Structure and Dynamics of the Tetra-A Loop and (A-A)-U Sequence Motif within the Coliphage GA Replicase RNA Operator

Andrew T. Chang, Michelle Tran, and Edward P. Nikonowicz*

Department of BioSciences, Rice University, Houston, TX 77251-1892

This work was supported by National Science Foundation grant 1412864 to E.P.N.

*Address correspondence to this author.

edn@rice.edu

fax (713) 348-5154

Key Words: RNA hairpin, protein-RNA interaction, bulge, tetraloop, RNA motif

Abbreviations: EDTA, ethylene-diamine-tetra-acetic acid; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; 2D, two-dimensional; 3D, three-dimensional; HSQC, heteronuclear single quantum coherence

Abstract

The three-dimensional structure of an RNA hairpin containing the RNA operator binding site for bacteriophage GA coat protein is presented. The phage GA operator contains the asymmetric (A-A)-U sequence motif and is capped by a four-adenine (tetra-A) loop. The uridine of the (A-A)-U motif preferentially pairs with the 5' proximal cross-strand adenine and the 3' proximal adenine stacks into the helix. The tetra-A loop is well-ordered with adenine residues 2-4 forming a 3' stack. This loop conformation stands in contrast to the structure of the 5'-AUUA loop of the related phage MS2 operator in which residues 1 and 2 form a 5' stack. The context dependence of the (A-A)-U sequence motif conformation was examined using structures of 74 unique occurrences from the PDB. The motif almost always has one adenine bulged and the other adenine adopting an A-U base pair. In the case where the (A-A)-U motif is flanked by only one Watson-Crick base pair, the adenine adjacent to the flanking base pair tends to bulge. 80% of motifs with a 3' flanking pair have a 3' bulged adenine and 84% of motifs with a 5' flanking pair have a 5' bulged adenine. The frequency of 3' and 5' proximal adenines bulging is 33% and 67%, respectively, when the (A-A)-U motif is flanked by base pairs on both sides. Although a 3' flanking cytidine correlates (88%) with bulging of the 5' proximal adenine, no strict dependence on flanking nucleotide identity was identified for the 5' side.

Introduction

RNA hairpins, or stem-loops, are ubiquitous secondary structure motifs that function as regulators of gene expression, scaffold elements for the assembly of large multi-domain RNA molecules, and sites for protein recognition. The terminal loops of hairpins can adopt well-organized folds of limited flexibility such as the tetra-loop motif[1-3] or they can be multi-conformational and only transiently occupy a single conformational state. The stems of RNA hairpins also exhibit structural variety, ranging from fully base paired helices to helices interrupted by non-canonical features such as bulged nucleotides or internal loops. This structural and dynamic diversity is central to the multitude of functional roles of RNA hairpins in cells.

The family of bacteriophage *Leviviridae* are single stranded positive sense RNA phage that infect *Escherichia coli* by adsorbing to conjugative F pili. Bacteriophage GA and MS2 are members of this family that have a bulge-containing 21 nucleotide stem-loop capped by a four-nucleotide loop in their RNA genomes which is the binding site for the phage capsid protein. This RNA hairpin, known as the operator, encompasses the ribosome binding site and start codon of the replicase gene so that capsid protein binding to the operator blocks translation of the replicase gene and nucleates the encapsidation of the RNA genome into phage particles[4, 5]. The phage GA operator hairpin has an unpaired adenine in the stem and a four-adenine terminal loop (Figure 1). Mutagenesis studies have shown that restrictions on the operator nucleotide sequence for GA coat protein binding are limited, but include an adenine and a purine at loop positions -4 and -7, respectively, and a bulged purine at either position -10 or -11 (Figure 1) [6]. The closely related MS2 operator hairpin has a bulged adenine at position -10 in the stem and a terminal loop of sequence 5'-AUUA. However, the coat protein binds hairpins containing a single bulged

adenine at position -10 three-fold tighter than hairpins containing the bulged adenine at position -11[6]. Although the amino acid sequences of GA and MS2 coat proteins are 67% identical, the RNA binding site requirements for the MS2 coat protein are more restrictive. MS2 coat protein binding additionally requires a pyrimidine at position -5 in the loop and the bulged purine must be present only at position -10 in the stem[7].

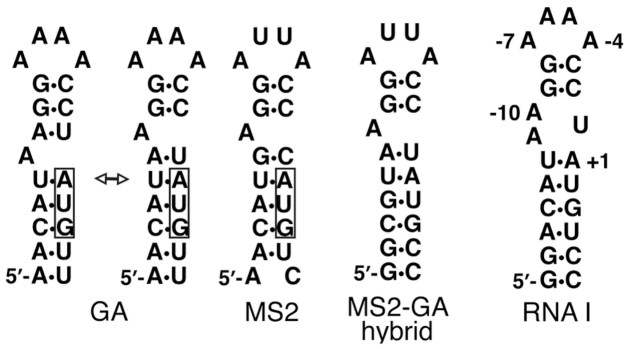

Figure 1. Sequence and secondary structures of the RNA operator from phage GA, phage MS2, an RNA hairpin with the stem of the GA operator and the loop of the MS2 operator, and the RNA molecule used in this study, RNA I. Residues in RNA I are numbered relative to the phage GA replicase gene start codon (boxed). Each secondary structure contains the (A-A)-U secondary structure motif. The (A-A)-U motif in the GA-MS2 hybrid molecule was shown to form a base triple through protonation of the A_{-10} N1.

The majority of RNA hairpins whose three-dimensional structures are known have terminal loops that contain seven or fewer nucleotides and those that contain four nucleotides (the tetraloops) are the most abundant[8]. Three families of these tetraloops, UNCG, GNRA, and CUUG, are especially common[9] and are thermodynamically very stable, owing to base-base and base-ribose hydrogen bonding and base stacking interactions. The structures of four-nucleotide loops with an adenine at the first position are less numerous but are distributed among different RNA types including rRNA, SAM-I and class II pre-Q1 riboswitches[10, 11], the binding site for the RNase III enzyme Rnt1p[12], the human Xist RNA A-repeat[13], and viral and phage RNAs[14-18]. A distinguishing feature of the phage GA loop is its all-adenine sequence, 5'-AAAA. Few structures of four-nucleotide all-adenine (tetra-A) loops have been reported[11, 19-22] and all are closed by base pairs A-U, C-G, or U-A (5' and 3' end,

respectively). Although the conformations of these tetra-A loops are varied, there is a tendency for adenines at position 1 and 4 to hydrogen bond and for adenines 2 and 3 to stack.

The bulge in the stem of the phage GA operator hairpin contains the common RNA sequence motif (A-A)-U, two adenine nucleotides opposite a single uridine nucleotide. In addition to the phage GA capsid protein binding site, the (A-A)-U motif is a central functional element in other RNA molecules such as the bacterial 16S rRNA binding site for ribosomal protein S8[23, 24], the recognition signal for dynein motor-based mRNA transport machinery of *Drosophila melanogaster*[25], the sensing domain of the *Vibrio cholerae gcvT* cooperative glycine riboswitch[26], and the influenza A virus RNA promoter[27, 28]. A variant of the (A-A)-U motif that contains pseudouridine, ψ, in place of uridine is present in the branch-point helix of the 5' splice site that forms as part of the first cleavage step of nuclear pre–mRNA splicing[29]. Interestingly, the conformations of the U- and ψ–containing motifs are different. The motif hydrogen bond network is different in each of these examples, suggesting a possible sequence context dependence to the conformation of the (A-A)-U motif.

We have used heteronuclear NMR spectroscopy to study the solution structure and dynamics of the RNA binding site for bacteriophage GA coat protein (Figure 1). In contrast to the loop sequence 5'-$A_{-7}U_{-6}U_{-5}A_{-4}$-3' of the MS2 operator hairpin in which the 5'-proximal adenine base stacks on the loop-closing base pair, the A_{-7} base of the GA hairpin exhibits motions on the intermediate timescale and appears disordered. Residues A_{-6}, A_{-5}, and A_{-4} of the loop form a well-ordered 3' stack that is contiguous with the closing base pair of the loop. The secondary structure of the (A-A)-U bulge in the stem is constrained by NOE data and consistent with relaxation data that indicate the 3' proximal adenine of the bulge is flexible and remains unpaired. Although the 5'-AAAA-3' of the phage GA operator increases the conformational

disorder and dynamics of the (A-A)-U motif relative to the 5'-AUUA-3' loop[15], the secondary structure of the bulge favors 5'-proximal adenine-uridine pairing.

Materials and Methods

All enzymes were purchased (Sigma) with the exception of T7 RNA polymerase which was prepared as described[30]. Deoxyribonuclease I Type II, pyruvate kinase, adenylate kinase, and nucleotide monophosphate kinase were obtained as powders and dissolved into 15% glycerol, 1 mM dithiothreitol and 10 mM Tris-HCl, pH 7.4 and stored at -20 °C. Guanylate kinase and nuclease P1 were obtained as solutions and stored at -20 °C. Unlabeled 5' nucleoside triphosphates (Sigma), phosphoenolpyruvate (potassium salt) (Bachem), and 99% [^{15}N] ammonium sulfate, 99% [$^{13}C_6$]-glucose, and 99% [$^{13}C_1$]-C2 sodium acetate (Cambridge Isotope Labs) were obtained as powders.

Preparation of RNA Samples. RNA I, (Figure 1), was prepared by in vitro transcription using T7 RNA polymerase and a synthetic DNA template[31]. Unlabeled RNA molecules were prepared from 12 ml transcription reactions using 4 mM 5'-NTPs. Isotopically enriched RNA molecules were prepared from 10 ml transcription reactions using 3 mM ^{15}N-enriched, ^{13}C /^{15}N-enriched, fractionally ^{2}H-enriched, and fractionally ^{13}C-enriched 5'-NTPs as described[32, 33]. Relaxation data were recorded using a $C_{+5}U$ variant of RNA I that reduced 1' overlap. The RNA molecules were purified (20% w/v preparative PAGE), electroeluted (Schleicher & Schuell), and ethanol precipitated. The RNA was dissolved in 1.0 M NaCl, 20 mM potassium phosphate, pH 6.8, and 2.0 mM EDTA and dialyzed extensively against 10 mM NaCl, 10 mM potassium phosphate, pH 6.8, and 0.05 mM EDTA using a Centricon-3 concentrator (Amicon Inc.). The samples were diluted with buffer to a volume of 0.32 ml and lyophilized to powders. For experiments involving the non-exchangeable protons, the samples were exchanged twice from

99.9% D_2O and dissolved in 0.32 ml of 99.96% D_2O. For experiments involving the exchangeable protons, the samples were dissolved in 0.32 ml of 90% H_2O/10% D_2O. The samples contained 100-130 A_{260} O.D. units in 0.32 ml (1.6-2.0 mM).

NMR Spectroscopy. Spectra were acquired on Varian Inova 500 MHz (^{1}H-[^{13}C, ^{15}N, ^{31}P] probe) and 600 MHz (^{1}H-[^{13}C, ^{15}N] cryoprobe) and Bruker AMX 500 MHz (^{1}H-[^{13}C, ^{15}N] probe) spectrometers and NMR spectra were processed and analyzed using Felix 2007 (Felix NMR Inc., San Diego, CA).

Two-dimensional (2D) ^{13}C-^{1}H HSQC spectra were collected to identify ^{13}C-^{1}H chemical shift correlations. Sugar spin systems were assigned using 3D HCCH-TOCSY (8 ms and 24 ms DIPSI-3 spin lock) experiments collected in D_2O. 2D HCN experiments were used to identify intra-residue base-ribose correlations. Pyrimidine C2 and C4 resonances were assigned from H6-C2 and H5-C4 correlations using 2D H(CN)C and 2D CCH-COSY experiments[34, 35]. 2D ^{15}N-^{1}H HSQC spectra optimized for 2-bond HN couplings were collected to identify purine N7 and adenine N1 and N3 resonances. Sequential assignments and distance constraints for the non-exchangeable resonances were derived at 25 °C from 2D ^{1}H -^{1}H NOESY spectra (t_m = 90, 180, and 360 ms) and 3D ^{13}C-edited NOESY spectra (t_m = 160 and 360 ms). Assignments and distance constraints for the exchangeable resonances were derived at 12 °C from 2D NOESY spectra (t_m = 160 and 360 ms) acquired in 90% ^1H_2O. The ^{31}P resonances were assigned using a ^{31}P-^{1}H hetero-TOCSY-NOESY spectrum (τ_{SL}=65 ms and τ_{NOE} =360 ms). $^3J_{H-H}$ and $^3J_{P-H}$ coupling constants were estimated using DQF-COSY and ^{31}P-^{1}H experiments, respectively. $^3J_{C-P}$ coupling constants also were estimated using the CECT-HCP experiment[36].

Distance and torsion angle constraints. Interproton distances were estimated from cross peak intensities in 2D NOESY and 3D ^{13}C-edited NOESY spectra. The covalently fixed

pyrimidine H5-H6 distance (≈2.4 Å) and the conformationally restricted sugar H1'-H2' distance (2.8-3.0 Å) give rise to the most intense peaks in the NOESY spectra and were used as a reference to assign upper bound distances to proton pairs based on the relative intensities of the corresponding cross peaks (qualitatively classified as strong, medium, weak, or very weak in the 90 and 160 ms mixing time spectra). None of the cross peaks are as intense as the pyrmidine H5-H6 peaks, thus upper bound distance constraints were set to 3.2, 4.2, 5.2, or 6.2 Å. Cross peaks observed only at mixing time 360 ms were classified as extremely weak and given 7.2 Å upper bound distance constraints to account for the possibility of spin diffusion. All distance constraints were given lower bounds of 1.8 Å. Only the intra-residue sugar-to-sugar constraints involving H5' and H5" resonances included in the calculations are considered conformationally restrictive. Distance constraints involving exchangeable protons were estimated from 360 ms mixing time NOESY spectra and were classified as medium, weak, very weak, or extremely weak.

Watson-Crick base pairs were identified by observation of a significantly downfield shifted NH or NH2 proton resonance and the observation of strong G-C NH–NH_2 or A-U H2–NH NOEs and by the chemical shifts of non-protonated base ^{15}N and ^{13}C carbonyl resonances. Hydrogen bonds were introduced as distance restraints of 2.9 ± 0.3 Å between donor and acceptor heavy atoms and 2.0 ± 0.2 Å between acceptor and hydrogen atoms.

Ribose ring pucker and backbone dihedral constraints were derived from $^3J_{HH}$, $^3J_{HP}$, and $^3J_{CP}$ couplings[36]. Residues with $^3J_{H1'-H2'} < 5$ Hz and C3' resonances between 70-74 ppm were constrained to C3'-*endo*. Ribose rings with $^3J_{H1'-H2'}$≈5 Hz and with C3' and C4' resonances between 74-76 and 84-86 ppm, respectively, were left unconstrained. The angle δ was constrained as 85° ± 30° and 160° ± 30° for C3'-*endo* and C2'-*endo* sugars respectively. For

residues -16 to -12, -10, -9, -3, -2, and 1-6, γ was constrained to the gauche+ conformation (60 ± 20°)[36]. γ was left unconstrained for the loop and bulge residues. Dihedral angle restraints for the β and ε torsion angles were derived from $^{3}J_{P\text{-}H5'}$, $^{3}J_{P\text{-}H5''}$, and $^{3}J_{P\text{-}H3'}$ couplings estimated in 2D ^{31}P-^{1}H HetCor (using a ≈75% random fractionally deuterated RNA hairpin) spectra and $^{3}J_{P\text{-}C2'/P\text{-}C4'}$ couplings measured in 2D ctHSQC spin-echo difference spectra. For stem residues, β was constrained to the *trans* conformation (180 ± 20°) if $^{3}J_{P\text{-}C4'}$ was > 5 Hz. ε was constrained to the *trans* conformation (-150 ± 20°) for residues with $^{3}J_{P\text{-}C2'}$ < 5 Hz and $^{3}J_{P\text{-}C4'}$ > 5 Hz. α and ζ were constrained to -65 ± 20° for the stem residues -16 to -12 and 1-6. Because a down-field shifted ^{31}P resonance is associated with the *trans* conformation of α or ζ and because no such shift is observed for any of the ^{31}P resonances in the RNA molecules, α or ζ were loosely constrained to exclude the *trans* conformation (0 ± 120˚) for residues -11 to -1. Although all base 6/8-1' intra-residue NOE cross peak intensities support the *anti* conformation about the glycosidic bond, no dihedral angle constraints were used for the angle χ.

^{1}H-^{13}C residual dipolar coupling constants (RDCs) were determined from the measured frequency difference between corresponding proton doublets in HSQC spectra acquired for isotropic and Pf1 phage-aligned samples. RDC values from base CH and ribose 1' CH vectors were obtained in this way. The axial and rhombic terms were determined within XPLOR-NIH using an extensive grid search[37], and yielded values of D_{aH}=-12.085 and R_{hH}=0.336.

Relaxation rates $R_C(C_z)$ and $R_C(C_{x,y})$, which yield T_1 and $T_{1\rho}$ values, respectively, were measured at 11.74 T (500 MHz ^{1}H) and 14.09 T (600 MHz ^{1}H) [38]. All experiments were recorded in D_2O at 25 ˚C. Recovery delays of 2.6 s were used between scans in the T_1 and $T_{1\rho}$ experiments and 180° ^{1}H pulses were applied during the ^{13}C relaxation periods to reduce the effects of cross-correlation. At 500 MHz, longitudinal (T_1) experiments were recorded with

inversion-recovery delays 20, 45, 100, 200, 300, 400, 500, 600, and 800 ms. Transverse ($T_{1\rho}$) experiments were separately optimized for ribose C1', C4', purine C8, and adenine C2 resonances and employed a ^{13}C spin lock (γB_2=3846 Hz). In these experiments, no C2 or C8 resonance was more than 600 Hz from the carrier frequency, which corresponds to a minimum effective tip angle of 81˚. $T_{1\rho}$ experiments were acquired with relaxation delays 4, 8, 16, 24, 36, 48, 64, and 88 ms. Five replicates of each of the T_1 and $T_{1\rho}$ data sets were collected. For rate measurements at 600 MHz, T_1 experiments had inversion-recovery delays 20 (x3), 100 (x3), 200, 300 (x3), 400, 500 (x3), 700, and 900 ms. $T_{1\rho}$ experiments had relaxation delays 4 (x3), 8 (x3), 12, 16 (x3), 24, 32 (x3), 40, 48, and 56 ms.

Structure calculations and refinement. An initial set of structures was calculated using a shortened version of the simulated annealing protocol (described below). A list of all proton pairs in the calculated structures closer than 5.0 Å (representing expected NOEs) was compared to the list of constraints. The NOESY spectra were then re-examined for predicted NOEs absent from the constraint list. In some cases, this allowed the unambiguous assignment of previously unidentified NOEs, but, in other cases, the predicted NOEs were obscured due to spectral overlap.

Structure refinement was carried out with simulated annealing and restrained molecular dynamics (rMD) calculations using XPLOR-NIH v2.44 with the RNA ff1 force field and potentials implemented[37, 39]. Starting RNA coordinates were generated as a single stranded oligoribonucleotide with C3'-*endo* ring puckers for all ribose moieties using 3DNA[40]. The structure calculations were performed in two stages. Beginning with the energy minimized starting coordinates, 50 structures were generated using 20 ps of rMD or 10000 steps, whichever is less, at 1200K with hydrogen bond, NOE-derived distance and base-pairing restraints. The

system then was cooled to 25 K over 0.2 ps of rMD or 10000 steps, whichever is less. During this stage, RDC constraints and repulsive van der Waals forces were introduced into the system and the SANI force constraint used for RDCs was gradually increased from 0.010 kcal mol^{-1} Hz^{-2} to 1.000 kcal mol^{-1} Hz^{-2}. Other force constants used for the calculations were increased—from 2 kcal mol^{-1} $Å^{-2}$ to 30 kcal mol^{-1} $Å^{-2}$ for the NOE and from 10 kcal mol^{-1} rad^{-2} to 100 kcal mol^{-1} rad^{-2} for the dihedral angle constraints. Each structure was then minimized with constraints at the end of the rMD. Eight structures were selected for the final refinement. The criteria for final structure selection included lowest energies, fewest constraint violations, and fewest predicted unobserved NOEs (^{1}H pairs less than 3.5 Å apart, but no corresponding cross peak in the NOE spectra). A second round of rMD was performed on these structures using protocols similar to those used in the first round of structure calculation. The major difference was the starting temperature of 300 K followed by cooling to 25 K over 28 ps of rMD. Eight refined structures for each model were analyzed using XPLOR-NIH and Pymol and UCSF Chimera.

Relaxation analysis. To avoid contributions from ^{13}C-^{13}C homonuclear relaxation, the T_1 and $T_{1\rho}$ experiments were acquired using fractionally ^{13}C enriched RNA samples. The adenine C2 and purine C8 positions were ≈85% ^{13}C enriched and the base C4, C5, and C6 positions were ≈10% enriched. Similarly, the C1' and C4' sites were ≈65% and 20% ^{13}C enriched, respectively, and the C2', C3', and C5' sites were <5% enriched. The ^{14}N quadrupolar contribution to relaxation of the ^{13}C nuclei was estimated to be 2% [41].

Cross peak volumes for purine C8 and adenine C2 resonances were measured from the T_1 and $T_{1\rho}$ spectra and were fit to a single exponential decay using the Levenburg-Marquardt nonlinear least squares fitting routine. Errors in the relaxation rates were calculated as standard

deviations from replicate experiments. Curve fitting was performed using the Relaxation Analysis Program[42].

Results

Spectral assignments. The non-exchangeable ^{1}H and ^{13}C resonances of the phage GA operator were assigned using standard heteronuclear methods [43, 44]. Most of the base and ribose resonances are resolved at 298 K and pH 6.8. The C_{-3} 1' resonance is broadened by chemical exchange and is not observed. The 1' and base 8 resonances of A_{-10} and A_{-7} also are broadened, indicative of intermediate exchange (Figure 2 and S1). The 23 ribose spin systems were identified using 3D HCCH-COSY and 3D HCCH-TOCSY experiments. Only the 2'-5' resonances of C_{-3} could not be assigned. Six of nine adenine intra-base H2-H8 correlations were identified via an HCCH-TOCSY experiment and the C2-H2 correlations of two adenine bases are broadened by chemical exchange. Intra-residue base-sugar correlations were identified from H6-N1, H8-N9, and H1'-N1/N9 cross peaks in 2D H(C)N spectra. All pyrimidine correlations and seven of 14 purine correlations were identified. The A_{-10}, G_{-9}, G_{-8}, A_{-7}, A_{-5}, and A_{-4} correlations were not observed. Sequential assignment of non-exchangeable resonances was accomplished using 2D NOESY and 3D ^{13}C-edited NOESY spectra. The sequential NOE

connectivities are continuous in the base-1' region (τ_{mix} = 180 ms) except at A_{-7} and C_{-2} (Figure 2). The sequence specific resonance assignments are listed in Table S1.

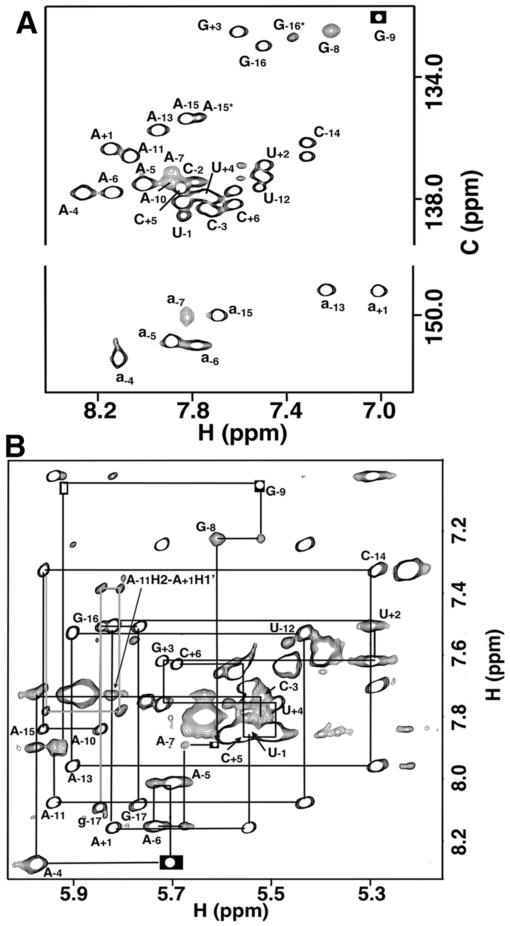

Figure 2. (A) Two-dimensional ^{13}C-^{1}H HSQC spectrum of the base C6/8 and C2 (lower case) regions of RNA I. The sequence-specific resonance assignments are shown. The a_{-7} C2H2 cross peak is significantly weakened by chemical exchange and the A_{-10} and A_{-11} C2H2 cross peaks are completely broadened by exchange. The G_{-9} C8H8 resonance is very broad and the A_{-7} and A_{-10} C8H8 resonances display moderate exchange broadening. (B) Sequential connectivities through the base-1' region of the 180 ms mixing time two-dimensional NOE spectrum. The gray lines trace the connectivities among a minor population at the 5' end. At this mixing time, the sequential connectivity is interrupted between steps A_{-10}-G_{-9}, G_{-8}-A_{-7}, and A_{-5}-A_{-4} (boxes). The complete set of chemical shifts are reported in Table S1.

The NH and NH_2 resonances were assigned using ^{1}H-^{1}H NOESY and HNCCH experiments. The NH resonances of the six neighboring base pairs of the lower stem yield NOE connectivities between each other. The G_{-9} and U_{-1} NH resonances were assigned based on ^{1}H and ^{15}N chemical shifts.

The inter-nucleotide phosphate ^{31}P resonances are clustered between -2.54 and -4.60 ppm and assignments were obtained using HCP or ^{31}P-^{1}H hetero-TOCSY-NOESY spectra [45]. Several H3'-P, H4'-P, H5'-P, and H5''-P cross peaks in the ^{31}P-^{1}H HetCor spectrum acquired using a random fractionally (≈75%) deuterated RNA molecule provided independent confirmation of the ^{31}P assignments. The fractional deuteration largely eliminated

^{1}H-^{1}H passive coupling, dramatically improved spectral quality, and facilitated extraction $^{3}J_{HP}$ coupling data.

Structure calculations and analysis. The structure of the phage GA operator was calculated using a restrained molecular dynamics routine starting from 50 sets of randomized coordinates. The simulated annealing and refinement calculations were performed with 311 conformationally restrictive NOE derived distance constraints, 180 torsional angle constraints,

Table 1. Summary of experimental distance and dihedral angle constraints and refinement statistics for phage GA operator RNA molecule.

Constraint		RNA I
NOE distance constraints		**311**
Intra-residue[a]		183
Inter-residue		128
Mean number per residue		13.5
NOE constraints by category		
Very strong	(1.8 – 2.5 Å)	11
Strong	(1.8 – 3.2 Å)	24
Medium	(1.8 – 4.2 Å)	133
Weak	(1.8 – 5.2 Å)	131
Very weak	(1.8 – 6.2 Å)	12
H-bond distance constraints		**21**
Dihedral angle constraints		**180**
Ribose ring[b]		54
Backbone		110
Base pairs		16
Mean number per residue		7.8
RDC constraints		**40**
Violations		**0**
Average distance constraints > 0.3 Å[c]		0
Average dihedral constraints > 0.5°[d]		0
Average RDC constraints > 1.5 Hz		0
RMSD from ideal geometry[e]		
Heavy Atoms (Å)		0.446

[a]Only conformationally restrictive constraints are included.

[b]Three torsion angles within each ribose ring were used to constrain the ring to either the C2'-*endo* or C3'-*endo* conformation.

[c]A distance violation of 0.3 Å corresponds to 5.0 kcal energy penalty.

[d]A dihedral angle violation of 0.5° corresponds to 0.05 kcal energy penalty.

[e]Calculated against the minimized average structure.

and 40 RDC constraints (Table 1) to produce eight converged structures (Figure 3). When the structures are arranged in order of increasing overall energy, the converged structures form a plateau with similarly low overall energies and constraint violation energies. The root mean square deviations (RMSDs) of the heavy atoms between the individual structures and the minimized mean structure is 0.446 Å. Structure statistics are summarized in Table 1.

Structure of the bulge and stem. The primary interactions of interest within the bulge region (U_{-12}–G_{-9} and C_{-2}–A_{+1}) are those that define the base pairing pattern of nucleotides A_{-11}, A_{-10}, and U_{-1}. A_{-10} and A_{-11} appear to have equal potential to base pair with U_{-1}, but only the A_{-11}-U_{-1} pairing is present in the converged structures. The U_{-1} N3 chemical shift of 156 ppm (confirmed by the U_{-1} N3-H5 correlation) is consistent with an A-U base pair, but rapid exchange of the U_{-1} NH and broad adenine H2 resonances prevent confirmation of either A_{-11} or A_{-10} as the sole U_{-1} base pair partner. Thus, no base pair constraints were introduced for residues U_{-1}, A_{-11}, and A_{-10} during the MD calculation of the

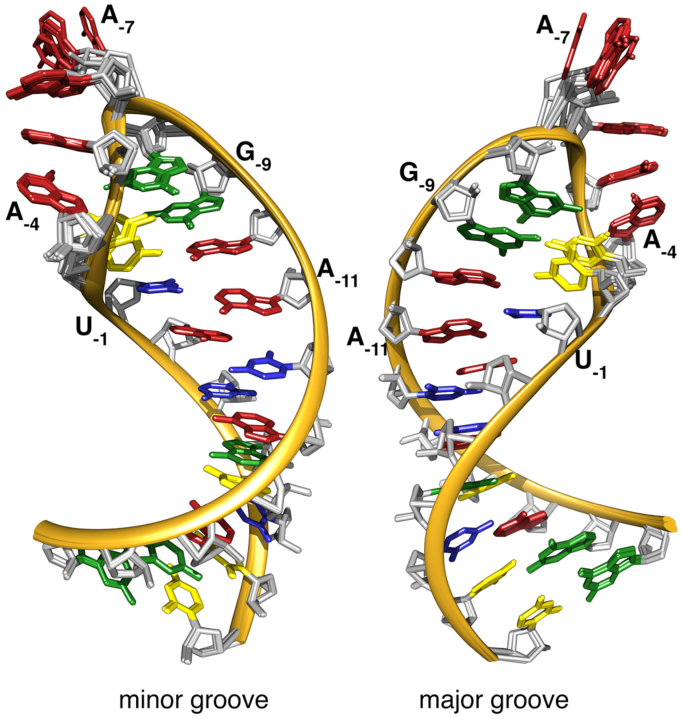

Figure 3. Superpositions of the eight converged structures of RNA I. Views are into the minor and major grooves at the center of the upper helices. Nucleotide bases are colored as: adenine (red), uridine (blue), guanine (green), and cytidine (yellow). The average RMSD between the individual structures and the average structure is 0.446 Å. The (A-A)-U sequence motif and adjacent regions are generally well defined. The disorder of the first nucleotide of the tetra-A loop, A_{-7}, reflects the lower number of constraints for this residue and is consistent with the $T_1\rho$ relaxation measurements.

structures. Cross strand NOEs identified involving the bulged nucleotides are A_{+1} H2–A_{-11} H1', A_{-11} H2-A_{+1} H1', and A_{+1} H2–A_{-11} H8 (τ_m=320 ms) which confirm the structural regularity of the stem up to the base of the bulge region but impose few constraints on the A_{-10} and U_{-1} conformations. However, sequential base-base and base-sugar NOEs in the bulge region, in particular those between U_{-1} H5 and H6 and A_{+1} H8, help to further define the bulge region structure and demonstrate the close base stacking between U_{-1} and A_{+1}. Thus, the NOE and RDC data establish important spatial limitations on the bulge region nucleotides and in the absence of base pair constraints, are sufficient to produce structures where A_{-11} and U_{-1} base pair. The A_{-11}-U_{-1} base pair configuration also is consistent with the intermediate time scale dynamics of the bulge region nucleotides. The base 2 and 8 and ribose 1' ^{13}C-^{1}H line widths of A_{-10} are significantly broadened by chemical exchange whereas only the base 2 resonance of A_{-11} is broad (Figure 2 and S1). Similarly, the base ^{1}H and ^{13}C resonance of U_{-1} do not display the effects of intermediate exchange. This pattern of resonance broadening points to the motional properties of A_{-10} as the primary source of chemical exchange effects on its own resonances and those of its

neighbors. Thus, although both A_{-10} and A_{-11} stack into the helix, the restricted flexibility of A_{-11} favors pairing with U_{-1} and is consistent with the calculated structures.

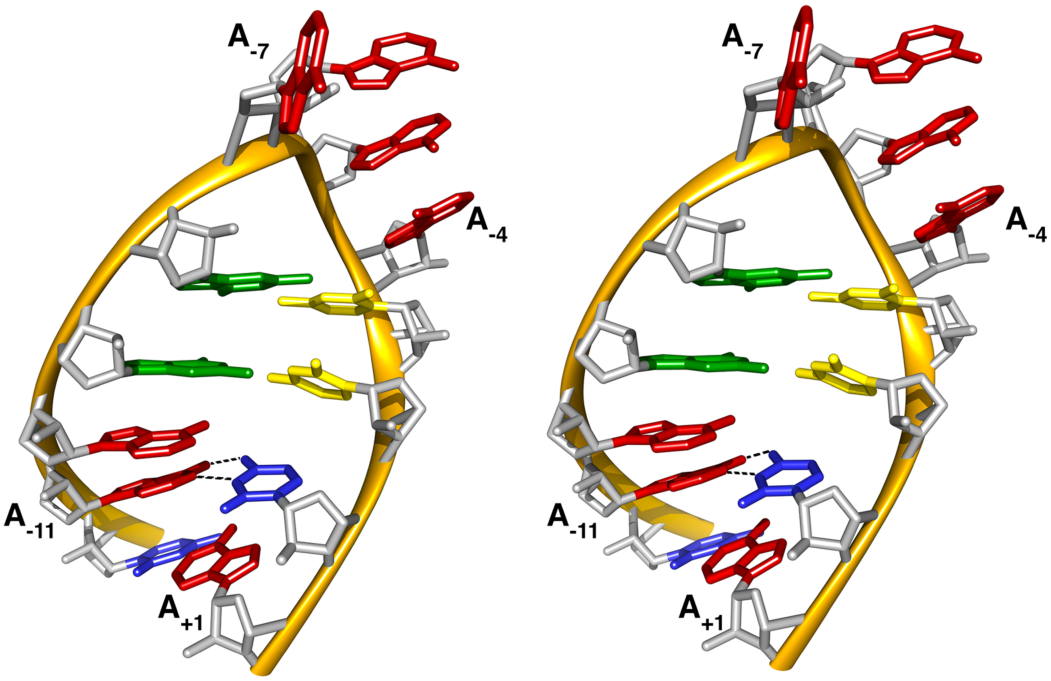

Figure 4. Stereoview of the upper portion of the phage GA operator. Nucleotide coloring is the same as in Figure 3 and hydrogen bonds between A_{-11} and U_{-1} are indicated by dotted black lines. The bases of A_{-6}-A_{-4} form a 3' stack with A_{-4} stacking above the pyrimidine ring of C_{-3}. A_{-11} forms a Watson-Crick pair with U_{-1} and the base of A_{-10} stacks into the helix between base pairs A_{-11}-U_{-1} and G_{-9}-C_{-2}.

Much of the sugar-phosphate backbone through the bulge and stem has A-form geometry (Figure 4). The ribose puckers of A_{-11} and A_{-10} are a mixture of C2'/C3'-*endo*; all other nucleotides in the stem and bulge have the C3'-*endo* conformation. Nucleotides A_{-10}, G_{-9}, and A_{+1} have $J_{P\text{-}C2'}$ and $J_{P\text{-}H3'}$ coupling constants that restrict the ε torsional angle to the *gauche*$^-$ conformation, but the coupling constants at the other internucleotide positions only exclude the *gauche*$^+$ conformation of the corresponding ε torsional angles. Small $^3J_{P\text{-}H5'/H5''}$ coupling constants measured for most stem and bulge nucleotides were used to restrict the β torsional angles to the *trans* conformation that is associated with A-form geometry. The β torsional angles of U_{-12}, A_{-11}, C_{-2}, and U_{-1} could not be determined due to the inability to clearly identify one or both P-H5'/H5"

cross peaks. The bulge region ^{31}P resonances are centered within the main cluster around -4.0 ppm, suggesting that none of the α- or ζ-torsional angles adopt the *trans* conformation. These torsion angles between nucleotides of the stem were constrained to exclude the *trans* conformation.

Structure of the loop. The average number of distance constraints per residue in the loop region (G_{-8} to C_{-3}) is less than the stem or bulge regions and can be largely attributed to the extremely weak A_{-7} H8. Many of the constraints are derived from sequential base-ribose NOEs but sequential base-base NOEs also are present in the loop except between A_{-7} and A_{-6}. A few constraints in the loop region are unusual and indicative of non-helical structure. The A_{-6} H2 resonance has cross peaks to the H1' resonances of G_{-8} and A_{-5} and the A_{-4} H2 resonance has a weak cross peak to the C_{-3} H5. A non-sequential i to i+2 NOE between A_{-6} H8 and A_{-4} H1' and nucleotides A_{-6} through A_{-4} have unusual $H2_i$–$H1'_{i+1}$ NOE cross peaks at long (320 ms) mixing time.

The sugar-phosphate backbone has non-A-form geometry at positions in the 5' region of the loop (Figure 4). Nucleotide A_{-6} adopts the C2'-*endo* conformation and nucleotides A_{-5} and A_{-4} adopt a mixture of C2'-*endo* and C3'-*endo* conformations. The ε torsional angles of nucleotides G_{-8}, A_{-7}, A_{-6}, and A_{-4} have the gauche$^-$ conformation but the $J_{P\text{-}C2'}$ and $J_{P\text{-}H3'}$ coupling constants of nucleotides A_{-5} and C_{-3} are consistent with either trans or *gauche*$^-$ ε conformations. The β torsional angles of A_{-7}–A_{-4} have the standard *trans* conformation and that of C_{-3} has the gauche conformation. The γ torsional angle of A_{-7} adopts the gauche$^+$ conformation and those of A_{-5} and C_{-3} adopt either the *trans* or *gauche*$^-$ conformations. No other γ torsional angles in the loop could

be determined. As with the bulge region, no constraints were imposed for torsion angles α and ζ even though the ^{31}P resonances of the loop region phosphates also are centered around -3.5 ppm.

The structure of the loop is remarkably well defined for residues A_{-6}-A_{-4} (Figure 3). Although the average RMSD of the loop is 0.287 Å, nucleotides A_{-6}, A_{-5} and A_{-4} are on the major groove side of the RNA helix above the C_{-3} ribose in all converged structures. The base of A_{-7} is restrained by relatively few NOEs. Its position varies but is consistently turned out towards the major grove of the upper helix and not stacked with G_{-8}. Although the G_{-8} imino proton is solvent accessible, the chemical shifts of the G_{-8} NH and C_{-3} NH_2 resonances support formation of the G_{-8}-C_{-3} base pair that closes the loop. The planarity of the G_{-8}-C_{-3} base pair is only slightly variable with a propeller twist up to 20 degrees.

Effect of pH on hairpin structure. The solution NMR study of another RNA hairpin containing the (A-A)-U bulge motif revealed that the 3' proximal adenine base (corresponding to A_{-10} of RNA I) was found to have an unusually high pKa of 6.1 [15]. The protonation of this adenine, indicated by an upfield shift of the base C2 resonance, suggested the N1 forms a non-standard hydrogen bond with the cross-strand uridine base [15]. Thus, the pH dependencies of the adenine C2 resonances were examined to identify residues in the bulge or loop regions that might be involved in unusual base-base interactions. Of the seven adenine C2 resonances between 152 and 156 ppm at pH 7.3, the A_{-4} C2 becomes broadened by chemical exchange below pH 6.7 and is unobservable below pH 5.2. An eighth C2 resonance begins to appear at 146 ppm, indicative of protonation, at pH 5.8, but intensifies only slightly by pH 4.8 and could not be uniquely assigned to A_{-11} or A_{-10}. Thus, although one adenine base has a slightly elevated pKa, the upper stem of the phage GA hairpin appears insufficiently stable to allow A_{-10} to form the tertiary contact.

Dynamics of the RNA hairpin. As indicated in HSQC spectra, some of the C1' and C8 resonances corresponding to nucleotides in the bulge and loop have line widths indicative of intermediate time scale ($t = 10^{-6}$–10^{-3} s) motions presumably due to chemical exchange. The half-height line widths of the G_{-9}, A_{-6}, A_{-5}, and A_{-4} C1' resonances are double those of stem nucleotides and the C1' resonances of A_{-10} and A_{-7} are broadened to the point that they are nearly unobservable in the HSQC spectrum (Figure S1). Similarly, within the base region of the spectrum, C8-H8 correlations for A_{-10}, G_{-9}, G_{-8}, and A_{-7} are broadened by chemical exchange.

The ^{13}C longitudinal and transverse relaxation rates, R_1 and $R_{1\rho}$, of the ribose C1', adenine C2, and purine C8 resonances were determined at 11.74 and 14.10 T (500 and 600 MHz ^1H) (Table S2). The longitudinal relaxation rates of the ribose C1' cluster between 1.8 to 2.5 s^{-1} at 500 MHz and between 1.3 to 1.8 s^{-1} at 600 MHz. At both field strengths, the most rapid rates are associated with nucleotides in the bulge and loop regions. This distribution is mirrored by the C2 and C8 nuclei which have longitudinal relaxation rates between 2.6 to 2.9 s^{-1} at 500 MHz and 2.1 to 2.7 s^{-1} at 600 MHz. The C1' transverse relaxation rates cluster between 18 and 26 s^{-1} at 500 MHz and between 20 and 29 s^{-1} at 600 MHz. Although the relaxation rates are relatively uniform throughout the hairpin, the A_{-11}, A_{-3}, and U_{-1} C1' relaxation rates are exceptionally rapid (>70 s^{-1}) as expected based on the relative broadness of the corresponding C1' peaks. Similarly, the C2 and C8 nuclei have relatively uniform relaxation rates—between 24 and 30 s^{-1} at 500 MHz and 20 and 24 s^{-1} at 600 MHz. Although none of the C8 nuclei have unusually fast or slow transverse relaxation rates, the average rate for the loop nucleotides is ~16% slower than that of the stem nucleotides, 21 s^{-1} to 25 s^{-1}, respectively, at 500 MHz.

Discussion

Structure and dynamics of the bulge. The coliphage GA and MS2 coat proteins bind as dimers to RNA hairpins containing single bulged adenine nucleotides in the stems and four nucleotides in the loops. These RNA secondary structures in complex with the coat protein regulate translation of the phage replicase genes and nucleate phage capsid assembly. The bulged adenine in the stem of the phage MS2 operator is flanked by G-C base pairs that restrict the bulge to position -10 and the upper stem to two base pairs (Figure 1). The bulge region nucleotides of the phage GA RNA hairpin, two adenines opposite a single uridine, constitute an (A–A)-U sequence motif that has the potential to form two different Watson-Crick A-U base pairs, resulting in a two or three base pair upper stem [6].

In the bulge region of the GA RNA hairpin, the base of A_{-10} is stacked into the helix and does not maintain any hydrogen bonds to the cross strand U_{-1} base (Figure 4), making it the most energetically favorable adenine in the bulge to turn out from the helix and interact with the coat protein. In binding studies, it was shown that a bulged adenine nucleotide at this position has the highest binding affinity to the GA coat protein [6]. Crystal structures of the phage MS2-RNA operator complex show the bulged adenine forms a hydrogen bond between Thr45 of one coat protein of the dimer and the N1 atom of A_{-10} [18, 46]. Additionally, the sugar-phosphate backbone at G_{-11} and A_{-10} faces the binding pocket to form a hydrogen bond between Lys61 and the phosphoryl oxygen and a water-mediated interaction between Ser47 and the G_{-11} 2'-OH.

Nucleotides in the bulge and upper stem exhibit dynamics approaching the intermediate time scale that leads to chemical exchange broadening of base and ribose resonances. The base 2 and 8 resonances of $A_{-10,}$ $G_{-9,}$ and G_{-8} are particularly broad in the HSQC spectrum (Figure 2) and indicate the instability of this region of the stem. Although the A_{-11} H2 peak was not observed, the base 8 and C1' resonances are present and have longitudinal and transverse relaxation times

similar to other lower stem nucleotides. Similarly, the base 5 and 6 resonances of C_{-3} and C_{-2} exhibit broadening, but to a lesser extent than their paired G_{-8} and G_{-9} bases. Interestingly, the mobility in the upper stem and bulge is not manifest when the tetra-A loop sequence is replaced with the loop nucleotides of phage MS2, U_{-6} and U_{-5}, to yield the sequence 5'-AUUA-3' [15]. This suggests the instability is a product of the structure of the loop.

Analysis of (A-A)-U motif structures. The (A-A)-U sequence motif is not uncommon and structures of other RNA molecules containing this motif could display possible correlations between flanking sequence identity or secondary structure and the tendency for the 5' or 3' proximal adenine to pair with the uridine or to bulge (stacked but unpaired or flipped out from the helix) (Figure 5). The program CoSSMos [8] was used to identify a total of 76 (A-A)-U

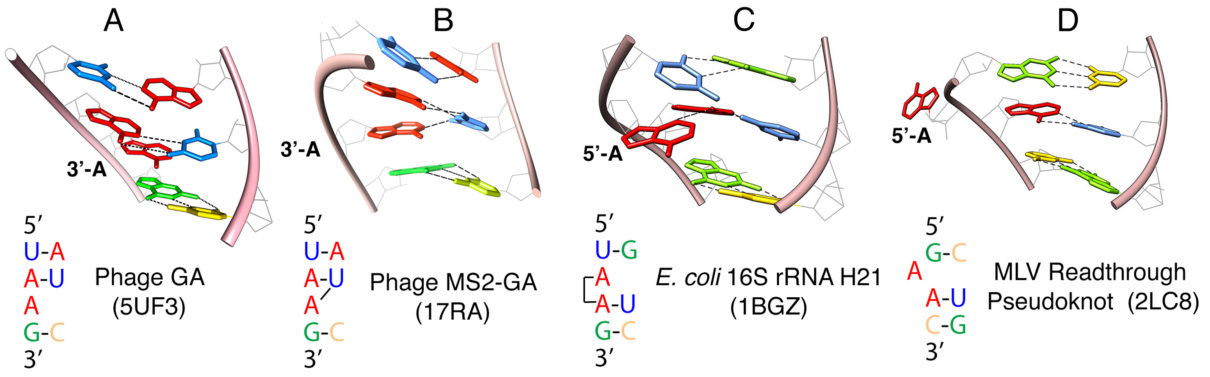

Figure 5. Comparison of (A-A)-U motif structures (PDB coordinates 5UF3 (this study), 17RA [15], 1BGZ [23], and 2LC8 [53]) showing the conformational variability of the motif. The four structures were determined using solution NMR and the unpaired adenine, 3'- or 5'-proximal, is labeled in each. The (A-A)-U motif flanking nucleotides are the same for A and B. In C, the U-A pair on the 5' side is replaced with U-G. In B, the N1 of the 3' proximal adenine of the motif has a pKa = 6.2 and protonates to form a hydrogen bond with the UO2 of the (A-A)-U motif. The A-to-G substitution in C appears conservative, but results in the 5' proximal adenine bulging and a base triple by hydrogen bonding with the 3' proximal adenine of the motif. In D, the 5' proximal adenine completely loops out of the helix.

sequence motifs extracted from 49 non-redundant protein data bank coordinate files representing high-resolution NMR and X-ray crystallographic structures. Statistical analysis of the structures (Supporting Information) supports trends between the conformations adopted by the motif and its

sequence and structural context. Overall, the 5'-A is bulged in 35 of the motifs and the 3'-A bulged in 41. However, of the 76 (A-A)-U motifs examined, 44 have base pairs flanking both ends of the motif whereas 32 of the motifs have a base pair flanking only one side of the motif. Among the latter group, the base pair appears adjacent to the 3'-A of the motif in 27 instances and adjacent to the 5' side of the motif only five times (Table S3).

Analysis of the group with a Watson-Crick base pair flanking only one side or the other of the (A-A)-U motif reveals a correlation between the position of the flanking base pair and the adenine that bulges. In 27 of the 32 motifs having a single flanking Watson-Crick base pair, the unpaired adenine is the adenine adjacent to the base pair. We also examined the group with two flanking base pairs for possible correlations between flanking nucleotide identity and preferential adenine bulging. The bulging of the 5'- or 3'-proximal adenine base was determined to be independent of the identity of the base flanking the 5' side of the motif. However, when a pyrimidine nucleotide flanks the 3' adenine of the motif, 90% of the time it is the 5' proximal adenine that bulges (16/18, Table S3C). In the 26 structures having a purine flanking the 3' adenine, however, there is no preference for the 5'- or 3'-proximal adenine base to bulge (13 in each case). Thus, although the tendency of a given adenine in the motif to pair or bulge displays some sequence context dependence when flanked 3' by a pyrimidine residue, additional factors affect the conformation that the motif adopts.

The motif structures were examined to identify the possible physical origins of the observed correlations. For the case in which the (A-A)-U motif is flanked on both sides by a base pair and a pyrimidine nucleotide is the residue adjacent to the 3' proximal adenine, the 3' proximal adenine preferentially pairs with the cross strand uridine. Favorable stacking interactions between the uridine base and the ring of the purine residue adjacent to the 5' side of

the uridine are optimal for the U-A (3' proximal adenine) pairing arrangement. The 3'-U:R-5' stacking appears to be the primary force behind the pairing arrangement. Neither pyrimidines flanking the 5' or 3' side of the uridine nor a purine 3' to the uridine can present a comparable stacking surface for the uridine base due to the curvature of the A-form helix (which is governed by the 3'-*endo* ring pucker of the ribose sugar). In the absence of a 3'–flanking Y-R pair, the dominant 3'-U:R-5' stacking force is lost and other factors combine that determine the configuration within the motif. For the case in which the motif is flanked by only one base pair, the adenine that is adjacent to that base pair preferentially bulges regardless of the identity of the nucleotide flanking the 3' proximal adenine. A structural basis for this preferred configuration is unclear, but the cross strand uridine and the adenine that pairs with it adopt reverse-Hoogsteen (70%) or Waston-Crick (30%) arrangements, albeit sometimes with less than perfect geometry. Interestingly, most examples involving the reverse-Hoogsteen base arrangement occur at sharp bends of the helix or at duplex-single strand junctions where the A-U pair creates a shorter and wider stacking surface than presented by the Watson-Crick base pair.

Structure and dynamics of the loop. Structure overlays clearly demonstrate that most of the loop residues converge to a single, well defined conformation (Figure 3). Residues A_{-6}, A_{-5}, and A_{-4} of the loop form a 3'-stack with the purine ring of A_{-4} above the cytidine base of the closing G-C base pair and slightly displaced towards the minor groove. These three loop nucleotides extend the helical architecture of the upper stem and the pattern of inter-base interactions is consistent with this arrangement. Conversely, the A_{-7} residue on the 5' side of the loop is minimally restricted and displays multiple conformations among the converged structures. The intrinsic flexibility, and consequent line broadening, of A_{-7} results in few structural constraints for this residue. Thus, the broad range of orientations adopted by A_{-7} in the converged structures may not

accurately reflect the true conformational state of the residue. The turn of the phosphate backbone through the loop is localized between residues G_{-8} and A_{-6} and resembles a U-turn motif. This structure is consistent with the mixed C3'-/C2'-*endo* sugar puckers of these residues and several strong inter-residue NOEs between their ribose protons.

The structure of the phage GA loop stands in contrast to the loop structure of the phage MS2 operator hairpin (Figure 6). The 5'-AUUA-3' loop sequence has a stable and generally well-

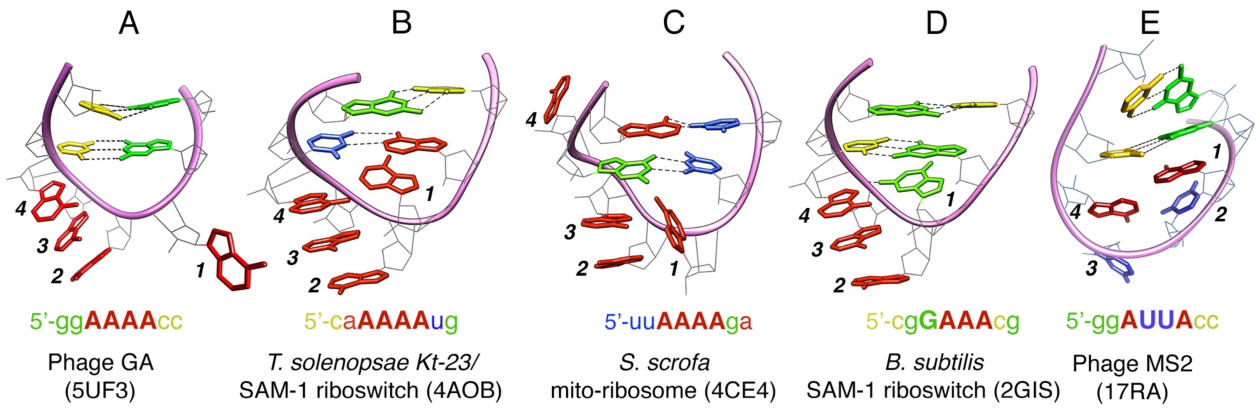

Figure 6. Structures showing the conformational variability of the tetra-A loop (A-C) from PDB coordinates 5UF3 (this study), 4AOB [11], and 4CE4 [20]. Included for comparison are structures of the canonical GNRA tetraloop 2GIS [51] (D) and the phage MS2 loop 17RA [15] (E). Structures A and E, B and D, and C were determined using solution NMR, X-ray crystallography, and cryo-EM, respectively. The loop nucleotides are numbered sequentially beginning with the 5' proximal nucleotide of the loop. Each of the tetra-A loops have different closing base pairs. Although adoption of a 3' stack may be the preferred conformation when the loop is closed by a 5'-R:Y-3' base pair, there are too few structures of the tetra-A loop to determine if such a correlation exists.

ordered structure in which A_{-7} and U_{-6} form a 5' stack and A_{-4} is positioned towards the helix axis. The phosphate backbone reverses between U_{-6} and A_{-4} and the position of the third loop nucleotide, U_{-5}, is ill-defined. The loop is further stabilized by a hydrogen bond between A_{-4} N6H and U_{-6} O2 [15]. The tetra-A loop in the phage GA RNA hairpin is well-ordered except for A_{-7} whose resonances are broadened by chemical exchange. The dynamical character of A_{-7} appears to propagate down the 5' strand of the upper stem leading to significant broadening of

the G_{-8} and G_{-9} bases resonances not observed when the stem is capped by the AUUA loop sequence.

Function and structure of tetra-A loops. In addition to the phage GA operator RNA, tetra-A loops also can be found in small and large ribosomal RNA subunits, group I and group II introns, and tmRNA molecules. Tetra-A loops have been considered unstructured or unstable and several investigations have used tetra-A loops as negative controls to test the functionality of other structured tetra-loops [1, 2, 47]. For instance, a GNRA tetra-loop present in a potato viroid RNA was replaced with the tetra-A loop to demonstrate the functional necessity of the GNRA tetra-loop for RNA to be processed [47]. Tetra-A loops also compose loop 2 of at least five naturally occurring hammerhead enzymes [48], with the other three belonging to the GNRA tetra-loop class. Loop 1 and loop 2 of the hammerhead ribozyme form a loop-loop interaction that includes a reverse Hoogsteen A-U base pair between the 3' A of loop 2 (GNRA class) to the 5' U of the loop 1. The tetra-A loop is expected to function (and fold) similar to the GNRA type loop 2. A functional tetra-A loop is also present in the domain IV of the Ll.LtrB intron in *Lactococcus lactis* [49]. Domain IV contains the Shine-Dalgarno sequence and the start codon for the intron-encoded and autogenously regulated protein LtrA. Mutations of the first two adenine nucleotides of the tetra-A loop in domain IV decrease the affinity of LtrA for the RNA by 33-50 fold [49] and leads to unregulated expression of LtrA.

There are few structures of tetra-A loops and most (PDB: 4CE4, 4V6W, 4V8M, 5AJ3, 4V4N) are from ribosomal RNA solved using cryo-EM [19, 21, 22]. The only high-resolution crystal structure of a tetra-A loop (PDB: 4AOB) is that from the SAM-I riboswitch with kink-turn 23 of *Thelohania solenopsae* [11]. As in the phage GA RNA, this tetra-A loop adopts a 3' stack and the phosphate backbone reverses direction between the first and second adenines (Figure 6). The

GAAA tetra-loops that cap the corresponding helices in other SAM-I riboswitches adopt the U-turn architecture [3, 50, 51]. The tetra-A loop of the SAM-I riboswitch also adopts the U-turn fold similar to the GNRA tetra-loops including the sheared purine-purine interaction between the 5'- and 3'-proximal nucleotides. Two of the ribosomal RNA tetra-A loops, (PDB: 4V6W and 4V8M) also adopt the U-turn motif. Although the folds of these tetra-A loops and the GAAA tetra-loop are the same, the guanine at the first position affords additional hydrogen bonding opportunities and this difference is evident from thermal melting studies that show the GAAA tetra-loop is approximately 0.5 kcal/mol more stable than the tetra-A loop [2]. The remaining three tetra-A loop structures (PDB: 4CE4, 5AJ3, 4V4N) do not adopt the characteristic U-turn fold and display limited base stacking (Figure 6).

Significance for binding to the GA coat protein dimer. The crystal structure of the phage MS2 coat protein dimer in complex with its operator RNA hairpin has been solved [18], but not that of the phage GA protein-operator complex. The sites of the phage MS2 and GA coat proteins that bind the nucleotides of the RNA loop region contain six conserved amino acid residues, V29, T45, S47, K61 and Y85, and it is likely that these systems share several identical RNA-protein contacts involving these residues [18, 52]. Although the phage GA coat protein can bind the MS2 RNA hairpin, the phage MS2 coat protein is unable to bind the GA RNA hairpin. In the MS2 coat protein, N87, which is in the loop nucleotide binding site but does not participate directly in RNA binding itself, would sterically clash with a purine at the third position in the RNA loop [46]. In the phage GA coat protein, residue 87 is a serine and so is able to accommodate either a purine or pyrimidine at the third position in the loop. In the bound state of the phage MS2 RNA hairpin, the A_{-7} is stacked between G_{-8} and U_{-5} and the U_{-6} and A_{-4} residues are flipped away from the top of the helix. In the complex, A_{-4} has contacts with V29, T45, S47,

K61 and the U_{-5}:A_{-7} base stack is capped by the Y85 side chain. The MS2 loop sequence requires only modest structural rearrangement to achieve the bound conformational state. The A_{-7} is already stacked above the closing G-C base pair and the A_{-4} nucleotide is projecting away from the helix axis on the minor groove side (Figure 6) [15]. Several of the RNA-protein contacts that would be predicted for the phage GA coat protein-RNA operator complex indicate that the GA RNA loop is not structurally pre-organized and that some nucleotides must rearrange to achieve the bound state conformation. The A_{-7} base must settle onto the loop-closing G-C base pair on the major groove side creating a 5' stack and possibly facilitating the turning out of A_{-4} from above the C_{-3} ribose and away from the major groove face of the loop. The A_{-5} base then is well positioned lie to down and stack on A_{-7}. Residue A_{-6}, like U_{-6} in the MS2 loop, is solvent exposed at the top of a 3' base stack and positioned to flip away from the loop unencumbered.

The conformations of the (A-A)-U motif in the contexts of the GA and MS2 loop sequences suggest the tetra-A loop may facilitate coat protein binding. High affinity binding requires that A_{-10} turn out of the helix and insert into a binding pocket in the coat protein [18]. When capped by the tetra-A loop of phage GA, A_{-10} stacks in the helix between A-U and G-C base pairs, but remains dynamic. However, when capped by the 5'-AUUA loop of phage MS2, A_{-10} not only stacks into the helix between A-U and G-C base pairs, but forms a hydrogen bond with the cross strand uridine. The A_{-10} residue in this latter context is more stable and potentially less easily displaced from the helix and into the coat protein binding pocket [15]. The R_1 and $R_{1\rho}$ relaxation data and broader line widths of the loop and bulge region nucleotide resonances indicate enhanced mobility of bulge and loop residues relative to nucleotides of the lower stem. A_{-10} and A_{-7} exhibit the most extensive intermediate time scale motions, although the calculated structures suggest the amplitudes of the motions may differ for the two (Figure 3). These

residues are two of three (A_{-5} being the third) expected to have significantly different conformations in the protein bound state. The dynamic properties of A_{-7} may enable the initiation of structural changes in the loop that are necessary for coat protein binding. Similarly, the mobility of A_{-10} allows the base to readily unstack and flip away from the helix, stabilizing coat protein binding. In this way, the (A-A)-U motif in combination with the tetra-A loop of phage GA may provide an additional level of orthogonality with phage MS2 by lowering the energetic barrier to turning out the A_{-10} base.

SUPPORTING INFORMATION

Additional details of experimental methods, one figure, three tables

ACCESSION NUMBERS

Coordinates have been deposited in the Protein Data Bank under accession number PBD ID: 5UF3. Chemical shifts have been deposited in the Biomolecular Magnetic Resonance Bank under accession number BMRB ID: 30224.

ACKNOWLEDGEMENTS

We thank Dr. Eric DeJong for spin relaxation analysis and early effort on the project and Malgorzata Michnicka for preparation of the T7 RNA polymerase and synthesis of the labeled 5'-nucleotide triphosphates. The 800 MHz NMR spectrometer was purchased with funds from the W.M. Keck Foundation and the John S. Dunn Foundation.

FUNDING

This work was supported by National Science Foundation grant CLP-1412864 to E.P.N.

REFERENCES

[1] Antao, V. P., Lai, S. Y., and Tinoco, I., Jr. (1991) A thermodynamic study of unusually stable

RNA and DNA hairpins, *Nucleic Acids Res 19*, 5901-5905.

[2] Garcia-Garcia, C., and Draper, D. E. (2003) Electrostatic interactions in a peptide--RNA complex, *J Mol Biol 331*, 75-88.

[3] Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops, *J Mol Biol 264*, 968-980.

[4] Eggen, K., and Nathans, D. (1969) Regulation of protein synthesis directed by coliphage MS2 RNA. II. In vitro repression by phage coat protein, *J Mol Biol 39*, 293-305.

[5] Hohn, T. (1969) Role of RNA in the assembly process of bacteriophage fr, *J Mol Biol 43*, 191-200.

[6] Gott, J. M., Wilhelm, L. J., and Uhlenbeck, O. C. (1991) RNA binding properties of the coat protein from bacteriophage GA, *Nucleic Acids Res 19*, 6499-6503.

[7] Witherell, G. W., Gott, J. M., and Uhlenbeck, O. C. (1991) Specific interaction between RNA phage coat proteins and RNA, *Prog Nucleic Acid Res Mol Biol 40*, 185-220.

[8] Vanegas, P. L., Hudson, G. A., Davis, A. R., Kelly, S. C., Kirkpatrick, C. C., and Znosko, B. M. (2012) RNA CoSSMos: Characterization of Secondary Structure Motifs--a searchable database of secondary structure motifs in RNA three-dimensional structures, *Nucleic Acids Res 40*, D439-444.

[9] Thapar, R., Denmon, A. P., and Nikonowicz, E. P. (2014) Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins, *Wiley Interdiscip Rev RNA 5*, 49-67.

[10] Kang, M., Eichhorn, C. D., and Feigon, J. (2014) Structural determinants for ligand capture by a class II preQ1 riboswitch, *Proc Natl Acad Sci U S A 111*, E663-671.

[11] Schroeder, K. T., Daldrop, P., McPhee, S. A., and Lilley, D. M. (2012) Structure and

folding of a rare, natural kink turn in RNA with an A*A pair at the 2b*2n position, *RNA 18*, 1257-1266.

[12] Wu, H., Henras, A., Chanfreau, G., and Feigon, J. (2004) Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III, *Proc Natl Acad Sci U S A 101*, 8307-8312.

[13] Duszczyk, M. M., Wutz, A., Rybin, V., and Sattler, M. (2011) The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization, *RNA 17*, 1973-1982.

[14] Lukavsky, P. J., Kim, I., Otto, G. A., and Puglisi, J. D. (2003) Structure of HCV IRES domain II determined by NMR, *Nat Struct Biol 10*, 1033-1038.

[15] Smith, J. S., and Nikonowicz, E. P. (1998) NMR structure and dynamics of an RNA motif common to the spliceosome branch-point helix and the RNA-binding site for phage GA coat protein, *Biochemistry 37*, 13486-13498.

[16] Staple, D. W., and Butcher, S. E. (2003) Solution structure of the HIV-1 frameshift inducing stem-loop RNA, *Nucleic Acids Res 31*, 4326-4331.

[17] Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., El Omari, K., Baudin, F., Hart, D. J., Beck, M., and Cusack, S. (2016) Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains, *Mol Cell 61*, 125-137.

[18] Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G., and Liljas, L. (1997) The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions, *J Mol Biol 270*, 724-738.

[19] Anger, A. M., Armache, J. P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., and Beckmann, R. (2013) Structures of the human and Drosophila 80S ribosome, *Nature 497*, 80-85.

[20] Greber, B. J., Bieri, P., Leibundgut, M., Leitner, A., Aebersold, R., Boehringer, D., and Ban, N. (2015) Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome, *Science 348*, 303-308.

[21] Greber, B. J., Boehringer, D., Leibundgut, M., Bieri, P., Leitner, A., Schmitz, N., Aebersold, R., and Ban, N. (2014) The complete structure of the large subunit of the mammalian mitochondrial ribosome, *Nature 515*, 283-286.

[22] Hashem, Y., des Georges, A., Fu, J., Buss, S. N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H. Y., Grassucci, R. A., Bajaj, C., Westhof, E., Madison-Antenucci, S., and Frank, J. (2013) High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome, *Nature 494*, 385-389.

[23] Kalurachchi, K., and Nikonowicz, E. P. (1998) NMR structure determination of the binding site for ribosomal protein S8 from *Escherichia coli* 16 S rRNA, *J Mol Biol 280*, 639-654.

[24] Wu, H., Jiang, L., and Zimmermann, R. A. (1994) The binding site for ribosomal protein S8 in 16S rRNA and spc mRNA from *Escherichia coli*: minimum structural requirements and the effects of single bulged bases on S8-RNA interaction, *Nucleic Acids Res 22*, 1687-1695.

[25] Bullock, S. L., Ringel, I., Ish-Horowicz, D., and Lukavsky, P. J. (2010) A'-form RNA helices are required for cytoplasmic mRNA transport in Drosophila, *Nat Struct Mol Biol 17*, 703-709.

[26] Huang, L., Serganov, A., and Patel, D. J. (2010) Structural insights into ligand recognition

by a sensing domain of the cooperative glycine riboswitch, *Mol Cell 40*, 774-786.

[27] Bae, S. H., Cheong, H. K., Lee, J. H., Cheong, C., Kainosho, M., and Choi, B. S. (2001) Structural features of an influenza virus promoter and their implications for viral RNA synthesis, *Proc Natl Acad Sci U S A 98*, 10602-10607.

[28] Lee, M. K., Bottini, A., Kim, M., Bardaro, M. F., Jr., Zhang, Z., Pellecchia, M., Choi, B. S., and Varani, G. (2014) A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication, *Chem Commun (Camb) 50*, 368-370.

[29] Newby, M. I., and Greenbaum, N. L. (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture, *RNA 7*, 833-845.

[30] Davanloo, P., Rosenberg, A. H., Dunn, J. J., and Studier, F. W. (1984) Cloning and Expression of the Gene for Bacteriophage-T7 RNA-Polymerase, *P Natl Acad Sci-Biol 81*, 2035-2039.

[31] Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA Polymerase and Synthetic DNA Templates, *Nucleic Acids Res. 15*, 8783-8789.

[32] Nikonowicz, E. P. (2001) Preparation and use of ^{2}H-labeled RNA oligonucleotides in nuclear magnetic resonance studies, *Methods Enzymol 338*, 320-341.

[33] Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M., and Pardi, A. (1992) Preparation of C-13 and N-15 Labeled RNAs for Heteronuclear Multidimensional NMR Studies, *Nucleic Acids Res. 20*, 4507-4513.

[34] Denmon, A. P., Wang, J., and Nikonowicz, E. P. (2011) Conformation effects of base modification on the anticodon stem-loop of *Bacillus subtilis* tRNA^{Tyr}, *J Mol Biol 412*, 285-303.

[35] Wang, J., and Nikonowicz, E. P. (2011) Solution structure of the K-turn and Specifier Loop domains from the *Bacillus subtilis tyrS* T-box leader RNA, *J Mol Biol 408*, 99-117.

[36] O'Neil-Cabello, E., Wu, Z., Bryce, D. L., Nikonowicz, E. P., and Bax, A. (2004) Enhanced spectral resolution in RNA HCP spectra for measurement of $^{3}J_{C2'P}$ and $^{3}J_{C4'P}$ couplings and ^{31}P chemical shift changes upon weak alignment, *J Biomol NMR 30*, 61-70.

[37] Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson. 160*, 65-73.

[38] Yamazaki, T., Muhandiram, R., and Kay, L. E. (1994) NMR Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly C-13,N-15-Labeled Proteins - Application to C-13(Alpha) Carbons, *J. Am. Chem. Soc. 116*, 8266-8278.

[39] Bermejo, G. A., Clore, G. M., and Schwieters, C. D. (2016) Improving NMR Structures of RNA, *Structure 24*, 806-815.

[40] Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res 31*, 5108-5121.

[41] Kojima, C., Ono, A., Kainosho, M., and James, T. L. (1998) DNA duplex dynamics: NMR relaxation studies of a decamer with uniformly ^{13}C-labeled purine nucleotides, *J Magn Reson 135*, 310-333.

[42] King, G. C., Harper, J. W., and Xi, Z. (1995) Isotope labeling for 13C relaxation measurements on RNA, *Methods Enzymol 261*, 436-450.

[43] Dieckmann, T., and Feigon, J. (1997) Assignment methodology for larger RNA oligonucleotides: application to an ATP-binding RNA aptamer, *J. Biomol. NMR 9*, 259-

272.

[44] Pardi, A. (1995) Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA, *Methods Enzymol 261*, 350-380.

[45] Kellogg, G. W., and Schweitzer, B. I. (1993) Two- and three-dimensional ^{31}P-driven NMR procedures for complete assignment of backbone resonances in oligodeoxyribonucleotides, *J Biomol NMR 3*, 577-595.

[46] Grahn, E., Stonehouse, N. J., Murray, J. B., van den Worm, S., Valegard, K., Fridborg, K., Stockley, P. G., and Liljas, L. (1999) Crystallographic studies of RNA hairpins in complexes with recombinant MS2 capsids: implications for binding requirements, *RNA 5*, 131-138.

[47] Schrader, O., Baumstark, T., and Riesner, D. (2003) A mini-RNA containing the tetraloop, wobble-pair and loop E motifs of the central conserved region of potato spindle tuber viroid is processed into a minicircle, *Nucleic Acids Res 31*, 988-998.

[48] Dufour, D., de la Pena, M., Gago, S., Flores, R., and Gallego, J. (2009) Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads, *Nucleic Acids Res 37*, 368-381.

[49] Singh, R. N., Saldanha, R. J., D'Souza, L. M., and Lambowitz, A. M. (2002) Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation, *J Mol Biol 318*, 287-303.

[50] Lu, C., Ding, F., Chowdhury, A., Pradhan, V., Tomsic, J., Holmes, W. M., Henkin, T. M., and Ke, A. (2010) SAM recognition and conformational switching mechanism in the

Bacillus subtilis yitJ S box/SAM-I riboswitch, *J Mol Biol 404*, 803-818.

[51] Montange, R. K., and Batey, R. T. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element, *Nature 441*, 1172-1175.

[52] Tars, K., Bundule, M., Fridborg, K., and Liljas, L. (1997) The crystal structure of bacteriophage GA and a comparison of bacteriophages belonging to the major groups of *Escherichia coli* leviviruses, *J Mol Biol 271*, 759-773.

[53] Houck-Loomis, B., Durney, M. A., Salguero, C., Shankar, N., Nagle, J. M., Goff, S. P., and D'Souza, V. M. (2011) An equilibrium-dependent retroviral mRNA switch regulates translational recoding, *Nature 480*, 561-564.