Published in final edited form as:

J Am Chem Soc. 2013 March 6; 135(9): 3528–3538. doi:10.1021/ja3108627.

A dynamic structural model of expanded RNA CAG repeats: A refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations

Ilyas Yildirim^{a,1}, Hajeung Park^b, Matthew D. Disney^{c,1}, and George C. Schatz^{a,1}

^aDepartment of Chemistry and International Institute for Nanotechnology, Northwestern University, Evanston, IL 60208

^bTranslational Research Institute, The Scripps Research Institute, Jupiter, FL 33458

^cDepartment of Chemistry, The Scripps Research Institute, Jupiter, FL 33458

Abstract

One class of functionally important RNA is repeating transcripts that cause disease through various mechanisms. For example, expanded r(CAG) repeats can cause Huntington's and other disease through translation of toxic proteins. Herein, crystal structure of r[5'UUGGGC(CAG)₃GUCC]₂, a model of CAG expanded transcripts, refined to 1.65 Å resolution is disclosed that show both *anti-anti* and *syn-anti* orientations for 1×1 nucleotide AA internal loops. Molecular dynamics (MD) simulations using Amber force field in explicit solvent were run for over 500 ns on model systems r(5'GCGCAGCGC)₂ (MS1) and r(5'CCGCAGCGG)₂ (MS2). In these MD simulations, both *anti-anti* and *syn-anti* AA base pairs appear to be stable. While *anti-anti* AA base pairs were dynamic and sampled multiple *anti-anti* conformations, no *syn-anti* ↔ *anti-anti* transformations were observed. Umbrella sampling simulations were run on MS2, and a 2D free energy surface was created to extract transformation pathways. In addition, over 800 ns explicit solvent MD simulation was run on r[5'GGGC(CAG)₃GUCC]₂, which closely represents the refined crystal structure. One of the terminal AA base pairs (*syn-anti* conformation), transformed to *anti-anti* conformation. The pathway followed in this transformation was the one predicted by umbrella sampling simulations. Further analysis showed a binding pocket near AA base pairs in *syn-anti* conformations. Computational results combined with the refined crystal structure show that global minimum conformation of 1×1 nucleotide AA internal loops in r(CAG) repeats is *anti-anti* but can adopt *syn-anti* depending on the environment. These results are important to understand RNA dynamic-function relationships and develop small molecules that target RNA dynamic ensembles.

¹To whom correspondence should be addressed. Ilyas Yildirim (i-yildirim@northwestern.edu) and George C. Schatz (schatz@chem.northwestern.edu) for theoretical inquiries; Matthew D. Disney (disney@scripps.edu) for experimental concerns. .

Supporting Information. Descriptions for each movie; minimization protocol and sample torsional restraint file used in umbrella sampling simulations; sugar and backbone torsional angles, distance and angle calculations for different base pairs, helical parameters calculated for base pairs and base pair steps, minor and major groove width calculations for r(3xCAG)^{X-RAY}; electron densities and stick models of 1×1 AA loop closing CG base pairs; 1D free energy profiles along each pathway predicted by umbrella sampling calculations; Na⁺ binding and rmsd analysis of AA base pairs for model systems MS1 and MS2; (χ, θ) states occupied by A12 in A12-A23 base pair in 3xCAG MD; stacking of AA internal loops on the loop closing CG base pairs in r(3xCAG)^{X-RAY}; secondary structure, refined structure, and crystal packing of the CAG repeat RNA construct, r(3xCAG)^{X-RAY}. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Keywords

Triplet disease; Huntington's disease; X-RAY structure of CAG repeat RNA construct; RNA loop dynamics; molecular dynamics simulations; umbrella sampling

INTRODUCTION

Many diseases are caused by transcripts that contain expanded repeats. The physiological consequences of these expansions, and thus the mechanism of disease, are varied. For example, r(CUG) expansions (r(CUG)^{exp}) in the 3' untranslated region (UTR) of the dystrophin myotonic protein kinase (*DMPK*) mRNA cause myotonic dystrophy Type 1 while myotonic dystrophy type 2 is caused by an expansion of r(CCUG) (r(CCUG)^{exp}) in intron 1 of the zinc-finger 9 (*ZNF9*) protein mRNA. These diseases are characterized by muscle weakness and slow relaxation of muscles after contraction.^{1,2} Fragile X-associated tremor ataxia syndrome (FXTAS) is caused by 50-200 CGG repeats (r(CGG)^{exp}) in the 5' UTR of the fragile X mental retardation 1 (*FMR1*) gene.³⁻⁵ In each of these cases, the RNAs are in a non-coding region and are not translated into protein. Thus, disease is caused by an RNA gain-of-function.^{1,6,7} In this mechanism, expanded repeats sequester proteins, typically ones involved in pre-mRNA splicing regulation, affecting their function. Muscleblind-like 1 protein (MBNL1) is sequestered and inactivated by r(CUG)^{exp} and r(CCUG)^{exp} while Sam68 is inactivated by r(CGG)^{exp}.

In contrast to the above cases, expansions of CAG repeats (r(CAG)^{exp}) that cause disease are typically found in the coding regions of mRNAs. For example, r(CAG)^{exp} have been found in Huntingtin (*HTT*), androgen receptor (*AR*), spinocerebellar ataxia (*SCA*), and atrophin-1 (*ATN1*) genes and cause Huntington disease (HD), Spinal and Bulbar Muscular Atrophy (SBMA), Spinocerebellar Ataxia Type 1 (SCA1), and Dentatorubral-pallidoluysian atrophy (DRPLA), respectively. Since r(CAG)^{exp} is in the coding region of these transcripts, it is translated into toxic polyglutamine (polyQ) proteins that cause disease.^{1,8-12}

More recently it has also been shown that r(CAG)^{exp} in coding regions can also sequester proteins such as MBNL1 and contribute to neurodegeneration.¹³⁻¹⁵ Moreover, expanded repeats in UTRs can be translated via a Repeat Associated Non-ATG Translation.¹⁶ Thus, although the disease mechanisms described above for each type of repeat are the most well established, it is not as clear-cut as it once appeared and the interplay of RNA and protein toxicities may ultimately cause disease.

In order to gain insights into the role of RNA structure in these diseases, various approaches have been used to gain structural information. For example, chemical and enzymatic probing has shown that triplet repeat expansions form expanded hairpin structures with 1×1 nucleotide internal loops closed by GC pairs.¹⁷ For several of these repeats, high-resolution structural information has been obtained through refinement of NMR spectral or X-ray crystallographic data.¹⁸⁻²³

The structure of a model of r(CAG)^{exp} has been reported by Kiliszek et al,²³ who studied the structure of the RNA duplex (GGCAGCAGCC)₂. They reported that the adenosines are in the anti conformations with a single C2-H2...N2 hydrogen bond, with significant distortions compared to a standard Watson-Crick paired duplex. This work did not identify fluxional behavior characteristic of large amplitude motions that might be expected for a 1×1 nucleotide internal loop. Information about such motions could improve our understanding of the conformations that toxic repeating transcripts adopt and how these conformations

affect the binding of protein or small molecules.^{24,25} Thus, the conformational flexibility of r(CAG) repeat-containing duplexes is of significant interest.²⁶

Computational technology is an emerging scientific tool that can make useful predictions of the atomic details of structure as well as of RNA dynamics. Such studies are described herein and utilized to study the conformational heterogeneity of r(CAG) repeats. Methods such as residue-centered force fields (coarse-grained),²⁷ atom-centered force fields (AMBER,²⁸ CHARMM,^{29,30} and GROMOS³¹), approximate quantum mechanics,^{32,33} and mixed quantum-mechanics/molecular mechanics (QM/MM),³⁴⁻⁴¹ are typically used for predicting the structure and dynamics of RNA. Molecular mechanics force fields are computationally economical such that molecular dynamics (MD) simulations on the microsecond or faster time scales are possible.⁴²⁻⁴⁶ Application of these methods allows for comparing the structural and thermodynamic properties of nucleic acids with experiments.⁴⁷⁻⁵² Free energy calculations are particularly important because free energies control a reaction coordinate. Combined with the weighted histogram analysis method (WHAM),^{53,54} umbrella sampling⁵⁵⁻⁵⁸ is a powerful computational approach to extract the free energy landscape along multiple reaction coordinates such as base opening⁵⁹ and base flipping⁶⁰⁻⁶² that can occur in DNA and RNA.

In this paper, we present the crystal structure of a self-complementary RNA duplex, r[5'UUGGGC(CAG)₃GUCC]₂, (a model of r(CAG)^{exp}) refined to 1.65 (r(3xCAG)^{X-RAY}) Å resolution. The crystal structure shows multiple conformations for 1×1 nucleotide AA internal loops (*anti-anti* and *syn-anti*) that suggest fluxional behavior. To provide insight into these results and probe more deeply into the structure of the AA internal loops, MD simulations using the Amber force field in explicit solvent were run on the model systems r(5'GCGCAGCGC)₂ (MS1), r(5'CCGCAGCGG)₂ (MS2), and r[5'GGGC(CAG)₃GUCC]₂ (3xCAG) including a detailed analysis of the dynamics of 1×1 nucleotide AA internal loops. After choosing two reaction coordinates, which mimic base flipping and base orientation with respect to sugar moieties for the 1×1 nucleotide AA internal loops, a 2D free energy landscape for *syn-anti*↔*anti-anti* transformation was generated by combining WHAM with umbrella sampling MD simulations. It was found that the AA base pairs are dynamic and can form stable structures both in *anti-anti* and *syn-anti* conformations with one hydrogen bond. The results indicate that the *anti-anti* and *syn-anti* 1×1 nucleotide AA internal loops are major and minor conformations, respectively. Multiple pathways are located for the *syn-anti*↔*anti-anti* transformation. Moreover, an electronegative binding pocket was determined around *syn-anti* AA base pairs where individual Na⁺ ions can bind for as long as 50 ns. These results are important to better understand how r(CAG)^{exp} contributes to disease, as well as to assist in the design of small molecules that bind r(CAG)^{exp} and ameliorate disease.

METHODS

Crystallization and structure refinements

The RNA duplex, r[5UUGGGC(CAG)₃GUCC]₂ (r(3xCAG)^{X-RAY}), at 1.2 mM concentration was dissolved in DEPC-water and folded by heating to 60 °C for 5 min and cooling to room temperature. Crystal of r(3xCAG)^{X-RAY} was grown by the sitting drop vapor diffusion method using 0.2 μL of 1.2 mM RNA and equal volume of precipitants. The precipitants for r(3xCAG)^{X-RAY} were 15 mM Mg(CH₃COO)₂, 50 mM Na cacodylate, pH 6.0, 1.7 M (NH₄)₂SO₄. A diffraction data set with Bragg spacings to 1.65 Å for r(3xCAG)^{X-RAY} was collected on a MAR325 CCD detector at beam line 12-2 of the Stanford Synchrotron Radiation Laboratory. The datasets were integrated and scaled using HKL2000.⁶³ Phases for RNA structure was obtained by molecular replacement using Phaser⁶⁴ in Phenix program interface⁶⁵ with PDB entry 3SYW as a search model.

Crystallographic refinement was performed using Phenix⁶⁵ and multiple rounds of manual model fittings were performed using Coot. Data collection and the crystallographic refinement statistics are summarized in Table 1.

Preparation of model systems for MD simulations

Model systems MS1, MS2, and 3xCAG (Figure 1) were prepared for MD simulations. MS1 and MS2 were modeled using the nucgen module of AMBER9⁶⁶ with AA base pairs in *anti-anti* and *syn-anti* conformations. 3xCAG (Figure 1) was modeled from crystal structure r(3xCAG)^{X-RAY} by trimming down the flanking uridines. In the model 3xCAG structure, one of the terminal AA base pair was modified to be in *anti-anti* conformation. This was done because the crystal structure of r(3xCAG)^{X-RAY} is symmetric and has both the terminal AA base pairs in *syn-anti* conformations that puts redundancy to the analysis of potential *syn-anti*↔*anti-anti* transformations. The systems were neutralized with Na⁺ ions⁶⁷ and solvated with TIP3P water molecules⁶⁸ in a truncated octahedral box. The MS1/MS2 and 3xCAG systems had 4025 and 8995 water molecules, respectively. The Amber force field²⁸ with revised χ ⁶⁹ and α/γ ⁷⁰ torsional parameters were used in molecular dynamics and umbrella sampling simulations.

The main reason why the flanking uridine bases were removed from the structure was to mimic the *in vivo* r(CAG)^{exp}. In the crystallization process, flanking uridines were included in order to crystallize the structure, which resulted in terminal AA base pairs in *syn-anti* conformations. The observation of this conformation, which was not seen before in RNA CAG repeats, motivated us to delve into the mechanism behind it. By trimming down the uridine bases, we disrupted the interaction between the flanking uridines and the terminal AA base pairs, giving us the opportunity to investigate the dynamics of AA base pairs when starting with *syn-anti* conformations in 3xCAG.

Molecular dynamics simulations

The systems were first minimized and equilibrated in two steps as described previously.^{71,72} In the production runs, constant pressure dynamics with isotropic positional scaling was used. The reference pressure was 1 atm with a pressure relaxation time of 2 ps. SHAKE⁷³ was turned on for bonds involving hydrogen atoms. An atom-based long-range hard cutoff of 8.0 Å was used in all simulations. A total of 854 ns and ~500 ns of MD were run at 300 K with a 1 fs time step for 3xCAG and MS1/MS2, respectively. Simulations were carried out with the PMEMD module in AMBER9.⁶⁶ Trajectory files were written at each 5000 fs time step. For all these calculations, over 210K CPU hours were used.

Umbrella sampling simulations

In order to predict the potential pathways for the *syn-anti*↔*anti-anti* transformation, a 2D potential of mean force (PMF) surface was calculated for the model system MS2. Rotation around the glycosidic bond is responsible for the syn and anti base orientations of adenosine with respect to the adjacent sugar. As a result, the χ torsional angle was chosen as one of the reaction coordinates to mimic the base orientation (Figure 2a). A second reaction coordinate, called center-of-mass pseudo-dihedral angle (θ), was chosen to mimic unstacking of adenosine as shown in Figure 2b. Similar reaction coordinates for base flipping studies of oligonucleotides were used previously.^{60,61,74} Each reaction coordinate was rotated with an increment of 10°, yielding 36×36=1296 conformations. Initial structures with different (χ , θ) combinations were created for the model system MS2.

Minimization—Each structure was minimized in two steps: 1) All residues except the adenosines were held fixed with a restraint force of 10 kcal/mol-Å². χ and θ dihedrals were held fixed in a particular combination using a square bottom well with parabolic sides that

had force constants of 50000 kcal/mol-rad². Steepest descent minimization of 2000 steps was followed by a conjugate gradient minimization of 2000 steps. 2) Similar to the first step, the χ and θ dihedrals were held fixed in their initial combination. Watson-Crick base pairing restraints were imposed on the GC base pairs neighboring the 1×1 nucleotide AA internal loop. Steepest descent minimization of 25000 steps was followed by a conjugate gradient minimization of 25000 steps (Tables S1 and S2).

Equilibration—After minimization, two steps of pressure equilibration were completed: 1) the RNA molecule was held fixed with a restraint force of 10 kcal/mol-Å². Constant volume dynamics with an atom-based long range hard cutoff of 8.0 Å was used. SHAKE⁷³ was turned on for bonds involving hydrogen atoms. The temperature was raised from 0 to 300 K in 20 ps. Langevin dynamics with a collision frequency of 1 ps⁻¹ was used. A total of 20 ps of MD were run with a 2 fs time step. 2) The above conditions were chosen, except the following: χ and θ dihedrals were held fixed in their initial combination with a force constant of 500 kcal/mol-rad². Watson-Crick base pairing restraints were imposed on GC base pairs neighboring 1×1 nucleotide AA internal loops. Constant pressure dynamics with isotropic positional scaling was turned on. The reference pressure was 1 atm with a pressure relaxation time of 2 ps. A total of 100 ps of MD were run with a 2 fs time step. Particle Mesh Ewald (PME)^{75,76} was used in all calculations.

Production runs—Each production run was similar to the second step of the pressure equilibration described above. No restraints except the ones imposed on χ and θ dihedrals were used in the simulations. A square bottom well with parabolic sides was used to restrain χ and θ dihedrals with force constants of 50 kcal/mol-rad². For each simulation, a total of 2 ns of MD were run with a 2 fs time step. (χ , θ) data were written at intervals of 10 fs. Simulations were carried out with the sander.MPI module in AMBER11. For umbrella sampling calculations, over 267K CPU hours were used.

Analysis—Dihedral and root-mean-square deviation (rmsd) analyses were completed using the ptraj module in the AMBER package.⁶⁶ Weighted Histogram Analysis Method^{77,78} (WHAM) was used to calculate the PMF surface using WHAM ver.2.0.4 written by Alan Grossfield (<http://membrane.urmc.rochester.edu/content/wham>). In the WHAM analysis, the last 1.5 ns data of each umbrella sampling simulation was used, and periodicity was turned on in both dimensions.

RESULTS AND DISCUSSION

Structure of r(3xCAG)^{X-RAY} (a model of r(CAG)^{exp})

Figure 3A shows the secondary structure of an RNA construct that has three copies of the 5'CAG/GAC motif. Crystal structure was refined to 1.65 Å resolution (r(3xCAG)^{X-RAY}). The central 5'CAG/GAC region of r(3xCAG)^{X-RAY} is in the *anti-anti* 1×1 nucleotide AA internal loop conformation while the terminal AA base pairs are in the *syn-anti* conformations (Figures 3B and 4). The duplex crystallized in double-stranded helical structure and made pseudoinfinite helices with the neighboring RNAs in the crystal lattice (Figures 3C). The duplex has A-form conformation and the overlay of r(3xCAG)^{X-RAY} and A-form RNA shown in Figure 3D displays almost identical backbone structures except the 5'UU dangling ends, which interact with the terminal 1×1 nucleotide AA internal loops (Figure 3E). Amino and 2'OH groups of A₁₂ and A₂₉ are in close contact with U₋₁₇ and U₋₂, respectively, which cause A₁₂ and A₂₉ to adopt *syn* conformations (Figures 3E and 4). Almost all riboses are C3'-endo conformation (sugar pucker) (Table S3). Dihedral angle values for α , β , γ , δ , ϵ , and ζ can be found in Table S3. Parallel to Figure 3D, global helical

parameters calculated for $r(3xCAG)^{X-RAY}$ show characteristic of A-form conformation with a minor groove width of 15.5 Å (Tables S4-S8).

Electron density maps of individual 1×1 AA internal loops for $r(3xCAG)^{X-RAY}$ are in line with *syn-anti* and *anti-anti* conformations for terminal and central 1×1 nucleotide AA internal loops, respectively (Figure 4). Temperature factor comparisons of the central *anti-anti* AA loop have higher average values in $r(3xCAG)^{X-RAY}$ than base paired regions (Figure 4 and S1). This is evidenced by the well-defined electron density for base paired regions while the central 1×1 nucleotide AA internal loops have less well-defined electron density, indicating dynamic nature (Figures 4 and S1). Electron densities for the CG closing base pairs are consistent with three hydrogen bonds (Figures S1 and S2). Interestingly, a similar dynamic character has been observed in a refined X-ray structure of the 1×1 nucleotide UU internal loops in $r(CUG)^{exp}$ models, suggesting that many repeats can sample multiple conformations.¹⁹ In contrast, a refined X-ray structure of $r(CGG)$ repeats shows that this RNA adopts a rather rigid structure, and each 1×1 nucleotide GG internal loop has a *syn-anti* conformation.¹⁸

Pathways for *syn-anti*↔*anti-anti* AA transformation

Umbrella sampling MD simulations were used to construct the 2D potential mean force (PMF) surface in order to visualize the potential pathways for the *syn-anti*↔*anti-anti* AA transformation (Figure 5). The x- and y-axes in Figure 5 are the reaction coordinates for unstacking (θ) and base orientation with respect to the sugar moiety (χ) of A5 in the 1×1 nucleotide AA internal loop (Figure 2). Three pathways are found for *syn-anti*↔*anti-anti* AA transformation (Figure 5). In path P1, A5 first unstacks from the helical axis through the minor groove direction (motion in x-axis). This motion relaxes the state and gives A5 enough space to make the *syn*↔*anti* transformation through a rotation around the glycosidic bond, χ (motion in y-axis). A5 then stacks back into the helical axis through the major groove direction to form the AA base pair. In P2, almost a direct *syn*↔*anti* transformation occurs in the system without A5 unstacking from the helical axis. In P3, a transformation similar to P1 is seen except that unstacking occurs through the major groove direction. P1 has two energy barriers in the *syn-anti* transformation pathway with a stable local minimum at around (−75, 60) (Figure 5). In addition, P1 has a broader space in its transformation pathway compared to P2 and P3, which makes the transformation more probable. To explicitly show the differences between the pathways, 1D free energy profiles along the most likely direction for each path were extracted (Figures S3 and S4). In Figure 5, the sampling range of the *syn-anti* region is very tight compared to the range of the *anti-anti* region. This means that the *anti-anti* region is dynamic and sample more than one conformation while *syn-anti* is much more rigid in structure compared to *anti-anti*.

In general, RNA loops are flexible, which gives them functional roles in the cell such as when AA base pairs form in RNA CAG repeats. This property, however, makes it hard to determine how small molecules bind to these flexible RNA regions, which was recently discussed by Al-Hashimi group at the University of Michigan at Ann Arbor.⁷⁹ While Figure 5 discloses different potential pathways for the *syn*↔*anti* transformation, there are some results that should not be overlooked. RNA residues have three significant motions that affect the global RNA structure: 1) Base opening/flipping, 2) Base orientation with respect to sugar, and 3) Sugar puckering. The first two motions will affect the RNA structure the most. In Figure 5, the x-axis represents base opening/flipping while the y-axis represents χ rotation. Even though the lowest energy regions are around (50,50) (adenosine stacked on the helical axis in a *syn* conformation) and (60,180) (adenosine stacked on the helical axis in an *anti* conformation) Figure 5 shows several other local minima regions in unstacked states such as around (−125,50), (−125,190), and (−75, 300). When AA base pairs are not interacting with anything, the structure prefers the *anti-anti* form while upon interacting with

something such as the flanking uridine bases they can adopt the *syn-anti* form. This interaction-driven conformational preference process can be applied to other local minima regions, too, as shown in Figure 5. It is possible that the local minima regions could be dominated by AA base pairs if a small molecule was designed to capitalize on these regions in binding to RNA.

Stability of *syn-anti* and *anti-anti* AA base pairs

Model systems MS1 and MS2 (Figures 1a and 1b) were designed in both *syn-anti* and *anti-anti* 1×1 nucleotide AA internal loop conformations in order to investigate the dynamics. For each case, at least 500 ns MD simulations were run. Each system has a total of 16 Na⁺ ions that neutralize the systems. Na⁺ binding and rmsd analysis show unique properties for each conformation (see Figure S5).

Rms deviations (rmsd) of 1×1 nucleotide AA internal loops (with respect to initial conformations) and distances of each Na⁺ ion from a loop as a function of time for each MD simulation were plotted (Figure S5). In the MD simulations, different 1×1 nucleotide AA internal loop types were observed (Figures 6 and S5). Different colors were used in Figure 6 to distinguish each 1×1 nucleotide AA internal loop type, and this color notation was conjoined with the rmsd analysis in Figure S5 to emphasize the structural transformations observed in the simulations. Additionally, different colors were used for each Na⁺ ion in the Na⁺ binding analysis to highlight the time spent by each ion around the 1×1 nucleotide AA internal loops (Figure S5). It was observed that both *syn-anti* and *anti-anti* 1×1 nucleotide AA internal loop conformations were stable in these model systems. While the *syn-anti* 1×1 nucleotide AA internal loop remained almost all the time in the initial configurations, multiple *anti-anti*→*anti-anti* 1×1 nucleotide AA internal loop transformations were observed (Figures 6 and S5).

In the MD simulations that were started with *syn-anti* 1×1 nucleotide AA internal loop, there was a Na⁺ binding pocket around the 1×1 nucleotide AA internal loop, which held different Na⁺ ions throughout the simulations (Figures 7 and S5, panels a1 and b1). Almost no structural change was seen in *syn-anti* AA base pair in these simulations. When 1×1 nucleotide AA internal loops were in *syn-anti* conformations, Na⁺ ions were bound in this pocket for as long as over 50 ns (Figure S5). In this state, Na⁺ was interacting with the electronegative groups of anti A and neighboring G (Figure 7). On average, distances between Na⁺ and A-N7, G-N7, and G-O6 were 2.4, 2.6, and 3.3 Å, respectively (Figure 7). This Na⁺-bound state is one reason why 1×1 nucleotide AA internal loop in the *syn-anti* conformation did not fluctuate too much from its initial state (Figure S5). Indeed, when there were fluctuations in *syn-anti* 1×1 nucleotide AA loops in the MD simulations, there were no Na⁺ ions bound to this binding pocket (Figure S5, panels a1 & a2 and b1 & b2). Around 125 and 440 ns in the MD simulation of model system MS1 (Figure S5a) and around 325 ns in the MD simulation of model system MS2 (Figure S5b), no Na⁺ ions were bound in the pocket. At these times, *syn-anti* AA transformed to a transition state for a short period of time that corresponds to the orange colored AA state shown in Figure 6d and highlighted with orange in the rmsd analysis (Figure S5, panels a1 & a2 and b1 & b2). Thus, the dynamics of *syn-anti* 1×1 nucleotide AA internal loops were controlled by the Na⁺ binding in this pocket.

In the MD simulations that were started with *anti-anti* 1×1 nucleotide AA internal loops, no Na⁺ binding pocket was found (Figure S5, c and d). Na⁺ ions were spending at most 10 ns around AA base pairs in these systems (Figure S5). On average, a Na⁺ ion was present within 5.5 Å of AA for 33% and 91% of the time in the *anti-anti* and *syn-anti* AA systems, respectively. The *anti-anti* 1×1 nucleotide AA internal loops, however, were dynamic. Multiple *anti-anti*↔*anti-anti* AA conformational transformations displayed in Figure 6 were

observed in the simulations (Figure S5, c and d). This result is in line with the umbrella sampling predictions that span a wide range for the *anti-anti* 1×1 nucleotide AA internal loops in the 2D PMF surface (Figure 5).

One explanation for not observing the *syn-anti*↔*anti-anti* AA transformation in these systems might be due to the energy barrier, which could be too high to overcome. In order to test this hypothesis, ten independent temperature jump MD simulations were completed on model system MS2, which were started both in *syn-anti* and *anti-anti* 1×1 nucleotide AA internal loop conformations. The temperature in each MD simulation was gradually increased from 300 to 400 K within 20 ns without any restraints, and final structures were analyzed. Five out of ten MD simulations that started with *syn-anti* 1×1 nucleotide AA internal loops ended up in *anti-anti* 1×1 nucleotide AA internal loop conformation while none of the simulations that started with *anti-anti* 1×1 nucleotide AA internal loops transformed to *syn-anti* AA. In model systems MS1 and MS2, there are four Watson-Crick GC base pairs on both sides of the 1×1 nucleotide AA internal loops while in the original triplet r(3xCAG)^{X-RAY} system there are only two GC base pairs between each consecutive 1×1 nucleotide AA internal loop. The results, therefore, suggest that model systems MS1 and MS2 have rigid stem regions that increase the energy barrier of the *syn-anti*↔*anti-anti* AA transformation.

Dynamics of model 3xCAG system

Model 3xCAG system (Figure 1c) resembles the original triplet RNA, r(3xCAG)^{X-RAY}, (Figure 3a) much more than the model systems MS1 and MS2 (Figures 1a and 1b). As a result, the dynamics of 3xCAG was investigated by doing an 854 ns long MD simulation. There were a total of 32 Na⁺ ions that neutralized the system.

A different rmsd and Na⁺ binding analysis was completed on 3xCAG (Figure 8). As before, each individual Na⁺ ion is highlighted with different colors (Figure 8, c/f/i). Similar to the analysis completed for MS1 and MS2, the rmsd's of each 1×1 nucleotide AA internal loop with respect to initial conformation were plotted (Figure 8, b/e/h). Colors representing unique 1×1 nucleotide AA internal loop types in Figure 6 were used in the rmsd analysis to emphasize the type of transformations seen in the MD simulation (Figure 8, b/e/h). Moreover, the rms deviations of backbone, and GC/CG stems located between the consecutive 1×1 nucleotide AA internal loops were plotted in order to investigate the rigidity of different regions of 3xCAG with respect to time (Figure 8, a/d/g).

Similar to the results of MS1 and MS2, it was observed that there was a Na⁺ binding pocket around the *syn-anti* type 1×1 nucleotide AA internal loop (see Figure 8i, first 125 ns). The initial conformation of A12-A23 was *syn-anti*, and this conformation was stable for the first 125 ns (Figure 8h). In this *syn-anti* state, different Na⁺ ions were binding to the pocket for as long as 30 ns (Figure 8, h and i). At around 175 ns, however, a full transformation from *synA12-antiA23* to *antiA12-antiA23* was observed (Figure 8h). This structural transformation was initiated when there were no Na⁺ ions bound to the pocket (Figure 8i). Without a Na⁺ ion in the pocket, *synA12* first unstacked through the minor groove side to transform itself into a transition state with two hydrogen bonds (Figure 6d and highlighted with orange color in Figure 8h). This particular transformation from the *synA12-antiA23* to the 'transition state' *synA12-antiA23* was seen a couple of times in the first 175 ns (highlighted with orange in Figure 8h). At around 175 ns, *synA12* transformed fully to *antiA12* to form the *antiA12-antiA23* base pairing (Figure 8h). Once the A12-A23 base pair was in the *anti-anti* form, multiple *anti-anti*→*anti-anti* transformations were observed until the end of the MD simulation (Figures 6 and 8h).

In order to explicitly define the *synA-antiA* ↔ *antiA-antiA* transformation seen in the MD simulation, the (χ , θ) states occupied by A12 between 125 and 175 ns were plot (Figure S6). The pathway P1 predicted by umbrella sampling calculations (Figure 5) was followed by A12 in the *synA12-antiA23* ↔ *antiA12-antiA23* transformation. Note that the prediction of pathway P1 for MS2 by umbrella sampling calculations is almost identical to the 3xCAG MD simulation result (Figures 5 and S6).

The other two 1×1 nucleotide AA internal loops, A6-A29 and A9-A26, which were started in *anti-anti* conformations, never transformed to a *syn-anti* state. Similar to the results of MS1 and MS2, no Na⁺ binding pocket was found when the 1×1 nucleotide AA internal loop was in the *anti-anti* conformation (Figure 8, c and f). As before, multiple *anti-anti* ↔ *anti-anti* AA transformations were observed for these internal loops (Figure 8, b and e).

Model 3xCAG is not as rigid as model systems MS1 and MS2. The two Watson-Crick GC/GC base pairs between each 1×1 nucleotide AA internal loop in 3xCAG were responsive to the environment. Rms deviations of G10C11/G24C25 at around 185 ns (right after the *synA* → *antiA* transformation) and 600 ns, and of G7C8/G27C28 at around 750 ns were over 3 Å (Figure 8, d and g). These fluctuations have a direct effect on the RNA backbone, giving an rmsd over 10 Å during these times (Figure 8a). Such flexible neighbors around 1×1 nucleotide AA internal loops were not observed in the model systems MS1 and MS2 that may contribute to lowering the free energy barrier for the *synA* ↔ *antiA* transformation.

Except for the transition state, all stable 1×1 nucleotide AA internal loops in the MD studies of 3xCAG, MS1, and MS2 have one hydrogen bond (Figure 6). These non-canonical pairs are weak compared to Watson-Crick base pairs, which have two or three hydrogen bonds.⁸⁰⁻⁸² As a result, any slight perturbation of the 1×1 nucleotide AA internal loop will have an effect on its conformation. Adenosine has three electronegative regions (N1, N3, N7) (Figure 2a) that can make non-covalent electrostatic interaction with cations such as Na⁺ and molecules. In a 1×1 nucleotide AA internal loop with one hydrogen bond, there are a total of five free electronegative groups on the adenosine bases that can interact with Na⁺ ions. Except for the Na⁺ ions bound to the pocket in *syn-anti* AA, the ions spent only short amounts of time near these electronegative regions (Figures 8c/f/i, and S5). It was observed that, depending on the momentum of the ions, they could assist 1×1 nucleotide AA internal loops in the structural transformations represented in Figure 6 (Movies S1-S3).

RNA CAG loops in literature

The 1×1 nucleotide AA internal loop conformations predicted by MD simulations (Figure 6) are also observed in RNA crystal structures. The Cate group at the University of California at Berkeley refined crystal structures of *Escherichia coli* (*E. coli*) ribosome bound to different complexes.⁸³⁻⁹⁰ In these structures, there are 5'CAG/3'GAC motifs with *syn-anti* 1×1 nucleotide AA internal loop conformations. Yet, the ribosomal structure is much more complex compared to the 3xCAG duplex structure that is presented herein. Thus, different long-range interactions can play roles to alter the conformation of the 5'CAG/3'GAC motif in the ribosome. For example, in the bacterial ribosome the conformation of one of the terminal GC base pairs in a 5'CAG/3'GAC motif does not form a canonical GC paired structure due to the formation of a pseudoknot. In the crystal structure of the ribosome-bound cricket paralysis virus, the 5'CAG/3'GAC motif has *anti-anti* AA conformation that is similar to Figures 6a and 6c.⁹¹ The crystal structure of r(GGCAGCAGGCC)₂ has two 5'CAG/3'GAC motifs that have both 1×1 nucleotide AA internal loops in *anti-anti* form similar to the structure presented in Figure 6b.²³ This crystal structure, however, has two sulfate (SO₄²⁻) molecules that interact with the amino groups of 1×1 nucleotide AA internal loops to stabilize the conformation. This observation further suggests that the interaction of ions or small molecules can affect the structures of the 1×1 nucleotide internal loops in

transcripts that contain expanded repeats in significant ways. These ligand-induced structural changes could also affect the binding of protein to these expanded repeats.

SUMMARY and CONCLUSIONS

Purines and pyrimidines have different electronic structures, and this difference brings different properties to their nucleotides. One such structural difference is seen in the preferred glycosidic dihedral angles, χ , which are responsible for the base orientation with respect to sugar. While pyrimidines prefer mostly the *anti* conformation, purines have a tendency to be in the *syn* state.⁶⁹ This main difference between pyrimidines and purines will have an effect on properties of different RNA triplet repeats. Recently, the χ torsional parameters for the Amber force field were revised, yielding better structural and thermodynamic predictions for unique RNA systems.^{69,71,72} Backed with the revised χ parameter sets, structural and thermodynamic calculations on other RNA triplet repeats such as CUG, CCUG, CCG, and CGG will produce important results that can be used against the genetic neurological diseases described above.

In this report, we described a refined crystal structure of r[5'UUGGGC(CAG)₃GUCC]₂ (r(3xCAG)^{X-RAY}), a model of r(CAG)^{exp}, refined to 1.65 Å resolution. In conjunction with MD simulations, the structure showed that the 1×1 nucleotide AA internal loops sample *anti-anti* and *syn-anti* conformations. Moreover, the 1×1 nucleotide AA internal loops have structures that are sensitive to their environment. Our MD studies provide key insights to the properties of the AA internal loops. While the 1×1 nucleotide AA internal loops in r(3xCAG)^{X-RAY} repeats prefer the *anti-anti* conformation the *syn-anti* form is another stable conformation. The latter conformation has structure that is affected and stabilized by a Na⁺ binding pocket formed near this internal loop. Parallel to this result, the crystal structure shows dangling uridine bases stabilizing the *syn-anti* conformations. These studies thus suggest that a small molecule that binds to 1×1 nucleotide AA internal loops can affect conformation. Such information is useful in understanding the potential structural consequences of ligand binding to repeating transcripts and designing compounds that affect biological function.

For diseases caused by r(CAG)^{exp}, small molecules that can inhibit protein binding thus modulating RNA gain-of-function toxicity and inhibit translation of toxic polyQ proteins would be ideal. Both of these active areas of research^{24,92-94} can be aided by a more thorough understanding of the structural nature of expanded repeats including information about dynamics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Amit Kumar for purification of materials, discussions, and review of the manuscript and Dr. Pengfei Fang, and Dr. Min Guo of Scripps Research Institute, FL, for preliminary data. We thank Dr. Brent M. Znosko of Saint Louis University for helpful discussions. All computations were done in Advanced Research Computing (QUEST) at the Northwestern University, and High Performance Computing (SEPA) at the Scripps Research Institute, FL. Dr. Ilyas Yildirim thanks the Faculty of Science, Akdeniz University, Antalya, Turkey, for providing a working environment while he was in Turkey. This work was supported by NIH Grant 1R01GM079235-01A2 (MDD) and the PS-OC Center of the NIH/NCI Grant 1U54CA143869-01 (GCS).

1 Abbreviations

C	cytidine
G	guanosine
A	adenosine
QM	quantum mechanical
MM	molecular mechanics
MD	molecular dynamics
WHAM	weighted histogram analysis method
PMF	potential of mean force
PME	Particle Mesh Ewald
rmsd	root-mean-square deviation
MS1	r(5'GCGCAGCGC) ₂
MS2	r(5'CCGCAGCGG) ₂
3xCAG	r[5'GGGC(CAG) ₃ GUCC] ₂
r(3xCAG)^{X-RAY}	r[5'UUGGGC(CAG) ₃ GUCC] ₂

REFERENCES

- (1). Ashley CT, Warren ST. Annu. Rev. Genet. 1995; 29:703. [PubMed: 8825491]
- (2). Emery AEH. Lancet. 2002; 359:687. [PubMed: 11879882]
- (3). Sellier C, Rau F, Liu YL, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, Hagerman PJ, Charlet-Berguerand N. EMBO J. 2010; 29:1248. [PubMed: 20186122]
- (4). Greco CM, Berman RF, Martin RM, Tassone F, Schwartz PH, Chang A, Trapp BD, Iwahashi C, Brunberg J, Grigsby J, Hessel D, Becker EJ, Papazian J, Leehey MA, Hagerman RJ, Hagerman PJ. Brain. 2006; 129:243. [PubMed: 16332642]
- (5). Jin P, Duan RH, Qurashi A, Qin YL, Tian DH, Rosser TC, Liu HJ, Feng Y, Warren ST. Neuron. 2007; 55:556. [PubMed: 17698009]
- (6). Verheij C, Bakker CE, Degraaff E, Keulemans J, Willemsen R, Verkerk A, Galjaard H, Reuser AJJ, Hoogeveen AT, Oostra BA. Nature. 1993; 363:722. [PubMed: 8515814]
- (7). Verkerk A, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DPA, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, Eussen BE, Vanommen GJB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST. Cell. 1991; 65:905. [PubMed: 1710175]
- (8). Klockgether T, Paulson H. Mov. Disord. 2011; 26:1134. [PubMed: 21626557]
- (9). Orr HT, Zoghbi HY. Annu. Rev. Neurosci. 2007; 30:575. [PubMed: 17417937]
- (10). Shoulson I, Young AB. Mov. Disord. 2011; 26:1127. [PubMed: 21626556]
- (11). Walker FO. Lancet. 2007; 369:218. [PubMed: 17240289]
- (12). Yazawa I, Nukina N, Hashida H, Goto J, Yamada M, Kanazawa I. Nat. Genet. 1995; 10:99. [PubMed: 7647802]
- (13). Li LB, Yu ZM, Teng XY, Bonini NM. Nature. 2008; 453:1107. [PubMed: 18449188]
- (14). Li LB, Bonini NM. Trends Neurosci. 2010; 33:292. [PubMed: 20398949]
- (15). Mykowska A, Sobczak K, Wojciechowska M, Kozłowski P, Krzyzosiak WJ. Nucleic Acids Res. 2011; 39:8938. [PubMed: 21795378]
- (16). Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MAC, Nan Z, Forster C, Low WC, Schoser B, Somia NV, Clark HB,

- Schmechel S, Bitterman PB, Gourdon G, Swanson MS, Moseley M, Ranum LPW. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:260. [PubMed: 21173221]
- (17). Sobczak K, Michlewski G, de Mezer M, Kierzek E, Krol J, Olejniczak M, Kierzek R, Krzyzosiak WJ. *J. Biol. Chem.* 2010; 285:12755. [PubMed: 20159983]
- (18). Kumar A, Fang PF, Park H, Guo M, Nettles KW, Disney MD. *ChemBioChem.* 2011; 12:2140. [PubMed: 21766409]
- (19). Kumar A, Park H, Fang PF, Parkesh R, Guo M, Nettles KW, Disney MD. *Biochemistry.* 2011; 50:9928. [PubMed: 21988728]
- (20). Mooers BHM, Logue JS, Berglund JA. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:16626. [PubMed: 16269545]
- (21). Parkesh R, Fountain M, Disney MD. *Biochemistry.* 2011; 50:599. [PubMed: 21204525]
- (22). Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. *Nucleic Acids Res.* 2009; 37:4149. [PubMed: 19433512]
- (23). Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. *Nucleic Acids Res.* 2010; 38:8370. [PubMed: 20702420]
- (24). Kumar A, Parkesh R, Sznajder LJ, Childs-Disney JL, Sobczak K, Disney MD. *ACS Chem. Biol.* 2012; 7:496. [PubMed: 22252896]
- (25). Parkesh R, Childs-Disney JL, Nakamori M, Kumar A, Wang E, Wang T, Hoskins J, Tran T, Housman D, Thornton CA, Disney MD. *J. Am. Chem. Soc.* 2012; 134:4731. [PubMed: 22300544]
- (26). Krzyzosiak WJ, Sobczak K, Wojciechowska M, Fiszer A, Mykowska A, Kozlowski P. *Nucleic Acids Res.* 2012; 40:11. [PubMed: 21908410]
- (27). Malhotra A, Harvey SC. *J. Mol. Biol.* 1994; 240:308. [PubMed: 7518524]
- (28). Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J. Am. Chem. Soc.* 1995; 117:5179.
- (29). Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. *J. Comput. Chem.* 1983; 4:187.
- (30). MacKerell, J.; A, D.; Brooks, B.; Brooks, CL., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *The Encyclopedia of Computational Chemistry*. Schleyer, P. v. R.; Allinger, NL.; Clark, T.; Gasteiger, J.; Kollman, PA.; Schaefer, I.; H, F.; Schreiner, PR., editors. Vol. Vol. 1. John Wiley & Sons; Chichester: 1998. p. 271
- (31). Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF. *J. Phys. Chem. A.* 1999; 103:3596.
- (32). Hobza P, Sponer J. *J. Am. Chem. Soc.* 2002; 124:11802. [PubMed: 12296748]
- (33). Kratochvil M, Sponer J, Hobza P. *J. Am. Chem. Soc.* 2000; 122:3495.
- (34). Nam KH, Gao JL, York DM. *J. Am. Chem. Soc.* 2008; 130:4680. [PubMed: 18345664]
- (35). Bash PA, Field MJ, Karplus M. *J. Am. Chem. Soc.* 1987; 109:8092.
- (36). Eichinger M, Tavan P, Hutter J, Parrinello M. *J. Chem. Phys.* 1999; 110:10452.
- (37). Freindorf M, Gao JL. *J. Comput. Chem.* 1996; 17:386.
- (38). Gao JL, Xia XF. *Science.* 1992; 258:631. [PubMed: 1411573]
- (39). Murphy RB, Philipp DM, Friesner RA. *J. Comput. Chem.* 2000; 21:1442.
- (40). Stanton RV, Hartsough DS, Merz KM. *J. Phys. Chem.* 1993; 97:11868.
- (41). Tunon I, MartinsCosta MTC, Millot C, RuizLopez MF, Rivail JL. *J. Comput. Chem.* 1996; 17:19.
- (42). Duan Y, Kollman PA. *Science.* 1998; 282:740. [PubMed: 9784131]
- (43). Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. *Curr. Opin. Struct. Biol.* 2009; 19:120. [PubMed: 19361980]
- (44). Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, Arkin IT, Jensen MO, Xu HF, Trbovic N, Friesner RA, Palmer AG, Shaw DE. *J. Phys. Chem. B.* 2008; 112:6155. [PubMed: 18311962]
- (45). Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossvary I,

- Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan YB, Spengler J, Theobald M, Towles B, Wang SC. *Commun. ACM*. 2008; 51:91.
- (46). Zagrovic B, Pande V. *J. Comput. Chem.* 2003; 24:1432. [PubMed: 12868108]
- (47). Beveridge DL, McConnell K. *J. Curr. Opin. Struct. Biol.* 2000; 10:182.
- (48). Cheatham TE. *Curr. Opin. Struct. Biol.* 2004; 14:360. [PubMed: 15193317]
- (49). Giudice E, Lavery R. *Acc. Chem. Res.* 2002; 35:350. [PubMed: 12069619]
- (50). Orozco M, Perez A, Noy A, Luque FJ. *Chem. Soc. Rev.* 2003; 32:350. [PubMed: 14671790]
- (51). Perez A, Blas JR, Rueda M, Lopez-Bes JM, de la Cruz X, Orozco M. *J. Chem. Theory Comput.* 2005; 1:790.
- (52). Vokacova Z, Budesinsky M, Rosenberg I, Schneider B, Sponer J, Sychrovsky V. *J. Phys. Chem. B*. 2009; 113:1182. [PubMed: 19128019]
- (53). Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. *J. Comput. Chem.* 1992; 13:1011.
- (54). Souaille M, Roux B. *Comput. Phys. Commun.* 2001; 135:40.
- (55). Bartels C, Karplus M. *J. Comput. Chem.* 1997; 18:1450.
- (56). Virnau P, Muller M. *J. Chem. Phys.* 2004; 120:10925. [PubMed: 15268122]
- (57). Babin V, Roland C, Darden TA, Sagui C. *J. Chem. Phys.* 2006; 125:204909. [PubMed: 17144742]
- (58). Kastner J. *Wiley Interdisciplinary Reviews Computational Molecular Science*. 2011; 1:932.
- (59). Giudice E, Varnai P, Lavery R. *Nucleic Acids Res.* 2003; 31:1434. [PubMed: 12595551]
- (60). Banavali NK, MacKerell AD. *J. Mol. Biol.* 2002; 319:141. [PubMed: 12051942]
- (61). Song K, Campbell AJ, Bergonzo C, de los Santos C, Grollman AP, Simmerling C. *J. Chem. Theory Comput.* 2009; 5:3105.
- (62). Hart K, Nystrom B, Ohman M, Nilsson L. *RNA*. 2005; 11:609. [PubMed: 15811914]
- (63). Otwinowski Z, Minor W. *Macromolecular Crystallography, Pt A*. 1997; 276:307.
- (64). Storoni LC, McCoy AJ, Read RJ. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:432. [PubMed: 14993666]
- (65). Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. *Acta Crystallogr D Biol Crystallogr.* 2002; 58:1948. [PubMed: 12393927]
- (66). Case, DA.; Darden, TA.; Cheatham, TEI.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Merz, KM.; Pearlman, DA.; Crowley, M.; Walker, RC.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, KF.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, DH.; Schafmeister, C.; Ross, WS.; Kollman, PA. *University of California San Francisco; San Francisco, CA*: 2006.
- (67). Joung IS, Cheatham TE. *J. Phys. Chem. B*. 2008; 112:9020. [PubMed: 18593145]
- (68). Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J. Chem. Phys.* 1983; 79:926.
- (69). Yildirim I, Stern HA, Kennedy SD, Tubbs JD, Turner DH. *J. Chem. Theory Comput.* 2010; 6:1520. [PubMed: 20463845]
- (70). Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. *Biophys. J.* 2007; 92:3817. [PubMed: 17351000]
- (71). Yildirim I, Stern HA, Tubbs JD, Kennedy SD, Turner DH. *J. Phys. Chem. B*. 2011; 115:9261. [PubMed: 21721539]
- (72). Yildirim I, Kennedy SD, Stern HA, Hart JM, Kierzek R, Turner DH. *J. Chem. Theory Comput.* 2012; 8:172. [PubMed: 22249447]
- (73). Ryckaert JP, Ciccotti G, Berendsen HJC. *J. Comput. Phys.* 1977; 23:327.
- (74). Priyakumar UD, MacKerell AD. *Chem. Rev.* 2006; 106:489. [PubMed: 16464016]
- (75). Sagui C, Pedersen LG, Darden TA. *J. Chem. Phys.* 2004; 120:73. [PubMed: 15267263]
- (76). Toukmaji A, Sagui C, Board J, Darden T. *J. Chem. Phys.* 2000; 113:10913.
- (77). Roux B. *Comput. Phys. Commun.* 1995; 91:275.

- (78). Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. *J. Comput. Chem.* 1995; 16:1339.
- (79). Stelzer AC, Frank AT, Kratz JD, Swanson MD, Gonzalez-Hernandez MJ, Lee J, Andricioaei I, Markovitz DM, Al-Hashimi HM. *Nat. Chem. Biol.* 2011; 7:553. [PubMed: 21706033]
- (80). Davis AR, Znosko BM. *Biochemistry.* 2007; 46:13425. [PubMed: 17958380]
- (81). Davis AR, Znosko BM. *Biochemistry.* 2010; 49:8669. [PubMed: 20681613]
- (82). Davis AR, Kirkpatrick CC, Znosko BM. *Nucleic Acids Res.* 2011; 39:1081. [PubMed: 20876693]
- (83). Berk V, Zhang W, Pai RD, Cate JHD. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:15830. [PubMed: 17038497]
- (84). Borovinskaya MA, Pai RD, Zhang W, Schuwirth BS, Holton JM, Hirokawa G, Kaji H, Kaji A, Cate JHD. *Nat. Struct. Mol. Biol.* 2007; 14:727. [PubMed: 17660832]
- (85). Borovinskaya MA, Shoji S, Fredrick K, Cate JHD. *RNA.* 2008; 14:1590. [PubMed: 18567815]
- (86). Borovinskaya MA, Shoji S, Holton JM, Fredric K, Cate JHD. *ACS Chem. Biol.* 2007; 2:545. [PubMed: 17696316]
- (87). Dunkle JA, Xiong L, Mankin AS, Cate JHD. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:17152. [PubMed: 20876128]
- (88). Llano-Sotelo B, Dunkle J, Klepacki D, Zhang W, Fernandes P, Cate JHD, Mankin AS. *Antimicrob. Agents Chemother.* 2010; 54:4961. [PubMed: 20855725]
- (89). Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JHD. *Science.* 2005; 310:827. [PubMed: 16272117]
- (90). Schuwirth BS, Day JM, Hau CW, Janssen GR, Dahlberg AE, Cate JHD, Vila-Sanjurjo A. *Nat. Struct. Mol. Biol.* 2006; 13:879. [PubMed: 16998486]
- (91). Schueler M, Connell SR, Lescoute A, Giesebrecht J, Dabrowski M, Schroeer B, Mielke T, Penczek PA, Westhof E, Spahn CMT. *Nat. Struct. Mol. Biol.* 2006; 13:1092. [PubMed: 17115051]
- (92). Pushechnikov A, Lee MM, Childs-Disney JL, Sobczak K, French JM, Thornton CA, Disney MD. *J. Am. Chem. Soc.* 2009; 131:9767. [PubMed: 19552411]
- (93). Hu JX, Matsui M, Corey DR, Rossi JJ, Gait MJ, Eckstein F. *Oligonucleotide Therapeutics.* 2009; Vol. 1175:24.
- (94). Hu JX, Matsui M, Gagnon KT, Schwartz JC, Gabillet S, Arar K, Wu J, Bezprozvanny I, Corey DR. *Nat. Biotechnol.* 2009; 27:478. [PubMed: 19412185]

Figure 1.
Sequences and structures of the RNA model systems used in molecular dynamics and umbrella sampling simulations.

Figure 2. Reaction coordinates (χ and θ) used in PMF calculations for model system MS2 (Figure 1). χ torsional angle, O4'-C1'-N9-C4, is responsible of base orientation with respect to sugar (a). Center-of-mass (COM) pseudo-dihedral angle, θ , is chosen in order to mimic base unstacking of A with respect to helical axis. θ is a pseudo-dihedral angle defined by COMs of four atom groups shown in (b).

Figure 3.

Secondary structure, refined structure, and crystal packing of the CAG repeat RNA construct ($r(3xCAG)^{X-RAY}$). (A), the secondary structure of the oligonucleotide $3x(CAG)$ duplex that was used for crystallization. (B), the central $5'CA_8G/3'GA_{25}C$ of the RNA model structure for $r(3xCAG)^{X-RAY}$. The electron density maps are contoured at 1.0σ . (C), Illustration of crystal packing of $r(3xCAG)^{X-RAY}$ (orange). Crystal packings shown in gray are crystallographic symmetry-related molecules. Origin (O), a -, and b -axes are labeled accordingly. c -axis is perpendicular to the plain of view. Each helix is extended to make a pseudo-infinite helix along the c -axis. (D), overlay of the backbone of $r(3xCAG)^{X-RAY}$ (orange) onto the backbone of an ideal A-form RNA duplex (green) in which the 1×1 nucleotide AA internal loops and GU pairs are replaced with AU and GC pairs, respectively. Note that $5'UU$ dangling ends interact with terminal AA base pairs from major groove sides in $r(3xCAG)^{X-RAY}$. (E) Magnified image of A_6A_{29} pair to show interaction with U_{-2} . Note that amino and $2'OH$ groups of A_{28} are in close contact with U_{-2} .

Figure 4.

Stick models and electron densities of the 1 $\times$ 1 nucleotide AA internal loops for r(3xCAG)^{X-RAY}. Electron density maps are contoured at 1.5 σ (red), 1.2 σ (blue), 1.0 σ (red mesh), and 0.6 σ (gray mesh). Hydrogen bonds, represented as dashed lines, and the distance values (in Å) are labeled.

Figure 5. 2D PMF surface for model system MS2 showing the three pathways for *syn-anti* ↔ *anti-anti* AA transformation. The x- and y-axis are the reaction coordinates for stacking (θ) and base orientation with respect to sugar (χ) of adenosine, respectively (Figure 2). *Anti* and *syn* regions as well as the minor and major groove directions are shown. 1D free energy profiles of each pathway along the most likely direction are shown in Figures S3 and S4.

Figure 6.

Different types of AA base pairings seen in MD simulations of model systems. Color notation was used to distinguish different AA types, which are also used in Figures 8, S5, and S6 to emphasize the type of transformations seen in the simulations. Arrow directions mean the type of transformations seen in the MD simulations. (a), (b), and (c) have *anti-anti* AA conformation while (d) and (e) have *syn-anti* AA conformation. (d) is a transition state with two hydrogen-bonds seen in *syn-anti*→*anti-anti* AA transformation (3xCAG). No *anti-anti*→*syn-anti* AA transformation was seen, while multiple *anti-anti*→*anti-anti* AA transformations were observed. Note that (a) and (c) are symmetric states.

Figure 7.
 Na^+ -bound state in *syn-anti* AA base pairs.

Figure 8.

Rmsd and Na^+ binding analysis of 3xCAG MD simulation. Residue numbering defined in Figure 1c is used in the analysis. Rmsd (in Å) of (a) backbone, (b) A6-A29 base pair, (d) G7C8/G27C28, (e) A9-A26 base pair, (g) G10C11/G24C25, and (h) A12-A23 base pairs with respect to initial structure. The system has a total of 32 Na^+ ions highlighted with different colors, and distances (in Å) between each Na^+ ion to AA base pairs are plot in (c), (f), and (i). In (b), (e), and (h) color schemes defined in Figure 6 for each AA type are used in order to highlight the transformations. Note that there is a Na^+ binding pocket when AA is in *syn-anti* conformation between 0-130 ns (i).

Table 1

Data collection and refinement statistics

	r(3xCAG)^{X-RAY}
Data collection	
Space group	H3
Unit cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	46.3, 46.3, 134.6
α , β , γ (°)	90, 90, 120
Resolution (Å)	50-1.65 (1.71-1.65) ^a
Total measured reflections	228,869
Unique reflections	12,952
R_{merge} [%] ^b	5.1 (49.5) ^a
$I/\sigma(I)$	40.4 (3.8) ^a
Completeness (%)	99.8 (100.0) ^a
Redundancy	5.7 (5.6) ^a
Refinement	
Resolution (Å)	22.43-1.65 (1.82-1.65) ^a
Number of reflections	12,318
R_{work} ^c / R_{free} ^d	16.6/17.9 (18.9/17.9) ^a
Number of atoms	
RNA	808
Ligand/ion	10
Solvent	112
Average <i>B</i> -factors	
RNA	50.5
Ligand/ion	23.7
Solvent	38.3
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.004
Bond angles (°)	0.697

^aValues in parentheses are for the highest resolution shell.

^b $R_{\text{merge}} = \frac{\sum_h \sum_l |I(h)_l - \langle I(h) \rangle|}{\sum_h \sum_l I(h)_l}$, where $I(h)_l$ is the *l*th observation of the reflection *h* and $\langle I(h) \rangle$ is the weighted average intensity for all observations *l* of reflection *h*.

^c $R_{\text{work}} = \frac{\sum_h |F_{\text{obs}}(h) - F_{\text{cal}}(h)|}{\sum_h F_{\text{obs}}(h)}$, where $F_{\text{obs}}(h)$ and $F_{\text{cal}}(h)$ are the observed and calculated structure factors for reflection *h* respectively.

^d R_{free} was calculated as R_{work} using the 5% of reflections which were selected randomly and omitted from refinement.