BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Crystallographic and Computational Analyses of AUUCU Repeating RNA That Causes Spinocerebellar Ataxia Type 10 (SCA10)

HaJeung Park,*[,†] Àlex L. González,[‖] Ilyas Yildirim,[⊥] Tuan Tran,[§] Jeremy R. Lohman,[§] Pengfei Fang,[‡] Min Guo,[‡] and Matthew D. Disney*[,§]

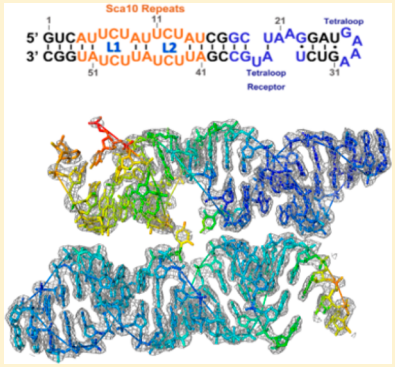
[†]Translational Research Institute, [‡]The Department of Cancer Biology, and [§]The Department of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way, Jupiter, Florida 33458, United States

[‖]Grup d'Enginyeria Molecular (GEM), Institut Químic de Sarrià (IQS)-Universitat Ramon Llull (URL), Barcelona 08017, Spain

[⊥]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Supporting Information

ABSTRACT: Spinocerebellar ataxia type 10 (SCA10) is caused by a pentanucleotide repeat expansion of r(AUUCU) within intron 9 of the *ATXN10* pre-mRNA. The RNA causes disease by a gain-of-function mechanism in which it inactivates proteins involved in RNA biogenesis. Spectroscopic studies showed that r(AUUCU) repeats form a hairpin structure; however, there were no high-resolution structural models prior to this work. Herein, we report the first crystal structure of model r(AUUCU) repeats refined to 2.8 Å and analysis of the structure via molecular dynamics simulations. The r(AUUCU) tracts adopt an overall A-form geometry in which 3 × 3 nucleotide $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ internal loops are closed by AU pairs. Helical parameters of the refined structure as well as the corresponding electron density map on the crystallographic model reflect dynamic features of the internal loop. The computational analyses captured dynamic motion of the loop closing pairs, which can form single-stranded conformations with relatively low energies. Overall, the results presented here suggest the possibility for r(AUUCU) repeats to form metastable A-from structures, which can rearrange into single-stranded conformations and attract proteins such as heterogeneous nuclear ribonucleoprotein K (hnRNP K). The information presented here may aid in the rational design of therapeutics targeting this RNA.



RNA repeat expansions cause various neuromuscular diseases, including spinocerebellar ataxia type 10 (SCA10), myotonic dystrophy (DM), Huntington's disease (HD), and frontotemporal dementia/amyotrophic lateral sclerosis (FTD/ALS). Repeat modules are generally three to six nucleotides in length.[1] For example, DM1 and HD are caused by triplet repeats (CUG and CAG, respectively) while longer repeats cause SCA10 (AUUCU). Repeat length scales with disease severity. For example, in SCA10, healthy individuals typically have <50 repeats while those afflicted with disease have up to ~5000 repeats.[2]

Studies have shown that the pathology of repeat expansion disorders is predominantly caused by two modes of RNA toxicity. Repeats bind to and sequester proteins involved in RNA biogenesis, leading to downstream defects in RNA processing, termed RNA gain of function. Also, expanded repeats initiate translation without the use of a start codon. Termed repeat-associated non-ATG (RAN) translation, this mode produces toxic homopolymeric proteins that accumulate as inclusion bodies and induce apoptosis.[3,4] In SCA10, r(AUUCU)^{exp} sequesters heterogeneous nuclear ribonucleoprotein K (hnRNP K), inducing translocation of protein kinase Cδ to mitochondria and caspase-3-mediated apoptosis of neuronal cells via RNA gain of function.[2]

Structural studies have been reported for various repeating transcripts[5−7] and revealed common structural features. For example, they adopt an overall A-form geometry, with variations in base pair and helical parameters. It is possible that repeating RNAs with longer repeat modules (>3) share similar features; however, high-resolution information for these RNAs is scarce. A biophysical study by Handa et al. suggests r(AUUCU)$_9$ forms a structured A-form helix via circular dichroism (CD) and nuclear magnetic resonance (NMR) analysis.[8] Their NMR studies revealed evidence of A-U and U-U base pairing, suggesting that r(AUUCU) repeats harbor 3 × 3 nucleotide $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ internal loops with two U-U noncanonical pairs and one C-C noncanonical pair.[8]

To capture structural characteristics of r(AUUCU) repeats, we have determined a crystal structure of a model RNA containing two copies of $^{5'}$AUUCU$^{3'}$/$^{3'}$UCUUA$^{5'}$ and thoroughly analyzed the dynamics of this structure with molecular dynamics (MD) simulations. The results indicate r(AUUCU) repeats form a metastable A-form RNA, and the dynamic characteristic is attributed to the internal

Received: May 20, 2015
Revised: June 2, 2015
Published: June 3, 2015

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

$^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ loop pairs. Overall, the results presented here provide structural evidence of the pathogenic mechanism of SCA10 caused by repeat expansion of r(AUUCU). This structure may also provide valuable information to guide the design of therapeutic modalities that target this RNA to ameliorate the disease.

MATERIALS AND METHODS

RNA Synthesis and Purification. A single-stranded DNA template for the AUUCU construct was purchased from Integrated DNA Technologies, Inc. (IDT). A double-stranded template suitable for *in vitro* transcription was generated by polymerase chain reaction as previously described[9] by using the following primers: forward primer 5′-d(**CTAATACGACTCACTATA**GCCCCTGCCTGCCTGCAGCTAAGGATG) (where bold nucleotides indicate a T7 RNA polymerase promoter) and reverse primer 5′-d(GCCCAGGCAGGCAGGCAGCATAGACTTTCATCCTTAGCTGCAGGCAGGCAG).[9] Transcription of the corresponding RNA was completed by runoff transcription with T7 RNA polymerase as previously described followed by purification by denaturing polyacrylamide gel electrophoresis.[10]

Crystallization. The RNA sample was dissolved in distilled water to afford a 1 mM solution and was folded by heating at 95 °C for 2 min and then cooled to room temperature. Screening of crystallization conditions was completed with a Nucleix Suite (Qiagen) using a Gryphon nanodrop crystallization robot (Art Robinson) in a 96-well sitting drop format. Large, reproducible crystals were grown in hanging drops with a 2 μL drop size at room temperature using a precipitant containing 100 mM ammonium acetate, 5 mM magnesium sulfate, 50 mM MES (pH 6.0), and 600 mM NaCl. The crystals were then mounted on a free mounting system (FMS) and dehydrated to a relative humidity of 75% to improve diffraction quality. Dehydrated crystals were coated with perfluoropolyether oil (Hampton Research) and flash-frozen in liquid nitrogen for synchrotron data collection.

Data Collection, Phasing, Refinement, and Analysis. X-ray diffraction data of the SCA10 repeat containing RNA were collected at beamline ID-G of LS-CAT in APS using a Mar300 CCD detector. The highest-quality crystal diffracted to Bragg spacings of 2.75 Å. The data set was then integrated and scaled using iMOSFLM.[11] Preliminary analysis of the diffraction data showed the crystal belongs to tetragonal space group *P*422. Detailed analysis, however, revealed that the crystal was merohedrally twinned. Thus, molecular replacement (MR) was conducted in the *P*4 space group using a tetraloop/tetraloop receptor core domain of Protein Data Bank (PDB) entry 4FNJ and Phaser in the Phenix suite.[12] The best MR solution with a log-likelihood-gain (LLG) gain and a translation function *Z*-score (TFZ) of 116.3 and 9.2, respectively, was obtained in space group $P4_1$. The electron density map of the MR solution showed base pair steps outside of the search model. Rigid body refinement against the MR solution also showed R_{work} and R_{free} values of 35 and 38%, respectively. Missing RNA bases were manually modeled using Coot.[13] Crystallographic refinement of the modeled RNA was performed by using Phenix (phenix.refine) and CCP4 (refmac) suites applying noncrystallographic symmetry (NCS) restraints along with a twin refinement protocol. The final R_{work} and R_{free} values of the structure were 17.8 and 22.4%, respectively. Figures were prepared with PyMol.[14] Helical parameter calculations were completed by 3DNA.[15] Data collection and refinement statistics are summarized in Table S-5 of the Supporting Information.

Model System Preparation. A model r(AUUCU) structure containing one 3 × 3 $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ internal loop was prepared by homology modeling. A symmetric system was designed with the sequence of r(CGAUUCUAUCG)$_2$, where CG and GC flanking pairs were included to increase the overall structural stability in the MD simulations. The system was prepared with ModeRNA[16] by extracting the (AUUCUAU)$_2$ fragment from the crystal structure and adding standard CG and GC base pairs from the rnaDB2005 database fragment library.[17] The system was neutralized with 20 Na^+ ions[18] and solvated with 5095 TIP3P water molecules[19] in a 12 Å truncated octahedral box. The AMBER force field[20] with revised χ[21] and α/γ[22]torsional parameters was used in all the MD simulations.

Replica Exchange Molecular Dynamics (REMD) Simulation. The system was minimized in two steps. First, all residues except solvent were held fixed with a restraint force of 500 kcal mol^{-1} Å^{-2} and minimized with 2500 steepest descent steps followed by 2500 conjugate gradient steps. A second minimization step was performed without positional restraints using 10000 steepest descent and 10000 conjugate gradient steps. REMD simulation was conducted under periodic boundary conditions, using 40 replicas at a constant volume. The temperature range was determined using the parallel tempering temperature predictor.[23] A temperature range from 272 to 363 K was spanned with uniform ratios for exchange of ∼30% between neighboring replicas. Each replica was slowly heated to its corresponding replica temperature in 150 ps while the solute was constrained with a force gradient from 8.0 to 0 kcal mol^{-1} Å^{-2}. Langevin dynamics with a collision frequency of 1 ps^{-1} was used. A 20 ps pressure equilibration step was then applied with isotropic scaling at 1 atm. Production runs were conducted at a constant volume using an exchange frequency of 2 ps. Chemical bonds involving hydrogen atoms were constrained with the SHAKE algorithm,[24] which allowed an integration step of 2 fs in the production runs. Particle mesh Ewald (PME)[25,26] was used in all calculations with a 9 Å long-range cutoff. A total of 5.03 μs of accumulated simulation time was obtained.

Steered Molecular Dynamics (SMD). To describe the instability of the RNA internal loop, we performed three independent SMD experiments by pulling away sequentially the U-U and C-C pairs. SMD simulations were performed with the same protocol described for the equilibration step and initiated with the final structure obtained from the equilibration step. First, a combination of velocities ranging from 0.01 and 1.0 Å/ns and a spring constant ranging from 5 to 20 kcal mol^{-1} Å^{-2} were tested using a 3^2 factorial design with 10 repeats. The pulling force was applied to the center of mass (COM) of the O2, O4, and N3 atoms of the U-U pairs, and O2, N3, and N4 atoms of the C-C pair, until a separation of 6 Å was achieved. Once the variables were optimized, a total of 25 successive SMD pulling experiments per pair were conducted with a constant velocity of 0.14 Å/ns and a spring constant of 10 kcal mol^{-1} Å^{-2}.

Simulation Analysis. Each replica window was analyzed with the *cpptraj* module (Amber 14).[20] Because the force field was parametrized at 300 K, only the replica at this temperature was subjected to further analysis. Structures were clustered using the average-linkage hierarchical agglomerative method with a distance cutoff ε of 3 Å. Symmetry-corrected root-mean-

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

square deviation (rmsd) clustering was performed on a subset of atoms using the Amber mask syntax (:1−22@P,C3′,C4′,C5′,O3′,O5′). All RNA structural parameters were calculated using the 3DNA suite[15] for the structures extracted from the MD trajectory at 20 ps intervals. Stiffness constants were computed according to the method of Lankas et al.[27] Once the covariance matrix (**C**) of the helical parameters was generated, the stiffness matrix (**K**) was computed with the following equation (eq 1):

$$\mathbf{K} = k_B T \times \mathbf{C}^{-1} \quad (1)$$

The potential of mean force (PMF) as a function of slide and helical twist was constructed as described previously for other biomolecular systems[28] by using observed and reference probability distributions. Grid inhomogeneous solvation theory (GIST)[29] analysis of solvent was performed using the algorithm incorporated in Amber 14.[20] Prior to the analysis, Na^+ atoms were removed from the system because GIST considers all nonsolute atoms as solvent molecules. The grid was centered at (27, 27, 29) and dimensioned to 20 Å × 20 Å × 40 Å with a spacing of 0.50 Å. The output was analyzed with VMD.[30]

RESULTS AND DISCUSSION

Overall Crystal Structure. RNA molecules tend to produce poor-quality crystals because of limited intermolecular

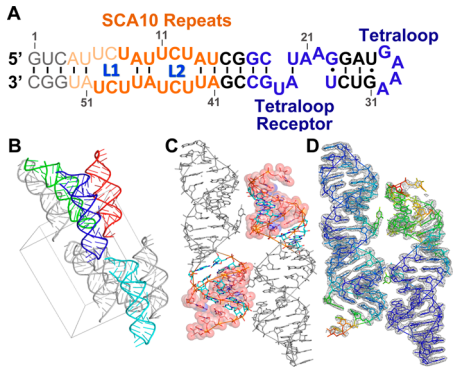


Figure 1. Structure of r(AUUCU) repeat-containing RNA. (A) Secondary structure of the crystallized RNA. The two loops are labeled L1 and L2. Bold fonts represent the modeled region. (B) Crystal packing environment. Two molecules, red and blue, are found in an asymmetric unit. Red and cyan molecules are coaxially aligned. Blue and green molecules interact through the tetraloop/tetraloop receptor interaction. (C) Overall structure of the asymmetric unit. The UCU loops are shown as orange sticks with a transparent sphere. AU closing pairs are colored cyan; remaining regions of the RNAs are shown as gray sticks. (D) Electron density colored gray at a contour level of 1.2σ. RNAs are colored according to temperature factor by rainbow colors from blue (low) to red (high).

contacts. The tetraloop/tetraloop receptor motif, which provides additional contacts, has been utilized to promote crystallization of various RNAs.[31,32] We employed this motif in our crystallization construct that contains two model SCA10 repeats (Figure 1A). The base pair alignment of the repeats was designed using the method of Handa et al.,[8] and additional G-C or G-U base pairs were included on both sides of the repeats to provide further stabilization.[8]

A molecular replacement search revealed two independent RNA molecules in the asymmetric unit (chains A and B) with a solvent content of 59% (Figure 1B,C). Although the packing

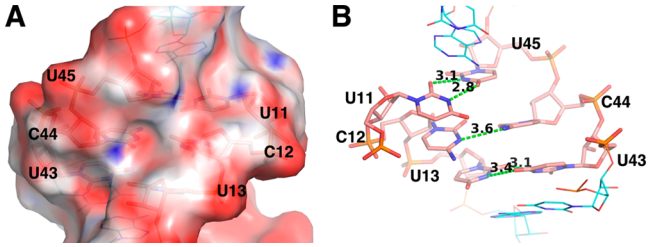


Figure 2. Overview of $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ loop in L2 of chain A. (A) Charge distribution of the minor groove at the loop shown as a transparent surface model. The electrostatic potential was calculated using APBS and contoured at ±25 kT/e. (B) Loop represented as a stick model. Hydrogen bonds are represented as green dashed lines.

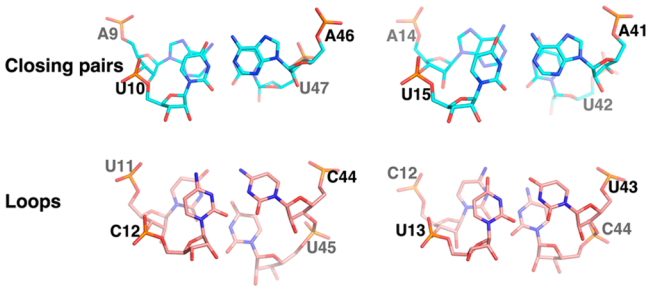


Figure 3. Base stacking of $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ loops and $^{5'}$AU$^{3'}$/$^{3'}$UA$^{5'}$ closing pairs in chain A.

environment is different, the two RNA chains are essentially identical as judged by the rmsd between them (1 and 0.65 Å for the whole molecule and SCA10 repeat region, respectively). The *B* factor for both RNAs is lower at the tetraloop region and increases throughout the repeat and additional base pairs, which suggests that the end of the hairpin stem is labile (Figure 1D). G1−C7 and A51−C55 of both chains were not modeled because of a lack of traceable electron density (Figure 1D). The disorder observed at the stem ends is due to multiple factors, including inherent instability of the SCA10 loop and crystal contacts (Figure 1D).

The two asymmetric RNA molecules are antiparallel to each other (Figure 1C), and the interaction is stabilized by multiple hydrogen bonds (Figure S-1 of the Supporting Information). The tetraloop/tetraloop receptor interaction, which formed between crystallographic symmetry mates, is the most significant packing interaction and is invariant compared to other tetraloop/tetraloop receptor structures. Although there is a coaxial "end to end" alignment of symmetry-related molecules, no base stacking between them is observed because of disorder in the stem ends (Figure 1B).

Although 3DNA analyses display the average base pair step and local helical parameters within the range of A-form RNA (Table S-2 of the Supporting Information), individual steps have wide variations indicating irregularities. Not surprisingly, the largest deviations occur within the loops. Helical twist values at the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ loops show increased average values (39°) relative to those of A-form RNA. Large variations in inclination and tip indicate disrupted base stacking interactions within the loop.

Internal $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ Loop Structures. A total of six potential U-U base pairs and two potential C-C base pairs are found in the two RNA molecules. The six carboxyl oxygens from the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ loops are all concentrated in the minor groove. This makes the minor groove of the SCA10

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

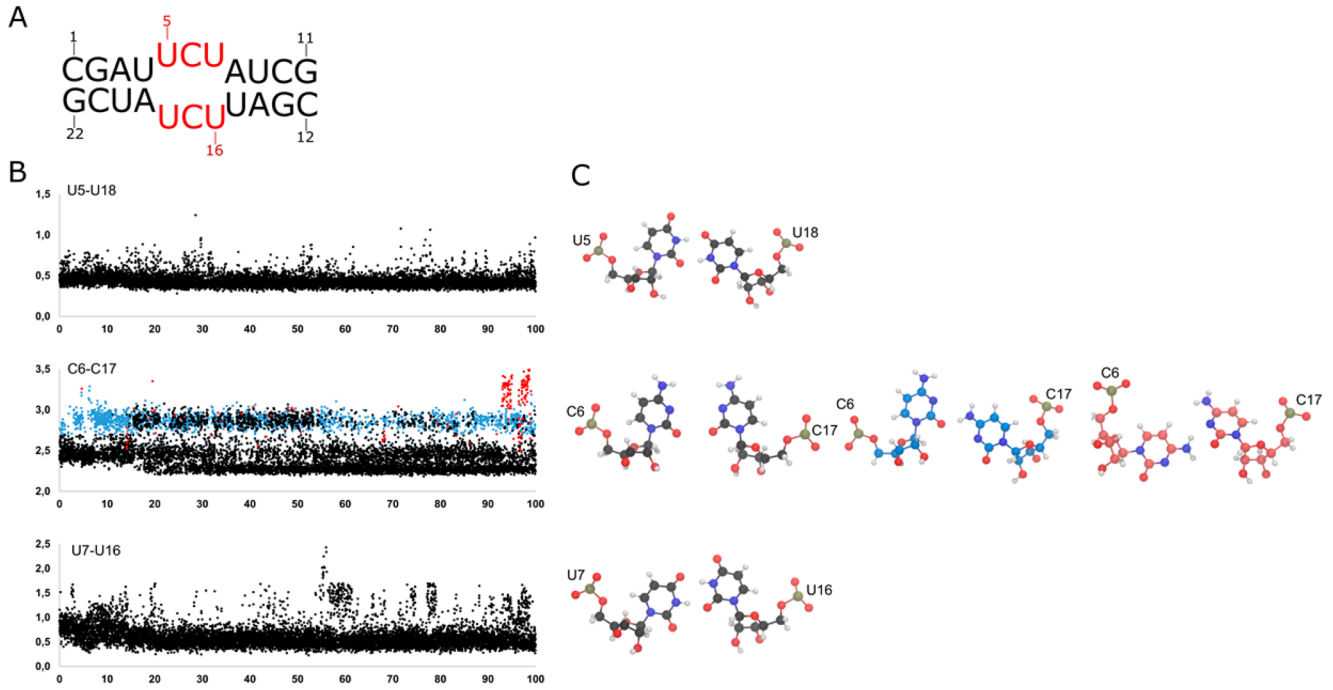


Figure 4. (A) Sequence of the model r(AUUCU) used in the computational studies. (B) Symmetry-corrected rmsd analysis of the U5-U18, C6-C17, and U7-U16 pairs. (C) Clustered conformations of the internal loop base pairs. Each data point in panel B corresponds to a cluster that is represented by a color code displayed in panel C. Only conformations with a >10% residence time were selected for the analysis.

repeats the most charge dense region among repeat expansion RNAs, which may have implications for the "druggability" of this RNA (Figure 2A).

The $^{5'}$AU$^{3'}$/$^{3'}$UA$^{5'}$ loop closing pairs have standard Watson–Crick geometries. In both chains, the U-U base pairs in L2 have a two-hydrogen bond geometry while the U8-U48 pair defined in L1 has a one-hydrogen bond geometry (Figure 2B and Figure S-2 of the Supporting Information). Interestingly, U-U internal loops in r(CUG) repeats have significant differences compared to the U-U pair geometries observed in the r(AUUCU) repeats. In particular, the U-U pairs in r(AUUCU) have larger propeller and buckle deviations, and the repeat overall has variations of roll and tilt larger than those of r(CUG) (Tables S-2 and S-3 of the Supporting Information).[6]

The central C-C base pair of chain A shows one hydrogen bond geometry, whereas that of chain B appears to be dynamic with poorly defined electron densities for both bases as well as parts of the phosphate backbone in C44 (Figures S-2 and S-3 of the Supporting Information). One hydrogen bond geometry of the C-C pair is similar to that previously observed in r(CCG) repeats.[7] Local base pair parameters propeller and buckle of the C-C pair are also comparable to those of r(CCG) repeats. However, the opening (36°) is larger than the one observed in r(CCG) repeats because C44 opens toward the major groove (Table S-3 of the Supporting Information).[7] Unusually short C1′–C1′ distances, ranging from 7.8 to 9 Å, found in $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ of L2 appear to contribute conformational deviation of U-U and C-C pairs from those found in r(CUG) and r(CCG) repeats (Figure S-2 of the Supporting Information).

As evidenced by increased helical twist values, stacking interactions among the bases throughout $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ were marginal to nonexistent. In contrast, the loop closing base pairs and their nearest neighbors ($^{5'}$AU$^{3'}$/$^{3'}$UA$^{5'}$) show extensive overlap, which is expected for adjacent Watson–Crick pairs (Figure 3). Some stacking is observed between the A of the closing pair and the neighboring U in the loop. However, this stacking interaction appears to be suboptimal because of a lack of planarity between the bases (Figure S-4 of the Supporting Information).

Dynamics of the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ Model System. Understanding the dynamics of biomolecules is often a computationally challenging process. Moreover, classical MD protocols are subject to sampling limitations, which are hard to overcome. The REMD method simulates several replicas at different temperatures with a certain probability of swapping periodically from one temperature replica to another to improve thermodynamic sampling in a simple and effective manner.[33] To achieve thorough sampling of the 3 × 3 internal loop conformational space, we employed REMD simulations with the hope that higher temperatures would destabilize the loop and provide possibilities of analyzing the dynamic features of this 3 × 3 RNA internal loop. A total of 40 replicas spanning a temperature range close to 300 K were used to study the intrinsic dynamics of the 3 × 3 internal loop. The initial coordinate of the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ model system is identical to the central fragment of the L1 crystal structure except for additional C-G and G-C pairs flanking the internal loop to enhance the structural stability.

The rmsd analysis was completed for each U-U and C-C pair contained in the 300 K replica individually (Figure 4). The rmsd values were plotted with respect to the average conformation, and each color represents an individual clustered conformation with a residence time that was >10% of the total simulation (Figure 4C). In good agreement with the crystal structure, the two U-U pairs in the model stay in a two-hydrogen bond conformation along the 300 K trajectory. However, some differences are observed in terms of stacking. For instance, the stacking surface area of $^{5'}$U4U5$^{3'}$/$^{3'}$A19U18$^{5'}$

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851–3859

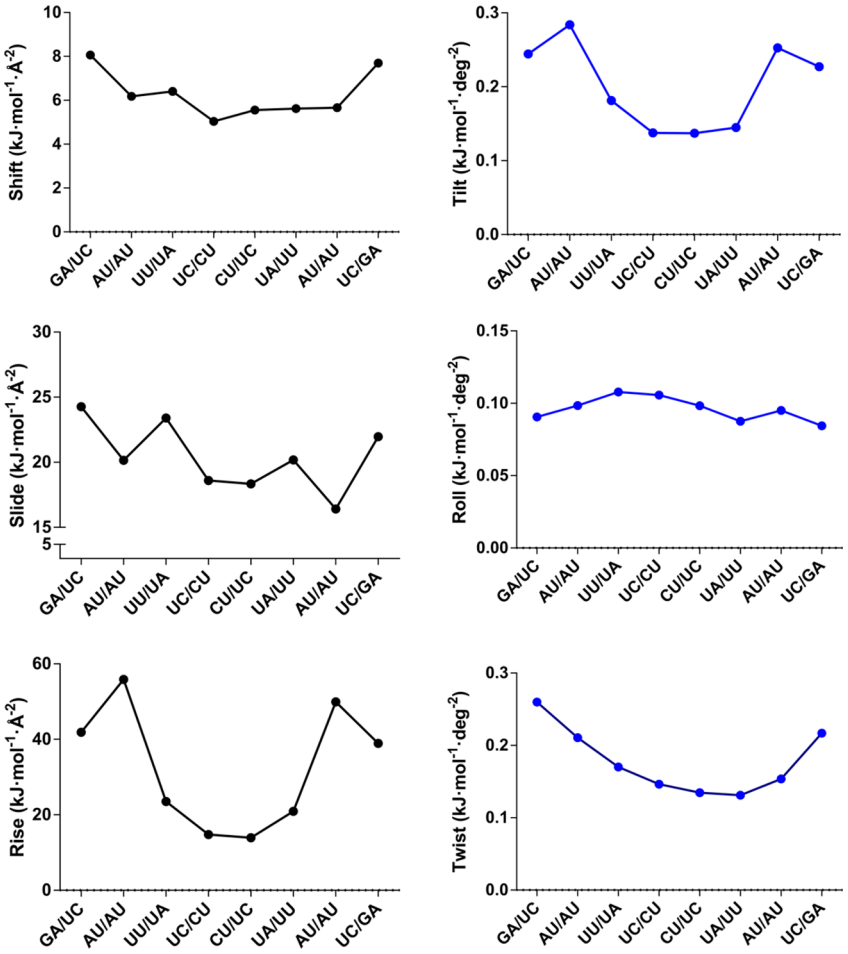


Figure 5. Stiffness force constants of the helical space (shift, slide, rise, tilt, roll, and twist) for the different nucleotides contained in the [GAUUCUAU] fragment. The most flexible regions correspond to those with associated low stiffness force constants. Overall, flexibility is especially pronounced in the [UCU] region as observed in the slide, rise, and tilt stiffness profiles.

is 0.56 Å^2 larger than that of $^{5'}$U7A8$^{3'}$/$^{3'}$U16U15$^{5'}$, so the stacking contribution is greatly diminished in the latter case.

The C-C pair is clearly the most dynamic mismatch, whereby a large number of conformations are equally distributed during the simulation. Of these, a total of three different relevant conformations were found for the C-C pair while the U5-U18 and U7-U16 pairs visited only one relevant conformation (Figure 4B). Interestingly, the C-C pair showed a zero-hydrogen bond conformation during 34% of the simulation (Figure 4C), while two different one-hydrogen bond conformations were present during 64% of the total time (Figure 4C). This dynamic behavior of the C-C pair during the simulation is in line with the mainly one-hydrogen bond state in the crystal structure as well as the poorly defined electron density. Previous studies of r(CUG) repeats defined the dynamic nature of U-U pairs; thus, it is not surprising the poor stacking of the C-C pairs with the flanking uridines and the concurrent C-C dynamic behavior. Several transitions between *cis* and *trans* Watson−Crick conformations of the C-C pair are observed along the simulations that altogether clearly affect the stiffness of the entire internal loop. The C1′−C1′ distances of the loop nucleotides remain in the range of 8.8−8.9 Å on average, as observed in the crystal structure, so no major backbone deviations were produced during the simulation.

The influence of the overall flexibility of r(AUUCU) by the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ internal loop was analyzed using both stiffness constants of a helical space defined by Lankas et al. and the helical parameters of 3DNA.[27] Most of the base pair and helical parameters are close to those in the crystal structure except for buckle, propeller, and slide, which showed significant deviations (Table S-6 of the Supporting Information). However, the fluctuations are considered to be small, so the effect of the Jacobian factor was neglected in the stiffness analysis. In general terms, the internal loop presents the lowest stiffness constant values along the structure (Figure 5), especially rise and tilt, which can be explained by the lack of optimal stacking into this region. Altogether, the labile stacking capability of the internal loop and its large charge density may confer to this RNA a unique binding region for small molecules or other ligands such as ions or proteins.

Stabilization of C-C Pairs by Water-Mediated Hydrogen Bonds. Although the C-C pair remains mainly in a one-hydrogen bond conformation, the MD results suggest the presence of a stable zero-hydrogen bond state. Moreover, two-dimensional potential of mean force analysis (2D-PMF) revealed that the zero-hydrogen bond state (C1 in Figure 4C) is the most stable among the clustered conformations (Figure 6). The PMF map was constructed using slide and helical twist as variables because of the high variability of these parameters in the crystal structure and during the simulation. The C1 coordinates are close to local minima (−1.9, −23) and deviated from the coordinates of the crystal structure in chain A

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

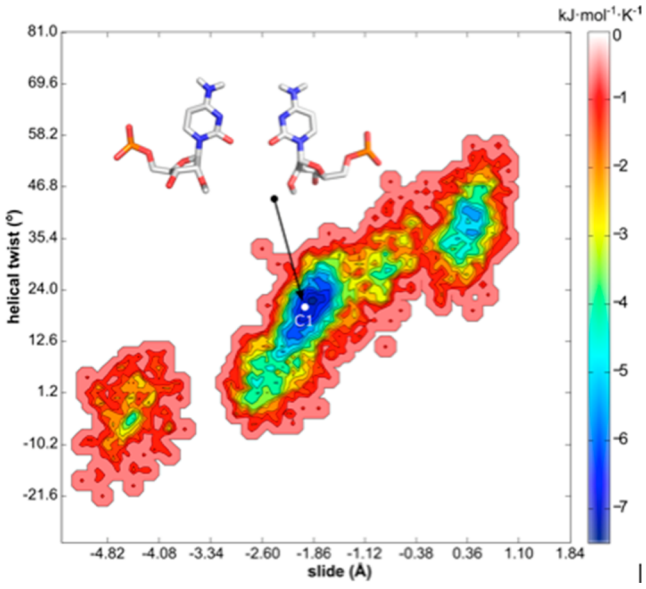


Figure 6. 2D-PMF surface showing the exploration of the C-C pair conformations along the MD simulation. The *x*- and *y*-axes represent the slide and helical twist values of the noncanonical CC pair, respectively. The zero-hydrogen bond CC conformation (C1 in Figure 4C) is displayed as a white dot in the figure that corresponds to a local minimum conformation.

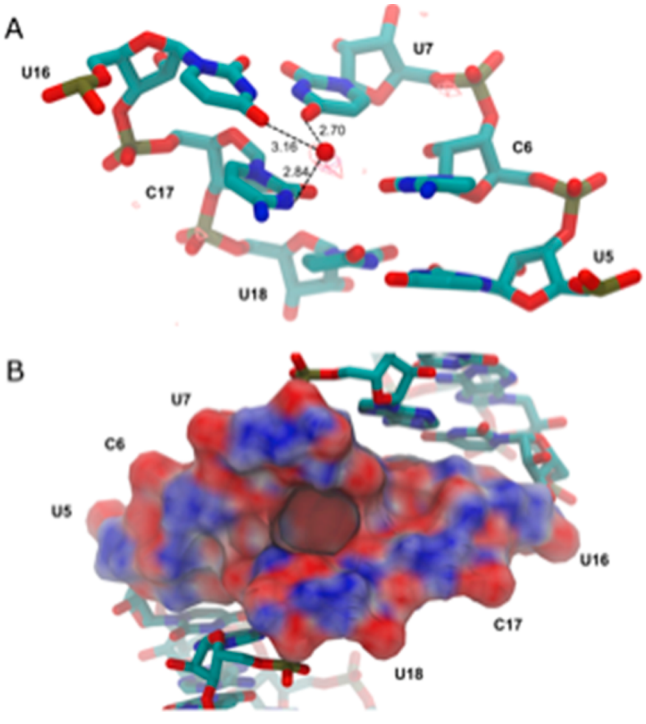


Figure 7. (A) Stick model representation of the internal loop and potential hydrogen bond with an explicit water molecule. GIST-computed water probability contours are represented as a grid. (B) Charge distribution of the minor groove at the loop shown as a surface model. The electrostatic potential was calculated using APBS and contoured at ±10 kT/e.

(−1.1, 36.2) and in chain B (−0.9, 40.1). This indicates an additional force should stabilize C1 to keep the zero-hydrogen bond conformation as a preferred conformation among all the sampled phase space. A closer inspection of this state suggests that the hydration pattern of the noncanonical pair plays an essential role in its stabilization. Effects of the structure and thermodynamics of the solvent were investigated using GIST (grid inhomogeneous solvation theory), which analyzes the three-dimensional water density around a certain area termed voxels with associated solvent properties.[29] Indeed, a highly occupied volume appears to be close to the C-C pair. Surprisingly, the water molecule occupies a space among C17, U7, and U16 and coordinates a potential hydrogen bond between atom O4 of U7 and atom N3 of C17 (Figure 7A). At the same time, the U7-U16 pair inclines toward the C-C pair and breaks the hydrogen bonding pattern, yielding a highly negatively charged cavity in the RNA minor groove (Figure 7B). This phenomenon is produced repeatedly and approximately one-third of the time in the MD simulation.

Inherent Instability of the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ Internal Loop. Our crystallographic and computational results presented here corroborate the findings of Handa et al. that the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ internal loop is metastable.[8] Therefore, it has been hypothesized that RNA unwinding can start through internal loop destabilization. Under this premise, we investigated the stability of each pair individually within the 3 × 3 internal loop context using steered molecular dynamics (SMD). SMD is a powerful technique used to gain insights into several mechanisms, such as unfolding pathways of macromolecules. This procedure exploits nonequilibrium sampling by applying time-dependent biasing forces to guide the system transformation. During SMD experiments, several pulls are simulated in one direction. Previous SMD studies have been proven to be successful in computing free energy profiles on realistic biomolecular systems.[34,35]

Herein, we conducted 25 successive SMD pullings per pair with a constant velocity of 0.14 Å/ns and a spring constant of 10 kcal mol^{-1} Å^{-2} and computed force of pulling and cumulative work profile for each system (Figure 8A). Each plot represents the work required to pull each pair from the bonded state with an ∼2.5−3.5 Å COM distance to 6.0 Å where hydrogen binding interactions are no longer operative. Pulling force peaked at ∼4 Å for U7-U16 and U5-U18 and 5 Å for C6-C17. The mean work performed in each pair clearly reflects the fact that the central C-C pair is the most probable starting point of unwinding (W = 7.88 kcal/mol). Indeed, the U-U pulling experiments yield very close work values (12.21 and 12.13 kcal/mol for U5-U18 and U7-U16 pairs, respectively).

Interaction energies of noncanonical pairs vanished smoothly and had relatively marginal impact over the conformation of the neighboring pairs; thus, the overall RNA structure remained intact during all the pulling experiments. While both U-U pairs did not change their orientation during the steering process, the C-C pair experienced a rotation along the α torsion that flipped C6 out of the RNA duplex. The rotational movement of the C6 base was consistently observed in all SMD trials. Therefore, we concluded that this should be the lowest-energy pathway for breaking the C-C pair. The thermodynamic stability of an RNA structure depends not only on base pairing but also on stable stacking interactions. Therefore, the observed unstacking of a cytosine base can further destabilize the $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ internal loop.

Biological Significance of the Structural Features. The crystallographic results herein confirm that the r(AUUCU) repeats indeed form an overall A-form geometry similar to that of other repeat expansion disease RNAs. However, unlike

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

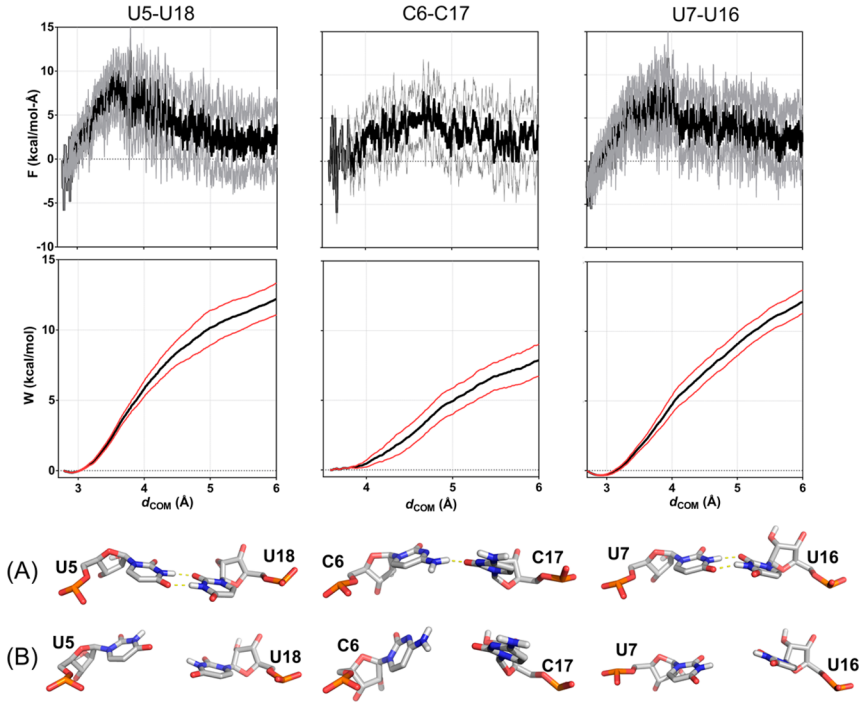


Figure 8. Pulling force and cumulative work distribution profiles vs distance from the center of mass (d_{COM}). For each base pair, initial (A) and final (B) representative conformations are depicted. The mean force and work are represented as black lines; the standard deviations of force and work profiles are depicted as gray and red lines, respectively.

trinucleotide repeats, which harbor two G-C closing pairs for every 1 × 1 nucleotide loop, the SCA10 repeat has two A-U pairs for every 3 × 3 nucleotide loop. Thus, the SCA10 repeat is likely less stable than r(CAG), r(CUG), or r(CGG) repeats because of (i) weaker closing base pairs (A-U vs G-C) and (ii) decreased loop stability. Thermodynamic studies of RNA duplexes containing two to four CNG repeats (where N represents any nucleotide) showed that C-C and U-U mismatches are less stable than A-A and G-G.[36] In SCA10, each $^{5'}$UCU$^{3'}$/$^{3'}$UCU$^{5'}$ forms two U-U mismatches and one C-C mismatch. Crystallographically, such instability is indirectly observed by the high temperature factors (and weak electron density map) (Figure S-3 of the Supporting Information). The accompanied computational simulations also explored the nature of the conformational flexibilities of the C-C and U-U mismatches. In addition, computational simulation observed that the central C-C mismatches spent significant simulation time in the zero-hydrogen bond state through stabilization by water-mediated hydrogen bonds near C17, U7, and U16.

Studies showed that SCA10 is caused when r(AUUCU)^{exp} binds and inactivates hnRNP K.[2] hnRNP K binds RNA through a KH domain, which is highly abundant particularly in proteins that regulate gene expression. Crystal structures have revealed that KH domains bind single-stranded RNA by forming favorable interactions with at least four nucleotides, N1−N4.[37] Interestingly, adenine and cytosine are preferred at N3 as they can most favorably hydrogen bond with a conserved hydrophobic residue in β-strand 2 (with the amino acid residue's carbonyl and a backbone amide oxygen).[37] In particular, structural studies of the KH3 domain of hnRNP K with single-stranded DNAs with cytosine at N3 determined the binding specificity through hydrogen bond interactions between O2 and N4 of cytosine and NH2 of Arg414 and the backbone carbonyl of Ile423, respectively. Additional bipartite hydrogen interaction between N3 of cytosine and the backbone amide of Ile423/NH2 of Arg414 is mediated through a water molecule.[38,39] It is conceivable that the other KH domains (KH1 and KH2) of hnRNP K have similar binding specificity for the single-stranded nucleic acids considering the high degree of sequence conservation among the three KH domains.[39] Our SMD study indicates the central C-C mismatch is particularly vulnerable to strand opening, which is likely done through rotational movement along the α torsion angle. Like base flipping of many DNA repair enzymes, hnRNP K may sample the weak C-C mismatch pair and (or) initiate unzipping of r(AUUCU)^{exp} by flipping out a base of the pair, i.e., C6 (Figure 4A). The unzipping event can be facilitated further by the disrupted base stacking as well as inherently weak interactions in the internal loop (as compared to a fully paired RNA). In such a scenario, the most likely register for r(AUUCU) binding is (N1−N4, where the fifth nucleotide of the repeat is denoted in parentheses) (A)^{1}U^{2}U^{3}C^{4}U. Taken together, the structural information presented herein provides a framework to allow the design of small molecule therapeutic agents.

■ ASSOCIATED CONTENT

Supporting Information

GIST input script, refinement and data collection statistics, and further analysis of the X-ray crystal structure. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.5b00551.

Accession Codes

The structure was deposited in the Protein Data Bank as entry 5BTM.

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

AUTHOR INFORMATION

Corresponding Authors

*E-mail: disney@scripps.edu. Phone: (561) 228-2203. Fax: (561) 228-2147.

*E-mail: hajpark@scripps.edu. Phone: (561) 228-2121. Fax: (561) 228-3067.

Funding

This work was funded by the National Institutes of Health (Grant R01-GM079235 to M.D.D.) and by The Scripps Research Institute.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Jessica Childs-Disney for discussions and critical review of the manuscript and the staff at the LS-CAT, APS, for synchrotron support.

ABBREVIATIONS

r(AUUCU)^{exp}, expanded r(AUUCU) repeat; DM, myotonic dystrophy; hnRNP K, heterogeneous nuclear ribonucleoprotein K; SCA10, spinocerebellar ataxia type 10.

REFERENCES

(1) Li, L. B., and Bonini, N. M. (2010) Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends Neurosci. 33*, 292−298.

(2) White, M. C., Gao, R., Xu, W., Mandal, S. M., Lim, J. G., Hazra, T. K., Wakamiya, M., Edwards, S. F., Raskin, S., Teive, H. A., Zoghbi, H. Y., Sarkar, P. S., and Ashizawa, T. (2010) Inactivation of hnRNP K by expanded intronic AUUCU repeat induces apoptosis via translocation of PKCδ to mitochondria in spinocerebellar ataxia 10. *PLoS Genet. 6*, e1000984.

(3) Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A. 108*, 260−265.

(4) Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., III, Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron 77*, 639−646.

(5) Kumar, A., Fang, P., Park, H., Guo, M., Nettles, K. W., and Disney, M. D. (2011) A crystal structure of a model of the repeating r(CGG) transcript found in fragile X syndrome. *ChemBioChem 12*, 2140−2142.

(6) Kumar, A., Park, H., Fang, P., Parkesh, R., Guo, M., Nettles, K. W., and Disney, M. D. (2011) Myotonic dystrophy type 1 RNA crystal structures reveal heterogeneous 1 × 1 nucleotide UU internal loop conformations. *Biochemistry 50*, 9928−9935.

(7) Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2012) Crystallographic characterization of CCG repeats. *Nucleic Acids Res. 40*, 8155−8162.

(8) Handa, V., Yeh, H. J., McPhie, P., and Usdin, K. (2005) The AUUCU repeats responsible for spinocerebellar ataxia type 10 form unusual RNA hairpins. *J. Biol. Chem. 280*, 29340−29345.

(9) Disney, M. D., Labuda, L. P., Paul, D. J., Poplawski, S. G., Pushechnikov, A., Tran, T., Velagapudi, S. P., Wu, M., and Childs-Disney, J. L. (2008) Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. *J. Am. Chem. Soc. 130*, 11185−11194.

(10) Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol. 180*, 51−62.

(11) Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D67*, 271−281.

(12) Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D66*, 213−221.

(13) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126−2132.

(14) DeLano, W. L. (2012) *The PyMOL Molecular Graphics System*, version 1.5.0.2, Schrödinger, LLC, Portland, OR.

(15) 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.

(16) Rother, M., Rother, K., Puton, T., and Bujnicki, J. M. (2011) ModeRNA: A tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res. 39*, 4007−4022.

(17) Richardson, J. S., Schneider, B., Murray, L. W., Kapral, G. J., Immormino, R. M., Headd, J. J., Richardson, D. C., Ham, D., Hershkovits, E., Williams, L. D., Keating, K. S., Pyle, A. M., Micallef, D., Westbrook, J., and Berman, H. M. (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA 14*, 465−481.

(18) Joung, I. S., and Cheatham, T. E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B 112*, 9020−9041.

(19) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926.

(20) Case, D. A., V. B, Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., B, Wu, X., and Kollman, P. A. (2014) *AMBER 14*, University of California, San Francisco.

(21) Yildirim, I., Stern, H. A., Kennedy, S. D., Tubbs, J. D., and Turner, D. H. (2010) Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theory Comput. 6*, 1520−1531.

(22) Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J. 92*, 3817−3829.

(23) Patriksson, A., and van der Spoel, D. (2008) A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys. 10*, 2073−2077.

(24) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327−341.

(25) Sagui, C., Pedersen, L. G., and Darden, T. A. (2004) Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys. 120*, 73−87.

(26) Toukmaji, A., Sagui, C., Board, J., and Darden, T. (2000) Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *J. Chem. Phys. 113*, 10913−10927.

(27) Lankas, F., Sponer, J., Langowski, J., and Cheatham, T. E. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J. 85*, 2872−2883.

(28) Paladino, A., and Zangi, R. (2013) Ribose 2′-hydroxyl groups stabilize RNA hairpin structures containing GCUAA pentaloop. *J. Chem. Theory Comput. 9*, 1214−1221.

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859

(29) Nguyen, C. N., Young, T. K., and Gilson, M. K. (2012) Erratum: Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril (The Journal of Chemical Physics (2012) 137 (044101)). *J. Chem. Phys. 137*, 044101.
(30) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graphics 14*, 33−38.
(31) Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) A general module for RNA crystallization. *J. Mol. Biol. 279*, 621−631.
(32) Coonrod, L. A., Lohman, J. R., and Berglund, J. A. (2012) Utilizing the GAAA tetraloop/receptor to facilitate crystal packing and determination of the structure of a CUG RNA helix. *Biochemistry 51*, 8330−8337.
(33) Sugita, Y., and Okamoto, Y. (1999) Replica exchange molecular dynamics method for protein folding. *Chem. Phys. Lett. 314*, 141−151.
(34) Patel, J. S., Berteotti, A., Ronsisvalle, S., Rocchia, W., and Cavalli, A. (2014) Steered molecular dynamics simulations for studying protein-ligand interaction in cyclin-dependent kinase 5. *J. Chem. Inf. Model. 54*, 470−480.
(35) Tekpinar, M., and Zheng, W. (2014) Unzipping of neuronal snare protein with steered molecular dynamics occurs in three steps. *J. Mol. Model. 20*, 2381.
(36) Broda, M., Kierzek, E., Gdaniec, Z., Kulinski, T., and Kierzek, R. (2005) Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry 44*, 10873−10882.
(37) Auweter, S. D., Oberstrass, F. C., and Allain, F. H. (2006) Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Res. 34*, 4943−4959.
(38) Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M., and Cusack, S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure 13*, 1055−1067.
(39) Braddock, D. T., Baber, J. L., Levens, D., and Clore, G. M. (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J. 21*, 3476−3485.

DOI: 10.1021/acs.biochem.5b00551
Biochemistry 2015, 54, 3851−3859