6380–6390 *Nucleic Acids Research, 2012, Vol. 40, No. 13* *Published online 6 April 2012*
doi:10.1093/nar/gks298

From dynamic combinatorial 'hit' to lead: *in vitro* and *in vivo* activity of compounds targeting the pathogenic RNAs that cause myotonic dystrophy

Leslie O. Ofori[1], Jason Hoskins[2], Masayuki Nakamori[2], Charles A. Thornton[2] and Benjamin L. Miller[3,*]

[1]Department of Chemistry, [2]Department of Medicine and [3]Department of Dermatology, University of Rochester, Rochester, NY 14642, USA

Received December 5, 2011; Revised March 19, 2012; Accepted March 21, 2012

ABSTRACT

The myotonic dystrophies (DM) are human diseases in which the accumulation of toxic RNA (CUG or CCUG) repeats in the cell causes sequestration of splicing factors, including MBNL1, leading to clinical symptoms such as muscle wasting and myotonia. We previously used Dynamic Combinatorial Chemistry to identify the first compounds known to inhibit (CUG)-MBNL1 binding *in vitro*. We now report transformation of those compounds into structures with activity *in vivo*. Introduction of a benzo[g]quinoline substructure previously unknown in the context of RNA recognition, as well as other modifications, provided several molecules with enhanced binding properties, including compounds with strong selectivity for CUG repeats over CAG repeats or CAG–CUG duplex RNA. Compounds readily penetrate cells, and improve luciferase activity in a mouse myoblast assay in which enzyme function is coupled to a release of nuclear CUG–RNA retention. Most importantly, two compounds are able to partially restore splicing in a mouse model of DM1.

INTRODUCTION

High affinity, sequence-selective recognition of RNA by synthetic molecules is increasingly recognized as a key strategic goal for the production of novel therapeutics and biochemical probes (1). The importance of this area is accentuated by the ever-increasing pace of discovery of new RNA sequences with biochemically important (and therefore potentially biomedically important) functions. Among the many recent advances in this area is the recognition that many non-coding RNA (ncRNA) elements present in the eukaryotic RNAome play a direct role in controlling cellular processes and disease (2). Human diseases believed to have an ncRNA origin include spinocerebellar ataxia, fragile X-syndrome, diabetes mellitus, myoclonus epilepsy and the myotonic dystrophies (DM) (3). However, to date only a relatively small number of compounds have been reported that bind specific RNA sequences and elicit a desired biological response. For these reasons, expanding the pool of sequence-selective RNA-targeted synthetic molecules presents a critically important but under-examined challenge in chemical biology.

The DM are central examples of a growing family of RNA-mediated diseases (3–5). Myotonic dystrophy type 1 (DM1) is the most common form of adult-onset muscular dystrophy, affecting ~1 in 8000 people (6). An autosomal dominant inherited disease, DM1 results from a CTG repeat expansion (CTG^{exp}) in the 3′-UTR of the DM protein kinase gene (*DMPK*) on chromosome 19q. The expanded CTG is transcribed into long CUG^{exp} repeat mRNA. These RNA repeats sequester RNA-binding proteins such as the MBNL (muscleblind-like) family of splicing regulators, retaining them in the nucleus as foci. This in turn leads to misregulated alternative splicing, or spliceopathy. Myotonic dystrophy type 2 (DM2) is caused by an unstable expansion of a CCTG repeat in intron 1 of the cellular nucleic acid-binding protein gene (*CNBP*, also known as the zinc finger protein 9 gene or *ZNF9*) on chromosome 3q. Transcription produces toxic mRNA containing hundreds to thousands of $CCUG^{exp}$. Like CUG^{exp}, these are also sequestered into foci, and deplete MBNL1 protein from the affected cell (7,8). Currently, there is no pharmaceutical therapy for either DM1 or DM2. However, the molecular understanding of these diseases suggests that displacement of MBNL1 from its CUG^{exp} or $CCUG^{exp}$ binding sites constitutes an attractive strategy for developing therapies targeting DM (8,9).

*To whom correspondence should be addressed. Tel: +1 585 275 9805; Fax: +1 585 273 1346; Email: benjamin_miller@urmc.rochester.edu

Experiments in animal models support this hypothesis. For example, a morpholino antisense oligonucleotide (CAG-25) complementary to CUG^{exp} RNA was able to hybridize with the CUG^{exp} RNA *in vitro*, and displace MBNL1. Intramuscular injection of CAG-25 in a transgenic DM1 mouse model partially restored chloride channel-1 (Clcn-1) protein expression, and DM1 symptoms lessened (10). The high cost and challenging pharmacological properties of oligonucleotide-based drugs suggest, however, that alternative approaches to targeting CUG^{exp} RNA are of value.

In the absence of well-defined rules guiding the design of sequence-selective RNA-binding molecules [such as those developed by Dervan *et al.* (11) for DNA recognition], the question arises as to how one successfully addresses the binding problem. Several years ago, a number of groups, including ours, developed the concept of Dynamic Combinatorial Chemistry (DCC) as a 'rapid prototyping' method for testing binding hypotheses and facilitating the identification of compounds with novel architectures and properties (12). Since that time, DCC has evolved in many directions, demonstrating its potential as a method for the identification of receptors for small-molecule analytes, catalysts, new materials, sensors and a broad range of compounds able to bind protein and nucleic acid targets. However, to our knowledge no DCC-derived hit compound has either directly or through analog production resulted in a structure with *in vivo* activity.

In 2008, we reported the first non-nucleic acid-based compounds (**1** is a representative structure) capable of binding CUG^{exp} RNA and competitively inhibiting CUG^{exp}-MBNL1 binding *in vitro* (13). This work relied on a resin-bound form of DCC, termed RBDCC, that we developed to facilitate the identification of sequence-selective DNA (14) and RNA (13) binding compounds. Several groups have subsequently demonstrated elegant and structurally varied approaches to binding CUG^{exp} and $CCUG^{exp}$ RNA (8,15–19). This recent upsurge of interest highlights the fact that DM1 and DM2 RNAs are important therapeutic targets, as well as valuable model systems for testing hypotheses regarding the factors influencing selectivity and affinity in RNA recognition. Despite these advances, demonstration of the restoration of MBNL1 activity *in vivo* by cell-permeable, highly selective CUG^{exp} RNA binders remains an important goal.

RBDCC hit compound **1** (Figure 1) and related molecules identified in our initial work provided a useful demonstration of feasibility, and set the stage for building toward a compound that would be suitable for further evaluation in the biological context. To accomplish that goal, we anticipated that replacing the disulfide bridge with an olefin bioisostere would not have a dramatic impact on affinity, based on results from parallel efforts in our lab targeting an RNA sequence involved in regulating −1 ribosomal frameshifting in HIV (20). Since disulfides are easily reduced in the cytoplasm, replacing the disulfide with an olefin or alkane would facilitate cellular studies. Second, molecules containing hydrocarbon bridges of varied length would allow us to examine the effect of linker length and configuration on binding ability and selectivity. Third, we wished to explicitly examine the importance of the amino acid sequence order. Finally, as quinolines are known intercalators, at least in the DNA-binding context (21), we hypothesized that increasing the pi surface area of this group would enhance affinity. In this regard, we were surprised to discover that despite the vast amount of research conducted into the nucleic acid recognition properties and biological activity of acridine derivatives, including the use of several acridines in humans as antimicrobials (22) and chemotherapeutic agents (23), we are only aware of one mention of the closely related benzo[g]quinoline heterocycle (i.e. **2**, Figure 1) in the nucleic acid recognition literature (24). Thus, synthesizing and testing derivatives incorporating this moiety would constitute the first examination of this heterocycle in the RNA binding context.

MATERIALS AND METHODS

Benzo[g]quinoline **2** was synthesized by condensation of methyl acetoacetate and 3-nitro-2-naphthaldehyde (25), using a one-pot procedure originally developed by us in the context of quinoline synthesis (26). Compounds **3**–**9** were synthesized on solid phase by analogy to methods previously reported by our group (20). For compounds **10** and **11**, L-pentenyl glycine was synthesized via asymmetric alkylation of pseudoephedrine glycinamide (27). Complete synthesis procedures and compound characterization are provided in Supplementary Data.

Surface plasmon resonance (SPR) analysis

All SPR experiments were conducted using a Biacore-X instrument (Biacore, Inc.). Both flow cells (FC1 and FC2) of a research grade carboxymethyl dextran coated sensor chip (CM5, GE Healthcare) were functionalized with streptavidin following activation by EDC/NHS. Unreacted NHS-ester was deactivated with ethanolamine. Next, a known amount (response unit, RU, between 200–1000) of 5′-biotin labeled RNA (IDT Inc.) in running buffer was captured in FC2; FC1 was blocked with biotin and used as a reference. The level of RNA immobilized was limited to reduce mass transfer effects on the association phase. Binding analyses were carried out by flowing various concentrations of compound in HBS-N buffer (0.01 M HEPES, 0.15 M NaCl, pH = 7.4) at 30 or 60 µl/min over the immobilized RNA and recording the reference-subtracted sensorgrams. For kinetic experiments, dissociation was monitored for at least 200 s after which the surface was regenerated, where necessary, with a pulse of either 0.5 or 1 M NaCl, followed by buffer wash to reestablish baseline. Each compound concentration was injected twice to verify consistency. Blank subtraction was performed by subtracting the result obtained from the injection of buffer alone. Equilibrium binding constants (K_D) and kinetic rate constants (k_{on}, k_{off}) were obtained by globally fitting sensorgrams to the 1:1 Langmuir equation using Biaevaluation software (Biacore, Inc.). Use of the 1:1 Langmuir equation assumes that each binding site on the RNA is identical; this is of course an oversimplification. Reported

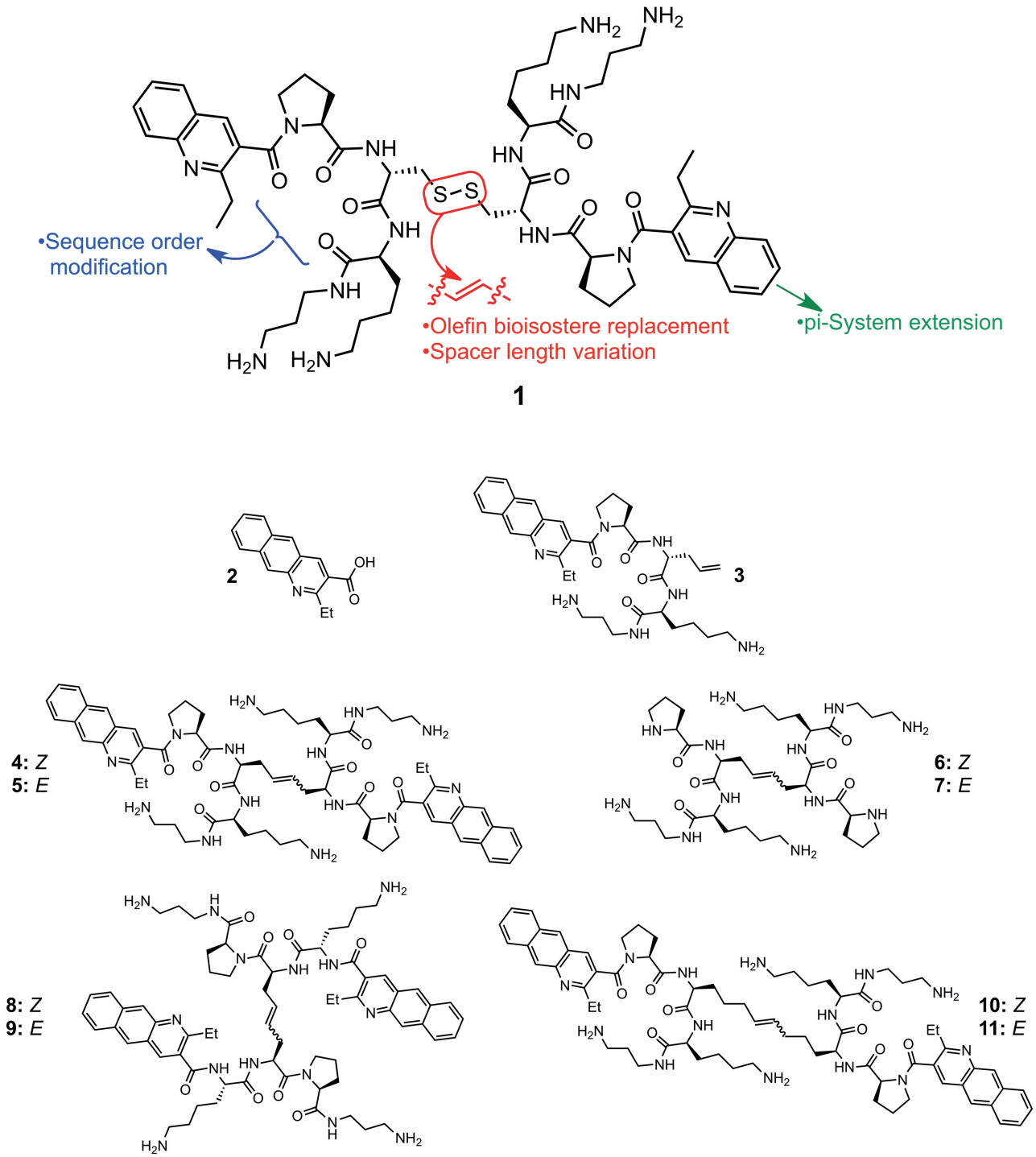


Figure 1. Hit compound **1** identified via RBDCC and molecules (**2–11**) synthesized in this work.

apparent K_D values are an average of two separate SPR experiments on different sensor chips, with different surface densities of immobilized RNA to verify that measured values were not impacted by surface RNA density. Dissociative half-lives $t_{1/2}$ were calculated from the dissociation rate constants k_d ($t_{1/2} = \ln 2/k_d$). Dissociative rates were calculated using a single-exponential decay. This methodology has been proven robust even for complexes with slower off rates than those observed in this study, and in cases where the SPR signal does not return to baseline within the measurement time frame (28). In part, this is because competing fast processes manifest as changes to early time points in the SPR dissociation curve. Stoichiometry n was assessed by repeating each SPR experiment with increasingly higher ligand concentration to saturation

(RU_{max}), followed by Scatchard analysis (r/C_{free} versus r) of the steady-state RUs. All error measurements were calculated as described in the literature (29,30) (Supplementary Methods, and Supplementary Tables S1–S44).

Fluorescence titrations

RNA (10 μM stock) in HBS-N buffer was titrated into a solution of 1 μM compound in HBS-N. After each addition of RNA, 10 min were allowed for equilibration before recording the change in fluorescence at 468.5 nm. Fluorescence units (FU) were then corrected for dilution and the FU after each addition was subtracted from the FU at zero RNA concentration to give ΔFU. Following saturation, ΔFU was plotted against RNA concentration using Origin 7 (OriginLab, Inc.), and fitted to a 1:1 Langmuir binding model to obtain apparent dissociation constants (K_D). The reported K_D are a composite fit from at least two repetitions.

Cell permeation

C2C12 mouse myoblasts grown to 80% confluence were exposed to compounds for 12 h in a 96-well tissue culture plate. After removal of the culture media [DMEM containing 10% FBS, 1% penicillin–streptomycin (GIBCO)], the cells were washed twice with PBS to remove excess compounds. Cells were then imaged while in buffer under a fluorescence microscope (Olympus IX70) in the 96-well plate using a 460-nm emission filter.

Cell toxicity

C2C12 mouse myoblasts were plated in a 96-well tissue culture plate in DMEM (10% fetal bovine serum, 1% penicillin–streptomycin) and allowed to grow to ~80% confluence at 37°C under CO_2. Varying compound concentrations (up to 1 mM) were incubated with cells (48 h, 37°C). Media was then removed, and 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) in media was added to each well and incubated at 37°C for 4 h. After removal of the MTT media, isopropanol (100 μl) was added, and absorbance was measured at 600 nm on a Modulus microplate reader (Turner Biosystems).

Activity in transformed mouse myoblasts

C2C12 mouse myoblasts were engineered to stably express a firefly luciferase transcript with or without ~800 uninterrupted CUG repeats in an *hDMPK* 3′-UTR (clones C5-14 and C1-S, respectively). Cells were grown in 100 μl growth media in 96-well plates to ~10% confluency. Compound was added at various concentrations in triplicate to both C1-S and C5-14 cultures, and incubated for 3 days. Culture media was replaced with fresh media containing 1% WST-1 reagent (Roche). After ~15–20 min of incubation, the WST-1 containing media was transferred to a clear 96-well plate, and the absorbances at 450 and 690 nm measured with a PerkinElmer EnVision Plate Reader. Cells were gently rinsed (1× PBS) before incubation at −20°C in 100 μl 1× Passive Lysis Buffer (Promega) for 10 min. Immediately before luminescence detection, 20 μl of each lysate was mixed with 50 μl of Luciferase Assay Reagent (Promega) in a fresh opaque-white 96-well plate. Luminescence values were normalized for well-to-well variations in viable cell numbers by dividing by the corresponding (A450–A690 nm) values of the WST-1 containing media.

Mouse studies

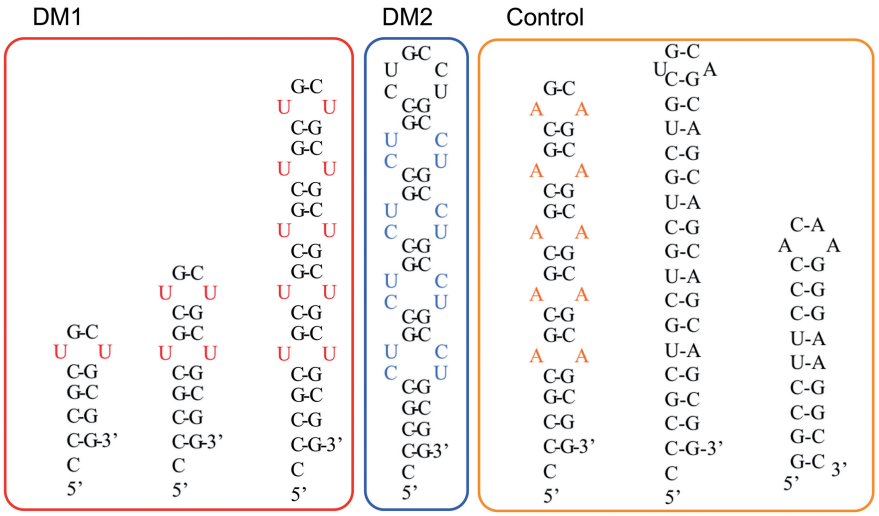
Mouse handling and experimental procedures were conducted in accordance with the Association for Assessment and Accreditation of Laboratory Animal Care. HSA^{LR} transgenic mice in line 20 b expressing human skeletal actin RNA with 220 CUG repeats in the 3′-UTR were previously described (while initially incorporating 250 CUG repeats, these have subsequently shortened) (31). Initial experiments were conducted with age- and gender-matched HSA^{LR} mice (12–16 weeks old), while follow-up experiments on compound **11** used age-matched mice, with four males and one female in each group. In each case, experimental animals were injected intraperitoneally with 40 mg/kg of compounds or saline alone once per day for 5 days. Mice were sacrificed 1 day after the last injection, and vastus muscle was obtained for splicing analysis. RNA extraction, cDNA preparation and RT-PCR were performed as described previously (31,32). Each RT-PCR reaction was performed in triplicate. The PCR products were separated on agarose gels, and scanned with a laser fluorimager (Typhoon, GE Healthcare). Quantitative analysis of amplified products was performed by ImageQuant software (Molecular Dynamics). Differences between two groups were evaluated by Student's *t*-test.

RESULTS

The binding characteristics of **3**–**11** were evaluated against pathogenic CUG^{exp} RNA and other RNA sequences by surface plasmon resonance (SPR). This technique allows the association and dissociation rates (k_a and k_d), equilibrium binding constants (K_D) and the binding stoichiometry (n) to be measured in a label-free format (33). RNA sequences, shown in Figure 2 along with their respective abbreviated names, were designed in order to test the effect of increasing numbers of CUG repeats ['$(CUG)_2$', '$(CUG)_4$' and '$(CUG)_{10}$'], differences between recognition of DM1 and DM2 RNA ['$(CCUG)_{10}$'], U–U versus A–A mismatches ['$(CAG)_{10}$'] and mismatched RNA versus a CUG–CAG duplex ('Duplex'). An RNA sequence derived from the *gag*/*pol* frameshift-stimulating region of HIV-1 ('HIV-1 FSS') (34) was employed as an off-target control.

Equilibrium affinities

Compound **3** showed no measurable affinity for $(CUG)_{10}$. In contrast, **4** and **5** bound $(CUG)_{10}$ with apparent K_D values of 39 and 68 nM, respectively, in both cases representing a roughly 50-fold improvement in affinity relative to **1** (6.7 μM, as measured by filter binding) (14). Scatchard analysis of the steady-state SPR response units (RU) for the binding between **4** or **5** and $(CUG)_{10}$



Compound	RNA Sequence						
	$(CUG)_2$	$(CUG)_4$	$(CUG)_{10}$	$(CCUG)_{10}$	$(CAG)_{10}$	Duplex	HIV-1 FSS
4	32 (n=2)	22.5 (n=4)	39 (n=10)	38 (n=10)	49 n=10	67	1240
5	21.5	38	68 (n=10)	15	55.5	307.5	759
6	ND	ND	263	21900	160.5	336.5	116
7	ND	ND	422	36100	312.5 (n=3)	ND	302.5 (n=1)
8	ND	ND	133.5	294	474.5	94	ND
9	ND	ND	120.5	138	263	77	67
10	ND	ND	81.5	278.5	ND	ND	ND
11	668	60 (n=1)	22.5 (n=5)	210	862	No binding	2000
Neomycin	ND	ND	409	1955	ND	ND	ND

Figure 2. Thermodynamic affinities (apparent K_D, nM) and selected stoichiometries (n) as measured by SPR. Binding constants are reported as an average of two complete titrations determined at differing surface RNA density; measured K_D values for each run and associated errors are provided in Supplementary Data.

yielded a binding stoichiometry of 10:1 for each isomer. This implies that one molecule of compound binds one (CUG) trinucleotide in the $(CUG)_{10}$ RNA hairpin stem and loop. The Scatchard plot for these data displays a strong convex curvature (Supplementary Data), and Hill coefficients of 1.9 (compound **4**) or 1.5 (compound **5**) were calculated, all consistent with positively cooperative binding (35,36). Both the apparent affinity and binding stoichiometry (the latter as a function of number of CUG repeats) remained consistent for **4** binding to $(CUG)_4$ and $(CUG)_2$, confirming that one molecule of **4** binds to each CUG. In contrast, the affinity of **5** actually appeared to strengthen slightly for $(CUG)_4$ and $(CUG)_2$. Reducing the number of CUG repeats in the target RNA yielded linear Scatchard plots for **4** (Supplementary Tables S3 and S4). This confirms an expected loss of cooperativity as the number of CUG repeats decreases. Likewise, calculated Hill coefficients were consistent with non-cooperative binding to short repeats. One possible explanation for this is that the longer $(CUG)_{10}$ repeat RNA has greater flexibility, and ligand binding induces a conformational change which then affects the mode of binding of subsequent ligand molecules. While **4** and **5** are strongly selective for binding CUG repeat RNAs relative to unrelated hairpins such as the HIV-1 FSS (apparent K_D 1240 and 759 nM, respectively), both $(CCUG)_{10}$ and $(CAG)_{10}$ were bound with similar affinities to the CUG repeat sequences. It is not surprising that differentiation among these sequences is particularly challenging. MBNL1 itself has been reported to bind CUG and CCUG repeats with only a 2-fold difference in affinity as measured by gel-shift assay (37). Likewise, MBNL1 was found to bind (CUG) and (CAG) repeats with similar affinity in a filter binding assay (38). The X-ray crystal structures of a $(CUG)_6$ repeat duplex and a (CAG) repeat containing RNA duplex showed that both adopt A-form RNA conformation with alternating stripes of positive and negative potential due to the G–C pairs (39–41). It has been predicted that the (CCUG) repeat RNA may also adopt a similar conformation (38). Binding of **4** to duplex CUG–CAG appears to be ~4.6-fold stronger than the *E* isomer, **5**.

To ascertain the importance of the benzo[g]quinoline moiety, we examined *cis* (**6**) and *trans* (**7**) peptides

lacking the heterocycle. While both compounds bound $(CUG)_{10}$ RNA with modest affinity, neither displayed selectivity for the CUG repeat. Likewise, benzo[g]quinoline-bearing compounds with a 'scrambled' peptide sequence (**8** and **9**) showed reduced affinity to $(CUG)_{10}$, and no selectivity for CUG or CCUG repeats relative to duplex CUG–CAG or HIV-1 FSS RNAs. Both observations suggest that a substantial fraction of the functionality in **4** and **5** must be present in the correct relative spatial orientation in order for high affinity recognition to be achieved.

A key aspect of the original RBDCC library design was that the disulfide was incorporated only to reversibly link RNA-binding modules together, rather than participating in RNA binding itself. Such a reversible linkage is a critical feature of any dynamic combinatorial library, but depending on the exchange reaction used may require subsequent re-engineering of the compound in a form not subject to exchange under physiological conditions. In this case, replacement of the labile disulfide bridge with an olefin allowed us to not only improve the biostability of the compound, but also provided an opportunity to test the effect of varying the spacing between modules on binding. This was also motivated by the complete lack of binding by monomer **3** and benzo[g]quinoline **2,** suggesting that a bidentate mode of binding stabilized by the two heterocyclic handles is required. For this hypothesis to hold, one would anticipate that changing the separation between the two halves of the ligand would alter selectivity and affinity. Thus, we synthesized 'extended' linker compounds **10** and **11**. These compounds were found to bind $(CUG)_{10}$ with an affinity similar to that displayed by **4** and **5**; however, affinity for **11** decreased in tandem with decreasing hairpin length [from an apparent K_D of 22.5 nM for $(CUG)_{10}$ to 668 nM for $(CUG)_2$]. Binding stoichiometry also changed, with 5:1 binding observed for $(CUG)_{10}$ and 1:1 binding observed for $(CUG)_4$. This more stringent structural requirement also manifested as a dramatic increase in sequence selectivity: **11** has a significant preference for $(CUG)_{10}$ over $(CCUG)_{10}$, and a 38-fold preference versus $(CAG)_{10}$. No measurable binding was observed by SPR to the duplex CUG–CAG sequence, while binding to the HIV hairpin was reduced 89-fold. As **10** was isolated as only a very minor product of the metathesis reaction, only a limited number of experiments were possible with this compound. However, SPR suggests it is also somewhat selective for $(CUG)_{10}$ over $(CCUG)_{10}$. Likewise, **10** binds $(CUG)_{10}$ with a 5:1 stoichiometry. Binding of **11** to $(CUG)_{10}$ is cooperative (Hill coefficient of 1.9). The 2-fold difference in stoichiometry for 'extended' compounds **10** and **11** relative to **4** and **5** is readily observable in the SPR trace. For example, an injection of excess **10** produces roughly half the steady-state response of the injection of an equivalent concentration of **5** (Supplementary Figure S2). These data are consistent with the requirement of a more distributed binding site. Neomycin, a well-studied aminoglycoside antibiotic with relatively low sequence selectivity, has been reported to bind (CUG) repeats (9), and represents a useful positive control from another structural class. We found that neomycin binds $(CUG)_{10}$ and $(CCUG)_{10}$ with much weaker affinity (apparent K_D of 409 and 1955 nM, respectively) than the best compounds described earlier.

Table 1. Binding constants as measured by fluorescence titration

Compound	Sequence	Apparent K_D (nM)
3	$(CUG)_{10}$	No binding
4	$(CUG)_{10}$	66 ± 2
5	$(CUG)_{10}$	70 ± 1
10	$(CUG)_{10}$	40.4 ± 0.3
11	$(CUG)_{10}$	57 ± 1
11	$(CCUG)_{10}$	73.6 ± 0.8

Values are a composite fit of at least two complete titrations ± 1 SD on the fit.

Selected equilibrium affinities for $(CUG)_{10}$ were confirmed by fluorescence titration. The benzo[g]quinoline moiety is fluorescent, with excitation and emission maxima at 362 and 439 nm, respectively, in methanol. The quantum yield of the ethylbenzo[g]quinoline carboxylic acid (**2**), relative to quinine sulfate, was determined to be 0.64 (Supplementary Figure S3). This unique feature allowed us to directly monitor RNA binding, and mammalian cell penetration and localization without the need for additional labeling. In all cases, compound binding resulted in saturable quenching of fluorescence (Supplementary Figures S4–S6). Apparent K_D values measured in this manner are consistent with those obtained by SPR (Table 1); that they are not identical to SPR-measured values likely results from the differing formats of the two methods. As a control, yeast tRNA was titrated into a mixture of **4** and **5**; no binding was observed.

Analysis of binding kinetics

While equilibrium binding constants are essential for understanding binding selectivity, they do not provide a complete picture of the recognition process. In the ligand–protein binding field, an emerging consensus holds that kinetic constants, and in particular the off-rate, are key determinants of selectivity (42–44). Similar observations have been reported for DNA-binding compounds (45–48), but to date there is little information addressing this hypothesis in the context of RNA binding (and no information with regard to how this affects compound behavior in a cellular environment or *in vivo*, where compounds must compete with other cellular constituents for target RNAs).

For SPR-derived kinetic constants, we observe that on-rates generally do not vary substantially for individual compounds across a series of RNAs. In contrast, off rates strongly reflect differences in selectivity, consistent with the above hypothesis. These data are perhaps easier to conceptualize in terms of dissociative half-life, or $t_{1/2}$, for the interaction with target sequences (Table 2). Comparisons among compounds and target sequences based on dissociative half-life are a useful addition to commonly employed metrics such as the selectivity

index, and may have some advantages since the former is an absolute measurement while the latter is defined based on a changeable set of RNAs. For the compounds studied, **11** displays the largest differences in $t_{1/2}$, consistent with its highest selectivity for CUG repeats: an ~25-fold difference between $(CUG)_{10}$ and $(CCUG)_{10}$, and an ~47-fold difference between $(CUG)_{10}$ and the HIV-1 FSS.

Prior to analyzing the ability of selected compounds to interfere with CUG^{exp} mediated effects in model cell lines, it was crucial to first establish that they are capable of crossing the cell membrane, and are non-toxic at experimentally relevant concentrations. Cell permeability was assessed in human fibroblasts, and in mouse C2C12 myoblasts. In both cases, penetration of the cell membrane was readily observed because of the fluorescence of the benzo[g]quinoline chromophore; compounds also appear to localize preferentially in the nucleus (Figure 3). Since the sequestration of MBNL1 by pathogenic CUG^{exp} RNA occurs entirely in the nucleus, this apparent selective localization even in non CUG^{exp}-containing cell lines is particularly notable. Of course, the mechanism of compound entry is not revealed by this analysis. Interestingly, benzo[g]quinoline **2** showed only a modest ability to enter cells, and 'scrambled' compound **9** was significantly less able to enter cells than **4**, **5** or **11** (Supplementary Figures S7–S10). Toxicity of **4** and **5** to mouse myoblasts was measured by MTT assay (49). No significant toxicity was observed at concentrations up to 100 μM. In contrast, mitomycin C, a commonly employed DNA-targeted cancer chemotherapeutic agent, showed significant toxicity at all concentrations tested (Figure 4). Similar results were obtained in preliminary experiments with human fibroblasts (Supplementary Figure S11). Benzo[g]quinoline **2** was also tested, and found to have no toxicity in fibroblasts at concentrations up to 500 μM (Supplementary Figure S10), an expected result given its limited ability to cross the cell membrane.

We then evaluated the ability of selected compounds to inhibit the nuclear retention of CUG^{exp} RNA, a phenomenon that depends on formation of RNA–MBNL1 complexes, in mouse myoblasts. An engineered cell line (C5-14) expressing a luciferase-encoding mRNA with 800 CUG repeats in the 3′-UTR was used. Formation of $(CUG)_{800}$ RNA–MBNL1 complexes in the C5-14 cell line suppresses translation of this mRNA, and therefore suppresses the cellular level of luciferase. Inhibition of protein complexation to the $(CUG)_{800}$ RNA allows translation to occur, restoring luciferase expression. A second cell line (C1-S) carrying an analogous luciferase construct lacking the CUG repeats in the 3′-UTR was used as a positive control for luciferase expression. After treatment with a morpholino antisense oligonucleotide (CAG-25)

Table 2. Residence time ($t_{1/2}$, s) for compound binding to RNA

Compound	$(CUG)_2$	$(CUG)_4$	$(CUG)_{10}$	$(CCUG)_{10}$	$(CAG)_{10}$	Duplex	HIV-1 FSS
4	757	1650	548	745	462	801	67
5	888	704	483	231	330	180	177
6	ND	ND	20	2	36	17	27
7	ND	ND	15	1	30	ND	35
8	ND	ND	230	253	227	256	ND
9	ND	ND	257	256	173	374	347
10	ND	ND	193	130	ND	ND	ND
11	109	806	1272	50	56	NB	27
Neomycin	ND	ND	14	4	ND	ND	ND

'NB' = no binding; 'ND' = not determined.

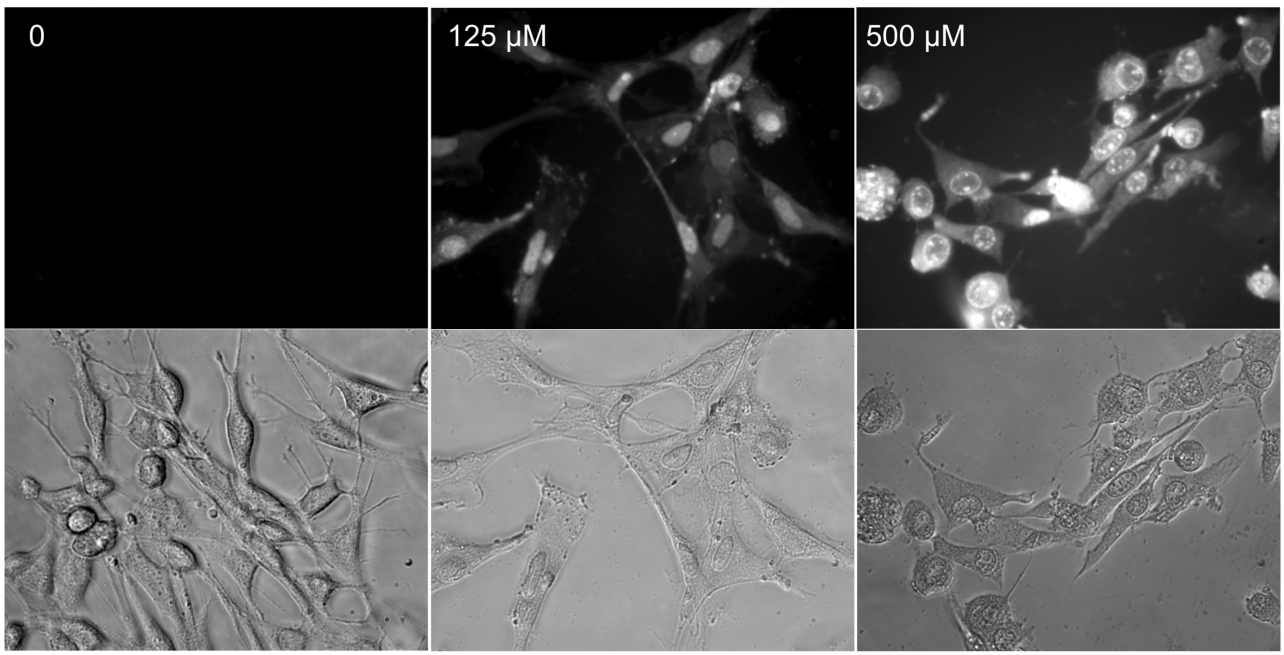


Figure 3. Compounds readily penetrate cell membranes and localize in the nucleus. Representative bright field (lower) and fluorescence (upper) images are shown for compound **5** in mouse myoblasts.

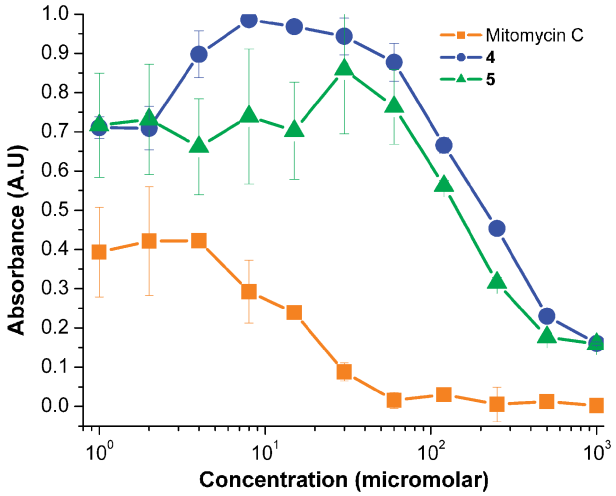


Figure 4. MTT assays indicate low toxicity of **4** and **5** in mouse myoblasts. The significantly higher toxicity of mitomycin C is shown for comparison. Error bars indicate single SDs of three replicates of each concentration.

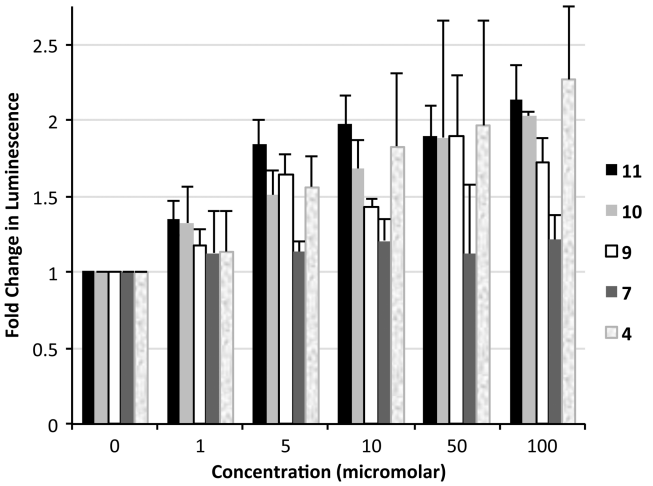


Figure 5. Compounds are able to promote expression of a $(CUG)_{800}$-containing luciferase construct in mouse myoblasts. Firefly luciferase activity is plotted as a ratio of the normalized luminescence from cells containing $(CUG)_{800}$ in the 3′-UTR of the luciferase mRNA (C5–14) to normalized luminescence of cells containing no CUG repeats (C1-S). Error bars indicate 1 SD of luminescence from the average of three replicate wells.

complementary to the CUG repeat RNA, C5-14 cells showed an increase in luciferase expression consistent with direct interaction with the $(CUG)_{800}$ mRNA in the nucleus (J. Hoskins *et al.*, manuscript in preparation). The increase in luciferase activity is accompanied by a disruption of CUG foci, as seen by fluorescence in-situ hybridization (FISH). We hypothesized that binding of our compounds to the $(CUG)_{800}$ repeat in C5-14 cells would likewise promote release of the luciferase transcripts from the nucleus, resulting in increased luciferase expression and activity. In contrast, luciferase activity in positive control (C1-S) cells should not be affected by the presence of CUG RNA binding compounds. Indeed, incubation of various concentrations of **4**, **10**, or **11** (0–100 μM) with C5-14 myoblasts resulted in concentration-dependent increases in luciferase activity, consistent with our hypothesis (Figure 5). Compound **7** (lacking selectivity to CUG^{exp} RNA) had no effect on luciferase activity when incubated with C5-14 cells under the same conditions. No statistically significant change in luciferase signal was observed for C1-S cells (which lack CUG repeats) treated with compounds **10** or **11**; however, we observed a concentration-dependent decrease in luciferase activity when these cells were incubated with **4**. Since the WST-1 viability results showed no observable toxicity of **4** to the C1-S cells at the assay concentrations, we speculated that this effect could result from an off-target effect of compound **4** causing a global decrease in protein synthesis. The concentration of total protein in cell lysates of both C5-14 and C1-S cells, as measured by Bradford assay (Supplementary Figures S12 and S13), indeed showed a decreasing level of protein expression with increasing concentration of **4**. The fact that this effect is not observed for **10** and **11** likely results from their greater selectivity for CUG repeats. Thus, CUG^{exp} binding compounds are capable of releasing the sequestered mRNA transcripts into the cytoplasm for translation.

Compounds partially restore splicing in a mouse model of DM1

Of course, a critical test of a compound's performance is its ability to function *in vivo*. DM1 has several well-studied mouse models (31). We examined the activity of compounds **4**, **9** and **11** in the HSA^{LR} mouse model, in which transgenic mice carry a long CTG repeat inserted into the human skeletal actin (*HSA*) gene in skeletal muscle (50). These mice exhibit several DM1-like phenotypic characteristics. They also display misregulated alternative splicing of genes, including *Clcn1* and *Atp2a1*, which are physiological targets of MBNL1 regulation. Not all compounds were tested due both to the complexity and cost of mouse studies, and due to the limited availability of compound **10**.

To examine effects on MBNL1-regulated splicing events we administered compounds **4**, **9** or **11** to HSA^{LR} transgenic mice by daily intraperitoneal injection of 40 mg/kg for 5 days. In the initial experiment, analysis of splicing in hind limb (quadriceps) muscle showed that compound **4** produced a modest but statistically significant ($P = 0.0379$) improvement of *Clcn1* splicing relative to saline-injected controls, whereas compound **9** did not (Figure 6). Statistically significant improvements in both *Clcn1* and *Atp2a1* ($P = 0.0419$ and $P = 0.0223$, respectively; Supplementary Figure S14) were observed following treatment with compound **11**. To confirm these results, we carried out a further study in which five age-matched mice were treated with compound **11** via interperitoneal injection at 40 mg/kg once daily for 5 days, and compared with five age-matched controls. Once again, **11** provided a statistically significant improvement in *Atp2a1* and *Clcn1* splicing. *Ttn*, another transcript dependent on MBNL1-mediated splicing, also showed improvement (Figure 7).

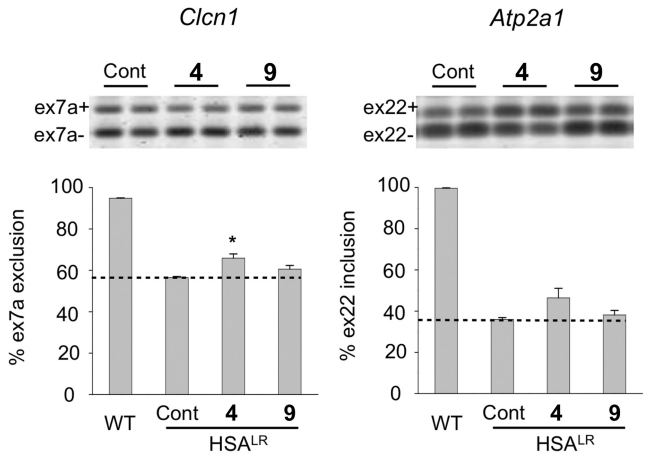


Figure 6. Compound **4** improves MBNL1-dependent splicing *in vivo*. Asterisk (*) indicates statistical significance ($P = 0.0379$); the dashed line is provided as a reference to splicing levels for HSA^{LR} mice ($n = 3$ for experimental, control and wild-type) in the absence of compound (control). The error bars indicate SEM.

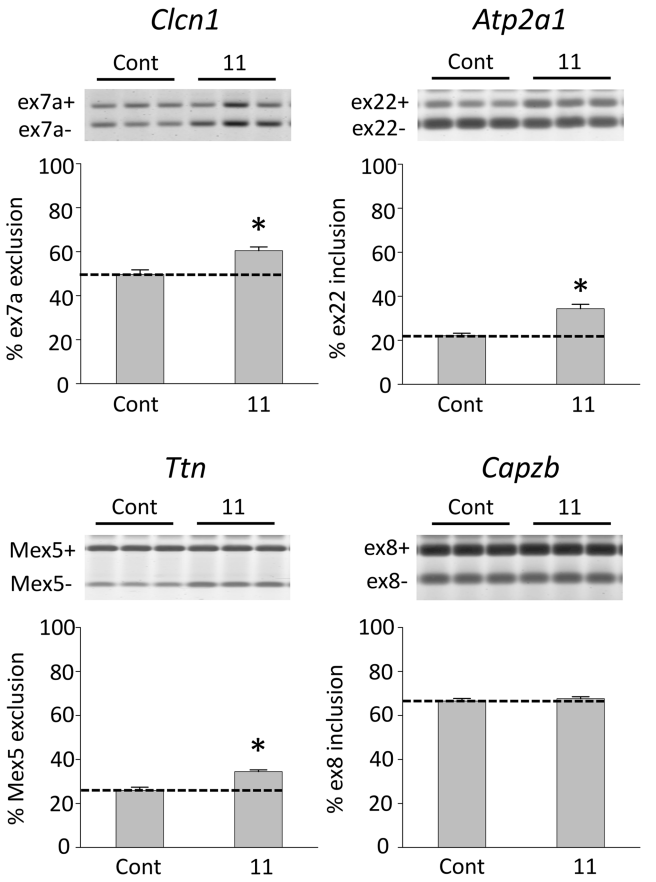


Figure 7. Compound **11** improves MBNL1-dependent splicing *in vivo*. Asterisks (*) indicate statistical significance ($P < 0.005$); the dashed line is provided as a reference to splicing levels for HSA^{LR} mice in the absence of compound ($n = 5$ for compound-treated and control). The level of an MBNL1-independent splicing event (*Capzb*) is unaffected by **11**. Error bars indicate SEM.

Splicing of *Capzb*, a developmentally regulated exon whose splicing does not depend on MBNL1, was not affected.

DISCUSSION

Using a moderate-affinity ligand identified from a resin-bound dynamic combinatorial library as a starting point, we have successfully developed next-generation lead compounds able to selectively bind DM1 and DM2 RNA with high affinity. These compounds represent to our knowledge the first use of the benzo[g]quinoline moiety in an RNA-binding context, an important advance in that this substructure allows for direct visualization of compounds in cells via fluorescence. The strong selectivity of **11** for CUG repeats is particularly notable, as is its enhanced affinity for longer CUG repeat sequences. Selectivity for longer repeats is potentially highly advantageous: since isolated and short CUG repeat sequences are found throughout the transcriptome, binders must differentiate between these and longer repeats in order to avoid off-target effects *in vivo*.

As stated at the outset, our inclusion of the benzo[g]quinoline moiety in second-generation compounds was predicated on the assumption that binding of lead compounds to CUG repeats involved intercalation of the quinoline moieties. Therefore, the stoichiometries for **4** and **5** binding these repeats as determined by SPR were surprising. In each case, binding of 1 molecule of **4** or **5** per CUG via a mode in which both benzo[g]quinolines intercalate would require violation of the neighbor-exclusion rule (51). Further studies will be essential to fully understanding the structural parameters of binding by these two compounds, as well as the others described in this article.

Several of the molecules synthesized show low toxicity in human fibroblast and mouse myoblast cell lines, suggesting the high sequence selectivity displayed by these structures is transferable to a biological context. Compounds examined in this work are able to release the nuclear retention of a CUG repeat-containing transcript in a concentration-dependent manner in mouse myoblasts, and improve splicing of MBNL1-dependent targets in a mouse model of DM1. Thus, we have demonstrated for the first time that DCC can serve as a starting point from which to develop high-affinity sequence-selective RNA binding compounds with desirable biological activity *in vitro* (in cell culture) and *in vivo* (in mice). Activity *in vivo* was observed to correlate with *in vitro* potency. Interestingly, we also observed that mice treated with **4** exhibited significant acute toxicity effects while those treated with **11** did not. While the higher (CUG^{exp}) sequence selectivity of **11** may contribute to its lower toxicity, further studies will be essential to understand these differences in detail. The amount of splicing improvement following treatment with **4** or **11** following daily interperitoneal injection at 40 mg/kg for 5 days is similar to that produced by pentamidine (8) at a dosage of 40 mg/kg once daily for 7 days, but we anticipate the greater selectivity of the compounds, we describe

(particularly **11**) may provide a more favorable pathway for future development.

In order to build on these initial successes, it will be necessary to increase the amount of compound-dependent splicing improvement observed in mouse models of DM1. It is possible that this goal can be at least partially accomplished by modification of the dosage and/or dosing schedule. Further improvement of RNA binding affinity, bioavailability and stability will also likely prove essential. Medicinal chemistry efforts based on compound **11** along those lines are currently in progress, as we continue to make strides toward the development of a viable anti-DM1 therapy.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–44, Supplementary Figures 1–14 and Supplementary Methods.

ACKNOWLEDGEMENTS

We thank Matthew Baker and Donald Curran for assistance with preliminary studies on the synthesis of benzo[g]quinoline **2**, and Prof. Brian R. McNaughton for helpful discussions. L.O.O. and B.L.M. designed all compounds. L.O.O. synthesized compounds **2–11**, carried out all SPR and fluorescence binding assays, carried out all MTT and cell permeability experiments and completed all assays in mouse myoblasts (C1-S and C5-14 cells) in tandem with J.H. M.N. carried out all *in vivo* mouse studies. Data analysis was conducted by all authors and all authors contributed to writing the manuscript.

FUNDING

National Institutes of Health (NINDS) [5R21NS071023, AR049077, U54NS48843]. Funding for open access charge: NIH – NINDS [5R21NS071023].

Conflict of interest statement. None declared.

REFERENCES

1. Thomas,J.R. and Hergenrother,P.J. (2008) Targeting RNA with small molecules. *Chem. Rev.*, **108**, 1171–1224.
2. Wahlestedt,C. (2006) Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug Discov. Today*, **11**, 503–508.
3. Gatchel,J.R. and Zoghbi,H.Y. (2005) Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.*, **6**, 743–755.
4. Todd,P.K. and Paulson,H.L. (2010) RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.*, **67**, 291–300.
5. Osborne,R.J. and Thornton,C.A. (2006) RNA-dominant diseases. *Hum. Mol. Genet.*, **1B5**, 162–169.
6. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P. and Hudson,T. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.
7. Machuca-Tzili,L., Brook,D. and Hilton-Jones,D. (2005) Clinical and molecular aspects of the myotonic dystrophies: a review. *Muscle Nerve*, **32**, 1–18.
8. Warf,M.B., Nakamori,M., Matthys,C.M., Thornton,C.A. and Berglund,J.A. (2009) Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl Acad. Sci. USA*, **106**, 18551–18556.
9. Wheeler,T. (2008) Myotonic dystrophy: therapeutic strategies for the future. *Neurotherapeutics*, **5**, 592–600.
10. Wheeler,T.M., Sobczak,K., Lueck,J.D., Osborne,R.J., Lin,X., Dirksen,R.T. and Thornton,C.A. (2009) Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science*, **325**, 336–339.
11. Dervan,P.B. and Bürli,R.W. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.*, **3**, 688–693.
12. Corbett,P.T., Leclaire,J., Vial,L., West,K.R., Wietor,J.L., Sanders,J.K. and Otto,S. (2006) Dynamic combinatorial chemistry. *Chem. Rev.*, **106**, 3652–3711.
13. Gareiss,P.C., McNaughton,B.R., Sobczak,K., Thornton,C.A. and Miller,B.L. (2008) Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA−MBNL1 interaction in vitro: discovery of lead compounds targeting myotonic dystrophy (DM1). *J. Am. Chem. Soc.*, **130**, 16254–16261.
14. McNaughton,B.R. and Miller,B.L. (2006) Resin-bound dynamic combinatorial chemistry. *Org. Lett.*, **8**, 1803–1806.
15. García-López,A., Llamusí,B., Orzáez,M., Pérez-Payá,E. and Artero,R.D. (2011) In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models. *Proc. Natl Acad. Sci. USA*, **108**, 11866–11871.
16. Arambula,J.F., Ramisetty,S.R., Baranger,A.M. and Zimmerman,S.C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl Acad. Sci. USA*, **106**, 16068–16073.
17. Wong,C.H., Fu,Y., Ramisetty,S.R., Baranger,A.M. and Zimmerman,S.C. (2011) Selective inhibition of MBNL1-CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2). *Nucleic Acids Res.*, **39**, 8881–8890.
18. Lee,M.M., Pushechnikov,A. and Disney,M.D. (2009) Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem. Biol.*, **4**, 345–355.
19. Lee,M.M., Childs-Disney,J.L., Pushechnikov,A., French,J.M., Sobczak,K., Thornton,C.A. and Disney,M.D. (2009) Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: targeting the RNAs that cause myotonic muscular dystrophy. *J. Am. Chem. Soc.*, **131**, 17464–17472.
20. Palde,P.B., Ofori,L.O., Gareiss,P.C., Lerea,J. and Miller,B.L. (2010) Strategies for recognition of stem-loop RNA structures by synthetic ligands: application to the HIV-1 frameshift stimulatory sequence. *J. Med. Chem.*, **53**, 6018–6027.
21. Varvaresou,A. and Iakovou,K. (2011) Molecular modeling study of intercalation complexes of tricyclic carboxamides with d(CCG GCGCCGG)$_2$ and d(CGCGAATTCGCG)$_2$. *J. Mol. Model.*, **17**, 2041–2050.
22. Wainwright,M. (2001) Acridine-a neglected antibacterial chromophore. *J. Antimicrob. Chemother.*, **47**, 1–13.
23. Horstmann,M.A., Hassenpflug,W.A., zur Stadt,U., Escherich,G., Janka,G. and Kabisch,H. (2005) Amsacrine combined with etoposide and high-dose methylprednisolone as salvage therapy in acute lymphoblastic leukemia in children. *Haematologica*, **90**, 1701–1703.
24. Petrov,O.E., Soprunova,N.Y., Binh,L.N., Kozyreva,N.P. and Bekhli,A.F. (1982) In vitro interaction of chick erythrocyte DNA with substituted 4-amino benzo[g]quinolines. *Khim.-Farmat. Zhurnal*, **16**, 1304–1306.
25. Kienzle,F. (1980) The reaction of phthalaldehydes with 3-nitropropionates A simple route to 3-nitro-2-naphthoic acids. *Helv. Chim. Acta*, **63**, 2364–2369.
26. McNaughton,B.R. and Miller,B.L. (2003) A mild and efficient one-step synthesis of quinolines. *Org. Lett.*, **5**, 4257–4259.

27. Myers,A.G., Gleason,J.L., Yoon,T. and Kung,D.W. (1997) Highly practical methodology for the synthesis of d- and l-α-amino acids, N-protected α-amino acids, and N-methyl-α-amino acids. *J. Am. Chem. Soc.*, **119**, 656–673.
28. Katsamba,P.S., Navratilova,I., Calderon-Cacia,M., Fan,L. *et al.* (2006) Kinetic analysis of a high-affinity antibody/antigen interaction performed by multiple Biacore users. *Anal. Biochem.*, **352**, 208–221.
29. Rich,R.L. and Myszka,D.G. (2009) Grading the commercial optical biosensor literature-Class of 2008: "The Mighty Binders." *J. Mol. Recognit.*, **23**, 1–64.
30. Rich,R.L., Papalia,G.A., Flynn,P.J., Furneisen,J. *et al.* (2009) A global benchmark study using affinity-based biosensors. *Anal. Biochem.*, **386**, 194–216.
31. Gomes-Pereira,M., Cooper,T.A. and Gourdon,G. (2011) Myotonic dystrophy mouse models: towards rational therapy development. *Trends Mol. Med.*, **17**, 506–517.
32. Lin,X., Miller,J.W., Mankodi,A., Kanadia,R.N., Yuan,Y., Moxley,R.T., Swanson,M.S. and Thornton,C.A. (2006) Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.*, **15**, 2087–2097.
33. Liu,Y. and Wilson,W.D. (2010) Quantitative analysis of small molecule-nucleic acid interactions with a biosensor surface and surface plasmon resonance detection. *Methods Mol. Biol.*, **613**, 1–23.
34. Staple,D.W. and Butcher,S.E. (2003) Solution structure of the HIV-1 frameshift inducing stem-loop RNA. *Nucleic Acids Res.*, **31**, 4326–4331.
35. Munde,M. *et al.* (2010) DNA minor groove induced dimerization of heterocyclic cations: compound structure, binding affinity, and specificity for a TTAA site. *J. Mol. Biol.*, **402**, 847–864.
36. Byers,L.D. (1977) Probe-dependent cooperativity patterns in Hill-plots. *J. Chem. Educ.*, **54**, 352.
37. Warf,M.B. and Berglund,J.A. (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA*, **13**, 2238–2251.
38. Yuan,Y., Compton,S.A., Sobczak,K., Stenberg,M.G., Thornton,C.A., Griffith,J.D. and Swanson,M.S. (2007) Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.*, **35**, 5474–5486.
39. Mooers,B.H., Logue,J.S. and Berglund,J.A. (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl Acad. Sci. USA*, **102**, 16626–16631.
40. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2010) Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res.*, **38**, 8370–8376.
41. Kiliszek,A., Kierzek,R., Krzyzosiak,W.J. and Rypniewski,W. (2009) Structural insights into CUG repeats containing the "stretched U-U wobble": implications for myotonic dystrophy. *Nucleic Acids Res.*, **37**, 4149–4156.
42. Lu,H. and Tonge,P.J. (2010) Drug-target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.*, **14**, 467–474.
43. Zhang,R. and Monsma,F. (2009) The importance of drug-target residence time. *Curr. Opin. Drug Discov. Devel.*, **12**, 488–496.
44. Copeland,R.A., Pompliano,D.L. and Meek,T.D. (2006) Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.*, **5**, 730–739.
45. Feigon,J., Denny,W.A., Leupin,W. and Kearns,D.R. (1984) Interactions of antitumor drugs with natural DNA: ^{1}H NMR study of binding mode and kinetics. *J. Med. Chem.*, **27**, 450–465.
46. Chaires,J.B., Dattagupta,N. and Crothers,D.M. (1985) Kinetics of the daunomycin-DNA interaction. *Biochemistry*, **24**, 260–267.
47. Phillips,D.R. and Crothers,D.M. (1986) Kinetics and sequence specificity of drug-DNA interactions: an *in vitro* transcription assay. *Biochemistry*, **25**, 7355–7362.
48. Chen,F.M. (1992) Binding specificities of actinomycin D to non-self-complementary -XGCY-tetranucleotide sequences. *Biochemistry*, **31**, 6223–6228.
49. Mosmann,T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods*, **65**, 55–63.
50. Mankodi,A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769–1773.
51. Crothers,D.M. (1968) Calculation of binding isotherms for heterogenous polymers. *Biopolymers*, **6**, 575–584.