BIOCHEMISTRY
including biophysical chemistry & molecular biology
Rapid Report

Biochemistry **2011,** *50,* 599–601 599
DOI: 10.1021/bi101896j

NMR Spectroscopy and Molecular Dynamics Simulation of r(CCGCUGCGG)$_2$ Reveal a Dynamic UU Internal Loop Found in Myotonic Dystrophy Type 1[†]

Raman Parkesh,[‡] Matthew Fountain,[§] and Matthew D. Disney*[,‡]

[‡]*The Kellogg School of Science and Technology and Department of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way #3A1, Jupiter, Florida 33458, United States, and* [§]*The Department of Chemistry, The State University of New York at Fredonia, 218 Houghton Hall, Fredonia, New York 14063, United States*

Received November 29, 2010; Revised Manuscript Received December 30, 2010

ABSTRACT: The NMR structure of an RNA with a copy of the 5′CUG/3′GUC motif found in the triplet repeating disorder myotonic dystrophy type 1 (DM1) is disclosed. The lowest energy conformation of the UU pair is a single-hydrogen bond structure; however, the UU protons undergo exchange indicating structural dynamics. Molecular dynamics simulations show that the single hydrogen bond structure is the most populated one but the UU pair interconverts among zero, one, and two hydrogen bond pairs. These studies have implications for the recognition of the DM1 RNA by small molecules and proteins.

RNA plays essential roles in many biological processes and is thus an important therapeutic target (*1*). In particular, recent studies have shown that an RNA gain of function causes myotonic dystrophy types 1 and 2 (DM1 and DM2,[1] respectively) (*2*, *3*). DM1 is caused by a trinucleotide CTG repeat in the 3′UTR of the DMPK gene. Once transcribed, expanded CUG repeats fold into an RNA hairpin with repeating 5′CUG/3′GUC motifs and bind CUG binding protein and the RNA splicing regulator muscleblind-like protein 1 (MBNL-1). The downstream effect of these interactions is mis-splicing of the main muscle chloride channel (*clc1*) and insulin receptor transcripts (*4*, *5*).

Oligonucleotides (*6*) and the small molecule pentamidine (*7*) target the DM1 CUG repeats and correct pre-mRNA splicing defects in animal models of disease. Other small molecules targeting DM1 RNAs have been developed, including triazine−acridine conjugates (*8*) and compounds selected via a dynamic diversity screening strategy (*9*). In addition, rational modular assembly strategies have been developed to provide nanomolar in vitro inhibitors (*10*−*13*). To gain insights into how small molecules and proteins bind DM1 RNAs, the refined NMR structure and a molecular dynamics (MD) simulation of an RNA that contains the 5′CUG/3′GUC motif are disclosed.

NMR spectra of r(CCGCUGCGG)$_2$ were recorded and nonexchangeable protons assigned by analysis of NOESY and DQF-COSY spectra in D_2O at 40 °C. The base−H1′ NOESY walk (Figure 1) and the observed cross-peaks and intensities from the inter- and intranucleotide base to H2′ and from the base to H3′ are indicative of A-form RNA. NOE connectivities between the C4 H1′ and U5 H6 and between U5 H1′ and G6 H8 are similar to other base−H1′ connectivities in the duplex; this

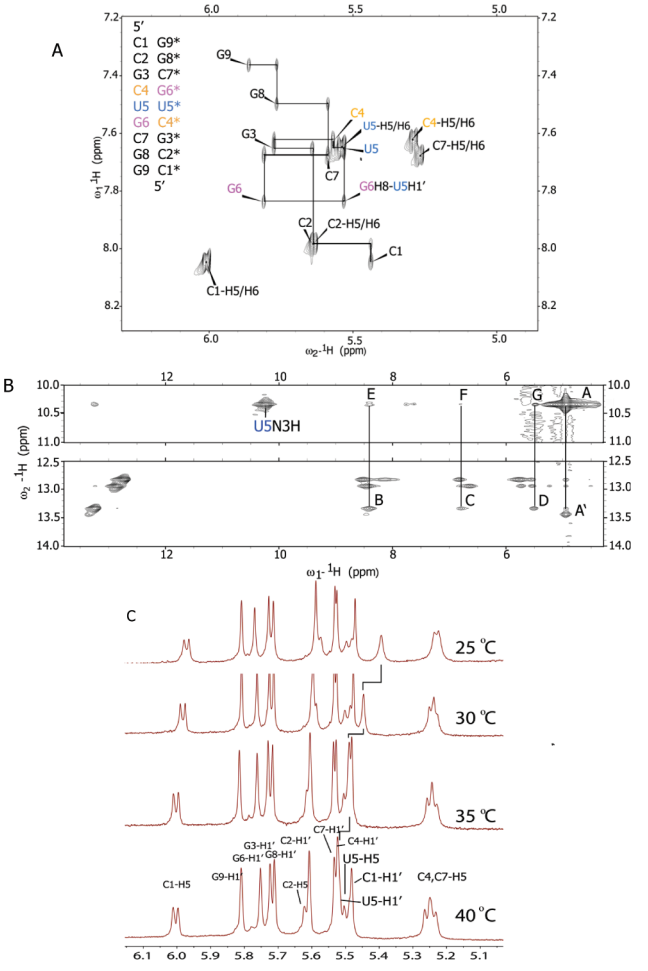

FIGURE 1: Two- and one-dimensional NMR spectra of r(CCGCUGCGG)$_2$. (A) The 200 ms NOESY spectrum showing base−H1′ connectivities. Labels identify intranucleotide base−H1′ NOESY cross-peaks. C and U H5−H6 cross-peaks are also identified. (B) The 100 ms NOESY spectrum recorded in a 90:10 H_2O/D_2O mixture at 10 °C. Label E identifies cross-peaks between the U5 N3H group and the G6 imino proton, while F and G identify cross-peaks between the U5 imino proton and the C4 and G6 imino protons, respectively. Strong cross-peaks between the U5 N3H group and the terminal G9 imino with water are identified in the spectrum as A and A′, respectively. (C) Temperature-dependent NMR spectra. An upfield shift of ~0.1 ppm in the U5−H1′ peak at lower temperatures indicates dynamic exchange.

[†]This work was funded by the National Institutes of Health (3R01GM-079235-02S1 and 1R01GM079235-01A2) and Research Corp. M.D.D. is a Dreyfus New Faculty Awardee, a Dreyfus Teacher-Scholar, and a Cottrell Scholar.

*To whom correspondence should be addressed. E-mail: Disney@scripps.edu. Phone: (561) 228-2203. Fax: (561) 228-2147.

[1]Abbreviations: *clc1*, main chloride ion channel; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; DMPK, dystrophia myotonia protein kinase; DQF-COSY, double-quantum-filtered correlation spectroscopy; MD, molecular dynamics; NOESY, nuclear Overhauser effect correlation spectroscopy; 3′UTR, 3′ untranslated region.

Published on Web 01/04/2011

FIGURE 2: Overlay of the 25 lowest energy structures of r(CCGCU-GGCC)$_2$: view of structures from the major groove (right) and view of the structure from the minor groove (left).

supports a structure in which stacking among C4, U5, and G6 is present.

Exchangeable proton assignments were made by analysis of a two-dimensional NOESY spectrum recorded in a 9:1 H_2O/D_2O mixture at 10 °C. All imino protons are observed. Imino−imino proton, imino−amino proton, and C−H5 NOEs are present for the internal GC base pairs. The G3 and G8 imino protons exhibit very little to no exchange with the solvent.

The terminal G9 imino proton (peak A′ in Figure 1) and the U5 imino proton (peak A in Figure 1) are broadened by exchange with solvent. The location of the U5 base is best defined by NOEs observed via water NOESY. The U5 imino proton exhibits an NOE to the G6 imino proton peak (E in Figure 1) and to the C4 and G6 amino protons (F and G, respectively, in Figure 1). The large NOE cross-peak between the U5 imino proton and water (A in Figure 1) shows that the UU pair, especially at the N3 interface, is dynamic. To further confirm dynamic exchange, NMR spectra (C in Figure 1) of the RNA duplex in D_2O were recorded at 25, 30, 35, and 40 °C. Line broadening and an upfield shift in the U5 H1′ resonance with decreased temperature indicate rapid exchange, thus confirming a structurally dynamic UU pair. An upfield shift of U5 H5 was observed as a function of temperature, indicating rapid exchange (*14*, *15*).

On the basis of $^1H-^1H$ NOE distances and Watson−Crick base pairing restraints, the structure of the duplex was refined using restrained molecular dynamics and Amber 11 (*16*). There are no NOE distance violations observed at >0.2 Å. Figure 2 shows the superposition of the 25 lowest free energy structures generated. The average root-mean-square deviation (rmsd) for the pairwise, all-atom superposition of the structures is 0.78 Å, indicating excellent convergence.

Globally, the structure adopts a well defined A-form geometry with a C3′-endo sugar pucker. The helical rise for each base pair step is less than 3 Å. Thus, the overall structure is similar to that of a standard A-form helix with moderate displacement of the UU pair.

Because NMR spectra indicated that the UU pair was in dynamic exchange as indicated by the NOE between U5 N3H and water (Figure 1), we sought to explore the nature of the dynamics via MD simulations. MD simulations were performed over a 1 ns period using the lowest energy, Amber-refined RNA structure as a starting point. During the course of the simulations, the global rmsd of the structure varied little relative to the initial starting structure, indicating equilibration and a stable simulation.

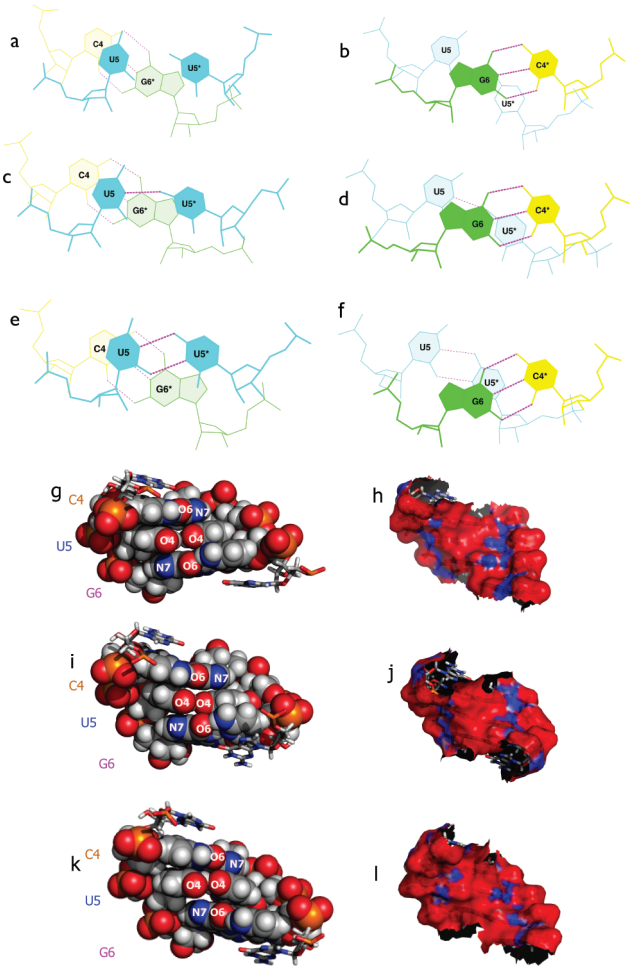

FIGURE 3: Top view of the stacking interactions for the zero, one, and two hydrogen bond structures, space filling model views from the minor groove, and electrostatic potential as viewed from the minor groove. (a−f) Stacking diagrams for the UU loop on the closing GC pairs for the structures with zero (a and b), one (c and d), and two (e and f) hydrogen bonds. (g−l) Views of the major groove as a CPK model and major groove electrostatic calculations for the structures with zero (g and h), one (i and j), and two (k and l) hydrogen bonds.

A series of more than 4000 discrete frames (structures) were generated during the MD simulation. These structures were analyzed to define the types of UU pairs present in the trajectory. The analysis showed UU pair structures that had zero, one, or two hydrogen bonds (using distance and angle cutoffs of 3.5 Å and 120°, respectively, to define the presence of hydrogen bonds) (*17*). The relative populations are as follows: 15.9% of the structures with zero hydrogen bonds, 76.5% with one hydrogen bond (from U5 N3 to U5 O4), and 7.6% with two hydrogen bonds (from U5 N3 to U5 O2 and from U5 N3 to U5 O4) (Figure 3).

To understand the features in the different UU pairs and to explain the relative population differences from MD, we studied the geometry of the pairings and their stacking interactions with the loop closing pairs. For the geometry, the interstrand C1′−C1′ distances were calculated for each type of structure. This analysis showed that UU pairs with zero, one, and two hydrogen bonds have distances of 10.7, 10.6, and 8.9 Å, respectively. Because the C1′−C1′ distance in standard A-form RNA helices is ~10.5 Å (*18*), the structures with zero and one hydrogen bonds do not induce a change in standard A-form helical geometry;

however, the structure with two hydrogen bonds does. Thus, despite the fact that the structure with two hydrogen bonds has a bonus in stability from an extra hydrogen bond, it is not the major structure. Evidently, formation of two hydrogen bonds would occur at the expense of the introduction of a distortion in the RNA helix.

Differences in stacking and electrostatic surface potentials in the three types of structures were also investigated (Figure 3). For all three structures, there is some overlap of the UU pair with the closing GC base pairs; however, there are not large differences in stacking among the three pairing types. Electrostatic potentials (Figure 3h,j,l) were calculated using the Poisson−Boltzmann equation (*19*). The electrostatic surface potential of the minor groove is largely negative, although there are isolated patches of positive surface area. Akin to stacking interactions, the electrostatic potentials are similar.

Two crystal structures of RNA molecules containing CUG repeats have been reported (*20*). In good agreement with these studies with six CUG repeats, the RNA molecule adopts an overall A-form structure. Moreover, there is no distortion in the RNA backbone relative to A-form to accommodate the UU pairs. A structure of $G(CUG)_2C$ was reported by Kiliszek et al. (*21*). In this structure, the 5′CUG/3′GUC motifs display the one-hydrogen bond pattern that was observed by NMR. The difference in the two crystal structures prompted Kiliszek et al. to reanalyze the structure of Mooers et al.; this reanalysis suggested that the structure of Mooers et al. is consistent with the one hydrogen bond structure. The observation that similar structures are observed with an isolated DM1 motif by NMR compared to crystal structures with multiple DM1 motifs suggests that the structural effects of having multiple DM1 motifs may be negligible.

Recently, an analysis of the structures of 1 × 1 nucleotide RNA internal loops deposited in the Protein Data Bank (PDB) was disclosed (*22*). Several of the structures in this analysis included UU pairs. A variety of structures are observed, including structures with one and two hydrogen bonds as well as structures in which the U nucleotides are flipped out of the RNA helix altogether. Together with the results reported here, it shows that UU pairs adopt multiple conformations without distorting the global conformation of the surrounding nucleotides. A movie of our MD simulations (Supporting Information) shows that the UU pairs in the 5′CUG/3′GUC motif sample structures with zero, one, and two hydrogen bonds without breaking loop-closing base pairs.

In summary, NMR spectroscopy and MD simulations show that the 5′CUG/3′GUC motif found in DM1 RNAs is dynamic. The NMR structures suggest that ligand binding may proceed through conformational selection of RNA structures. These investigations provide a foundation for studying structures of ligands bound to DM1 RNAs.

ACKNOWLEDGMENT

We acknowledge Jon French and Amit Kumar for help with NMR experiments and Doug Kojetin, Anthony Carvalloza, and Andrew Davis for help with Amber.

SUPPORTING INFORMATION AVAILABLE

Protocols and a movie of the MD trajectory simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Thomas, J. R., and Hergenrother, P. (2008) *Chem. Rev. 108*, 1171–1224.
2. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) *Science 293*, 864–867.
3. Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., and Thornton, C. A. (2000) *Science 289*, 1769–1772.
4. Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001) *Nat. Genet. 29*, 40–47.
5. Lueck, J. D., Mankodi, A., Swanson, M. S., Thornton, C. A., and Dirksen, R. T. (2007) *J. Gen. Physiol. 129*, 79–94.
6. Wheeler, T. M., Sobczak, K., Lueck, J. D., Osborne, R. J., Lin, X., Dirksen, R. T., and Thornton, C. A. (2009) *Science 325*, 336–339.
7. Warf, M. B., Nakamori, M., Matthys, C. M., Thornton, C. A., and Berglund, J. A. (2009) *Proc. Natl. Acad. Sci. U.S.A. 106*, 18551–18556.
8. Arambula, J. F., Ramisetty, S. R., Baranger, A. M., and Zimmerman, S. C. (2009) *Proc. Natl. Acad. Sci. U.S.A 106*, 16068–16073.
9. Gareiss, P. C., Sobczak, K., McNaughton, B. R., Palde, P. B., Thornton, C. A., and Miller, B. L. (2008) *J. Am. Chem. Soc. 130*, 16254–16261.
10. Lee, M. M., Pushechnikov, A., and Disney, M. D. (2009) *ACS Chem. Biol. 4*, 345–355.
11. Pushechnikov, A., Lee, M. M., Childs-Disney, J. L., Sobczak, K., French, J. M., Thornton, C. A., and Disney, M. D. (2009) *J. Am. Chem. Soc. 131*, 9767–9779.
12. Disney, M. D., Lee, M. M., Pushechnikov, A., and Childs-Disney, J. L. (2010) *ChemBioChem 11*, 375–382.
13. Lee, M. M., Childs-Disney, J. L., Pushechnikov, A., French, J. M., Sobczak, K., Thornton, C. A., and Disney, M. D. (2009) *J. Am. Chem. Soc. 131*, 17464–17472.
14. Nonin, S., Leroy, J. L., and Gueron, M. (1995) *Biochemistry 34*, 10652–10659.
15. Schroeder, K. T., Skalicky, J. J., and Greenbaum, N. L. (2005) *RNA 11*, 1012–1016.
16. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) *J. Comput. Chem. 26*, 1668–1688.
17. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) *Comput. Phys. Commun. 92*, 1–41.
18. Limer, S. (1997) *Prog. Nucleic Acid Res. Mol. Biol. 57*, 1–39.
19. Fogolari, F., Brigo, A., and Molinari, H. (2002) *J. Mol. Recognit. 15*, 377–392.
20. Mooers, B. H., Logue, J. S., and Berglund, J. A. (2005) *Proc. Natl. Acad. Sci. U.S.A. 102*, 16626–16631.
21. Kiliszek, A., Kierzek, R., Krzyzosiak, W. J., and Rypniewski, W. (2009) *Nucleic Acids Res. 37*, 4149–4156.
22. Davis, A. R., Kirkpatrick, C. C., and Znosko, B. M. (2010) *Nucleic Acids Res. 28*, 1–14.