Published in final edited form as:
ACS Chem Biol. 2014 February 21; 9(2): 538–550. doi:10.1021/cb4007387.

Structure of the Myotonic Dystrophy Type 2 RNA and Designed Small Molecules That Reduce Toxicity

Jessica L. Childs-Disney[#a], **Ilyas Yildirim**[#b], **HaJeung Park**[a,c], **Jeremy R. Lohman**[a], **Lirui Guan**[a], **Tuan Tran**[a], **Partha Sarkar**[d], **George C. Schatz**[b], and **Matthew D. Disney**[a,*]

[a] Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458

[b] Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3113

[c] Translational Research Institute, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458

[d] Department of Neurology, University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-0539

[#] These authors contributed equally to this work.

Abstract

Myotonic dystrophy type 2 (DM2) is an untreatable neuromuscular disorder caused by a r(CCUG) expansion (r(CCUG)^{exp}) that folds into an extended hairpin with periodically repeating 2×2 nucleotide internal loops (5'C<u>CU</u>G/3'G<u>UC</u>C). We designed multivalent compounds that improve DM2-associated defects using information about RNA-small molecule interactions. We also report the first crystal structure of r(CCUG)^{exp} refined to 2.35 Å. Structural analysis of the three 5'C<u>CU</u>G/3'G<u>UC</u>C repeat internal loops (L) reveals that the CU pairs in L1 are each stabilized by one hydrogen bond and a water-mediated hydrogen bond while CU pairs in L2 and L3 are stabilized by two hydrogen bonds. Molecular dynamics (MD) simulations reveal that the CU pairs are dynamic and stabilized by Na^+ and water molecules. MD simulations of the binding of the small molecule to r(CCUG) repeats reveal that the lowest free energy binding mode occurs via the major groove, in which one C residue is unstacked and the cross-strand nucleotides are displaced. Moreover, we modeled the binding of our dimeric compound to two 5'C<u>CU</u>G/3'G<u>UC</u>C motifs, which shows that the scaffold on which the RNA-binding modules are displayed provides an optimal distance to span two adjacent loops.

[*] author to whom correspondence is addressed; Disney@scripps.edu.

AUTHOR CONTRIBUTIONS

M.D.D., H.P., I.Y., G.C. and J.L.C. designed experiments; J.L.C., I.Y., H.P., and M.D.D. wrote the manuscript; J.L.C. completed biological assays and biophysical measurements; I.Y. completed computational analyses; H.P. crystallized r(CCUG) repeats, collected diffraction data, and helped refine those data; J.R. refined diffraction data; T.T. synthesized RNA for crystallization studies; L.G. synthesized the control compound **3N-4**: P.S. cloned r(CCUG)$_{300}$ construct used in biological studies.

Accession codes

Coordinate and diffraction data have been deposited in the Protein Data Bank under accession code 4K27.

Supporting Information. Methods for computational analysis and MD simulations, movie descriptions, crystallographic data collection and refinement parameters, supplementary tables and figures for MD and crystallographic studies, and representative gel images. This material is available free of charge via the Internet at http://pubs.acs.org.

INTRODUCTION

The cellular function of RNA is most commonly modulated using oligonucleotides that recognize the RNA's sequence and induce RNase H cleavage (antisense) or the RNA interference (RNAi) pathway. Oligonucleotide-based therapeutics are advantageous for many reasons, including ease of design using simple base pairing rules and the availability of modifications that impart greater binding affinity and resistance to nucleases. Indeed, oligonucleotide-based therapeutics that treat hypercholesterolemia (mipomersen)[1] and cytomegalovirus (CMV) retinitis (fomivirsin)[2] have been approved by the FDA while others are in clinical trials. Oligonucleotides have been successfully employed to study biology and to improve disease-associated defects in cellular and animal models of disease[3-6]. Despite these benefits and successes, oligonucleotides have sub-optimal pharmacokinetic properties, the most significant of which is poor delivery[7-9]. Small molecules can also be useful modalities to target RNA, as has been demonstrated for antibacterials that target ribosomal RNA[10-12]. These compounds have found applications as lead therapeutics and as chemical probes of function that elucidated the intricacies of the translational machinery[13-16].

Despite the great interest in developing small molecules that modulate non-ribosomal RNA function in cells, these RNA targets remain largely untapped[10, 11]. Often, a validated target is subjected to screening to identify lead compounds[17-21]. However, such approaches generally have lower hit rates for RNA targets compared to protein counterparts and yield hits that bind with modest affinity. An alternative to this "top-down" approach is to use information about the binding of small molecules to RNA secondary structural elements (or motifs) to inform compound design (Figure 1A). We developed a microarray-based selection method that screens chemical and RNA motif space simultaneously and identifies privileged RNA space for each small molecule (those that bind with high affinity and selectivity)[22, 23]. The motifs that comprise an RNA target are then compared with our identified interactions to afford lead small molecules (Figure 1A).

In one of our original selections, we discovered that 6'-*N*-5-hexynoate kanamycin A (**K-alkyne**; Figure 1A) binds pyrimidine-rich, 2×2 nucleotide internal loops (5'UU/3'UC or 5'CU/3'UU) with high affinity and selectivity[22, 24]. Interestingly, the RNA that causes myotonic dystrophy type 2 (DM2) contains regularly repeating internal loops that are similar to those identified from our selection (Figure 1A). Thus, **K-alkyne** is a lead small molecule for the RNA that causes DM2 (Figure 1B).

DM2 is caused via a toxic gain-of-function by an expanded tetranucleotide repeat. The expanded repeat, or r(CCUG)^{exp}, resides in intron 1 of the zinc finger 9 (*ZNF9*) pre-mRNA; disease occurs when >75 repeats are present (range: >75 – 11,000)[25]. r(CCUG)^{exp} folds into a hairpin with multiple copies of a 5'C<u>CU</u>G/3'G<u>UC</u>C internal loop that binds and inactivates the pre-mRNA splicing regulator muscleblind-like 1 protein (MBNL1) (Figure 1B). Sequestration of MBNL1 by r(CCUG)^{exp} causes pre-mRNA splicing defects[25]. Another defect associated with DM2 is the formation of nuclear foci caused by r(CCUG)^{exp}-protein aggregates[26, 27]. Importantly, this disease model points to a therapeutic approach for DM2: a small molecule drug binds r(CCUG)^{exp} and inhibits the binding of proteins, in particular MBNL1, to rescue their normal cellular functions (Figure 1B).

Because r(CCUG)^{exp} contains regularly repeating 5'C<u>CU</u>G/3'G<u>UC</u>C internal loops, it is an ideal target for a modular assembly approach (Figure 1B). That is, compounds that display copies of a small molecule (**K-alkyne**) with the same periodicity as 5'C<u>CU</u>G/3'G<u>UC</u>C can bind adjacent loops simultaneously (Figure 1B), resulting in enhanced affinity and selectivity for r(CCUG)^{exp} relative to a monomeric small molecule[28]. We previously reported the design of multivalent small molecules that display **K-alkyne** and characterized

them *in vitro* (Figure 1)[28]. By using an *in vitro* assay that measures r(CCUG)$_{24}$-MBNL1 complex formation, we determined the optimal distance between **K** RNA-binding modules (where **K** indicates the conjugated form of **K-alkyne**). The optimal compounds bind r(CCUG) repeats with nanomolar affinities and inhibit the r(CCUG)$_{24}$-MBNL1 interaction *in vitro* with nM IC_{50}'s[28].

Another commonly employed tool to discover protein inhibitors, and to a lesser extent RNA inhibitors, is structure-based design. Structure-based design requires the structure of the biomolecule as determined by NMR spectroscopy or X-ray crystallography. Small molecule ligands are then docked into the structure and output in rank order in terms of binding affinity. Such an approach has been used to design compounds that bind riboswitches and r(CUG) repeats, the causative agent of myotonic dystrophy type 1 (DM1)[29, 30]. Unfortunately, the structure of r(CCUG) repeats has not been determined. Therefore, we sought to determine their structure(s) by X-ray crystallography.

In this report, we describe the results of two important studies: (i) the reduction of r(CCUG)^{exp} toxicity by our designed compounds in cellular models of DM2; and (ii) the first structure of r(CCUG)^{exp}. In particular, designed modularly assembled compounds improve DM2-associated dysregulation of alternative pre-mRNA splicing. In a second series of studies, a structure of a model of r(CCUG)^{exp}, refined to 2.35 Å, provides the first atomic-level picture of these toxic repeats. The 5'C<u>CU</u>G/3'G<u>UC</u>C motifs form ordered 2×2 nucleotide internal loops; each CU pair is stabilized by two hydrogen bonds. Additionally, high negative electrostatic potential is found in the minor groove of r(CCUG)^{exp}, which is neutralized by the bridging water molecules. The presence of bridging water molecules is supported by MD simulations and provides a rationale for the high affinity binding and selectivity of our designed, modularly assembled aminoglycosides. Moreover, free energy calculations of **K-acetyl** binding to r(CCUG) repeats using MD simulations reveal that the lowest free energy binding mode occurs via the major groove. In this binding mode, one C residue is unstacked while the cross-strand nucleotides are displaced to allow for favorable small molecule-RNA interactions. Using this binding mode, we modeled the binding of our optimal dimeric compound to two 5'C<u>CU</u>G/3'G<u>UC</u>C motifs. The compound, **2K-4**, displays two **K** RNA-binding modules displayed on a peptoid backbone. The distance between **K**'s afforded by the peptoid backbone closely mimics the distance between the 5'C<u>CU</u>G/3'G<u>UC</u>C motifs, with the peptoid lying along the major groove and each **K** interacting with a CU/UC loop.

RESULTS & DISCUSSION

Designed small molecules improve DM2-associated defects

Myotonic dystrophy type 2 (DM2) is caused by an expanded r(CCUG) repeat (r(CCUG)^{exp}) that binds and sequesters the pre-mRNA splicing regulator muscleblind-like 1 protein (MBNL1). We previously designed multivalent compounds that bind r(CCUG) repeats with nanomolar affinities and inhibit r(CCUG)$_{24}$-MBNL1 complex formation with nanomolar IC_{50}'s[28]. These compounds are comprised of a peptoid backbone that displays multiple copies of **K-alkyne**, a 5'C<u>CU</u>G/3'G<u>UC</u>C-binding module. Using an *in vitro* assay that measures r(CCUG)$_{24}$-MBNL1 complex formation, we determined that the optimal distance between **K-alkyne** RNA-binding modules is afforded by four propylamine spacing modules[28]. That is, four propylamines most closely mimic the periodicity of 5'C<u>CU</u>G/3'G<u>UC</u>C in r(CCUG)^{exp}. The compounds are thus referred to as **nK-4** where **nK** indicates valency and **4** indicates four propylamine spacing modules. The binding of the designer compounds to DM2 repeats is highly selective. For example, **3K-4** binds to r(CCUG) repeats with >250-fold selectivity over base paired RNAs and with >50-fold selectivity over RNAs with modest, single nucleotide modifications to the internal loop[28]. Furthermore,

3K-4 binds to r(CCUG)^{exp} repeats with 15-fold higher affinity than MBNL1[28]. The **nK-4** compounds are cell permeable in the absence of a transfection agent[28]. Collectively, these studies suggest that designer small molecules have high potential to ameliorate DM2 defects in cells, however, a model cellular system for studying DM2 pre-mRNA splicing defects was not available to study their bioactivity when these compounds were originally reported.

As mentioned above, various defects are associated with DM2 including the formation of nuclear foci comprised of r(CCUG)^{exp}-protein complexes and the dysregulation of alternative pre-mRNA splicing due to sequestration of MBNL1. Recently, it was shown that the alternative splicing of bridging integrator-1 (*BIN1*) pre-mRNA splicing is dysregulated in skeletal muscle of DM1 and DM2 patients[31]. BIN1 is required for biogenesis of muscle tubules, specialized skeletal muscle membrane structures required for excitation-contraction coupling. In DM2-affected cells, exon 11 is skipped too frequently, resulting in the production of an inactive BIN1 protein[31]. The intronic region surrounding exon 11 contains several UCG steps that are known to bind MBNL1, which regulates *BIN1* alternative splicing[31].

Thus, **K-alkyne**, **2K-4**, and **3K-4** were studied for improving the *BIN1* pre-mRNA splicing defect in a DM2 model cellular system (Figure 2). Briefly, C2C12 cells were co-transfected with plasmids that express r(CCUG)$_{300}$ and a *BIN1* mini-gene that reports on alternative splicing (Figure 2A)[31]. In this model system, exon 11 has an inclusion rate of ~40% in healthy cells (do not express r(CCUG)$_{300}$) and only ~20% in DM2-like cells (express r(CCUG)$_{300}$) (Figure 2). This difference provides a readout to assess the bioactivity of designer small molecules.

After transfection, cells were treated with **K-alkyne**, **2K-4**, or **3K-4** in medium containing 2% horse serum (no transfection vehicle). After 2 days, *BIN1* alternative splicing patterns were determined by RT-PCR. Each kanamycin-containing compound improves the dysregulation of *BIN1* alternative splicing (Figure 2B). The most modest improvement is observed for **K-alkyne**, which improves the defect by ~50% when cells are treated with 20 M compound. Increasing the dosage does not further improve dysregulation. Increased potency is observed for **2K-4** and **3K-4**: treatment with 20 M of **2K-4** improves *BIN1* pre-mRNA splicing pattern by ≈80% (as compared to cells that do not express r(CCUG)$_{300}$) while treatment with 20 μM of **3K-4** restores the *BIN1* alternative splicing pattern to that observed in healthy cells (Figure 2B). Improvement in splicing patterns is observed when cells are treated with as little as 2.5 μ? of **3K-4** (Figure 2B). None of the compounds affects *BIN1* alternative splicing in cells that do not express r(CCUG)$_{300}$ (Figure 2C) nor do they affect the alternative splicing of *SMN2* pre-mRNA, which is controlled by Sam68[32] (Figure 2D). Thus, bioactivity is due to disruption of the r(CCUG)^{exp}-MBNL1 complex, freeing MBNL1 and restoring its function. A control compound was also studied, **3N-4**, a peptoid to which 6'-*N*-5-hexynoate neamine was coupled (Figure 1C). Previous studies determined that **3N-4** binds weakly to r(CCUG) repeats and is a poor inhibitor of the r(CCUG)$_{24}$-MBNL1 interaction[28]. As expected, this compound is inactive (Figure 2C). Thus, we have demonstrated that by using a "bottom-up" design approach that the bioactive small molecules can be developed for the RNA that cause DM2, the first ones reported.

Biophysical analysis of the binding of small molecules and r(CCUG) repeats

The binding of bioactive small molecules to r(CCUG) was characterized by using bio-layer interferometry (BLI) (a method similar to surface plasmon resonance (SPR) spectroscopy) (Table 1). These analyses determined that **K-alkyne**, **2K-4**, and **3K-4** bind r(CCUG)$_{12}$ with K_ds of 4400, 27, and 5 nM, respectively. Thus, affinity for the RNA target increases as valency increases, as expected. Interestingly, the on and off rates (k_{on} and k_{off}) are also

affected by valency. **2K-4** and **3K-4** reside on r(CCUG)$_{12}$ ~20-times longer than **K-alkyne**. The k_{on} rates are ~7- and ~50-fold faster for **2K-4** and **3K-4**, respectively, than **K-alkyne**. Importantly, designer modularly assembled compounds have faster k_{on} and slower k_{off} rates for binding r(CCUG) repeats than MBNL1 does. MBNL1 has a k_{off} of 0.013 s^{-1} and a k_{on} of 1.3×10^5 $M^{-1}s^{-1}$, or ~5-fold faster k_{off} and a ~5-fold slower k_{on} than **3K-4**. The residence time for **3K-4** on r(CCUG)^{exp} is ~3-fold longer than the residence time for MBNL1.

These studies point to a binding model for multivalent compounds in which the k_{on} and k_{off}, are enhanced due to a "hinge effect" (Figure 1C). That is, binding of one RNA-binding module decreases the space that the second RNA-binding module can sample, increasing its effective concentration and acting as a hinge for interactions with additional 5'C<u>CU</u>G/3'G<u>UC</u>C motifs. Collectively, they help to establish a mechanistic framework for why modularly assembled compounds more effectively modulate the biological activity of r(CCUG)^{exp} than monomeric **K-alkyne** and how they can outcompete MBNL1. These conclusions are likely applicable to the binding of other multivalent compounds to repeating transcripts. Additionally, the faster k_{on} for designer small molecules as compared to MBNL1 lends support to differences in their molecular recognition of r(CCUG)^{exp}. Previous studies have suggested that MBNL1 binding induces a chain-reversal trajectory and unzipping of the bound RNA[33, 34]. Small molecules, such as **3K-4,** most likely do not have to open up the RNA's structure for binding to occur. This translates into a faster k_{on} as structural rearrangements are not required.

Structural studies on models of r(CCUG)^{exp}

In order to gain insight into the structure of r(CCUG) repeats, a model RNA containing three copies of 5'C<u>CU</u>G/3'G<u>UC</u>C embedded adjacent to a tetraloop-tetraloop receptor was subjected to crystallization (Figure 3A). The GAAA tetraloop and its receptor have been utilized previously to promote crystallization of RNA constructs[35, 36]. The crystal diffracted to a resolution of 2.35 Å, and the structure was determined by the molecular replacement method (Figure 3B and Table S-1). All bases were well resolved with an exception of U36, which is part of the tetraloop receptor. The most significant packing contacts are derived from the tetraloop-tetraloop receptor interaction and base stacking interactions between end GC pairs of neighboring symmetry molecules. Analysis of helical parameters, inclinations and helical twist revealed the average values of these parameters throughout the CCUG repeat are within those of A-form RNA. However, significant deviations from A-form RNA were observed around the central 5'CU/3'UC loop L2 (Figure 3A; Tables S-2 & S-3). Base pair axis analysis reveals that the r(CCUG) repeat helix is bent approximately 18.5°. In conjunction with this bending, L1 shows widening of the both major and minor grooves while L2 and L3 show narrowing of both grooves (Table S-4). Electrostatic potential surface analysis reveals the impact of the four carbonyl oxygens of the 5'CU/3'UC pairs, which appear as patches of prominent electronegative surfaces in the minor groove (Figure 3C).

Central CU pairs and closing GC pairs

Among the three CCUG repeats, the 5'CCUG repeat (L1; Figure 3A) has a different loop conformation than L2 and L3, which have similar conformations. The CU pairs of L1 are each stabilized by one hydrogen bond between N4H(C) and O4(U) and a bridging water molecule wedged between N3(C) and N3H(U) (Figure 4A). This type of CU pairing with a bridging water molecule was first observed in a crystal structure of RNA duplex containing a central 5'U<u>UC</u>G/3'G<u>CU</u>U internal loop[37, 38]. NMR spectroscopy and molecular dynamics (MD) simulations also predicted the water-mediated interaction between CU pairs in a RNA duplex[39, 40]. While the bridging water molecules are not necessary for CU pairs to be incorporated into a paired region[41], weak hydrogen bonds can be further stabilized by bridging water molecules. The pairing interaction stabilized by a bridging water molecule

causes significant opening of the minor groove with C1'-C1' distance increased by ~1 Å (Figure 4A and Table S-4). Moreover, this induces strain on the closing GC base pairs and neighboring base pairs (observed as a propeller twist and buckling) (Table S-5).

While the CU pairs of L1 form one hydrogen bond, the CU pairs of L2 and L3 each make two hydrogen bonds: between N4H(C) and O4(U) and between N3(C) and N3H(U) (Figure 4B). This type of base pairing is similar to the previously observed CU pairs in NMR structures[42, 43]. The hydrogen bond between N3(C) and N3H(U) brings the two carbonyl oxygens of cytosine and uracil in close proximity with an average distance of 3.2 Å. Similar to CU pairs observed in L1, the CU pairs in L2 and L3 also show prominent propeller twist to alleviate the electrostatic repulsion stress caused by neighboring carbonyls of cytosine and uracil. There is spurious electron density between U14 and C42 of L3 (Figure 4C). Although it could not be unambiguously assigned, it may be a water molecule or another component of the crystallization matrix that helps to alleviate electrostatic repulsion. Curves analysis of L2 and L3 shows narrowing of both major and the minor grooves by an average of −0.7 Å and −1.6 Å, respectively with average C1'-C1' distance of 8.6 Å (Figure 4B and Table S-4). The analysis also shows an ~14° opening toward the major groove at CU pairs. Contraction and propeller twist with concomitant opening of the CU pairs in L2 and L3 cause significant strain on the glycosidic bonds. Analysis of backbone torsion angles reveals large deviations from the values of average A-form RNA in L2 and L3 (Table S-6 and S-7). Surprisingly, sugar puckering of G15 and G44 were C2' endo rather than C3' endo, which is unusual for paired regions of RNA as the C2' hydroxyl disfavors this conformation. Globally, the distortion makes the phosphate backbone appear pinched at U47. Deformation of the phosphate backbone in this region appears to be the prime cause of bending of the helix through shifting of the helical axis (Figure S-2).

While U10, G11 and U47 of L2 and U14 and G15 of L3 are involved in crystal packing, C41 to G44 of L3 are completely open to the solvent channel and do not form interactions with neighboring bases. None of the base pairs in L1 is involved in crystal packing. Thus, the base pairings observed here do not appear to be the result of a packing artifact.

CU base pairs are dynamic as determined by MD simulations

Next, we investigated the dynamics of the CU pairs using MD simulations. Two model systems were designed using the X-ray crystal structure described: 2×CCUG, r(GCCUGCCUGC)$_2$; and 2×CCUG$_{inf}$, r(5'CCCCUGCCUGC/3'GGGUCCGUCCG) (Figure 5). 2×CCUG was designed and prepared to mimic an independent RNA molecule solvated in a truncated octahedral box while 2×CCUG$_{inf}$ was specifically designed and prepared to investigate an infinitely long RNA structure with CU/UC loops (Figures 5). Both systems were neutralized with Na^+ ions[44] and solvated with TIP3P water molecules[45]. Simulations were carried out with the PMEMD module in AMBER11[46] where revised χ[47] and α/γ[48] torsional parameters were used to define parameters for RNA residues. Eight conformations of the CU pairs were observed in MD trajectories (Figures 5 and S-4 – S-6 & Table S-8). Each CU pair was distinguished using a different color in molecular surface representations in order to emphasize transformations seen in the root mean square deviation (RMSD) analysis (Figures 5 & S-4).

The major conformation (58%) of the CU pair observed in the MD trajectories was CU_a (shown in black, Figures 5A, S-5, and S-6 & Table S-8), which is stabilized by two hydrogen bonds between the bases. The two free carbonyl groups in the minor groove side interact with either water molecules or Na^+ ions (Figures 5A & S-4). Na^+ ions spend as long as 25 ns in bound state while water molecules are bound for much shorter amounts of time (Figures S-7 – S-14). The preference of Na^+ ion over a water molecule to the free carbonyl groups was also predicted by SwS.[49] This hydrogen bonding pattern, supported by water

bridges and Na^+ ions, makes CU_a the major conformation (Movie S-1). Thus, cations enhance the stability of CU base pairs, providing a rationale for the binding of **K**, **2K-4**, and **3K-4**.

The second major conformation observed in MD trajectories is CU_b (16%; shown in red in Figures 5B, S-5, and S-6 & Table S-8). The conformation is same as L1 in our crystal structure (Figure 4). The bridging water molecule that stabilizes the conformation of the CU base pair and remains in the binding pocket for as long as 3-4 ns during the simulation (Figures S-7 – S-14 and Movie S-2).

The other CU conformations were observed less than 10% of the time in the MD trajectories (Figure S-4 and Table S-8). The CU pairs in these conformations had either one (CU_g in Figure S-4) or no hydrogen bonds with Na^+ ions and water molecules bridging the bases of C and U (CU_c in Figure S-4; see Movie S-3). Due to the flexible nature of the CU pairs, they can open up and lose all the hydrogen bonds (CU_f and CU_h in Figure S-4), which can give rise to flow of water molecules and ions between the bases (CU_e in Figure S-4) and, in some cases, create stable conformations (CU_d in Figure S-4). While transition state conformations are difficult to capture in crystal structures, the dominance of CU_a (58% in MD, and twice in crystal structure) over CU_b (16% in MD and once in crystal structure) corroborate each other and suggest that CU_a is the preferred conformation for the DM2 motif.

Docking of a K-alkyne derivative (K-acetyl) into r(CCUG) repeats

In order to gain insight into the molecular recognition of r(CCUG) repeats by **K-alkyne**, we completed docking studies (Figure S-15). For these studies, a model RNA system was employed that contains 2×2 nucleotide CU/UC internal loops, or r(5'CCCCUGGG/ 3'GGGUCCCC). The hexynoate in **K-alkyn**e was replaced with an acetyl group, affording **K-acetyl** (Figure 6A), to aid computational analysis. The GAFF force field[50] was used to define the parameters for **K-acetyl** and binding modes were simulated essentially as previously described (see the Supplementary Information for details)[51]. Briefly, MD simulations of a total of 3 μs were run in implicit solvent to sample the simulation space. Cluster analysis was completed on the MD trajectories, which showed over 100 binding modes for **K-acetyl**-internal loop complex (Tables S-9 & S-10).

The binding affinity of each binding mode was calculated by using free energy analysis (MM-PBSA approach[46, 52]) (Tables S-9 & S-10). In the lowest free energy binding mode, which is 2.9 kcal/mol more stable than the second lowest free energy binding mode (Table S-9), the loop residue $C_{i'}$ is pushed out by **K-acetyl**'s ring II and becomes unstacked from neighboring bases (Figure 6A). The adjacent CU pair (C_i-U_{i+1}) is opened towards the minor groove side, forming a hydrogen bond with **K-acetyl.** The conformation of this CU pair is similar to the conformation of L1 in our crystal structure (Figure 6A). No contacts are made between the acetyl group and the RNA.

There are extensive attractive electrostatic interactions between **K-acetyl** and the 5'CU/ 3'UC loop in the lowest free energy structure (Figure 6A). Two charged amino groups (N_4 and N_1 in ring I and ring II, respectively)[53] interact with backbone phosphate groups (Figure 6A). In addition, amino group N_1 makes a hydrogen bond with the carbonyl group of the unstacked cytosine, C_i (Figure 6A). In this binding mode, all three hydroxyl groups in ring III of **K-acetyl** interact with the phosphate groups on the opposite RNA strand (Figure 6A). The lone hydroxyl group in ring II is hydrogen bonded to C_{i+1}'s amino group, which is crosswise to the unstacked cytosine, $C_{i'}$. N_2 of ring II makes a hydrogen bond with the uracil, U_{i+1}, which is opposite to the unstacked cytosine, C_i. There are no interactions between the other uracil residue, U_{i+1} and **K-acetyl** (Figure 6A). It is noteworthy to

emphasize that the global RNA structure of the lowest free energy binding mode has one of the lowest RMSDs from A-form conformation (Table S-9). Thus, the global RNA conformation does not change significantly from A-form upon binding of **K-acetyl**.

Modeling of 2K-4 binding to r(CCUG) repeats

Next, we modeled the binding of **2K-4** to two 5'CCUG/3'GUCC repeats (Figure 6B). Briefly, two **K-acetyl**-5'CU/3'UC structures (lowest free energy structure shown in 6A) were placed adjacent to each other separated by two GC base pairs (5'CCUGCCUG/3'GUCCGUCC). The peptoid scaffold was then added to link the two **K** RNA-binding modules together. The trajectory of the peptoid scaffold was calculated through MD simulation while positional constraints were imposed on the **K** RNA-binding modules and the RNA. MD simulations showed that the peptoid scaffold is mostly linear, located at the center along the major groove between the two CU/UC loops. Two of the propyl side chains make weak van der Waals interactions with the GC closing pair proximal to the unstacked $C_{i'}$. Importantly, the distance afforded by the four propylamine spacing modules closely mimics the distance between the two CU/UC loops, providing a rationale for its better *in vitro* potency compared to other **2K-n** compounds[28].

Relevance of short r(CCUG) repeat structure to disease

DM2 is caused by the presence of at least 75 r(CCUG) repeats; thus, how relevant are the structures for three repeats described herein? It is likely that the internal loops in $r(CCUG)^{exp}$ are dynamic *in vivo*, interconverting between the conformations observed in the crystal structure and MD simulations, with CU_a as the preferred conformation. This dynamic nature may provide insight into MBNL1 binding. MBNL1 interacts specifically with RNA via four zinc finger (ZnF) domains. A crystal structure of the ZnF3/4 domain complexed with two copies of single stranded r(CGCUGU) was recently reported[34]. Both zinc fingers interact with one molecule of RNA with ZnF3 forming contacts to the 5′GC step and ZnF4 forming contacts to 5′GCU. The RNA molecules are oriented antiparallel to each other. It is likely that MBNL1 opens the r(CCUG) hairpin stem to afford two single stranded regions, which is supported by the fact that MBNL1 binds weakly to fully base-paired RNAs[54] and by breaking of cTNT pre-mRNA structure upon MBNL1 binding[33]. The dynamic nature of the 5'CU/3'UC loops may facilitate MBNL1 binding and remodeling of the RNA's structure. It should be noted that the crystal structure of short r(CUG) repeats[55] was used successfully to design small molecule binders[29]. It is likely that the structures of our short r(CCUG) repeats can also facilitate structure-based design. Information about the structures and the dynamic trajectories of these internal loops are of paramount importance to design small molecules that target them specifically over base pairs. The MD studies described herein suggest that knowledge of structural properties including dynamics can aid in understanding of RNA-small molecule interactions.

Comparison of the binding of K and r(CCUG) to kanamycin and the A-site

We completed *in silico* binding experiments in order to understand how **K-alkyne** selectively recognizes r(CCUG) repeats. The results show that **K-acetyl** interacts with CU/UC loop via major groove through extensive mixed hydrogen bond and electrostatic interactions with backbone phosphate groups and loop bases (Figure 6A). Noticeably, there is no interaction between the RNA and acetyl group of **K-acetyl** (Figure 6A), in agreement with our experimental results in which derivatization of the 6-NH_2 (Figure 6A) has no effect on binding affinity[56]. It is also intriguing to note that the conformation of one of the CU loops in the lowest free energy binding mode of **K-acetyl**-5'CU/3'UC is similar to the conformation of L1 conformation observed in our crystal structure (Figures 4 & 6A). Our crystal structure along with MD results reveals that the 2×2 CU/UC loop conformations are

dynamic, which could be exploited by a small molecule such as **K-acetyl**. Our results suggest that the binding of **K-alkyne** and modularly assembled compounds thereof is facilitated by the inherent instability and dynamic character of the 2×2 CU/UC loops.

A crystal structure of kanamycin A complexed with an A-site mimic by Francois *et al.* showed that kanamycin displaces two adenosine residues in the internal loop[57]. That is, kanamycin's ring I interacts with the 5' adenosine of the 1×2 A/AA RNA internal loop while the remainder of the aminoglycoside is inserted into the major groove (displacing the 3' loop nucleotides) and stabilized by electrostatic interactions[57]. Although this structure does not represent the dynamic nature of small molecule binding in solution, it provides clues about conformational changes that occur. Similarly, **K-acetyl** unstacks loop residues, forming a complex series of electrostatic interactions with loop residues and the RNA backbone (Figure 6A).

Binding of **K-acetyl** via the minor groove was also observed in our computational experiments, albeit with much weaker binding free energies as compared to binding through the major groove. (The lowest free energy structure that interacts through the minor groove side (set_3; Table S-9) is 29.58 kcal/mol less stable than the lowest free energy structure that interacts via the major groove side (set_22; Table S-9 & Figure 6A).) If minor groove binding does occur, it is likely that **K-alkyne**, **2K-4**, and **3K-4** bind preferentially to the regularly repeating, negatively charged electrostatic patches in the minor groove. Baranger, Zimmerman, and co-workers recently reported a triaminopyrimidine that selectively binds r(CCUG) repeats[58]. The compound's three amino groups could interact with the four carbonyls of the CU pairs in the minor groove.

We previously modified the spacing module in the **2K-4** peptoid scaffold and studied the effect on *in vitro* potency, cellular permeability, and toxicity[59]. The propylamine spacing module was replaced with methylamine, ethanolamine, isobutylamine, lysine, glycine, arginine, tyramine, 2-phenylethanamine, 2-(pyridine-2-yl)ethanamine, 2-(4-fluorophenyl)ethanamine, and 3,3-diphenylpropan-1-amine. Most substitutions decreased potency (by ~2.8-fold to >250-fold), suggesting that propylamine is optimal[59]. Interestingly, substitution with arginine or lysine increased potency by approximately 2-fold. In our model of **2K-4** binding to r(CCUG) repeats, the propyl groups point towards the major groove; substitution with arginine or lysine could allow formation of electrostatic interactions. Unfortunately, arginine spacing modules decrease cellular permeability; both arginine and lysine substitutions alter subcellular localization to mainly the cytoplasm and perinuclear region, which is unfavorable for targeting r(CCUG)^{exp} as it is largely nuclear[59]. Our model could also explain why bulky substituents negatively affect potency as they could cause a steric clash with the RNA.

Therapeutic potential of 2K-4 and 3K-4

Our modularly assembled compounds, **2K-4** and **3K-4**, significantly improve *BIN1* splicing patterns when cells are treated with as little as 5 μM and 2.5 μM compound, respectively. Thus, they have therapeutic potential. We recently reported that related compounds that target r(CUG)^{exp}, the causative agent of DM1, improve splicing defects in a mouse model[60]. It should be noted that the compound was delivered by intraperitoneal (IP) injection; thus our compound was delivered systemically. Because of the similarity of the DM1 compound to the compounds described herein for DM2 (both display aminoglycoside binding modules on a peptoid scaffold), it is likely that **2K-4** and **3K-4** could improve DM2-associated alternative splicing defects in animals

Oligonucleotides have been used to target various repeating RNAs, in particular r(CUG)^{exp} and r(CAG)^{exp}. Typically, gapmers complementary to the repeat are transfected into cells

and induce RNase H cleavage[3, 61-63]. Oligonucleotides that target r(CUG)^{exp} improve DM1-associated defects in a mouse model; however, electrostimulation was required[3, 5] and one caused muscle damage[3]. More recently, an antisense oligonucleotide was developed that binds downstream of r(CUG)^{exp} in the context of the human skeletal actin (h*ACTA1*) transgene[6]. The oligo is effective when delivered by subcutaneous injection (no electrostimulation), and improvement in DM1-associated defects was observed in transgenic mice up to 1 year post treatment[6]. These effects are traced to the catalytic nature of the oligonucleotides as they bind the r(CUG)^{exp}-containing RNA and induce cleavage of the target by RNase H; the oligo is then free to bind another mRNA[6]. Although the oligonucleotide binds a sequence in the h*ACTA1* mRNA, not a sequence in h*DMPK* (the natural context of r(CUG)^{exp} in human DM1), these studies suggest that inducing cleavage of mRNAs downstream of expanded repeats could be a therapeutic option.

CONCLUSION

Herein, we report the first crystal structure of r(CCUG) repeats, the causative agent of DM2 and an analysis of the dynamics of this structure by MD simulations. Also, we described the first known compounds that modulate toxicity of r(CCUG)^{exp} in cellular models. It is notable that these compounds were designed via a bottom-up versus traditional routes. Moreover, we investigated potential binding modes of the RNA-binding module computationally. Modeling of the binding of **2K-4** to r(CCUG) repeats indicated that the distance between **K** binding modules closely mimics the distance between 5'CU/3'UC internal loops. These studies could aid DM2 therapeutic development. In general, these studies suggest that the bottom-up design strategy may prove useful to exploit the myriad of other RNA targets in the transcriptome.

METHODS

Cell Culture & Assaying Alternative Pre-mRNA Splicing

The C2C12 cell line was maintained as monolayers at 37 °C / 5% CO_2 in 1X DMEM supplemented with 10% (v/v) FBS and 1X Glutamax (Invitrogen). Cells were plated in 96-well plates and transfected with Lipofectamine 2000 (Invitrogen). Each transfection contained 1 □L of lipofectamine and 200 ng of plasmid (100 ng of a plasmid encoding r(CCUG)$_{300}$ (or empty vector) and 100 ng of a plasmid encoding a mini-gene that reports on alternative splicing). Approximately 5 h post-transfection, the compound of interest was added in 1X DMEM supplemented with 2% (v/v) horse serum and 1X Glutamax. Total RNA was harvested 48 h later using a GenElute Mammalian Total RNA Mini-prep Kit (Sigma) including an on-column DNase digestion.

Splicing patterns were assayed by RT-PCR as previously described. Splicing isoforms were separated on a 3% (w/v) agarose gel stained with SYBR Gold (Invitrogen) and imaged with a Molecular Dynamics Typhoons phosphorimager. Bands were quantified using BioRad's QuantityOne software.

RT-PCR primers used were as follows. *Bin1:* forward – 5′-CATTCACCACATTGGTGTGC-3′; reverse - 5′-AAGTGATCCTAGACTAGCCGCC-3′; *SMN2* (regulated by Sam68): forward - 5' GGTGTCCACTCCCAGTTCAA-3'; reverse - GCC TCA CCA CCG TGC TGG-3'.

Statistical significance was calculated using a two-tailed student t-test. Treated samples were compared to untreated samples derived from cells that express r(CCUG)^{exp} as indicated in Figure 2 with "*" (p<0.05) or "**" (p<0.01).

Bio-Layer Interferometry (BLI)

The binding of small molecules and MBNL1 to r(CCUG) repeats was characterized by BLI using a ForteBio Octet RED system. Briefly, 5'-biotinylated $(CCUG)_{12}$ was folded in 1X Kinetics Buffer (Pall Life Sciences) by heated to 95 °C for 3 min followed by slowly cooling to room temperature. Streptavidin SA biosensors (Pall Life Sciences) were used to monitor the binding of compounds or protein to the RNA. Seven concentrations of compound or protein were studied and their binding curves were fit to a 1:1 model (Octet Data Analysis software) to afford k_{on}, k_{off}, and K_d.

Crystallization, Data Collection and Refinement

The RNA containing three 5'CCUG/3'GUCC was screened for crystallization against the Nucleix suite (Qiagen) in 96-well hanging drop plates using a Gryphon crystallization robot (Art Robinsons) at room temperature. Crystals with approximate sizes of 0.1 × 0.05 × 0.05 mm appeared from multiple conditions. However, consistent reproducible crystals were obtained only from precipitants containing 50 mM MES, pH 6.0, 100 mM ammonium acetate, 5 mM Mg sulfate, and 600 mM NaCl. Although the sizes of the crystals were reasonable, the crystals did not diffract well. Improvement of diffraction was achieved by dehydration of the crystals using a free mounting system (Proteros).

A diffraction dataset with Bragg spacing of 2.3 Å was obtained on a PILATUS detector at beam line 11-1 of SLAC, SSRL. The dataset was processed with MOSFLM. The structure of the RNA was determined by molecular replacement using Phaser[64]. The core tetraloop and tetraloop receptor domain of the PDB ID 4FNJ[35] was used as the search model. Refinement was performed in the Phenix suite[65]. Data processing and refinement statistics are given in Table S-1. Figures were prepared with PyMol[66], and RNA helical parameters were calculated with Curves+[67]. Electrostatic charge distribution was calculated with APBS[>68] and visualized in PyMol[66]. The structure of the r(CCUG) repeats was deposited in the Protein Databank, PDB ID 4K27.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the Muscular Dystrophy Association (Grant ID# 254929 to MDD), the National Institutes of Health (R01 GM079235 to MDD), the Scripps Research Institute, and the PS-OC Center of the NIH/NCI Grant 1U54CA143869-01 (to IY and GCS).

REFERENCES

1. Two new drugs for homozygous familial hyperchelesterolemia. Med. Lett. Drugs Ther. 2013; 55:25–27. [No authors listed]. [PubMed: 23545581]
2. Geary RS, Henry SP, Grillone LR. Fomivirsen: clinical pharmacology and potential drug interactions. Clin. Pharmacokinet. 2002; 41:255–260. [PubMed: 11978144]
3. Lee JE, Bennett CF, Cooper TA. RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:4221–4226. [PubMed: 22371589]
4. Mulders SA, van den Broek WJ, Wheeler TM, Croes HJ, van Kuik-Romeijn P, de Kimpe SJ, Furling D, Platenburg GJ, Gourdon G, Thornton CA, Wieringa B, Wansink DG. Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:13915–13920. [PubMed: 19667189]

5. Wheeler TM, Sobczak K, Lueck JD, Osborne RJ, Lin X, Dirksen RT, Thornton CA. Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. Science. 2009; 325:336–339. [PubMed: 19608921]
6. Wheeler TM, Leger AJ, Pandey SK, MacLeod AR, Nakamori M, Cheng SH, Wentworth BM, Bennett CF, Thornton CA. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature. 2012; 488:111–115. [PubMed: 22859208]
7. Alvarez-Salas LM. Nucleic acids as therapeutic agents. Curr. Top. Med. Chem. 2008; 8:1379–1404. [PubMed: 18991725]
8. Lebleu B, Robbins I, Bastide L, Vives E, Gee JE. Pharmacokinetics of oligonucleotides in cell culture. Ciba Found. Symp. 1997; 209:47–59. [PubMed: 9383568]
9. Nesterova M, Cho-Chung YS. Killing the messenger: antisense DNA and siRNA. Curr. Drug Targets. 2004; 5:683–689. [PubMed: 15578949]
10. Thomas JR, Hergenrother PJ. Targeting RNA with small molecules. Chem. Rev. 2008; 108:1171–1224. [PubMed: 18361529]
11. Guan L, Disney MD. Recent advances in developing small molecules targeting RNA. ACS Chem. Biol. 2012; 7:73–86. [PubMed: 22185671]
12. Gallego J, Varani G. Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges. Acc. Chem. Res. 2001; 34:836–843. [PubMed: 11601968]
13. Yoshizawa S, Fourmy D, Puglisi JD. Recognition of the codonanticodon helix by ribosomal RNA. Science. 1999; 285:1722–1725. [PubMed: 10481006]
14. 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. [PubMed: 11014183]
15. Borovinskaya MA, Pai RD, Zhang W, Schuwirth BS, Holton JM, Hirokawa G, Kaji H, Kaji A, Cate JH. Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. Nat. Struct. Mol. Biol. 2007; 14:727–732. [PubMed: 17660832]
16. Moazed D, Noller HF. Interaction of antibiotics with functional sites in 16S ribosomal-RNA. Nature. 1987; 327:389–394. [PubMed: 2953976]
17. Swayze EE, Jefferson EA, Sannes-Lowery KA, Blyn LB, Risen LM, Arakawa S, Osgood SA, Hofstadler SA, Griffey RH. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. J. Med. Chem. 2002; 45:3816–3819. [PubMed: 12190303]
18. Seth PP, Miyaji A, Jefferson EA, Sannes-Lowery KA, Osgood SA, Propp SS, Ranken R, Massire C, Sampath R, Ecker DJ, Swayze EE, Griffey RH. SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. J. Med. Chem. 2005; 48:7099–7102. [PubMed: 16279767]
19. Rogers JT, Mikkilineni S, Cantuti-Castelvetri I, Smith DH, Huang X, Bandyopadhyay S, Cahill CM, Maccecchini ML, Lahiri DK, Greig NH. The alpha-synuclein 5′untranslated region targeted translation blockers: anti-alpha synuclein efficacy of cardiac glycosides and Posiphen. J. Neural Transm. 2011; 118:493–507. [PubMed: 21221670]
20. Chapman RL, Stanley TB, Hazen R, Garvey EP. Small molecule modulators of HIV Rev/Rev response element interaction identified by random screening. Antiviral Res. 2002; 54:149–162. [PubMed: 12062388]
21. Payton S, Cahill CM, Randall JD, Gullans SR, Rogers JT. Drug discovery targeted to the Alzheimer's APP mRNA 5′-untranslated region: the action of paroxetine and dimercaptopropanol. J. Mol. Neurosci. 2003; 20:267–275. [PubMed: 14501007]
22. Childs-Disney JL, Wu M, Pushechnikov A, Aminova O, Disney MD. A small molecule microarray platform to select RNA internal loop-ligand interactions. ACS Chem. Biol. 2007; 2:745–754. [PubMed: 17975888]
23. Disney MD, Labuda LP, Paul DJ, Poplawski SG, Pushechnikov A, Tran T, Velagapudi SP, Wu M, Childs-Disney JL. Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. J. Am. Chem. Soc. 2008; 130:11185–11194. [PubMed: 18652457]
24. Disney MD, Childs-Disney JL. Using selection to identify and chemical microarray to study the RNA internal loops recognized by 6′-N-acylated kanamycin A. Chembiochem. 2007; 8:649–656. [PubMed: 17394189]

25. Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001; 293:864–867. [PubMed: 11486088]
26. Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, Brook JD. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. Hum. Mol. Genet. 2002; 11:805–814. [PubMed: 11929853]
27. Mankodi A, Urbinati CR, Yuan QP, Moxley RT, Sansone V, Krym M, Henderson D, Schalling M, Swanson MS, Thornton CA. Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. Hum. Mol. Genet. 2001; 10:2165–2170. [PubMed: 11590133]
28. Lee MM, Pushechnikov A, Disney MD. Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. ACS Chem. Biol. 2009; 4:345–355. [PubMed: 19348464]
29. 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. [PubMed: 19805260]
30. Daldrop P, Reyes FE, Robinson DA, Hammond CM, Lilley DM, Batey RT, Brenk R. Novel ligands for a purine riboswitch discovered by RNA-ligand docking. Chem. Biol. 2011; 18:324–335. [PubMed: 21439477]
31. Fugier C, Klein AF, Hammer C, Vassilopoulos S, Ivarsson Y, Toussaint A, Tosch V, Vignaud A, Ferry A, Messaddeq N, Kokunai Y, Tsuburaya R, de la Grange P, Dembele D, Francois V, Precigout G, Boulade-Ladame C, Hummel MC, de Munain AL, Sergeant N, Laquerriere A, Thibault C, Deryckere F, Auboeuf D, Garcia L, Zimmermann P, Udd B, Schoser B, Takahashi MP, Nishino I, Bassez G, Laporte J, Furling D, Charlet-Berguerand N. Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. Nat. Med. 2011; 17:720–725. [PubMed: 21623381]
32. Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, Hagerman PJ, Charlet-Berguerand N. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J. 2010; 29:1248–1261. [PubMed: 20186122]
33. Fu Y, Ramisetty SR, Hussain N, Baranger AM. MBNL1-RNA recognition: contributions of MBNL1 sequence and RNA conformation. Chembiochem. 2012; 13:112–119. [PubMed: 22106026]
34. Teplova M, Patel DJ. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. Nat. Struct. Mol. Biol. 2008; 15:1343–1351. [PubMed: 19043415]
35. 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. [PubMed: 23025897]
36. Ferre-D'Amare AR, Zhou K, Doudna JA. A general module for RNA crystallization. J. Mol. Biol. 1998; 279:621–631. [PubMed: 9641982]
37. Holbrook SR, Cheong C, Tinoco I, Jr. Kim SH. Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. Nature. 1991; 353:579–581. [PubMed: 1922368]
38. Cruse WB, Saludjian P, Biala E, Strazewski P, Prange T, Kennard O. Structure of a mispaired RNA double helix at 1.6-A resolution and implications for the prediction of RNA secondary structure. Proc. Natl. Acad. Sci. U. S. A. 1994; 91:4160–4164. [PubMed: 7514296]
39. Schneider C, Brandl M, Suhnel J. Molecular dynamics simulation reveals conformational switching of water-mediated uracilcytosine base-pairs in an RNA duplex. J. Mol. Biol. 2001; 305:659–667. [PubMed: 11162082]
40. Tanaka Y, Kojima C, Yamazaki T, Kodama TS, Yasuno K, Miyashita S, Ono A, Kainosho M, Kyogoku Y. Solution structure of an RNA duplex including a C-U base pair. Biochemistry. 2000; 39:7074–7080. [PubMed: 10852704]
41. Anderson AC, O'Neil RH, Filman DJ, Frederick CA. Crystal structure of a brominated RNA helix with four mismatched base pairs: An investigation into RNA conformational variability. Biochemistry. 1999; 38:12577–12585. [PubMed: 10504226]

42. Du Z, Yu J, Ulyanov NB, Andino R, James TL. Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. Biochemistry. 2004; 43:11959–11972. [PubMed: 15379536]
43. Ohlenschlager O, Wohnert J, Bucci E, Seitz S, Hafner S, Ramachandran R, Zell R, Gorlach M. The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. Structure. 2004; 12:237–248. [PubMed: 14962384]
44. 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. [PubMed: 18593145]
45. 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.
46. Case, DA.; Darden, TA.; Cheatham, I.; T. E.; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Walker, RC.; Zhang, W.; Merz, KM.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M-J.; Cui, G.; Roe, DR.; Mathews, DH.; Seetin, NG.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, PA. AMBER 11. University of California; San Francisco: 2010.
47. 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. [PubMed: 20463845]
48. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophys. J. 2007; 92:3817–3829. [PubMed: 17351000]
49. Auffinger P, Hashem Y. SwS: a solvation web service for nucleic acids. Bioinformatics. 2007; 23:1035–1037. [PubMed: 17324939]
50. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J. Comput. Chem. 2004; 25:1157–1174. [PubMed: 15116359]
51. Childs-Disney JL, Stepniak-Konieczna E, Tran T, Yildirim I, Park H, Chen CZ, Hoskins J, Southall N, Marugan JJ, Patnaik S, Zheng W, Austin CP, Schatz GC, Sobczak K, Thornton CA, Disney MD. Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. Nat. Commun. 2013; 4:2044. [PubMed: 23806903]
52. Rastelli G, Del Rio A, Degliesposti G, Sgobba M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J. Comput. Chem. 2010; 31:797–810. [PubMed: 19569205]
53. Walter F, Vicens Q, Westhof E. Aminoglycoside-RNA interactions. Curr. Opin. Chem. Biol. 1999; 3:694–704. [PubMed: 10600721]
54. Warf MB, Berglund JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. RNA. 2007; 13:2238–2251. [PubMed: 17942744]
55. Mooers BH, 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. [PubMed: 16269545]
56. Disney MD, Lee MM, Pushechnikov A, Childs-Disney JL. The role of flexibility in the rational design of modularly assembled ligands targeting the RNAs that cause the myotonic dystrophies. Chembiochem. 2010; 11:375–382. [PubMed: 20058255]
57. Francois B, Russell RJ, Murray JB, Aboul-ela F, Masquida B, Vicens Q, Westhof E. Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. Nucleic Acids Res. 2005; 33:5677–5690. [PubMed: 16214802]
58. Wong CH, Fu Y, Ramisetty SR, Baranger AM, Zimmerman SC. Selective inhibition of MBNL1-CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2). Nucleic Acids Res. 2011; 39:8881–8890. [PubMed: 21768123]
59. Lee MM, French JM, Disney MD. Influencing uptake and localization of aminoglycoside-functionalized peptoids. Mol. Biosyst. 2011; 7:2441–2451. [PubMed: 21611644]

60. 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. [PubMed: 23130637]

61. Gonzalez-Barriga A, Mulders SA, van de Giessen J, Hooijer JD, Bijl S, van Kessel ID, van Beers J, van Deutekom JC, Fransen JA, Wieringa B, Wansink DG. Design and analysis of effects of triplet repeat oligonucleotides in cell models for myotonic dystrophy. Mol. Ther. Nucleic Acids. 2013; 2:e81. [PubMed: 23511335]

62. Evers MM, Pepers BA, van Deutekom JC, Mulders SA, den Dunnen JT, Aartsma-Rus A, van Ommen GJ, van Roon-Mom WM. Targeting several CAG expansion diseases by a single antisense oligonucleotide. PLoS One. 2011; 6:e24308. [PubMed: 21909428]

63. Gagnon KT, Pendergraff HM, Deleavey GF, Swayze EE, Potier P, Randolph J, Roesch EB, Chattopadhyaya J, Damha MJ, Bennett CF, Montaillier C, Lemaitre M, Corey DR. Allele-selective inhibition of mutant huntingtin expression with antisense oligonucleotides targeting the expanded CAG repeat. Biochemistry. 2010; 49:10166–10178. [PubMed: 21028906]

64. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. Acta Crystallogr. D Biol. Crystallogr. 2004; 60:432–438. [PubMed: 14993666]

65. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. 2002; 58:1948–1954. [PubMed: 12393927]

66. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.3r1. 2010

67. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+ Nucleic Acids Res. 2009; 37:5917–5929. [PubMed: 19625494]

68. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:10037–10041. [PubMed: 11517324]

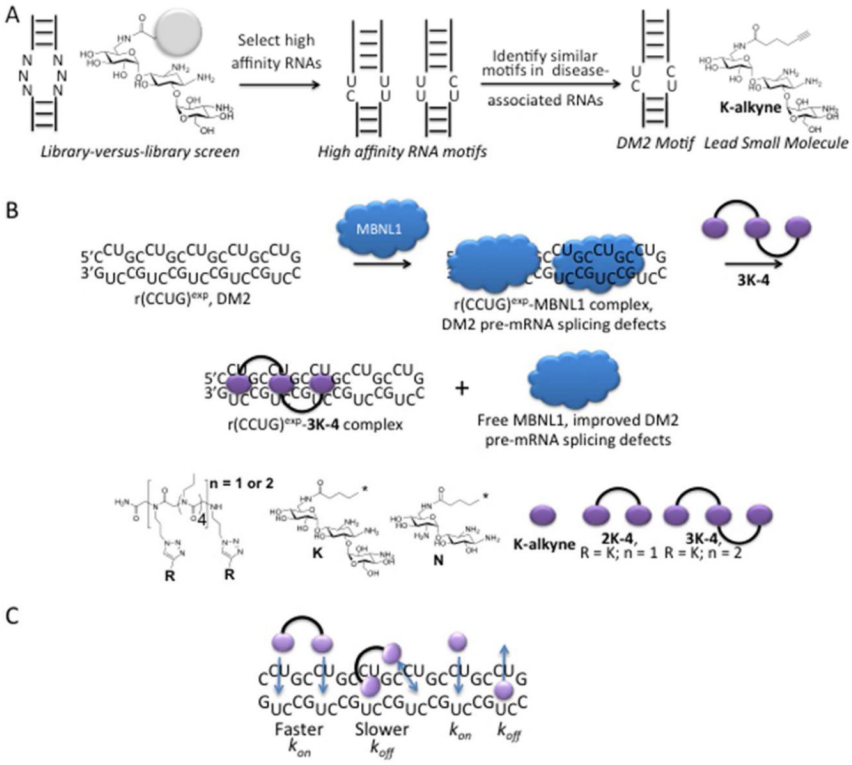

Figure 1.

The design of small molecules that reduce r(CCUG)^{exp} toxicity in cellular model systems of DM2. A, a resin-based selection completed in our laboratory determined that 2×2 nucleotide pyrimidine-rich internal loops bind to kanamycin A with the highest affinity. These loops are similar to those found in r(CCUG)^{exp}. B, DM2 is caused by an expanded r(CCUG) repeat (r(CCUG)^{exp}) found in intron 1 on the *ZNF9* pre-mRNA. r(CCUG)^{exp} binds and inactivates MBNL1, leading to dysregulation of alternative pre-mRNA splicing. The kanamycin A-like module (**K**) identified from our resin-based selection binds the repeating motif in r(CCUG)^{exp}, 5'C<u>CU</u>G/3'G<u>UC</u>C, with high affinity and selectivity. We previously showed that displaying the **K** module multiple times on a peptoid scaffold potently inhibits the r(CCUG)^{exp}-MBNL1 complex *in vitro*. Herein, we demonstrate that the optimal compound identified from *in vitro* studies, **3K-4**, improves DM2-associated defects in cellular model systems. C, the enhanced bioactivity of **3K-4** is due to a hinge effect. That is, multivalency affects k_{on} and k_{off} when binding to the RNA target.

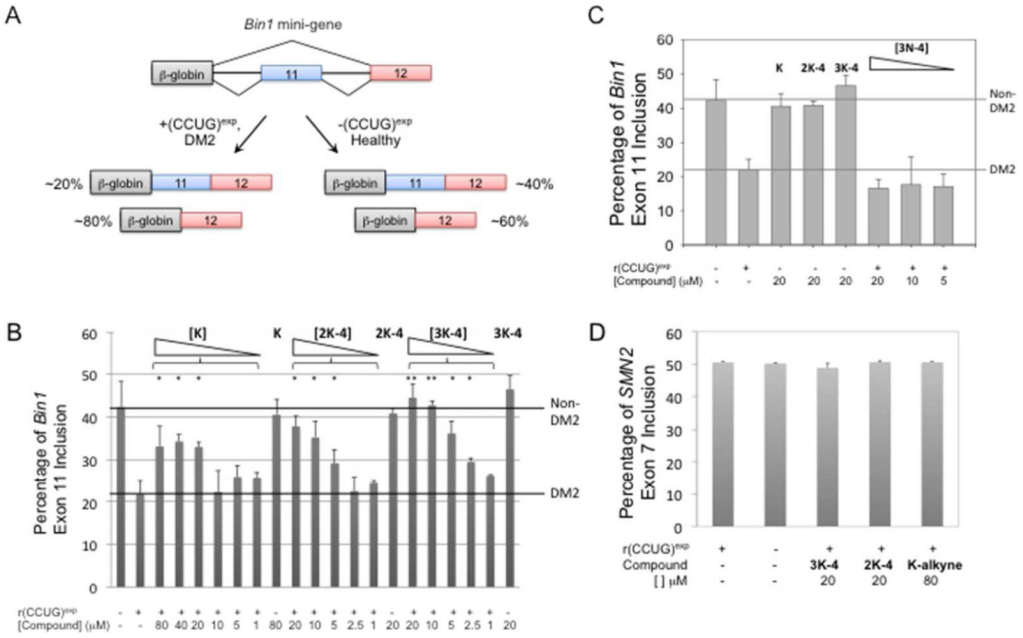

Figure 2.

A designed, modularly assembled compound improves dysregulation of alternative pre-mRNA splicing in a DM2 model cellular system. A, schematic of the *BIN1* mini-gene used to study dysregulation of pre-mRNA splicing due to r(CCUG)^{exp}. The construction of the mini-gene has been previously described[31]. B, **K**, **2K-4**, and **3K-4** improve *BIN1* alternative splicing defects to varying extents. **3K-4** is the most potent restoring *BIN1* splicing patterns to wild type at 10 M. Splicing patterns were determined by RT-PCR. C, Importantly, **K**, **2K-4**, and **3K-4** do not affect the splicing of *BIN1* pre-mRNA in the absence of r(CCUG)^{exp}. Moreover, **3N-4**, a peptoid similar to **3K-4** but displays a neamine-like (**N**) module, does not improve DM2-associated splicing defects. **3N-4** does not bind r(CCUG)^{exp} and is a poor inhibitor of the r(CCUG)^{exp}-MBNL1 complex *in vitro*[28]. D, **K**, **2K-4**, and **3K-4** do not affect the alternative splicing of *SMN2* mRNA, which is regulated by Sam68[32], indicating that the compounds exhibit specificity *in vivo*. Error bars represent the standard deviation of at least three measurements. A "*" indicates $p<0.05$ while "**" indicates $p<0.01$ as determined by a two-tailed student t-test in which the treated samples were compared to untreated samples derived from cells that express r(CCUG)^{exp}.

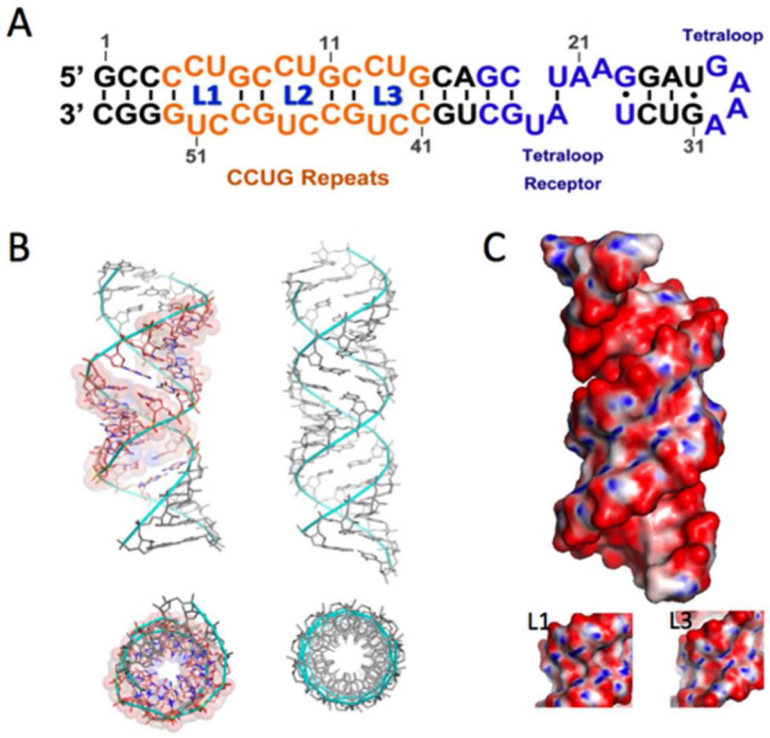

Figure 3.

Crystal structure of r(CCUG) repeat-containing RNA. (A) Secondary structure of the RNA that was subjected to crystallization. Tetraloop-tetraloop receptor elements are colored blue and CCUG repeats are colored orange with each loop numbered L1, L2, and L3. (B) Overall crystal structure of the RNA showing only the r(CCUG) repeats and closing pairs (left) and a standard A-form RNA with the same length and base pair composition except CU pairs are replaced with 5'CU/3'GA (right). The repeats are shown as orange sticks with transparent surfaces while the closing pairs are shown as grey sticks. (C) Electropotential surfaces of r(CCUG) repeats and closing pairs in the crystal structure (A). The insets show the electropotential surfaces of L1 (left) and L3 (right). The electropotential was contoured at ± 20 *KT/e*.

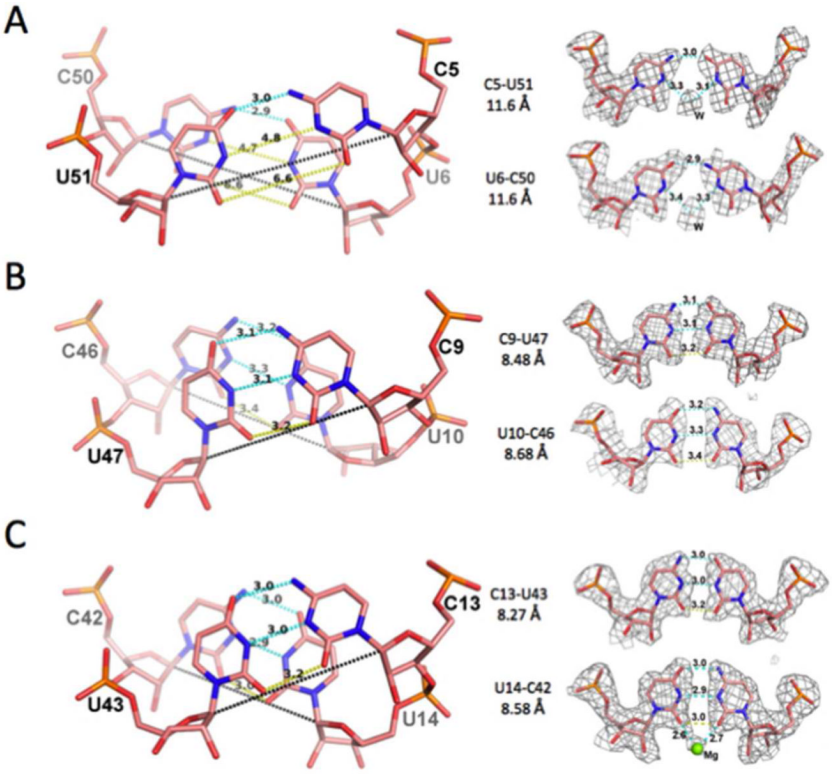

Figure 4.

Hydrogen bonding pattern and base stacking of CU pairs in the crystal structure. Interatomic bond distances (cyan dashes) and non-bonded distances (yellow dashes) between subject atoms of L1 (A), L2 (B), and L3 (C) are shown. The C1'-C1' distances (black dash lines) are indicated under each CU pair. $2F_o$-F_c electron density map covering each CU pair is contoured at 1 σ.

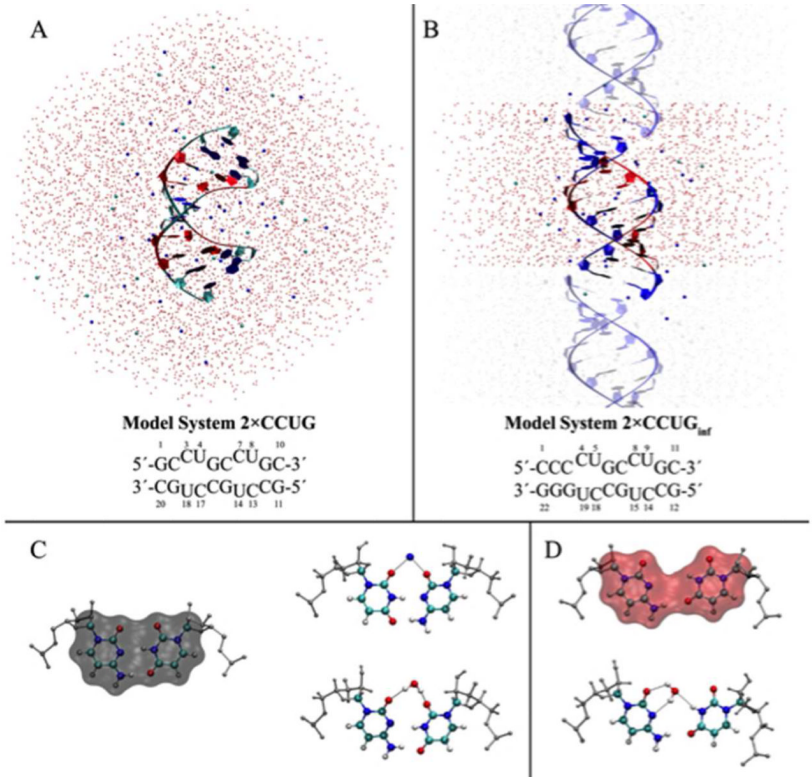

Figure 5.

Model systems (A) 2×CCUG and (B) 2×CCUG$_{inf}$ used in MD simulations, and two most frequently seen conformations, (C) CU_a and (d) CU_b, in the MD trajectories. 2×CCUG and 2×CCUG$_{inf}$ are solvated in truncated octahedral and cubic boxes, respectively. In (A) and (B), RNA molecules are shown in new ribbon representations with the CU/UC loops shown in red. Oxygen atoms of water molecules in (A) and (B) are shown as red spheres while Na^+ and Cl^- ions are shown as blue and cyan spheres, respectively. In (B), the RNA sequence and system were designed in order to study an infinitely long RNA molecule with CU/UC loops when periodicity was turned on in the MD simulations. In (B), the unit cell is shown in opaque colors while the nearest neighboring cells are shown in transparent colors to illustrate how RNA appears in the MD simulation. In (C) and (D), different colors were used to emphasize structural transformations observed in RMSD plots (Figure S-5). In (C) and (D), CU base pair conformations interacting with Na^+ and/or water molecules are shown next to the structure. All conformations of the CU pairs observed in the MD trajectories are shown in Figure S-5 (Table S-8). Note that in (C) and (D) the top and bottom parts of each CU conformation represent the minor and major grooves, respectively.

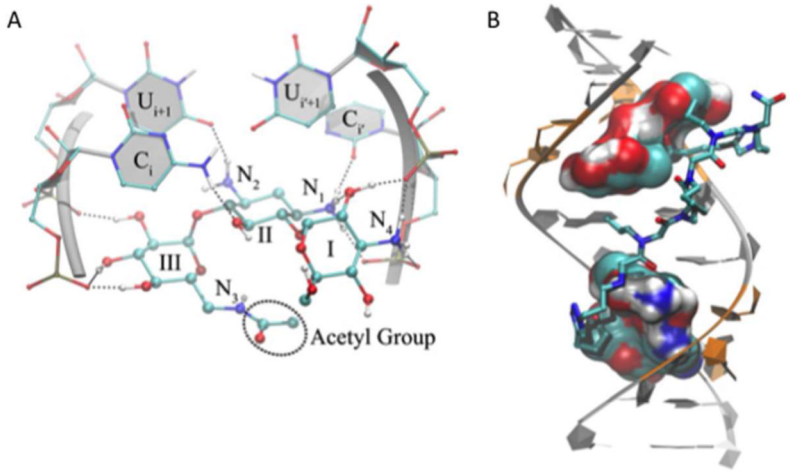

Figure 6.

Binding modes of **K-acetyl** (**K-alkyne** mimic) and **2K-4**. (A) Lowest free energy binding mode of **K-acetyl**-r(5'CCCCUGGG/3'GGGUCCCC). Only the 2×2 CU/UC loop region is displayed. The RNA molecule is shown in new ribbon and CPK representations to distinguish it from **K-acetyl**. Dashed lines represent attractive electrostatic interactions between **K-acetyl** and loop residues. (B) The binding mode of **2K-4** to 2×CCUG modeled from the lowest free energy binding mode of **K-acetyl**-5'CCUG/3'GUCC displayed in (A). RNA loop residues and Watson-Crick GC base pairs are highlighted in orange and silver, respectively. The two **K** RNA-binding modules in **2K-4** are displayed in molecular surface representations. Note that the peptoid backbone is linear in geometry.

Table 1

Kinetic analysis of small molecule-r(CCUG) complexes via SPR spectroscopy. Errors are the standard deviations in the measurements.

	k_{off} (s^{-1})	Residence Time (s)	k_{on} ($M^{-1}s^{-1}$)	K_{obs} (M)
K-alkyne	6.2 (±2.3) × 10^{-2}	1.1 (±0.04) × 10^{1}	1.4 (±0.23) × 10^{4}	4.4 (±1.8) × 10^{-6}
2K-4	2.6 (±0.76) × 10^{-3}	2.7 (±0.29) × 10^{2}	9.5 (±1.83) × 10^{4}	2.7 (±0.96) × 10^{-8}
3K-4	3.6 (±0.55) × 10^{-3}	1.9 (±0.15) × 10^{2}	7.1 (±1.3) × 10^{5}	5.1 (±1.2) × 10^{-9}
3N-4	1.5 (±0.14) × 10^{-2}	4.5 (±0.09) × 10^{1}	2.8 (±1.3) × 10^{3}	5.4 (±2.5) × 10^{-6}
MBNL1	1.3 (±0.36) × 10^{-2}	5.5 (±0.27) × 10^{1}	1.3 (±0.75) × 10^{5}	1.0 (±0.64) × 10^{-7}