BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

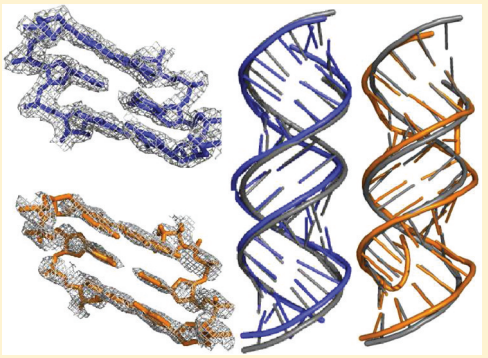
Myotonic Dystrophy Type 1 RNA Crystal Structures Reveal Heterogeneous 1 × 1 Nucleotide UU Internal Loop Conformations

Amit Kumar,[†] HaJeung Park,[‡] Pengfei Fang,[§] Raman Parkesh,[†] Min Guo,[§] Kendall W. Nettles,*[,§] and Matthew D. Disney*[,†]

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

Supporting Information

ABSTRACT: RNA internal loops often display a variety of conformations in solution. Herein, we visualize conformational heterogeneity in the context of the 5′CUG/3′GUC repeat motif present in the RNA that causes myotonic dystrophy type 1 (DM1). Specifically, two crystal structures of a model DM1 triplet repeating construct, 5′r[UUGGGC(CUG)$_3$GUCC]$_2$, refined to 2.20 and 1.52 Å resolution are disclosed. Here, differences in the orientation of the 5′ dangling UU end between the two structures induce changes in the backbone groove width, which reveals that noncanonical 1 × 1 nucleotide UU internal loops can display an ensemble of pairing conformations. In the 2.20 Å structure, CUGa, the 5′ UU forms a one hydrogen-bonded pair with a 5′ UU of a neighboring helix in the unit cell to form a pseudoinfinite helix. The central 1 × 1 nucleotide UU internal loop has no hydrogen bonds, while the terminal 1 × 1 nucleotide UU internal loops each form a one-hydrogen bond pair. In the 1.52 Å structure, CUGb, the 5′ UU dangling end is tucked into the major groove of the duplex. While the canonically paired bases show no change in base pairing, in CUGb the terminal 1 × 1 nucleotide UU internal loops now form two hydrogen-bonded pairs. Thus, the shift in the major groove induced by the 5′ UU dangling end alters noncanonical base patterns. Collectively, these structures indicate that 1 × 1 nucleotide UU internal loops in DM1 may sample multiple conformations in vivo. This observation has implications for the recognition of this RNA, and other repeating transcripts, by protein and small molecule ligands.



RNA is an important target for small molecules;[1] however, most RNA targets have not been exploited as such because little is known about the RNA motifs that specifically recognize small molecules and the small molecules that specifically bind RNA. One notable class of RNAs that could be targeted by small molecules consists of triplet and tetranucleotide repeating transcripts.[2] As of 2011, a variety of untreatable diseases are caused by expanded repeating transcripts, including myotonic dystrophy types 1 and 2 (DM1 and DM2, respectively), spinocerebellar ataxia type 3 (SCA3), Fragile X syndrome, and Huntington's disease (HD).[3]

In contrast to most diseases caused by repeating transcripts, DM1 and DM2 repeating RNAs are not translated into protein. DM1 is caused by an expansion of a CUG repeat in the 3′ untranslated region (UTR) of the dystrophia myotonica protein kinase (DMPK) mRNA,[4−6] while DM2 is caused by an expansion of a CCUG repeat in intron 1 of the zinc finger 9 protein (ZNF9) pre-mRNA.[7] A model for how these RNAs contribute to DM1 and DM2 has been developed and is centered on an RNA gain of function that occurs upon expansion. Long, toxic repeats fold into hairpins and bind the RNA splicing regulator muscleblind-like protein 1 (MBNL1). Sequestration of MBNL1 by the repeating RNAs causes splicing defects in a subset of pre-mRNAs, including the insulin receptor and the muscle main chloride ion channel.[4−6]

Analysis of this disease model provides a strategy for treating DM1 or DM2: a small molecule binds to long CUG repeats and either displaces MBNL1 or inhibits its binding, thereby restoring normal function. Several groups have explored such a strategy. Small molecules have been developed to target DM1 repeats, including compounds identified through screening[8] and an acridine−triazole conjugate.[9] Notably, a designer modular assembly strategy, which could be generally applied to all repeating transcripts, provided potent nanomolar in vitro inhibitors of the DM1 RNA−MBNL1 complex.[10−13] Morpholino oligonucleotides[14] and pentamidine[15] correct splicing defects in a DM1 mouse model.

Previously, structural studies have been completed on model RNA systems containing CUG repeats.[16,17] In these structures, the 1 × 1 nucleotide UU internal loops adopt either a zero- or a one-hydrogen bond pairing structure. A refined NMR structure and molecular dynamics simulation of 5′r(CCGCUGCGG)$_2$ showed that the 1 × 1 nucleotide UU internal loop prefers a

Received: August 16, 2011
Revised: October 10, 2011
Published: October 11, 2011

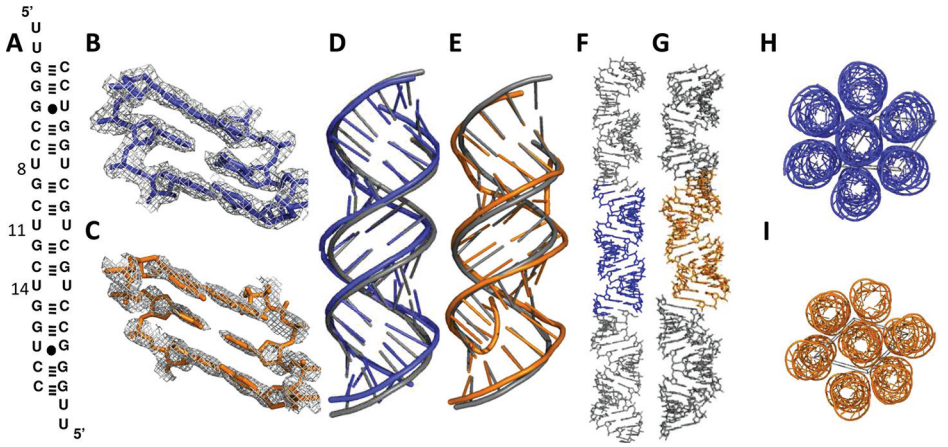


Figure 1. Secondary structure, refined structure, and crystal packing of the RNA construct. (A) Secondary structure of the oligonucleotide r(CUG) duplex model that crystallizes. (B and C) Central 5′CUG/3′GUC motifs of the RNA, including the electron density map at 1.0σ for CUGa and CUGb, respectively. (D and E) Overlays of the backbone of CUGa (blue) and CUGb (orange) onto the backbone of a model RNA duplex in which the 1 × 1 nucleotide UU internal loops are replaced with AU pairs (gray). (F and G) Views of the stacking of pseudoinfinite or stacked helices on each other in the crystal packing for CUGa and CUGb, respectively. (H and I) Views of the stacking of helices next to each other for CUGa and CUGb, respectively.

one-hydrogen bond structure, but it is dynamic and can interconvert among zero-, one-, and two-hydrogen bond pairs without breaking the loop's closing base pairs.[18]

In this study, two crystal structures of a self-complementary duplex with three copies of the DM1 5′CUG/3′GUC motif at 2.20 and 1.52 Å resolution are disclosed. The structures have several notable differences from the structures previously reported. For example, the UU pairs adopt different conformations, including pairing geometries that are consistent with zero-, one-, and two-hydrogen bond pairs depending upon their position in the helix. The structures of the external 1 × 1 nucleotide UU loops are different in the two structures because of differences in the structures of the 5′ UU dangling ends. For example, a 1 × 1 nucleotide internal loop with two hydrogen bonds is observed when the dangling end is tucked into the groove, while a one-hydrogen bond pair is observed when the dangling ends form a pseudoinfinite helix. Evidently, the structure of the dangling end allows for conformational selection of different pairings in the 1 × 1 nucleotide UU internal loops in the crystal structure. However, in both structures, the central 1 × 1 nucleotide UU internal loop adopts a zero-hydrogen bond conformation. Collectively, the available information about CUG repeat structures indicates that the 1 × 1 nucleotide UU internal loops could sample multiple conformations in vivo, which has implications for the recognition of this RNA by protein and small molecule ligands.

MATERIALS AND METHODS

Purification of RNA. Deprotected and desalted RNA was purchased from Integrated DNA Technologies, Inc. The RNA was dissolved in water and purified by HPLC on a Waters HPLC instrument with an attached UV−vis detector that monitored absorbance at 220 and 254 nm. An XTerra Prep MS C18 column (7.8 mm × 150 mm, 5 μm) was used. A linear gradient was applied [from 100 to 0% 10 mM triethylammonium acetate (pH 7.0) in acetonitrile over 55 min] with a flow rate of 2 mL/min (t_R = 25 min). Fractions containing RNA were lyophilized, dissolved in DEPC-treated water, and desalted by using a Sephadex PD-10 prepacked size exclusion column. Fractions containing RNA were combined and lyophilized. The RNA sample was redissolved in DEPC-treated water, and the concentration was determined by its absorbance at 260 nm and 95 °C. Molar extinction coefficients were determined by using the Hyther server (N. Peyret and J. SantaLucia, Jr., Wayne State University, Detroit, MI),[19,20] which uses parameters based on the molar absorptivity of RNA nearest neighbors.[21]

Crystallization of CUG Oligonucleotides. A 1.2 mM solution of RNA duplex (Figure 1A) was prepared in DEPC-treated water. The RNA was folded by being heated at 60 °C for 5 min and slowly cooled to room temperature by being placed on the benchtop. The sitting drop vapor diffusion method and a Qiagen Nucleix Suite kit were used to screen conditions that provided high-quality crystals. Samples were kept at 18 °C. The CUGa crystals that diffracted to 2.20 Å resolution were produced from the reservoir solution containing 10 mM magnesium sulfate, 50 mM sodium cacodylate (pH 6.0), and 1.8 M lithium sulfate, whereas the CUGb crystals that diffracted to 1.52 Å resolution were produced from the reservoir solution containing 15 mM magnesium acetate, 50 mM sodium cacodylate (pH 6.0), and 1.7 M ammonium sulfate. Both crystals appeared within 2−3 days.

Data Collection, Structure Determination, and Refinement. Crystals used for data collection were flash-frozen by immersion in liquid nitrogen. The complete diffraction data sets were collected on beamline 9-1 at SSRL or beamline LS-CAT (21-ID) at the Advanced Photo Source (Argonne National Laboratory, Argonne, IL) under a cryostream at 100 K using ADSC CCD detectors. Data were processed and scaled using HKL2000.[22] The structural solutions of both crystal forms were determined by molecular replacement using PHASER,[23] a component of the PHENIX suite,[24] using a search model of a 17 bp standard A-form RNA generated with Coot. The atomic models were refined with the PHENIX suite. The statistics for data collection and processing and refinement are listed in Table 1 of the Supporting Information. The structures were deposited in the PDB [entries 3SZX (CUGa) and 3SYW (CUGb)].

Calculation of Electrostatic Potentials. The electrostatic potentials of the CUGa and CUGb structures were calculated from the corresponding PDB files; electrostatic potentials for

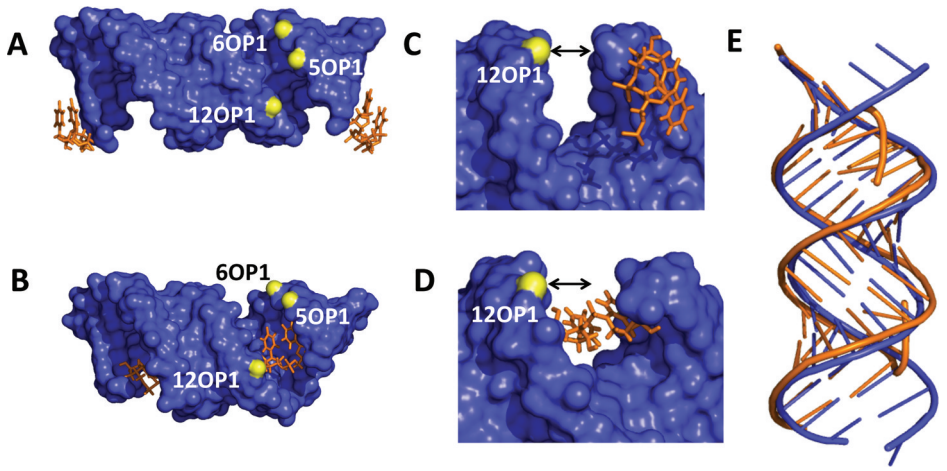


Figure 2. Orientation of the 5′ UU dangling ends in the crystal structures and their effect on groove width. (A and B) Views down the major and minor grooves of CUGa and CUGb, respectively. (C and D) Close-up views of the 5′ UU dangling ends in CUGa and CUGb, respectively. (E) Overlay of the backbone of CUGa (blue) on the backbone of CUGb (orange). Black arrows are used to illustrate the relative difference in the size of the groove.

fully paired RNAs containing AU and GC base pairs were calculated from duplexes that were rebuilt using Amber topology parameters. Hydrogen atoms were added to the RNAs and positioned on the basis of a previously described algorithm.[24] After construction, hydrogen atoms were checked for steric conflicts. Atom partial charges and atomic radii were assigned on the basis of the Amber99 force field using AMBER.[25] The RNA molecule was treated as a low-dielectric medium within the volume enclosed by its solvent-accessible surface (probe radius of 1.4 Å). A dielectric constant of 2 was used to account for electronic polarizability effects. The surrounding solvent was treated as a continuum with a dielectric constant of 80. To account for ion size on the RNA molecule surface, a 2.0 Å ion exclusion radius was added. Ten grid points per square angstrom were used to construct molecular surfaces.

All electrostatic calculations were completed at 298 K. To solve the numerical equation for calculating the electrostatic surface potential, we used a focusing method.[26] The equation was first solved using a coarse grid, which was then refined to provide a more accurate, finer grid using the Dirichlet boundary condition.[27] Electrostatic potentials depict the solvent-excluded molecular surfaces. As RNA molecules are highly negatively charged molecules, most of the molecular surface has a negative electrostatic potential with patches of positive charge.

Calculation of Structural Parameters. Helical parameters, groove widths, and torsion angles were calculated using 3DNA.[25] To avoid computational artifacts arising from noncanonical base pairing, we took sequence-independent measurements using a vector connecting the C1′ atoms.

RESULTS AND DISCUSSION

Three-Dimensional Refined Structures of an RNA Construct with Three Copies of the 5′CUG/3′GUC Motif Found in DM1 RNA. The secondary structure of the RNA construct that was crystallized is shown in Figure 1A. Duplex regions adjacent to the 5′CUG/3′GUC repeats were added to stabilize the ends of the duplex and have also been shown to noncovalently bind heavy atoms that can be used for phasing.[26] In the case of the structures presented here, models that fit the diffraction data could be built (Figure 1B–I) without having to collect data on heavy atom derivatives. The 5′ UU dangling end was used because it allows crystallization by forming pseudoinfinite helices with neighboring RNAs in a unit cell (Figure 1F–I).[27,28] The highest-quality crystals appeared in 2–3 days, diffracting to 2.20 and 1.52 Å (CUGa and CUGb, respectively).

Both of the duplexes crystallize in double-stranded helical structures that form a paired region that flanks three 1 × 1 nucleotide UU internal loops. The secondary structure is similar for both crystal forms, and both crystals display extended helical structures in the packing (Figure 1F–I). The similarity of the structure to canonically paired, A-form RNA is illustrated by an overlay of the backbone from the CUGa structure on a duplex in which the 1 × 1 nucleotide UU loops were replaced with AU pairs (Figure 1D,E). Globally, these structures have very similar shapes.

An overlay of CUGa and CUGb and the effect of the orientation of the 5′ UU end is shown in Figure 2. The two structures have significant differences in the width of the grooves because of very different orientations of the 5′ UU dangling ends (Figure 2A–D). In CUGa, each U in the 5′ UU dangling ends forms a one-hydrogen bond pair with a 5′ UU dangling end of a neighboring duplex in the unit cell to form a pseudoinfinite helix as expected (Figures 1F and 2A,C).[27,28] In CUGb, the 5′ UU dangling end curves into the major groove of the RNA duplex (Figures 1G and 2B,D); this orientation is driven by an intramolecular hydrogen bond formed between N3 of U2 and a nonbridging oxygen atom on O2P of G5 (Supporting Information)

Comparison of the orientation of the dangling ends in the two structures shows that there is an increase in the width of the major groove when the 5′ UU end is tucked into it (Figure 2A–D). The distance between the phosphate atoms in P6 and P12 on either end of the groove increases from 12.6 to 17.3 Å, while the distance between the phosphate atoms in P5 to P12 increases from 9.4 to 14.5 Å (Figure 2C,D). Such a difference in width can affect solvent and ionic interactions that may not be visible in each of these structures at the reported resolution. These effects could drive the differences in the orientation of the UU loops.

This shift in the backbone grooves is also associated with contraction of the C1′–C1′ distances for CUGb. For CUGa

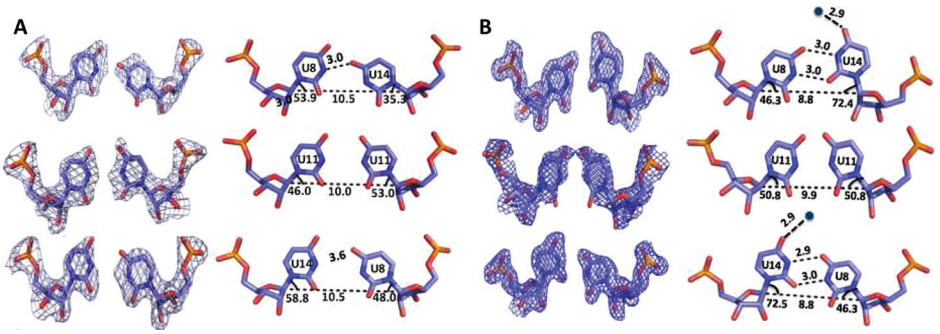


Figure 3. Three-dimensional structures of the 1 × 1 nucleotide UU internal loops. (A and B) Density maps at 1σ and structures of the 1 × 1 nucleotide UU internal loops in CUGa and CUGb, respectively. Different types of 1 × 1 nucleotide UU Internal loops are observed in both structures. The dashed lines indicate the lengths of either the C1′−C1′ or hydrogen bond distance, and the numbers above these lines are the measured distances.

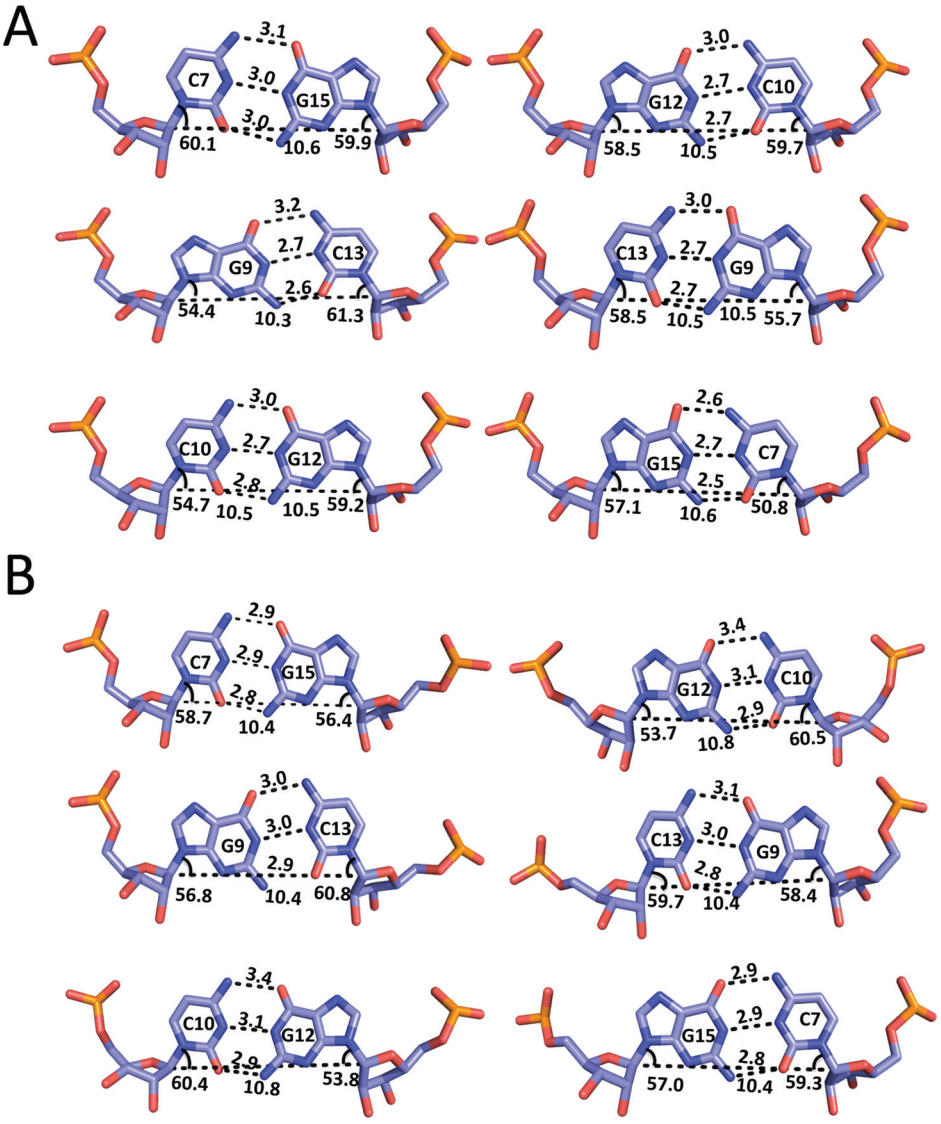


Figure 4. Three-dimensional structures of the 1 × 1 nucleotide UU internal loop closing base pairs in CUGa (A) and CUGb (B). Each of the loop closing pairs has a geometry consistent with that of Watson−Crick GC base pairs.

(Figure 3A), distances range from 10.0 to 10.5 Å, which are similar to the C1′−C1′ distance of ∼10.5 Å observed in a standard A-form RNA helical geometry. For CUGb, the C1′−C1′ distance for the two hydrogen-bonded UU pairs is 8.8 Å, indicating local contraction (Figure 3B), induced by the widening of the adjacent major groove by the dangling UU pair.

Structures of the 1 × 1 Nucleotide UU Internal Loops. Our trapping of the 5′CUG/3′GUC motif into different conformers reveals that 1 × 1 nucleotide UU internal

dx.doi.org/10.1021/bi2013068 | *Biochemistry* 2011, 50, 9928−9935

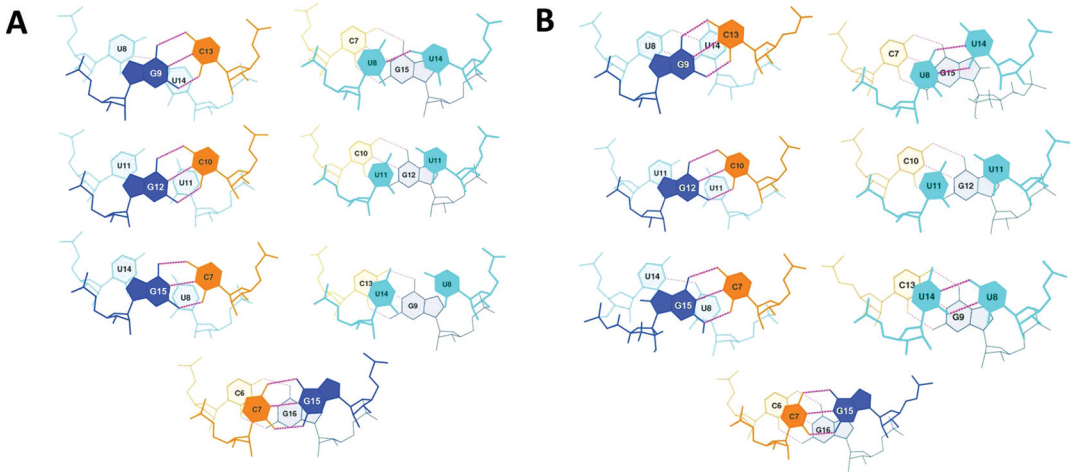


Figure 5. Stacking of the 1 × 1 nucleotide UU internal loops on the loop closing base pairs. (A and B) Structures for CUGa and CUGb, respectively. For each type of 1 × 1 nucleotide UU internal loop, there is little stacking on the closing pairs. However, the 3′ closing GC pairs of one 5′CUG/3′GUC motif stack well on the 5′ closing GC pairs of the next 5′CUG/3′GUC motif.

loops display an ensemble of base pairings between the two structures. Interestingly, there are three different pairing geometries for the 1 × 1 nucleotide UU internal loops in CUGa (Figure 3A). The two UU pairs in 5′CUG/3′GUC motifs at the ends of the duplex have geometries consistent with one- and zero-hydrogen bond structures (Figure 3A). In the one-hydrogen bond geometry, the hydrogen bond is between the U8 N3 and U14 O4 and is 3.0 Å long. The zero-hydrogen bond geometry on the other end has a distance of 3.6 Å (between U14 N3 and U8 O4), which is longer than the standard hydrogen bond distance of 3.2 Å. In both of these structures, the geometry is similar in that U8 O4 is inclined toward the U14 N3 atom; however, differences in the degree of the base's incline determine whether a hydrogen bond forms. In contrast, the central 1 × 1 nucleotide UU internal loop in CUGa has no hydrogen bonds and the carbonyl oxygen atoms are not inclined toward each other.

In CUGb (Figure 3B), the pairings of the 1 × 1 nucleotide UU internal loops are similar and different than those in CUGa. For example, the central 1 × 1 nucleotide UU internal loop has no hydrogen bonds and has an orientation very similar to that of the central 1 × 1 nucleotide UU internal loop in CUGa. In contrast, the two terminal 1 × 1 nucleotide UU internal loops have two hydrogen bonds between the U bases (U8 N3 with U14 O2 and U8 O4 with U14 N3) and an additional hydrogen bond between the U14 O4 atoms and a water molecule; in CUGa, these pairs had zero and one hydrogen bond (Figure 3A).

Canonical Base Pairs and Stacking of 1 × 1 Nucleotide UU Internal Loops on Closing Base Pairs. The structures of the G-C base pairs that close each of the 1 × 1 nucleotide UU internal loops have a standard Watson−Crick geometry (Figure 4A). In addition, the terminal base paired regions at the end of the duplex have hydrogen bonding and stacking interactions that are standard for canonically paired duplexes.

Figure 5 depicts the stacking interactions for each 1 × 1 nucleotide UU internal loop on the loop's closing GC pairs and representative stacking plots for a GC step in the duplex. There is little difference in the stacking of the internal loop U residues on adjacent G-C pairs, for the different types of loops (zero, one, or two hydrogen bonds), although there is more overlap between the region around U8 N3 and the exocyclic amine of a stacked guanine in the two-hydrogen bond structure. The 3′ closing G-C pair of one 5′CUG/3′GUC motif stacks well on the 5′ closing C-G pair of the next motif (Figure 5). Evidently, differences in stacking of the 1 × 1 nucleotide UU internal loops do not account for the observed differences in geometry.

Electrostatics of the Duplexes in Comparison to Canonically Paired RNAs. Figure 6 provides a comparison

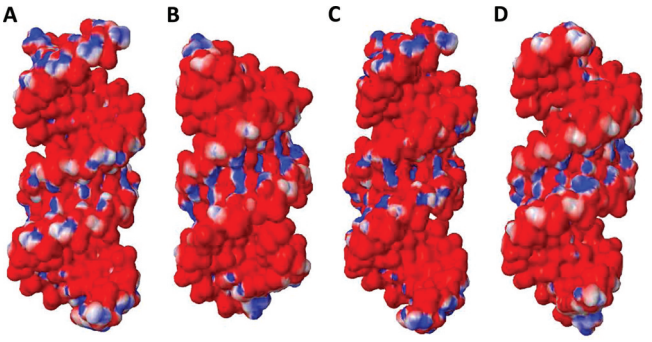


Figure 6. Comparison of the electrostatic charge distributions for CUGa and CUGb to canonically paired duplexes. (A and B) Electrostatic charge distributions of refined CUGa and CUGb, respectively. (C and D) Electrostatic charge distributions of a structure in which the 1 × 1 nucleotide UU internal loops in the CUG construct were replaced with AU pairs and GC pairs, respectively.

between the electrostatic charge distribution of CUGa and CUGb and fully paired RNA duplexes (the 1 × 1 nucleotide UU internal loops are replaced with either A-U or G-C base pairs). There is an alternating pattern of positive and negative charge distribution in the minor groove in all structures. The density of the partial positive charge distribution is higher in CUGa and CUGb structures than in the paired structures; however, CUGb has the greatest density of partial positive charge.

Comparison of CUGa to CUGb. Globally, the two structures reported herein are similar. However, two significant differences are the orientation of the 5′ double dangling UU end and the orientation of the external 1 × 1 nucleotide UU internal loops (Figures 1−3 and Supporting Information). That

dx.doi.org/10.1021/bi2013068|*Biochemistry* 2011, 50, 9928−9935

is, the orientation of the 5′ UU ends is associated with different structures of the internal loops. If the 5′ UU dangling ends are in the major groove of the flanking duplexes, then the external 1 × 1 nucleotide UU internal loops form two hydrogen-bonded pairs. If, however, the 5′ UU dangling ends form base pairs with the adjacent helix, the 1 × 1 nucleotide UU internal loops have a zero- or one-hydrogen bond geometry. Thus, the shift in the width of the backbone allows for conformational selection of different types of 1 × 1 nucleotide UU pairs in the crystal structure, even though canonical base pairing and stacking are not disrupted (Figures 4 and 5). Conformational selection has been used by small molecules to select different orientations of noncanonically paired or unpaired RNA loops.[29−31] In those cases, small molecules make direct contacts to loop nucleotides.

Comparison to Other Structures of CUG Systems. Two crystal structures of RNAs containing the DM1 motif have been previously reported. In the structure of r[(CUG)$_6$]$_2$, the majority of the 1 × 1 nucleotide UU internal loops (five of six) have a zero-hydrogen bond conformation, similar to the central 1 × 1 nucleotide UU loops in both CUGa and CUGb.[16] There is evidence of a minor population of a one-hydrogen bond conformation in r[(CUG)$_6$]$_2$ (one of six) that is similar to the structure observed for a terminal UU pair in a 5′CUG/3′GUC motif in CUGa.[16]

In the previously reported structure of r[G(CUG)$_2$C]$_2$, each UU pair forms a one-hydrogen bond conformation that is considered a UU "stretched wobble" pair.[17] In this stretched U-U pair, two water molecules were found to interact with the UU pairs by forming contacts to O2 and O4 carbonyl atoms in the major and minor grooves. The pair has an orientation similar to that of the one-hydrogen bond UU pair that was found in CUGa; however, we could not observe specifically bound water molecules in CUGa, which may be due to the lower resolution of this structure (2.20 Å vs 1.23 Å). A single bound water molecule was observed, however, for two of the three UU pairs in the CUGb structure. The bound water molecules in CUGb interact with the same functional groups in the major groove that were observed in r[G(CUG)$_2$C]$_2$.

An NMR spectroscopy and molecular dynamics study of an RNA duplex with one copy of 5′CUG/3′GUC or 5′r-(CCGCUGCGG)$_2$ has also been published.[18] These studies showed that the lowest-energy pairing for the 5′CUG/3′GUC motif is a single-hydrogen bond structure that is similar to the UU pair found in CUGa. However, the 1 × 1 nucleotide UU internal loop in this structure is dynamic and samples among zero-, one-, and two-hydrogen bond structures without breaking the loop closing base pairs.

Comparisons of each of these structures with the CUGa and CUGb structures reported herein support a model in which the 1 × 1 nucleotide UU internal loops in the expanded CUG repeats sample multiple conformations. Collectively, each structure may provide unique snapshots of the ensemble of structures that DM1 repeats sample that could be exploited as binding sites for small molecules. Furthermore, the structures of the central 1 × 1 nucleotide UU internal loops in both CUGa and CUGb as well as the majority of 1 × 1 nucleotide internal loops in the structure of r[(CUG)$_6$]$_2$ have a zero-hydrogen bond conformation. Because each of the internal 1 × 1 nucleotide UU internal loops is surrounded by 5′CUG/3′GUC motifs on each slide as would be observed in CUG repeats in vivo, the zero-hydrogen bond conformation may be the most populated one in vivo.

Considerations for High-Affinity Recognition of DM1 CUG Repeating Transcripts. The structures of the CUG repeats could allow for the development of selective and nontoxic small molecules targeting DM1 RNA repeats. In the different hydrogen-bonded structures, hydrogen bond acceptors and donors are presented to the major groove that could bind to a small molecule. In the zero-, one-, and two-hydrogen bond 1 × 1 nucleotide UU internal loop structures, a carbonyl oxygen that could serve as a hydrogen bond acceptor for a small molecule is available. Because the electrostatic potential of the CUG repeats is different from that of base paired RNA, these differences could also be exploited for selective small molecule binding.

Interestingly, several types of small molecules have been found to bind CUG repeats. In general, these compounds are rich in hydrogen bond donors that could form contacts with the available hydrogen bond-accepting carbonyl groups discussed above. For example, Hoechst 33258,[11] kanamycin derivatives,[10,12,13] and pentamidine[15] are all positively charged ligands that could form direct contacts to carbonyl atoms in the CUG repeats. Interestingly, Hoechst 33258 has been shown to interact with DNA by forming contacts between its benzimidazole side chains and the O2 atoms of T in AT rich regions.[32]

Another interesting series of compounds that has been shown to bind CUG repeats consists of triazine−acridine conjugates.[9] On the basis of the crystal structures of the central 1 × 1 nucleotide UU internal loops in CUGa and CUGb, one could envision the triazine moiety easily forming a full complement of three hydrogen bonds to the 1 × 1 nucleotide UU internal loop with a zero-hydrogen bond structure. Such interactions could also form with the one- and two-hydrogen bond UU pairs; however, it would be at the expense of breaking the hydrogen bonds between the U residues, which are each worth ∼1 kcal/mol.[33]

Considerations for Specific Targeting of DM1 CUG Repeating Transcripts. The most significant barrier for specifically targeting CUG repeats that cause DM1 is the presence of 1 × 1 nucleotide UU internal loops in other, highly abundant RNA transcripts. Notable is the presence of a 1 × 1 nucleotide UU internal loop in the *Homo sapiens* rRNA A-site. Interestingly, analysis of structural data for this UU pair shows that it exists in multiple conformations, including one- and two-hydrogen bond pairs. In structures of isolated cytoplasmic and mitochondrial A-sites,[34,35] the UU loop has a one- or two-hydrogen bond conformation. The *H. sapiens* A-site is an accessible target for small molecules; aminoglycoside antibiotics bind the human A-site and stimulate read-through translation.[36] In a crystal structure of the entire yeast ribosome in the ratcheted state, the 1 × 1 nucleotide UU internal loop forms a one-hydrogen bond pair[37] (see the Supporting Information). Additionally, many cellular tRNAs also contain 1 × 1 nucleotide UU internal loops that could bind small molecules that are intended to bind DM1 CUG repeats (Supporting Information).

Selective binding of the CUG hairpin by small molecules is compounded by the relative expression levels of RNA in cells.[38−40] Highly abundant rRNAs and tRNAs comprise ≈80 and ≈15%, respectively, of the total cellular RNA, whereas total mRNA constitutes only ≈5%. Therefore, specific targeting of a single mRNA is a considerable challenge because the level of a single mRNA or other low-abundance RNA is ≪1% of the total RNA content of cells.

Despite these challenges, strategies for designing specific ligands that target the DM1 RNA, which has regularly repeating copies of the 5′CUG/3′GUC motif, are being developed. Via modular assembly of ligands that bind multiple 5′GUC/3′GUC motifs simultaneously, compounds that target the DM1 RNA with affinities and specificities that exceed that of MBNL1 by orders of magnitude can be designed.[10−13] Assembled compounds also have improved potencies for inhibition of the DM1 RNA−MBNL1 interaction by sequestering amounts of surface area on the target RNA larger than that sequestered by a lower-molecular weight compound.[10−13] When using modularly assembled structures, however, a careful balance between potency and size must be considered, as a higher-molecular weight compound is likely to be more potent but less cell permeable.

■ ASSOCIATED CONTENT

Supporting Information

Crystal parameters and table of analysis of the structures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*M.D.D.: Department of Chemistry, Scripps Florida, 130 Scripps Way, 3A1, Jupiter, FL 33458; phone, (561) 228-2203; fax, (561) 228-2147; e-mail, disney@scripps.edu. K.W.N.: Department of Cancer Biology, Scripps Florida, 110 Scripps Way, 2C1, Jupiter, FL 33458; phone, (561) 228-3209; fax, (561) 228-3071; e-mail, knettles@scripps.edu.

Funding

This work was funded by the National Institutes of Health (3R01GM079235-02S1 and 1R01GM079235-01A2 to M.D.D.) and by The Scripps Research Institute. M.D.D. is a Camille and Henry Dreyfus New Faculty Awardee, a Camille and Henry Dreyfus Teacher-Scholar, and a Research Corp. Cottrell Scholar.

Funding

These structures have been deposited in the Protein Data Bank as entries 3SZX (CUGa) and 3SYW (CUGb).

■ ABBREVIATIONS

DEPC, diethyl pyrocarbonate; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; DMPK, dystrophia myotonica protein kinase; HD, Huntington's disease; HPLC, high-performance liquid chromatography; MBNL1, muscleblind-like 1 protein; MD, molecular dynamics; mRNA, messenger RNA; PDB, Protein Data Bank; rRNA, ribosomal RNA; SCA3, spinocerebellar ataxia type 3; tRNA, transfer RNA; UTR, untranslated region; ZNF9, zinc finger 9 protein.

■ REFERENCES

(1) Thomas, J. R., and Hergenrother, P. J. (2008) Targeting RNA with small molecules. *Chem. Rev. 108*, 1171−1224.

(2) Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G. Jr., Nelson, D. L., and Kuhl, D. P. (1992) Triplet repeat mutations in human disease. *Science 256*, 784−789.

(3) Orr, H. T., and Zoghbi, H. Y. (2007) Trinucleotide repeat disorders. *Annu. Rev. Neurosci. 30*, 575−621.

(4) Kanadia, R. N., Johnstone, K. A., Mankodi, A., Lungu, C., Thornton, C. A., Esson, D., Timmers, A. M., Hauswirth, W. W., and Swanson, M. S. (2003) A muscleblind knockout model for myotonic dystrophy. *Science 302*, 1978−1980.

(5) Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., and Thornton, C. A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science 289*, 1769−1773.

(6) Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science 280*, 737−741.

(7) Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science 293*, 864−867.

(8) Gareiss, P. C., Sobczak, K., McNaughton, B. R., Palde, P. B., Thornton, C. A., and Miller, B. L. (2008) Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA-MBNL1 interaction in vitro: Discovery of lead compounds targeting myotonic dystrophy (DM1). *J. Am. Chem. Soc. 130*, 16254−16261.

(9) Arambula, J. F., Ramisetty, S. R., Baranger, A. M., and Zimmerman, S. C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl. Acad. Sci. U.S.A. 106*, 16068−16073.

(10) Lee, M. M., Childs-Disney, J. L., Pushechnikov, A., French, J. M., Sobczak, K., Thornton, C. A., and Disney, M. D. (2009) Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: Targeting the RNAs that cause myotonic muscular dystrophy. *J. Am. Chem. Soc. 131*, 17464−17472.

(11) Pushechnikov, A., Lee, M. M., Childs-Disney, J. L., Sobczak, K., French, J. M., Thornton, C. A., and Disney, M. D. (2009) Rational Design of Ligands Targeting Triplet Repeating Transcripts That Cause RNA Dominant Disease: Application to Myotonic Muscular Dystrophy Type 1 and Spinocerebellar Ataxia Type 3. *J. Am. Chem. Soc. 131*, 9767−9779.

(12) Lee, M. M., Pushechnikov, A., and Disney, M. D. (2009) Rational and Modular Design of Potent Ligands Targeting the RNA that Causes Myotonic Dystrophy 2. *ACS Chem. Biol. 4*, 345−355.

(13) Disney, M. D., Lee, M. M., Pushechnikov, A., and Childs-Disney, J. L. (2010) The Role of Flexibility in the Rational Design of Modularly Assembled Ligands Targeting the RNAs that Cause the Myotonic Dystrophies. *ChemBioChem 11*, 375−382.

(14) Wheeler, T. M., Sobczak, K., Lueck, J. D., Osborne, R. J., Lin, X., Dirksen, R. T., and Thornton, C. A. (2009) Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science 325*, 336−339.

(15) Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A., and Berglund, J. A. (2009) Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. U.S.A. 106*, 18551−18556.

(16) Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. U.S.A. 102*, 16626−16631.

(17) Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2009) Structural insights into CUG repeats containing the 'stretched U-U wobble': Implications for myotonic dystrophy. *Nucleic Acids Res. 37*, 4149−4156.

(18) Parkesh, R., Disney, M. D., and Fountain, M. (2011) NMR Spectroscopy and Molecular Dynamics Simulation of r-$(CCGCUGCGG)_2$ Reveal a Dynamic UU Internal Loop Found in Myotonic Dystrophy Type 1. *Biochemistry 50*, 599−601.

(19) Peyret, N., Seneviratne, P. A., Allawi, H. T., and SantaLucia, J. Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches. *Biochemistry 38*, 3468−3477.

(20) SantaLucia, J. Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.

(21) Puglisi, J. D., and Tinoco, I. Jr. (1989) Absorbance melting curves of RNA. *Methods Enzymol. 180*, 304−325.

(22) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.

(23) Storoni, L. C., McCoy, A. J., and Read, R. J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D60*, 432−438.
(24) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D58*, 1948−1954.
(25) 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.
(26) Keel, A. Y., Rambo, R. P., Batey, R. T., and Kieft, J. S. (2007) A general strategy to solve the phase problem in RNA crystallography. *Structure 15*, 761−772.
(27) Vicens, Q., and Westhof, E. (2002) Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding a site. *Chem. Biol. 9*, 747−755.
(28) Wahl, M. C., Rao, S. T., and Sundaralingam, M. (1996) The structure of r(UUCGCG) has a 5′-UU-overhang exhibiting Hoogsteen-like trans U·U base pairs. *Nat. Struct. Biol. 3*, 24−31.
(29) Lu, J., Kadakkuzha, B. M., Zhao, L., Fan, M., Qi, X., and Xia, T. (2011) Dynamic ensemble view of the conformational landscape of HIV-1 TAR RNA and allosteric recognition. *Biochemistry 50*, 5042−5057.
(30) Stelzer, A. C., Frank, A. T., Kratz, J. D., Swanson, M. D., Gonzalez-Hernandez, M. J., Lee, J., Andricioaei, I., Markovitz, D. M., and Al-Hashimi, H. M. (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol. 7*, 553−559.
(31) Stelzer, A. C., Kratz, J. D., Zhang, Q., and Al-Hashimi, H. M. (2010) RNA Dynamics by Design: Biasing Ensembles Towards the Ligand-Bound State. *Angew. Chem., Int. Ed. 49*, 5731−5733.
(32) Teng, M. K., Usman, N., Frederick, C. A., and Wang, A. H. (1988) The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG). *Nucleic Acids Res. 16*, 2671−2690.
(33) SantaLucia, J. Jr., Kierzek, R., and Turner, D. H. (1992) Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science 256*, 217−219.
(34) Kondo, J., Urzhumtsev, A., and Westhof, E. (2006) Two conformational states in the crystal structure of the *Homo sapiens* cytoplasmic ribosomal decoding A site. *Nucleic Acids Res. 34*, 676−685.
(35) Kondo, J., and Westhof, E. (2008) The bacterial and mitochondrial ribosomal A-site molecular switches possess different conformational substates. *Nucleic Acids Res. 36*, 2654−2666.
(36) Kondo, J., Hainrichson, M., Nudelman, I., Shallom-Shezifi, D., Barbieri, C. M., Pilch, D. S., Westhof, E., and Baasov, T. (2007) Differential selectivity of natural and synthetic aminoglycosides towards the eukaryotic and prokaryotic decoding A sites. *ChemBioChem 8*, 1700−1709.
(37) Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010) Crystal structure of the eukaryotic ribosome. *Science 330*, 1203−1209.
(38) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2007) *Biochemistry*, 6th ed., W. H. Freeman and Co., New York.
(39) Johnson, L. F., Abelson, H. T., Penman, S., and Green, H. (1977) The relative amounts of the cytoplasmic RNA species in normal, transformed and senescent cultured cell lines. *J. Cell. Physiol. 90*, 465−470.
(40) Johnson, L. F., Williams, J. G., Abelson, H. T., Green, H., and Penman, S. (1975) Changes in RNA in relation to growth of the fibroblast. III. Posttranscriptional regulation of mRNA formation in resting and growing cells. *Cell 4*, 69−75.