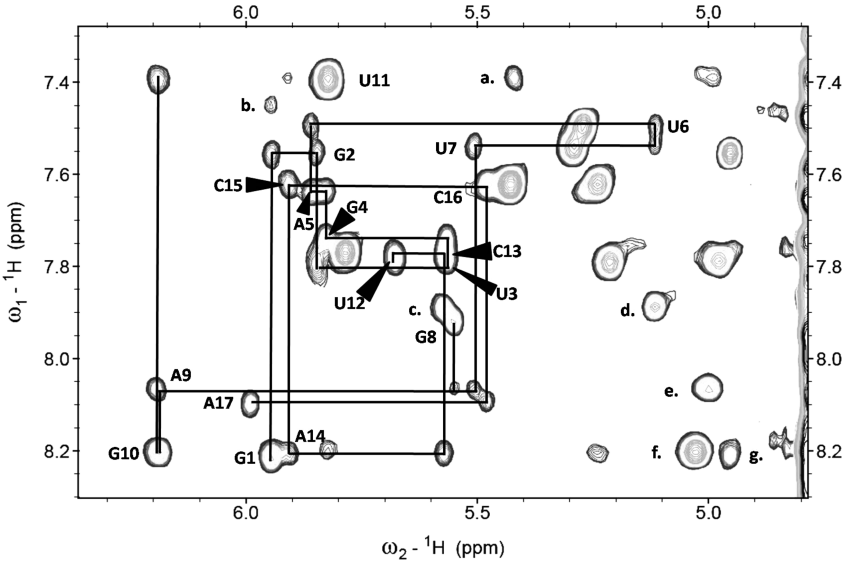

FIGURE 3: Base to H1′ and H5 region of the 400 ms D_2O NOESY spectrum recorded at 20 °C for r(5′GGUGAUUGAGUUCACCA). The base−H1′ NOESY walk is indicated by solid lines. Labels represent intranucleotide aromatic−H1′ NOEs. The other labels are as follows: (a) A14 H2−C15 H1′, (b) A17 H2−G1 H1′, (c) A5 H2−C13 H1′, (d) A5 H2−U6 H1′, (e) A9 H8−A9 H3′, (f) G10 H8−G10 H2′, (g) G1 H8−G1 H2′, and (h) G10 H1′−H8.

H_2O NOESY. In the imino proton spectra, only eight of the ten possible imino resonances were observed between 9 and 15 ppm (Figures 1 and 2 and Figure 5 of the Supporting Information). The G8 and G10 imino resonances were not seen. In the final model, G8 and G10 are solvent-exposed and do not form hydrogen bonds, thus allowing imino protons to exchange with water. The two most downfield imino protons at 14.43 and 13.9 ppm were assigned to U12 and U3, respectively; the imino resonances at 12.32 and 11.71 ppm were assigned to G2 and G1, respectively. These assignments were based on chemical shifts and the pattern of NOEs expected for the stem duplex (*52*, *53*). The remaining three imino protons at 11.63, 11.21, and 10.71 ppm were assigned to U11, U6, and U7, respectively. The strong NOE between U6 and U11 is consistent with the formation of a U-U mismatch base pair (*54*, *55*). These bases are also stacked on the closing A-U base pair, and this is confirmed by the observation of both bases having NOEs to U12. The remaining imino proton was assigned to U7. U7 H3 gives an NOE to U11 which is consistent with U7 stacking on the U6-U11 pair. Tables of complete proton assignments and all NOEs are included as Supporting Information.

D_2O NOESY. Sequential assignment of ribose and aromatic protons was conducted using the D_2O NOESY spectra and standard procedures (*53*, *56*) (Figure 3 and Table 1 of the Supporting Information). A normal NOESY sequential walk is observed for bases G1-U7 and U12-A17. In the hairpin loop, U7 shows no NOEs to G8, but U7 does show $N + 2$ NOEs to A9. Interestingly, NOEs between G8 and A9 occur. A9 shows $N + 2$ NOEs to U11. These NOEs define the U-turn in the hairpin. The 50 ms NOESY spectrum shows a strong H1′−H8 NOE similar in intensity to that of a pyrimidine H5−H6 cross-peak (Figure 7 of the Supporting Information). This NOE was assigned to G10. The possible assignment of this peak as an A9 H1′−G10 H8 NOE was considered much less likely because no NOE was observed between A9 H2′ and G10 H8 even at NOESY mixing times of 400 ms, which is long enough for either a direct NOE or secondary NOE to occur through a strong H1′−H2′ NOE and a strong H1′−H8 NOE.

DQFCOSY. An unusually strong and downfield-shifted H1′−H2′ NOE was observed and assigned to G10 which indicates a partial C2′ endo conformation (Figure 1 of the Supporting Information). In the preliminary structure calculations,

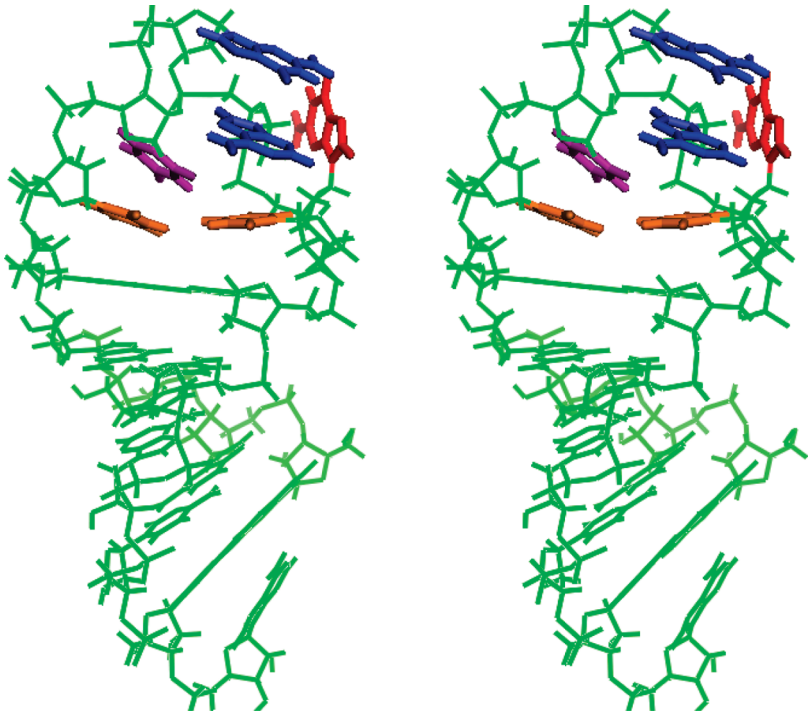

FIGURE 4: Wall-eye stereo image of the average structure of the prohead RNA E-loop hairpin, r(GGUGAUUGAGUUCACCA). The stem bases are colored green. U6 and U11 are colored orange. U7 is colored purple. G8 and A9 are colored blue. G10 is colored red.

the G10 sugar pucker was unrestrained and the ensembles of structures showed a mixture of C2′-endo and C3′-endo. The structure was then further refined with restraints for G10 to be C2′-endo to improve the convergence of the hairpin nucleotides. The COSY experiment also confirms that the sugar of A17 exhibits C2′-endo characteristics as expected for a 3′-dangling end nucleotide (*23*).

$^{1}H-^{13}C$ HSQC. The $^{1}H-^{13}C$ HSQC spectrum confirms key assignments for sugar proton resonances that were overlapped or had unusual chemical shifts. G10 H1′ and A9 H1′ both resonate at 6.18 ppm, but the carbon chemical shifts for A9 C1′ and G10 C1′ occur at 87.9 and 93.3 ppm, respectively (Figure 3 of the Supporting Information). U11 H1′ resonates at ~4.8 ppm in the $^{1}H-^{13}C$ HSQC spectrum. G10 H2′ and G1 H2′ are shifted downfield relative to the HDO peak.

The H2 protons of A5, A9, A14, and A17 were distinguished from all other protons by their carbon chemical shift (Figure 2 of the Supporting Information). Adenine H2 assignments were based on NOEs observed in both the H_2O and D_2O NOESY spectra. In the H_2O NOESY spectrum, both A5 and A14 H2 protons give cross-strand NOEs to U12 and U3 H3 protons, respectively. The A5 H2 proton also displays a cross-strand NOE to C13 H1′. The A9 H2 proton is shifted downfield and gives a NOE to its own H1′ sugar proton as well as an $N + 2$ NOE to U11 H4′. The A14 H2 proton gives a NOE to its own H1′ sugar proton and also to the C15 H1′ proton. The A17 base was added to the end of the sequence to improve the stability of the duplex and increase the probability of forming the terminal base pair (*23*). The A17 H2 proton gives a cross-strand NOE to G1 H1′, which verifies that A17 is stacked on the terminal G1-C16 base pair.

$^{1}H-^{31}P$ HETCOR. The phosphorus chemical shifts are within the normal range for A-form RNA except for A9 P and A17 P. In the final ensemble of structures, the A9 backbone angle ranges from −99.6° to −162.1°. A9 H3′ is shifted downfield relative to the HDO peak (Figure 5 of the Supporting Information). This assignment was also confirmed by the observation of a single peak having a carbon chemical shift of an H3′ proton in the $^{1}H-^{13}C$ HSQC spectrum and an A9 H8−A9 H3′ NOE at 8.065 and 5 ppm in the 100 ms D_2O NOESY.

Hairpin Structure. Figure 4 shows the U-U pair, syn G10, and U-turn in the ϕ29 E-loop hairpin. The structure was calculated from 195 measured NOE restraints (Table 1) as well as hydrogen bonding and dihedral angle restraints for the A-form stem helix. Thirty-one structures were calculated and had an overall rmsd of 1.50 Å from starting structures with an rmsd of 6.08 Å. When G8, G10, and A17 are excluded from the rmsd calculation, the overall rmsd is 1.24 Å. Figure 8A of the Supporting Information shows the overlay of 10 lowest-energy structures. Thus, the ensemble of calculated structures converges well, has a low final energy, and contains no NOE violations. All the expected NOEs in the average structure are observed in the NMR spectra. The hairpin turn is defined by a total of 33 internucleotide and 46 intranucleotide NOEs for six nucleotides and an all-atom rmsd of 1.76 Å (Table 3 of the Supporting Information). The U-turn occurs in all members of the ensemble. Table 1 lists the number of intranucleotide NOEs, internucleotide NOEs, and the all-atom rmsd for each nucleotide. The U-U pair is well-defined by 26 internucleotide NOEs and has an all-atom rmsd of 1.23 and 1.45 Å for U6 and U11, respectively. Figure 8B of the Supporting Information focuses on the U-U pairs in the overlay of 10 lowest-energy structures. Fewer internucleotide NOEs define the positions of G8, A9, and G10, which is consistent with the Watson−Crick edges of these bases facing the solvent. G8 and G10 positions are defined by mainly intranucleotide NOEs. These nucleotides are likely to be somewhat structurally dynamic in solution. The partial overlap of H8 and C8 resonances for G8, A9, and G10 combined with the G8 and G10 imino protons in fast exchange with water precludes further T1 or residual dipolar coupling measurements to further define the positions and dynamics of these

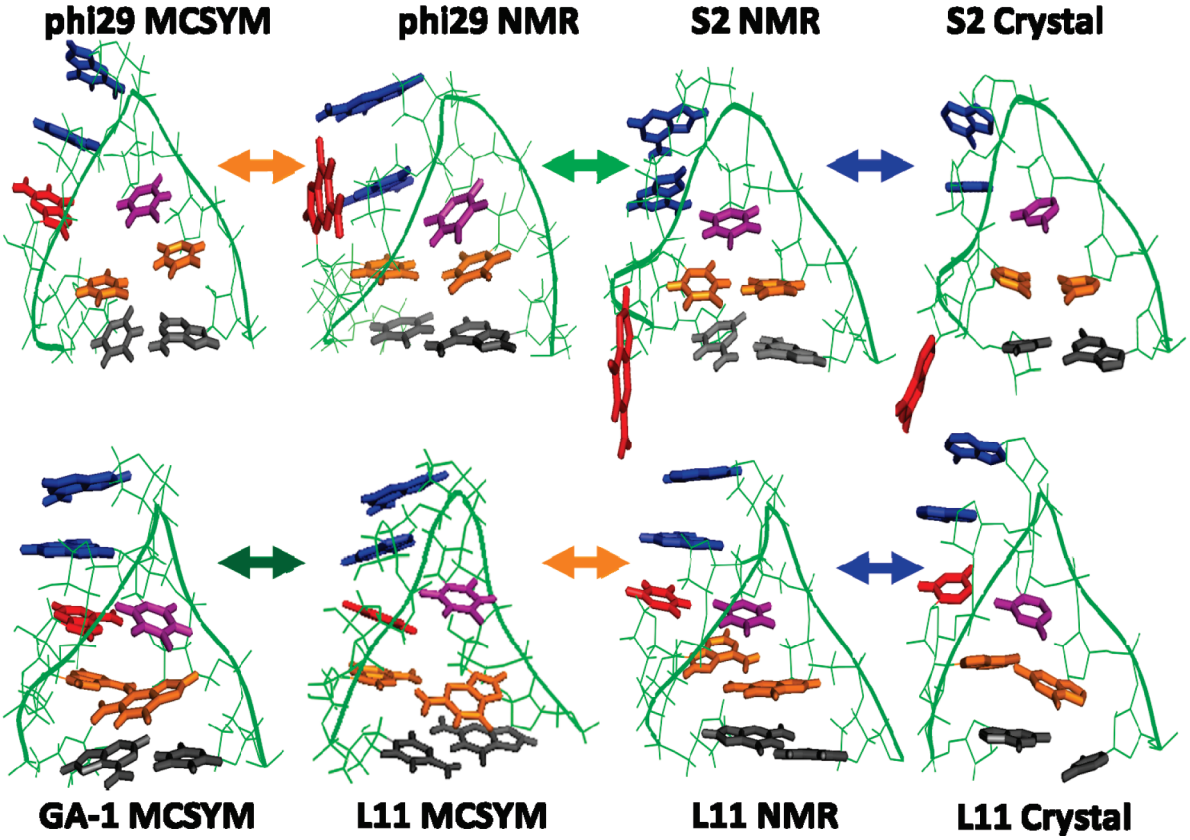

FIGURE 5: Comparison of prohead and rRNA hairpin three-dimensional structures. Blue arrows compare hairpins determined by NMR and crystallography. Orange arrows compare experimentally determined and predicted structures. The light and dark green arrows compare hairpins with different sequences and similar three-dimensional structures. The colors of the bases correspond to nucleotides in the same sequence position as the ϕ29 E-loop. The position of the G10 base in the S2 NMR hairpin is poorly defined, and only one average position is shown. Protein Data Bank entries 2WRJ (*64*) and 3HUW (*65*) were used to generate images of S2 and L11 hairpins in ribosome crystal structures.

bases. G10 projects out of the hairpin and into the solvent, and the ensemble positions of G10 are consistent with the strong intranucleotide H1′−H8 NOE defining the syn conformation and the H1′−H2′ COSY cross-peak defining the C2′-endo sugar pucker.

DISCUSSION

Structural Features of the ϕ29 E-Loop Hairpin. The U-U mismatch in the hairpin loop is confirmed by NOEs in the H_2O NOESY spectrum (Figure 5 of the Supporting Information). The U6 H6 proton follows the NOESY sequential walk observed in the D_2O NOESY spectrum , which is consistent with U6 stacking on A5 (*38*). U6 and U11 form a 2-carbonyl-N3, 4-carbonyl-N3 interaction or shifted *cis* Watson−Crick base pair (*20*). This type of U-U pair is similar to that observed in the S2 hairpin, r(5′GCGUUAAGUCGCA), found in the 30S ribosome (*22*, *24*, *26*). The stacking of U7 on the preceding U-U mismatch, which is confirmed by the cross-strand NOE observed in the H_2O NOESY spectrum, is also similar to the S2 hairpin. The ϕ29 E-loop hairpin has an A-U closing base pair. There is no observed effect of stacking the U-U pair on the closing A-U pair when compared to the S2 hairpin with a closing G-C base pair.

G10 does not pair within the loop but projects into solution. G10 has a syn conformation and a partial C2′-endo sugar pucker revealed in 50 ms NOESY and COSY spectra, respectively. The internucleotide NOEs for G10 include only G10 H8−U11 H4′ and G10 H5′− and G10 H5″−U11 H6 NOEs. The few internucleotide NOEs to G10 and the NOEs between A9 H2 and U11 H4′ confirm that the G10 base projects into solution (Figure 6 of the Supporting Information). The A9 H2−U11 H4′ NOE indicates that A9 stacks on the U11 sugar, and this stacking interaction causes an upfield shift of the U11 H1′ resonance.

The uridine turn observed in the ϕ29 E-loop hairpin occurs in many hairpins (*57*, *58*). The characteristic UNR sequence occurs in the ϕ29 E-loop (5′U6-U7-G8-A9-G10-U11), the S2 hairpin, and the anticodon and T-loop of yeast tRNA (*59*, *60*). The U7 H1′ and U7 H2′ sugar protons both give $N + 2$ NOEs to the A9 H8 proton which causes a kink in the backbone and is consistent with the location of the U-turn between U7 and G8.

Comparison of Prohead and Ribosomal Hairpin Structures. Comparisons of prohead RNA E loop hairpins and rRNA hairpins suggest that the NMR structure of the ϕ29 E loop will likely represent the functional structure of the RNA in the ϕ29 packaging motor. The S2 hairpin is similar in sequence to the bacteriophage ϕ29, M2/NF, and SF5 prohead E loop sequences (Figures 5 and 6, Table 2, and Table 7 of the Supporting Information). The L11 hairpin is similar in sequence to the GA1 prohead E-loop sequence (Figures 5 and 6, Table 2, and Figure 7 of the Supporting Information). The hairpins in rRNA motifs that bind small ribosomal subunit protein S2 and large ribosomal subunit protein L11 have been studied by NMR as isolated hairpins in solution and by crystallography of entire ribosomal subunits (*21*−*27*). The rRNA conformations are exactly the same in isolated hairpins and in rRNA−protein complexes. In these cases, a small section of the rRNA does represent the functional structure in the larger context of the ribosome. Thus, the rRNA sequence determines the hairpin structure, which does not change upon protein binding or tertiary RNA interactions.

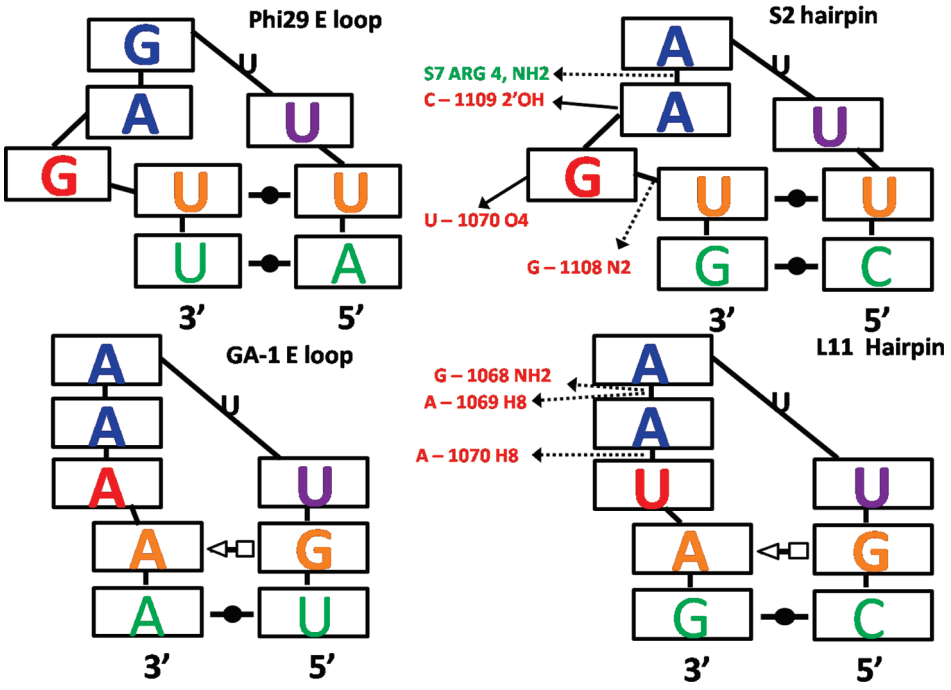

FIGURE 6: Diagram of base stacking in prohead and rRNA hairpins. Bases are represented by boxes, and the colors of the bases correspond to nucleotides in the same sequence position as the ϕ29 E-loop. Lines represent the ribose and phosphodiester backbone between bases. The U-turn location is indicated by a U. Arrows indicate hydrogen bonding in protein or RNA tertiary interactions in the crystal structures of ribosomal hairpins (*21*−*27*, *64*, *65*). The structure of the GA-1 E-loop hairpin was predicted by MC-SYM.

Table 2: Comparison of Prohead and Ribosomal Hairpin Structures[a]

hairpin	first pair	C2′-endo	syn
ϕ29 NMR	U-U cis-WC/WC	G10	G10
ϕ29 MC-SYM	U-U cis-WC/WC	none	U7
S2 NMR[b]	U-U cis-WC/WC	G8,G10	none
L11 NMR[c]	G-A trans-S/H	U9,G10	none
GA-1 MC-SYM	G-A trans-S/H	none	none

[a]The first pair conformations are described by the isostericity matrix nomenclature (*20*). C2′-endo refers to hairpin nucleotides with the C2′-endo sugar pucker conformation. Syn refers to the conformation of the glycosidic bond between the ribose and base. All numbering is consistent with the ϕ29 NMR construct. [b]From refs *24* and *63*. [c]From ref *23*.

The structures in Figures 5 and 6 show the conserved features in hairpins with similar sequences. All of the hairpins in the comparison are characterized by a noncanonical base pair stacked on the helix and a uridine stacked on the noncanonical base pair. All of the hairpins also incorporate the U-turn motif, which causes the two subsequent bases to stack upon one another. In the ϕ29 E-loop and the S2 hairpin, the G10 base is protruding into solution, but the GA-1 E-loop prediction and the L11 protein binding RNA have this base stacked under 5′ purines. This difference in purine stacking may be the result of the different backbone orientation for the first cis U-U or trans G-A pair.

Comparison of ϕ29 NMR Hairpin Structure to MC-SYM Structure Predictions. The lowest-energy MC-SYM predicted structure has an overall rmsd of 2.48 Å with respect to the experimentally determined NMR structure of the ϕ29 E-loop hairpin (Figure 6). Both structures have a U-U pair, a U stacked on the U-U pair, and a U-turn motif. The differences occur in the stacking of G10, the nucleotide with a syn conformation, and the subtleties of the U-U pair hydrogen bonding. The energetically unfavorable syn conformation of G10 is balanced by a very favorable conformation of U6, U7, and U11 in the NMR structure. In contrast in the MC-SYM predicted structure, G10 is stacked in a favorable conformation and the U-U pair is in a less favorable conformation. The predicted structure shows G10 stacked under G8 and A9 rather than extending into solution. The U-U mismatch in the MC-SYM prediction shows the N-3 imino proton of the U bases directly facing each other. This orientation does not allow the C-2 or C-4 carbonyl close enough to the imino protons to form a hydrogen bond (4 Å). In the ϕ29 NMR structure, the shifted cis-Watson−Crick, Watson−Crick U-U pair has two hydrogen bonds (2-carbonyl-N3 and 4-carbonyl-N3). In the MC-SYM prediction, U7 is oriented in the syn conformation, which is not observed in the NMR structure. The U7 imino−U11 imino NOEs in the H_2O NOESY spectrum define the U7 stacking orientation in the ϕ29 NMR structure. The MC-SYM predicted structure violates only two NOEs: U7 H3−U11 H3 and A9 H2−U11 H4′. The U7 H3−U11 H3 NOE is not compatible with a U7 syn conformation, and the A9 H2−U11 H4′ NOE is not compatible with G10 being stacked below the 5′ purines. If these two NOEs are included in the MC-SYM simulation, then the predicted structure has an rmsd of 1.94 Å with the NMR structure (Figure 7 of the Supporting Information).

A comparison of ribosomal hairpins and MC-SYM predictions demonstrates the generally accurate predictive ability of MC-SYM. The rmsd was 1.5 Å between the lowest-energy structure for the MC-SYM predicted L11 hairpin and the crystal structure (Figure 6). The lowest-energy S2 hairpin structure compared to the crystal structure has an rmsd of 1.4 Å (Figure 7 of the Supporting Information) for the loop nucleotides. These comparisons provide a basis for evaluating the MC-SYM predicted structures relative to the ϕ29 NMR structure and confidence in the predicted structure for the GA1 E-loop hairpin.

Improvements in the force field parameters for RNA and inclusion of more diverse structures in the database could further improve MC-SYM predictions. The AMBER force field uses parameters

optimized for the DNA phosphodiester backbone and deoxyribose angles, rather than RNA backbone conformations (*47*−*49*). A more polarizable force field optimized for RNA could improve predictions of the ways in which RNA hairpin backbones fold (*61*, *62*). As the diversity of experimentally determined RNA structures increases, the database for optimizing MC-SYM scoring functions expands, thus improving prediction capabilities for novel structures. While research to improve RNA force fields and structural databases continues, the incorporation of experimental data, such as NOEs, can refine MC-SYM predictions.

Comparison of Hairpin Structures Determined by NMR to Predictions from Isostericity Matrices. Because the ϕ29 and GA1 pRNAs have a conserved secondary structure and function, one hypothesis may be that the tertiary structure may be conserved as well. Predicting base pair structures with isostericity matrices is based on the assumption that RNAs with similar functions will retain a similar structure. The comparisons of the experimental and predicted structures for ribosomal and prohead hairpins, however, suggest different structures for the ϕ29 and GA1 E-loop hairpins. The only known isosteric U-U and G-A pair forms hydrogen bonds only between the ribose edges and leaves the Watson−Crick edges solvent exposed, which is inconsistent with the NOEs between imino protons in U6 and U11 in the ϕ29 E-loop. There are no known isosteric pairs that conserve the orientation and C1′−C1′ distance for U-U and G-A pairs with hydrogen bonding at the Watson−Crick edges. The U-U pair in the ϕ29 E-loop is in the cis-Watson−Crick/Watson−Crick (2-carbonyl-N3, 4-carbonyl-N3) form, and the G-A pair in the predicted GA1 E-loop is in the trans-sugar/Hoogsteen (N7-amino, amino-N3 sheared) form (*20*, *38*). However, the experimentally determined ϕ29 U-U pair has a longer than expected C1′−C1′ distance of 10 Å, rather than the 8.1 Å C1′−C1′ distance expected from isostericity matrices. The predicted GA1-GA pair has a C1′−C1′ distance of 9.0 Å, which is similar to the 9.7 Å distance expected from isostericity matrices. Thus, both pRNA E-loop hairpins have C1′−C1′ distances within 1 Å for the first pair. The significant difference occurs in the cis or trans backbone conformation of the first pair in the hairpin loop. This difference may cause the G10 nucleotide to flip out in a syn conformation in the ϕ29 hairpin and the adenine to stack under the other adenines in the GA1 predicted structure. Although the overall backbone structure may not be conserved between the ϕ29 and GA1 E-loops, the U-turn and the uridine stacking on the first mismatch will likely be conserved. Differences between the gp10 proteins and surrounding RNA tertiary interactions may accommodate the different shapes of the ϕ29 and GA1 hairpins.

CONCLUSIONS

This research presents the NMR structure of the ϕ29 E-loop hairpin and compares the experimentally determined structure to predictions from MC-SYM and isostericity matrices. The comparisons between ϕ29 and GA1 and ribosomal hairpins provide benchmarks for testing RNA structure prediction methods. All the hairpins contain a well-predicted U-turn motif. The ϕ29 hairpin is more similar in structure and sequence to the ribosomal S2 hairpin than to predicted structures for the GA1 hairpin, which is similar in sequence and structure to the ribosomal L11 hairpin. Thus, these two pRNA hairpins are not predicted to be isosteric despite sharing the same functional role in the same type of RNA. The first difference appears to be whether the first pair is a U-U pair with cis backbone orientations or a G-A pair with trans backbone orientations. The second difference is whether the G10 nucleotide in ϕ29 or its equivalent purine projects out into solution or stacks under the 5′ purines. The differences between the ϕ29 NMR structure and the MC-SYM predictions suggest improvements in estimating the stabilities of non-A-form backbone conformations and syn conformations is necessary for future improvements in RNA three-dimensional structure predictions.

ACKNOWLEDGMENT

We thank Scott Kennedy for ^{13}C and ^{31}P NMR data collection and Marc Parisien for helpful instructions and discussions about MC-SYM.

SUPPORTING INFORMATION AVAILABLE

Detailed NMR data collection and structure calculation methods; tables of proton, carbon, and phosphorus chemical shift assignments; a table of NOE distance restraints; a table comparing structure calculations with and without G10 backbone restraints; a COSY spectrum; ^{13}C HSQC spectra; a ^{31}P HETCOR spectrum; a H_2O NOESY spectrum; sections of the D_2O NOESY spectrum showing A9 H2 NOEs and G10 NOEs; a figure showing the overlay of the lowest-energy structures in the calculated ensemble of structures; and a figure showing the superposition of the hairpin structures compared in Figures 5 and 6. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Moffitt, J. R., Chemla., Y. R., Aathavean, K., Grimes, S., Jardine, P. J., Anderson, D. L., and Bustamante, C. (2009) Intersubunit coordination in a homomeric ring ATPase. *Nature 457*, 446–451.
2. Lee, T. J., Zhang, H., Chang, C. L., Savran, C., and Guo, P. (2009) Engineering of the fluorescent-energy conversion arn of ϕ29 DNA packaging motor for single molecule studies. *Small 5*, 2453–2459.
3. Guo, P., Grimes, S., and Anderson, D. (1986) A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage ϕ29. *Proc. Natl. Acad. Sci. U.S.A. 83*, 3505–3509.
4. Guo, P., Bailey, S., Bodley, J. W., and Anderson, D. (1987) Characterization of the small RNA of the bacteriophage ϕ29 DNA packaging machine. *Nucleic Acids Res. 15*, 7081–7090.
5. Guo, P. X., Erickson, S., and Anderson, D. (1987) A small viral RNA is required for in vitro packaging of bacteriophage ϕ29 DNA. *Science 236*, 690–694.
6. Meijer, W. J. J., Horcajadas, J. A., and Salas, M. (2001) ϕ29 Family of Phages. *Microbiol. Mol. Biol. Rev. 65*, 261–287.
7. Bailey, S., Wichitwechkarn, J., Johnson, D., Reilly, B. E., Anderson, D. L., and Bodley, J. W. (1990) Phylogenetic Analysis and Secondary Structure of *Bacillus subtilis* Bacteriophage RNA Required for DNA Packaging. *J. Biol. Chem. 265*, 22365–22370.
8. Zhang, F., and Anderson, D. (1998) In vitro selection of bacteriophage ϕ29 prohead RNA aptamers for prohead binding. *J. Biol. Chem. 273*, 2947–2953.
9. Chen, C., Zhang, C., and Guo, P. (1999) Sequence requirements for hand-in-hand interaction in formation of RNA dimers and hexamers to gear ϕ29 DNA translocation motor. *RNA 5*, 805–818.
10. Morais, M., Koti, J., Bowman, V., Reyes-Aldrete, E., Anderson, D. L., and Rossman, M. G. (2008) Defining molecular and domain boundaries in the bacteriophage ϕ29 DNA packaging motor. *Structure 16*, 1267–1274.
11. Reid, R. J. D., Bodley, J. W., and Anderson, D. (1994) Characterization of the Prohead-pRNA Interaction of Bacteriophage ϕ29. *J. Biol. Chem. 269*, 5157–5162.
12. Xiao, F., Moll, W.-D., Guo, S., and Guo, P. (2005) Binding of pRNA to the N-terminal 14 amino acids of connector protein to bacteriophage ϕ29. *Nucleic Acids Res. 33*, 2640–2649.
13. Garver, K., and Guo, P. (1997) Boundary of pRNA functional domains and minimum pRNA sequence requirement for specific connector binding and DNA packaging of phage ϕ29. *RNA 3*, 1068–1079.

14. Guasch, A., Pous, J., Ibarra, B., Gomis-Ruth, F. X., Valpuesta, J. M., Sousa, N., Carrascosa, J. L., and Coll, M. (2002) Detailed Architecture of a DNA Translocating Machine: The High-Resolution Structure of the Bacteriophage ϕ29 Connector Particle. *J. Mol. Biol. 315*, 663–676.
15. Reid, R. J., Bodley, J. W., and Anderson, D. (1994) Identification of Bacteriophage ϕ29 Prohead RNA Domains Necessary for in vitro DNA-gp3 Packaging. *J. Biol. Chem. 269*, 9084–9089.
16. Kitamura, A., Jardine, P. J., Anderson, D. L., Grimes, S., and Matsuo, H. (2008) Analysis of intermolecular base pair formation of prohead RNA of the phage ϕ29 DNA packaging motor using NMR spectroscopy. *Nucleic Acids Res. 36*, 839–848.
17. Trottier, M., Mat-Arip, Y., Zhang, C., Chen, C., Sheng, S., Shao, Z., and Guo, P. (2000) Probing the structure of monomers and dimers of the bacterial virus ϕ29 hexamer RNA complex by chemical modification. *RNA 6*, 1257–1266.
18. Parisien, M., and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature 452*, 51–55.
19. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res. 30*, 3497–3531.
20. Stombaugh, J., Zirbel, C. L., and Leontis, N. B. (2009) Frequency and Isostericity of RNA Base Pairs. *Nucleic Acids Res. 37*, 2294–2312.
21. Conn, G. L., Draper, D. E., Lattman, E. E., and Gittis, A. (1999) Crystal Structure of a Conserved Ribosomal Protein-RNA Complex. *Science 284*, 1171.
22. Zhang, H., Fountain, M., and Krugh, T. (2001) Structural Characterization of a Six-Nucleotide RNA Hairpin Loop Found in *Escherichia coli*, r(UUAAGU). *Biochemistry 40*, 9879–9886.
23. Fountain, M. A., Serra, M. J., Krugh, T. R., and Turner, D. H. (1996) Structural Features of a Six-Nucleotide RNA Hairpin Loop found in Ribosomal RNA. *Biochemistry 35*, 6539–6548.
24. Huang, S., Wang, Y.-X., and Draper, D. E. (1996) Structure of a Hexanucleotide RNA Hairpin Loop Conserved in Ribosomal RNAs. *J. Mol. Biol. 258*, 308–321.
25. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å. *Science 289*, 905–919.
26. Wimberly, B. T., Broderson, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature 407*, 327–339.
27. Selmer, M., Dunham, C. M., Murphy, F. V., Weizlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R., and Ramakrishnan, V. (2006) Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science 313*, 1935.
28. Morais, M. C., Tao, Y., Olson, N. H., Grimes, S., Jardine, P., Anderson, D. L., Baker, T. S., and Rossman, M. G. (2001) Cryoelectron-Microscopy Image Reconstruction of Symmetry Mismatches in Bacteriophage ϕ29. *J. Struct. Biol. 135*, 38–46.
29. Morais, M. C., Kanamaru, S., Badasso, M. O., Kotis, S. J., Owen, B. A., McMurray, C. T., Anderson, D. L., and Rossman, M. G. (2003) Bacteriophage ϕ29 scaffolding protein gp7 before and after prohead assembly. *Nat. Struct. Mol. Biol. 10*, 572–576.
30. Simpson, A. A., Tao, Y., Leiman, P. G., Badasso, M. O., He, Y., Jardine, P. J., Olson, N. H., Morais, M. C., Grimes, S., Anderson, D. L., Baker, T. L., and Rossman, M. G. (2000) Structure of the Bacteriophage ϕ29 DNA Packaging Motor. *Nature 408*, 745–750.
31. Morais, M. C., Choi, K. H., Kotis, J. S., Chipman, P. R., Anderson, D. L., and Rossman, M. G. (2005) Conservation of the capsid structure in tailed dsDNA bacteriophages: The pseudoatomic structure of ϕ29. *Mol. Cell 18*, 149–159.
32. Gausch, A., Pous, J., Ibarra, B., Gomis-Ruth, F. X., Valpuesta, J. M., Sousa, N., Carrascosa, J. L., and Coll, M. (2002) Detailed Architecture of a DNA Translocating Machine: The High Resolution Crystal Structure of the Bacteriophage ϕ29 Connector Particle. *J. Mol. Biol. 315*, 663–676.
33. Xiang, Y., Leiman, P. G., Li, L., Grimes, S., Anderson, D. L., and Rossman, M. G. (2009) Crystallographic insights into the autocatalytic assembly mechansim of a bacteriophage tail spike. *Mol. Cell 34*, 375–386.
34. Xiang, M., Morais, M. C., Cohen, D. N., Bowman, V. D., Anderson, D. L., and Rossman, M. G. (2008) Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage ϕ29 tail. *Proc. Natl. Acad. Sci. U.S.A. 105*, 9552–9557.
35. Zhang, F., Lemieux, S., Wu, X., St.-Arnaud, D., McMurray, C., Major, F., and Anderson, D. (1998) Function of Hexameric RNA in Packaging of Bacteriophage ϕ29 DNA In Vitro. *Mol. Cell 2*, 141–147.
36. Hoeprich, S., and Guo, P. (2002) Computer Modeling of Three-dimensional Structure of DNA-packaging RNA (pRNA) Monomer, Dimer, and Hexamer of ϕ29 DNA Packaging Motor. *J. Biol. Chem. 277*, 20794–20803.
37. McDowell, J. A., and Turner, D. H. (1996) Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of $(rGAGGUCUC)_2$ by Two-Dimensional NMR and Simulated Annealing. *Biochemistry 35*, 14077–14089.
38. Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr. (2000) Nucleic Acids Structures, Properties and Functions, University Science Books, Sausalito, CA.
39. Richardson, J. S., Schneider, B., Murray, L. W., Kapral, G. J., Immormino, R. M., Headd, J. J., Richardson, D. C., Ham, D., Hershkovits, E., Williams, L. D., Keating, K. S., Pyle, A. M., Micallef, D., Westbrook, J., and Berman, H. M. (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA 14*, 465–481.
40. Tolbert, B. S., Kennedy, S. D., Schroeder, S. J., Krugh, T. R., and Turner, D. H. (2006) NMR Structures of $(rGCUGAGGCU)_2$ and $(rGCGGAUGCU)_2$: Probing the Structural Features That Shape the Thermodynamic Stability of GA Pairs. *Biochemistry 45*, 1511–1522.
41. Goddard, T. D., and Kneller, D. G. (2006) SPARKY 3, University of California, San Francisco.
42. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kuntsleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Crystallogr. D54*, 905–921.
43. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA.
44. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science 244*, 48–52.
45. 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.
46. 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.
47. Wang, J., Cieplak, P., and Kollman, P. A. (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem. 21*, 1049–1074.
48. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *J. Comput. Chem. 25*, 1157–1174.
49. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986) An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem. 7*, 230–252.
50. Williams, D. J., and Hall, K. B. (1996) Thermodynamic Comparison of the Salt Dependence of Natural RNA Hairpins and RNA Hairpins with Non-Nucleotide Spacers. *Biochemistry 35*, 14665–14670.
51. Vieregg, J., Cheng, W., Bustamante, C., and Tinoco, I., Jr. (2007) Measurement of the effect of monovalent cations on RNA hairpin stability. *J. Am. Chem. Soc. 129*, 14966–14973.
52. Lukavsky, P. (2007) Basic Principles of RNA NMR Spectroscopy. In Structure and Biophysics: New Technologies for Current Challenges in Biology and Beyond, Springer, Dordrecht, The Netherlands.
53. Roberts, G. C. K. (1993) NMR of macromolecules: A practical approach, Oxford University Press, New York.
54. Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry 39*, 9257–9274.
55. Santa Lucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) Stabilities of consecutive A·C, C·C, G·G, U·C, and U·U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U·U and $C{\cdot}C^+$ pairs. *Biochemistry 30*, 8242–8251.
56. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.
57. Gutell, R. R., Cannone, J. J., Konings, D., and Gautheret, D. (2000) Predicting U-turns in Ribosomal RNA with Comparative Sequence Analysis. *J. Mol. Biol. 300*, 791–803.

58. Jucker, F. M., and Pardi, A. (1995) GNRA tetraloops make a U-turn. *RNA 1*, 219–222.
59. Ashrah, S. S., Ansari, G., Guenther, R., Sochacka, E., Maliewicz, A., and Agris, P. F. (1999) The uridine in "U-turn": Contributions to tRNA-ribosomal binding. *RNA 5*, 503–511.
60. Quigley, G. J., and Rich, A. (1976) Structural domains of transfer RNA molecules. *Science 194*, 796–806.
61. Yildirim, I., and Turner, D. H. (2005) RNA CHallenges for Computational Chemists. *Biochemistry 44*, 13225–13234.
62. Yildirim, I., Stern, H. A., Sponer, J., Spackova, N., and Turner, D. H. (2009) Effects of Restrained Sampling Space and Nonplanar Amino Groups on free-Energy Predictions for NRA with IMino and Sheared Tandem GA Base Pairs Flanked by a GC, CG, iGiC, or iCiG Base Pairs. *J. Chem. Theory Comput. 5*, 2088–2100.
63. Zhang, H., Fountain, M., and Krugh, T. (2001) Structural characterization of a six-nucleotide RNA hairpin found in *Escherichia coli*, r(UUAAGU). *Biochemistry 40*, 9879–9886.
64. Gao, Y.-G., Selmer, M., Dunham, C. M., Weixlbaumer, A., Kelley, A. C., and Ramakrishnan, V. (2009) The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science 326*, 694–699.
65. Blaha, G., Stanley, R. E., and Steitz, T. A. (2009) Formation of the First Peptide Bond: The Structure of EF-P Bound to the 70S Ribosome. *Science 325*, 966–970.