2012

Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch

Catherine D. Eichhorn
The University Of Michigan

Jun Feng
The University Of Michigan

Krishna C. Suddala
The University Of Michigan

Nils G. Walter
The University Of Michigan

Charles L. Brooks III
The University Of Michigan, brookscl@umich.edu

See next page for additional authors

Follow this and additional works at: https://digitalcommons.unl.edu/chemfacpub

Part of the Analytical Chemistry Commons, Medicinal-Pharmaceutical Chemistry Commons, and the Other Chemistry Commons

Eichhorn, Catherine D.; Feng, Jun; Suddala, Krishna C.; Walter, Nils G.; Brooks, Charles L. III; and Al-Hashimi, Hashim M., "Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch" (2012). Faculty Publications -- Chemistry Department. 180.
https://digitalcommons.unl.edu/chemfacpub/180

This Article is brought to you for free and open access by the Published Research - Department of Chemistry at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications -- Chemistry Department by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Unraveling the structural complexity in a single-stranded RNA tail: implications for efficient ligand binding in the prequeuosine riboswitch

Catherine D. Eichhorn¹, Jun Feng²,³, Krishna C. Suddala²,³, Nils G. Walter², Charles L. Brooks III¹,²,³,⁴,* and Hashim M. Al-Hashimi¹,²,³,*

¹Chemical Biology Doctoral Program, ²Department of Chemistry, ³Department of Biophysics, University of Michigan, Ann Arbor, MI 48109 and ⁴Center for Theoretical Biological Physics, University of California San Diego, San Diego, CA 92037, USA

Received June 16, 2011; Revised September 19, 2011; Accepted September 20, 2011

ABSTRACT

Single-stranded RNAs (ssRNAs) are ubiquitous RNA elements that serve diverse functional roles. Much of our understanding of ssRNA conformational behavior is limited to structures in which ssRNA directly engages in tertiary interactions or is recognized by proteins. Little is known about the structural and dynamic behavior of free ssRNAs at atomic resolution. Here, we report the collaborative application of nuclear magnetic resonance (NMR) and replica exchange molecular dynamics (REMD) simulations to characterize the 12 nt ssRNA tail derived from the prequeuosine riboswitch. NMR carbon spin relaxation data and residual dipolar coupling measurements reveal a flexible yet stacked core adopting an A-form-like conformation, with the level of order decreasing toward the terminal ends. An A-to-C mutation within the polyadenine tract alters the observed dynamics consistent with the introduction of a dynamic kink. Pre-ordering of the tail may increase the efficacy of ligand binding above that achieved by a random-coil ssRNA. The REMD simulations recapitulate important trends in the NMR data, but suggest more internal motions than inferred from the NMR analysis. Our study unmasks a previously unappreciated level of complexity in ssRNA, which we believe will also serve as an excellent model system for testing and developing computational force fields.

INTRODUCTION

Single-stranded RNAs (ssRNAs), typically located at the ends of RNA hairpins and consisting of more than three unpaired residues, serve diverse structural and functional roles. They can fold onto neighboring RNA hairpins to form pseudoknots, essential architectural RNA elements involved in ribosomal frameshifting (1,2), hepatitis C internal ribosomal entry site (IRES) recognition (3,4) and telomerase activity (5). Messenger RNA (mRNA) degradation is prevented or promoted by 3' addition of a polyadenylated tail, which recruits essential protein cofactors (6). Cleavage of the 5' transfer RNA (tRNA) leader by RNase P is a key step in tRNA maturation (7). In riboswitches, ssRNA links the ligand-binding aptamer domain to the expression platform, providing the basis for communication between the two (8–10).

Much of our understanding of the conformational behavior of ssRNA comes from high-resolution NMR and X-ray structures of RNA, in which ssRNA directly engages in tertiary or RNA–protein interactions. However, the atomic-level structural and dynamic behavior of these elements in the absence of these interactions remains unclear, in large part due to their high degree of flexibility. Several studies suggest that ssRNA polynucleotides adopt stacked and partially helical conformations, particularly adenine-rich sequences; however, the biological relevance of these structures is unclear (11–17). Atomic-resolution studies of ssRNA are scarce: at present only one iso-sequential ssRNA and ssDNA sequence has been characterized by homonuclear NMR methods and shown to possess properties reminiscent of A-form and B-form helices, respectively (18). Few MD studies have been...
performed on ssRNA, the majority of which use the AMBER force field (19,20) to explore the impact of chemical modifications such as peptide nucleic acids (PNA) and O2'-methylation (21–24).

The class I prequeuosine riboswitch (*queC*), typically found in firmicute bacterial species, is commonly located in the 5'-untranslated region (UTR) of the queCDEF operon, which expresses proteins directly involved in the queuosine biosynthetic pathway (25). The aptamer binds preQ1, an intermediate in queuosine synthesis, with high affinity to attenuate protein expression at either the transcription or translation level (25). This class has the smallest minimal aptamer domain (34 nucleotides, nt) discovered to date, consisting of a small hairpin followed by a 12 nt ssRNA tail (Figure 1A). Upon ligand recognition, the highly conserved adenine-rich tail condenses into a pseudoknot, forming a host of interactions to both the hairpin and ligand, including A-minor ‘kissing’ interactions between the ssRNA polyadenine tract and the minor groove (26–30). The activity of transcription-regulating riboswitches, such as the *Bacillus subtilis queC* riboswitch, has been shown to depend on the kinetics of ligand binding as well as the rate of transcription (8). Notably, the very small size of the *queC* riboswitch leaves very little time, in comparison to other switches, for ligand binding to take place prior to formation of the anti-terminator helix which, when formed, prevents terminator helix formation, thereby allowing gene expression to continue. For example, the *B. subtilis* FMN riboswitch, which is highly dependent upon the rate of polymerase and contains sites that locally pause polymerase to lengthen the ligand-binding window, has ~70 nt between the minimal aptamer sequence and complete formation of the anti-terminator helix (8). In comparison, the ligand-binding window for the *queC* riboswitch is ~20 nt (26,27). How efficient ligand binding is achieved is unclear given that the ssRNA tail is thought to be highly disordered, and therefore capable of sampling a wide range of competing conformations.

Here, we use NMR chemical shifts, spin relaxation, and residual dipolar couplings (RDCs) in conjunction with REMD simulations using the recently updated CHARMM27 nucleic acid force field (31,32) to explore the conformational properties of the 12 nt ssRNA tail from the *queC* aptamer domain and the impact of a single A-to-C mutation targeting the polyadenine tract. Our study unmasks a previously unappreciated level of complexity in ssRNA and suggests that these structures can serve as excellent model systems for testing and developing computational force fields.

**MATERIALS AND METHODS**

**Sample preparation**

Uniformly $^{13}$C/$^{15}$N-labeled *queC*36 and C14U/C17U constructs were prepared by *in vitro* transcription as described previously (33). Unlabeled wild-type (WT, 5'-AUAAAAAAACUAA-3') and A29C (5'-AUAAACAAAAACUAA-3') RNAs were purified from Integrated DNA Technologies (IDT) and purified using a C18 column (Waters) followed by lyophilization and reconstitution in NMR buffer (15 mM sodium phosphate, pH 6.4; 25 mM sodium chloride, 0.1 mM EDTA) containing 10% D$_2$O by volume. 100% D$_2$O samples were prepared by repeatedly lyophilizing the sample and replacing with 99.99% pure D$_2$O (Sigma) three times. RNA concentrations ranged from 1.5 to 2.8 mM. AMP, UMP and CMP (Sigma) were directly dissolved into NMR buffer with no additional purification to 5 mM. For RDC measurements, samples were dialyzed into Millipore-purified ddH$_2$O using 1 kDa dialysis tubing (Spectrum Labs), lyophilized, and reconstituted into 52.4 mg/ml *Pf1* phage solution (34–36) in NMR buffer with 100% D$_2$O (Asla Biotech). RNA concentrations in *Pf1* phage ranged from 1.5 to 2 mM.

**UV/Vis melting**

RNA samples (0.25–0.5 μM) were prepared in NMR buffer and the melting profiles measured between 275 K and 368 K using a Varian Bio 300 UV/Vis instrument equipped with a Cary Temperature Controller. The absorbance at 260 nm was recorded every 0.5 s with a ramp rate of 0.5 K/min. The two-state helix to coil melting transition was analyzed using

\[
A = A_C + \left( A_H - A_C \right) \frac{e^{(\Delta S/RT) - (\Delta H/RT)}}{1 + e^{(\Delta S/RT) - (\Delta H/RT)}},
\]

where \(A\) is the absorbance value at a given temperature \(T\), \(A_H\) is the absorbance of the fully helical ssRNA, \(A_C\) is the absorbance of the fully random coil ssRNA, \(\Delta S\) and \(\Delta H\) are the entropy and enthalpy of the melting transition respectively, and \(R\) is the gas constant (37,38). Absorbance values were fitted to the above equation using the non-linear least squares fitting function in Origin 7 to determine thermodynamic parameters. The melting temperature \(T_m\) was determined by dividing the enthalpy by the entropy.

**NMR experiments**

All NMR experiments were performed on an Avance Bruker 600 MHz NMR spectrometer equipped with a triple-resonance 5-mm cryogenic probe. NOEY experiments were performed at 277 K and 298 K using a mixing time of 350 ms (39). $^{13}$C spin relaxation experiments were performed at natural abundance and 298 K (40). Relative order parameters were calculated by normalizing \((2R_2-R_1)\) to either A31 (C8) or C33 (C6). Relaxation parameters were computed using HydroNMR (41,42), assuming an idealized A-form structure, to obtain diffusion tensor parameters \((\tau_m\) and \(D_{ratio}\)), and in-house written software was used to compute \(R_2/R_1\) values as previously reported (33,40). Motionally averaged bond lengths of 1.104 Å were used for both C8 and C6 moieties as previously described (40,43). The following experimentally derived CSAs \((\sigma_{xx}, \sigma_{yy}, \sigma_{zz}\) were used in the analysis: (89, 15, –104); (80, 5, –85); and (98,4, 9.2, –107.5) for C2, C8 and C6 moieties (43,44). IP-COSY experiments were performed at 277 K and 298 K to observe relative $^3$J$_{HH}$ scalar coupling crosspeak intensities (45). Base and sugar $^1$H-$^{13}$C splittings were measured from the
difference between the upfield and downfield components of the $^1$H-$^1$H-$^1$C doublet along the $^1$H component using the narrow transverse relaxation-optimized spectroscopy (TROSY) component in the $^1$H-$^1$C dimension as implemented in 2D $^1$H-$^1$H-$^1$C S$^3$T-heteronuclear single quantum correlation (HSQC) experiments (46). $^2$H splittings were 71 and 69 Hz for WT and A29C, respectively. Idealized A-form structures were constructed using Insight II (Molecular Simulations, Inc.) correcting the propeller twist angles from $+15^\circ$ to $-15^\circ$ using an in-house program, as previously described (47). The complementary strand was removed and the resulting ssRNA used in NMR data analysis. B-form helices were constructed using W3DNA (48).

Computational methods

Simulation. REMD simulations were performed with the CHARMM simulation package (49) using the recently updated CHARMM27 nucleic acid force field (31,32) and the MMTSB (50) tool set. Each REMD simulation comprised 40 replicas exponentially distributed over a temperature range from 278 K to 330 K, resulting in an average exchange acceptance ratio of 30%. Each replica was first equilibrated for 0.5 ns, restraining nucleotide heavy atoms, and subsequently run without any restraints for 10 ns, with exchange moves attempted every 0.5 ps.

Both WT and A29C RNAs were initially built in an ideal A-form helical configuration and served as the starting conformation in every simulation of REMD. The RNA was solvated in an 80-A˚ cubic box of pre-equilibrated TIP3P water (approximately 50 000 atoms). Twelve pairs of sodium chloride with an additional 11 sodium ions were added to the box, corresponding to the experimental ionic concentration of 40 mM.

Analysis. We utilized the last 5 ns of the REMD trajectory at 298 K for the following analysis. Base stacking energies were defined as the electrostatic and van der Waals interaction energies between the adjacent bases. The molecular orientation was expressed by the order parameters $S^2$ of the C–H bond vectors employing the model-free approach of Lipari and Szabo (51). After a translational and rotational fit of each RNA snapshot to the ideal A-form helical structure, the order parameters were taken from the plateau phase of the correlation function, given by $C(t) = \langle P_2(\mu(0) \cdot \mu(t)) \rangle$, where $P_2$ is the second order Legendre polynomial and $\mu$ is the unit vector along the C–H dipole. Additionally, from the atomic coordinates we constructed the RDC values by first orienting an idealized C–H dipole. Additionally, from the atomic coordinates we constructed the RDC values by first orienting an idealized C–H dipole. Furthermore, from the atomic coordinates we constructed the RDC values by first orienting an idealized C–H dipole. Additionally, from the atomic coordinates we constructed the RDC values by first orienting an idealized C–H dipole.

The average structure of the ssRNA was calculated as the structure with the minimal root-mean-square deviations from all RNA conformations in the 5 ns REMD trajectory.

RESULTS AND DISCUSSION

NMR chemical shift and NOE-based analysis of the ssRNA tail conformation

Previous studies have shown that in the absence of ligand, the queC aptamer domain folds into a non-native hairpin, in which the 5’-strand frame-shifts to allow the first two guanine residues to base pair, with the 12 nt ssRNA tail lacking any tertiary interactions (26). The 2D C–H NMR spectra of the 36 nt queC minimal aptamer domain (Figure 1A), in the absence of ligand, show severe resonance overlap and large variations in resonance intensities indicating a highly disordered conformation (Figure 1B). Excess imino proton resonances as well as $^1$H-$^1$N NOE data indicate that the unbound queC aptamer domain is in equilibrium between native and non-native hairpin conformations (data not shown), consistent with previous NMR studies (26).

The NMR spectra suggest that the unbound 36 nt queC minimal aptamer domain is highly disordered and that the ssRNA tail is not involved significantly in any tertiary interactions. To test this hypothesis further, we compared NMR spectra of the isolated 12 nt ssRNA tail with the corresponding spectra of the unbound queC aptamer. Remarkably, NMR spectra of the isolated 12 nt ssRNA tail overlay almost perfectly with the queC aptamer domain and specifically onto the highly intense resonances corresponding to highly disordered residues (Figure 1B). The only significant deviations are observed for A25 and U26, which are located at the junction site between the hairpin and the tail (Figure 1B). This indicates that in the absence of ligand, the ssRNA tail is not involved in any significant tertiary interactions under the NMR conditions.

Figure 1. (A) Sequence and secondary structure of the B. subtilis queC riboswitch minimal aptamer with PreQ1 inset, (B) 2D C–H NMR chemical shifts show near-identical agreement between single-stranded tail (SS tail), shown in red, and unbound minimal aptamer (queC36), (C) NOE crosspeaks at 277 K enable resonance assignment and indicate base stacking within single strand.
(1 mM RNA, 25 mM sodium chloride, 15 mM sodium phosphate, pH 6.4, 0.1 mM EDTA, 298 K).

Similarly to the *Fusobacterium nucleatum queC* riboswitch, the *B. subtilis queC* aptamer forms kissing dimers, as observed in non-denaturing polyacrylamide gels (Supplementary Figure S1) (52). To ensure that the dimer does not obstruct hairpin–tail interactions, we compared a mutant C14U/C17U construct characterized previously by Kang and coworkers to generate a ligand-bound solution NMR structure (26) to the WT aptamer. MFold predicts the C14U/C17U mutations will reduce the dimer stability from −6.1 kcal/mol to −0.9 kcal/mol (53). While we observe removal of the kissing dimer, chemical shifts overall overlay extremely well between the WT *queC* aptamer and the C14U/C17U mutant (Supplementary Figure S1). Specifically, tail chemical shifts correspond extremely well to the 12 nt ssRNA, further suggesting that the tail does not participate in tertiary interactions in the absence of ligand under our NMR conditions.

Strikingly, the spectra of the 12 nt ssRNA are well resolved, indicating that it does not adopt a completely random conformation (Figure 1B and Supplementary Figure S2). This stands in stark contrast to corresponding spectra of a 12 nt polyuridine (polyU) ssRNA, well established to have a random-coil conformation (16), which exhibits severe spectral overlap indicative of a highly disordered conformation (Supplementary Figure S2). This structural order is observed in the ssRNA despite the lack of any observable imino protons and therefore any base structural order is observed in the ssRNA despite the lack of ordered conformation (Supplementary Figure S2). This is consistent with increased formation of stacking interactions (59) (data not shown). In contrast polyU has near-identical (∆=0.1 ppm) chemical shifts to UMP (Supplementary Figure S2). Thus, consistent with NOE data, the chemical shift data suggest a comparatively stacked core with a growing level of disorder towards the terminal ends. Normalized resonance intensities (33) further support these observations, which gradually increase towards the terminal ends, consistent with a higher level of pico- to nanosecond motions (Supplementary Figure S2).

**Thermal stability by experiment and REMD computation**

The abundance of NOEs indicates significant base stacking interactions, which likely contribute to ordering of the tail. To probe the thermodynamic stability of the tail, we performed UV/Vis melting experiments to determine the melting temperature of the helix to coil transition. Consistent with previous studies of single-stranded nucleic acids, the melting profile of the ssRNA is extremely broad, characteristic of a non-cooperative transition (Figure 2A) (37). Previous studies of a 7 nt polyadenine ssRNA in similar buffer conditions yield analogous melting temperatures to those observed (∼35°C compared to 31.7 ± 1.90°C) (37).

We then used our REMD simulations to explore the temperature dependence of base stacking compared to the described UV/Vis melting curves. Base stacking energies from the REMD simulation between temperatures 278–330 K show a similar gradual decrease with increasing temperature and a similar, although reduced, *T_m* value (experimental 31.7 ± 1.90°C compared towards the center of the tail (57). This is further supported by chemical shift perturbations in a trajectory toward the NMPs with increasing temperature (data not shown). Alternatively, addition of magnesium up to 4 mM results in slight chemical shift perturbations farther from NMPs, consistent with previous studies suggesting that increases in ionic strength stabilize ssRNA stacking interactions (59) (data not shown). In contrast polyU has near-identical (∆=0.1 ppm) chemical shifts to UMP (Supplementary Figure S2). Thus, consistent with NOE data, the chemical shift data suggest a comparatively stacked core with a growing level of disorder towards the terminal ends. Normalized resonance intensities (33) further support these observations, which gradually increase towards the terminal ends, consistent with a higher level of pico- to nanosecond motions (Supplementary Figure S2).

**Figure 2.** (A) UV melting profile (black) compared to base stacking energy calculated from REMD simulations (gray), (B) NMR (closed) and REMD (open) relative order parameters suggest central polyadenine residues are more ordered with flexible terminal ends, (C) NMR ∆13C spin relaxation *R_2/R_1* values, with HYDRONMR-predicted values assuming an order parameter (S^2) of 0.45 shown as a gray bar, (D) REMD-calculated order parameters.
to computed 20–25°C, as estimating the $T_{50}$ value from melting curve, Figure 2A). However, the calculated base stacking energy plateaus around 320 K while the experimental slope begins to plateau around 330 K, indicating that stacking energies may be under-estimated in the REMD simulation or that additional unaccounted-for factors contribute to the ssRNA stability. Nevertheless, our data suggest that base stacking is the guiding force behind ssRNA stability, consistent with previous studies.

**Picosecond to nanosecond dynamics by NMR spin relaxation and comparison with REMD simulations**

To gain further insights into the dynamic properties of the ssRNA at pico- to nanosecond timescales, we measured longitudinal ($R_1$) and transverse ($R_2$) carbon relaxation data for the nucleobases (C2 C6 C8) using 2D $^{13}$C relaxation $R_1$ and $R_{1p}$ NMR experiments (40), where $R_1$ and $R_2$ values are determined using in-house software (Supplementary Figure S4). These measurements represent the first nucleobase $^{13}$C relaxation measurements performed on a ssRNA. The measured $R_1$ and $R_2$ values were used to compute order parameters (51) using $S^2 = (2R_2 - R_1)$ (60), and normalized to yield a relative order parameter ($S^2_{rel}$) describing the relative degree of order within a molecule ranging from 0 to 1, where 0 and 1 represent minimum and maximum order, respectively. The $S^2_{rel}$ values were normalized against central residues A31 (C8) and C33 (C6). Resonance overlap prevented the normalization of C2 spins. Again, we observe a gradual reduction in $S^2_{rel}$ indicating higher levels of disorder moving from central polyadenine residues (A28–C33) towards the terminal ends (Figure 2B).

We also computed the $S^2_{rel}$ values based on the REMD simulation described above. The REMD simulations reproduce the general trends observed in the experiments; however, the simulations show significantly increased dynamics at the terminal ends compared to experimental values, with $S^2_{rel}$ values approaching the dynamic limit (Figure 2B). Additionally, while experimental values have similar relative order parameters from A28–C33, large variations are observed in the REMD simulation, with A29–A30 being more ordered and A32 less ordered than experimentally observed (Figure 2B). These differences may reflect shortcomings in the force field and/or mismatch in the experimental/computational timescales since the REMD simulations likely probe fluctuations that extend beyond the picosecond timescales sensed by spin relaxation data.

The high level of disorder and motional coupling in the ssRNA prevents quantitative analysis of relaxation data using the model-free formalism, which assumes that internal and overall motions are decoupled from one another (51). This makes it difficult if not impossible to assess the absolute level of disorder in the ssRNA; one can only make qualitative assessments about the relative disorder across different residues. However, it is noteworthy that even the comparatively high $R_2/R_1$ values measured in the rigid core (~2.9, Figure 2C) remain significantly lower than values predicted for a perfectly rigid helical ssRNA (~6.4, Supplementary Figure S4) as estimated using the program HYDRONMR (41,42). If we assume an overall diffusion tensor predicted by HYDRONMR, we find that central polyadenine residues are highly flexible with an estimated average NMR spin relaxation order parameter $S^2$ of ~0.45 (Figure 2C and Supplementary Figure S4). Interestingly, similar though slightly smaller absolute $S^2$ values are calculated from the REMD simulations (on average $S^2$ ~0.36 for core residues, Figure 2D). These data indicate that despite measurable stacking interactions and a helical-like average conformation, the polyadenine core is highly disordered with residues experiencing fluctuations on the order of a ± 40° cone angle (61) at pico- to nanosecond timescales.

**Overall conformation and sub-millisecond dynamics by NMR residual dipolar couplings and comparison with REMD simulations**

To further probe the conformation of the ssRNA and extend the NMR timescale sensitivity to milliseconds, we measured RDCs (62,63) using 52.4 mg/ml Pf/ phage as an ordering medium. While most RNAs align optimally in ~25 mg/ml of phage, a much higher concentration of phage was used for the ssRNA to ensure optimal alignment. To our knowledge, these are the first RDC measurements reported on a single-stranded nucleic acid. The RDCs measured between two nuclei depend on $H_1$ where $\theta$ is the angle between the inter-nuclear vector and the magnetic field and the angular bracket denotes a time-average over all orientations sampled at sub-millisecond timescales (62,63). RDCs were measured for base C5H5, C6H6, C8H8, C2H2 and sugar C1’H1’ moieties (47).

In general, isotropic motions tend to reduce the observed RDC value, approaching zero at the limit of spatially unrestricted isotropic motions (61,64,65). In the ssRNA, large base C–H RDCs are measured in the polyadenine tract residues that gradually decrease at the termini (Figure 3A). Although small RDC values can also arise from static placement of the bond vector near the magic angle relative to the principal direction of order, the overall trends observed are consistent with NMR chemical shift and $S^2_{rel}$ data suggesting that the RDCs indicate increased dynamic averaging at the termini (Figure 3A). Interestingly, the near-zero RDCs measured at terminal residues (Figure 3A and Supplementary Figure S5) agree more closely to the REMD simulations compared to the $S^2_{rel}$ values, indicating that the discrepancy between the measured and computed $S^2_{rel}$ values may be due to truncation of the $S^2$ sensitivity to motions faster than nanoseconds. These results add to a growing number of NMR studies on different types of RNA showing that RDC data are capable of probing motions that are incompletely sensed by spin relaxation due to truncation of the time-sensitivity by overall correlation time of the molecule (64,66–68). Unfortunately, severe spectral overlap, particularly pronounced in the Pf/ phage sample, prevented measurement of several C1’H1’ RDCs for the polyadenine core.

We subjected the RDCs (excluding RDCs for the two flexible residues from the terminal ends) to an order tensor
analysis (47,69,70) assuming different input structures including single strands derived from idealized A-form and B-form helices, the REMD-averaged structure, and available ligand-bound X-ray and NMR structures (26,27). Despite the relatively small number of RDCs used in this analysis, we clearly observe a better fit with an A-form geometry (Q-factor 4.77%) as compared to all other conformations (Q-factor ≥ 16%) (Figure 3B). This is consistent with independently observed $^3J_{H1-H2}$ scalar coupling crosspeaks, which indicate a C3'-endo sugar conformation for core residues in the tail, suggesting an A-form (and not B-form) helical geometry. The RDCs are in strong disagreement with preQ1-bound X-ray and NMR structures (PDBID: 3FU2 and 2L1V) indicating that the tail must undergo a transition from an A-form helical geometry towards the distinct helical conformation observed in the X-ray and NMR structures in which the A-form geometry is perturbed at the hairpin-tail junction, likely due to torsional strain from the ssRNA folding back upon the hairpin. The REMD-averaged structure has a Q-factor of 30%, indicating a better fit than ligand-bound structures, but is still outside the range considered to represent a good fit. Together, these data suggest that, on average, the ssRNA tail adopts an A-form like conformation. The good RDC fit to the A-form structure also suggests that averaging of the RDCs due to internal motions is largely isotropic in nature, causing a semi-uniform attenuation of the RDCs relative to values expected for an A-form structure. The dynamics could involve exchange between a stacked ordered conformation and unstacked highly disordered conformation, or local isotropic motions about the average A-form conformation.

As a further check on the accuracy of the A-form structure, we compared the principal direction of alignment ($S_{zz}$) determined experimentally using RDCs assuming a ssRNA A-form structure with the orientation predicted by PALES (71) using a ssRNA A-form structure. Surprisingly, we find that the experimentally determined $S_{zz}$ deviates from the helix axis by ~19.8° (Figure 3C). Interestingly, PALES predicts a principal direction of order that deviates from the helix axis by 14.4°; the $S_{zz}$ orientation predicted using PALES is in good agreement from that measured experimentally (deviation ~5°). The deviation from the helix axis can be attributed to the absence of the complementary strand, resulting in an overall shape with a long axis that is not coincident with the helical axis, as reported previously for a quadruplex DNA topology (72).

To further test the conformational distribution from the REMD simulations, we used a number of simplifying assumptions to compute RDCs from the REMD trajectory. Snapshots from the REMD simulations were superimposed onto an idealized A-form helix oriented in the principal axis system determined using the experimental RDCs and the order tensor fit. RDCs were then arbitrarily scaled by $-82/r^3$, in which $r$ is the C–H bond length and accounts for bond length variations during the dynamics. We find excellent agreement between experimental and computed nucleobase RDCs; however, computed C1'H1' RDCs fail to reproduce observed RDCs, particularly for A32: while the magnitude is similar (18 Hz compared to ~30 Hz) the sign differs, suggesting the orientation of the C1'H1' bond vector differs between experiment and simulation (Supplementary Figure S5). C1'H1' RDCs are generally opposite in sign to base RDCs in a double-stranded A-form helix. However, back-calculated C1'H1' RDCs from the order tensor analysis assuming a ssRNA A-form helix are positive in sign, suggesting the C1'H1' orientation in the REMD simulations deviates from an A-form structure (Supplementary Figure S5).

**Impact of A-to-C mutation within polyadenine core**

Taken together, the data show that the polyadenine tract is relatively ordered at 298 K, with a gradual reduction in order approaching the termini and that the base stacking interactions are the guiding force behind this order. To determine whether disrupting the polyadenine tract will destabilize the global structure, we substituted A29 within the polyadenine tract with a cytosine residue (referred to as A29C). Other types of mutations involving placements of uridine were not explored as these were expected to yield partially base paired conformations. As with the WT construct, we observed no imino protons, indicating the absence of any detectable base pairing and secondary structure (Supplementary Figure S7).
The 2D C–H spectra for the A29C mutant remain highly dispersed, and the chemical shift perturbations relative to WT are clustered around the site of mutation (A28 and A30) (Figure 4A and Supplementary Figure S6). However, small but significant chemical shift perturbations relative to WT are also observed at more distant residues, including A27, A31, C33 and U34. These perturbations diminish when moving away from the center of the ssRNA and are basically absent in the highly flexible terminal residues (Figure 4A and Supplementary Figure S6). Such longer-range perturbations suggest that the mutation may have a long-range effect possibly by influencing the stacking interactions of several nucleobases. A perturbation to stacking interactions is also supported by distinct NOE connectivities in A29C, which show weakened cross peaks to C29, and new crosspeaks between A28 (H2) and A30 (H1′) that indicate C29 partially loops out to allow A28 to stack onto A30 (Supplementary Figure S7). The melting temperature of the mutant is reduced by ~5°C, and the base stacking energies are computed to be ~2 kcal/mol lower compared to WT, indicating that the mutation likely destabilizes the stacking interactions (Supplementary Figure S6).

Interestingly, many of the residues that experience chemical shift perturbations following the A29 to C29 mutation also exhibit a greater degree of dynamics as assessed by normalized resonance peak intensities in 2D C–H HSQC spectra (Supplementary Figure S6) and carbon relaxation data (R1 and R2) (Figure 4B and Supplementary Figure S8). In particular, severe line broadening consistent with a slow exchange process occurring at micro- to millisecond timescales manifesting as reduced resonance intensities in 2D spectra and higher R2 values is observed for C29 in the A29C mutant (Supplementary Figures S6 and S8). This is not observed for A29 in WT. Smaller but significant line broadening is also observed for residues A31, A32 and U34 (Supplementary Figure S8). This line broadening across several residues may reflect exchange between stacked and unstacked conformations. Higher intensities as well as reduced S2rel values are observed for residues A27 and A28, indicating a greater degree of fast pico- to nanosecond dynamics (Supplementary Figures S6 and S8). Note that the high R2 and weak signal intensity leads to a higher error in the R2/R1 measurements, particularly for C29.

Although the A29C RDCs are generally in good agreement with the WT RDCs, variations are observed for a number of residues (U26, A27, A31) that indicate differences in conformation and/or dynamic behavior (Figure 4C). Though an order tensor analysis of 13 RDCs shows best agreement with an A-form structure, the quality of the fit is not as good as that observed for WT (Q-factor = 8.77%, Figure 4D). The Szz direction measured for A29C when assuming an A-form structure deviates substantially from that predicted using PALES (~1°, Supplementary Figure S9). These data suggest that A29C deviates from an idealized A-form structure as compared to WT. These deviations may reflect static and/or dynamic bending about the C29 pivot point, possibly arising from looping out of this residue from the helical stack. Such a conformation is observed in the REMD simulations of A29C ~1% but not in WT (data not shown).

In general, the REMD simulations predict the NMR data measured for A29C with reduced quality to that noted for WT. Interestingly, the computed absolute S2 values indicate a global reduction in order for A29C, particularly for residues A27–C33 (Supplementary Figure S8), whereas NMR relaxation parameters between WT and A29C are more similar, suggesting comparable global order parameters. The REMD simulations reveal enhanced dynamics at C29 consistent with the NMR chemical exchange data. The REMD simulation also suggests increased dynamics at A32, which is not observed experimentally: although slightly reduced, the S2rel is within error of A29–A31 values (S2rel of 1) (Supplementary Figure S8). Computed RDCs agree reasonably with measured RDCs, although the C1′H1′ RDCs are opposite in sign as observed in the comparison between WT NMR and REMD-calculated RDCs. The Q-factor comparing the average REMD structure to measured RDCs is 70%; however, removal of A28 C8H8, A30 C2H2 and A30 C1′H1′ RDCs improves the Q-factor significantly. This improvement is observed only for the REMD structure (Figure 4D), indicating that these residues, localized about the mutation site, adopt non-A-form conformations and likely experience perturbations from the increased dynamics at C29. The difference in timescales between the REMD simulations and NMR may be another factor leading to the observed discrepancies. Nevertheless, MD and NMR data both indicate significant dynamics at the mutation site with perturbations extending toward the 3′ end of the ssRNA.
ssRNA tail conformation and dynamics optimized for ligand docking in queC aptamer

One of the main questions we set out to explore during the course of our studies was how the queC aptamer manages to efficiently bind its cognate ligand despite the small commitment time available in the kinetic switch and the large conformational space that may be available to a highly disordered ssRNA, which would have to search many competing conformations before arriving at the ligand bound pseudoknot conformation. Our study reveals that the ssRNA is not entirely disordered, but rather, has the character of a stacked A-form-like helical conformation which may effectively reduce the conformational search of the ssRNA, promoting efficient docking onto the hairpin to form the pseudoknot. Moreover, our study uncovers a greater degree of flexibility towards the terminal ends, particularly the 5'-end which forms the pivot point for docking the ssRNA tail onto the hairpin.

The NMR data clearly show the absence of any pre-existing tertiary interactions involving the ssRNA tail in the unbound queC aptamer domain. This together with our findings regarding the conformational behavior of the unbound ssRNA tail suggests the following model for ligand binding (Figure 5). In the absence of ligand, the ssRNA tail is disordered but on average forms an A-form helix-like conformation, which can efficiently explore conformational space about a highly flexible junction. The ligand may transiently form encounter complexes when the tail is close in space to the P1 hairpin, and possibly with the help of divalent ions such as calcium (27,73), triggering the necessary conformational changes required to form the pseudoknot and binding pocket. This finding is consistent with computational modeling of the ligand binding mechanism in which A-minor tertiary interactions form first, followed by pseudoknot formation (30) and may explain the fast ligand binding rate observed in the related F. nucleatum queC riboswitch (52). Our results, including the observation of greater dynamics in the mutant, provide a framework for more rigorous testing of this proposed model with future in vitro and in vivo studies.

CONCLUSION

Our study shows that ssRNA can exhibit complex conformational behavior, including variable levels of stacking and propensities to form an A-form helical conformation across the polynucleotide chain, and also, the ability to interrupt stacked residues by introducing sequence-specific kinks and/or distortions. While it has been known for some time that polyadenine stretches tend to stack and form helical conformations (13,14,16,18,37), the details of this helical geometry were difficult to decipher based solely on NOE-based NMR data. Our RDC measurements on the ssRNA, together with scalar coupling constant measurements, strongly suggest that the polyadenine tract forms an A-form-like conformation in the WT ssRNA. Our results also unveil dynamic complexity in ssRNA, including a gradual increase in disorder occurring towards the terminal ends that is reminiscent of unfolded polypeptide chains (74), and also, slower sequence-specific dynamics occurring at micro- to millisecond timescales that may involve transient stacking/unstacking motions that may result in kinking of the ssRNA. Altogether, our studies show that 'structured' ssRNA exhibits exquisite quality spectra and can be studied quantitatively using NMR-based structure and dynamics measurements.

The REMD simulations recapitulate many of the key features and trends observed based on the melting and NMR data, including the existence of stacking interactions that are weakened by the A29C mutation, the formation of helical geometry that may be kinked in A29C at the mutation site, and an increase in dynamic disorder towards the terminal ends and localized about C29 in the mutant. However, the REMD simulations showed weaker agreement with sugar RDCs or sugar conformation, particularly in the A29C mutant, and had increased dynamics compared to the NMR data. Prior studies on HIV-1 TAR RNA noted higher levels of dynamics in CHARMM simulations compared to NMR measurements (75). Our studies indicate that suboptimal base stacking energies may be a source of these excess dynamics. However, a quantitative assessment of the simulations requires the application of domain-elongation methods to rigorously decouple internal and overall motions, and make it possible to quantitatively predict NMR measurements (33,75–77). In addition, MD simulations that retain aspects of time are required to compare the rates of dynamics observed by relaxation and exchange broadening type measurements. The simplicity of ssRNA offers a much needed model system for such studies directed at rigorously examining currently used nucleic acid force fields.

Finally, our results suggest that the conformational properties of the ssRNA tail are optimized to allow the queC riboswitch to efficiently bind ligands within the short commitment time available to this kinetic switch. In particular, the pre-stacked ssRNA tail can efficiently rotate about a flexible hinge against the hairpin loop, and explore conformational space efficiently for rapid ligand binding. This pre-stacking about dynamic hinges may be a
general feature of many ssRNAs that can play different architectural roles in a variety of RNA contexts.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures S1–S9.

ACKNOWLEDGEMENTS

C.D.E. and H.M.A. thank members of the Al-Hashimi lab for insightful discussions and Dr Alex Kurochkin for his expertise and maintenance of the NMR instruments. The authors gratefully acknowledge the Michigan Economic Development Cooperation and the Michigan Technology Tri-Corridor for support in the purchase of the 600 MHz spectrometer.

FUNDING

National Science Foundation (0644278 to H.M.A., PHY0216576 to C.L.B.); National Institutes of Health National Science Foundation (0644278 to H.M.A., FUNDING the 600 MHz spectrometer. Technology Tri-Corridor for support in the purchase of Development Cooperation and the Michigan authors gratefully acknowledge the Michigan Economic Development and maintenance of the NMR instruments. The C.D.E. and H.M.A. thank members of the Al-Hashimi lab

REFERENCES


