Published in final edited form as:

Biochemistry. 2017 July 11; 56(27): 3463–3474. doi:10.1021/acs.biochem.7b00252.

Structure and Dynamics of RNA Repeat Expansions That Cause Huntington's Disease and Myotonic Dystrophy Type 1

Jonathan L. Chen[†], **Damian M. VanEtten**[‡], **Matthew A. Fountain**[‡], **Ilyas Yildirim**[*,†,§], and **Matthew D. Disney**[*,†,§]

[†]Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States

[‡]Department of Chemistry and Biochemistry, State University of New York at Fredonia, Fredonia, New York 14063, United States

[§]Department of Chemistry and Biochemistry, Florida Atlantic University, Jupiter, Florida 33458, United States

Abstract

RNA repeat expansions cause a host of incurable, genetically defined diseases. The most common class of RNA repeats consists of trinucleotide repeats. These long, repeating transcripts fold into hairpins containing 1 × 1 internal loops that can mediate disease via a variety of mechanism(s) in which RNA is the central player. Two of these disorders are Huntington's disease and myotonic dystrophy type 1, which are caused by r(CAG) and r(CUG) repeats, respectively. We report the structures of two RNA constructs containing three copies of a r(CAG) [r(3×CAG)] or r(CUG) [r(3×CUG)] motif that were modeled with nuclear magnetic resonance spectroscopy and simulated annealing with restrained molecular dynamics. The 1 × 1 internal loops of r(3×CAG) are stabilized by one-hydrogen bond (*cis* Watson–Crick/Watson–Crick) AA pairs, while those of r(3×CUG) prefer one- or two-hydrogen bond (*cis* Watson–Crick/Watson–Crick) UU pairs. Assigned chemical shifts for the residues depended on the identity of neighbors or next nearest neighbors. Additional insights into the dynamics of these RNA constructs were gained by molecular dynamics simulations and a discrete path sampling method. Results indicate that the global structures of the RNA are A-form and that the loop regions are dynamic. The results will be useful for understanding the dynamic trajectory of these RNA repeats but also may aid in the development of therapeutics.

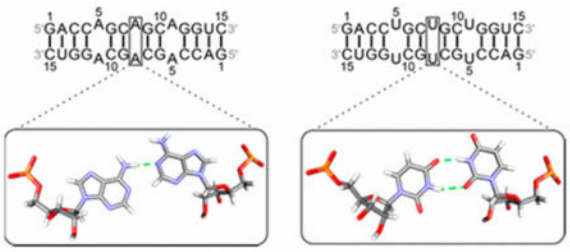
Graphical Abstract

[*] iyildirim@fau.edu. Phone: (561) 799-8325. [*] disney@scripps.edu. Phone: (561) 228-2203.

Accession Codes

NMR chemical shifts for r(3×CAG) and r(3×CUG) were deposited in the Biological Magnetic Resonance Data Bank as entries 30282 and 30283, respectively. Atomic coordinates for these constructs were deposited in the RCSB Protein Data Bank as entries 5VH7 and 5VH8, respectively.

The authors declare no competing financial interest.



RNA structure is involved in many cellular processes, such as catalyzing reactions,(1) synthesizing proteins,(2, 3) guiding chemical modifications of RNA,(4) and regulating gene expression.(5) RNA structure is also important in diseases, including cancers,(6) viral diseases,(7, 8) and neurodegenerative diseases.(9, 10) Long, tandem RNA repeats fold into hairpins that may undergo expansions and cause neurological diseases.(11) r(CAG) repeat expansions [r(CAG)^{exp}], located in exon 1 of the *HD* gene, are translated into polyglutamine tracts that form toxic aggregates in neurons in Huntington's disease.(12–15) The progression of these diseases may also be caused by RNA toxicity, such as mechanisms by which r(CAG)^{exp} sequesters SRSF6, causing aberrant splicing in the *HD* gene.(16) r(CUG) repeat expansions [r(CUG)^{exp}] located in the 3′ UTR of the *DMPK* gene(17–19) sequester MBNL1 or CUGBP1, causing aberrant splicing of genes such as the insulin receptor (IR) and cardiac troponin T (cTNT) in myotonic dystrophy type 1 (DM1).(20–23) r(CAG)^{exp} and/or r(CUG)^{exp} is also associated with several types of spinocerebellar ataxias (SCAs).(24–27) Targeting these repeat expansions with therapeutics may restore normal splicing and reverse the pathogenesis of these diseases.

Current treatments for diseases caused by r(CAG)^{exp} or r(CUG)^{exp} consist of symptom management.(28–31) As potential drugs, antisense oligonucleotides (ASOs) have had problems with efficacy, off-target effects, delivery, and side effects, in addition to enhanced toxicity from generation of small RNAs that target other repeat-containing transcripts.(32, 33) Furthermore, many mRNAs are not accessible to ASOs because they are highly structured or bound to proteins.(34) An alternative class of drugs consists of small molecules that bind to repeat expansions. Kumar et al.(35) identified a compound that inhibited formation of the MBNL1–r(CAG)^{exp} complex with micromolar potency. Pushechnikov et al.(36) and Childs-Disney et al.(37) reported dimeric and trimeric derivatives of Hoechst 33258 on peptoid backbones that selectively bound to r(CUG)^{exp} with nanomolar affinity and inhibited MBNL1 binding *in vitro. In vivo* activities of multimers of this compound on optimized scaffolds were later demonstrated.(38, 39) A chemical similarity search for compounds with improved bioactivity led to compound H1, which displaced MBNL1 from r(CUG)^{exp} with 20-fold improvement over that of Hoechst 33258.(40) Subsequently, Rzuczek et al.(41) reported small molecule modules that catalyze synthesis of a multivalent compound at adjacent sites in r(CUG)^{exp} for enhanced potency. Other compounds that target r(CUG)^{exp} include actinomycin D, aminoglycosides, and triaminotriazine derivatives.(42–46) Knowledge of the detailed structure and dynamics of these RNA repeats may assist with structure- or ligand-based design of drugs against these RNAs.(47)

Structures of r(CAG)^{exp} and r(CUG)^{exp} refined by X-ray crystallography and/or molecular dynamics simulations were previously reported. Kiliszek et al.(48) reported a crystal structure of an RNA duplex with two r(CAG) motifs, where each AA pair is *anti–anti* and forms one hydrogen bond. In a structure of an RNA construct with three copies of r(CAG), Yildirim et al.(49) found *anti–anti* AA pairs in the central r(CAG) motif and *syn–anti* AA pairs in the terminal r(CAG) motifs, flanked by Watson–Crick GC pairs. In comparison, several crystal structures of r(CUG) repeats have been reported. Mooers et al. reported an A-form r(CUG)$_6$ crystal structure with no direct hydrogen bonds within the UU loops and poor stacking with closing base pairs.(50) Kiliszek et al.(51) identified one hydrogen bond, "stretched UU wobble" base pairs, in detwinned crystal structures of r(CUG)$_6$. Coonrod et al.(52) utilized a GAAA tetraloop and receptor interaction to facilitate crystal packing of RNAs with r(CUG)$_2$, where the "stretched UU wobble" pair also predominated. Kumar et al. (53) developed crystals of RNA constructs with r(CUG)$_3$ and observed one- or two-hydrogen bond UU pairs depending on the orientation of the 5′ UU dangling ends. Molecular dynamics simulations on this construct indicated that the UU loops prefer one- or two-hydrogen bond states to a no-hydrogen bond state.(54)

Even though X-ray structures of r(CAG)^{exp} and r(CUG)^{exp} were previously reported, crystal packing forces may influence X-ray structures of RNAs.(1, 48, 52, 55) Furthermore, X-ray structures would represent the frozen state of RNA where dynamic properties would be missing. An alternative method for obtaining three-dimensional structures of these RNAs is nuclear magnetic resonance (NMR) spectroscopy. A solid state NMR study of a construct with 97 r(CUG) repeats revealed A-form geometry with C3′-*endo* geometry and an *anti* glycosidic torsion angle (χ).(56) NMR spectroscopy can also provide information about dynamics, which may be used to design therapeutics to target these RNA repeats.(57) Although conformations of RNAs in solution would not be influenced by crystal packing, signal overlap may complicate full assignment of resonances. Previously, NMR analysis of an RNA construct with one r(CUG) motif found the one- and two-hydrogen bond states to be preferred.(58) However, no NMR structures of r(CAG)^{exp} or r(CUG)^{exp} have yet been reported.

Herein, we report NMR structures of RNAs with three copies of r(CAG) or r(CUG). Furthermore, we utilized molecular dynamics (MD) simulations and a discrete path sampling (DPS) method to predict atomistic details of structure and dynamics of these model RNA systems. Results indicate that 5′C<u>A</u>G/3′G<u>A</u>C and 5′C<u>U</u>G/3′G<u>U</u>C loops interconvert between multiple states, but that the one-hydrogen bond state was preferred for the former and one- and two-hydrogen bond states were preferred for the latter. Furthermore, we discovered the NMR chemical shift data of RNA 5′C<u>A</u>G/3′G<u>A</u>C and 5′C<u>U</u>G/3′G<u>U</u>C mimicking the corresponding repeat expansions that can be used to design NMR experiments to study binding of small molecules to RNA repeat expansions in solutions. The results might be also applied to other challenging aspects of RNA repeats such as MBNL1–RNA 5′C<u>A</u>G/3′G<u>A</u>C and 5′C<u>U</u>G/3′G<u>U</u>C binding.

Methods

Preparation of Samples

The following RNA sequences were purchased from GE Dharmacon, Inc., and deprotected and desalted according to the manufacturer's recommended protocol: r(GACCAGCAGCAGGUC)$_2$[r(3×CAG)], r(GACCUGCUGCUGGUC)$_2$ [r(3×CUG)], r(GACCU^{5F}GCUGCUGGUC)$_2$, r(GACCUGCUGCU^{5F}GGUC)$_2$, r(GACCU^{5F}GCU^{5F}GCUGGUC)$_2$, and r(GACCU^{5F}GCUGCU^{5F}GGUC)$_2$. Desalted RNAs were then dissolved in NMR buffer [5 mM KH_2PO_4/K_2HPO_4 and 0.25 mM EDTA (pH 6.0)] and reannealed by being heated to 70 °C for 3 min and then slowly cooled to room temperature.

NMR Spectroscopy

NMR spectra of samples in Shigemi tubes (Shigemi, Inc.) were acquired on Bruker Avance III 700 and 850 MHz spectrometers. For samples in H_2O, one- and two-dimensional spectra were acquired with excitation sculpting to suppress the water signal.(59) Two-dimensional (2D) NOESY spectra with different mixing times were acquired at 25 and 3 °C or 5 °C for both constructs. Proton chemical shifts were referenced to the frequency of water. 2D NMR spectra were processed with Bruker BioSpin 3.2 and NMRPipe,(60) and resonances were assigned with SPARKY.(61)

Methods for Obtaining Distance and Dihedral Restraints

Distance restraints for pairs of hydrogen atoms were obtained by integrating nuclear Overhauser effect (NOE) volumes with SPARKY(61) or manually assigned to a range of distances based on the relative intensities of the NOEs. NOE volumes were referenced to volumes from fixed distances: H2′–H1′ (2.75 Å) and cytosine or uracil H5–H6 (2.45 Å).(62) Hydrogen bonds in canonical base pairs were assigned to distances of 2.1 ± 0.3 Å. The NOE intensities between the H1′ and H6/H8 base protons indicated that all nucleotides adopted an *anti* conformation. Therefore, the χ dihedral angle was constrained between 170° and 340° (*anti*) for all residues except the terminal residues and loop adenines or uridines.

Modeling Methods

Structures were calculated with a simulated annealing protocol(63, 64) on a starting structure generated with Nucleic Acid Builder.(65) The RNA was solvated with the generalized Born implicit solvent model and 0.1 M NaCl.(66) The system was heated from 0 to 3000 K in 5000 steps for 5 ps, cooled to 100 K in 93000 steps for 93 ps, and then cooled to 0 K in 2000 steps for 2 ps. Force constants were 20 kcal mol^{-1} Å^{-2} for NOE distance restraints and 20 kcal mol^{-1} rad^{-2} for dihedral angle restraints. Restrained molecular dynamics simulations were performed with AMBER(67)using the parm99χ_YIL force field.(68) The simulated annealing procedure was repeated with different initial velocities to generate an ensemble of 40 structures for r(3×CAG) and r(3×CUG). The 35 r(3×CAG) structures with no distance violations of >0.1 Å and all 40 r(3×CUG) structures were refined with the same simulated annealing procedure except that they were heated to 600 K. Force constants were 30 kcal mol^{-1} Å^{-2} for NOE distance restraints and 30 kcal mol^{-1} rad^{-2} for

dihedral angle restraints. The 20 lowest-energy structures that agreed with NMR experimental restraints were selected as the final ensemble of structures. Structures were analyzed with PyMOL, X3DNA, and DSSR.(69–71) Root-mean-square deviations (RMSDs) of the ensemble of structures were calculated with VMD.(72) Images for publication were generated with Discovery Studio Visualizer.(73)

Molecular Dynamics (MD) Simulations

Preparation of Model Systems—Model systems of r(GACCUGCUGCUGGUC) [r(3×CAG)] and r(GACCUGCUGCUGGUC) [r(3×CUG)] were prepared for MD simulations using the nucgen module of AMBER 16.(67) Initial structures were first minimized by restraining only the heavy atoms in Watson–Crick base pairs. The systems were then neutralized with 28 Na^+ ions(74) and solvated with 8813 TIP3P(75) water molecules in a truncated octahedral box. The Na^+ concentrations in both systems were 0.17 M once they had reached equilibration. The AMBER force field(76) with revised χ(68) and α/γ(77)torsional parameters was used in the MD simulations.

Molecular Dynamics Simulations—The structures were first minimized and then equilibrated each in two steps as described previously.(49, 78) After minimization and equilibration, a 1 μs MD simulation with a 2 ps time step was run for each system using *NPT* dynamics with isotropic positional scaling. The reference pressure was set to 1 atm with a pressure relaxation time of 2 ps. SHAKE was turned on for constraining bonds involving hydrogen atoms.(79) An atom-based long-range hard cutoff of 8.0 Å was used in the production runs. The reference temperature was set to 300 K.

Discrete Path Sampling—We previously utilized the discrete path sampling (DPS)(80, 81) method to study the energy landscape of 1 × 1 UU internal loops in $r(CUG)^{exp}$.(54) DPS is an approach to scanning configurational space efficiently to discover global and local minimum states of RNA systems, and transformation pathways that cannot be accessed with regular MD simulations. In the work presented here, DPS was applied to study the 1 × 1 AA internal loops. For the DPS calculations, initial conformations were extracted from the MD trajectory of r(3×CAG). The middle 1 × 1 AA internal loop region of r(3×CAG), $(5'\text{-}G_6C_7A_8G_9C_{10}\text{-}3')_2$, was extracted from the MD trajectory at intervals of 200 ps, yielding 5000 conformations. A modified version of the LBFGS algorithm(82)was used to build the initial database with a convergence criterion of 10^{-6} kcal mol^{-1} for the root-mean-square gradient. Between each minimum, connection attempts were made in a pairwise fashion. A disconnectivity graph(83–85) analysis was performed at the end that displayed the stable conformational states. The UNTRAP scheme(86) was utilized to further refine the database. For all these calculations, the OPTIM and PATHSAMPLE software was used (see Table S1 for sample input files). The final stationary point database contains −35K minima and −48K transition states. Free energies were estimated from the database by employing the harmonic superposition approximation.(87)

Analysis—The ptraj module of AMBER 16(67) was used for root-mean-square deviation (RMSD) and cluster analyses. X3DNA(70, 88) was used to extract global structural properties of RNA duplexes from the MD trajectories. A disconnectivity graph(83, 84) was

plotted to visualize the free energy landscapes for RNA 1 × 1 AA internal loops.(85, 89) The disconnectivity graph was constructed from the stationary point databases using the disconnectionDPS code (https://wikis.ch.cam.ac.uk/ro-walesdocs/wiki/index.php/DisconnectionDPS)(90) (see Table S1 for a sample input file).

Results

NMR Analysis of the r(3×CAG) Duplex

Nonexchangeable protons were assigned with NOESY spectra recorded in D_2O. A sequential H6/H8–H1′ walk through all residues of r(3×CAG) is shown (Figure 1 and Figure S1) as well as interresidue NOEs between H6/H8 and H2′, indicative of an A-form structure.(91) Key NOEs between the adenine H2 protons and nearby H1′ protons provided valuable information about the structure of r(3×CAG). Specifically, NOEs were observed from A2H2 to C3H1′ and C15H1′, A5H2 and A11H2 to G6H1′ and G12H1′, and A8H2 to G9H1′ (Figure 1 and Figure S1). Altogether, these NOEs suggest the helix adopts A-form geometry with stacking among the 5′CAG/3′GAC bases. Intra- and interresidue NOEs between equivalent H6/H8 and H1′ protons within each r(CAG) motif overlap are consistent with the three r(CAG)s having a repetitive structure.

The exchangeable proton assignments were made using NOESY spectra collected in a 95% H_2O/5% D_2O solvent. All imino proton resonances were observed with imino–amino NOEs present for all internal GC pairs (Figures S2 and S3). NOEs from G1H1 to one of the C15 amino protons and from A2H2 to U14H3 were also observed and used to assign and confirm amino and imino resonances. Intrastrand and interstrand guanine H1–3′ H1′ NOEs were observed for guanines in each r(CAG) motif. Specifically, NOEs were observed from G6H1 to C7H1′ and A11H1′, G9H1 to C10H1′ and A8H1′, and G12H1 to G13H1′ and A5H1′. These NOEs were valuable in defining the base pairing and the structural features of the r(CAG) triplet repeats.

NMR Analysis of the r(3×CUG) Duplex

NOEs corresponding to H1′–H6 and H5–H6 of U5, U8, U11, and C4 significantly overlap in the NOESY spectra of r(3×CUG) in D_2O. Nevertheless, a sequential H6/H8–H1′ walk was assigned through all nucleotides except C4 with the assistance of sequential aromatic base–base proton NOEs (Figure 2 and Figure S4). Same- and cross-strand NOEs exist between A2H2 and C3H1′ and C15H1′. As seen for r(3×CAG), intra- and interresidue NOEs between equivalent H6/H8 and H1′ protons of each r(CUG) motif overlap, indicating that all r(CUG)s have very similar structures. The NOEs and their intensities in the NOESY spectra clearly indicate an A-form RNA-like helix.

The NOESY spectrum recorded in a 95% H_2O/5% D_2O solvent shows imino proton resonances and imino–amino NOEs for all Watson–Crick base pairs (Figures S5 and S6). Additionally, NOEs from G1H1 to the C15 amino protons and A2H2 to U14H3 were observed. The G6, G9, and G13 imino proton resonances overlap. The two upfield imino proton peaks between 10.21 and 10.28 ppm (Figure 3) are consistent with UU base pairs.(58, 92, 93) The broader resonance at 10.28 ppm was assigned to the U5 and U8 imino

protons, with the U11 imino proton at 10.21 ppm. These peaks are broad, indicating exchange with solvent, and may be dynamic. NOEs from these U imino protons to G12H1′, G6H1′, and G9H1′ were observed, supporting stacking of the r(CUG) bases and providing important information for positioning the UU mismatch base pair.

The uridine imino proton resonance assignments were validated by systematically substituting uridines with 5-fluorouridine (U^{5F}). The electronegativity of the 5-fluorine atom is expected to lower the pK_a of the UH3 proton and increase its rate of exchange with the solvent, thus broadening its resonance.(94, 95) In the 1D spectra shown in Figure 3, single U^{5F} substitutions at U5 and U11 result in the loss of the peak at 10.21 ppm and the appearance of a peak at 9.88 ppm. U^{5F}substitution at both U5 and U8 results in the loss of the 10.28 and 10.21 ppm peaks and the appearance of a peak at 9.88 ppm. Having U^{5F} substitutions at U5 and U11 results in the loss of the resonance at 10.21 ppm, with the resonance at 10.28 ppm remaining. On the basis of the data and peak areas mentioned above, U11H3 is at 10.21 ppm, and the broader and larger resonance at 10.28 ppm must be U8 and U5. The resonances between 12.6 and 13.6 ppm are sensitive to the U^{5F} substitution at U5 and, to a lesser extent, U11 and U8. The remaining imino proton assignments were confirmed by NOEs observed in the 2D NOESY spectra, where NOEs from U11H3 to G12H1′ and G6H1′ were observed for r(3×CUG) with U^{5F} substituted at U5 and U8 (Figure S7) and an NOE from U8H3 to G9H1′ was observed for r(3×CUG) with U^{5F} substituted at U5 and U11 (Figure S8).

NMR Analysis of a Second Conformation for r(3×CAG) and r(3×CUG)

Additional cross-peaks were observed in the r(3×CAG) and r(3×CUG) NOESY spectra (Figures S2 and S4). Because of overlapping resonances and weak cross-peaks, the only protons that were assigned in these minor conformations were the aromatic protons of G1, A2, and C3, and H1 and H3 for r(3×CUG) and H6/H8 to H1′ and some H2′ protons of G1–G6, A11, and G13–C15 for r(3×CAG) (Figure S2). From this limited set of NMR data, the minor conformation in both repeats appears to be in slow exchange with the major conformation due to the absence of exchange cross-peaks between resonances in the spectra. Both systems have the same complementary ends, 5′GACC/3′CUGG, which could form a stem–loop structure with a four-nucleotide stem and a seven-nucleotide loop. On the basis of the limited set of NMR data, the second conformation in both systems may be a stem–loop structure. Evidence of a stem–loop structure is clearer for r(3×CUG) in that the imino proton resonances for U5 and U11 exhibit an NOE to each other, with one of the U imino protons exhibiting an NOE to an adjacent base pair amino proton, while the U8 imino proton at 11.58 ppm is extremely weak and exhibits no observable NOEs to adjacent nucleotides (Figure S5). If the minor conformation was a duplex, one would expect the U5, U8, and U11 imino protons to have similar chemical shift values and intensities, which the minor conformation lacks. The difference in chemical shift values for the U5, U8, and U11 imino protons and the weak U8 imino proton suggests that U5 and U11 are interacting with each other and U8 is most likely oriented externally at the top of the loop. Thus, a stem–loop structure is plausible. From the limited NOESY data given above and the lack of proton assignments for the loop nucleotides, the existence of a stem–loop structure can only be

hypothesized. Therefore, only the NMR results for the duplex conformation will be discussed further.

Modeling Study of r(3×CAG)

The simulated annealing procedure provided structures consistent with the NMR data (Figure S9). The average RMSD for the top 20 lowest-energy structures is 1.34 Å (Table 1), indicating good convergence. The average helical rise is 2.7 Å, and the helical twist is 30.6°, slightly less compressed than average for A-form RNA (2.8 Å and −32°).(96, 97) In the ensemble of the 20 lowest-free energy structures of r(3×CAG), the conformationally stable AA mismatches adopt *cis*Watson–Crick/Watson–Crick pairs(98) flanked by canonical GC pairs (Figure 4). All nucleotides in the structure adopt *anti* χ torsion angles, which is consistent with the presence of H6/H8–H1′ sequential NOE connectivity through all residues, without large H6/H8–H1′ NOEs that are found in the *syn* orientation. Some stacking of AA pairs on the neighboring GC pairs was observed, typical for an A-form helix (Figure 5).

The helix is unwound at either the 5′CA/GA or 5′AG/AC steps for all of the 5′CAG/3′GAC motifs, with helical twist values of less than −25°. In all of the structures, the average C1′–C1′ distance for the AA pairs is 12.2–12.4 Å. For typical A-form RNA, the C1′–C1′ distance is −10.5 Å.(58) The increased C1′–C1′ distance corresponds to an increased helical radius of ≳10 Å from −9.2 Å for A-form RNA.(71, 99) Near the internal 5′CAG/3′GAC motif, the major groove is the widest, at >20 Å on average, and the minor groove is the narrowest, at <16 Å on average. The AA pairs are characterized by large propeller twists (up to an average of −18.6°) from rotation approximately around the AN1–AN6 bond. The closing GC pairs of the central r(CAG) motif were buckled more (8.4° and −8.6°) than those of either of the outer r(CAG) motifs. These nonplanarities may enhance intrastrand base stacking of the mismatch adenines on neighboring bases.(100, 101)Taken together, the AA pairs induce distortions in the A-form helix.

Modeling Study of r(3×CUG)

From the simulated annealing calculations, 26 of the 40 structures adopt A-form helices and agreed with the NMR data (Figure S10). The average RMSD for all heavy atom superposition of the ensemble of structures is 1.51 Å (Table 1), again indicating good convergence. Similar to the case for r(3×CAG), the average helical rise is 2.7 Å and the average helical twist is 30.8°. In the ensemble of the 20 lowest-potential energy structures that agree with the NMR data, all of the UU mismatches contain *cis* Watson–Crick/Watson–Crick UU base pairs(98) closed by Watson–Crick GC pairs (Figure 6). The average number of hydrogen bonds in all of the UU base pairs is 1.5 ± 0.5. Although the UO4–UH3 distances in all of the UU pairs are 1.8–1.9 Å, the UO2–UH3 distances vary to a greater extent (Figure 7). Among the 20 lowest-potential energy structures, UO2–UH3 distances are <2.5 Å in 40% of UU pairs, between 2.5 and 3.4 Å in 40% of UU pairs, and >3.4 Å in the remaining 20%. The GC pairs adjacent to the internal r(CUG) motif are not buckled more than other base pair steps toward the end of the helix, but they have standard deviations for buckle larger than the deviations of those outside of the external r(CUG) motifs. Some stacking of one- and two-hydrogen bond UU pairs on neighboring GC pairs was observed.

Shorter C1′–C1′ distances were observed for UU pairs with shorter UO2–UH3 distances. For example, in one structure with UO2–UH3 distances of 2.10, 2.85, and 3.56 Å, the C1′–C1′ distances were 9.10, 10.10, and 10.79 Å, respectively. The helix is underwound at either the 5′CU/GU or 5′UG/UC steps of the 5′CUG/3′GUC motifs. A greater range of twist angles was observed in the 5′CU/3′GU and 5′UG/3′UC steps of r(3×CUG) than in the 5′CA/3′GA and 5′AG/3′AC steps of r(3×CAG), with larger standard deviations for each of those steps in r(3×CUG) than for equivalent steps in r(3×CAG). The major groove widens between the outer r(CUG) motifs, although not to the extent seen in r(3×CAG). Propeller twisting in the UU pairs, around the UN3–UO4 bond, is not as significant as in AA pairs of r(3×CAG). The helical radius for r(3×CUG) is typically ≳10 Å, similar to that of r(3×CAG). Altogether, the UU pairs induce distortions in the helix and have greater flexibility than AA pairs.

Molecular Dynamics Results

Figure 8 displays the RMSD analyses of 1 × 1 internal loops with respect to zero-, one, and two hydrogen bond states for r(3×CAG) and r(3×CUG) extracted from the MD trajectories. It was observed that 1 × 1 AA internal loops sampled the one- and zero-hydrogen bond states 37 and 47% of the time, respectively (Figure 8A–C), values similar to those previously reported.(49)Comparison of the terminal and middle 1 × 1 AA internal loops did not show any significant differences. As a result of this, the GC/CG Watson–Crick base pairs neighboring the AA mismatch are relatively flexible compared to GC pairs beyond the r(CAG) repeats.

On the other hand, the 1 × 1 UU internal loops sampled from the MD trajectory exhibited one-, two-, and zero-hydrogen bond states 44, 33, and 18% of the time, respectively, with the middle 1 × 1 UU internal loop sampling the two-hydrogen bond state 42% of the time (Figure 8D–F). When 1 × 1 UU internal loops preferred the two-hydrogen bond states, the neighboring GC Watson–Crick base pairs were observed to be distorted. Similar results were previously seen.(54) These results indicate that in $r(CUG)^{exp}$, 2 × 2 GC/CG base pairs between 1 × 1 UU internal loops can deform because of the formation of UU loops with two-hydrogen bond states. We should note, however, that the predictions might not be physical and that further testing of the force field will improve the prediction of mismatched base pairs.

We also performed base pair step analyses on these two systems (Figures S11 and S12). As expected, because of 1 × 1 AA and UU internal loops preferring more than one state, the loop regions are flexible. This behavior causes the loop regions to have large standard deviations compared to those of the Watson–Crick regions (see error bars in Figures S11 and S12). Specifically, analyses of slide and twist show that the loop regions have both high fluctuations and values, which are relatively slipped compared to the Watson–Crick regions (Figures S11 and S12). Nevertheless, combined with the groove width data, these results indicate both r(3×CAG) and r(3×CUG) systems have predominantly A-form structures with dynamic 1 × 1 internal loops, which correlate well with the observed dynamics in the NMR data.

Thermodynamic Properties of 1 × 1 AA Internal Loops

Utilizing a DPS approach, a disconnectivity graph that represents the hyperdimensional free energy landscape of 1 × 1 AA internal loops in r(CAG)^{exp} was generated (Figure 9). Figure 9displays both the global minima and local minimum states of 1 × 1 AA internal loops. This corresponds well with the NMR structures that indicate the one-hydrogen bond states are the global minima while the zero-hydrogen bond state is a local minimum with a 0.45 kcal mol $^{-1}$ less favorable free energy (Figure 9). Because this system is symmetric, both one-hydrogen bond states are observed (Figure 9). The transformation from the one-hydrogen bond state to a symmetric one-hydrogen bond state passes through the zero-hydrogen bond state (Movie S1). Similar results were predicted previously, which utilized MD and umbrella sampling calculations in the study of r(CAG)^{exp}.(49)

Discussion

NMR Chemical Shifts of Triplet Repeat Regions

There have been a number of studies that describe the use of proton chemical values to validate and help determine the structure of RNA.(102–106) Analysis of the chemical shifts of the r(3×CAG) and r(3×CUG) constructs indicates that the triplet repeat structure is repetitive. In r(3×CAG), C7 and C10 have very similar chemical shift values for their nonexchangeable H1′ (Δ0.001 ppm), H2′ (Δ0.002 ppm), H5 (Δ0.005 ppm), H6 (Δ0.006 ppm), and exchangeable amino protons (Figure 1 and Table S2). Nonexchangeable and imino proton resonances of the guanines to which they are base paired, G6 and G9, are also similar. For example, the H8 chemical shifts of G6 and G9 differ by <0.01 ppm, while their H1′ resonances have a slightly larger difference of 0.04 ppm. The nonexchangeable and exchangeable proton resonances of C4 and G12 differ because they are within a different sequence of triplets: 5′C<u>C</u>A/3′G<u>G</u>A for C4 and 5′A<u>G</u>C/3′A<u>C</u>G for G12. The proton chemical shifts for the adenines in all three triplets are also very similar: H1′ (Δ0.007 ppm), H2′ (Δ0.04 ppm), H2 (Δ0.04 ppm), and H8 (Δ0.03 ppm), indicating that they are located in nearly identical structural environments.

The repeating chemical shift values for the protons in these triplet repeats can be attributed to repetitive sequence-dependent stacking and/or electrostatic interactions, where G6 and C7, and G9 and C10, are located within the same triplet of base pairs.(102, 103, 105, 106) These triplets are 5′G<u>C</u>W/3′C<u>G</u>W for cytidines, 5′C<u>W</u>G/3′G<u>W</u>C for adenosines or uridines, and 5′W<u>G</u>C/3′W<u>C</u>G for the guanosines, where W is A or U. However, slight differences among chemical shifts of nucleotides within the same triplets may result from next nearest neighbor effects. Considering that the identity of the next nearest neighbors of each 5′C<u>A</u>G/3′G<u>A</u>C or 5′C<u>U</u>G/3′G<u>U</u>C motif are different in the modeled RNAs, the NMR chemical shifts, and to a larger extent the structure, of the central 5′C<u>A</u>G/3′G<u>A</u>C or 5′C<u>U</u>G/3′G<u>U</u>C motif may be the best approximation of those observed in repeat expansion diseases.

Comparison to Other r(CAG) Structures

Two RNA structures with r(CAG) repeats have been examined by X-ray crystallography and molecular dynamics modeling.(48, 49) Hydrogens were added to the structure of Kiliszek et

al. with REDUCE,(107) and helical parameters for both structures were calculated with X3DNA.(70)In the structure of Kiliszek et al., the AA pairs were *anti–anti* and stabilized by a single hydrogen bond from C2H2 to N1, but the AH6–AN1 distances were too long (3.4–3.5 Å) to form hydrogen bonds.(48) The distances between AC2 and AN1 in r(3×CAG) were similar to that reported by Kiliszek et al. (3.41 Å),(48) but the adenines were tilted to allow AH6 and AN1 in r(3×CAG) to be close enough to form a hydrogen bond. However, carbon is a poor hydrogen bond donor, so the N6–N1 hydrogen bonds likely stabilize the AA pairs more in r(3×CAG).

The average helical twist of the structures of Kiliszek et al.(48) and Yildirim et al.(49) are 28.6° and 27.7°, respectively, in close agreement with those of r(3×CAG). Helical rise is also consistent among the structures, in the range of 2.8–3.0 Å. Undertwisting around the AA pairs was observed in all of the structures. In the structure of Yildirim et al., *syn–anti* AA pairs in the terminal r(CAG) motifs formed from interactions of the amino and 2′-OH groups of the adenines with dangling uridines, which were included to crystallize the RNAs. Little stacking on closing base pairs was observed for each type of AA pair. The backbone of closing GC pairs of the r(CAG) motifs was flexible in the MD simulations of Yildirim et al. of an RNA with three r(CAG) motifs, likely facilitating *syn–anti* transitions of the loop adenines.(49) This flexibility was not observed in the RNA constructs with single r(CAG) motifs. Additionally, the 5′CG/GC and 5′GC/CG steps on either side of the internal r(CAG) motif are buckled more than either step outside of the other two r(CAG) motifs. Thus, the internal r(CAG) motif of r(3×CAG) is likely a better approximation of $r(CAG)^{exp}$.

Analyses of the MD trajectory of r(3×CAG) display dynamics in the 1 × 1 AA internal loops. One- and zero-hydrogen bond states are the most frequent structures observed in the MD simulation, while 10% of the structures were in other forms (Figure 8A–C). DPS results verify this prediction, too, where the one-hydrogen bond state is 0.45 kcal mol^{-1} more favorable than the zero-hydrogen bond state (Figure 9). The 1 × 1 AA internal loops in $r(CAG)^{exp}$ are symmetric. Thus, there are constant transformations from a one-hydrogen bond state to its symmetric version through the zero-hydrogen bond state (Movie S1 and Figure 8A–C). These motions, thus, create fluctuations in the base pair steps observed in the analyses (Figure S11). Despite these motions, the neighboring GC base pairs of 1 × 1 AA internal loops are not distorted as much as in r(CUG) repeats. This might be due to the favored π–π stacking observed between the adenosine and GC base pairs, which causes the system to be less affected by the fluctuations.

Comparison to Other r(CUG) Structures

In comparison to the model of r(3×CAG), the internal loops of r(3×CUG) are dynamic. The first crystal structure of r(CUG) repeats was reported by Mooers et al. in 2005.(50) The group reported that the UU pairs form water-mediated hydrogen bonds in the minor and major grooves. In their $r(CUG)_6$, the UU pairs adopt conformations to stabilize the helix by maximizing 5′GC/CG stacking and minimizing their own distortion, preventing formation of direct hydrogen bonding interactions. At the same time, 5′CU/GU stacking is minimized. Kiliszek et al. detwinned the X-ray data of $r(CUG)_6$ from the Berglund group and found that the UU pairs formed single-hydrogen bond “stretched UU wobble” base pairs, stabilized by

water molecules in the major and minor grooves.(51) In the r(3×CUG) structure, the 5′GC/CG step maintains relatively large intrastrand stacking like in A-form RNA. The one-hydrogen bond UU pairs observed in r(3×CUG) may be similar to the “stretched UU wobble”, although no explicit water molecules were used in the simulated annealing. Alternatively, water-mediated hydrogen bonding in one-hydrogen bond UU pairs may occur only in the minor groove.(93) In the UU pair with two hydrogen bonds, one uracil favors stacking on the cross-strand guanine rather than its adjacent cytosine while the other favors intrastrand stacking on its 5′-cytosine (Figures 6 and 7). There appears to be no correlation between the overlap area of the uracils on adjacent bases and the number of hydrogen bonds in the UU pairs. No UU pairs with an enol tautomeric form of an uridine were observed,(93, 108)although formation of this type of base pair may depend on the sequence.(93, 108)

The average helical rise twist of r(3×CUG) is consistent with those of previously published structures.(53, 58) Similar to the NMR structures, the two-hydrogen bond structure that Kumar et al. reported has a shorter C1′–C1′ distance (8.8 Å), which was attributed to widening of the adjacent major groove by the dangling UU residues, although this type of base pair observed in the crystal structure of Kumar et al. contains the same distortion of the residues as in the NMR structures.(53) C1′–C1′ distances of the structures with one or zero hydrogen bonds ranged from 10.0 to 10.5 Å in the crystal structure, consistent with one-hydrogen bond UU pairs in the NMR-refined structure. In the NMR structures of r(CUG) of Parkesh et al., the C1′–C1′ distance was also shorter in the two-hydrogen bond UU pair (8.9 Å) than in the one-hydrogen bond UU pair (10.6 Å).(58) Consistent with observations in r(3×CUG), this indicates that structures with a stronger hydrogen bond between UO2 and UH3 have greater distortion in the helix. The results from our overlap area analysis show our structures are in agreement with the reported crystal structures, with 5′CU/GU and 5′UG/UC steps of the r(CUG) motifs having considerably less overlap than the 5′GC/CG steps (Supporting Information). Altogether, these data support a model of r(3×CUG) that favors one- and two-hydrogen bond states, with dynamic UU pairs closed by structurally stable GC pairs.

A similar result was observed in the MD simulation of r(3×CUG). One- and two-hydrogen bond states of the noncanonical UU base pairs were the most frequent states observed in the MD trajectory, while the zero-hydrogen bond state was the least frequent (Figure 8C–E). Furthermore, the formation of the two-hydrogen bond state of the UU base pairs appears to induce a distortion of the neighboring GC base pairs to accommodate the less flexible UU base pair. The competition between base stacking (**π**–**π** interaction) and hydrogen bond formation (Coulomb interaction) is probably the main reason why we see dynamic behavior in 1 × 1 UU internal loops. Even though the general shape of r(3×CUG) is A-form, the local properties around 1 × 1 UU internal loops will direct how drugs, ligands, and proteins interact with the r(CUG) repeats. As an example, if the neighboring 2 × 2 GC base pairs next to 1 × 1 UU internal loops in $r(CUG)^{exp}$ are less stable than regular Watson–Crick GC base pairs, it might be possible that when a protein such as MBNL1 binds to $r(CUG)^{exp}$ or $r(CAG)^{exp}$, it distorts the neighboring GC base pairs to enhance its binding. The same results were observed in the disconnectivity graph of a model 1 × 1 UU internal loop (Figures S13 and S14).(54) This observation, however, needs to be experimentally tested in detail.

Conclusion

In summary, r(3×CAG) forms relatively stable single-hydrogen bond base pairs. On the other hand, the UU pairs of r(3×CUG) are dynamic and exchange between one- and two-hydrogen bond states. The AA and UU pairs induce distortions to A-form RNA of their helices. These results provide insight into the structure and dynamics of $r(CAG)^{exp}$ and $r(CUG)^{exp}$ and may be used to design therapeutics that reverse the symptoms of DM1, HD, or the SCAs that involve $r(CAG)^{exp}$and/or $r(CUG)^{exp}$ repeats. Furthermore, we discovered the NMR chemical shifts of 5′CAG/3′GAC and 5′CUG/3′GUC regions, which mimic the corresponding RNA repeat expansions in solution. Combined with NMR experiments, the NMR chemical shift data of r(CAG) and r(CUG) motifs can be utilized to design experiments to investigate how small molecules and ligands interact with RNA repeats. Finally, the results might be applied to future studies such as how disease-associated proteins interact with $r(CAG)^{exp}$ and $r(CUG)^{exp}$ and yield invaluable information about the binding mechanism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Information

This work was supported by the National Ataxia Foundation (J.L.C.), the Department of Chemistry and Biochemistry, Florida Atlantic University (I.Y.), National Institutes of Health Grant DP1NS096898 (M.D.D.), and Muscular Dystrophy Association Grant 380467 (M.D.D.).

We thank Dr. Xiangming Kong and Dr. Chunhua Yuan for assistance with NMR experiments. Minimization by simulated annealing and molecular dynamics simulations were performed with computational resources provided by The Scripps Research Institute and Florida Atlantic University. This study made use of the Campus Chemical Instrument Center NMR facility at The Ohio State University.

Abbreviations

ID	one-dimensional
2D	two-dimensional
ASO	antisense oligonucleotide
cTNT	cardiac troponin T
CUGBP1	CUG-binding protein 1
DM1	myotonic dystrophy type 1
DMPK	dystrophia myotonica protein kinase
DPS	discrete path sampling
HD	Huntington’s disease

IR	insulin receptor
MBNL1	muscleblind-like 1 protein
MD	molecular dynamics
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
RMSD	root-mean-square deviation
SCA	spinocerebellar ataxia
SRSF6	serine and arginine rich splicing factor 6

References

This article references 108 other publications.

1. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. Crystal structure of a group I ribozyme domain: Principles of RNA packing. Science. 1996; 273:1678.doi: 10.1126/science.273.5282.1678 [PubMed: 8781224]
2. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature. 2000; 407:340–348. DOI: 10.1038/35030019 [PubMed: 11014183]
3. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JHD, Noller HF. Crystal structure of the ribosome at 5.5 Å resolution. Science. 2001; 292:883.doi: 10.1126/science.1060089 [PubMed: 11283358]
4. Samarsky DA, Fournier MJ, Singer RH, Bertrand E. The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. EMBO J. 1998; 17:3747.doi: 10.1093/emboj/17.13.3747 [PubMed: 9649444]
5. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003; 425:415–419. DOI: 10.1038/nature01957 [PubMed: 14508493]
6. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature. 2005; 435:834–838. DOI: 10.1038/nature03702 [PubMed: 15944708]
7. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009; 460:711–716. DOI: 10.1038/nature08237 [PubMed: 19661910]
8. Moss WN, Priore SF, Turner DH. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. RNA. 2011; 17:991–1011. DOI: 10.1261/rna.2619511 [PubMed: 21536710]
9. Marquis Gacy A, Goellner G, Juranić N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. Cell. 1995; 81:533–540. DOI: 10.1016/0092-8674(95)90074-8 [PubMed: 7758107]
10. Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S, Chakraverty S, Isaacs A, Grover A, Hackett J, Adamson J, Lincoln S, Dickson D, Davies P, Petersen RC, Stevens M, de Graaff E, Wauters E, van Baren J, Hillebrand M, Joosse M, Kwon JM, Nowotny P, Che LK, Norton J, Morris JC, Reed LA, Trojanowski J, Basun H, Lannfelt L, Neystat M, Fahn S, Dark F, Tannenberg T, Dodd PR, Hayward N, Kwok JBJ, Schofield PR, Andreadis A, Snowden J, Craufurd D, Neary D, Owen F, Oostra BA, Hardy J, Goate A, van Swieten J, Mann D, Lynch T,

Heutink P. Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature. 1998; 393:702–705. DOI: 10.1038/31508 [PubMed: 9641683]

11. Sobczak K, de Mezer M, Michlewski G, Krol J, Krzyzosiak WJ. RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Res. 2003; 31:5469–5482. DOI: 10.1093/nar/gkg766 [PubMed: 14500809]
12. MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, Barnes G, Taylor SA, James M, Groot N, MacFarlane H, Jenkins B, Anderson MA, Wexler NS, Gusella JF, Bates GP, Baxendale S, Hummerich H, Kirby S, North M, Youngman S, Mott R, Zehetner G, Sedlacek Z, Poustka A, Frischauf A-M, Lehrach H, Buckler AJ, Church D, Doucette-Stamm L, O'Donovan MC, Riba-Ramirez L, Shah M, Stanton VP, Strobel SA, Draths KM, Wales JL, Dervan P, Housman DE, Altherr M, Shiang R, Thompson L, Fielder T, Wasmuth JJ, Tagle D, Valdes J, Elmer L, Allard M, Castilla L, Swaroop M, Blanchard K, Collins FS, Snell R, Holloway T, Gillespie K, Datson N, Shaw D, Harper PS. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. 1993; 72:971–983. DOI: 10.1016/0092-8674(93)90585-E [PubMed: 8458085]
13. Orr HT, Chung MY, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LP, Zoghbi HY. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nat Genet. 1993; 4:221–226. DOI: 10.1038/ng0793-221 [PubMed: 8358429]
14. Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, Lawton M, Trottier Y, Lehrach H, Davies SW, Bates GP. Exon 1 of the *HD* gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell. 1996; 87:493–506. DOI: 10.1016/S0092-8674(00)81369-0 [PubMed: 8898202]
15. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, Bates GP, Davies SW, Lehrach H, Wanker EE. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates *in vitro* and *in vivo*. Cell. 1997; 90:549–558. DOI: 10.1016/S0092-8674(00)80514-0 [PubMed: 9267034]
16. Gipson TA, Neueder A, Wexler NS, Bates GP, Housman D. Aberrantly spliced *HTT*, a new player in Huntington's disease pathogenesis. RNA Biol. 2013; 10:1647–1652. DOI: 10.4161/rna.26706 [PubMed: 24256709]
17. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion J-P, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell. 1992; 68:799–808. DOI: 10.1016/0092-8674(92)90154-5 [PubMed: 1310900]
18. Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Earle-MacDonald J, De Jong PJ, Wieringa B, Korneluk RG. Myotonic dystrophy mutation: An unstable CTG repeat in the 3′ untranslated region of the gene. Science. 1992; 255:1253–1255. DOI: 10.1126/science.1546325 [PubMed: 1546325]
19. Davis BM, McCurrach ME, Taneja KL, Singer RH, Housman DE. Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. Proc Natl Acad Sci U S A. 1997; 94:7388–7393. DOI: 10.1073/pnas.94.14.7388 [PubMed: 9207101]
20. Philips AV, Timchenko LT, Cooper TA. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science. 1998; 280:737–741. DOI: 10.1126/science.280.5364.737 [PubMed: 9563950]
21. Timchenko NA, Cai ZJ, Welm AL, Reddy S, Ashizawa T, Timchenko LT. RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. J Biol Chem. 2001; 276:7820–7826. DOI: 10.1074/jbc.M005960200 [PubMed: 11124939]
22. Savkur RS, Philips AV, Cooper TA. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. Nat Genet. 2001; 29:40–47. DOI: 10.1038/ng704 [PubMed: 11528389]
23. Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, Cooper TA. Muscleblind proteins regulate alternative splicing. EMBO J. 2004; 23:3103–3112. DOI: 10.1038/sj.emboj.7600300 [PubMed: 15257297]

24. Trottier Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, Saudou F, Weber C, David G, Tora L, Agid Y, Brice A, Mandel J-L. Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. Nature. 1995; 378:403–406. DOI: 10.1038/378403a0 [PubMed: 7477379]
25. Mykowska A, Sobczak K, Wojciechowska M, Kozlowski P, Krzyzosiak WJ. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. Nucleic Acids Res. 2011; 39:8938–8951. DOI: 10.1093/nar/gkr608 [PubMed: 21795378]
26. Tsoi H, Lau TCK, Tsang SY, Lau KF, Chan HYE. CAG expansion induces nucleolar stress in polyglutamine diseases. Proc Natl Acad Sci U S A. 2012; 109:13428–13433. DOI: 10.1073/pnas.1204089109 [PubMed: 22847428]
27. Daughters RS, Tuttle DL, Gao W, Ikeda Y, Moseley ML, Ebner TJ, Swanson MS, Ranum LPW. RNA gain-of-function in spinocerebellar ataxia type 8. PLoS Genet. 2009; 5:e1000600.doi: 10.1371/journal.pgen.1000600 [PubMed: 19680539]
28. Logigian EL, Martens WB, Moxley RT, McDermott MP IV, Dilek N, Wiegner AW, Pearson AT, Barbieri CA, Annis CL, Thornton CA, Moxley RT III . Mexiletine is an effective antimyotonia treatment in myotonic dystrophy type 1. Neurology. 2010; 74:1441–1448. DOI: 10.1212/WNL.0b013e3181dc1a3a [PubMed: 20439846]
29. Mulders SAM, van Engelen BGM, Wieringa B, Wansink DG. Molecular therapy in myotonic dystrophy: Focus on RNA gain-of-function. Hum Mol Genet. 2010; 19:R90–R97. DOI: 10.1093/hmg/ddq161 [PubMed: 20406734]
30. Fan HC, Ho LI, Chi CS, Chen SJ, Peng GS, Chan TM, Lin SZ, Harn HJ. Polyglutamine (polyQ) diseases: Genetics to treatments. Cell Transplant. 2014; 23:441–458. DOI: 10.3727/096368914X678454 [PubMed: 24816443]
31. Turner C, Hilton-Jones D. Myotonic dystrophy: Diagnosis, management and new therapies. Curr Opin Neurol. 2014; 27:599–606. DOI: 10.1097/WCO.0000000000000128 [PubMed: 25121518]
32. Rayburn ER, Zhang R. Antisense, RNAi, and gene silencing strategies for therapy: Mission possible or impossible? Drug Discovery Today. 2008; 13:513–521. DOI: 10.1016/j.drudis.2008.03.014 [PubMed: 18549978]
33. Yu Z, Teng X, Bonini NM. Triplet repeat-derived siRNAs enhance RNA-mediated toxicity in a Drosophila model for myotonic dystrophy. PLoS Genet. 2011; 7:e1001340.doi: 10.1371/journal.pgen.1001340 [PubMed: 21437269]
34. Gallego J, Varani G. Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges. Acc Chem Res. 2001; 34:836–843. DOI: 10.1021/ar000118k [PubMed: 11601968]
35. Kumar A, Parkesh R, Sznajder LJ, Childs-Disney JL, Sobczak K, Disney MD. Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. ACS Chem Biol. 2012; 7:496–505. DOI: 10.1021/cb200413a [PubMed: 22252896]
36. Pushechnikov A, Lee MM, Childs-Disney JL, Sobczak K, French JM, Thornton CA, Disney MD. 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. 2009; 131:9767–9779. DOI: 10.1021/ja9020149 [PubMed: 19552411]
37. Childs-Disney JL, Tsitovich PB, Disney MD. Using modularly assembled ligands to bind RNA internal loops separated by different distances. ChemBioChem. 2011; 12:2143–2146. DOI: 10.1002/cbic.201100298 [PubMed: 21830289]
38. Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, Disney MD. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. ACS Chem Biol. 2012; 7:856–862. DOI: 10.1021/cb200408a [PubMed: 22332923]
39. Rzuczek SG, Gao Y, Tang ZZ, Thornton CA, Kodadek T, Disney MD. Features of modularly assembled compounds that impart bioactivity against an RNA target. ACS Chem Biol. 2013; 8:2312–2321. DOI: 10.1021/cb400265y [PubMed: 24032410]
40. Parkesh R, Childs-Disney JL, Nakamori M, Kumar A, Wang E, Wang T, Hoskins J, Tran T, Housman D, Thornton CA, Disney MD. Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database and chemical similarity searching. J Am Chem Soc. 2012; 134:4731–4742. DOI: 10.1021/ja210088v [PubMed: 22300544]

41. Rzuczek SG, Colgan LA, Nakai Y, Cameron MD, Furling D, Yasuda R, Disney MD. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. Nat Chem Biol. 2017; 13:188–193. DOI: 10.1038/nchembio.2251 [PubMed: 27941760]
42. Arambula JF, Ramisetty SR, Baranger AM, Zimmerman SC. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. Proc Natl Acad Sci U S A. 2009; 106:16068–16073. DOI: 10.1073/pnas.0901824106 [PubMed: 19805260]
43. Childs-Disney JL, Parkesh R, Nakamori M, Thornton CA, Disney MD. Rational design of bioactive, modularly assembled aminoglycosides targeting the RNA that causes myotonic dystrophy type 1. ACS Chem Biol. 2012; 7:1984–1993. DOI: 10.1021/cb3001606 [PubMed: 23130637]
44. Wong CH, Nguyen L, Peh J, Luu LM, Sanchez JS, Richardson SL, Tuccinardi T, Tsoi H, Chan WY, Chan HYE, Baranger AM, Hergenrother PJ, Zimmerman SC. Targeting toxic RNAs that cause myotonic dystrophy type 1 (DM1) with a bisamidinium inhibitor. J Am Chem Soc. 2014; 136:6355–6361. DOI: 10.1021/ja5012146 [PubMed: 24702247]
45. Siboni RB, Nakamori M, Wagner SD, Struck AJ, Coonrod LA, Harriott SA, Cass DM, Tanner MK, Berglund JA. Actinomycin D specifically reduces expanded CUG repeat RNA in myotonic dystrophy models. Cell Rep. 2015; 13:2386–2394. DOI: 10.1016/j.celrep.2015.11.028 [PubMed: 26686629]
46. Luu LM, Nguyen L, Peng S, Lee J, Lee HY, Wong C-H, Hergenrother PJ, Chan HYE, Zimmerman SC. A potent inhibitor of protein sequestration by expanded triplet (CUG) repeats that shows phenotypic improvements in a Drosophila model of myotonic dystrophy. ChemMedChem. 2016; 11:1428–1435. DOI: 10.1002/cmdc.201600081 [PubMed: 27245480]
47. Angelbello AJ, González ÀL, Rzuczek SG, Disney MD. Development of pharmacophore models for small molecules targeting RNA: Application to the RNA repeat expansion in myotonic dystrophy type 1. Bioorg Med Chem Lett. 2016; 26:5792–5796. DOI: 10.1016/j.bmcl.2016.10.037 [PubMed: 27839685]
48. Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. Atomic resolution structure of CAG RNA repeats: Structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic Acids Res. 2010; 38:8370–8376. DOI: 10.1093/nar/gkq700 [PubMed: 20702420]
49. Yildirim I, Park H, Disney MD, Schatz GC. A dynamic structural model of expanded RNA CAG repeats: A refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. J Am Chem Soc. 2013; 135:3528–3538. DOI: 10.1021/ja3108627 [PubMed: 23441937]
50. Mooers BHM, Logue JS, Berglund JA. The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. Proc Natl Acad Sci U S A. 2005; 102:16626–16631. DOI: 10.1073/pnas.0505873102 [PubMed: 16269545]
51. Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. Structural insights into CUG repeats containing the 'stretched U-U wobble': implications for myotonic dystrophy. Nucleic Acids Res. 2009; 37:4149–4156. DOI: 10.1093/nar/gkp350 [PubMed: 19433512]
52. Coonrod LA, Lohman JR, Berglund JA. Utilizing the GAAA tetraloop/receptor to facilitate crystal packing and determination of the structure of a CUG RNA helix. Biochemistry. 2012; 51:8330–8337. DOI: 10.1021/bi300829w [PubMed: 23025897]
53. Kumar A, Park H, Fang P, Parkesh R, Guo M, Nettles KW, Disney MD. Myotonic dystrophy type 1 RNA crystal structures reveal heterogeneous 1 × 1 nucleotide UU internal loop conformations. Biochemistry. 2011; 50:9928–9935. DOI: 10.1021/bi2013068 [PubMed: 21988728]
54. Yildirim I, Chakraborty D, Disney MD, Wales DJ, Schatz GC. Computational investigation of RNA CUG repeats responsible for myotonic dystrophy 1. J Chem Theory Comput. 2015; 11:4943–4958. DOI: 10.1021/acs.jctc.5b00728 [PubMed: 26500461]
55. Shah SA, Brunger AT. The 1.8 Å crystal structure of a statically disordered 17 base-pair RNA duplex: Principles of RNA crystal packing and its effect on nucleic acid structure1. J Mol Biol. 1999; 285:1577–1588. DOI: 10.1006/jmbi.1998.2385 [PubMed: 9917398]
56. Riedel K, Herbst C, Häfner S, Leppert J, Ohlenschläger O, Swanson MS, Görlach M, Ramachandran R. Constraints on the structure of $(CUG)_{97}$ RNA from magic-angle-spinning solid-state NMR spectroscopy. Angew Chem, Int Ed. 2006; 45:5620–5623. DOI: 10.1002/anie.200600769

57. Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee JH, Andricioaei I, Markovitz DM, Al-Hashimi HM. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat Chem Biol. 2011; 7:553–559. DOI: 10.1038/nchembio.596 [PubMed: 21706033]
58. Parkesh R, Fountain M, Disney MD. NMR spectroscopy and molecular dynamics simulation of r(CCGCUGCGG)$_2$ reveal a dynamic UU internal loop found in myotonic dystrophy type 1. Biochemistry. 2011; 50:599–601. DOI: 10.1021/bi101896j [PubMed: 21204525]
59. Hwang TL, Shaka AJ. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. J Magn Reson, Ser A. 1995; 112:275–279. DOI: 10.1006/jmra.1995.1047
60. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6:277–293. DOI: 10.1007/BF00197809 [PubMed: 8520220]
61. Goddard, TD., Kneller, DG. SPARKY, NMR Assignment and Integration Software, version 3. University of California; San Francisco, San Francisco: 2004.
62. Kennedy SD, Kierzek R, Turner DH. Novel conformation of an RNA structural switch. Biochemistry. 2012; 51:9257–9259. DOI: 10.1021/bi301372t [PubMed: 23134175]
63. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science. 1983; 220:671–680. DOI: 10.1126/science.220.4598.671 [PubMed: 17813860]
64. Schmitz M, Steger G. Description of RNA folding by “Simulated Annealing”. J Mol Biol. 1996; 255:254–266. DOI: 10.1006/jmbi.1996.0021 [PubMed: 8568872]
65. Macke, TJ., Case, DA. Molecular Modeling of Nucleic Acids. American Chemical Society; Washington, DC: 1997. Molecular Modeling of Nucleic Acids; p. 379-393.
66. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc. 1990; 112:6127–6129. DOI: 10.1021/ja00172a038
67. Case, DA., Betz, RM., Botello-Smith, W., Cerutti, DS., Cheatham, TEI., Darden, TA., Duke, RE., Giese, TJ., Gohlke, H., Goetz, AW., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, TS., LeGrand, S., Li, C., Lin, T., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, KM., Monard, G., Nguyen, H., Nguyen, HT., Omelyan, I., Onufriev, A., Roe, DR., Roitberg, A., Sagui, C., Simmerling, CL., Swails, J., Walker, RC., Wang, J., Wolf, RM., Wu, X., Xiao, L., York, DM., Kollman, PA. AMBER 16. University of California; San Francisco, San Francisco: 2016.
68. Yildirim I, Stern HA, Kennedy SD, Tubbs JD, Turner DH. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. J Chem Theory Comput. 2010; 6:1520–1531. DOI: 10.1021/ct900604a [PubMed: 20463845]
69. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; San Carlos, CA: 2002.
70. Lu XJ, Olson WK. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 2003; 31:5108–5121. DOI: 10.1093/nar/gkg680 [PubMed: 12930962]
71. Lu X-J, Bussemaker HJ, Olson WK. DSSR: An integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Res. 2015; 43:e142.doi: 10.1093/nar/gkv716 [PubMed: 26184874]
72. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graphics. 1996; 14:33–38. DOI: 10.1016/0263-7855(96)00018-5
73. Discovery Studio Visualizer. Dassault Systèmes BIOVIA, version 16.1.0.15350. Dassault Systèmes; San Diego: 2015.
74. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B. 2008; 112:9020–9041. DOI: 10.1021/jp8001614 [PubMed: 18593145]
75. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926–935. DOI: 10.1063/1.445869

76. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc. 1995; 117:5179–5197. DOI: 10.1021/ja00124a002
77. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE III, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. Biophys J. 2007; 92:3817–3829. DOI: 10.1529/biophysj.106.097782 [PubMed: 17351000]
78. Yildirim I, Stern HA, Sponer J, Spackova N, Turner DH. Effects of restrained sampling space and nonplanar amino groups on free-energy predictions for RNA with imino and sheared tandem GA base pairs flanked by GC, CG, iGiC or iCiG base pairs. J Chem Theory Comput. 2009; 5:2088–2100. DOI: 10.1021/ct800540c [PubMed: 20090924]
79. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys. 1977; 23:327–341. DOI: 10.1016/0021-9991(77)90098-5
80. Wales DJ. Discrete path sampling. Mol Phys. 2002; 100:3285–3305. DOI: 10.1080/00268970210162691
81. Wales DJ. Some further applications of discrete path sampling to cluster isomerization. Mol Phys. 2004; 102:891–908. DOI: 10.1080/00268970410001703363
82. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Prog. 1989; 45:503–528. DOI: 10.1007/BF01589116
83. Becker OM, Karplus M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. J Chem Phys. 1997; 106:1495–1517. DOI: 10.1063/1.473299
84. Wales DJ, Miller MA, Walsh TR. Archetypal energy landscapes. Nature. 1998; 394:758–760. DOI: 10.1038/29487
85. Krivov SV, Karplus M. Free energy disconnectivity graphs: Application to peptide models. J Chem Phys. 2002; 117:10894–10903. DOI: 10.1063/1.1517606
86. Strodel B, Whittleston CS, Wales DJ. Thermodynamics and kinetics of aggregation for the GNNQQNY peptide. J Am Chem Soc. 2007; 129:16005–16014. DOI: 10.1021/ja075346p [PubMed: 18052168]
87. Strodel B, Wales DJ. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. Chem Phys Lett. 2008; 466:105–115. DOI: 10.1016/j.cplett.2008.10.085
88. Lu XJ, Olson WK. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat Protoc. 2008; 3:1213–1227. DOI: 10.1038/nprot.2008.104 [PubMed: 18600227]
89. Evans DA, Wales DJ. Free energy landscapes of model peptides and proteins. J Chem Phys. 2003; 118:3891–3897. DOI: 10.1063/1.1540099
90. Miller, M., Wales, DJ. [accessed February 15, 2017] disconnectionDPS. http://www-wales.ch.cam.ac.uk/software.html
91. Fürtig B, Richter C, Wohnert J, Schwalbe H. NMR spectroscopy of RNA. ChemBioChem. 2003; 4:936–962. DOI: 10.1002/cbic.200300700 [PubMed: 14523911]
92. SantaLucia J, Kierzek R, Turner DH. Stabilities of consecutive A·C, C·C, G·G, U·C, and U·U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U·U and C·C+ pairs. Biochemistry. 1991; 30:8242–8251. DOI: 10.1021/bi00247a021 [PubMed: 1714301]
93. Wang Y-X, Huang S, Draper DE. Structure of a U·U pair within a conserved ribosomal RNA hairpin. Nucleic Acids Res. 1996; 24:2666–2672. DOI: 10.1093/nar/24.14.2666 [PubMed: 8758993]
94. Hennig M, Scott LG, Sperling E, Bermel W, Williamson JR. Synthesis of 5-fluoropyrimidine nucleotides as sensitive NMR probes of RNA structure. J Am Chem Soc. 2007; 129:14911–14921. DOI: 10.1021/ja073825i [PubMed: 17990877]
95. Puffer B, Kreutz C, Rieder U, Ebert M-O, Konrat R, Micura R. 5-fluoro pyrimidines: Labels to probe DNA and RNA secondary structures by 1D ^{19}F NMR spectroscopy. Nucleic Acids Res. 2009; 37:7728–7740. DOI: 10.1093/nar/gkp862 [PubMed: 19843610]

96. Fessl T, Lilley DMJ. Measurement of the change in twist at a helical junction in RNA using the orientation dependence of FRET. Biophys J. 2013; 105:2175–2181. DOI: 10.1016/j.bpj.2013.09.042 [PubMed: 24209863]

97. Liebl K, Drsata T, Lankas F, Lipfert J, Zacharias M. Explaining the striking difference in twist-stretch coupling between DNA and RNA: A comparative molecular dynamics analysis. Nucleic Acids Res. 2015; 43:10143–10156. DOI: 10.1093/nar/gkv1028 [PubMed: 26464435]

98. Leontis NB, Stombaugh J, Westhof E. The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. 2002; 30:3497–3531. DOI: 10.1093/nar/gkf481 [PubMed: 12177293]

99. Nikolova EN, Zhou H, Gottardo FL, Alvey HS, Kimsey IJ, Al-Hashimi HM. A historical account of hoogsteen base-pairs in duplex DNA. Biopolymers. 2013; 99:955–968. DOI: 10.1002/bip.22334 [PubMed: 23818176]

100. Dickerson RE, Klug A. Base sequence and helix structure variation in B and A DNA. J Mol Biol. 1983; 166:419–441. DOI: 10.1016/S0022-2836(83)80093-X [PubMed: 6854650]

101. Šponer J, Leszczynski J, Hobza P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. Biopolymers. 2001; 61:3–31. DOI: 10.1002/1097-0282(2001)61:1<3::AID-BIP10048>3.0.CO;2-4 [PubMed: 11891626]

102. Cromsigt JAMTC, Hilbers CW, Wijmenga SS. Prediction of proton chemical shifts in RNA - Their use in structure refinement and validation. J Biomol NMR. 2001; 21:11–29. DOI: 10.1023/A:1011914132531 [PubMed: 11693565]

103. Barton S, Heng X, Johnson BA, Summers MF. Database proton NMR chemical shifts for RNA signal assignment and validation. J Biomol NMR. 2013; 55:33–46. DOI: 10.1007/s10858-012-9683-9 [PubMed: 23180050]

104. Sripakdeevong P, Cevec M, Chang AT, Erat MC, Ziegeler M, Zhao Q, Fox GE, Gao X, Kennedy SD, Kierzek R, Nikonowicz EP, Schwalbe H, Sigel RKO, Turner DH, Das R. Structure determination of noncanonical RNA motifs guided by ^{1}H NMR chemical shifts. Nat Methods. 2014; 11:413–416. DOI: 10.1038/nmeth.2876 [PubMed: 24584194]

105. Brown JD, Summers MF, Johnson BA. Prediction of hydrogen and carbon chemical shifts from RNA using database mining and support vector regression. J Biomol NMR. 2015; 63:39–52. DOI: 10.1007/s10858-015-9961-4 [PubMed: 26141454]

106. Chen JL, Bellaousov S, Tubbs JD, Kennedy SD, Lopez MJ, Mathews DH, Turner DH. Nuclear magnetic resonance-assisted prediction of secondary structure for RNA: Incorporation of direction-dependent chemical shift constraints. Biochemistry. 2015; 54:6769–6782. DOI: 10.1021/acs.biochem.5b00833 [PubMed: 26451676]

107. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol. 1999; 285:1735–1747. DOI: 10.1006/jmbi.1998.2401 [PubMed: 9917408]

108. Sheng J, Larsen A, Heuberger BD, Blain JC, Szostak JW. Crystal structure studies of RNA duplexes containing s^{2}U:A and s^{2}U:U base pairs. J Am Chem Soc. 2014; 136:13916–13924. DOI: 10.1021/ja508015a [PubMed: 25188906]

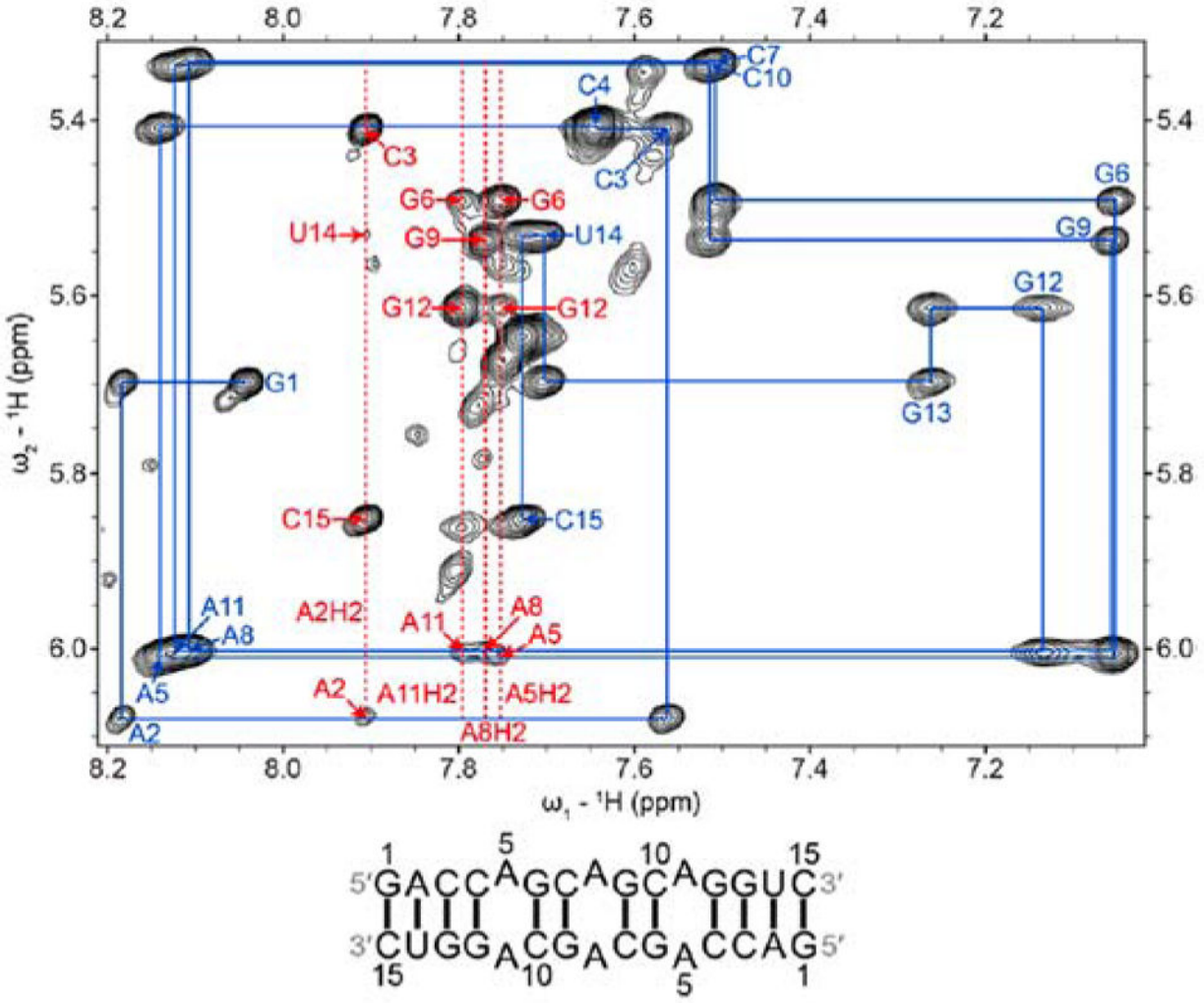


Figure 1.

H1′–H6/H8 region of a 2D proton NOESY spectrum of r(3×CAG) showing a sequential proton walk with blue lines. H1′–H6/H8 walk NOEs are labeled in blue. Adenine H2 signals are labeled with red dashed hnes. H1′–adenine H2 NOEs are labeled in red with only the label of the residue for H1′.

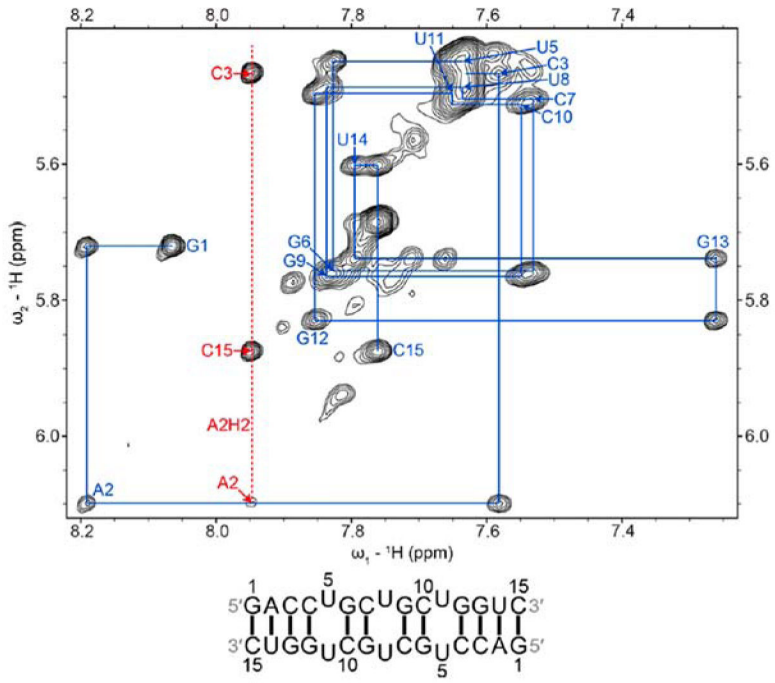


Figure 2.

H1′–H6/H8 region of a 2D proton NOESY spectrum of r(3×CUG) showing a sequential proton walk with blue lines. H1′–H6/H8 walk NOEs are labeled in blue. Adenine H2 signals are labeled with red dashed lines. H1′–adenine H2 NOEs are labeled in red with only the label of the residue for H1′. Cross-peaks from C4H1′ to C4H6 and U5H6 were not observed in the spectrum.

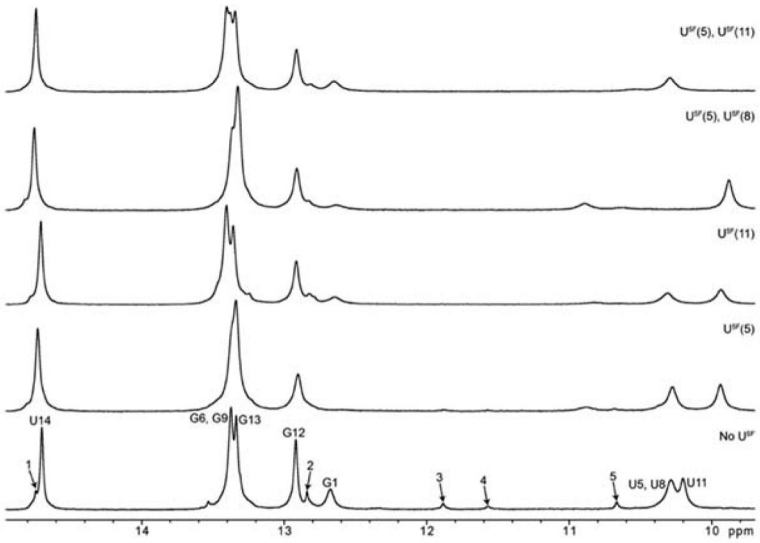


Figure 3.

NMR spectra of unsubstituted and U^{5F}-substituted r(3×CUG) RNAs. Spectra were recorded at 3 °C. The UU pair imino protons are at 10.2–10.3 ppm: (bottom) unmodified RNA with no U^{5F}, (middle) U^{5F} at positions 5 and 8, and (top) U^{5F} at positions 5 and 11. Peaks labeled 1–5 correspond to the minor conformation in solution: (1) U14, (2) G12, (3) U5 or U11, (4) U8, and (5) U5 or U11.

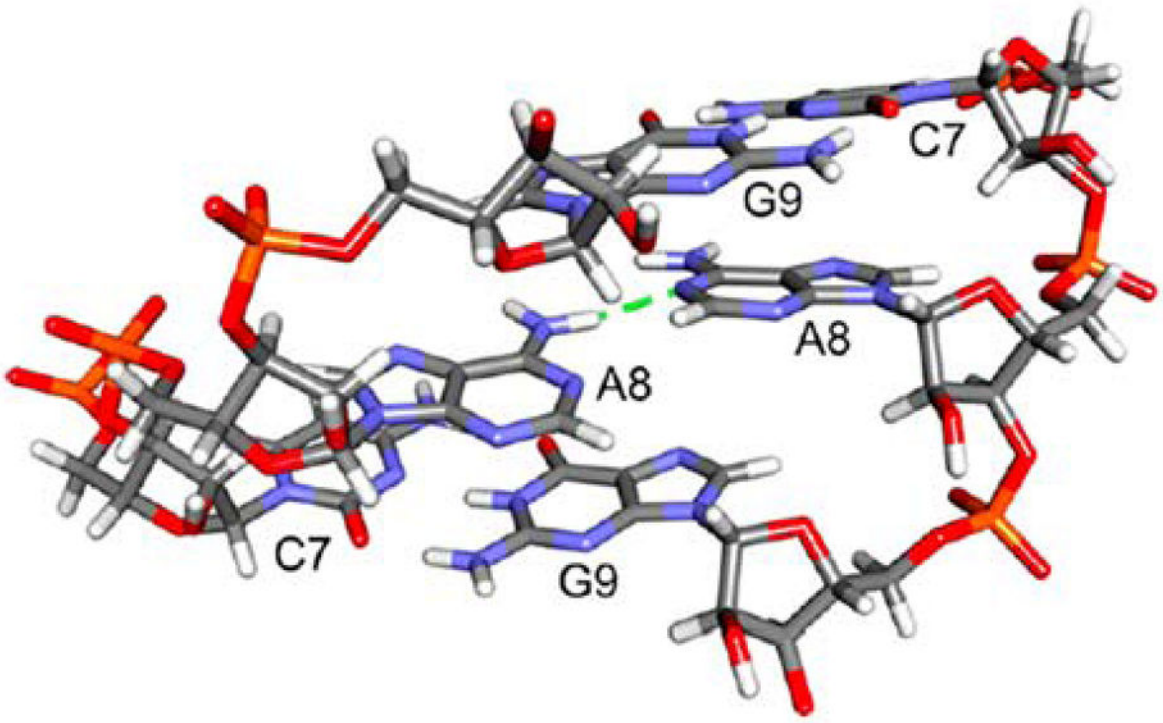


Figure 4.
5′CAG/3′GAC motif from r(3×CAG).

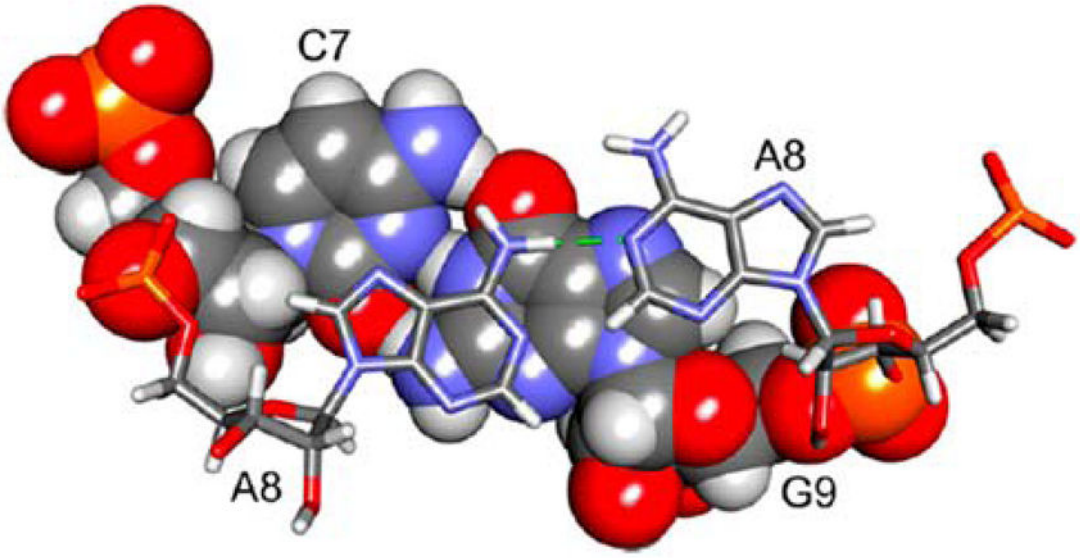


Figure 5.
Stacking diagrams for an AA pair of r(3×CAG) on a closing GC pair.

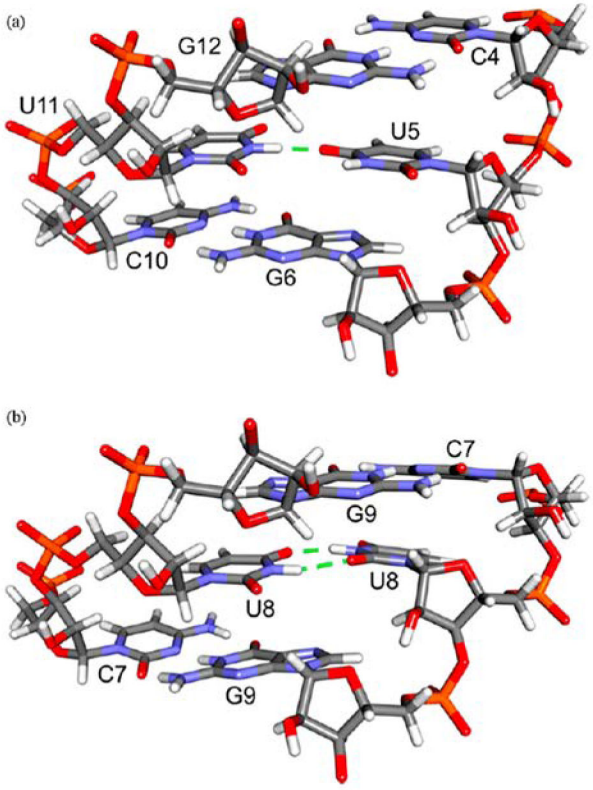


Figure 6.
5′CUG/3′GUC motifs from r(3×CUG). (a) One-hydrogen bond UU loop. (b) Two-hydrogen bond UU loop.

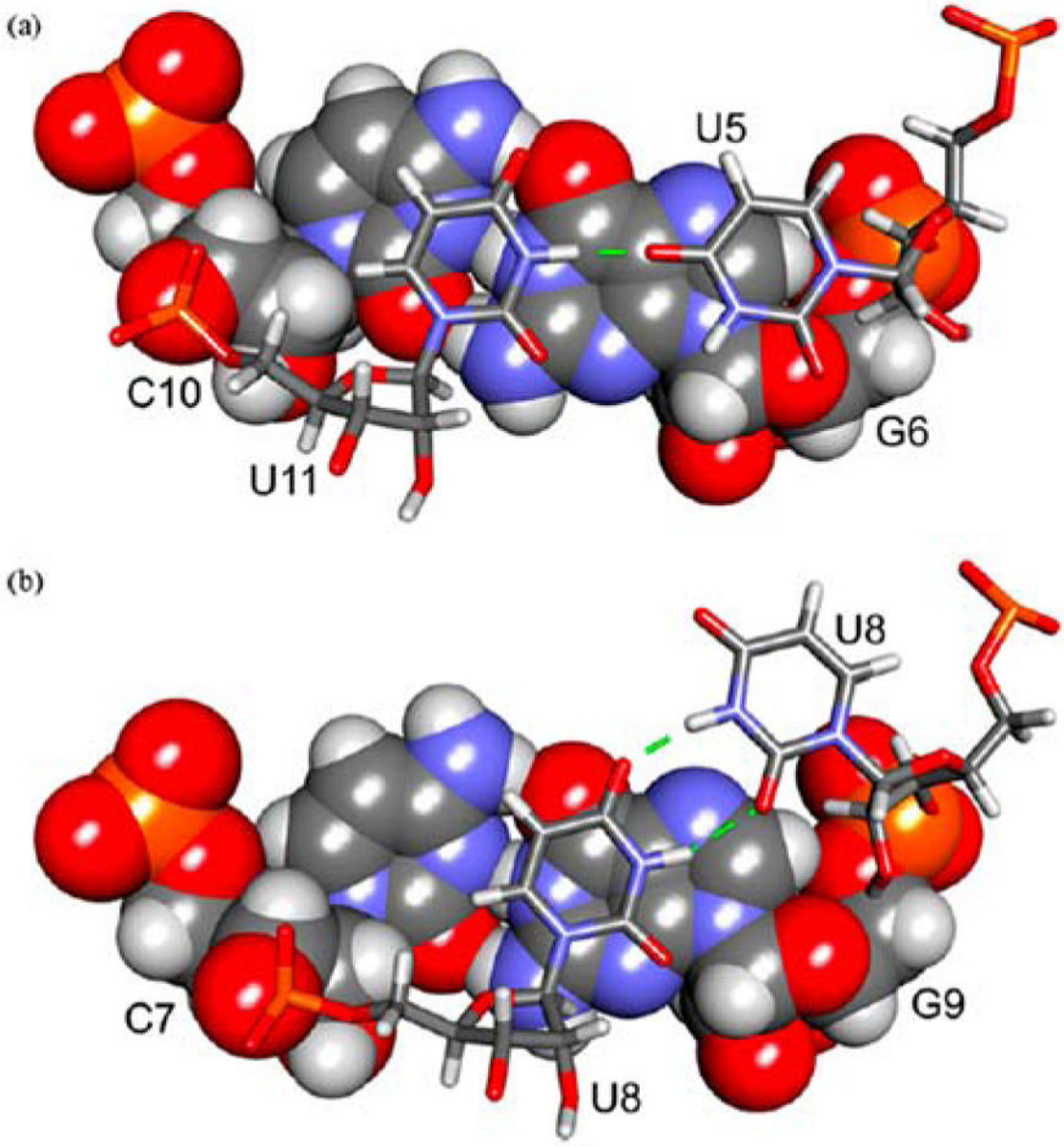


Figure 7.
Stacking diagrams for UU pairs of r(3×CUG) on dosing GC pairs. (a) One-hydrogen bond UU loop. (b) Two-hydrogen bond UU loop.

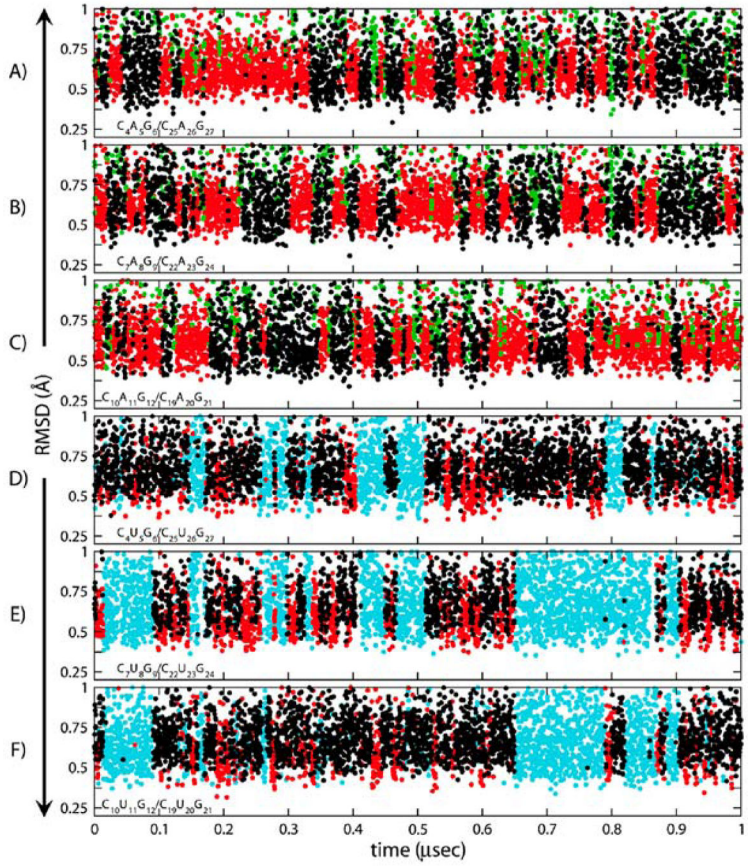


Figure 8.

Structures observed in the 1 μs MD trajectories of r(3×CAG) and r(3×CUG). Time vs RMSD plots of the (A) first, (B) second, and (C) third loop regions of r(3×CAG) and the (D) first, (E) second, and (F) third loop regions of r(3×CUG) were prepared to display the dynamics (see the inset in each plot for details). Color notation was used to highlight the one-hydrogen bond (black), zero-hydrogen bond (red), and two-hydrogen bond (cyan) states. Green represents other conformations observed in the r(3×CAG) trajectory.

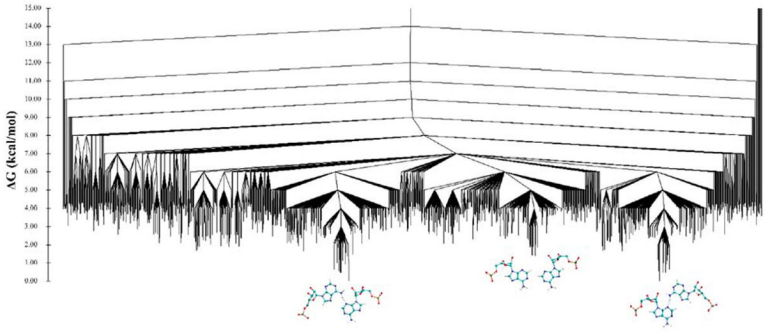


Figure 9.

Disconnectivity graph showing the free energy landscape of RNA 1×1 AA internal loops in r(CAG)^{exp}. Structures below the minima display the conformation of a 1 × 1 AA internal loop. Note that the system is symmetric and that the global minimum represents the one-hydrogen bond state of a 1 × 1 AA pair, while the zero-hydrogen bond state is a local minimum.

Table 1

Structural Refinement Statistics of r(3×CAG) and r(3×CUG) for the Average of 20 Structures of Each RNA Construct

	r(3×CAG) (PDB ID 5VH7)	r(3×CUG) (PDB ID 5VH8)
no. of restraints		
all distance restraints, including hydrogen bonds	254	218
all NOE restraints	220	184
intraresidue	131	72
sequential residues	78	96
long-range	10	16
hydrogen bond	34	34
dihedral restraints	140	140
RMSD of experimental restraints		
distances (Å)	1.0×10^{-3}	9.1×10^{-4}
dihedral angles (deg)	1.5	0.2
RMSD of structures for heavy atoms (Å)		
all residues	1.37 ± 0.30	1.55 ± 0.36
helices (excluding AA or UU loops)	1.32 ± 0.32	1.48 ± 0.39
outer helices (residues 1–3 and 13–15)	1.18 ± 0.36	1.31 ± 0.40
AA or UU loops	1.29 ± 0.26	1.45 ± 0.20
triplet repeats (residues 4–12)	1.13 ± 0.22	1.18 ± 0.18