BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

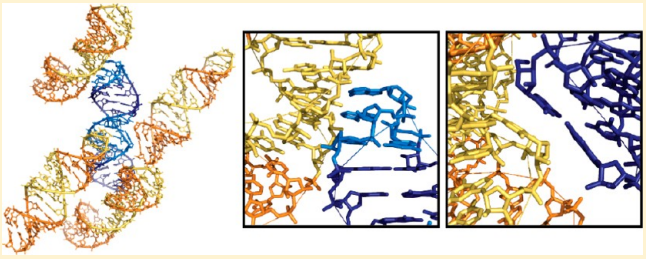
Utilizing the GAAA Tetraloop/Receptor To Facilitate Crystal Packing and Determination of the Structure of a CUG RNA Helix

Leslie A. Coonrod, Jeremy R. Lohman, and J. Andrew Berglund*

Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403, United States

Supporting Information

ABSTRACT: Myotonic dystrophy type 1 (DM1) is a microsatellite expansion disorder caused by the aberrant expansion of CTG repeats in the 3′-untranslated region of the *DMPK* gene. When transcribed, the toxic RNA CUG repeats sequester RNA binding proteins, which leads to disease symptoms. The expanded CUG repeats can adopt a double-stranded structure, and targeting this helix is a therapeutic strategy for DM1. To improve our understanding of the 5′CUG/3′GUC motif and how it may interact with proteins and small molecules, we designed a short CUG helix attached to a GAAA tetraloop/receptor to facilitate crystal packing. Here we report the highest-resolution structure (1.95 Å) to date of a GAAA tetraloop/receptor and the CUG helix it was used to crystallize. Within the CUG helix, we identify two different forms of noncanonical U-U pairs and reconfirm that CUG repeats are essentially A-form. An analysis of all noncanonical U-U pairs in the context of CUG repeats revealed six different classes of conformations that the noncanonical U-U pairs are able to adopt.



Microsatellite expansion diseases are the result of aberrant expansion of short repeating segments of DNA (between 1 and 10 bases).[1] These expansions can occur either in coding regions where they result in abnormal proteins[2] or in noncoding regions of the genome where they, upon transcription, act through a toxic RNA gain-of-function mechanism.[3,4] Myotonic dystrophy (DM), which is the most common form of adult onset muscular dystrophy, affecting ∼1 in 8000 individuals, is the result of microsatellite expansions (reviewed in refs 1, 3, and 5). DM type 1 (DM1) is caused by a CTG expansion in the 3′-untranslated region in the *DMPK* gene.[6] When transcribed to RNA, these CUG repeats sequester RNA binding proteins within nuclear foci.[7,8] Members of one family of sequestered proteins that colocalizes with expanded CUG repeats, the Muscleblind-like (MBNL) proteins, are regulators of alternative splicing, resulting in the missplicing of target transcripts, which leads to disease symptoms.[9−14] A related disease, myotonic dystrophy type 2 (DM2), is very similar to DM1, except instead of a trinucleotide repeat, it is a tetranucleotide repeat (CCTG) in the first intron of the *ZNF9* gene.[15]

The expanded toxic CUG repeat RNA can form stable stem−loop structures,[16] and a handful of CUG repeat structures have been published demonstrating an essentially A-form structure, with the noncanonical U-U pairs forming only one hydrogen bond without distorting the backbone.[17−19] However, two of these structures contained some degree of disorder, and all of the structures had very tight crystal packing, leaving little to no potential room for cocrystallization with ligands (either small molecules or short peptides). To overcome these issues, we decided to use an RNA crystallization motif to design specific intermolecular contacts and minimize close packing and introduce "free spaces" within the crystal.

We chose the GAAA tetraloop and its conserved 11-nucleotide receptor as the core for the crystal contacts. This tetraloop/receptor was initially identified as a tertiary motif in group I and II self-splicing introns.[20−22] The three adenosines in the loop flip outward and form specific hydrogen bonds with the 11-nucleotide receptor,[22] making it an ideal candidate for engineering crystal contacts. This motif has been utilized to aid in the crystallization of larger RNA constructs, such as the hepatitis delta virus, among others.[23,24] Here, we utilized the tetraloop/receptor to crystallize a short RNA helix composed of CUG repeats, allowing us to obtain crystals that diffracted to high resolution. Additionally, we were able to use the previously determined structure of the tetraloop/receptor as our search model, allowing us to quickly and easily determine the structure using molecular replacement.

MATERIALS AND METHODS

Purification of RNA. RNA was purchased from Dharmacon and desalted per the manufacturer's instructions. RNA was then resuspended in 1× denaturing dye [25% (w/w) formamide, 1× TBE, 0.1% bromophenol blue, and 0.1% xylene cyanol] and run on a denaturing 8% polyacrylamide (19:1) gel with 7 M urea. RNA was located using UV shadowing, excised, and eluted overnight using the Elutrap Electroelution System (Whatman) in 1× TBE. RNA was then ethanol precipitated and

Received: June 20, 2012
Revised: September 4, 2012
Published: October 1, 2012

dx.doi.org/10.1021/bi300829w | *Biochemistry* 2012, 51, 8330−8337

resuspended in ddH_2O and desalted using a Micro Bio-Spin column with Bio-Gel P2 Gel (Bio-Rad). RNA was brought to a final concentration of 0.5 mM in a solution of 15 mM NaCl, 5 mM Tris (pH 7.5), and 5 mM $MgCl_2$.

RNA Crystallization. The RNAs were annealed by being heated to 70 °C for 5 min and rapidly cooled to 4 °C. RNAs were screened for crystallization using the Mosquito nanoliter high-throughput robot (TTP labtech) and the Natrix screen (Hampton Research) by the hanging drop vapor diffusion method at room temperature. Promising hits were screened in 4 μL hanging drops. The best crystals of trCUG-3 were grown from a mixture of 2 μL of RNA solution and 2 μL of well solution containing 4 mM $MgSO_4$, 50 mM Tris (pH 8.5), and 30% (w/v) 1,6-hexanediol. Crystals appeared in approximately 1 week.

Data Collection. Crystals of at least 0.02 mm in the smallest dimension were mounted in rayon loops and flash-frozen in liquid nitrogen. Experimental data were collected at Advanced Light Source BL 5.0.1 under a cryostream. The X-ray data were integrated, merged, and scaled using the HKL-2000 program suite[25] and converted to structure factors using the CCP4i GUI[26] for the CCP4 program suite.[27] Data collection statistics are listed in Table 1.

Table 1. Summary of Data Collection and Refinement Statistics

space group	*R*3
unit cell dimensions	
a, *b*, *c* (Å)	70.34, 70.34, 68.24
α, *β*, *γ* (deg)	90, 90, 120
resolution range (Å)	22.75−1.95
total no. of reflections	98787
no. of unique reflections	8621
average redundancy	5.5
completeness (%)	99.3 (98.9)[a]
$I/\sigma I$	37.0 (2.2)[a]
R_{merge}[b]	0.068 (0.60)[a]
average *B* factor (no. of atoms)	
nucleotide	54.58 (744)[a]
solvent	60.53 (44)[a]
R_{free}[c]	0.266
R_{work}[c]	0.208

[a]Values in parentheses represent those for the highest-resolution shell. [b]$R_{merge} = (\sum|I - \langle I\rangle|)/(\sum\langle I\rangle)$, where I is the observed intensity and $\langle I\rangle$ is the average of intensities obtained from multiple observations of symmetry-related reflections. [c]$R_{factor} = (\sum||F_o| - |F_c||)/(\sum|F_o|)$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

Structure Determination. The structure was determined using the molecular replacement method in molrep,[28] a part of the CCP4 program suite,[27] with the tetraloop/receptor from Protein Data Bank (PDB) entry 1GID, the structure of the tetraloop/receptor of the group I ribozyme domain,[22] as the search model. The 11-nucleotide receptor and the four-nucleotide loop (plus the two adjacent base pairs) were successfully used as our search model. The model was manually rebuilt using Coot[29] and refined with refmac.[30] Coot and PyMOL[31] were used to generate figures. Refinement statistics are listed in Table 1. Simulated annealing omit maps were created using CNS[32] by deleting all four uridines in the CUG portion of the structure while all other nucleotides remained fixed during annealing. The structure was deposited in the PDB as entry 4FNJ. Structural parameters were calculated using 3DNA.[33]

RESULTS

Design of RNA Constructs for Crystallization. To overcome crystallographic disorder present in previous CUG repeat structures,[17,18] we designed 10 different constructs utilizing the GAAA tetraloop/receptor to facilitate intermolecular contacts (Figure 1). Each construct contained either four or six CUG repeats, one or two stabilizing base pairs, the receptor, a short upper helix, and the GAAA tetraloop. The upper helix was kept short to favor intermolecular rather than intramolecular interactions. Only trCUG-8 failed to form any crystals, possibly because of the 3′ overhanging nucleotide. Of the remaining constructs, only trCUG-2, -3, -6, and -10 grew crystals of sufficient size to mount. Neither trCUG-2 nor trCUG-10 diffracted to better than 10 Å. Both trCUG-3 and trCUG-6 diffracted to better than 3 Å resolution; however, the 3′ overhanging G of trCUG-6 was disordered, rendering the models essentially the same.

Structure Determination Using the Tetraloop/Receptor as a Model for Molecular Replacement. Because the sequence, and presumably the structure, of the tetraloop/receptor in trCUG-3 was the same as that of the tetraloop/receptor in the group I ribozyme domain structure determined by Cate et al.[22] (with the exception of the first base pair, which was reversed), this portion of their structure was successfully used for molecular replacement. The unmodeled base pairs were clearly visible in the maps, and the nucleotides were easily modeled into the density (Figure 2A). Five peaks were observed in the maps that were unlikely to be waters because of geometric considerations. On the basis of previous structures of the tetraloop/receptor[22,34,35] and their proximity to the phosphate backbone, we modeled the peaks as magnesium ions (Figure 2B). The final model has 35 nucleotides, 44 water molecules, and five magnesium ions with R_{work} and R_{free} values of 0.208 and 0.266, respectively.

Structure of trCUG-3. The 35-nucleotide trCUG-3 RNA crystallized in space group *R*3. As designed, the RNA formed a hairpin with a GAAA tetraloop, a short 5 bp stem, the tetraloop/receptor, and an 8 bp CUG-containing stem (Figure 2B). The upper stem begins and ends with a G·U wobble base pair. The lower stem contains two different noncanonical U-U pairs (Figure 2C), while the G-C base pairs formed Watson−Crick interactions. Each RNA makes specific crystal contacts with four other RNAs: two of these interactions were predicted tetraloop/receptor interactions, and the other two were the unanticipated stacking of the base of the helices against the opposite sides of the receptors of the symmetry mates (Figure 2D).

Within the CUG-containing stem, two different conformations of the noncanonical U-U pair were observed (Figure 2C). The lower U2-U34 noncanonical pair formed two hydrogen bonds, which resulted in a shorter C1′−C1′ distance (8.6 Å) compared to that of the upper U5-U31 noncanonical pair, which formed one hydrogen bond and had a C1′−C1′ distance of 10.6 Å (Table 1 of the Supporting Information). Both noncanonical pairs were inclined toward the minor groove ($\lambda_1 < 54°$) and, as a result, do not appear to stack with the preceding C. The incline was more severe in the U5-U31 noncanonical pair, which is most likely a result of the single hydrogen bond. *B* factors for the atoms that composed these

dx.doi.org/10.1021/bi300829w | *Biochemistry* 2012, 51, 8330−8337

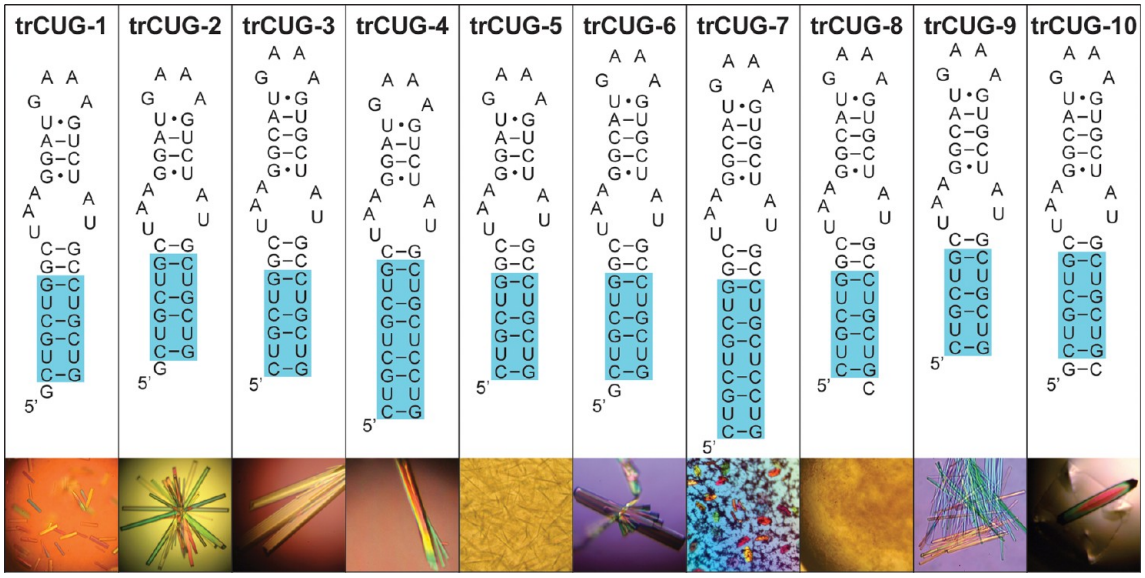


Figure 1. Sequence and predicted secondary structure of RNAs containing the CUG repeats and the tetraloop/receptor. Each RNA construct consists of four or six CUG repeats (highlighted in blue), the 11-nucleotide receptor, a short upper stem, and the GAAA tetraloop. Below each sequence is a representative picture of how well each RNA crystallized. The only sizable trCUG-6 crystal grew off a fiber.

pairs were not noticeably different from those of other atoms within the structure.

Examining the overall structure of the CUG portion of the lower helix [hereafter termed $(CUG)_2$] revealed that it is primarily an A-form helix. The average helical rise of $(CUG)_2$ is 2.39 Å (Table 2 of the Supporting Information), which is closer to an A-form conformation (2.83 Å) than a B-form conformation (3.29 Å).[36] The average helical twist is 35.9°, which is greater than the expected twist of 32.5° for A-form, and not much less than the expected twist of 36.5° for B-form. Base pairs in $(CUG)_2$ are also steeply inclined (18.6°), as is expected for A-form helices. Local dimer step parameters, including roll, twist, slide, and most importantly z_P and z_P(h), also match well with average A-form values (Table 2 of the Supporting Information). Interestingly, the structure did not seem to be affected, beyond the first CU/UG step, by the crystal packing interaction that occurred at the base of the helix (Figure 2D). This base pair step was strongly tipped (20.6°), tilted (−11.3°), and rolled (19.4°) as compared to the other base pairs in the structure.

Within the structure, five suspected magnesium ions were identified. Two were near the GAAA tetraloop; two more flanked the 11-nucleotide receptor, and the final magnesium ion was near the center of the CUG portion of the lower helix (Figure 2B). Only the third and fourth magnesium ions were close enough to the RNA to make direct contacts: the rest interacted via water molecules. None of the suspected magnesium ions appeared to be fully coordinated, and the first and second magnesium ions near the loop interacted with four water molecules.

DISCUSSION

Comparison to Other CUG Helices. To date, four other crystal structures of CUG repeat RNA crystal structures have been published.[17−19] The first, and longest, to be crystallized was a $(CUG)_6$ duplex.[17] Unfortunately, this construct crystallized with 2-fold translational disorder, complicating analysis. The next structure to be released was another duplex, $G(CUG)_2C$.[18] This structure contained three different duplexes within the same unit cell: the A+B duplex, the C+D duplex, and the E+E* duplex, which were related to one another by crystallographic symmetry. The authors of this publication also detwinned the data from Mooers et al. The final two structures are different conformations of a third duplex, $UUGGGC(CUG)_3UCC$.[19] Using 3DNA, we compared the helical parameters of $(CUG)_2$ to the $(5'CUG/3'GUC)_x$ portions of these six different helices observed in four separate crystal structures. [For the sake of simplicity, the detwinned $(CUG)_6$ data were used.]

Overall, all seven CUG helices appeared to be more similar to an A-form helix than a B-form helix, especially when parameters that demonstrate the greatest distinction between A-form and B-form are considered [roll, slide, inclination, *x*-displacement, z_P, and z_P(h)] (Table 2 of the Supporting Information). Although $(CUG)_2$ from trCUG-3 has the greatest helical twist of all the published CUG helices (35.9°), it also has the smallest helical rise (2.39 Å). In general, it appears that the $(CUG)_2$ model is consistent with existing models.

Noncanonical U-U Pairs Are Dynamic. In addition to comparing helical parameters of other CUG structures, we also examined all the noncanonical U-U pairs, in the context of CUG repeats that were available (Figure 3 and Table 1 of the Supporting Information). Noncanonical U-U pairs from all the CUG crystal structures[17−19] and an NMR structure[37] were compared. Examining these pairs revealed that the noncanonical U-U pair in the 5′CUG/3′GUC motif is dynamic, and at least six unique conformations have been captured in the structures. These six conformations were determined by considering the potential number of hydrogen bonds and the direction of inclination (toward the major or minor groove, or not inclined). There is one other potential type of noncanonical U-U pair that has not been observed in the currently available data, which is a noncanonical pair with no hydrogen bonds inclined toward the minor groove. It is likely that this type of noncanonical pair occurs in the context of CUG repeats (as a counterpart to type VI, which also forms no hydrogen bonds

dx.doi.org/10.1021/bi300829w | *Biochemistry* 2012, 51, 8330−8337

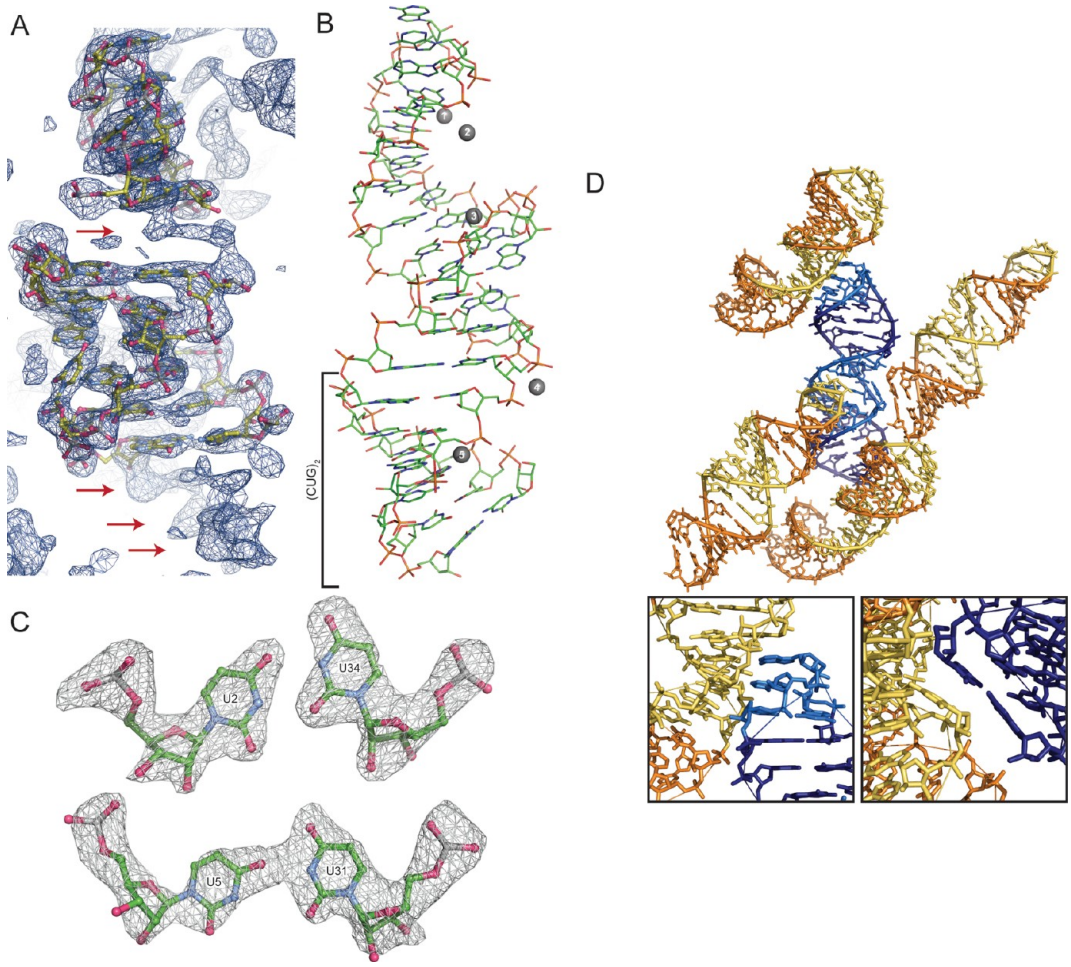


Figure 2. Structure and crystal packing of trCUG-3. (A) Molecular replacement solution and $2F_o - F_c$ map (contoured at 1.5σ) after one round of refinement. The density of the base pairs not yet modeled, both between the tetraloop/receptor and directly below the receptor, can clearly be observed in the density (red arrows). (B) Tertiary structure of trCUG-3. The $(CUG)_2$ region, containing four CUG repeats at the base of the hairpin, is denoted by a bracket. Magnesium ions are represented as numbered gray spheres. Waters are not shown. (C) Simulated annealing omit map ($F_o - F_c$) of both noncanonical U-U pairs contoured at 5σ. (D) Crystal packing (or quaternary structure) of trCUG-3. Each RNA interacts with four other RNAs in the structure: at the loop, at the receptor, on the opposite side of the receptor, and at the bottom of the helix. The tetraloop and receptor are denoted by lighter colors. The inset gives a closer view of each interaction.

but is inclined toward the major groove) and has simply not been observed to date.

Within the trCUG-3 structure, we found two of the six types: type I and type II (Figure 3 and Table 1 of the Supporting Information). The type I noncanonical pair forms two hydrogen bonds, shortening the C1′−C1′ distance, and is inclined toward the minor groove. One other type I noncanonical pair has been observed, in UUGGGC-$(CUG)_3$UCC (Figure 3E),[19] suggesting this conformation is not simply caused by potential distortion of the helix of $(CUG)_2$ due to the crystal packing at the base of the trCUG-3 helix. The other noncanonical pair we observed was a type II noncanonical pair, the second most frequent type of non-canonical U-U pair. The type II noncanonical pair is also inclined toward the minor groove but forms only one hydrogen bond. It was observed twice in $(CUG)_6$ (Figure 3B), twice in the G$(CUG)_2$C A+B duplex, and once in the E+E* duplex (Figure 3D). The most commonly observed noncanonical pair, type IV, is the counterpart to type II and forms one hydrogen bond while being inclined toward the major groove. It was found four times in $(CUG)_6$ (Figure 3B), in the NMR structures (Figure 3C), twice in the G$(CUG)_2$C C+D duplex, once in the E+E* duplex (Figure 3D), and once in the UUGGGC$(CUG)_3$UCC duplex (Figure 3F). Of course, in solution, it is probable that the noncanonical U-U pairs found in the 5′CUG/3′GUC motif sample all these conformations, and that these "types" of noncanonical U-U pairs are not discrete but are rather found in a continuum of different states. These data are consistent with the flexibility of the non-canonical U-U pairs observed in the NMR structure of the CCGCUGCGG duplex (Figure 3C), where the pairs have the ability to sample various conformations without disrupting the structure of the surrounding nucleotides.[37]

Our findings confirm the observation by Kumar et al. that the U-U pairs found in the 5′CUG/3′GUC motif sample more conformations than the two forms of "stretched U-U wobble" (a noncanonical U-U pair characterized by one hydrogen bond interaction between the uridines) identified by Kiliszek et al. However, our analysis suggests that the most common U-U interaction found in the 5′CUG/3′GUC motif is most likely of the stretched U-U wobble variety. This is in contrast to the work of Kumar et al., who conclude that the zero-hydrogen bond conformation is predominant. The difference between our analysis and Kumar's is the use of the detwinned $(CUG)_6$

dx.doi.org/10.1021/bi300829w | *Biochemistry* 2012, 51, 8330−8337

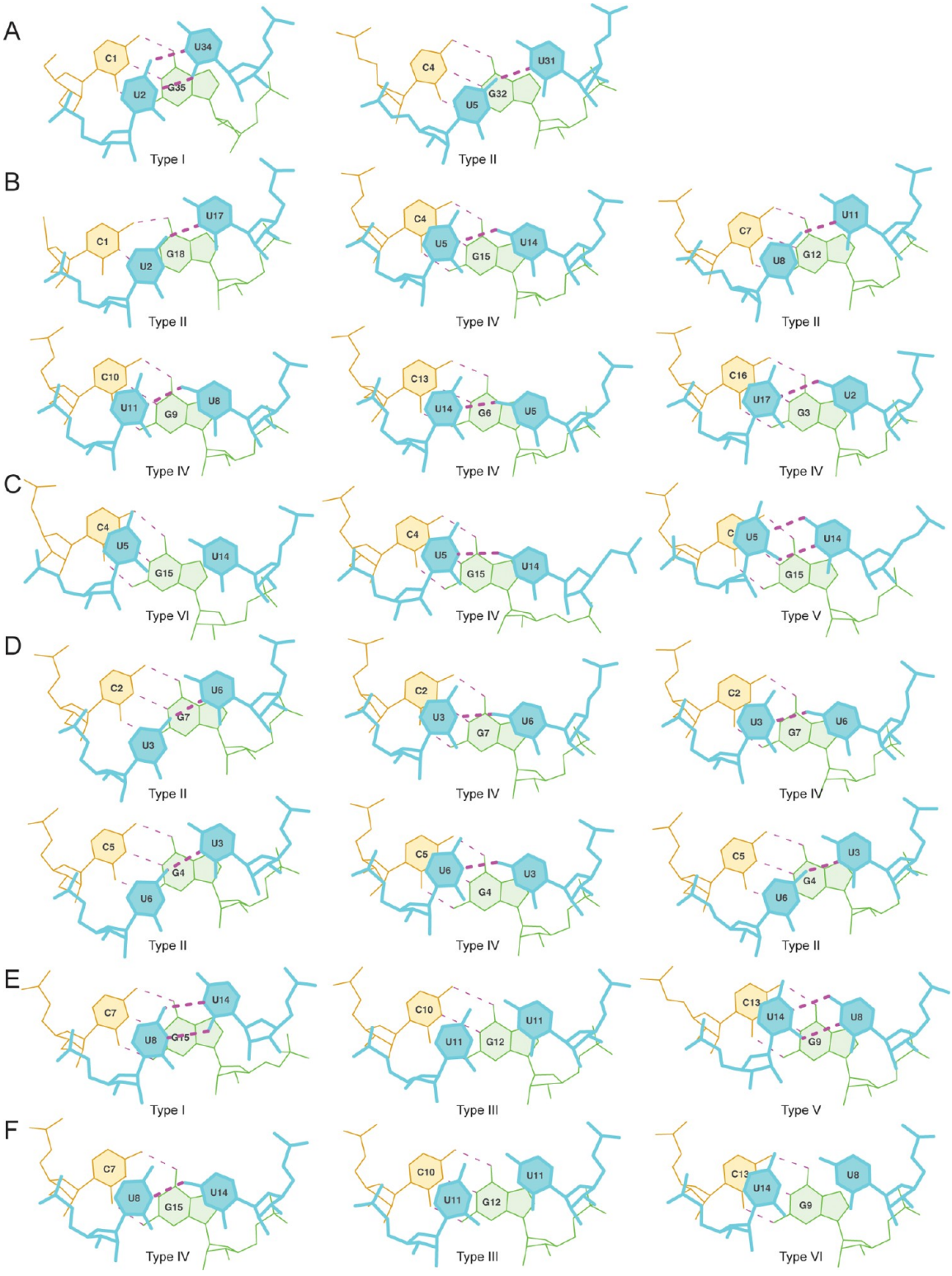


Figure 3. Comparison of noncanonical U-U pairs in the context of CUG repeats. (A) Lower and upper noncanonical U-U pairs of trCUG-3. The upper U5-U31 noncanonical pair forms the more typical (in the context of CUG repeats) number of hydrogen bonds (one), while the lower U2-U34 noncanonical pair forms two hydrogen bonds, shortening the C1′−C1′ distance. (B) Noncanonical U-U pairs of the detwinned $(CUG)_6$ structure. All noncanonical U-U pairs in this structure form one hydrogen bond, maintaining the C1′−C1′ distance typical of a Watson−Crick interaction. (C) Noncanonical U-U pairs from the NMR structure of CCGCUGCGG, forming zero, one, or two hydrogen bonds. (D) Noncanonical U-U pairs from the $G(CUG)_2C$ crystal structure. The first column contains the noncanonical U-U pairs in duplex A+B, the second column C+D, and the third column E+E*. Every noncanonical pair within the $G(CUG)_2C$ crystal structure forms one hydrogen bond. (E and F) Noncanonical U-U pairs from $UUGGGC(CUG)_3CUCC$ of PDB entries 3SYW (E) and 3SZX (F).

dx.doi.org/10.1021/bi300829w | *Biochemistry* 2012, 51, 8330−8337

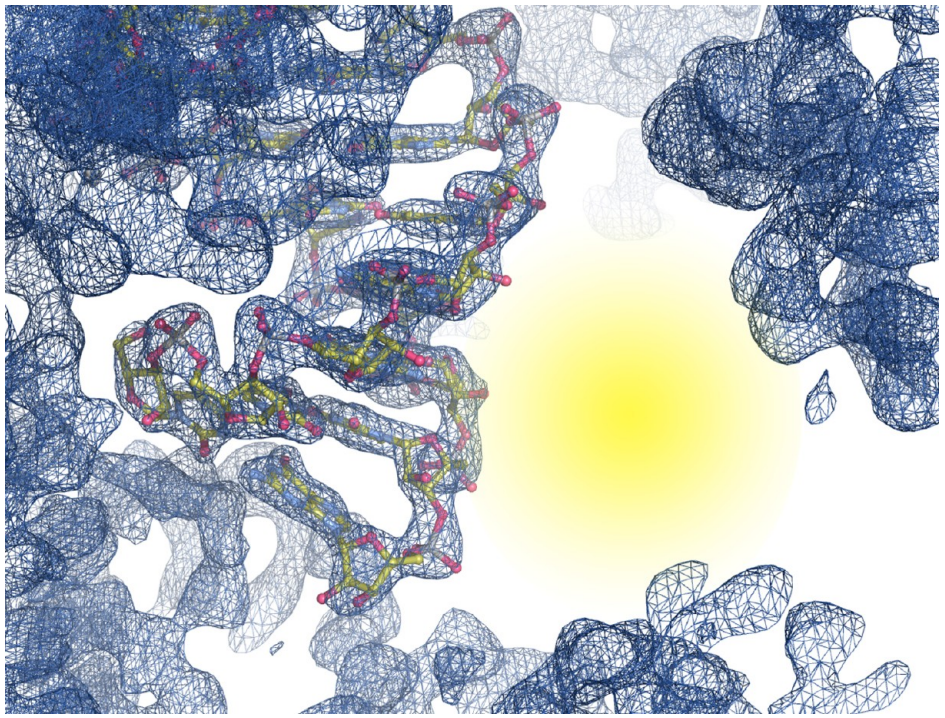


Figure 4. Structure and $2F_o - F_c$ map contoured at 1.5σ near the CUG repeat stem. The yellow area highlights the large open space near the minor groove of the $(CUG)_2$ stem.

data[18] versus the original twinned data modeled at half-occupancy.[17] In the original model, five of the six U-U pairs formed zero hydrogen bonds and the final pair was a stretched U-U wobble, while in the detwinned model, all the U-U pairs were stretched U-U wobbles. The higher proportion of stretched U-U wobbles is consistent with the molecular dynamics simulations performed by Parkesh et al. in which they found 76.5% of noncanonical U-U pairs in the 5′CUG/3′GUC motif formed one hydrogen bond.[37] Interestingly, our structure is the first crystal structure to contain a stretched U-U wobble as well as a U-U pair that forms two hydrogen bonds.

Metal Ions in the Structure. Within the trCUG-3 crystal structure, we identified five potential magnesium ions. It is unsurprising that magnesium ions would be identified within the structure, especially the tetraloop/receptor region, because it has previously been shown that the tetraloop/receptor interaction is dependent on magnesium[38] or other divalent metal ions.[39] Several other structures of the GAAA tetraloop/receptor have been published, so we compared our structure and metal ion sites with several previously published structures (Figure 1 of the Supporting Information). Comparing trCUG-3 and PDB entry 1GID,[22] we observed our first three magnesium ions were in a similar location as compared to the cobalt hexamine ion which was found to contact the GAAA loop (Figure 1B of the Supporting Information). Our fourth magnesium ion near the receptor has a nearby counterpart in PDB entry 1GID. When compared to PDB entry 1HR2[34] (Figure 1C of the Supporting Information), we observed a similar scenario: our first three magnesium ions were near one magnesium ion contacting the loop. The second, third, and fourth magnesium ions had counterparts in the structure of PDB entry 1HR2. The additional ions observed in our structure may be due to the proximity of our tetraloop and receptor as compared to that in either PDB entry 1GID or PDB entry 1HR2, or it could simply be that the magnesium ions are important for electrostatic screening, allowing close contact between RNA strands.[40]

Engineered RNA Contacts as a Strategy for Crystallization. It is clear that utilizing the GAAA tetraloop/receptor was a successful strategy for crystallizing a short helix and prevents the pseudosymmetry problems associated with duplex RNA. Additionally, examining the crystal packing of this particular construct, we observed a large "open" space in the crystal lattice (Figure 4), which could potentially allow for cocrystallization of small molecules, short peptoids, and possibly small zinc-finger domains in complex with the CUG stem. This open space is unique to our tetraloop/receptor construct as compared to other CUG repeat-containing crystal structures:[17−19] these structures all crystallized in pseudoinfinite helices with very tight packing between helices, leaving little to no room for cocrystallization of an additional molecule.

This GAAA tetraloop/receptor framework can also be utilized to crystallize other short repeat RNA stems, such as the CCUG repeats responsible for DM2,[8,15] CAG repeats responsible for Huntington's disease and several spinocerebellar ataxias, and CGG repeats responsible for fragile X-associated tremor ataxia syndrome, among many others (reviewed in ref 41). We have already crystallized a CCUG helix utilizing the tetraloop/receptor construct, which diffracted to a resolution of better than 3 Å (data not shown).

We have crystallized a CUG-containing RNA construct that has captured two different orientations of the clearly mobile noncanonical U-U pair. Via examination of other crystal and NMR structures, it is clear that noncanonical U-U pairs, in the context of CUG repeats, can flex between many different conformations, from fully paired with one another, to only sharing a single hydrogen bond, to being too distant to interact with one another (Figure 3). Additionally, these noncanonical U-U pairs can go from being inclined toward the minor groove, through no inclination, all the way to being inclined toward the

dx.doi.org/10.1021/bi300829w | *Biochemistry* 2012, 51, 8330−8337

major groove. In our interpretation, this conformational heterogeneity suggests a weak interaction caused by an apparent reduction in the level of hydrogen bonding (averaged over all potential U-U pair conformations) and, more importantly, less efficient base stacking compared to the stacking of the GC/GC steps within the helix (Figure 2 of the Supporting Information). The weak interaction between the uridines and neighboring bases may play an important role in MBNL proteins gaining access to single-stranded CUG repeats to bind the Watson−Crick face of the GC dinucleotide as found in the MBNL1-RNA crystal structure (PDB entry 3D2S).[42] Interestingly, a recently published structure of CCG repeats shows that the noncanonical C-C base pair appears to adopt fewer conformations than the U-U pair does (zero or one hydrogen bond in comparison to zero to two hydrogen bonds in the case of U-U).[43] Because of the arrangement of hydrogen bond donors and acceptors on the Watson−Crick face of cytosine, it is hard to envision the C-C pair adopting some of the conformations observed with the U-U pairs, but the C−C pairs do show similarly poor base stacking[43] and similar thermodynamic stability in the context of the CNG helix.[44] Additionally, in the context of DM2, we predict that the noncanonical C-U pairs of the 5′CCUG/3′GUCC motif will have similarly weak and/or dynamic interactions as compared to the noncanonical U-U pairs found in the CUG helix, again allowing MBNL proteins to gain access to the Watson−Crick faces of the GC dinucleotide.

■ ASSOCIATED CONTENT

Supporting Information

Tables of noncanonical U-U pair structural parameters and helical parameters, a figure showing the comparison of the trCUG-3 tetraloop/receptor and previously crystallized tetraloop/receptors, and a figure showing the stacking interactions of the three different base steps in the 5′CUG/3′GUC motif. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: aberglun@uoregon.edu. Phone: (541) 346-5097.

Funding

These studies were supported by Grant AR0599833 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank B. Nolen, J. Purcell, and members of the Berglund lab for helpful discussions and comments about the manuscript. Portions of this research were conducted at the Advanced Light Source, a national user facility operated by Lawrence Berkeley National Laboratory, on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The Berkeley Center for Structural Biology is supported in part by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences.

■ REFERENCES

(1) Cooper, T. A., Wan, L., and Dreyfuss, G. (2009) RNA and Disease. *Cell 136*, 777−793.
(2) Li, Y.-C., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2002) Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Mol. Ecol. 11*, 2453−2465.
(3) Ranum, L. P., and Cooper, T. A. (2006) RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci. 29*, 259−277.
(4) O'Rourke, J. R., and Swanson, M. S. (2009) Mechanisms of RNA-mediated Disease. *J. Biol. Chem. 284*, 7419−7423.
(5) Cho, D. H., and Tapscott, S. J. (2007) Myotonic dystrophy: Emerging mechanisms for DM1 and DM2. *Biochim. Biophys. Acta 1772*, 195−204.
(6) Lee, J. E., and Cooper, T. A. (2009) Pathogenic mechanisms of myotonic dystrophy. *Biochem. Soc. Trans. 37*, 1281−1286.
(7) Mankodi, A., Urbinati, C. R., Yuan, Q. P., Moxley, R. T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M. S., and Thornton, C. A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet. 10*, 2165−2170.
(8) Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet. 11*, 805−814.
(9) Ho, T. H., Charlet-B, N., Poulos, M. G., Singh, G., Swanson, M. S., and Cooper, T. A. (2004) Muscleblind proteins regulate alternative splicing. *EMBO J. 23*, 3103−3112.
(10) Hino, S. I., Kondo, S., Sekiya, H., Saito, A., Kanemoto, S., Murakami, T., Chihara, K., Aoki, Y., Nakamori, M., Takahashi, M. P., and Imaizumi, K. (2007) Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. *Hum. Mol. Genet. 16*, 2834−2843.
(11) Warf, M. B., and Berglund, J. A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA 13*, 2238−2251.
(12) Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., Hall, M. P., Shiue, L., Swanson, M. S., Thornton, C. A., and Ares, M. (2010) Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol. 17*, 187−193.
(13) Sen, S., Talukdar, I., Liu, Y., Tam, J., Reddy, S., and Webster, N. J. G. (2010) Muscleblind-like 1 (Mbnl1) promotes insulin receptor exon 11 inclusion via binding to a downstream evolutionarily conserved intronic enhancer. *J. Biol. Chem. 285*, 25426−25437.
(14) Gates, D. P., Coonrod, L. A., and Berglund, J. A. (2011) Autoregulated Splicing of muscleblind-like 1 (MBNL1) Pre-mRNA. *J. Biol. Chem. 286*, 34224−34233.
(15) 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.
(16) Napierała, M., and Krzyzosiak, W. J. (1997) CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J. Biol. Chem. 272*, 31079−31085.
(17) Mooers, B. H. M., 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.
(18) 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.
(19) 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.
(20) Murphy, F. L., and Cech, T. R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol. 236*, 49−63.

(21) Costa, M., and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J. 14*, 1276−1285.
(22) Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science 273*, 1678−1685.
(23) Ferré-D'Amaré, A. R., Zhou, K., and Doudna, J. A. (1998) A general module for RNA crystallization. *J. Mol. Biol. 279*, 621−631.
(24) Reiter, N. J., Osterman, A., Torres-Larios, A., Swinger, K. K., Pan, T., and Mondragón, A. (2010) Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature 468*, 784−789.
(25) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307−326.
(26) Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D59*, 1131−1137.
(27) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D67*, 235−242.
(28) Vagin, A., and Teplyakov, A. (1997) MOLREP: An Automated Program for Molecular Replacement. *J. Appl. Crystallogr. 30*, 1022−1025.
(29) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D66*, 486−501.
(30) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D53*, 240−255.
(31) *The PyMOL Molecular Graphics System*, version 1.5.0.1 (2010) Schrodinger, LLC, New York.
(32) Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D54*, 905−921.
(33) Lu, X.-J., and Olson, W. K. (2008) 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc. 3*, 1213−1227.
(34) Juneau, K., Podell, E., Harrington, D. J., and Cech, T. R. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA−solvent interactions. *Structure 9*, 221−231.
(35) Davis, J. H., Foster, T. R., Tonelli, M., and Butcher, S. E. (2006) Role of metal ions in the tetraloop-receptor complex as analyzed by NMR. *RNA 13*, 76−86.
(36) Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol. 313*, 229−237.
(37) Parkesh, R., Fountain, M., and Disney, M. D. (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.
(38) Qin, P. Z., Butcher, S. E., Feigon, J., and Hubbell, W. L. (2001) Quantitative analysis of the isolated GAAA tetraloop/receptor interaction in solution: A site-directed spin labeling study. *Biochemistry 40*, 6929−6936.
(39) Downey, C. D., Fiore, J. L., Stoddard, C. D., Hodak, J. H., Nesbitt, D. J., and Pardi, A. (2006) Metal ion dependence, thermodynamics, and kinetics for intramolecular docking of a GAAA tetraloop and receptor connected by a flexible linker. *Biochemistry 45*, 3664−3673.
(40) Draper, D. E., Grilley, D., and Soto, A. M. (2005) Ions And RNA Folding. *Annu. Rev. Biophys. Biomol. Struct. 34*, 221−243.
(41) Wojciechowska, M., and Krzyzosiak, W. J. (2011) Cellular toxicity of expanded RNA repeats: Focus on RNA foci. *Hum. Mol. Genet. 20*, 3811−3821.
(42) Teplova, M., and Patel, D. J. (2008) Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat. Struct. Mol. Biol. 15*, 1343−1351.
(43) Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2012) Crystallographic characterization of CCG repeats. *Nucleic Acids Res. 40*, 8155−8162.
(44) 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.