

AWARD NUMBER: W81XWH-19-1-0719

TITLE: Design and Study of Small Molecules That Cleave the RNA That Causes Myotonic Dystrophy Type 1 (DM1)

PRINCIPAL INVESTIGATOR: Eric T Wang, Ph.D.

CONTRACTING ORGANIZATION: University of Florida, Gainesville, FL

REPORT DATE: October 2021

TYPE OF REPORT: ANNUAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE October 2021		2. REPORT TYPE ANNUAL		3. DATES COVERED 01SEPT2020 - 31AUG2021	
4. TITLE AND SUBTITLE Design and Study of Small Molecules That Cleave the RNA That Causes Myotonic Dystrophy Type 1 (DM1)				5a. CONTRACT NUMBER W81XWH-19-1-0719	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Eric T. Wang, Ph.D Disney, Matthew, D., Ph.D. Paegel, Brian, M., Ph.D.				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF FLORIDA DIVISION OF SPONSORED RESEARCH 1523 UNION RD RM 207 GAINESVILLE FL 32611-1941				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT: Myotonic dystrophy type 1 (DM1) is a genetic disorder characterized by multisystemic wasting of muscle function, including organ wasting that leads to cardiac disease, respiratory impairment, cataracts, and a host of other significant problems. In particular, DM1 is caused by an RNA repeat expansion [r(CUG)exp where "exp" denotes an expanded repeat] harbored in the 3' untranslated region (UTR) of the dystrophin myotonia protein kinase (DMPK) mRNA. This r(CUG)exp is toxic via a gain of function mechanism; the repeat forms a structure recognized by various RNA binding proteins, in particular muscleblind-like 1 which controls the alternative splicing of various transcripts. Mis-splicing of the muscle-specific chloride ion channel due to MBNL1 sequestration can be directly linked to myotonia. Over more than a decade, the Disney Laboratory has designed small molecules that bind and deactivate r(CUG)exp in patient derived cells and mouse models. Indeed, we have developed nM and pM inhibitors of r(CUG)exp dysfunction using innovative approaches: onsite drug synthesis in which r(CUG)exp catalyzes the synthesis of its own inhibitor, engendering small molecules with the ability to cross link to its target, and engendering small molecules with the ability to cleave disease causing RNAs directly or by recruiting an endogenous nuclease. Here, bolstered by in vivo studies in which a small molecule cleaves r(CUG)exp selectively, rescues 134 of 138 MBNL1 regulated splicing events, normalizes the transcriptome, and improves myotonia, we propose to develop antisense- or CRISPR- like small molecules into preclinical candidates for the treatment of the root cause of DM1. Importantly, these studies are directly applicable to other microsatellite diseases including amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and fragile X associated tremor ataxia syndrome (FXTAS).					
15. SUBJECT TERMS NONE LISTED					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unclassified	18. NUMBER OF PAGES 84	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	4
2. Keywords	4
3. Accomplishments	4 - 40
4. Impact	40 - 41
5. Changes/Problems	41 - 42
6. Products	42- 43
7. Participants & Other Collaborating Organizations	43 - 47
8. Special Reporting Requirements	47
References	48 - 52
9. Appendices	53 - 84

1. INTRODUCTION:

Myotonic dystrophy type 1 (DM1) is a genetic disorder characterized by multisystemic wasting of muscle function, including organ wasting that leads to cardiac disease, respiratory impairment, cataracts, and a host of other significant problems. In particular, DM1 is caused by an RNA repeat expansion [r(CUG)^{exp} where “exp” denotes an expanded repeat] harbored in the 3' untranslated region (UTR) of the dystrophin myotonia protein kinase (*DMPK*) mRNA. This r(CUG)^{exp} is toxic via a gain of function mechanism; the repeat forms a structure recognized by various RNA binding proteins, in particular muscleblind-like 1 which controls the alternative splicing of various transcripts. Mis-splicing of the muscle-specific chloride ion channel due to MBNL1 sequestration can be directly linked to myotonia. Over more than a decade, the Disney Laboratory has designed small molecules that bind and deactivate r(CUG)^{exp} in patient derived cells and mouse models. Indeed, we have developed nM and pM inhibitors of r(CUG)^{exp} dysfunction using innovative approaches: onsite drug synthesis in which r(CUG)^{exp} catalyzes the synthesis of its own inhibitor, engendering small molecules with the ability to cross link to its target, and engendering small molecules with the ability to cleave disease causing RNAs directly or by recruiting an endogenous nuclease. Here, bolstered by *in vivo* studies in which a small molecule cleaves r(CUG)^{exp} selectively, rescues 134 of 138 MBNL1 regulated splicing events, normalizes the transcriptome, and improves myotonia, we propose to develop antisense- or CRISPR- like small molecules into preclinical candidates for the treatment of the root cause of DM1. Importantly, these studies are directly applicable to other microsatellite diseases including amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and fragile X associated tremor ataxia syndrome (FXTAS).

2. KEYWORDS:

Myotonic dystrophy type 1; RNA repeat expansion; r(CUG)^{exp}; RNA binding proteins; muscleblind-like 1; myotonia; small molecule cleavage; splicing events

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Below we list the major goals of the project per our approved SOW.

SPECIFIC AIM 1a: Optimize our cleaving compound for binding to r(CUG)^{exp} by using a massively parallel screening approach enabled by using DNA-encoded libraries (DEL), focusing on the linker that tethers RNA-binding modules.

Major Task 1: Design and synthesis of a DNA-encoded small molecule library

Subtask 1: Design and synthesis of DEL libraries. **100% complete**

Months 1-4: 09/01/19-12/31/19

Prof. Paegel's laboratory: design, synthesis, QC

Prof. Disney's laboratory: design

(continued below)

Major Task 2: Screen the DEL to identify improved compounds that bind selectively to a model of $r(\text{CUG})^{\text{exp}}$, $r(\text{CUG})_{12}$. **100% complete**

Subtask 1: Development of FACS $r(\text{CUG})_{12}$ binding assay **100% complete**

Months 3-4: 11/01/19-12/31/19

Subtask 2: Multiplex FACS screen of DEL using $r(\text{CUG})_{12}$ and two competing RNAs ($r(\text{CUG})_2$ and fully paired RNA). **100% complete**

Months 4-9: 12/31/19-05/31/20

Subtask 3: Lead refinement by competition with **1** using FACS or with long residence times

Was not necessary

Months 9-12: 05/31/20-08/31/20

Subtask 4: Synthesis and characterization of hit compounds. **100% complete**

Months 9-12: 05/31/20-08/31/20

Major Task 3: In vitro evaluation of hit compounds from DEL. **100% complete**

Subtask 1: Hit validation: inhibit $r(\text{CUG})_{12}$ -MBNL1 complex. **100% complete**

Months 12-15: 09/01/20-11/30/20

Subtask 2: In vitro evaluation of most potent compounds from Subtask 1: affinity, k_{on} , k_{off} residence time by BL

Months 15-18: 11/01/20-02/28/21

Major Task 4: Evaluate compounds from **Maj. Task 3** in cells **100% complete**

Subtask 1: Assess cell permeability & cytotoxicity of compounds **100% complete**

Months 15-18: 11/01/20-02/28/21

Subtask 2: Study non-toxic, cell permeable compounds for improving splicing defects and foci in patient-derived cells. **100% complete**

Months 18-24: 02/01/21-08/31/21

Subtask 3: Study transcriptome-wide effects of compound treatment by RNA-seq. **100% complete**

Months 18-24: 02/01/21-08/31/21

SPECIFIC AIM 1b: Optimize the bleomycin A5 cleavage module and attach to lead compounds

Major Task 5: Synthesis of bleomycin conjugates. **100% complete**

Subtask 1: Synthesis of bleomycin derivatives and conjugates. **100% complete**

Months 1-12: 09/01/19-08/31/20

(continued below)

Major Task 6: In vitro evaluation. **100% complete**

Subtask 1: Study cleavage of r(CUG)₁₂ by small molecule-bleomycin conjugates by gel electrophoresis. **100% complete**

Months 12-18: 08/01/20-02/28/21

Subtask 2: Compare extent of cleavage of r(CUG)₁₂ by small molecule-bleomycin conjugates to DNA and other RNAs. **100% complete**

Months 12-18: 08/01/20-02/28/21

Major Task 7: Cellular evaluation of conjugates. **100% complete**

Subtask 1: Assess cell permeability and cytotoxicity

Months 18-21: 02/01/21-05/31/21

Subtask 3: Study non-toxic, cell permeable compounds for selective cleavage of r(CUG)^{exp} (patient-derived cells). **100% complete**

Months 21-28: 05/01/21-12/31/21

Subtask 4: As a counter screen, study compounds in Subtask 3 for inducing DNA damage in patient-derived cells. **100% complete**

Months 21-28: 05/01/21-12/31/21

SPECIFIC AIM 2: Rigorously evaluate optimized small molecule-bleomycin conjugates in cells and *in vivo*.

Major Task 1: Comprehensive *in cellulis* evaluation of cleaving compounds emanating from Aim 1.

Subtask 1: Study selective cleave of r(CUG)^{exp} in DM1 cell lines. **100% Complete**

Months 24-30: 08/01/21-02/28/22

Subtask 2: Study if compounds induce DNA damage. **100% Complete**

Months 24-30: 08/01/21-02/28/22

Subtask 3: Study compounds for improving DM1-associated splicing defects and reducing nuclear foci in multiple cell lines. **100% Complete**

Months 24-30: 08/01/21-02/28/22

Subtask 4: Complete a comprehensive analysis of transcriptome-and proteome-wide effects.

Months 24-30: 08/01/21-02/28/22

Major Task 2: Complete DMPK studies.

Subtask 1: In vitro DMPK analysis

Months 1-30: 09/01/19-02/28/22

(continued below)

Subtask 2: Mouse pharmacokinetics and tissue distribution

Months 1-30: 09/01/19-02/28/22

Subtask 3: Study lung fibrosis (not expected)

Months 1-30: 09/01/19-02/28/22

Major Task 3: Determine the optimal dosing regimen of optimal bleomycin conjugates and RIBOTAC probes.

Subtask 1: Study myotonia over different dosages and treatment periods, informed by

Specific Aim 2, Major Task 2.

Months 6-36: 02/01/20-08/30/22

Subtask 2: Study improvement of splicing defects and formation of foci over different dosages and treatment periods.

Months 6-36: 02/01/20-08/30/22

Subtask 3: Study lung fibrosis (not expected; bleomycin conjugates) over different dosages and treatment periods.

Months 6-36: 02/01/20-08/30/22

Subtask 4: Complete a comprehensive analysis of transcriptome- and proteome-wide effects.

Months 6-36: 02/01/20-08/30/22

Subtask 5: Compare effects of small molecule-bleomycin conjugates and RIBOTACS probes in vivo to ASOs.

Months 6-36: 02/01/20-08/30/22

SPECIFIC AIM 3: Targeted small molecule recruitment of a nuclease to r(CUG)^{exp}.

Major Task 1: Synthesis of RIBOTAC probes

Subtask 1: Synthesis of different RNase L recruiting modules and small molecule-RNase L conjugates. **100% complete – currently undergoing re-optimization**

Months 1-6: 09/01/19-02/28/20

Major Task 2: Assess if small molecule-RNase L conjugates recruit RNase L in vitro

Subtask 1: Evaluate ability of small molecule-RNase L conjugates to cleave r(CUG)₁₂ - FRET-based assay. **100% complete – currently undergoing re-optimization**

Months 6-18: 02/01/20-02/28/21

Subtask 2: Rigorously evaluate the ability of small molecule-RNase L conjugates to recruit RNase L and cleave r(CUG)₁₂. **Ongoing**

Months 6-18: 02/01/20-02/28/21

(continued below)

Major Task 3: Cellular evaluation of RIBOTAC probes

Subtask 1: Assess cell permeability & cytotoxicity of compounds.

Months 12-15: 08/01/20-11/30/20

Subtask 2: Study non-toxic, cell permeable compounds for selective cleavage of r(CUG)^{exp} (patient-derived cells).

Months 15-21: 11/01/20-05/31/21

Major Task 4: Comprehensive in cellulis evaluation of RIBOTAC probes

Subtask 1: Determine if RIBOTACs selectively cleave r(CUG)^{exp} in multiple DM1 patient-derived cell lines.

Months 18-30: 02/01/21-02/28/22

Subtask 2: Study compounds for improving DM1-associated splicing defects and reducing foci in multiple cell lines.

Months 18-30: 02/01/21-02/28/22

Subtask 3: Complete a comprehensive analysis of transcriptome-wide and proteome-wide (including immune system) effects of compound treatment.

Months 18-30: 02/01/21-02/28/22

What was accomplished under these goals?

SPECIFIC AIM 1a: Optimize our cleaving compound for binding to r(CUG)^{exp} by using a massively parallel screening approach enabled by using DNA-encoded libraries (DEL), focusing on the linker that tethers RNA-binding modules. [Note: parts of the summary below are also related to SPECIFIC AIM 1b.]

Note: the studies described below were completed by the Disney Laboratory.

Major Tasks 1 & 2 were completed during the previous funding period. Please see our previous report for details. Below we summarize our progress on Major Tasks 3 & 4, which are now complete. Prior to discussing progress on Major Tasks 3 & 4, below please find an update on our studies to optimize the linker in dimeric molecules that bind r(CUG)^{exp} and inhibit hallmarks of disease in DM1 patient-derived cells, further fulfilling Specific Aim 1a.

Massively parallel optimization of the linker domain in small molecule dimers targeting a toxic r(CUG) repeat expansion.

Targeting RNA with synthetic molecules was long thought challenging both because of RNA's limited local diversity from only four nucleotides and its flexible and dynamic nature. Many RNAs, however, adopt stable three-dimensional (3D) structures, which are potential binding sites for small molecules.¹⁻⁶ The discovery of selective small molecules targeting disease-causing RNA has historically been accomplished through selection methods⁷ or in various targeted screens for

specific RNA structures.^{8, 9} Suitable RNA targets for small molecules were identified in a myriad of diseases, such as cancers caused by oncogenic microRNA (miR)-21,^{10, 11} or spinal muscular atrophy.¹² To identify and optimize novel chemical matter targeting toxic RNA structures, robust combinatorial methods are required. Advanced combinatorial chemistry methods, such as DNA encoded libraries,¹³⁻¹⁶ one-bead-one-compound libraries (OBOC),^{17, 18} phage display,^{19, 20} and affinity selection-mass spectrometry^{21, 22} have been successfully applied to proteins and can be adapted for RNA. Notably, Patel and coworkers reported a one-bead-two-compounds (OBTC) strategy to discover macrocycles targeting the long non-coding RNA Growth Arrest Specific 5 (GAS5), with implications for the treatment of type 2 diabetes mellitus.²³ The application of combinatorial methods to disease-causing RNAs enables the discovery of novel chemical probes and optimization of ligands targeting these RNA structures in a massively parallel fashion.

One class of toxic RNAs is repeat expansions that cause greater than 40 genetic diseases including myotonic dystrophy type 1 (DM1). DM1 is caused by a r(CUG) repeat expansion [r(CUG)^{exp}] located in the 3' untranslated region (UTR) of the *dystrophia myotonica protein kinase* (DMPK) mRNA.²⁴ DM1 is an incurable neuromuscular disease with a prevalence of ~1 in 8,000 people. Those afflicted with DM1 have symptoms that include muscle weakness and myotonia, heart abnormalities, cataracts, and insulin resistance.^{25, 26} The r(CUG)^{exp} folds into a highly structured hairpin, the stem of which comprises a periodic array of 5'CUG/GUC 1×1 internal loops. These loops bind and sequester the pre-mRNA splicing regulator muscleblind-like 1 (MBNL1) protein as well as other RNA-binding proteins in nuclear foci (**Figure 1A**). Sequestration of MBNL1 by r(CUG)^{exp} prevents its interaction with its natural substrates, therefore leading to dysregulation of alternative pre-mRNA splicing and manifestation of disease.^{27, 28} Different types of modalities can bind to r(CUG)^{exp} and improve disease defects in DM1-affected cells,²⁹⁻³¹ including monomeric small molecules,³¹⁻³³ modularly assembled dimeric compounds that occupy multiple internal loops simultaneously,^{30, 34} pseudopeptides,³⁵ and small molecules that cleave r(CUG)^{exp}.³⁶ These compounds were discovered by various approaches, such as a bead-based screening method called resin-bound dynamic combinatorial chemistry (RBDCC)³⁵ and by fragment-based target-guided synthesis.³⁴

Our group has previously reported the modular assembly of RNA-binding modules targeting consecutive 5'CUG/GUC 1×1 internal loops to disrupt the toxic MBNL1- r(CUG)^{exp} complex and improve DM1-associated defects (**Figure 1B** and **1C**).³⁷ A dimeric compound named **2H-4** was built by a *N*-propylglycine peptoid bridge to separate the binding modules at a specific distance to enable simultaneous binding (**Figure 1D**). Efforts to improve the binding affinity and potency of **2H-4** have involved changing the nature of its backbone including an *N*-methyl alanine linker,³⁸ a proline linker,³⁹ and macrocyclization.⁴⁰ These modifications have improved binding affinity and cellular activity, however thus far the optimization of the linker between RNA-binding modules has not been attempted on a large scale.

Peptoids, oligomers of *N*-substituted glycine building blocks, offer access to a large chemical diversity by simple incorporation of primary amine building blocks. Further, peptoids have desirable pharmacological properties, such as cellular permeability and resistance to proteolytic degradation.^{41, 42} Since their syntheses are fully compatible with OBOC library approaches,⁴³ we envisioned a simple method to introduce a wide variety of functional groups in the linker domain of **2H-4**, which is described herein.

Briefly, the OBOC library was synthesized on 90 μm Tentagel microbeads with four variable positions between the two 5'CUG/GUC internal loop binding modules, as this is the optimal linker length identified from previous studies (Figure 2A).³⁸ A peptide tag, BBRGYM, was incorporated at the C-terminal of each compound, allowing release from the beads by selective cyanogen bromide (CNBr) cleavage of the M residue for facile deconvolution of compound identity by tandem mass spectrometry (MS-MS). Indeed, fragmentation of the amide bonds generates a mass ladder that represents the nature of the different building blocks and their sequential order in the compound. The peptide tag provides a known starting point in the MS-MS spectrum from which unknown building blocks are identified. A set of 24 different building blocks were incorporated by split-and-pool methodology (Figure 2B), generating a theoretical diversity of 331,776 compounds. Building blocks were selected to cover various chemotypes, such as aliphatic, cationic, aromatic or heteroaromatic moieties. As a quality control step, random beads from the library were selected and processed for compound deconvolution, and indeed MS-MS analysis enabled structure deconvolution unambiguously.

To identify compounds that avidly bind $\text{r(CUG)}^{\text{exp}}$, the OBOC library was incubated with 150 nM of an established structural model of $\text{r(CUG)}^{\text{exp}}$, r(CUG)_{12} containing a 5' biotin tag,⁴⁴ in the presence of 15 μM of **2H-4** (100-fold molar excess as compared to the RNA's concentration). The 2H-4 competitor was added to increase the stringency of the screen and to allow identification of compounds that bind more avidly to r(CUG)_{12} than the competitor. Beads bound to 5'-biotin- r(CUG)_{12} were isolated with streptavidin-coated Dynabeads (Figure 2C) and identified by MS-MS analysis, affording 142 hit compounds.

To triage the hits to a more manageable number for further investigation, we calculated the enrichment and the statistical confidence thereof (Z-score) of each building block at each variable position (Figure 2D). (A Z-score of 1.96 corresponds to $P = 0.05$ and hence statistical significance.) The Z-score at each position was also summed for each hit compound (Figure S2). Several trends could be observed in selection amongst the building blocks, including (i) a depletion of tertiary amines, particularly building blocks #19-22, at positions 1 – 3 ($P = 0.013$ to 0.039) and (ii) enrichment of secondary and primary amines, building blocks #13 and #15, at all positions ($P < 0.0001$).

We observed that the sum of Z-scores for the building blocks of the hit compounds correlated with positive charge. Since the RNA backbone is negatively charged, this is not surprising. To further investigate this observation as well as to maintain chemical diversity, we binned hit compounds by their linker charge and then selected those in each bin with Z-scores in the top 30%. This first triage provided a reasonable number of compounds ($n = 32$) to synthesize by solid-phase parallel synthesis on Rink amide polystyrene resin, measure avidity for bind $\text{r(CUG)}^{\text{exp}}$, and study for rescue of DM1-associated defects.

In previous studies of dimeric compounds targeting $\text{r(CUG)}^{\text{exp}}$, we observed that cell uptake and localization can change as a function of the structure of the linker.^{39, 40} Furthermore, charged backbones are known to affect localization.⁴⁵ As nuclear localization is crucial for biological activity as $\text{r(CUG)}^{\text{exp}}$ is sequestered in nuclear foci, we studied the compounds' cellular permeability and localization.

DM1-patient derived myotubes containing 1300 r(CUG) repeats were treated with 5 μ M of each compound and localization was monitored by using the inherent fluorescence of the H RNA-binding modules. Interestingly, **2H-4** localizes almost completely in the nucleus, as do 12 of the 32 hit OBOC compounds, which were thus prioritized for further study. A closer inspection of building block distribution provided insights into the relationship between compound structure and nuclear localization. Amongst the nuclearly localized compounds, only the polar building block #14 (*N*-hydroxyethyl substituent) was found to be significantly enriched at position 2 (17%, $P = 0.03$). Within the compounds with poor nuclear localization, building blocks #2 (*N*-propyl substituent) and #13 (*N*-methylaminobutyl substituent) were found to be significantly enriched at positions 3 (20%, $P = 0.03$) and 2 (33%, $P = 0.049$), respectively. These data suggest that the position where polar and hydrophobic moieties are incorporated within the dimer linker influence localization. MBNL1 protein self-regulates the alternative splicing of its exon 5, which is included too frequently in DM1 cells.⁴⁶ Thus, we evaluated which of the 12 compounds rescue the *MBNL1* exon 5 splicing defect in DM1-patient derived myotubes and compared their activities to **2H-4** (Figure 3).

DM1 cells were treated with a 5 μ M dose for **2**, **5**, **10**, **13**, **15**, **18**, **21**, and **25** for 48 h. Changes in cellular morphology were observed upon treatment with 5 μ M of **7**, **14**, **22** and **32**, and therefore these compounds were evaluated at 1 μ M, where no changes in morphology were observed. Of the eight compounds evaluated at 5 μ M, only two slightly improved the *MBNL1* exon 5 splicing defect, by ~20%, which is less than the effect observed upon treatment with **2H-4** (~30% improvement) (Figure 3A). The four compounds evaluated at 1 μ M all improved the *MBNL1* exon 5 splicing defect, more potently than **2H-4** which is inactive at the 1 μ M dose (Figure 3B), although the percent rescue by **32** was not statistically significant. Notably, **7** and **22** improved *MBNL1* exon 5 splicing dose-dependently, with **22** appearing to be the more potent of the two (7% rescue for **22** vs. 2% rescue for **7** at 0.2 μ M) (Figure 3C). As **14** improved splicing similarly at all concentrations tested, it was eliminated from further study (Figures 3C). Since **22** was the most potent, it was studied in more detail, both *in vitro* and in DM1 patient-derived myotubes.

We first measured the binding of **22** to r(CUG)₁₂ *in vitro*, affording an EC₅₀ of 106 $\pm$ 4 nM (Figure 4A), ~10-fold more avid than **2H-4** (EC₅₀ = 1140 $\pm$ 31 nM).⁴⁰ Importantly, no saturable binding was observed for **22** and a base paired control RNA that does not contain the 1 $\times$ 1 nucleotide UU internal loops. This avid binding of **22** to r(CUG) repeats translated to inhibition of the formation of the r(CUG)^{exp}-MBNL1 complex *in vitro* with an IC₅₀ of 2.8 $\pm$ 0.2 μ M in a TR-FRET assay,⁴⁴ a ~10-fold improvement over **2H-4** (IC₅₀ of 32.2 $\pm$ 4.3 μ M; determined by the same TR-FRET assay),⁴⁰ which correlates with its enhanced affinity.

To rationalize the improved *in vitro* binding of **22** to r(CUG)₁₂ when compared to **2H-4**, molecular modeling studies were carried out. Models of dimer **22** bound to a model RNA with two 1 $\times$ 1 nucleotide UU internal loops in r(CUG)^{exp} [r(5'-CCGCUGCUGCGG/3'GGCGUCGUCGCC)] were generated using molecular dynamics simulations and compared with a previously published model of **2H-4**. The best model, as defined by the lowest free binding energy, generated for **22** was 10 kcal/mol lower than **2H-4** (-61.7 vs. -51.7 kcal/mol).

Inspection of the interactions of **22** with the internal loops shows increased number of interactions between the dimer's linker region and the RNA backbone, in addition to the interactions of RNA-binding modules with the RNA, which are similar to **2H-4** (**Figure 4E**). The imidazole group of the building block in position 1 hydrogen bonds with NH₂ of G21 [r(5'-CCGCUGCUGCGG/3'GGCGUCGUCGCC)] and stacks with the RNA-binding module. The carbonyl groups in the backbone undergo a number of hydrogen bonding interactions, including intramolecular hydrogen bonds with the NH-CH₃ side chain of the building block in position 3 and NH₂ of the side chain of the building block in position 4. Intermolecular H-bonding interactions of the dimer backbone with the RNA also contribute significantly to the binding of the dimer, including: (i) interaction of building block 2's carbonyl with 2' OH of G6; (ii) H-bond formed between the terminal amide CONH₂ with 2' OH of C7; (iii) interaction of building block 3's carbonyl with G6's and G8's NH₂; and (iv) H-bond between the linker's terminal amide NH₂ with phosphate backbone of U8. These interactions, in addition to the Van der Waals and π -stacking interactions formed throughout the dimer-RNA complex, attribute to the observed 10 kcal/mol improvement in the binding energy when compared to the bound structure of **2H-4**.

The selectivity and activity of **22** was further investigated in DM1 patient-derived myotubes. Importantly, the observed rescue of *MBNL1* exon 5 pre-mRNA splicing defect was not due to transcriptional inhibition of *DMPK*, as its RNA levels were unaffected (**Figure 4B**), suggesting direct r(CUG)^{exp} target engagement. Importantly, **22** specifically improved DM1-associated defects as *MAP4K4* exon 22a splicing, a NOVA-regulated splicing event,⁴⁷ was not affected in DM1 myotubes. Furthermore, *MBNL1* exon 5 splicing was not affected in wild-type myotubes (from healthy donors) treated with **22**. Compound **22** also reduced the number of r(CUG)^{exp}-MBNL1 foci in DM1 myotubes by ~20%, which correlates with the observed improvement in splicing (**Figures 4C & 4D**). Thus, through specific binding to r(CUG)^{exp}, **22** potently and specifically improves DM1-associated defects in patient-derived myotubes.

In conclusion, the OBOC library methodology provides a facile means to optimize the linker domain of a dimeric compound targeting r(CUG)^{exp}. A simple affinity-based selection strategy enabled the screening of >330,000 compounds and subsequent hit identification via MS-MS sequencing. Through subsequent analysis of the bioactivity of hit compounds, we identified compound **22** which bound r(CUG) repeats ten times more avidly than **2H-4** by forming additional interactions between the target and optimized linker and more potently rescued disease-associated defects in DM1 patient-derived myotubes. Thus, OBOC library synthesis and screening can be used to identify high affinity binders to r(CUG)^{exp}. Importantly, the methodology developed herein is likely to be general, applicable to numerous other RNA targets, to aid in the identification of high-affinity small molecules.

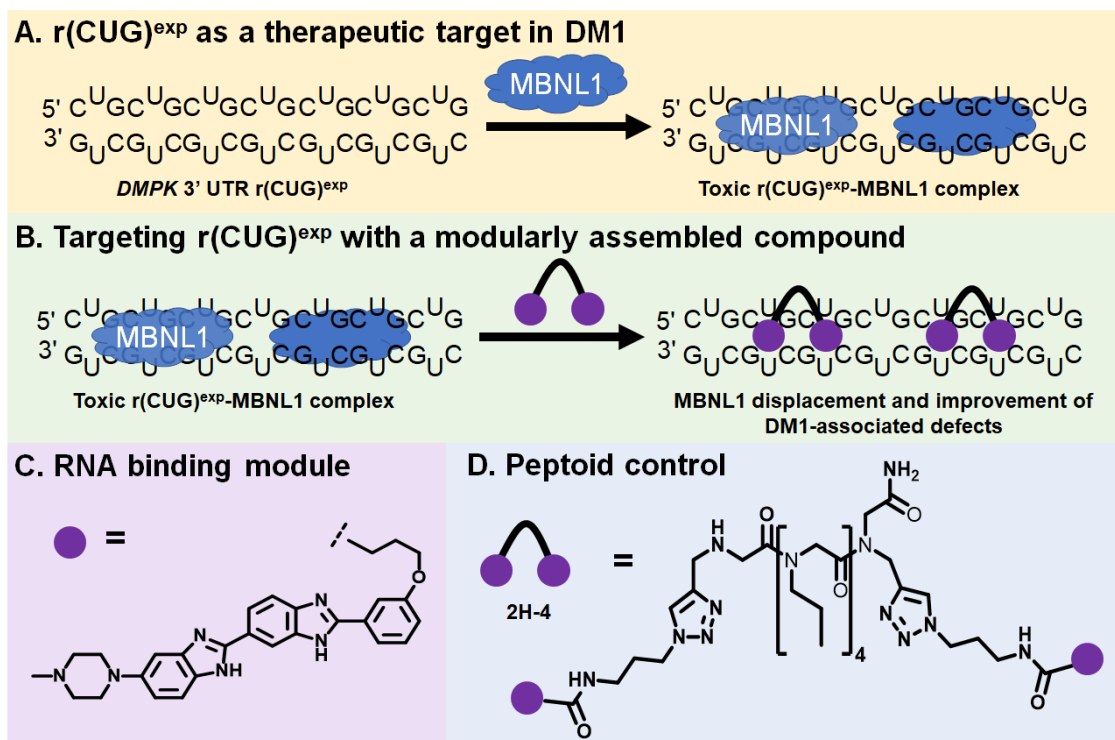


Figure 1. Small molecule targeting approach for the toxic RNA repeat expansion, r(CUG)^{exp}, that causes myotonic dystrophy type 1 (DM1). (A) The toxic r(CUG)^{exp} present in the 3' UTR of the *DMPK* mRNA sequesters MBNL1 protein, a regulator of alternative pre-mRNA splicing, with high affinity, leading to splicing defects. (B) Modularly assembled RNA-binding modules bind with high affinity to r(CUG)^{exp}, liberating MBNL1 and rescuing DM1-associated splicing defects. (C) Chemical structure of the 5'CUG/3'GUC (loops present in r(CUG)^{exp}) RNA-binding module used in this study. (D) Chemical structure of a previously developed dimeric compound that rescues DM1-associated defects, **2H-4**. **2H-4** was used as a starting point to develop the OBOC methodology described herein.

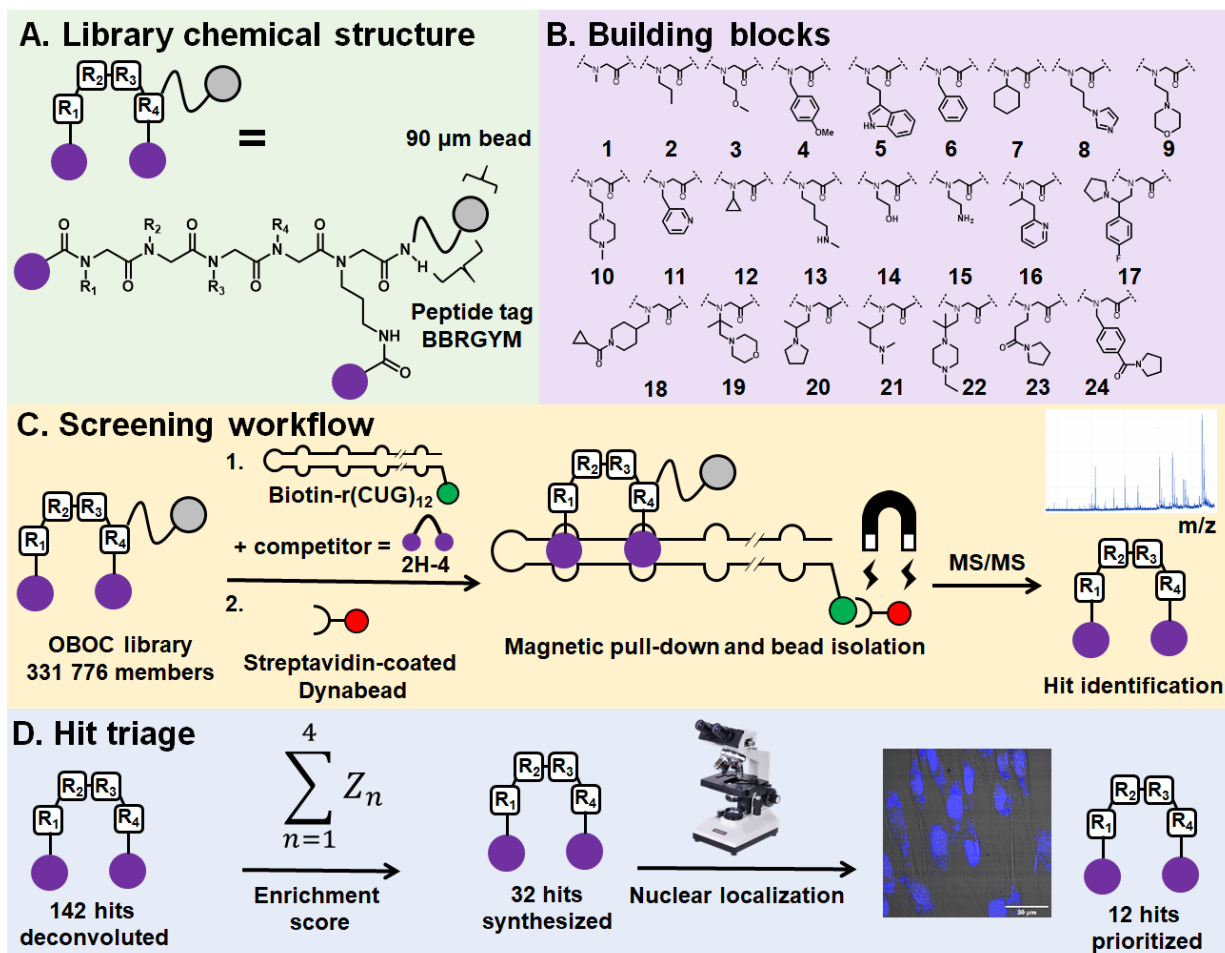


Figure 2. OBOC library and screening method to optimize dimeric compounds targeting $r(\text{CUG})^{\text{exp}}$. (A) Chemical structure of the library with R_{1-4} indicating variable residues. (B) Building blocks incorporated to generate the chemical diversity of the 331,776-member library. (C) Screening workflow using compounds supported on 90 μm Tentagel beads. 1) Incubation with biotinylated $r(\text{CUG})_{12}$; 2) magnetic pull-down using streptavidin coated Dynabeads; 3) bead isolation and cleavage; 4) hit structure deconvolution by tandem mass spectrometry. (D) Hit triage of the 142 hits by enrichment score and nuclear localization in DM1 myotubes.

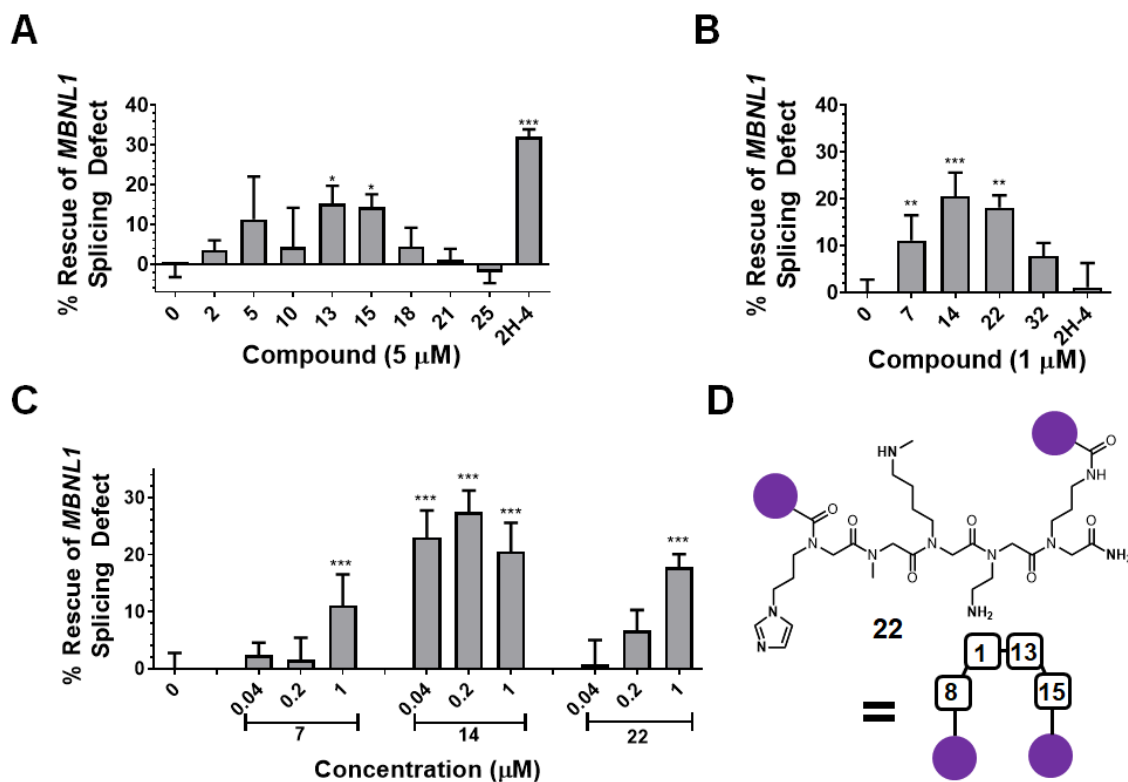


Figure 3. Evaluation of hit compound activity in DM1 patient-derived myotubes. (A) Activity of hit compounds with no toxicity at 5 μ M, as assessed by rescue of the *MBNL1* exon 5 splicing defect in DM1 myotubes (n = 3). (B) Activity of hit compounds with no toxicity at 1 μ M, as assessed by rescue of the *MBNL1* exon 5 splicing defect in DM1 myotubes (n = 3). (C) Dose response analysis of compounds with statistically significant activity at 1 μ M (n = 6). For all panels: error bars represent SD, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, as determined by one-way ANOVA. (D) Chemical structure of compound 22.

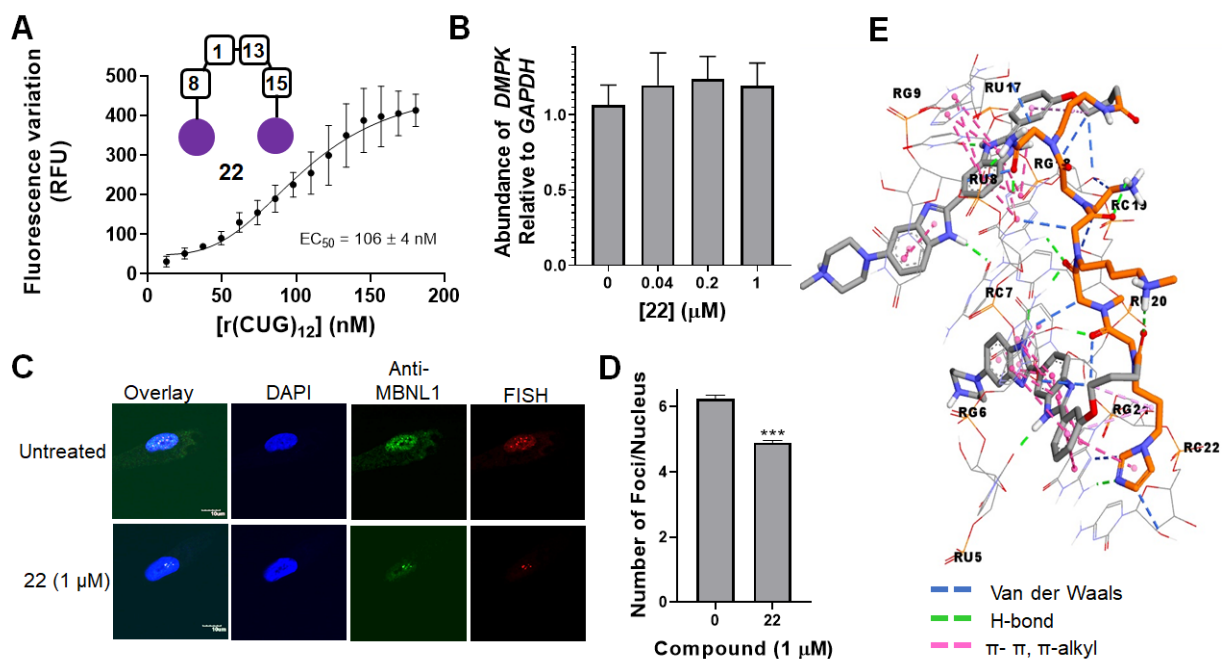


Figure 4. Compound **22** binds avidly to r(CUG)₁₂ *in vitro* and selectively improves disease-associated defects in DM1 myotubes. (A) Representative direct binding curve of **22** and r(CUG)₁₂ hairpin using the inherent fluorescence of **22** (n = 3). (B) Evaluation of *DMPK* levels in DM1 myotubes treated with **22** via RT-qPCR (n = 3). (C) Representative confocal microscopy images of nuclear foci. MBNL1 was imaged by immunofluorescence while r(CUG)^{exp} was imaged by RNA fluorescence *in situ* hybridization (FISH). (D) Quantification of the number of foci/nucleus in DM1 myotubes treated with **22** (n = 3; 40 nuclei counted/replicate). *** *P* < 0.001 Student's *t*-test. (E) Interactions between **22** and two 5' CUG/3' GUC internal loops present in r(CUG)^{exp} from a model generated by molecular dynamics simulation. Dashed lines represent interactions between the linker residues of **22** and the RNA.

Macrocycles are privileged scaffolds displaying desirable pharmacological properties.⁴⁸⁻⁵³ They provide a rigid scaffold that bind targets with high affinity and selectivity and have a propensity for high cellular uptake.⁵⁴⁻⁵⁸ A few studies reported the design of cyclic ligands targeting 3D RNA motifs.^{59, 60} The effect of macrocyclization on ligands that bind r(CUG)^{exp} is thus the purview of this report (Figure 5A). The bioactive compound **2H-4**, previously designed to bind r(CUG)^{exp}, comprises two RNA-binding modules (H) separated by four *N*-propylglycine spacing units, the optimal distance to span the two base pairs between r(CUG)^{exp}'s internal loops (Figure 5B & 5C).⁶¹ To mimic the periodicity between the RNA-binding modules in a macrocyclic topology, we developed a solid phase synthesis approach to construct a library of dimeric compounds that display two H modules.

A collection of macrocycles with ring sizes from 26 to 35 atoms was synthesized (Figure 5D). The size of the macrocycle and the distance between the H RNA-binding modules was systematically altered. An allyl ester was incorporated as a side chain to provide a palladium-labile protected carboxylic acid to afford macrocyclization via amide bond formation. Final treatment of the resin with trifluoroacetic acid yielded the desired diamine macrocycles, which were then coupled to two H RNA-binding modules (Figure 5D).

The library of macrocycles was studied for inhibition of the r(CUG)^{exp}-MBNL1 complex *in vitro*.^{44, 62} An initial study was completed at 5 μ M compound (Figures 5E & 5F). The macrocycle with the shortest linker, **2H-2C2** showed the greatest amount of inhibition (~50%). A full dose-response revealed that **2H-2C2** had an IC₅₀ of ~3 μ M, a 10-fold improvement over linear **2H-4** (IC₅₀ = 32.2 $\pm$ 4.3 μ M, Figure 5G). To assess the selectivity of macrocycle **2H-2C2**, we developed a high throughput selectivity screen that monitors the binding of a base paired RNA and the protein DiGeorge Syndrome Critical Region Gene 8 (DGCR8D), a general RNA-binding protein.⁶³ In this assay, **2H-2C2** and **2H-4** had IC₅₀s of >20 μ M and >100 μ M, respectively, indicating selectivity for r(CUG)^{exp} (Figure S1). Thus, this newly developed assay can be used in High-throughput screening (HTS) format to define ligand selectivity for RNA. We next measured the affinity and selectivity of **2H-2C2** using direct RNA binding assays using a model r(CUG)₁₂ construct. The macrocycle bound r(CUG)₁₂ with an EC₅₀ of 81 $\pm$ 8 nM, compared to 1,100 $\pm$ 30 nM for linear **2H-4**. Neither compound exhibited measurable affinity for the fully base-paired RNA.

Modeling studies were performed to study conformational differences between **2H-4** and **2H-2C2** by using a model of r(CUG)^{exp} containing two copies of the 5'CUG/3'GUC motif. Conformational searching of the unbound state generated 35 conformers for **2H-4** and only 10 conformers for **2H-2C2**.⁶⁴ These results suggest conformational restriction of the macrocycle and its pre-organization to bind r(CUG)^{exp}, as has been observed previously with other macrocyclic ligands.^{48, 65, 66} Next, we generated a model of the H module bound to a single 5'CUG/3'GUC motif by using dynamic docking. Based on the lowest energy structure, homology models of **2H-4**, and **2H-2C2** bound to an RNA with two 5'CUG/3'GUC loops [r(CUG)₄] were generated. Explicit solvent MD simulations, cluster analysis, and additional MD simulations were used to study binding energies of the bound complexes.

The most probable structure for **2H-4** (32% probability; Figure 6A) had a free energy of -46 kcal/mol, whereas **2H-2C2** (38% probability; Figure 6B) had a free energy of -51 kcal/mol. Interestingly, both complexes have the same number of interactions ($n = 24$). Closer inspection revealed important interactions unique to the **2H-2C2**-r(CUG)₄ complex occur to the macrocyclic linker. Thus, both entropic (pre-organization) and enthalpic contributions may enhance binding of **2H-2C2** to r(CUG)^{exp}, as compared to **2H-4**.

As aforementioned, macrocyclization can enhance cellular uptake.^{48, 67, 68} In comparison to linear **2H-4**, the macrocycle **2H-2C2** entered DM1 patient-derived myotubes at a 10-fold higher concentration (Figure 7A). Subcellular localization studies showed that **2H-4** was predominantly nuclear while **2H-2C2** is nuclear and cytoplasmic. We next compared the ability of both compounds to rescue DM1-associated defects. We first studied reduction of nuclear foci containing r(CUG)^{exp} and MBNL1 by RNA fluorescence *in situ* hybridization (FISH) and immunohistochemistry (IHC). **2H-2C2** reduced the number of foci per cell from 4.5 ± 0.4 to 3.2 ± 0.2 in DM1 myotubes ($P < 0.01$) while linear **2H-4** was less effective, reducing the number of foci to 3.8 ± 0.1 per cell ($P < 0.05$) (Figures 7B & C). Sequestration of r(CUG)^{exp} in nuclear foci reduces its nuclear export, causing reduced translation of *DMPK* mRNA.⁶⁹

Therefore, we measured if **2H-2C2** could allow for nucleocytoplasmic transport of the r(CUG)^{exp}-containing mRNA and hence translation; that is, if **2H-2C2** affects mRNA subcellular localization. Experiments were completed by using a luciferase reporter fused to a 3'UTR harboring r(CUG)₈₀₀ or control reporter lacking r(CUG) repeats (Figure 8A).⁶¹ When cells stably expressing luciferase-r(CUG)₈₀₀ were treated with macrocyclic **2H-2C2**, luciferase activity was increased, indicating increased transport to the cytoplasm (Figure 8B). In contrast, no change in luciferase activity was observed upon treatment of cells with the reporter lacking r(CUG) repeats (Figure 8C). Further, **2H-2C2** was 10-fold more potent than linear compound **2H-4** in this cellular assay.

Rescue of nuclear foci and nucleocytoplasmic transport defects suggests that **2H-2C2** liberates MBNL1 and may also rescue dysregulation of alternative pre-mRNA splicing. We assessed alleviation of two splicing defects, *MBNL1* exon 5 and nuclear receptor corepressor 2 (*NCOR2*) exon 45a, both of which are regulated by MBNL1, in a DM1-patient derived cell line.⁷⁰ Indeed, **2H-2C2** (200 nM) rescued aberrant splicing of *MBNL1* exon 5 by $25 \pm 4\%$, which is 10-fold more potent than **2H-4** (Figure 9D). Similar results were obtained for *NCOR2* exon 45a. Importantly, neither compound affected the NOVA-regulated alternative splicing of mitogen-activated protein kinase kinase kinase kinase 4 (*MAP4K4*) exon 22a.⁷¹ To ensure that this improvement in MBNL1-regulated splicing events was due to binding of **2H-2C2** to r(CUG)^{exp} and not reduction of *DMPK* mRNA abundance, we measured *DMPK* levels by RT-qPCR in DM1 myotubes. No effect was observed as expected for a selective RNA-targeted ligand. Furthermore, **2H-2C2** has no effect on *DMPK* levels and *MBNL1* exon 5 splicing in healthy fibroblasts as the target, r(CUG)^{exp}, is not present. Thus, **2H-2C2**'s mechanism of action is binding to toxic r(CUG)^{exp} and freeing sequestered proteins, thereby improving DM1-associated defects, such as formation of nuclear foci, defects in nucleocytoplasmic transport of mutant *DMPK* mRNA, or aberrant splicing.

In summary, we investigated the impact of macrocyclization of a bivalent ligand targeting r(CUG)^{exp} in DM1 patient-derived myotubes. A small library of macrocycles was generated with

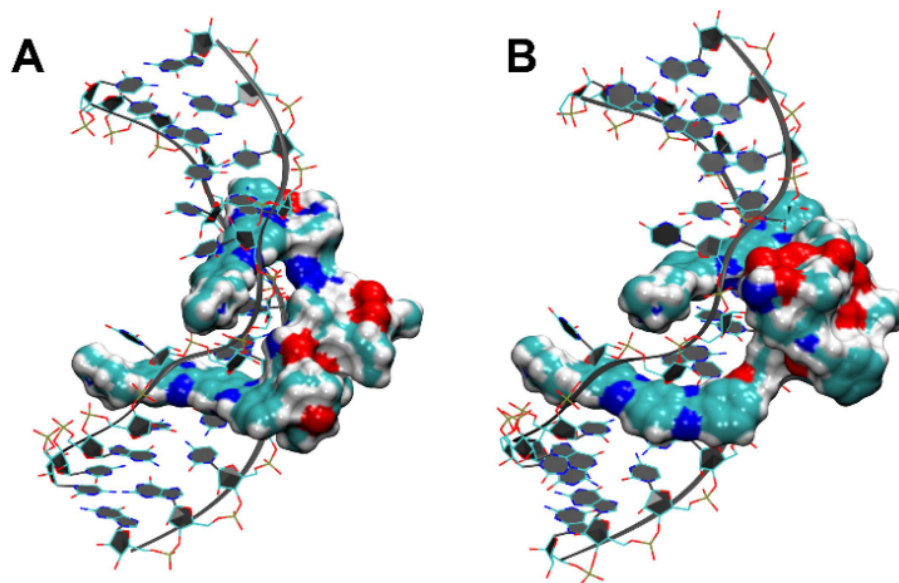


Figure 6. The bound state with the lowest binding energy for (A) **2H-4/r(CUG)₄** and (B) **2H-2C2/r(CUG)₄**. RNA molecules are displayed in gray NewRibbon representation, while the dimeric compounds are displayed in MSMS model (green).

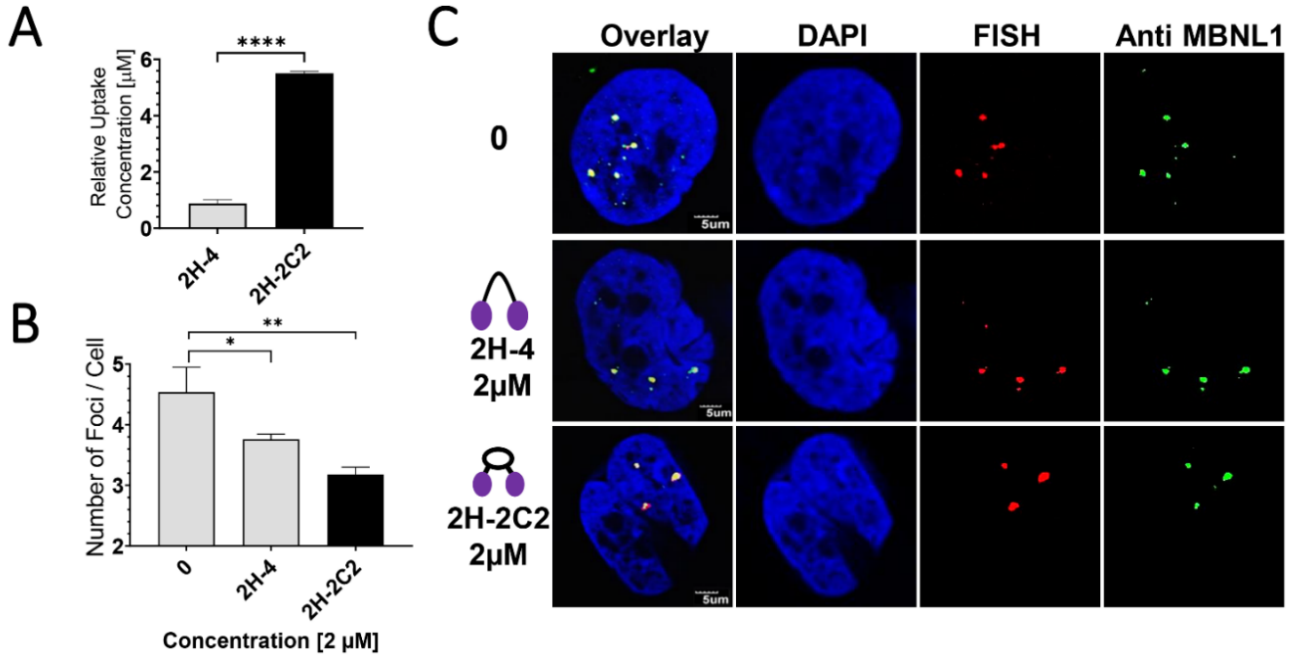


Figure 7. Cellular permeability of **2H-2C2** relative to **2H-4** and rescue of nuclear foci formation. (A) Evaluation of cellular permeability using the intrinsic fluorescence of **2H-2C2** or **2H-4** in DM1 myotubes (n = 3). (B) Quantification of r(CUG)^{exp}-MBNL1 foci/nucleus (n = 3 biological replicates, 40 nuclei counted per replicate). (C) Representative microscopic images for the reduction of nuclear foci by **2H-2C2** and **2H-4**, imaged by RNA FISH and IHC using a fluorescently labeled complementary oligonucleotide and an anti-MBNL1 antibody, respectively.⁷² Error bars represent SD. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$, as determined by a one-way ANOVA.

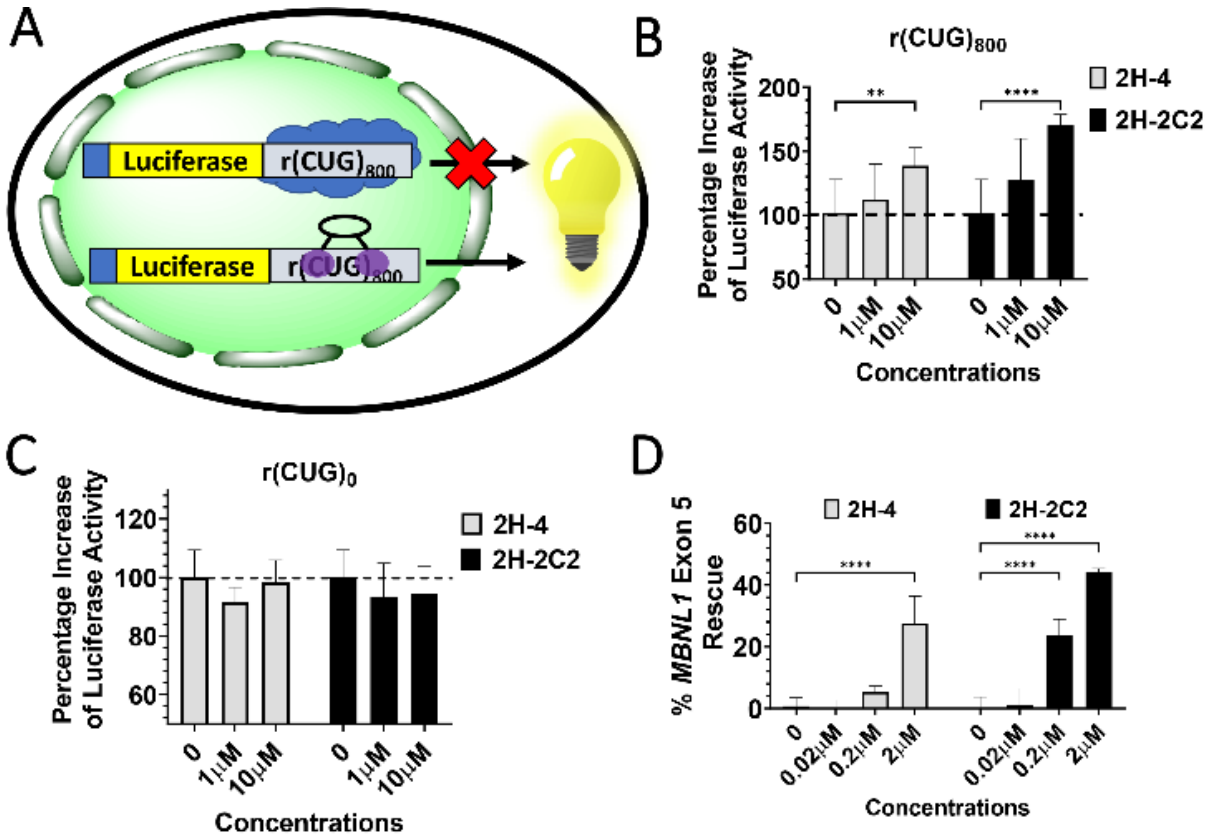


Figure 8. 2H-2C2 alleviates DM1-associated defects. (A) Schematic of the cellular model used to study translational defects, a C2C12 mouse myoblast cell line that stably expresses firefly luciferase mRNA with r(CUG)₈₀₀ in the 3' UTR. r(CUG)₈₀₀ causes the transcript to be retained mostly in the nucleus and thus it is not efficiently translated (low luciferase activity). If the macrocycle binds to r(CUG)₈₀₀ and displaces or inhibits MBNL1 binding, then the transcript is more efficiently exported from the nucleus and translated in the cytoplasm (increased luciferase activity). (B) Quantification of the effect of 2H-2C2 and 2H-4 on the DM1 translational defect, as measured by luciferase activity in a C2C12 cell line with r(CUG)₈₀₀ (n = 6). (C) Quantification of the effect of 2H-2C2 and 2H-4 on luciferase activity in a control C2C12 cell line that does not harbor the disease-causing repeats [r(CUG)₀] (n = 6). (D) 2H-2C2 and 2H-4 rescue the MBNL1 exon 5 splicing defect in DM1 patient-derived myotubes (n = 3). Error bars represent SD. ***P* < 0.01, *****P* < 0.0001 as determined by a one-way ANOVA.

Major Task 3: *In vitro* evaluation of hit compounds from DEL.

- Subtask 1: Hit validation: inhibit r(CUG)₁₂-MBNL1 complex.
- Subtask 2: *In vitro* evaluation of most potent compounds from Subtask 1: affinity, k_{on} , k_{off} , residence time by BLI

Major Task 4: Evaluate compounds from **Major Task 3** in cells

- Subtask 1: Assess cell permeability & cytotoxicity of compounds
- Subtask 2: Study non-toxic, cell permeable compounds for improving splicing defects and foci in patient-derived cells.
- Subtask 3: Study transcriptome-wide effects of compound treatment by RNA-seq

In our previous report, we summarized our efforts to design and study a DNA-encoded library for binding r(CUG)^{exp}. In brief, we successfully synthesized the DEL, where all building blocks ($n = 3$) were selected for druglikeness. We rapidly screened the DEL library for selective binding of r(CUG) repeats by fluorescence activated cell sorting (FACS).

To complete a FACS screen for selective binders, the r(CUG) repeat RNA and a fully base paired control were labeled with a different fluorophore, Cy-5 and Alexa-750 dyes, respectively. The DEL beads were incubated simultaneously with the two RNAs. We were only interested in compounds selective for r(CUG)₁₂; that is, they had high fluorescence signal derived from r(CUG)₁₂ and little or no fluorescence derived from the fully paired RNA, affording a hit rate of 0.5%.

Next Generation Sequencing (NGS) was used to decode each building block position via its DNA tag. Commonalities between the hits were then assessed. In building block position #1, D-phenylalanine and L-4-chloro-phenylalanine were most highly represented; L-4-chloro-phenylalanine was also the most highly represented at position 2 while a phenyl-substituted benzimidazole was enriched in position 3 (**Figure 9A**). We next applied three filters to refine the hits: (i) total number of reads; (ii) high quality reads; and (iii) redundant hits. The intersection of all three criteria afforded 23 hit compounds from the DEL screen (**Figure 9B**). Of these, 16 were successfully re-synthesized with high purity (>95%) and were evaluated for displacing MBNL1 from a r(CUG)₁₂-MBNL1 complex *in vitro*. The percentage of MBNL1 displaced ranged from 0 – 40% by 50 μ M of the compound of interest, the most potent of which is compound **23** (**Figure 9C**).

We next tested the nine compounds that showed activity *in vitro* in a cell-based assay to study a molecular defect in DM1, decreased nucleocytoplasmic transport and hence translation of mutant DMPK. Here, we used a previously reported luciferase reporter in which r(CUG)₈₀₀ or no repeats fused to luciferase as a 3' UTR is stably expressed in the mouse myoblast cell line C2C12. Active compounds will bind the repeats and facilitate transport to the cytoplasm where luciferase is translated; thus, luciferase activity should increase dose dependently. Indeed, a dose dependent increase in luciferase was observed upon treatment of C2C12 cells expressing luciferase-r(CUG)₈₀₀ with compound **3** and **23**, albeit to a lesser extent (**Figure 10A**). We therefore further studied **3**, which had no activity in C2C12 cells stably expressing the luciferase reporter expressing

no r(CUG) repeats, suggesting a selective interaction (**Figure 10B**). In DM1 patient-derived myotubes, **3** improves the MBNL1 exon 5 splicing defect (deregulated in DM1 due to sequestration of r(CUG)^{exp} by MBNL1) by ~30% at a 10 μ M dose, with no effect on *DMPK* levels (**Figure 10C**). These data support a mechanism of action whereby **3** binds the repeats and displaced MBNL1 to resume its normal function, not transcriptional inhibition (DNA level). Further, these cell permeable compounds were not toxic to DM1 myotubes at the active concentration.

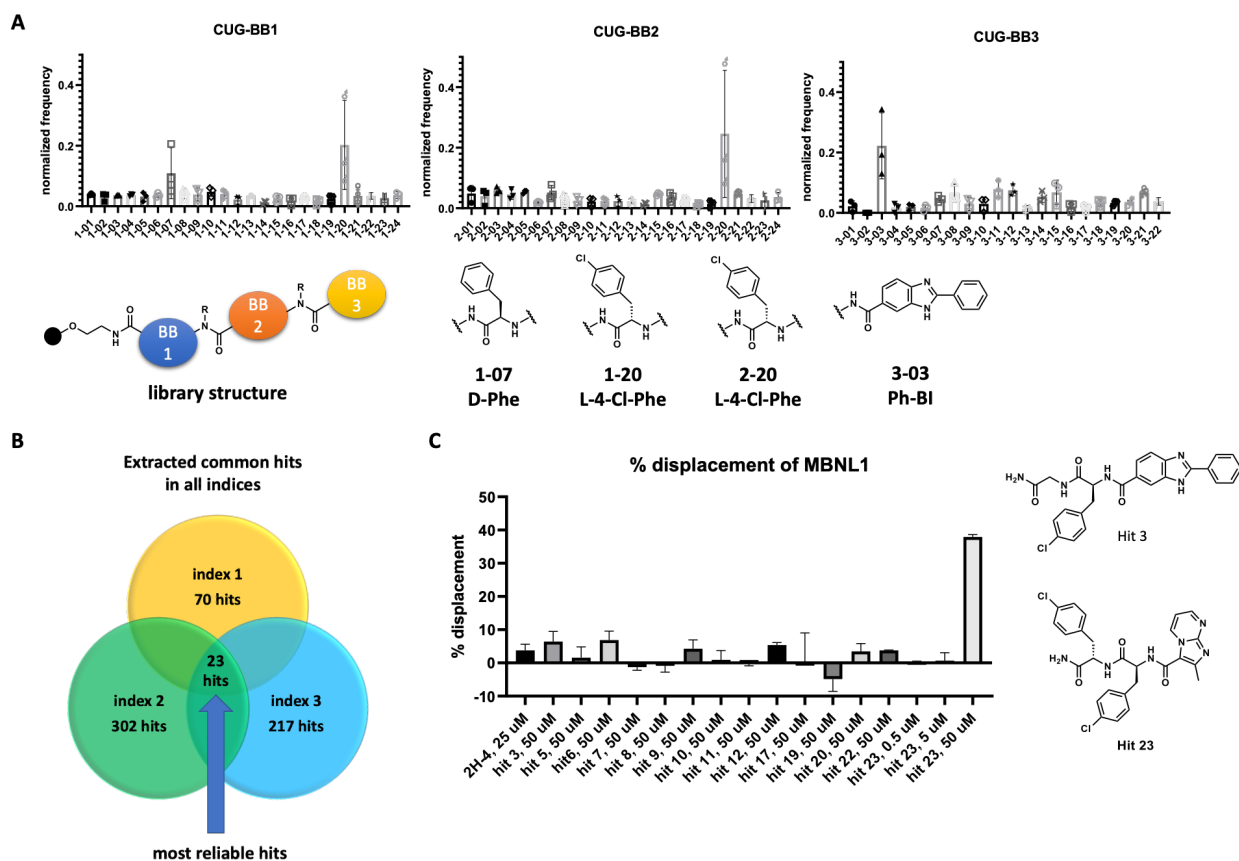


Figure 9: Screening of a DNA-encoded library (DEL) to identify new chemical matter that binds r(CUG) repeats. **A**) Frequency of DEL building blocks observed at each position of compounds that selectively bind r(CUG) repeats vs. a fully paired RNA. **B**) Schematic of the refinement process for DEL hit compounds. **C**) Displacement of MBNL1 from r(CUG) repeats in vitro for DEL hit compounds and the structures of two lead compounds.

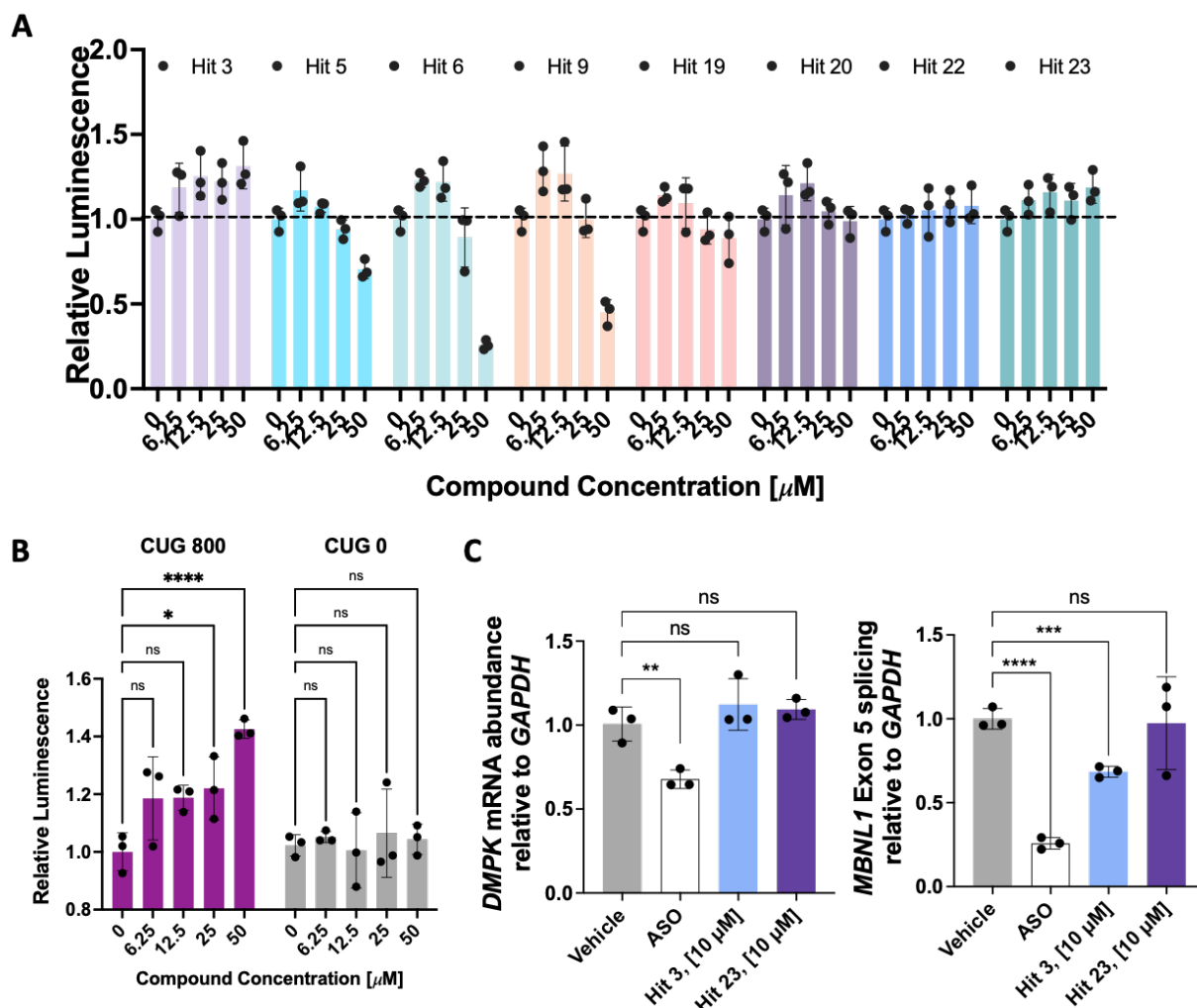


Figure 10: Cellular activity of DEL hit compounds. **A)** Activity of DEL compounds from improving the nucleocytoplasmic transport defect of mutant DMPK using a luciferase-based reporter. **B)** Dose dependent rescue of the nucleocytoplasmic transport defect of mutant DMPK by **3** and its inactivity in a control luciferase reporter containing no r(CUG) repeats. **C)** Effect of **3** on DMPK mRNA abundance and on rescue of the MBNL1 exon 5 splicing defect in DM1 patient-derived myotubes.

SPECIFIC AIM 1b: Optimize the bleomycin A5 cleavage module and attach to lead compounds

and

SPECIFIC AIM 2: Rigorously evaluate optimized small molecule-bleomycin conjugates in cells and *in vivo*.

Note: The section below summarizes activities completed by the Disney Laboratory (TSRI).

The use of bleomycin (BLM) analogs to specifically cleave r(CUG)^{exp} offers an attractive method to enhance RNA cleavage selectivity by further diminishing off-target DNA cleavage. In our previous technical report, we showed that the carbohydrate domain of BLM is not essential for permeability or cleavage of r(CUG)^{exp} when attached to r(CUG)^{exp}-binding small molecules. Notably, it is necessary for the efficient cleavage of DNA⁷³ and cellular permeability,⁷⁴ the disaccharide, enhancing selectivity for the RNA target by reducing DNA damage that occurs with high concentrations of BLM-conjugated small molecules. Below, we report additional work we have completed on bleomycin-small molecule conjugates.

In particular, we appended **22** (reported above in progress for Specific Aim 1a) with bleomycin A5 using its terminal amine as attachment at this site is known to decrease bleomycin's affinity for DNA and ablate off-target effects (**Figure 11A**).³⁶ The compound **22-Bleo** was first evaluated for its ability to selectively bind and cleave r(CUG)₁₂ *in vitro*. Indeed **22-Bleo** selectively bound r(CUG)₁₂ with an EC₅₀ of 120 ± 7 nM (similar to the EC₅₀ of **22**, 106 ± 4 nM; note: Fe²⁺, required for cleavage, is absent in binding assays), with no binding observed to a base paired RNA (**Figure 11B**). The RNA degrader also cleaved the r(CUG) repeats *in vitro*, and >70% of the RNA was cleaved at a 5 µM dose (**Figure 11C**). Thus, **22-Bleo** can selectively recognize r(CUG)^{exp} and cleave it *in vitro*.

We next studied whether the *in vitro* cleavage activity of **22-Bleo** translated to cleavage of r(CUG)^{exp} in DM1 myotubes. Indeed, a ~35% reduction in the abundance of *DMPK* was observed after treatment with 1 µM of **22-Bleo** (**Figure 11D**). Importantly, this cleavage was specific to the disease-causing, *DMPK* allele harboring r(CUG)^{exp} as *DMPK* levels were not affected in wild-type cells. A competition experiment between **22-Bleo** and parent compound **22** was then completed to confirm the former's mode of action (cleavage rather than transcriptional inhibition) and the latter's direct engagement of r(CUG)^{exp}. As expected, co-treatment of DM1 myotubes with varying concentration of **22** and a constant concentration of **22-Bleo** afforded a dose dependent rescue of *DMPK* levels (**Figure 11G**).

Specific cleavage of r(CUG)^{exp} resulted in improvement of DM1-associated defects including ~30% rescue of the *MBNL1* exon 5 splicing defect and ~30% reduction in the number of r(CUG)^{exp}-MBNL1 nuclear foci (**Figures 11E-F**). A small but statistically significant improvement in splicing was observed at 0.2 µM indicating that **22-Bleo** is more potent than **22** (**Figure 11E**). Similar to the parent compound, **22-Bleo** did not affect *MBNL1* exon 5 splicing in wild-type myotubes or *MAP4K4* exon 22a splicing in DM1 myotubes, indicating specific effects.

Importantly, the cleavage mode of action of **22-Bleo** allows for direct profiling of potential off-targets. We assessed the levels of all transcripts containing short, non-pathological r(CUG) repeats upon treatment of DM1 myotubes with **22-Bleo**. None of these genes was significantly affected indicating that **22-Bleo** can specifically recognize and cleave the disease-causing repeat expansion (**Figure 11H**). We have previously shown that these RNAs containing shorter r(CUG) repeats do not fold into a structure containing repeating 1×1 U/U internal loops, the source of the observed selectivity.³⁶

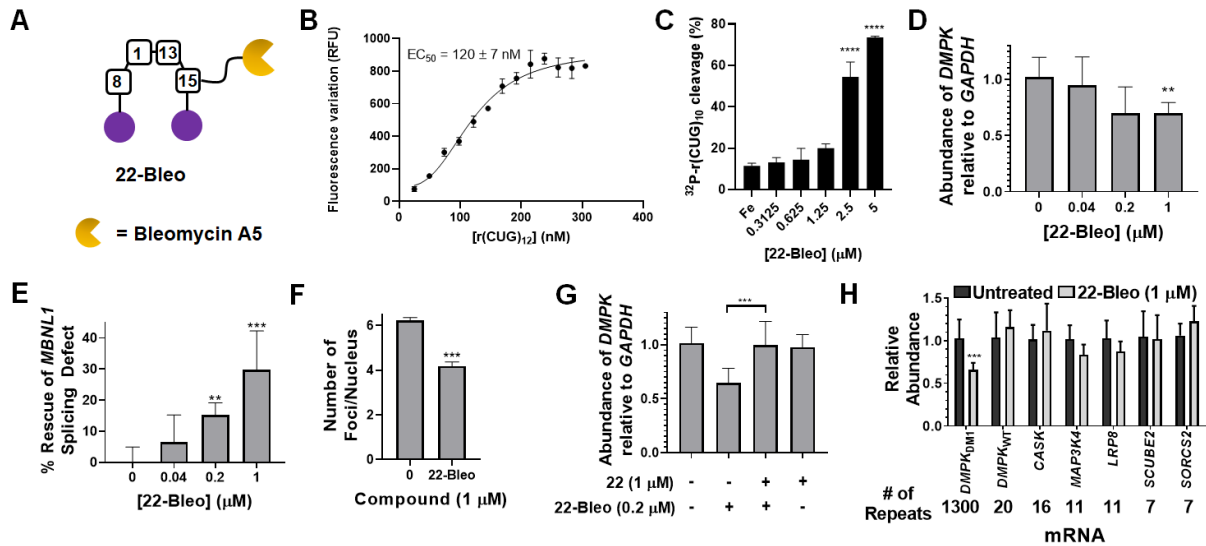


Figure 11. Compound **22-Bleo** cleaves r(CUG)^{exp} *in vitro* and in DM1 myotubes, rescuing disease-associated defects. (A) Conjugation of **22** to bleomycin A5 afforded cleaving compound **22-Bleo**. (B) Direct binding assay of **22-Bleo** with (CUG)₁₂ hairpin using the fluorescence of **22** (n = 3). Note: assays were completed in the absence of Fe²⁺, required for cleavage. (C) *In vitro* cleavage activity of **22-Bleo** using radioactively labeled r(CUG)₁₀ and analysis of fragments by gel electrophoresis (n = 3); **** *P* < 0.0001, as determined by one-way ANOVA. (D) Evaluation of *DMPK* levels upon compound treatment via RT-qPCR (n = 6); ** *P* < 0.01, as determined by one-way ANOVA. (E) Ability of **22-Bleo** to improve the *MBNL1* exon 5 splicing defect in DM1 myotubes (n = 6); ** *P* < 0.01, *** *P* < 0.001, as determined by one-way ANOVA. (F) Quantification of the number of foci/nucleus in DM1 myotubes treated with **22-Bleo** (n = 3; 40 nuclei counted/replicate); *** *P* < 0.001, as determined by a Student's *t*-test. (G) Competitive cleavage experiment between **22** and **22-Bleo** where excess **22** prevents cleavage of *DMPK* by **22-Bleo**. Levels of *DMPK* mRNA were measured by RT-qPCR (n = 6); *** *P* < 0.001, as determined by one-way ANOVA. (H) Abundance of r(CUG) repeat-containing transcripts upon treatment with **22-Bleo**, as measured by RT-qPCR (n = 6); *** *P* < 0.001, as determined by a Student's *t*-test. For all panels, error bars represent SD.

Complementarily to these studies, we have also identified new chemical matter, a low molecular weight fragment, that binds to r(CUG)^{exp}, the conjugation of which to bleomycin affords a potent cleaver of r(CUG)^{exp} in patient-derived cells. Relative to conventional ligand discovery screens which utilize larger libraries of higher molecular weight compounds, fragment-based ligand discovery (FBLD) utilizes libraries of lower molecular weight compounds to more efficiently explore chemical space⁷⁵⁻⁷⁷. However, a major challenge of FBLD is the development of assays that detect low affinity binding or short residence times of fragments^{76, 78-82}. Recent work has demonstrated that this obstacle can be overcome through the installation of photoaffinity groups onto small molecule fragments, to capture and identify bound protein targets directly in cells via mass spectrometry-based proteomics^{83, 84}. These fully functionalized fragments (FFFs) thus can be used to provide binding site information, to broadly assess proteome-wide “ligandability”, and even be advanced to compounds that selectively modulate protein function.

Identification of CUG targeting small molecules in vitro: Previous work has demonstrated the efficacy of utilizing FFFs to elucidate RNA-small molecule interactions.⁸³⁻⁸⁵ By using FFF screening methods, information can be gathered on a small molecule’s binding site, selectivity, and even proteome-wide drugability. We therefore screened a library of FFFs to identify potential binders of r(CUG)^{exp}. Using our previously reported method, chemical cross-linking and isolation by pull-down (Chem-CLIP),^{86, 87} radiolabeled r(CUG)₁₀ was incubated with a FFF library consisting of ~500 functionalized small molecule fragments (**Figure 12A**). In our preliminary screen, of the 473 compounds tested, 18 showed significant enrichment of the 5'- ³²P-labeled r(CUG)₁₀ after pulldown when treated at 100 μ M of compound (**Figure 12B**). Compounds that demonstrated significant enrichment were then tested in full dose response. Of the 18 initial hits, six dose dependently enriched 5'- ³²P-labeled r(CUG)₁₀ (**Figure 12C**), with compound **1** (**Figure 12D**) having the most significant enrichment.

Synthesis of compound 2: As lead fragment **1** demonstrated the most significant enrichment of ³²P-r(CUG)₁₀ *in vitro*, we proceeded with the synthesis of a simple binding compound **2** (**Figure 13A**), which lacks the diazirine cross-linking module, replacing it with a propanamide group. To assess if compounds **1** and **2** compete for the same site, a competitive chem-CLIP (C-Chem-CLIP) experiment was performed wherein ³²P-r(CUG)₁₀ was treated with 50 μ M compound **1** along with varying concentrations of **2** ranging from 0 to 500 μ M. Indeed, compound **2** was able to compete **1** from the RNA, resulting in a dose-dependent reduction in the amount of radioactive r(CUG)₁₀ pulled down (**Figure 13B**), thereby confirming they bind to the same site.

Affinity measurements for lead compounds: To confirm the molecular recognition of compounds **1** and **2**, its affinity for r(CUG)₁₂, and r(CAG-CUG)₁₂ (fully paired) was measured by microscale thermophoresis (MST). Indeed, both compounds **1** and **2** demonstrated saturable binding to r(CUG)₁₂ with K_d of 237 $\pm$ 129 nM and 34 $\pm$ 23 nM respectively, while having no measurable binding to the fully paired RNA (**Figure 13C**). Binding of **1** was further confirmed by NMR spectroscopic analysis.

Compound **1** and the corresponding control compound **3**, which contains the diazirine and alkyne but lacks the RNA-binding module, were examined for their ability to cross-link to *DMPK* in cells using the Chem-CLIP approach. DM1 and wild type myotubes were treated with 5 μ M of **1** or **3** for 24 h. Following compound treatment, cells were irradiated with UV light to cross-link

the Chem-CLIP probes to any bound RNA. The alkyne handle is clicked to biotin, and the cross-linked RNA is isolated and purified with streptavidin beads. The abundance of *DMPK* was assessed before and after pull-down for each treatment group by RT-qPCR. In DM1 myotubes, *DMPK* abundance was enriched by ~7-fold upon treatment 5 μ M of **1** (**Figure 13D**), while no enrichment was observed with **3**. Importantly neither compound significantly enriched *DMPK* abundance in wild type myotubes. When DM1 myotubes were co-treated with **1** and **2**, we observed dose-dependent competition as measured by decreased enrichment of *DMPK* by **1** (**Figure 13E**), in agreement with in vitro experiments.

Synthesis of 037Bleo: As aforementioned, we have previously demonstrated that r(CUG)^{exp} can be successfully targeted with small molecule-bleomycin conjugates to elicit targeted cleavage of the repeat containing *DMPK* transcript and rescue MBNL1 splicing defects.⁸⁸ To this end, **2** was functionalized with bleomycin A5 to afford small molecule **4** (**Figure 14A**).

Target validation of compound **4** was first assessed by *in vitro* C-Chem-CLIP. Radiolabeled r(CUG)₁₀ was co-incubated 50 μ M of **1**, and varying concentrations of **4** in a solution lacking iron, which is required for cleavage. With increasing concentration of **4**, a dose dependent decrease in the percent of 5'-³²P-r(CUG)₁₀ pulled down by **1** was observed (**Figure 14B**), indicating that **4** indeed binds the target RNA. We verified cleavage by **4** using radioactively labeled r(CUG)₁₀ and analysis of the resulting fragments by gel electrophoresis. Significant cleavage of r(CUG)₁₀ was observed upon incubation with **4**, with ~50% of the RNA cleaved at as little as 2.5 μ M, and 75 $\pm$ 14% cleavage observed at 10 μ M (**Figure 14C**). No statistically significant cleavage of r(CUG)₁₀ was observed with the parent compound **2**, which does not have the bleomycin functionality. Collectively, these data confirm that **4** indeed binds the 1 $\times$ 1 internal U/U loops formed by the expanded r(CUG) repeat, eliciting targeted cleavage of the RNA *in vitro*.

Bioactivity of 037Bleo: We next sought to measure the ability of **4** to selectively cleave r(CUG)^{exp} and improve DM1-associated defects in cells (**Figure 15A**). In DM1 patient-derived myotubes,⁸⁹ **4** cleaved 40 $\pm$ 8% of r(CUG)^{exp}-containing *DMPK* when treated at 5 μ M (**Figure 15B**). Notably, the observed decrease in *DMPK* abundance was selective for r(CUG)^{exp}, as WT myotubes, which express r(CUG)₂₀, were unaffected by treatment with **3**. This selectivity can be traced to differences in structure – r(CUG)^{exp} folds into an array of 1 $\times$ 1 UU internal loops whereas r(CUG)₂₀ does not. Additionally, treatment with 5 μ M of **4** was able to improve the dysregulation of MBNL1 exon 5 splicing observed in DM1 myotubes by 60 $\pm$ 6% (**Figure 15B**), while having no effect on MBNL1 splicing in wildtype myotubes.

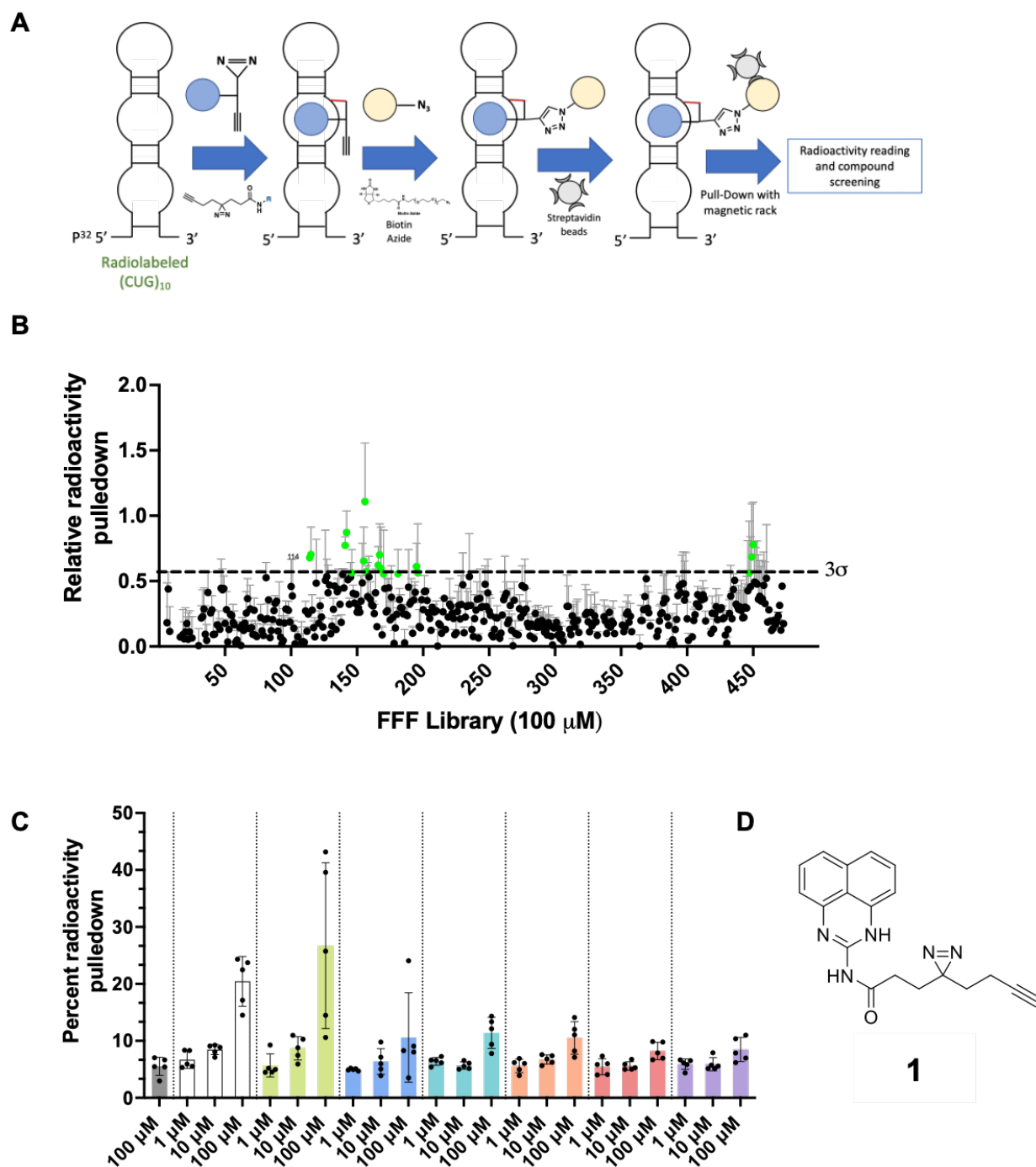


Figure 12. Screening of Diazirine small molecule library. **A)** Schematic of screening FFFs for binding to r(CUG) repeats. **B).** *In vitro* Chem-CLIP screening of FFF library for binding to r(CUG)₁₀. The screen identified 18 lead fragments enriching radioactivity by $>3\sigma$ from the mean of all FFFs studied. **C)** Secondary validation of hit compounds by measuring dose dependent enrichment of r(CUG)₁₀ by *in vitro*, with lead fragment **1** showing the greatest extent of enrichment. **D)** Chemical structure of lead fragment **1**.

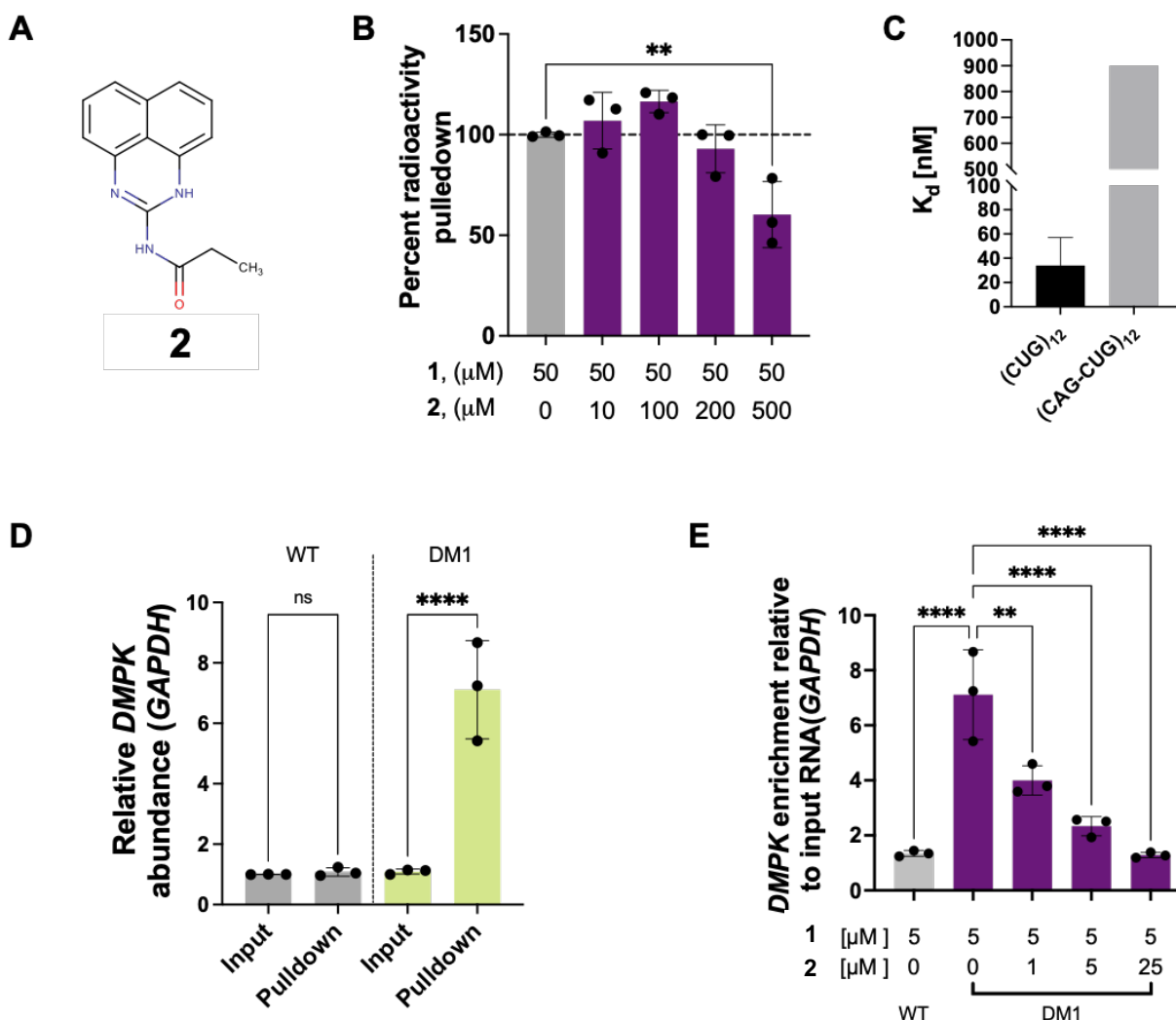


Figure 13. Validation of hit compound 037 target engagement *in vitro* and patient-derived cells. **A)** Chemical structure of parent compound 2, which lacks the diazirine functionalization, replacing it with a propanamide group. **B)** Competitive Chem-CLIP experiment performed *in vitro* between compounds 1 and 2 as demonstrated by a decrease in the percent radioactivity pulled down by 1 with increasing concentration of the competitor, 2 ($n = 3$); ** $P = 0.0053$, as determined by a One-way ANOVA with multiple comparisons. **C)** Binding affinity (K_d) of compounds 1 and 2 as determined by MST to r(CUG)₁₂ and a control nucleic acid, r(CAG-CUG)₁₂. **D)** Chem-CLIP pulldown of 1 in differentiated DM1 and wildtype myotubes. Compound 1 significantly enriches the r(CUG)^{exp}-containing *DMPK* gene selectively in DM1 cells ($n = 3$); **** $P < 0.0001$, as determined by a One-way ANOVA with multiple comparisons.. **E)** Competitive Chem-CLIP experiment performed in patient-derived myotubes. Compounds 1 and 2 compete for the same binding site in cells as observed by a decrease in the abundance of *DMPK* gene enriched by compound 1 upon addition of the competition, 2 ($n = 3$); ** $P = 0.0022$; **** $P < 0.0001$, as determined by a One-way ANOVA with multiple comparisons.

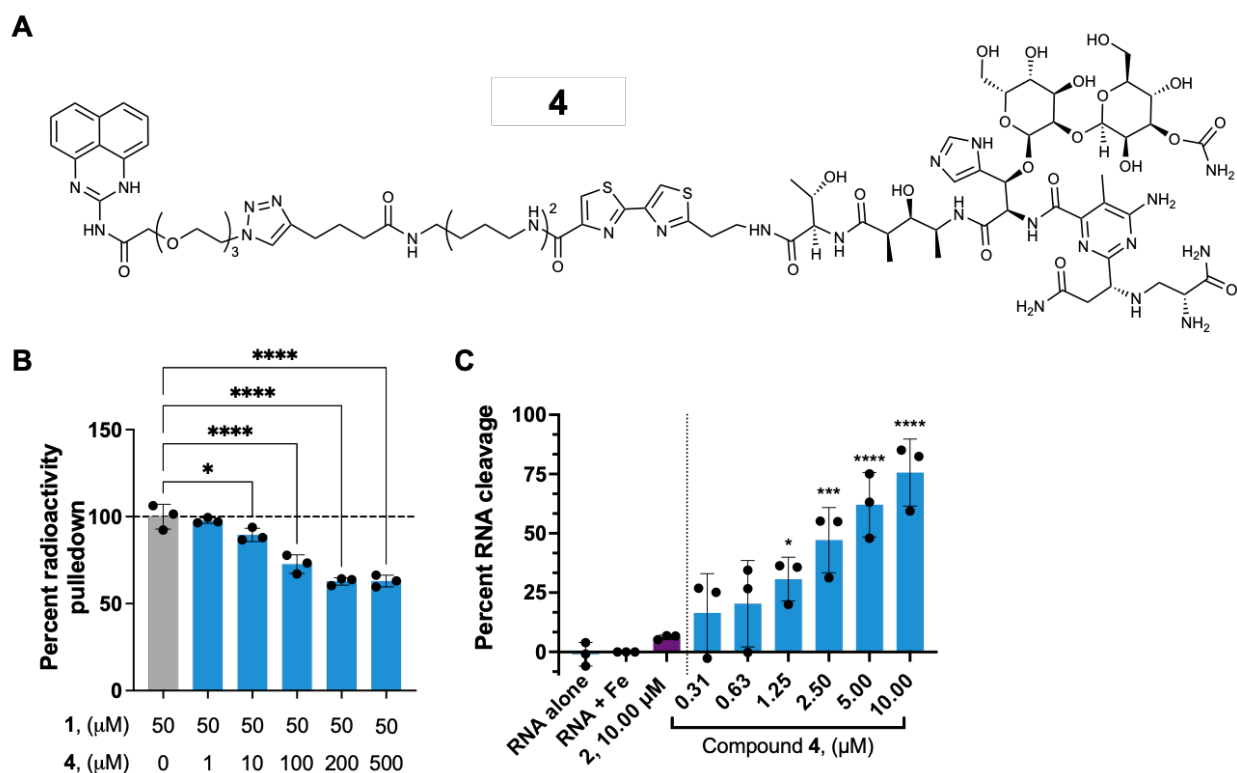


Figure 14. Design and validation of monomeric cleaver module that cleaves toxic r(CUG)^{exp}. **A)** Structure of compound **4**, functionalized with bleomycin A5, capable of directed RNA cleavage. **B)** Competitive Chem-CLIP experiment performed *in vitro* between compounds **1** and **4**, as demonstrated by a decrease in the percent radioactivity pulled down by **1** with increasing concentrations of the competitor, **4** ($n = 3$); * $P = 0.044$; **** $P < 0.0001$, as determined by a One-way ANOVA with multiple comparisons. **C)** *In vitro* cleavage of r(CUG)₁₀ by **4**, as determined using 5'-³²P-RNA and gel electrophoresis. Quantification of gel autoradiograms are reported as percent of r(CUG)₁₀ cleaved for each treatment group relative to vehicle-treated samples ($n = 3$); * $P = 0.0332$; *** $P = 0.0009$; **** $P < 0.0001$, as determined by a One-way ANOVA with multiple comparisons.

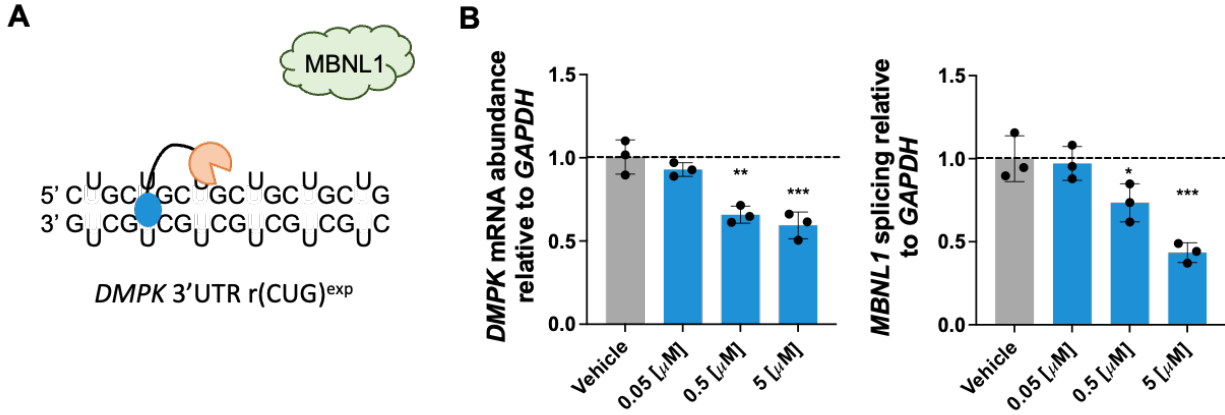


Figure 15. Design and validation of monomeric cleaver in patient-derived muscle cells. **A)** Effect of **4** on *DMPK* abundance, which harbors r(CUG)^{exp}, in DM1 myotubes as determined by RT-qPCR (n = 3 biological replicates). **B)** Effect of **4** on *DMPK* levels and *MBNL1* exon 5 inclusion in DM1 myotubes, as determined by RT-qPCR. For the latter, primers span the *MBNL1* exon 5 – exon 6 junction (n = 3); * $P = 0.0406$; ** $P = 0.001$; *** $P < 0.0005$, as determined by a One-way ANOVA with multiple comparisons.

Wang Lab Progress Report: University of Florida

For this project period, we have focused on tasks outlined in the statement of work from months 12-24. For our site, we proposed to:

- Aim 2, Task 3, Subtask 4: Complete a comprehensive analysis of transcriptome-and proteome-wide effects (months 6-36)
- Aim 2, Task 3, Subtask 7: Compare effects of small molecule-bleomycin conjugates and RIBOTACS probes in vivo to ASOs (months 6-36)

We have generated RNA-seq libraries from DM1 patient-derived and WT unaffected myotubes treated with a variety of compounds. These include cells treated with **037-Bleo**, **2H2-C2**, ASOs, and DMSO (vehicle), of which the first two have been described above. We have generated 21 RNA-seq libraries in total. We describe in detail analyses of a triplicate set of DM1 and WT myotubes treated with DMSO and **037-Bleo** (12 libraries total):

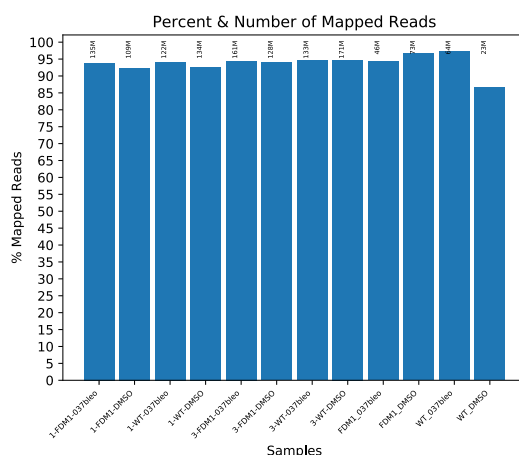


Figure 16. Percent and number of mapped reads for a subset of RNA-seq libraries in which WT or DM1 myotubes were treated with DMSO or 037-Bleo.

Reads were mapped to the hg19 genome, gene expression was quantitated by Kallisto, and isoforms were quantitated by MISO. Plotted below are MA plots for each pairwise comparison between samples; the high degree of correlations between samples both support a high degree of reproducibility as well as the idea that global gene expression changes are not extremely dramatic in response to small molecule treatment.

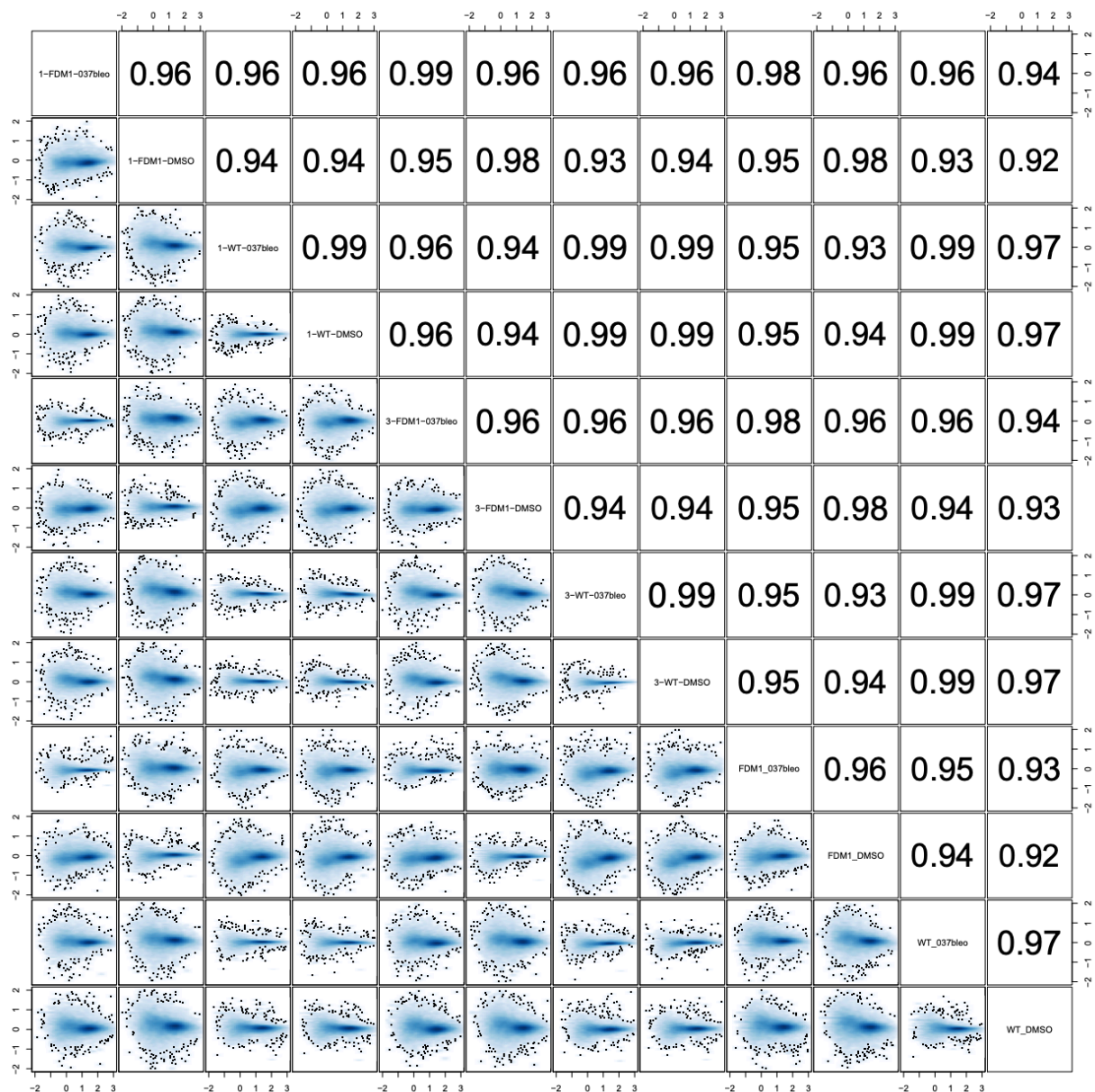


Figure 17. MA Plots for pairwise comparisons between all samples using TPM (transcript per million) values as estimated from RNA-seq. Correlation values are shown above the diagonal.

To analyze splicing patterns, we computed psi values for all alternative splicing events; we used all three replicates for each condition for this analysis. Using stringent thresholds ($|\Delta \text{psi}| > 0.1$, MISO Bayes Factor > 5), we found 122 events significantly regulated between WT DMSO and DM1 DMSO myotubes, and 100 events significantly regulated between WT DMSO and DM1 037-Bleo myotubes.

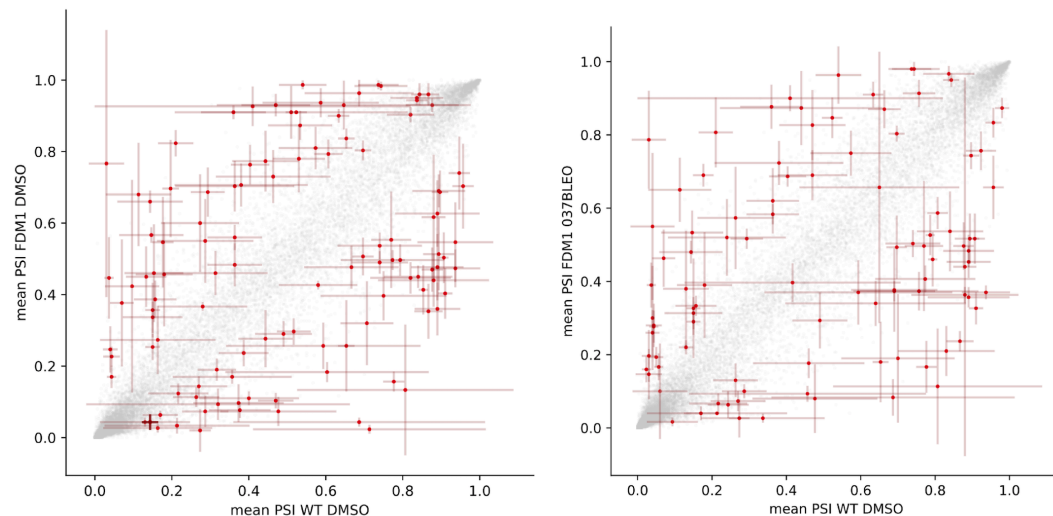


Figure 18. Scatter of mean psi between selected conditions. The mean and standard deviation in psi for significantly regulated events is highlighted in red; non-regulated events are plotted in gray.

Future studies for this project will continue to evaluate transcriptome-wide responses to small molecule treatments in DM1 cells and DM1 animal models. Our infrastructure and analytical pipelines are well equipped to assess both on- and off-target effects in these systems in collaboration with the Disney lab.

SPECIFIC AIM 3: Targeted small molecule recruitment of a nuclease to r(CUG)^{exp}.

Note: The section below summarizes activities completed by the Disney Laboratory (TSRI).

A potential approach to direct malfunctioning RNAs down endogenous decay pathways is to exploit ribonucleases (RNases) and recruit them to specific transcripts with a small molecule. RNase L, an integral part of the viral immune response, is present in minute quantities in all cells as an inactive monomer. Upon activation of the immune system, RNase L is upregulated and 2'-5' oligoadenylate [2'-5'poly(A)] is synthesized; binding of 2'-5'poly(A) dimerizes and activates RNase L.⁹⁰ Due to the ubiquitous nature of this system, we sought to assemble active RNase L onto a specific RNA target to cleave it, akin to antisense.⁹¹ Indeed, we have shown that this approach, dubbed Ribonuclease targeting chimeras (RIBOTACs), allows for the selective cleavage of a desired transcript. Herein, we propose to develop RIBOTACs to selectively degrade r(CUG)^{exp} by appending compounds identified in Specific Aim 1a (simple binding compounds) with RNase L-recruiting modules. These RIBOTACs will be evaluated in DM1 patient-derived fibroblasts and myotubes and *in vivo*.

During the last funding period, we have focused our efforts on optimizing the RNase L recruiting molecule, as we were seeing much less cleavage of r(CUG)^{exp} by the RIBOTAC than we have observed for other targets. To that end, we have completed a high throughput screen to identify RNase L binding and activating small molecules. In brief, we employed an *in vitro*, fluorescence-based assay in which an RNA with multiple RNase L cleavage site is labeled with a fluorophore and a quencher on the 5' and 3' ends, respectively (**Figure 19A**). Upon cleavage, the fluorophore and quencher are no longer proximal and fluorescence increases. Here, we took a 2-pronged approach, screening derivatives of previous RNase L recruiter as well as new chemical matter (**Figure 19B**), affording 15 hit compounds (increased fluorescence <3 σ of the mean). In secondary validation, four of these compounds increased fluorescence dose dependently (**Figure 19C**) and oligomerized RNase L (**Figure 19D**). The structures of the four compounds are shown in Figure XE.

We are currently synthesizing RIBOTACs in which the RNase L recruiting molecule is systematically changed to one of these four molecules (Figure XE) and other analogs of our original recruiter. To assess their ability to cleave r(CUG)^{exp}, we will: (i) use the fluorescence-based, *in vitro* assay described above with r(CUG) repeats labeled with a fluorophore and quencher on the 5' and 3' ends, respectively; (ii) verify cleavage and nucleotides that are cleaved by using 5'-³²P-labeled r(CUG) repeats and gel electrophoresis; and (iii) measure *DMPK* transcript levels upon treatment of DM1 patient-derived myotubes.

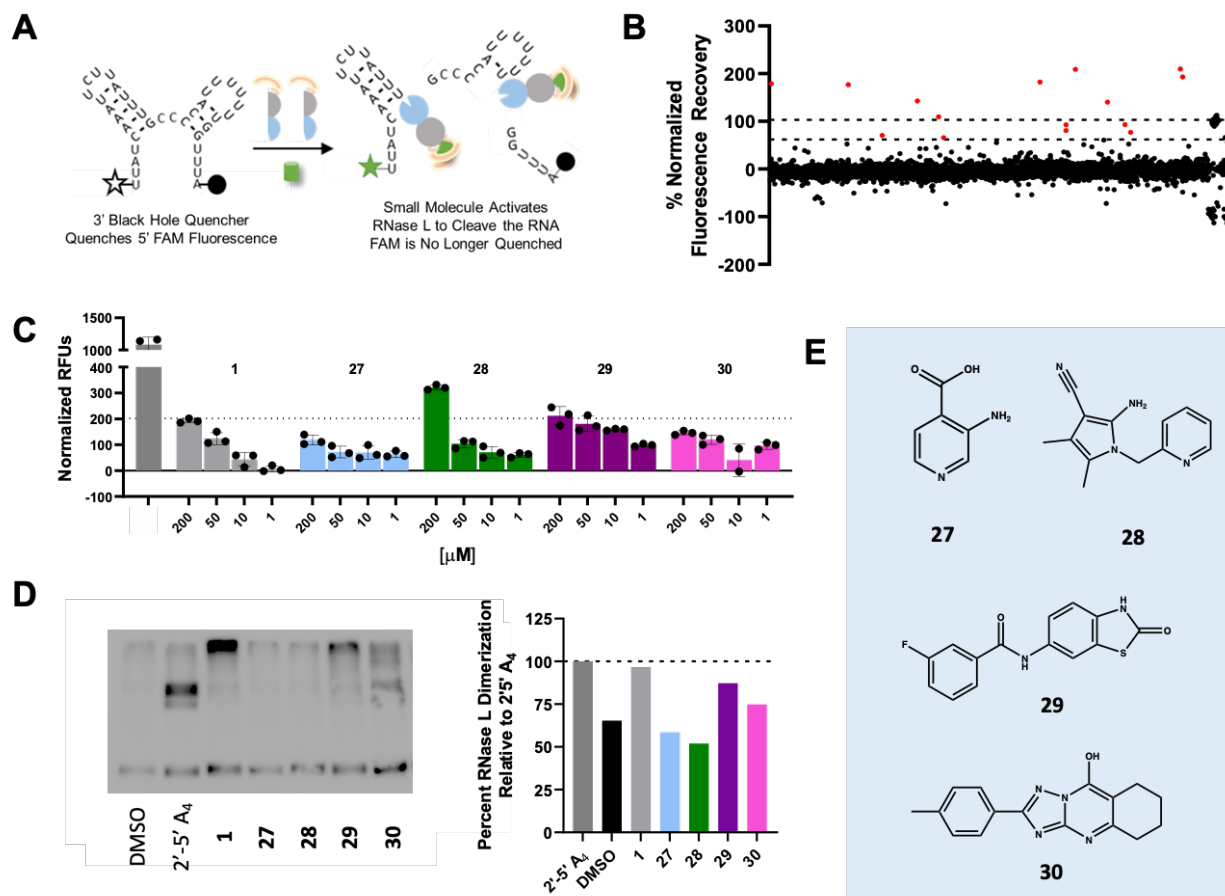


Figure 19. Chemical fragment library screening to identify novel chemical matter capable of recruiting and activating RNase L. **A)** Schematic depicting the fluorescence-based *in vitro* RNA cleavage assay that was utilized in this high-throughput screen (Z factor = 0.52). **B)** Fifteen hit compounds were identified by the high-throughput screen (>3 S.D. hit cut-off; 0.5% hit rate). **C)** Four hit compounds showed dose-dependent RNA cleavage in the fluorescence-based *in vitro* cleavage assay. **D)** The same four hits oligomerized RNase L *in vitro*, as determined by Western blotting. **E)** Structures of the four hits identified by the high-throughput screen that showed dose-dependent RNA cleavage and *in vitro* RNase L oligomerization.

What opportunities for training and professional development has the project provided?

The Scripps Research Institute (Disney): Annual performance reviews and Individual Development Plans (IDPs) are widely recognized as effective tools for setting and achieving Ph.D.-level training goals. They also encourage productive communication between trainees and their mentors. The Scripps Research Institute strongly encourages postdocs such as Dr. Raphael Benhamou to create and revisit IDPs, and to seek regular feedback on their performance from their mentor (Prof. Matthew Disney). IDP templates are available from the institute's Career and Postdoctoral Services Office website and are provided to trainees as part of their initial onboarding process. The Career and Postdoctoral Services Office also arranges biannual IDP workshops to help trainees such as Dr. Benhamou interpret self-assessment information, explore career options, and set goals using myIDP from AAAS/ScienceCareers.org. Dr. Disney worked with Dr. Benhamou to create his own personalized IDP for this project. To provide necessary feedback, TSRI also recommends that mentors conduct annual reviews with their assigned postdocs to discuss lab obligations, research goals, skills development, and career planning. Dr. Disney and Dr. Benhamou have consulted together frequently in this regard.

The University of Florida (Wang): This award provided opportunities for staff to maintain the skills required to generate RNA-seq libraries as well as to perform computational analysis of data.

How were the results disseminated to communities of interest?

Results were disseminated in the usual way by publishing in peer-reviewed journals.

What do you plan to do during the next reporting period to accomplish the goals?

Specific Aim 1: In the next year, we plan to initiate and/or complete the following tasks and subtasks in Aim 1:

Major Task 3: In vitro evaluation of hit compounds from DEL.

- Subtask 2: In vitro evaluation of most potent compounds from Subtask 1: affinity, k_{on} , k_{off} , residence time by biolayer interferometry; **Disney Laboratory**

Specific Aim 2: In the next year, we plan to initiate and/or complete the following tasks and subtasks in Aim 2:

Major Task 2: Complete DMPK studies of lead molecules.

- Subtask 1: In vitro DMPK analysis; **Months 1-30: 09/01/19-02/28/22; Disney Laboratory**
- Subtask 2: Mouse pharmacokinetics and tissue distribution; **Months 1-30: 09/01/19-02/28/22 Disney Laboratory**

- Subtask 3: Study lung fibrosis (not expected); **Months 1-30: 09/01/19-02/28/22 Disney Laboratory**

Major Task 3: Determine the optimal dosing regimen of optimal bleomycin conjugates.

- Subtask 1: Study myotonia over different dosages and treatment periods, informed by **Specific Aim 2, Major Task 2; Months 6-36: 02/01/20-08/30/22; Disney Laboratory**
- Subtask 2: Study improvement of splicing defects and formation of foci over different dosages and treatment periods; **Months 6-36: 02/01/20-08/30/22; Disney Laboratory**
- Subtask 3: Study lung fibrosis (not expected; bleomycin conjugates) over different dosages and treatment periods; **Months 6-36: 02/01/20-08/30/22; Disney Laboratory**
- Subtask 4: Complete a comprehensive analysis of transcriptome- and proteome-wide effects; **Months 6-36: 02/01/20-08/30/22; Wang Laboratory**
- Subtask 5: Compare effects of small molecule-bleomycin conjugates and RIBOTACS probes in vivo to ASOs; **Months 6-36: 02/01/20-08/30/22; Wang Laboratory**

Specific Aim 3: In the next year, we plan to initiate and/or complete the following tasks and subtasks in Aim 3:

Major Task 4: Comprehensive in cell evaluation of RIBOTAC probes

- Subtask 1: Determine if RIBOTACs selectively cleave r(CUG)^{exp} in multiple DM1 patient-derived cell lines; **Months 18-30: 02/01/21-02/28/22; Disney Laboratory**
- Subtask 2: Study compounds for improving DM1-associated splicing defects and reducing foci in multiple cell lines; **Months 18-30: 02/01/21-02/28/22; Disney Laboratory**
- Subtask 3: Complete a comprehensive analysis of transcriptome-wide and proteome-wide (including immune system) effects of compound treatment; **Months 18-30: 02/01/21-02/28/22; Disney Laboratory**

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

It is estimated that there are 30,000 non-redundant human mRNAs⁹² and >700 novel conserved RNA structures, almost all with unknown function, in the mammalian transcriptome.⁹³ An extraordinarily challenging problem is developing general methods to target selectively defective or malfunctioning RNAs that cause disease. Current therapeutic strategies to target RNAs are based on specific sequence recognition by oligonucleotides. However, many human disorders are caused by highly structured RNAs not readily targetable by conventional base pairing. We are implementing a radical new approach for reprogramming or eliminating the highly structured RNA

that causes DM1 with a small molecule directly or by recruiting endogenous cellular machinery to degrade the RNA.

Our chemical approach employs bi-functional, cell-permeable small molecules composed of structural RNA-binding modules coupled to a small molecule cleaving or protein-recruiting module. This modular design will enable rapid assembly and identification of molecules capable directly cleaving the DM1 RNA or of recruiting endogenous enzymes to eliminate the RNA by nucleolytic cleavage. Such studies are advancing new therapeutic strategies for the treatment of DM1.

What was the impact on other disciplines?

Our highly unconventional approach has the potential to establish a completely new paradigm for studying human disease pathology caused by toxic structured RNAs. Such diseases include numerous other expanded microsatellite repeat disorders such as, Huntington's disease (HD), fragile X-associated tremor ataxia syndrome (FXTAS), and amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD), as well as viral infections such as HIV and HCV.

What was the impact on technology transfer?

NOTHING TO REPORT

What was the impact on society beyond science and technology?

Collectively, our studies are changing how pharmaceutical companies view RNA, traditionally as poor drug targets, and how they approach tackling diseases caused by RNA. We have expanded the mode of action of drugs from simple binding to binding and cleavage, eliminating toxic RNAs directly or by recruiting a cellular protein to degrade it. That is, our studies are opening up new therapeutic pipelines, approaches, and strategies for the treatment of many diseases, which will have a positive impact on human health in the form of precision medicines.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

NOTHING TO REPORT

Actual or anticipated problems or delays and actions or plans to resolve them

NOTHING TO REPORT

Changes that had a significant impact on expenditures

NOTHING TO REPORT

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

N/A

Significant changes in use or care of vertebrate animals

NOTHING TO REPORT

Significant changes in use of biohazards and/or select agents

NOTHING TO REPORT

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal Publications

New During Current Funding Period

1. Vezini-Dawod S, Angelbello AJ, Choudhary S, Won Wang K, Yildirim I, **Disney MD**. Massively parallel optimization of the linker domain in small molecule dimers targeting a toxic r(CUG) repeat expansion. *ACS Med Chem Lett* (2021), 12, 907-914. doi: 10.1021/acsmchemlett.1c00027; acknowledges federal support (yes).
2. Baisden T Jared, Childs-Disney, Lucas S Ryan, **Disney MD**. Affecting RNA biology genome-wide by binding small molecules and chemically induced proximity. *Curr Opin Chem Biol* (2021), 62, 119-129. doi: 10.1016/j.cbpa.2021.03.006; acknowledges federal support (yes).
3. Benhamou RI, Abe M, Choudhary S, Meyer SM, Angelbello AJ, **Disney MD**. Optimization of the linker domain in a dimeric compound that degrades an r(CUG) repeat expansion in cells. *J Med Chem* (2020), 63, 7827-7839; published; acknowledges federal support (yes).

Reported Last Funding Period

4. Angelbello AJ, DeFeo ME, Glinkerman CM, Boger DL, **Disney MD**. Precise targeted cleavage of a r(CUG) repeat expansion in cells by using a small-molecule-deglycobleomycin conjugate. *ACS Chem Biol* (2020), 15, 849-855; published; acknowledges federal support (yes).

5. Costales MG, Childs-Disney JL, Haniff HS, **Disney MD**; How we think about targeting RNA with small molecules. *J Med Chem*, (2020), 63, 8880-8900; acknowledges federal support (yes).

6. Meyer SM, Williams CC, Akahori Y, Tanaka T, Aikawa H, Tong Y, Childs-Disney JL, **Disney MD**. Small molecule recognition of disease-relevant RNA structures. *Chem Soc Rev* (2020), 49, 7167-7199; acknowledges federal support (yes).

7. Ursu A, Childs-Disney JL, Andrews RJ, O'Leary CA, Meyer SM, Angelbello AJ, Moss WN, **Disney MD**. Design of small molecules targeting RNA structure from sequence. *Chem Soc Rev* (2020), 49, 7252-7270; acknowledges federal support (yes).

8. Fitzgerald PR & Paegel BM; *DNA-Encoded Chemistry: Drug Discovery From a Few Good Reactions*; ACS Chemical Reviews, (2020), 121, 7155-7177; acknowledges federal support (yes).

Books or other non-periodical, one-time publications.

NOTHING TO REPORT

Other publications, conference papers and presentations.

NOTHING TO REPORT

• **Website(s) or other Internet site(s)**

NOTHING TO REPORT

• **Technologies or techniques**

NOTHING TO REPORT

• **Inventions, patent applications, and/or licenses**

NOTHING TO REPORT

• **Other Products**

NOTHING TO REPORT

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: Matthew D. Disney

Project Role: Principal Investigator (Initiating PI)

Researcher Identifier: 0000-0001-8486-1796

Nearest Person Month Worked: 1 calendar month

Contribution to Project: Professor Disney oversees all aspects of the Initiating project.

Funding Support: in addition to this award, active awards include W81XWH2010727; UG3/UH3 NS116921; R35 NS116846; R01 CA249180; and P01 NS099114.

Name: Shruti Choudary

Project Role: Postdoctoral Associate, Disney Lab

Researcher Identifier: 0000-0002-3373-0080

Nearest Person Month Worked: 7.3 calendar months

Contribution to Project: Completed experimental work associated with Disney Laboratory milestones/tasks

Funding Support: N/A

Name: Xueyi Yang

Project Role: Postdoctoral Associate, Disney Lab

Researcher Identifier: 0000-0003-3876-1852

Nearest Person Month Worked: 1.2 calendar months

Contribution to Project: Completed experimental work associated with Disney Laboratory milestones/tasks

Funding Support: N/A

Name: Samantha Meyer

Project Role: Graduate Student- Div. 3, Disney Lab

Researcher Identifier: 0000-0003-3353-6939

Nearest Person Month Worked: 0.92 calendar months

Contribution to Project: Completed experimental work associated with Disney Laboratory milestones/tasks

Funding Support: N/A

Name: Christopher Williams

Project Role: Graduate Student- Div. 3, Disney Lab

Researcher Identifier: 0000-0001-7503-0149

Nearest Person Month Worked: 0.92 calendar months

Contribution to Project: Completed experimental work associated with Disney Laboratory milestones/tasks

Funding Support: N/A

Name: Yuquan Tong

Project Role: Graduate Student- Div. 3, Disney Lab

Researcher Identifier: 0000-0003-0793-7747

Nearest Person Month Worked: 0.92 calendar months

Contribution to Project: Completed experimental work associated with Disney Laboratory milestones/tasks

Funding Support: N/A

Name: Eric Wang

Project Role: Principal Investigator (Partnering PI)

Researcher Identifier: 0000-0003-2655-5525

Nearest Person Month Worked: 1 calendar month

Contribution to Project: Professor Wang oversees all aspects of the Partnering project.

Funding Support: in addition to this award, active awards include R01 AR076060; R01 AG058636; R01 AR060209; R01 GM121862; R01 NS112291; and R01 NS114253.

Name: Hailey Olafson

Project Role: Computational Biologist, Wang Lab

Researcher Identifier: N/A

Nearest Person Month Worked: 8 calendar months

Contribution to Project: RNA-seq and proteomics analysis of small molecules targeting $r(\text{CUG})^{\text{exp}}$

Funding Support: N/A

Name: Kendra McKee

Project Role: Laboratory Manager, Wang Lab

Researcher Identifier: N/A

Nearest Person Month Worked: 7 calendar months

Contribution to Project: RNA-seq and proteomics analysis of small molecules targeting $r(\text{CUG})^{\text{exp}}$

Funding Support: N/A

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

New or ended since last year in budget period

Matthew D. Disney

New Awards Received Throughout Budget Period (2020-2021):

New

W81XWH2010727 (Disney/Lairson) 09/01/20-08/31/23 0.76 calendar (6.33%)

DoD/PRMRP

Disney lab share: (total direct costs for project) | (total costs for project) **Small Molecules That Target the RNAs That Cause Pulmonary Fibrosis and Polycystic Kidney Disease**

Department of the Army

US Army Medical Research Acquisition Activity

820 Chandler Street

Fort Detrick, MD 21702-5014

Program Official: Christopher L. Baker, christopher.l.baker132.civ@mail.mil

This project aims to provide a potentially transformative approach to treating pulmonary fibrosis and polycystic kidney disease (PCKD) by developing and optimizing novel small molecule compounds for use in RNA-targeting therapeutics.

Specific Aim 1: Develop and study small molecules that cleave the pri-miR-17/92 cluster, the upregulation of which causes PCKD.

Specific Aim 2: Develop and study small molecules that cleave pre-miR-21, the upregulation of which causes PCKD and lung fibrosis.

Specific Aim 3: Study the optimal compounds emerging from Aims 1 & 2 for improving PCKD and lung fibrosis in animal models of each disease.

Overlap: None

Special Funding Project 0205 12/22/2020-12/22/2022 0.12 calendar (1.00%)

(total costs for project)

Developing Pre-Clinical Candidates Against miRNAs Implicated in Autoimmune Diseases

EMD Serono

1 Technology Place

Rockland, MA 02370

Program Official: Sophia Perides, Sophia.perides@emdserono.com

New or ended since last year in budget period

Eric Wang

R01 AR076060 (Lott/Wang) 08/01/2020-06/01/2023 0.12 calendar (1.00%)

Wang Lab: per year for 3 years

Development of MRI, Alternative Splicing, and Functional Abilities as Biomarkers in Myotonic Dystrophy Type 1

NIAMS

6701 Democracy Blvd

Building: Democracy I, Suite 800

Bethesda MD 20892-4872

Program Official: Emily Carifi, emily.carifi@nih.gov

The objective of this proposal is to develop splicing and MRI biomarkers for DM1.

Overlap: None

Service Agreement (Wang) 03/05/2021-03/04/2022 No effort requested
(total costs for project)

Evaluation of peptide-oligonucleotide conjugates for myotonic dystrophy

Entrada Therapeutics

50 Northern Avenue

Boston, MA 02210

Program Official: Dipal Doshi, dipal@entradatx.com

What other organizations were involved as partners?

<i>Organization Name:</i>	The University of Florida
<i>Location of Organization:</i>	Center for NeuroGenetics Dept. of Molecular Genetics & Microbiology 2033 Mowry Road Gainesville, FL 32610
<i>Partner's Contribution:</i>	Collaboration (academic worksite of Partnering PI Eric Wang)

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: In this collaborative award, Initiating and Partnering PI are submitting a joint progress report.

QUAD CHARTS: N/A

9. APPENDICES: Please see attached appendices with published articles.

REFERENCES

1. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706-709 (2014).
2. Lu, Z. & Chang, H.Y. Decoding the RNA structurome. *Curr. Opin. Struct. Biol.* **36**, 142-148 (2016).
3. Andrews, R.J., Baber, L. & Moss, W.N. RNAStructuromeDB: A genome-wide database for RNA structural inference. *Sci. Rep.* **7**, 17269 (2017).
4. Disney, M.D. Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *J. Am. Chem. Soc.* **141**, 6776-6790 (2019).
5. Donlic, A. & Hargrove, A.E. Targeting RNA in mammalian systems with small molecules. *Wiley Interdiscip. Rev.: RNA* **9**, e1477 (2018).
6. Connelly, C.M., Moon, M.H. & Schneekloth, J.S., Jr. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem. Biol.* **23**, 1077-1090 (2016).
7. Childs-Disney, J.L., Wu, M., Pushechnikov, A., Aminova, O. & Disney, M.D. A Small Molecule Microarray Platform To Select RNA Internal Loop–Ligand Interactions. *ACS Chem. Biol.* **2**, 745-754 (2007).
8. Ursu, A., Vézina-Dawod, S. & Disney, M.D. Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discov. Today* **24**, 2002-2016 (2019).
9. Haniff, H.S., Knerr, L., Chen, J.L., Disney, M.D. & Lightfoot, H.L. Target-Directed Approaches for Screening Small Molecules against RNA Targets. *SLAS Discov.* **25**, 869-894 (2020).
10. Costales, M.G. et al. Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 2406-2411 (2020).
11. Velagapudi, S.P. et al. Approved Anti-cancer Drugs Target Oncogenic Non-coding RNAs. *Cell Chem. Biol.* **25**, 1086-1094.e1087 (2018).
12. Groen, E.J.N., Talbot, K. & Gillingwater, T.H. Advances in therapy for spinal muscular atrophy: promises and challenges. *Nat. Rev. Neurol.* **14**, 214-224 (2018).
13. Kodadek, T., Paciaroni, N.G., Balzarini, M. & Dickson, P. Beyond protein binding: recent advances in screening DNA-encoded libraries. *Chem. Commun.* **55**, 13330-13341 (2019).
14. Madsen, D., Azevedo, C., Micco, I., Petersen, L.K. & Hansen, N.J.V. in *Progress in Medicinal Chemistry*, Vol. 59. (eds. D.R. Witty & B. Cox) 181-249 (Elsevier, 2020).
15. Litovchick, A. et al. Novel Nucleic Acid Binding Small Molecules Discovered Using DNA-Encoded Chemistry. *Molecules* **24**, 2026 (2019).
16. Ottl, J., Leder, L., Schaefer, J.V. & Dumelin, C.E. Encoded Library Technologies as Integrated Lead Finding Platforms for Drug Discovery. *Molecules* **24**, 1629 (2019).
17. Komnatnyy, V.V., Nielsen, T.E. & Qvortrup, K. Bead-based screening in chemical biology and drug discovery. *Chem. Commun.* **54**, 6759-6771 (2018).
18. Liu, T., Qian, Z., Xiao, Q. & Pei, D. High-Throughput Screening of One-Bead-One-Compound Libraries: Identification of Cyclic Peptidyl Inhibitors against Calcineurin/NFAT Interaction. *ACS Comb. Sci.* **13**, 537-546 (2011).
19. Sohrabi, C., Foster, A. & Tavassoli, A. Methods for generating and screening libraries of genetically encoded cyclic peptides in drug discovery. *Nat. Rev. Chem.* **4**, 90-101 (2020).

20. Simonetti, L. & Ivarsson, Y. Genetically encoded cyclic peptide phage display libraries. *ACS Cent. Sci.* **6**, 336-338 (2020).
21. Quartararo, A.J. et al. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. *Nat. Commun.* **11**, 3183 (2020).
22. Annis, D.A., Nickbarg, E., Yang, X., Ziebell, M.R. & Whitehurst, C.E. Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Curr. Opin. Chem. Biol.* **11**, 518-526 (2007).
23. Shi, Y. et al. Stabilization of lncRNA GAS5 by a Small Molecule and Its Implications in Diabetic Adipocytes. *Cell Chem. Biol.* **26**, 319-330.e316 (2019).
24. Brook, J.D. et al. 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 (1992).
25. Mathieu, J. & Prévost, C. Epidemiological surveillance of myotonic dystrophy type 1: A 25-year population-based study. *Neuromuscular Disord.* **22**, 974-979 (2012).
26. Machuca-Tzili, L., Brook, D. & Hilton-Jones, D. Clinical and molecular aspects of the myotonic dystrophies: A review. *Muscle Nerve* **32**, 1-18 (2005).
27. Taneja, K.L., McCurrach, M., Schalling, M., Housman, D. & Singer, R.H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* **128**, 995-1002 (1995).
28. Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T. & Thornton, C.A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13**, 3079-3088 (2004).
29. Li, J. et al. A dimeric 2,9-diamino-1,10-phenanthroline derivative improves alternative splicing in myotonic dystrophy type 1 cell and mouse models. *Chemistry* **24**, 18115-18122 (2018).
30. Rzuczek, S.G. et al. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **13**, 188 (2016).
31. Jenquin, J.R. et al. Furamidine rescues myotonic dystrophy type I associated mis-splicing through multiple mechanisms. *ACS Chem. Biol.* **13**, 2708-2718 (2018).
32. Rzuczek, S.G., Southern, M.R. & Disney, M.D. Studying a drug-like, RNA-focused small molecule library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. *ACS Chem. Biol.* **10**, 2706-2715 (2015).
33. Warf, M.B., Nakamori, M., Matthys, C.M., Thornton, C.A. & Berglund, J.A. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18551-18556 (2009).
34. Hagler, L.D. et al. Expanded DNA and RNA trinucleotide repeats in myotonic dystrophy type 1 select their own multitarget, sequence-selective inhibitors. *Biochemistry* **59**, 3463-3472 (2020).
35. Gareiss, P.C. et al. 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 (2008).
36. Angelbello, A.J. et al. Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7799-7804 (2019).

37. Pushechnikov, A. et al. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J. Am. Chem. Soc.* **131**, 9767-9779 (2009).
38. Rzuczek, S.G. et al. Features of modularly assembled compounds that impart bioactivity against an RNA target. *ACS Chem. Biol.* **8**, 2312-2321 (2013).
39. Benhamou, R.I. et al. Optimization of the linker domain in a dimeric compound that degrades an r(CUG) repeat expansion in cells. *J. Med. Chem.* **63**, 7827-7839 (2020).
40. Benhamou, R.I. et al. Macrocyclization of a ligand targeting a toxic RNA dramatically improves potency. *ChemBioChem* **21**, 3229-3233 (2020).
41. Fowler, S.A. & Blackwell, H.E. Structure–function relationships in peptoids: Recent advances toward deciphering the structural requirements for biological function. *Org. Biomol. Chem.* **7**, 1508-1524 (2009).
42. Miller, S.M. et al. Proteolytic studies of homologous peptide and N-substituted glycine peptoid oligomers. *Bioorg. Med. Chem. Lett.* **4**, 2657-2662 (1994).
43. Gao, Y., Amar, S., Pahwa, S., Fields, G. & Kodadek, T. Rapid lead discovery through iterative screening of one bead one compound libraries. *ACS Comb. Sci.* **17**, 49-59 (2015).
44. Chen, C.Z. et al. Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* **402**, 1889-1898 (2012).
45. Lee, M.M., French, J.M. & Disney, M.D. Influencing uptake and localization of aminoglycoside-functionalized peptoids. *Mol. Biosyst.* **7**, 2441-2451 (2011).
46. Gates, D.P., Coonrod, L.A. & Berglund, J.A. Autoregulated splicing of muscleblind-like 1 (MBNL1) pre-mRNA. *J. Biol. Chem.* **286**, 34224-34233 (2011).
47. Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **37**, 844-852 (2005).
48. Yudin, A.K. Macrocycles: lessons from the distant past, recent developments, and future directions. *Chem. Sci.* **6**, 30-49 (2015).
49. White, T.R. et al. On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat. Chem. Biol.* **7**, 810-817 (2011).
50. Stotani, S. & Giordanetto, F. Overview of Macrocycles in Clinical Development and Clinically Used. In *Practical Medicinal Chemistry with Macrocycles* (eds E. Marsault and M.L. Peterson). <https://doi.org/10.1002/9781119092599.ch16> (2017).
51. Gao, Y. & Kodadek, T. Direct comparison of linear and macrocyclic compound libraries as a source of protein ligands. *ACS Comb. Sci.* **17**, 190-195 (2015).
52. Celine, A. & Claudio, S. Converting a Peptide into a Drug: Strategies to Improve Stability and Bioavailability. *Curr. Med. Chem.* **9**, 963-978 (2002).
53. Mallinson, J. & Collins, I. Macrocycles in new drug discovery. *Future Med. Chem.* **4**, 1409-1438 (2012).
54. Rezai, T., Yu, B., Millhauser, G.L., Jacobson, M.P. & Lokey, R.S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J. Am. Chem. Soc.* **128**, 2510-2511 (2006).
55. Ahlbach, C.L. et al. Beyond cyclosporine A: conformation-dependent passive membrane permeabilities of cyclic peptide natural products. *Future Med. Chem.* **7**, 2121-2130 (2015).
56. Traboulsi, H. et al. Macrocyclic cell penetrating peptides: a study of structure-penetration properties. *Bioconj. Chem.* **26**, 405-411 (2015).
57. Qian, Z. et al. Discovery and mechanism of highly efficient cyclic cell-penetrating peptides. *Biochemistry* **55**, 2601-2612 (2016).

58. Tian, Y. et al. Achieving enhanced cell penetration of short conformationally constrained peptides through amphiphilicity tuning. *Chem. Sci.* **8**, 7576-7581 (2017).
59. Shortridge, M.D. et al. A macrocyclic peptide ligand binds the oncogenic microRNA-21 precursor and suppresses Dicer processing. *ACS Chem. Biol.* **12**, 1611-1620 (2017).
60. Mukherjee, S. et al. Structural insights into synthetic ligands targeting A–A pairs in disease-related CAG RNA repeats. *Nucleic Acids Res.* **47**, 10906-10913 (2019).
61. Childs-Disney, J.L., Hoskins, J., Rzuczek, S.G., Thornton, C.A. & Disney, M.D. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol.* **7**, 856-862 (2012).
62. Kumar, A. et al. Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. *ACS Chem. Biol.* **7**, 496-505 (2012).
63. Roth, B.M., Ishimaru, D. & Hennig, M. The core microprocessor component DiGeorge Syndrome Critical Region 8 (DGCR8) is a nonspecific RNA-binding protein. *J. Biol. Chem.* **288**, 26785-26799 (2013).
64. **Schrödinger Release 2020-1**: ConfGen, S., LLC, New York, NY, 2020. (
65. Khan, A.R. et al. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **37**, 16839-16845 (1998).
66. Driggers, E.M., Hale, S.P., Lee, J. & Terrett, N.K. The exploration of macrocycles for drug discovery — an underexploited structural class. *Nat. Rev. Drug Discovery* **7**, 608-624 (2008).
67. Kwon, Y.U., Kodadek, T., Kwon, Y.U. & Kodadek, T. Quantitative comparison of the relative cell permeability of cyclic and linear peptides. *Chem. Biol.* **14**, 671-677 (2007).
68. Yang, N.J. & Hinner, M.J. in *Site-Specific Protein Labeling: Methods and Protocols*. (eds. A. Gautier & M.J. Hinner) 29-53 (Springer New York, New York, NY; 2015).
69. Pettersson, O.J., Aagaard, L., Jensen, T.G. & Damgaard, C.K. Molecular mechanisms in DM1 — a focus on foci. *Nucleic Acids Res.* **43**, 2433-2441 (2015).
70. Konieczny, P., Stepniak-Konieczna, E., Taylor, K., Sznajder, Ł.J. & Sobczak, K. Autoregulation of MBNL1 function by exon 1 exclusion from MBNL1 transcript. *Nucleic Acids Res.* **45**, 1760-1775 (2016).
71. Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **37**, 844-852 (2005).
72. Holt, I. et al. Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells* **12**, 1035-1048 (2007).
73. Boger, D.L.a.C., H. Bleomycin: synthetic and mechanistic studies. *Angew. Chem. Int. Ed.* **38**, 448-476 (1999).
74. Schroeder, B.R. et al. The disaccharide moiety of bleomycin facilitates uptake by Cancer cells. *J. Am. Chem. Soc.* **136**, 13641-13656 (2014).
75. Bembenek, S.D., Tounge, B.A. & Reynolds, C.H. Ligand efficiency and fragment-based drug discovery. *Drug Discov. Today* **14**, 278-283 (2009).
76. Scott, D.E., Coyne, A.G., Hudson, S.A. & Abell, C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **51**, 4990-5003 (2012).
77. Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug. Discov.* **15**, 605-619 (2016).
78. Lamoree, B. & Hubbard, R.E. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **61**, 453-464 (2017).

79. Erlanson, D.A., Davis, B.J. & Jahnke, W. Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. *Cell Chem. Biol.* **26**, 9-15 (2019).
80. Pahl, A., Waldmann, H. & Kumar, K. Exploring Natural Product Fragments for Drug and Probe Discovery. *Chimia (Aarau)* **71**, 653-660 (2017).
81. Nasiri, H.R. et al. Targeting a c-MYC G-quadruplex DNA with a fragment library. *Chem. Commun.* **50**, 1704-1707 (2014).
82. Mortenson, P.N., Erlanson, D.A., de Esch, I.J.P., Jahnke, W. & Johnson, C.N. Fragment-to-lead medicinal chemistry publications in 2017. *J. Med. Chem.* **62**, 3857-3872 (2019).
83. Parker, C.G. et al. Ligand and target discovery by fragment-based screening in human cells. *Cell* **168**, 527-541 e529 (2017).
84. Wang, Y. et al. Expedited mapping of the ligandable proteome using fully functionalized enantiomeric probe pairs. *Nat. Chem.* **11**, 1113-1123 (2019).
85. Suresh, B.M. et al. A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 33197-33203 (2020).
86. Velagapudi, S.P., Li, Y. & Disney, M.D. A cross-linking approach to map small molecule-RNA binding sites in cells. *Bioorg. Med. Chem. Lett.* **29**, 1532-1536 (2019).
87. Guan, L. & Disney, M.D. Covalent small-molecule-RNA complex formation enables cellular profiling of small-molecule-RNA interactions. *Angew. Chem. Int. Ed. Engl.* **52**, 10010-10013 (2013).
88. Angelbello, A.J., DeFeo, M.E., Glinkerman, C.M., Boger, D.L. & Disney, M.D. Precise targeted cleavage of a r(CUG) repeat expansion in cells by using a small-molecule-deglycobleomycin conjugate. *ACS Chem. Biol.* **15**, 849-855 (2020).
89. Arandel, L. et al. Immortalized human myotonic dystrophy muscle cell lines to assess therapeutic compounds. *Dis. Model Mech.* **10**, 487-497 (2017).
90. Silverman, R.H. Viral Encounters with 2' ,5' -Oligoadenylate Synthetase and RNase L during the Interferon Antiviral Response. *J. Virol.* **81**, 12720-12729 (2007).
91. Torrence, P.F. et al. Recruiting the 2-5A system for antisense therapeutics. *Antisense Nucleic Acid Drug Dev.* **7**, 203-206 (1997).
92. Furey, T.S. et al. Analysis of human mRNAs with the reference genome sequence reveals potential errors, polymorphisms, and RNA editing. *Genome Res.* **14**, 2034-2040 (2004).
93. Parker, B.J. et al. New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. *Genome Res.* **21**, 1929-1943 (2011).

Optimization of the Linker Domain in a Dimeric Compound that Degrades an r(CUG) Repeat Expansion in Cells

Raphael I. Benhamou, Masahito Abe, Shruti Choudhary, Samantha M. Meyer, Alicia J. Angelbello, and Matthew D. Disney*

Cite This: *J. Med. Chem.* 2020, 63, 7827–7839

Read Online

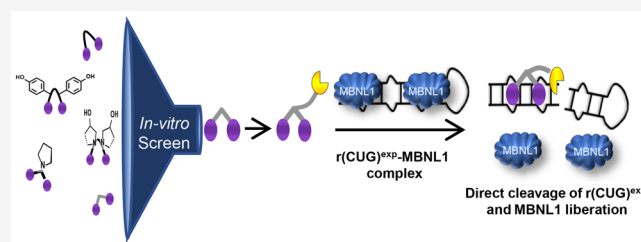
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: RNA repeat expansions are responsible for more than 30 incurable diseases. Among them is myotonic dystrophy type 1 (DM1), the most common form of adult on-set muscular dystrophy. DM1 is caused by an r(CUG) repeat expansion [r(CUG)^{exp}] located in the 3' untranslated region (UTR) of the dystrophin myotonic protein kinase gene. This repeat expansion is highly structured, forming a periodic array of 5'CUG/3'GUC internal loop motifs. We therefore designed dimeric compounds that simultaneously bind two of these motifs by connecting two RNA-binding modules with peptoid linkers of different geometries and lengths. The optimal linker contains two proline residues and enhances compound affinity. Equipping this molecule with a bleomycin A5 cleaving module converts the simple binding compound into a potent allele-selective cleaver of r(CUG)^{exp}. This study shows that the linker in modularly assembled ligands targeting RNA can be optimized to afford potent biological activity.



INTRODUCTION

To date, much of chemical biology and drug discovery for human diseases has been centered on targeting proteins.¹ However, recent estimates have suggested that the number of potential RNA drug targets far exceeds the number of potential protein drug targets.^{2,3} Currently, the most common approach for targeting RNA and affecting its biological function is by using oligonucleotide-based modalities to recognize sequences through base pairing.⁴ Many RNAs, however, contribute to human biology by forming particular structures,^{5–7} which are generally not targetable using oligonucleotides. In contrast, RNA structures are best targeted by small molecules that can bind to the shapes and features these structures display, much like a small molecule binding to a pocket in a protein.⁸

One major class of disease-causing RNA targets is repeat expansions. Repeat expansions have diverse and varied biological dysfunction and contribute to greater than 30 human diseases.⁹ In these microsatellite diseases, the RNA repeats generally fold into stable structures that undergo various processes contributing to disease pathology. For example, they can be translated without the use of a canonical start codon,¹⁰ participate in gain-of-function pathways by binding to and sequestering proteins involved in RNA biogenesis including pre-mRNA splicing,¹¹ and participate in other deleterious functions. One disease that is mediated by an RNA gain-of-function mechanism is myotonic dystrophy type 1 (DM1), the most common cause of adult on-set muscular dystrophy. DM1 is caused by a non-coding expansion of r(CUG) [hereafter, r(CUG)^{exp}] harbored in the 3' untrans-

lated region (UTR) of the dystrophin myotonic protein kinase (DMPK) gene.¹² This r(CUG)^{exp} folds into a highly stable structure with regularly repeating 5'CUG/3'GUC 1 × 1 nucleotide internal loops (Figure 1A). This structure binds to and sequesters the muscleblind-like 1 (MBNL1) protein, which regulates pre-mRNA splicing. Sequestration of MBNL1 by r(CUG)^{exp} results in pre-mRNA splicing defects in various genes, and these defects contribute to DM1 pathology.¹³

Previously, small molecules have been designed to selectively target r(CUG)^{exp}, liberating MBNL1 and thereby improving DM1-associated disease defects.^{14,15} One strategy to deactivate this target is the selective recognition of structural motifs in r(CUG)^{exp} with small molecules, which can be further derivatized into dimers that target multiple motifs simultaneously.¹⁵ These dimeric molecules have been optimized for binding repeating targets by altering the identity of the linker connecting the two RNA-binding motifs (N-methyl alanine or propylamine linkers)^{14,15} or attaching cellular uptake tags such as lysine and arginine,¹⁶ thus enhancing the uptake and localization of the compounds. Subsequently, uptake tags⁹ have also been used as spacing modules.¹⁷ The enhanced permeability of these previously developed uptake tags has

Received: April 13, 2020

Published: July 13, 2020



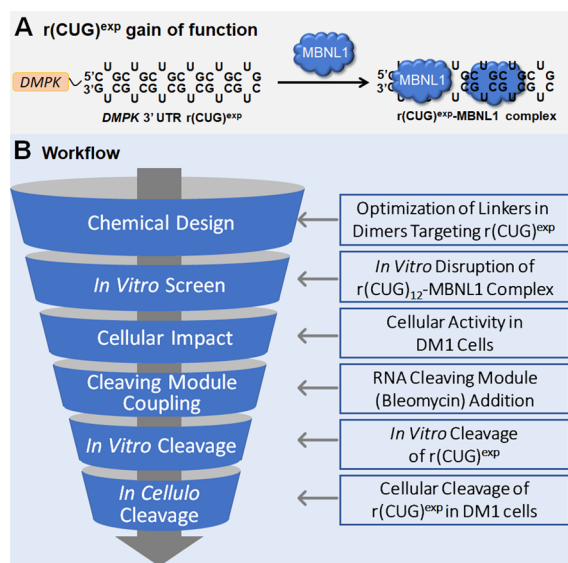


Figure 1. Design of potent dimers targeting $r(\text{CUG})^{\text{exp}}$ in DM1. (A) $r(\text{CUG})^{\text{exp}}$ in the 3' UTR of DMPK folds into a hairpin displaying a periodic array of internal loops that bind and sequester MBNL1, resulting in pre-mRNA splicing defects and the formation of nuclear foci. (B) Step-by-step methodology for the design of compounds targeting $r(\text{CUG})^{\text{exp}}$.

been traced to polycationic interactions with cell-surface heparin sulfate.¹⁸ However, the increased basicity of these compounds has also been correlated with promiscuity and off-target binding.^{19,20} Thus, the cationic nature of a compound must be properly balanced in order to maintain selectivity. Other methods to inhibit DM1 biology include targeting the encoding DNA and thus inhibiting transcription^{21,22} and degrading the RNA by appending cleaving modules (such as bleomycin A5) to structure-binding small molecules, which enhances both potency and selectivity.^{14,23}

Herein, we designed and synthesized a library of dimeric compounds by modifying the linker connecting the two RNA-binding modules. We show that a shorter and more rigid linker can improve both the affinity for $r(\text{CUG})^{\text{exp}}$ and cellular potency without increasing the cationic nature of the compound. The optimized binder for $r(\text{CUG})^{\text{exp}}$ was then appended with bleomycin A5, further enhancing potency, as compared to the parent compound in DM1 patient-derived myotubes (Figure 1B).

Results and Discussion. Previously, dimeric compounds comprising two copies of an RNA-binding module and a linker moiety were shown to be biologically active against $r(\text{CUG})^{\text{exp}}$. These molecules contain a peptoid backbone,¹⁵ which was optimized with an *N*-methyl peptide as a spacer between the two RNA-binding modules (2H-K4NMeS, Figure 2A).²⁴ Both the first generation compound and 2H-K4NMeS improve

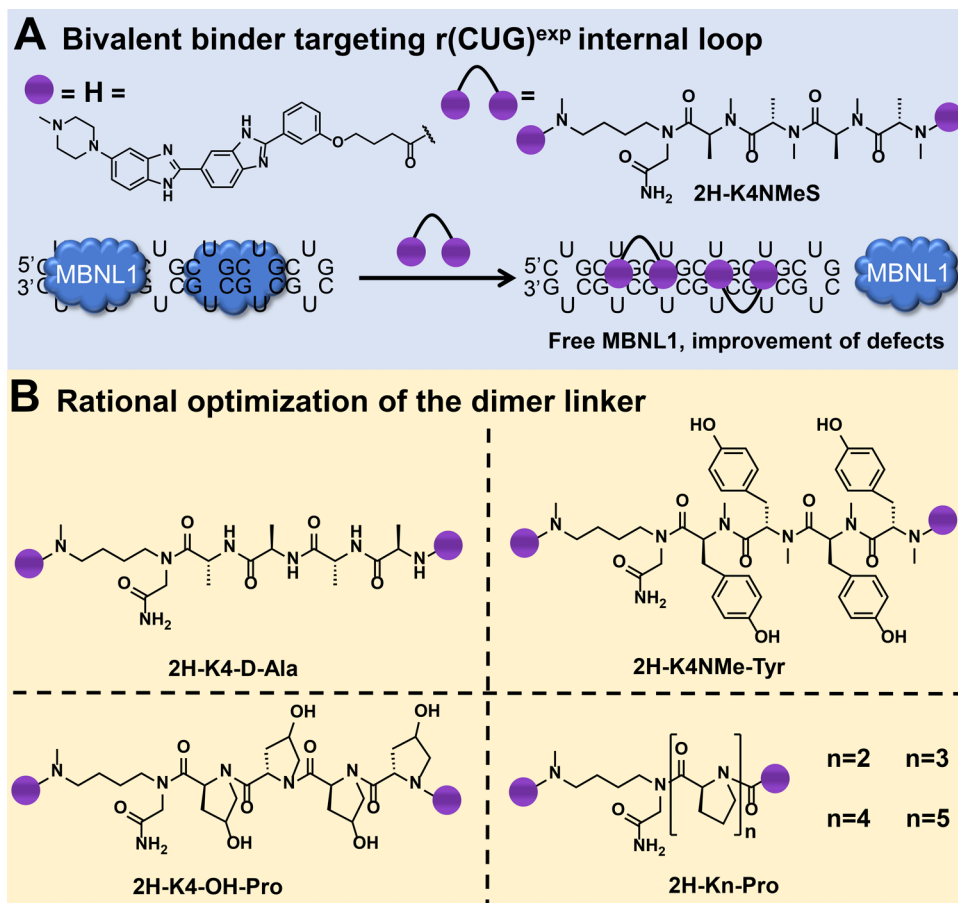


Figure 2. Design of small molecules that selectively target $r(\text{CUG})^{\text{exp}}$. (A) Bivalent small molecules linked together using an *N*-methyl peptide scaffold (2H-K4NMeS) with two H RNA-binding modules (purple spheres). The dimer binds two $r(\text{CUG})^{\text{exp}}$ loops simultaneously and releases MBNL1, improving DM1-associated defects. (B) Optimization of the peptoid linker connecting the two binding modules, generating seven new derivatives.

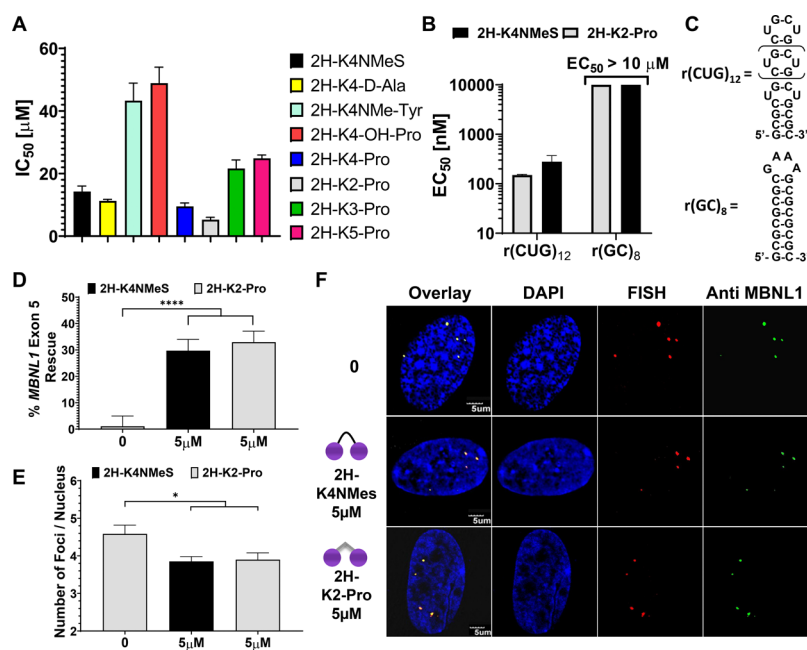


Figure 3. Studying the ability of designer compounds to inhibit the r(CUG)^{exp}-MBNL1 complex *in vitro* and assessment of compound activity in DM1 myotubes. (A) *In vitro* IC₅₀s for disruption of the r(CUG)^{exp}-MBNL1 complex by new dimer derivatives, compared to the parent compound ($n = 3$ replicates; 2 independent experiments). (B) Binding affinities (EC₅₀s) of 2H-K2-Pro and 2H-K4NMeS for a model of r(CUG)^{exp}, r(CUG)₁₂ ($n = 3$ replicates; 2 independent experiments), and a base-paired control ($n = 2$ replicates; 2 independent experiments). (C) Secondary structures of the r(CUG)₁₂ and r(GC)₈ constructs used for binding assays. (D–F) Studying the ability of 2H-K2-Pro to alleviate DM1-associated defects in DM1 myotubes, compared to 2H-K4NMeS. (D) Studying the ability of 2H-K2-Pro to rescue the MBNL1 exon 5 splicing defect ($n = 3$ replicates). (E) Quantification of the number of r(CUG)^{exp}-MBNL1 foci/nucleus. Error bars represent SD, $n = 3$ biological replicates, 40 nuclei counted/replicate. * $P < 0.05$, **** $P < 0.0001$, as determined by a one-way ANOVA. (F) Representative images of r(CUG)^{exp}-MBNL1 foci in DM1 myotubes treated with 2H-K2-Pro or 2H-K4NMeS.

various DM1-associated disease processes in cells; a bleomycin conjugate of 2H-K4NMeS provided a selective, bioactive compound in a pre-clinical animal model.²³ Given the success for improving compound properties by changing the spacer module and of other work that has shown that the identity of the spacer can affect affinity, uptake, and compound localization in various disease cell models,^{24,25} we further optimized the linker moiety by changing its length and geometry.

Design and *In Vitro* Evaluation of Dimeric Derivatives. A library of dimeric compounds displaying a 5' CUG/3' GUC RNA motif binding module (H) linked to D-alanine (D-Ala), tyrosine (Tyr), hydroxyproline (OH-Pro), and proline (Pro) spacers was synthesized using the optimal four-spacer module of previously reported 2H-K4NMeS (Figure 2A,B).¹⁴ These linker spacers were chosen in order to investigate the roles of the side-chain orientation with D-Ala, the rigidity of the linker with Tyr, OH-Pro, and Pro spacers, or the addition of hydrogen bond donors in 2H-K4-Tyr or 2H-K4-OH-Pro. The *in vitro* IC₅₀s of these dimers were measured as assessed by disruption of the r(CUG)₁₂-MBNL1 complex using a previously reported time-resolved fluorescence resonance energy transfer (TR-FRET) assay.^{26,27} Two dimers were modestly more potent than 2H-K4NMeS, namely 2H-K4-D-Ala and 2H-K4-Pro (Figure 3A).

Compared to other amino acids in a peptide backbone that prefer a *trans* orientation of the amide bond, proline undergoes an equilibrium between *cis*- and *trans*- isomers as a result of its structure where the residue itself forms a part of the backbone.^{28,29} The implications of this *cis*-*trans* isomerization in biological systems translates to distinct conformational flexibility, hydrogen bonding interactions, hydrophobic inter-

actions, and solvation,³⁰ all of which influence the binding of a ligand to its target and hence its bioactivity.

We therefore explored the proline linker further by varying the number of proline spacers ($n = 2, 3$, and 5) in the linker (Figure 2B) and measuring their IC₅₀s in the TR-FRET assay (Figure 3A). Notably, 2H-K2-Pro containing two prolines in the linker was the most potent with an IC₅₀ of $5.3 \pm 0.8 \mu\text{M}$, an ~ 3 -fold improvement over 2H-K4NMeS (Figure 3A). Interestingly, 2H-K2-Pro has a shorter linker and is more rigid than the original dimer 2H-K4-Pro, contributing to a favorable bioactive conformation. Moreover, physical properties of the compounds, such as logP, hydrogen bond donors, and topological polar surface areas (TPSAs), were calculated, and no significant differences were observed among the proline derivatives (Supporting Information, Dataset 1).

The binding affinity and selectivity of 2H-K2-Pro were next evaluated using a direct binding assay, measuring the change in the inherent fluorescence of the RNA-binding modules as a function of the RNA concentration, with either r(CUG)₁₂ or an RNA hairpin containing eight GC base pairs [r-(GCGCGCGCGAAAGCGCGCGC); dubbed r(GC)₈]. The 2H-K2-Pro bound to r(CUG)₁₂ with an EC₅₀ of $150 \pm 5 \text{ nM}$, while binding was not observed to the GC base-paired control RNA (Figure 3B and Figure S1). Additionally, 2H-K2-Pro showed an ~ 2 -fold greater affinity for r(CUG)₁₂ than the parent compound 2H-K4NMeS (EC₅₀ = $280 \pm 90 \text{ nM}$), while having less atoms in the linker.¹²

***In Cellulis* Evaluation of the Designed Dimer to Alleviate DM1-Associated Defects.** Given that 2H-K2-Pro is a potent and specific binder to r(CUG)^{exp} and capable of disruption of a pre-formed complex between r(CUG)^{exp} and

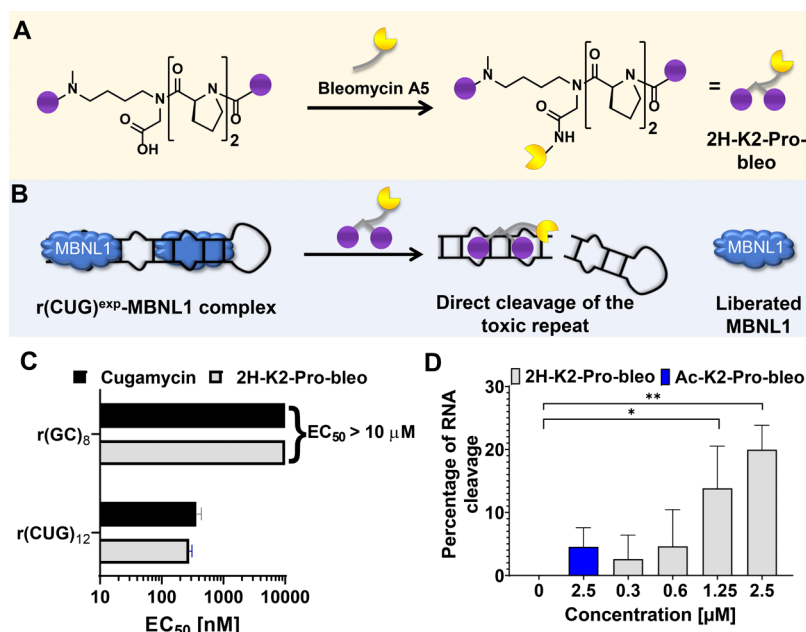


Figure 4. Design and *in vitro* evaluation of small molecules that cleave $r(CUG)^{exp}$. (A) The binder **2H-K2-Pro** was conjugated to bleomycin to afford the cleaver **2H-K2-Pro-bleo**. (B) **2H-K2-Pro-bleo** binds to $r(CUG)^{exp}$, cleaves the toxic RNA repeat, and displaces MBNL1, thereby improving DM1-associated defects. (C) Binding affinities of **2H-K2-Pro-bleo** and Cugamycin to $r(CUG)_{12}$ ($n = 3$ replicates; 2 independent experiments) and $r(GC)_8$ ($n = 2$ independent experiments with two replicates each). (D) *In vitro* RNA cleavage of $r(CUG)_{12}$ ($n = 2$ replicates; 2 independent experiments). * $P < 0.05$, ** $P < 0.01$ as determined by a one-way ANOVA relative to 0.

MBNL1 *in vitro*, the compound was evaluated in DM1-patient derived cells for rescue of two disease-associated defects: (i) dysregulation of alternative pre-mRNA splicing regulated by MBNL1 and (ii) the presence of $r(CUG)^{exp}$ -containing nuclear foci.^{31–33} We therefore studied the rescue of these DM1-associated defects by **2H-K2-Pro**. DM1 patient-derived fibroblasts harboring 1300 repeats were differentiated into myotubes³⁴ as a robust model of human disease. We first assessed improvement of the MBNL1-dependent splicing of its own exon 5;³⁵ in DM1-affected cells, exon 5 is included too frequently, 50% versus 15% in healthy cells as previously observed.²³ Notably, **2H-K2-Pro** rescued splicing of MBNL1 exon 5 in a dose-dependent fashion (Figure S2). At the 5 μ M dose, both **2H-K2-Pro** and **2H-K4NMeS** improved the MBNL1 exon 5 splicing defect similarly by ~30% (Figure 3D and Figure S2). Likewise, both compounds reduced the number of foci containing $r(CUG)^{exp}$ [imaged by RNA fluorescence *in situ* hybridization (FISH)] and MBNL1 (imaged using immunofluorescence) similarly when DM1 myotubes were treated with 5 μ M compound (Figure 3E,F).

A potential reason that the ~3-fold enhancement in **2H-K2-Pro**'s *in vitro* potency was not recapitulated in cells could be the differences in cell permeability or subcellular localization. Thus, the cellular uptake of **2H-K2-Pro** and **2H-K4NMeS** was measured using the inherent fluorescence of the RNA-binding modules. After treatment, DM1 myotubes were washed and lysed followed by measurement of fluorescence in the lysate. A standard curve was created by spiking in varying concentrations of the compound into lysate from untreated myotubes. Interestingly, the uptake of the two compounds was similar (Figure S2C,D). Next, we studied subcellular localization via live-cell fluorescence microscopy. Microscopy studies revealed that **2H-K4NMeS** was predominantly localized in the nucleus, whereas **2H-K2-Pro** was both cytoplasmic and nuclear (Figure S2C), which could contribute to a reduction in the bioactivity

of **2H-K2-Pro** as $r(CUG)^{exp}$ is sequestered in the nucleus in foci.

As the Pro-spacing module changed the subcellular localization of **2H-K4NMeS** from primarily nuclear to nuclear and cytoplasmic, we investigated how other spacing modules affected cell permeability, localization, and bioactivity, namely, **2H-K4-D-Ala**, **2H-K4-OH-Pro**, and **2H-K4-Tyr**. **2H-K4-Tyr** was found to be toxic in DM1 myotubes at 5 μ M and was not investigated further (Figure S3A). Interestingly, **2H-K4-OH-Pro** was taken up into cells at a higher concentration than that of **2H-K4NMeS**, while the permeability of **2H-K4-D-Ala** was similar to that of **2H-K4NMeS** (Figure S3B). Further investigation of the cellular localization of **2H-K4-D-Ala** and **2H-K4-OH-Pro** via live-cell fluorescence microscopy revealed that **2H-K4-OH-Pro** was localized only in the cytoplasm while **2H-K4-D-Ala** was localized in the cytoplasm and the nucleus (Figure S3C). Neither of these compounds improved MBNL1 exon 5 splicing in DM1 myotubes, likely due to the combination of the less favorable subcellular localization and linker length (Figure S3D). Thus, although **2H-K2-Pro** is localized in both the cytoplasm and the nucleus, its improved *in vitro* potency contributes to its cellular activity.

Design of Compounds that Cleave $r(CUG)^{exp}$. To enhance the activity of **2H-K2-Pro**, we converted the binder into an RNA degrader by using a bleomycin conjugation strategy.^{14,23,36,37} That is, **2H-K2-Pro** was conjugated to bleomycin A5 via its amine (known to contribute to DNA binding³⁸) to yield **2H-K2-Pro-bleo** (Figure 4A,B).²³ The affinity of **2H-K2-Pro-bleo** was similar to that of **2H-K2-Pro** with an EC_{50} of 280 ± 30 nM. [Note: binding assays were completed in the absence of Fe(II) such that the compound does not cleave the RNA.] No binding of **2H-K2-Pro-bleo** was observed to the GC-paired control RNA, $r(GC)_8$ (Figure 4C and Figure S4).

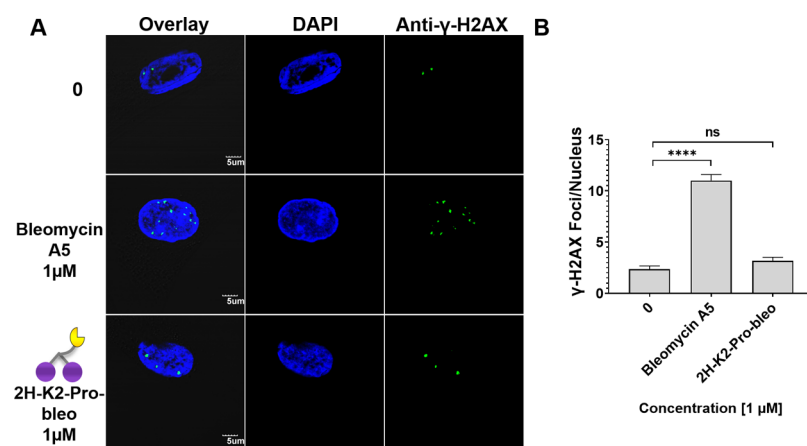


Figure 5. Effects of small molecules on the DNA damage response pathway. (A) Representative images from γ -H2AX immunofluorescence to assess DNA damage in DM1 myotubes upon treatment with 2H-K2-Pro-bleo and Bleomycin A5. (B) Quantification of the number of γ -H2AX foci/nucleus. Error bars represent SD; $n = 3$ biological replicates with 40 nuclei counted/replicate. **** $P < 0.0001$ as determined by one-way ANOVA; “ns” denotes not statistically significant.

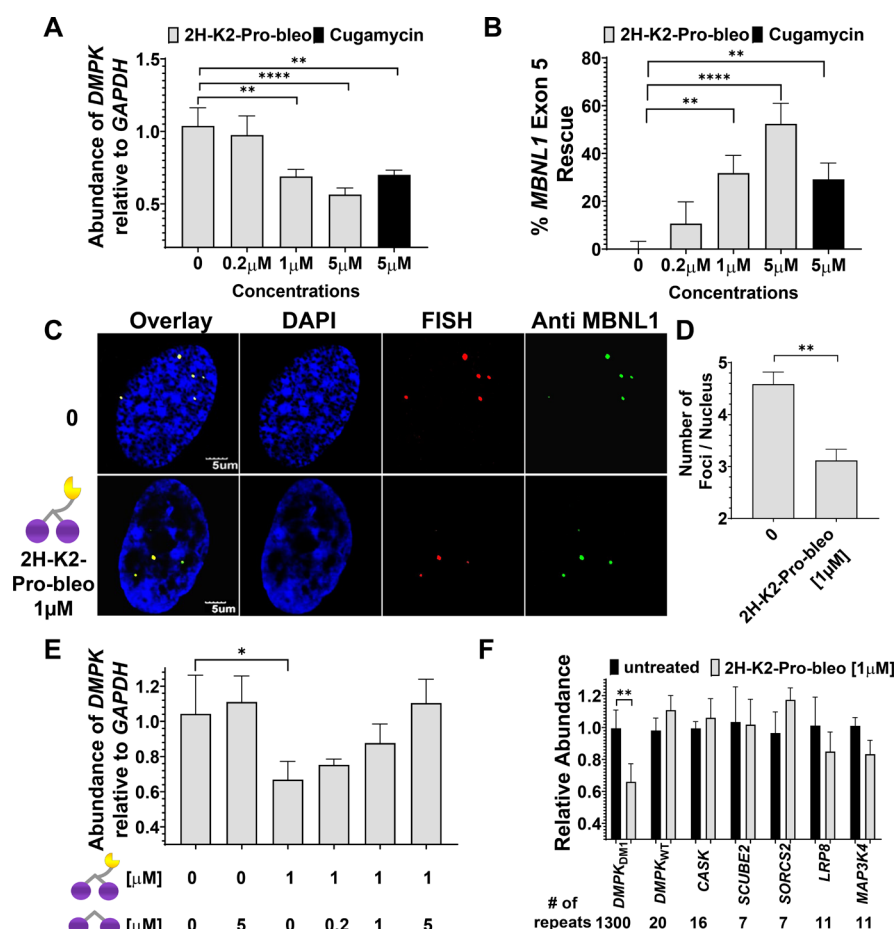


Figure 6. Biological activity of 2H-K2-Pro-bleo in DM1 patient-derived myotubes. (A) Effect of 2H-K2-Pro-bleo on r(CUG)^{exp}-containing DMPK levels in DM1 myotubes compared to Cugamycin, as determined by RT-qPCR. Error bars represent SD, $n = 3$ biological replicates. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$ (one-way ANOVA). (B) 2H-K2-Pro-bleo rescue of the MBNL1 exon 5 splicing defect compared to Cugamycin, as measured by RT-qPCR. Error bars represent SD, $n = 3$ biological replicates. ** $P < 0.01$, **** $P < 0.0001$ (one-way ANOVA). (C) Representative images of r(CUG)^{exp}-MBNL1 foci in DM1 myotubes treated with 2H-K2-Pro-bleo. (D) Quantification of the r(CUG)^{exp}-MBNL1 foci/nucleus. Error bars represent SD, $n = 3$ biological replicates, 40 nuclei counted/replicate. ** $P < 0.01$ (Student t -test). (E) Results of a cellular competitive cleavage assay between 2H-K2-Pro-bleo and 2H-K2-Pro on DMPK levels, as measured by RT-qPCR. Error bars represent SD, $n = 3$ biological replicates. * $P < 0.05$ (one-way ANOVA). (F) Effect of 2H-K2-Pro-bleo (1 μ M) on the abundance of mRNAs containing more than 6 but less than 20 r(CUG) repeats expressed in DM1 myotubes, as determined by RT-qPCR. Error bars represent SD, $n = 3$ biological replicates; 2 independent experiments. ** $P < 0.01$, (Student t -test).

In Vitro Cleavage of r(CUG)₁₀. In order to assess the ability of **2H-K2-Pro-bleo** to cleave r(CUG)^{exp} *in vitro*, a control compound, **Ac-K2-Pro-bleo**, was synthesized where bleomycin A5 was coupled to the peptoid backbone lacking the RNA-binding modules (Supporting Information, Synthetic Procedures). **2H-K2-Pro-bleo** dose-dependently cleaved r(CUG)₁₀, cleaving ~20% of the RNA at the highest concentration tested, 2.5 μ M (Figure 4D and Figure S5A,B). No significant RNA cleavage was observed with **Ac-K2-Pro-bleo** as expected (Figure 4D and Figure S5A,B).^{14,23}

Evaluation of DNA Damage in Cells. We have previously shown that by attaching r(CUG)^{exp}-binding modules to the C-terminal amine of bleomycin A5 a key positive charge on bleomycin A5 that contributes to DNA binding is eliminated along with its ability to cleave DNA.²³ To confirm that a similar effect is observed for **2H-K2-Pro-bleo**, we measured the amount of the phosphorylated form of H2A histone family member X (γ -H2AX) foci, formed in response to DNA double strand breaks.³⁹ Indeed, **2H-K2-Pro-bleo** did not cause a significant increase in γ -H2AX in DM1 myotubes, whereas bleomycin A5 alone showed an ~6-fold increase in the number of foci observed per cell (Figure 5 and Figure S6) as expected based on its ability to cause DNA damage in cancer cells.³⁹ Thus, DNA recognition and cleavage by bleomycin A5 are significantly reduced when conjugated to **2H-K2-Pro**.

Biological Evaluation of Compounds that Cleave r(CUG)^{exp}. After confirming *in vitro* activity, the ability of **2H-K2-Pro-bleo** to improve DM1-associated defects in myotubes was assessed. First, cleavage of the mutant allele [*DMPK* mRNA harboring r(CUG)₁₃₀₀ in the 3' UTR] by **2H-K2-Pro-bleo** was measured by RT-qPCR and compared to the previously reported cleaver, Cugamycin (**2H-K4NMeS-bleo**).²³ Notably, **2H-K2-Pro-bleo** reduced *DMPK* levels by 45 $\pm$ 4% at 5 μ M, whereas only an ~30 $\pm$ 3% decrease was observed with Cugamycin at 5 μ M (Figure 6A). Importantly, no effect was observed on *DMPK* levels in healthy myotubes that only express wild-type (WT) *DMPK* [r(CUG)₂₀ in the 3' UTR] (Figure S7), indicating that **2H-K2-Pro-bleo** specifically recognizes and cleaves the mutant r(CUG)^{exp}-containing *DMPK* transcript. To further explain the gain in potency by **2H-K2-Pro-bleo**, we evaluated its and Cugamycin's cellular uptake into DM1 myotubes. Interestingly, a statistically significant increase in permeability was observed for **2H-K2-Pro-bleo** ($P < 0.05$; Figure S8). However, compared to the parent binders, both bleomycin A5 conjugates were ~4-fold less permeable (Figure S8).

Next, the ability of **2H-K2-Pro-bleo** to rescue DM1-associated splicing defects and the formation of nuclear foci were investigated. **2H-K2-Pro-bleo** rescued the *MBNL1* exon 5 splicing defect by 50 $\pm$ 8% at 5 μ M and 30 $\pm$ 7% at 1 μ M (Figure 6B), the latter of which is similar to 5 μ M **2H-K2-Pro** (Figure 3D). Thus, in this assay, **2H-K2-Pro-bleo** is approximately 5 times more potent than **2H-K2-Pro**. Importantly, **2H-K2-Pro-bleo** did not affect the alternative splicing of mitogen-activated protein kinase kinase kinase 4 (*MAP4K4*) exon 22a, which is NOVA-, not *MBNL1*-, dependent (Figure S9). Likewise, **2H-K2-Pro-bleo** did not change *MBNL1* exon 5 splicing patterns in wild-type myotubes (Figure S7). Altogether, these control experiments indicate specificity for r(CUG)^{exp}. Next, we evaluated the ability of **2H-K2-Pro-bleo** (1 μ M) to reduce the number of r(CUG)^{exp}-*MBNL1* foci in cells. A significant decrease in the number of foci per cell was observed (Figures 6C,D and Figure S10).

To assess target engagement, competitive cleavage assays between the cleaver, **2H-K2-Pro-bleo**, and either **2H-K4NMeS** or **2H-K2-Pro** were completed in DM1 myotubes. In these experiments, cells were co-treated with excess **2H-K4NMeS** or **2H-K2-Pro** and 1 μ M **2H-K2-Pro-bleo**; if **2H-K4NMeS** and **2H-K2-Pro** bind to the same site in r(CUG)^{exp} as **2H-K2-Pro-bleo**, then cleavage of mutant *DMPK* should be reduced. As expected, both binding compounds reduced the cleavage of *DMPK* dose-dependently. However, **2H-K2-Pro** (EC_{25} = 0.2 μ M) inhibited *DMPK* cleavage more efficiently than **2H-K4NMeS** (EC_{25} = 1 μ M) and restored levels to those observed in untreated cells (Figure 6E and Figure S11).

To evaluate selectivity, the ability of **2H-K2-Pro-bleo** to discriminate between disease-causing r(CUG)^{exp} and short non-pathological r(CUG) repeats in other transcripts was measured. Notably, only the *DMPK* levels from DM1-patient derived myotubes were significantly decreased after treatment with the compound, while the levels of the other transcripts remained unchanged (Figure 6F). It is important to note that the r(CUG) repeat number found in these other transcripts is less than 20 so they do not fold into a hairpin structure. That is, the 3D structures of these RNAs are different than those of r(CUG)^{exp}, as shown in our previous folding analysis.²³ As a control, we also evaluated **Ac-K2-Pro-bleo**, the analog of **2H-K2-Pro-bleo** lacking the RNA-binding modules, in DM1 myotubes. As expected, no effect was observed on *DMPK* levels nor *MBNL1* splicing (Figure S12). Thus, cleavage of r(CUG)^{exp} is driven by the RNA-binding modules in **2H-K2-Pro-bleo**. **2H-K2-Pro-bleo** is able to selectively target the expanded disease-driving allele of *DMPK*, which in terms of selectivity is advantageous over sequence-based recognition of r(CUG)^{exp} with oligonucleotides as previously shown.²³

CONCLUSIONS

Small-molecule RNA cleavers may offer potent and selective alternatives to antisense oligonucleotides, which can have off-target effects as they recognize the RNA's sequence, not its structure. Additionally, RNA cleavers such as **2H-K2-Pro-bleo** offer a generalized approach for selectively affecting r(CUG)^{exp} disease biology across several disorders, such as myotonic dystrophy type 1 (DM1), Fuchs endothelial corneal dystrophy (FECD),⁴⁰ and Huntington's disease-like 2 (HDL-2).⁴¹ In contrast, oligonucleotide-based modalities do not target r(CUG)^{exp}; instead, they target the coding region of genes that harbor the repeat expansion. Thus, oligonucleotides have to be customized for each disease even though they are caused by the same repeat. Ligands targeting RNA structure, however, could prove to be general across these repeat-associated diseases.

Overall, this study demonstrated that altering the peptoid backbone of dimeric RNA-targeting compounds can alter bioactivity, thus creating an easy avenue for optimization. In addition, this work further supports the idea that appending a cleaver to RNA-binding small molecules can be used broadly to target and cleave structured RNAs associated with repeat expansion disorders selectively.

EXPERIMENTAL SECTION

General Synthetic Procedures. Bleomycin A5 (Bleocin) was purchased from EMD Millipore and used without further purification. Hoechst carboxylic acid, 4-(3-(6-(4-methylpiperazin-1-yl)-1H,3'H-[2,5'-bibenzo[d]imidazol]-2'-yl)phenoxy)butanoic acid (**6**) (Ht-CO₂H), intermediate (**10**), and **2H-K4NMeS** were synthesized as

reported previously.¹⁴ Peptide synthesis reactions were monitored by a chloranil test. Preparative HPLC was performed using a Waters 1525 Binary HPLC pump equipped with a Waters 2487 dual absorbance detector system and a Waters Sunfire C18 OBD 5 μ m 19 $\times$ 150 mm column. Absorbance was monitored at 345 and 220 or 254 nm. A gradient of methanol in water with 0.1% TFA varied in each purification. Purity was assessed by analytical HPLC using a Waters Symmetry C18 5 μ m 4.6 $\times$ 150 mm column, and a linear gradient of 0–100% methanol in water with 0.1% TFA over 60 min. Absorbance was monitored at 345 and 254 or 220 nm. All compounds evaluated had $\geq$ 95% purity (see characterization of small molecules). Mass spectrometry was performed with an Applied Biosystems MALDI ToF/ToF Analyzer 4800 Plus using an α -cyano-4-hydroxycinnamic acid matrix or an Agilent 1260 Infinity LC system coupled to an Agilent 6230 TOF (HR-ESI) with a Poroshell 120 EC-C18 column (Agilent, 50 mm $\times$ 4.6 mm, 2.7 μ m).

Synthesis of 2. A mixture of *tert*-butyl (4-(methylamino)butyl)-carbamate (3.00 g, 14.8 mmol), NEt₃ (4.13 mL, 29.7 mmol), and AllocCl (1.89 mL, 17.8 mmol) in CH₂Cl₂ (60 mL) was stirred at room temperature overnight. After completion of the reaction, the reaction was diluted with CH₂Cl₂ (100 mL), extracted with brine, and washed with 5% HCl aq (v/v, 15 mL) twice and aqueous NaHCO₃ (15 mL) once. The organic layer was concentrated to afford allyl (4-((*tert*-butoxycarbonyl)amino)butyl)(methyl)carbamate. To the intermediate, 90 mL of a 1:1 mixture of TFA:CH₂Cl₂ was added, and the reaction was stirred at room temperature overnight. The reaction mixture was concentrated, and then NaHCO₃ (aq; 15 mL) was added. The mixture was then extracted with CH₂Cl₂ (5 $\times$ 15 mL) to give **2** (759 mg, 27% yield over 2 steps). This compound was used for the next reaction without further purification. ¹H NMR (600 MHz, CDCl₃) δ ppm: 5.95–5.87 (m, 1H), 6.01–5.61 (br, 2H), 5.28 (d, 1H, *J* = 17.2 Hz), 5.20 (d, 1H, *J* = 10.3 Hz), 4.58–4.50 (m, 2H), 3.32–3.23 (m, 2H), 2.98–2.92 (m, 2H), 2.89 (s, 3H), 1.69–1.55 (m, 2H) ¹³C NMR (150 MHz, CDCl₃) δ ppm: 154.8, 131.0, 115.5, 64.3, 46.3, 37.6, 32.1, 22.6, 22.4 HR-MS (ESI): Calculated for [C₉H₁₉N₂O₂]⁺, 187.1441; found, 187.1445.

Synthesis of 3. Rink amide resin (**1**) (2.00 g, 1.0 mmol) was swollen in DMF (16 mL) at room temperature for 10 min and then deprotected with a solution of 20% piperidine in DMF (16 mL, 2 $\times$ 20 min). The resin was washed with DMF (3 $\times$ 8 mL). To the resin, bromoacetic acid (0.50 M, 10 mL, 5.0 mmol), DIC (0.774 mL, 5.0 mmol), and oxyma (0.711 g, 5.0 mmol) in DMF (4 mL) were added. The mixture was shaken at room temperature for 2 h. After, the resin was washed with CH₂Cl₂ (3 $\times$ 8 mL) and DMF (3 $\times$ 8 mL). To the resin, DMF (4 mL) and **2** (0.447 g, 2.40 mmol) in DMF (4 mL) was added. The mixture was shaken at room temperature overnight. The resin was then washed with CH₂Cl₂ (3 $\times$ 8 mL) and DMF (3 $\times$ 8 mL). Again, to the resin, DMF (4 mL) and **2** (0.447 g, 2.40 mmol) in DMF (4 mL) was added. The mixture was shaken at room temperature overnight. The resin was washed with CH₂Cl₂ (3 $\times$ 8 mL) and DMF (3 $\times$ 8 mL) to afford intermediate **3**.

Synthesis of 2H-K4-D-Ala. Resin **3** (20 mg, 10 μ mol) was swollen in DMF (500 μ L) for 10 min, and DMF was removed. To the resin, a pre-mixed solution of Fmoc-D-Ala-OH (27.0 mg, 50 μ mol), HATU (19.0 mg, 50 μ mol), HOAt (6.8 mg, 50 μ mol), and DIPEA (16.5 μ L, 100 μ mol) in DMF (500 μ L) was added, and the reaction was shaken at room temperature for 30 min. The resin was then washed with DMF (5 $\times$ 1 mL). To the resin, 20% piperidine in DMF (500 μ L) was added, and the resin was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 $\times$ 1 mL). This cycle was repeated an additional 2 times. Following this, a pre-mixed solution of Fmoc-D-Ala-OH (27.0 mg, 50 μ mol), HATU (19.0 mg, 50 μ mol), HOAt (6.8 mg, 50 μ mol), and DIPEA (16.5 μ L, 100 μ mol) was added to the resin, which was then shaken at room temperature for 30 min. The resin was washed with DMF (5 $\times$ 1 mL). To this resin, Pd(PPh₃)₄ (5.78 mg, 5.0 μ mol) in CH₂Cl₂ (500 μ L) and PhSiH₃ (24.7 μ L, 200 μ mol) and CH₂Cl₂ (1 mL) were added, and the mixture was shaken at room temperature for 40 min. This step was repeated once more. After that, the resin was washed with CH₂Cl₂ (5 $\times$ 1 mL, 30 s), 0.5% DIPEA (v/v) in DMF (2 $\times$ 1

mL, 30 s) and 0.5% sodium diethyl dithiocarbamate (w/v) in DMF (2 $\times$ 1 mL, 30 s), and DMF (5 $\times$ 1 mL, 30 s). To the resin, 20% piperidine in DMF (500 μ L) was added, and the reaction was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 $\times$ 1 mL). These two steps were repeated. To the resin, a pre-mixed solution of (Ht-CO₂H) (**6**) (25.5 mg, 50 μ mol), HATU (19.0 mg, 50 μ mol), HOAt (6.8 mg, 50 μ mol), and DIPEA (16.5 μ L, 100 μ mol) in 500 μ L of DMF was added, and it was shaken at 45 $^{\circ}$ C for 40 min. The resin was washed with DMF (5 $\times$ 1 mL). To this resin, 1 mL of TFA/H₂O (95/5) was added. The mixture was shaken at room temperature for 1 h; after which, the solution was collected and concentrated. To the mixture was added 10 times the volume of Et₂O, and the resulting precipitate was collected. The precipitate was washed with Et₂O (5 $\times$ volume), and the supernatant was removed by decantation. This crude product was purified by HPLC (41–81% MeOH+0.1% TFA vs H₂O + 0.1% TFA, over 20 min) to give **2H-K4-D-Ala** (372 nmol, 1.86% yield); HR-MS (MALDI): Calculated for [C₇₇H₉₄N₁₉O₉]⁺, 1428.7476; found, 1428.7456.

Synthesis of 2H-K4NMe-Tyr. Resin **3** (20 mg, 10 μ mol) was swollen in DMF (500 μ L) for 10 min, and then the DMF was removed. To the resin, a pre-mixed solution of Fmoc-N-Me-Tyr(*t*Bu)-OH (23.7 mg, 50 μ mol), DIC (7.7 μ L, 50 μ mol), and oxyma (7.1 mg, 50 μ mol) in DMF (500 μ L) was added, and the resin was shaken at room temperature for 2 h. After washing the resin with DMF (5 $\times$ 1 mL), 20% piperidine in DMF (500 μ L) was added, followed by shaking at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 $\times$ 1 mL). This cycle was repeated an additional 2 times. To the resin, a pre-mixed solution of Fmoc-N-Me-Tyr(*t*Bu)-OH (23.7 mg, 50 μ mol), DIC (7.7 μ L, 50 μ mol), and oxyma (7.1 mg, 50 μ mol) in DMF (500 μ L) was added, and it was shaken at room temperature for 2 h. The resin was washed with DMF (5 $\times$ 1 mL). To this resin, Pd(PPh₃)₄ (5.78 mg, 5.0 μ mol) in CH₂Cl₂ (500 μ L), PhSiH₃ (24.7 μ L, 200 μ mol), and CH₂Cl₂ (1 mL) were added, and the mixture was shaken at room temperature for 40 min. This deprotection step was repeated once more. After that, the resin was washed with CH₂Cl₂ (5 $\times$ 1 mL, 30 s), 0.5% DIPEA (v/v) in DMF (2 $\times$ 1 mL, 30 s) and 0.5% sodium diethyl dithiocarbamate (m/v) in DMF (2 $\times$ 1 mL, 30 s), and DMF (5 $\times$ 1 mL, 30 s). To the resin, 20% piperidine/DMF (500 μ L) was added, and it was shaken at room temperature for 10 min twice. The resin was washed with DMF (5 $\times$ 1 mL). To the resin, a pre-mixed solution of (Ht-CO₂H) (**6**) (25.5 mg, 50 μ mol), HATU (19.0 mg, 50 μ mol), HOAt (6.8 mg, 50 μ mol), and DIPEA (16.5 μ L, 100 μ mol) in 500 μ L of DMF was added, and it was shaken at room temperature overnight. The resin was washed with DMF (5 $\times$ 1 mL) followed by addition of 1 mL of 1:1 TFA: H₂O. The mixture was shaken at room temperature for 1 h. After that, the solution was collected and concentrated. To the mixture was added 10 $\times$ the volume of Et₂O, and the resultant precipitate was collected. The precipitate was washed with Et₂O (5 $\times$ volume), and the supernatant was removed by decantation. This crude product was purified by HPLC (50–100% MeOH + 0.1% TFA/H₂O + 0.1% TFA, over 20 min) to give **2H-4N-Me-Tyr** (60 nmol, 0.60% yield) HR-MS (ESI): Calculated for [C₁₀₅H₁₁₉N₁₉O₁₃]²⁺, 926.9612; found, 926.9590.

Synthesis of 2H-K4-OH-Pro. Resin **3** (20 mg, 10 μ mol) was swollen in DMF (500 μ L) for 10 min, and then DMF was removed. To the resin, a pre-mixed solution of Fmoc-Hyp(*t*Bu)-OH (20.5 mg, 50 μ mol), DIC (7.7 μ L, 50 μ mol), and oxyma (7.1 mg, 50 μ mol) in DMF (500 μ L) was added followed by shaking at room temperature for 2 h. The resin was washed with DMF (5 $\times$ 1 mL), and then 20% piperidine in DMF (500 μ L) was added. The resin was shaken at room temperature for 10 min and then washed with DMF (5 $\times$ 1 mL). This cycle was repeated an additional 2 times. To the resin, a pre-mixed solution of Fmoc-Hyp(*t*Bu)-OH (20.5 mg, 50 μ mol), DIC (7.7 μ L, 50 μ mol), and oxyma (7.1 mg, 50 μ mol) was added, and it was shaken at room temperature for 2 h. The resin was washed with DMF (5 $\times$ 1 mL), and then Pd(PPh₃)₄ (5.78 mg, 5.0 μ mol) in CH₂Cl₂ (500 μ L), PhSiH₃ (24.7 μ L, 200 μ mol), and CH₂Cl₂ (1 mL) were added. The mixture was shaken at room temperature for 40 min.

This deprotection step was repeated once more. After that, the resin was washed with CH_2Cl_2 ($5 \times 1 \text{ mL}$, 30 s each), 0.5% DIPEA (v/v) in DMF ($2 \times 1 \text{ mL}$, 30 s each) and 0.5% sodium diethyl dithiocarbamate (m/v) in DMF ($2 \times 1 \text{ mL}$, 30 s each), and DMF ($5 \times 1 \text{ mL}$, 30 s each). To the resin, 20% piperidine in DMF (500 μL) was added, and it was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF ($5 \times 1 \text{ mL}$) followed by addition of a pre-mixed solution of (Ht- CO_2H) (**6**) (25.5 mg, 50 μmol), HATU (19.0 mg, 50 μmol), HOAt (6.8 mg, 50 μmol), and DIPEA (16.5 μL , 100 μmol) in 500 μL of DMF. The resin was shaken at room temperature overnight and then washed with DMF ($5 \times 1 \text{ mL}$). To this resin, 1 mL of 95:5 TFA/ H_2O was added. The mixture was shaken at room temperature for 1 h. After that, the solution was collected and concentrated. The mixture was added to 10 $\times$ the volume of Et_2O , and the resultant precipitate was collected. The precipitate was washed with 5 $\times$ the volume of Et_2O , and its supernatant was removed by decantation. The crude product was purified by HPLC (45–75% MeOH + 0.1% TFA vs H_2O + 0.1% TFA, over 20 min) to give **2H-K4-OH-Pro** (233 nmol, 2.33% yield). HR-MS (ESI): Calculated for $[\text{C}_{85}\text{H}_{103}\text{N}_{19}\text{O}_{13}]^{2+}$, 798.8986; found, 798.9019.

Synthesis of 2H-K2-Pro. Resin **10** (400 mg, 200 μmol) was swollen in DMF (8 mL) for 10 min, and DMF was removed. To this resin, a pre-mixed solution of Fmoc-L-Pro-OH (337 mg, 1000 μmol), HATU (380 mg, 1000 μmol), HOAt (136 mg, 1000 μmol), and DIPEA (331 μL , 2000 μmol) in DMF (4 mL) was added, and the mixture was shaken at room temperature for 30 min. The resin was washed with DMF ($5 \times 5 \text{ mL}$). To the resin, 20% piperidine in DMF (5 mL) was added, and the resin was shaken at room temperature for 10 min. These two steps were repeated. The resin then was washed with DMF ($5 \times 10 \text{ mL}$). This addition of Fmoc-L-Pro-OH was repeated. To the resin, 4.0 mL of 30% TFA in CH_2Cl_2 was added. The mixture was shaken at room temperature for 10 min to allow the peptoid to be cleaved from the resin. The solution was collected and precipitated in 10 $\times$ the volume of Et_2O . The precipitate was washed with Et_2O ($5 \times \text{v/v}$), and the supernatant was removed by decantation. To a solution of the product in DMF (2.0 mL), a mixture of Ht- CO_2H (**6**) (215 mg, 420 μmol), HATU (160 mg, 420 μmol), HOAt (57.2 mg, 420 μmol), and DIPEA (139 μL , 840 μmol) in DMF (2.0 mL) was added at room temperature. The mixture was stirred at 60 $^\circ\text{C}$ under microwave irradiation for 1 h. The mixture was concentrated under vacuum and then purified by flash chromatography on a C18 column (120 g size column, MeOH in H_2O , 0.1% TFA as an additive, 3 times) and then by HPLC (41–81% MeOH + 0.1% TFA/ H_2O + 0.1% TFA, over 40 min) to give **2H-K2-Pro** (192 nmol, 0.96%) HR-MS (MALDI): Calculated for $[\text{C}_{75}\text{H}_{88}\text{N}_{17}\text{O}_7]^+$, 1338.7047; found, 1338.7058.

Synthesis of 2H-K3-Pro. Resin **3** (20 mg, 10 μmol) was swollen in DMF for 10 min, and DMF was removed. To the resin, a pre-mixed solution of Fmoc-L-Pro-OH (10.1 mg, 30 μmol), HATU (11.4 mg, 30 μmol), HOAt (4.1 mg, 30 μmol), and DIPEA (9.9 μL , 60 μmol) in DMF (1 mL) was added, and the resin was shaken at room temperature for 30 min. The resin was washed with DMF ($5 \times 1 \text{ mL}$). To the resin, 20% piperidine in DMF (500 μL) was added, and the resin was shaken at room temperature for 10 min. The resin was washed with DMF ($5 \times 1 \text{ mL}$). This cycle was repeated an additional 2 times. To the resin, a pre-mixed solution of Fmoc-L-Pro-OH (10.1 mg, 30 μmol), HATU (11.4 mg, 30 μmol), HOAt (4.1 mg, 30 μmol), and DIPEA (9.9 μL , 60 μmol) was added followed by shaking at room temperature for 30 min. The resin was washed with DMF ($5 \times 1 \text{ mL}$). To this resin, $\text{Pd}(\text{PPh}_3)_4$ (5.78 mg, 5.0 μmol) in CH_2Cl_2 (500 μL) and PhSiH_3 (24.7 μL , 200 μmol) and CH_2Cl_2 (1 mL) were added, and the mixture was shaken at room temperature for 40 min. This deprotection step was repeated once more. After, the resin was washed with CH_2Cl_2 ($5 \times 1 \text{ mL}$, 30 s each), 0.5% DIPEA (v/v) in DMF ($2 \times 1 \text{ mL}$, 30 s each) and 0.5% sodium diethyl dithiocarbamate (w/v) in DMF ($2 \times 1 \text{ mL}$, 30 s each), and DMF ($5 \times 1 \text{ mL}$, 30 s each). To the resin, 20% piperidine in DMF (500 μL) was added, and the mixture was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF ($5 \times 1 \text{ mL}$) followed

by addition of a pre-mixed solution of Ht- CO_2H (**6**) (25.5 mg, 50 μmol), HATU (19.0 mg, 50 μmol), HOAt (6.8 mg, 50 μmol), and DIPEA (16.5 μL , 100 μmol) in 500 μL of DMF. The resin was then shaken at 45 $^\circ\text{C}$ for 40 min. The resin was washed with DMF ($5 \times 1 \text{ mL}$). To this resin, 1 mL of 95:5 TFA/ H_2O was added. The mixture was shaken at room temperature for 1 h. After that, the solution was collected and concentrated. To the mixture was added 10 $\times$ the volume of Et_2O , and the resultant precipitate was collected. The precipitate was washed with Et_2O (5 $\times$ volume), and the supernatant was removed by decantation. This crude product was purified by HPLC (42–82% MeOH + 0.1% TFA vs H_2O + 0.1% TFA, over 20 min) to give **2H-K3-Pro** (106 nmol, 1.06%) HR-MS (MALDI): Calculated for $[\text{C}_{80}\text{H}_{95}\text{N}_{18}\text{O}_8]^+$, 1435.7575; found, 1435.7566.

Synthesis of 2H-K4-Pro. Resin **10** (400 mg, 200 μmol) was swollen in DMF (8 mL) for 10 min, and DMF was removed. To this resin, a pre-mixed solution of Fmoc-L-Pro-OH (337 mg, 1000 μmol), HATU (380 mg, 1000 μmol), HOAt (136 mg, 1000 μmol), and DIPEA (331 μL , 2000 μmol) in DMF (1 mL) was added, and the mixture was shaken at room temperature for 30 min. The resin was washed with DMF ($5 \times 5 \text{ mL}$). To the resin, 20% piperidine in DMF (5 mL) was added, and the reaction was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF ($5 \times 10 \text{ mL}$). This cycle was repeated an additional three times. To 190 μmol of the resin, 4.0 mL of 30% TFA in CH_2Cl_2 was added. The mixture was shaken at room temperature for 10 min. After, the solution was collected and then added to 10 $\times$ the volume of Et_2O . The resulting precipitate was collected and washed with 5 $\times$ the volume of Et_2O . The supernatant was removed by decantation. The precipitate was dissolved in DMF (1.9 mL), and then a mixture of Ht- CO_2H (**6**) (204 mg, 399 μmol), HATU (152 mg, 399 μmol), HOAt (54.3 mg, 399 μmol), and DIPEA (132 μL , 798 μmol) in DMF (1.9 mL) was added at room temperature. The mixture was stirred at 50 $^\circ\text{C}$ under microwave irradiation for 1 h. The mixture was concentrated under vacuum and then purified by flash chromatography using a C18 column (120 g size column, MeOH in H_2O , 0.1% TFA as an additive) to give **2H-K4-Pro** (12.0 μM , 6.00%) HR-MS (MALDI): Calculated for $[\text{C}_{85}\text{H}_{102}\text{N}_{19}\text{O}_9]^+$, 1532.8102; found, 1532.8101.

Synthesis of 2H-K5-Pro. Resin **10** (50 mg, 25 μmol) was swollen in DMF (1 mL) for 10 min, and DMF was removed. To this resin, a pre-mixed solution of Fmoc-L-Pro-OH (42.2 mg, 125 μmol), HATU (47.5 mg, 125 μmol), HOAt (17.0 mg, 125 μmol), and DIPEA (41.3 μL , 250 μmol) in DMF (1 mL) was added, and the mixture was shaken at room temperature for 30 min. The resin was washed with DMF ($5 \times 1 \text{ mL}$), and then 20% piperidine in DMF (1 mL) was added with shaking for 10 min. This step was repeated. The resin was washed with DMF ($5 \times 1 \text{ mL}$). This cycle was repeated an additional 4 times. To 12.5 μmol of the resin, 1.0 mL of 30% TFA in CH_2Cl_2 was added. The mixture was shaken at room temperature for 10 min. After, the solution was collected and added to 10 $\times$ the volume of Et_2O . The resultant precipitate was collected and washed with Et_2O (5 $\times$ volume). The supernatant was removed by decantation. To a solution of this product in DMF (0.25 mL), a mixture of Ht- CO_2H (**6**) (16.0 mg, 31.3 μmol), HATU (11.9 mg, 31.3 μmol), HOAt (4.3 mg, 31.3 μmol), and DIPEA (10.3 μL , 62.5 μmol) in DMF (0.25 mL) was added at room temperature. The mixture was stirred at 50 $^\circ\text{C}$ under microwave irradiation for 30 min and then concentrated under vacuum. The mixture was purified by HPLC (49–69% MeOH + 0.1% TFA vs H_2O + 0.1% TFA, over 30 min) to give **2H-K5-Pro** (64 nmol, 0.26%) HR-MS (MALDI): Calculated for $[\text{C}_{85}\text{H}_{102}\text{N}_{19}\text{O}_9]^+$, 1629.8630; found, 1629.8601.

Synthesis of 13. A mixture of *tert*-butyl (4-(methylamino)butyl)-carbamate (2.00 g, 9.89 mmol), NEt_3 (2.76 mL, 19.8 mmol), and NaCl (2.27 g, 11.9 mmol) in CH_2Cl_2 (49 mL) was stirred at room temperature overnight. After completion of the reaction, the reaction was diluted with CH_2Cl_2 (100 mL) and washed with 5% HCl aq (v/v, 15 mL) twice and then aqueous NaHCO_3 (15 mL). The organic layer was concentrated. The crude material was purified by column chromatography (silica gel, ethyl acetate in hexanes) to give *tert*-butyl (4-((*N*-methyl-4-nitrophenyl)sulfonamido)butyl)-carbamate. A mix-

ture of this material in 1:1 TFA/CH₂Cl₂ (98 mL) was stirred at room temperature for 3 h. The mixture was concentrated followed by addition of aq NaHCO₃ (15 mL) and extraction with CH₂Cl₂ (5 × 15 mL) to give **13** (1.521 g, 54%, 2 steps). This was used for the next reaction without further purification. ¹H NMR (400 MHz, CDCl₃) δ ppm: 7.96–7.91 (m, 1H), 7.72–7.68 (m, 2H), 7.59–7.62 (m, 1H), 3.25 (t, 2H, *J* = 6.9 Hz), 2.87 (s, 3H), 2.79 (t, 2H, *J* = 6.8 Hz), 1.68–1.61 (m, 2H), 1.59–1.52 (m, 2H). ¹³C NMR (100 MHz, CDCl₃) δ ppm: 148.2, 133.6, 132.0, 131.8, 130.5, 124.1, 50.1, 40.4, 34.4, 28.2, 24.9; HR-MS (ESI) Calculated for [C₁₁H₁₈N₃O₄S]⁺, 288.1013; found, 288.1015.

Synthesis of 2H-K2-Pro-bleo. The 2-chlorotrityl resin (**12**) (500 mg, 730 μmol) was washed with CH₂Cl₂, which was then removed. To this resin, CH₂Cl₂ (6.6 mL) and HCl in dioxane (4 M, 3.0 mL) were added. After shaking for 30 min, the resin was washed with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). To the resin, bromoacetic acid (1.0 M, 3650 μL, 730 μmol) and DIPEA (636 μL, 3650 μmol) were added. The mixture was shaken at room temperature for 2 h. After, the resin was washed with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). To the resin, **13** (1.049 g, 3650 μmol) in DMF (3 mL) was added. The mixture was shaken at room temperature for 1 h followed by washing with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). To the resin, a pre-mixed solution of Fmoc-L-Pro-OH (1232 mg, 3650 μmol), HATU (1388 mg, 3650 μmol), HOAt (497 mg, 3650 μmol), and DIPEA (1207 μL, 7300 μmol) in DMF (10 mL) was added, and the reaction was shaken at room temperature for 30 min. The resin was then washed with DMF (5 × 10 mL). To the resin, 20% piperidine in DMF (10 mL) was added, and it was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 × 10 mL). This cycle was repeated additionally. To 365 μmol of this resin, DBU (273 μL, 1825 μL) and 2-mercaptoethanol (256 μL, 3650 μL) in DMF (5 mL) were added, and the mixture was shaken at room temperature for 15 min. This deprotection step was repeated once more. After, the resin was washed with CH₂Cl₂ (5 × 10 mL) and DMF (5 × 10 mL). To the resin, a pre-mixed solution of (Ht-CO₂H) (**6**) (399 mg, 1460 μmol), DIC (113 μL, 1460 μmol), and oxyma (104 mg, 1460 μmol) was added, and it was shaken at room temperature overnight. The resin was washed with DMF (5 × 10 mL). To this resin, 10 mL of 30% TFA in CH₂Cl₂ was added. The mixture was shaken at room temperature for 10 min. The solution was collected and concentrated. This crude product was initially purified by HPLC (20–8% CH₃CN + 0.1% TFA/H₂O + 0.1% TFA, over 60 min) to afford **17** (83.0 mg, ca. 80% purity).

Compound **17** (13.0 mg) was washed with Et₂O (3 × 1 mL), and the supernatant was removed by decantation. To a solution of this product in DMF (181 μL), a mixture of HATU (11.0 mg, 29.0 μmol), HOAt (29.0 μL, 1.0 M in DMF, 29.0 μmol), and DIPEA (21.5 μL, 130.3 μmol) in DMF (181 μL) was added at room temperature. The mixture was stirred at room temperature for 2 min. To this solution, bleomycin A5 (217.2 μL, 0.20 M in DMF, 43.4 μmol) was added. The mixture was stirred at room temperature for 2 h. The mixture was concentrated under vacuum and then purified by HPLC [step 1: 0.1 mM EDTA in water (pH 6.3) for 15 min; step 2: 100% water for 15 min; step 3: 20–80% MeOH + 0.1% TFA vs H₂O + 0.1% TFA, over 90 min; and step 4: after concentrating the product in step 3, 8–58% MeOH + 0.1% TFA vs H₂O + 0.1% TFA, over 60 min] to give **2H-K2-Pro-bleo** (1.02 μmol, 0.89% yield); HR-MS (MALDI): Calculated for [C₁₃₂H₁₇₄N₃₅O₂₈S₂]⁺, 2761.2704; found, 2761.2832.

Synthesis of Ac-K2-Pro-bleo. The 2-chlorotrityl resin (0.5 g, 0.73 mmol) was washed three times with CH₂Cl₂ followed by treatment with a mixture of 6 mL of CH₂Cl₂ and 3.0 mL of 4 M HCl in dioxane. After shaking for 30 min at room temperature, the resin was washed with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). To the resin, a mixture of bromoacetic acid (1.0 M, 3.65 mL, 3.65 mmol) and DIPEA (626 μL, 3650 μmol) in CH₂Cl₂ (3.65 mL) were added. The mixture was shaken at room temperature for 3 h, and then the resin was washed with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). To the resin, *N*-(4-aminobutyl)-*N*-methyl-4-nitrobenzenesulfonamide (0.525 g, 0.285 mmol) in DMF (4 mL) was added. The mixture

was shaken at room temperature overnight followed by washing with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). To the resin, a pre-mixed solution of Fmoc-Pro-OH (1.232 g, 3.65 mmol), HATU (1.388 g, 3.65 mmol), HOAt (0.497 g, 3.65 mmol), and DIPEA (1251 μL, 7.30 mmol) in DMF (8 mL) was added, and the mixture was shaken at room temperature for 30 min. The resin was washed with DMF (5 × 10 mL). After these washing steps, 20% piperidine in DMF (10 mL) was added, and the mixture was shaken at room temperature for 10 min. The solvent was removed, and this step was repeated. The resin was then washed with CH₂Cl₂ (3 × 10 mL) and DMF (3 × 10 mL). The coupling reaction with Fmoc-Pro-OH was repeated once more, and the process was repeated as mentioned above.

After washing the resin with CH₂Cl₂ and DMF, DBU (546 μL, 3.65 mmol) and HO(CH₂)₂SH (511 μL, 7.30 mmol) in DMF (10 mL) were added, and the mixture was shaken at room temperature for 15 min. This deprotection step was repeated once more. The resin was washed with CH₂Cl₂ (5 × 10 mL) and DMF (5 × 10 mL). To the resin, a 1:1 solution of acetic anhydride and DIPEA (total volume 4 mL) was added, and the mixture was shaken at room temperature for 1 h. The resin was washed with CH₂Cl₂ (5 × 10 mL) and DMF (5 × 10 mL), and then 10 mL of 30% TFA in CH₂Cl₂ was added. The mixture was shaken at room temperature for 10 min, and the supernatant was collected and concentrated under vacuum. The resultant product was purified by HPLC, and the intermediate (0.010 g, 22.8 μmol), HATU (0.017 g, 45.6 μmol), HOAt (6.0 mg, 45.6 μmol), DIPEA (20 μL, 114 μmol), and bleomycin A5 (0.055 g, 34.2 μmol) were stirred at room temperature overnight. The reaction mixture was purified using HPLC to give the target compound in a 2.1% yield (472 nmol). HR-MS (MALDI): Calculated for [C₇₈H₁₂₂N₂₃O₂₆S₂]⁺, 1860.8367; found, 1860.8273.

Cell Lines. Compounds were tested in two cell lines that could be differentiated into myotubes:³⁴ (i) a DM1 [1300 r(CUG) repeats] conditional MyoD-fibroblast cell line and (ii) a WT conditional MyoD-fibroblast cell line.

In Vitro IC₅₀ Measurements. The ability of compounds to disrupt the r(CUG)^{exp}–MBNL1 complex was completed using a previously reported TR-FRET assay^{27,42} with minor modifications. Briefly, 5'-biotinylated r(CUG)₁₂ was folded in a 1× folding buffer (20 mM HEPES, pH 7.5, 110 mM KCl, and 10 mM NaCl) by heating at 60 °C for 5 min followed by cooling to room temperature. This buffer was adjusted to a 1× assay buffer (20 mM HEPES, pH 7.5, 110 mM KCl, 10 mM NaCl, 2 mM MgCl₂, 2 mM CaCl₂, 5 mM DTT, 0.1% BSA, and 0.5% Tween-20). Next, MBNL1-His₆ was added, and the samples were incubated at room temperature for 15 min. The compound of interest was added, and the samples were incubated for another 15 min at room temperature. The final concentrations of r(CUG)₁₂ and MBNL1 were 80 and 60 nM, respectively. A solution of streptavidin-XL665 and anti-His₆-Tb antibody was then added to final concentrations of 40 nM and 0.44 ng/μL, respectively, in a total volume of 10 μL. The samples were incubated for 30 min at room temperature and added to a well of a white 384-well plate where time-resolved fluorescence resonance energy transfer was measured on a Molecular Devices SpectraMax M5 plate reader. The ratios of fluorescence intensity at 545 and 665 nm were calculated, and ratios in the absence of a compound and RNA were used to calculate the percent disruption. The resulting curves were fit to eq 1 to determine IC₅₀ values:

$$y = B + \frac{A - B}{1 + \left(\frac{IC_{50}}{x}\right)^{hillslope}} \quad (1)$$

where *y* is the ratio of fluorescence intensities at 545 and 665 nm (F545/F665), *x* is the concentration of a compound, *B* is the F545/F665 value at the max FRET effect (the solution has RNA and protein but no compound added), *A* is the F545/F665 value at the min FRET effect (the solution has antibodies but no RNA, protein, or compound), and the IC₅₀ is the concentration of a compound where half of the protein is displaced by a compound. *n* = 3 replicates; 2 independent experiments.

Affinity Measurements. The affinity of ligands for various RNAs was measured by monitoring fluorescence intensity as a function of the RNA concentration as previously reported.¹⁴ Briefly, nucleic acids were folded in a 1× binding buffer (8 mM Na₂HPO₄, pH 7.0, 185 mM NaCl, and 1 mM EDTA) for 5 min at 60 °C. The solution was cooled to room temperature, and bovine serum albumin (BSA) was added to a final concentration of 40 μg/mL. Binding assays with r(CUG)₁₂ were completed by titrating folded RNA into 5 μM of either 2H-K2-Pro or 2H-K2-Pro-bleo in a 1× binding buffer containing 40 μg/mL BSA. After each addition of RNA, the samples were incubated for 1 min, and the intrinsic fluorescence intensity of the H RNA-binding modules was measured using a BioTek FLX-800 fluorescence plate reader (excitation wavelength: 360/40 nm; emission wavelength: 460/40 nm; sensitivity = 90). Binding assays with r(GC)₈ were completed by serial dilution (1:2) of the RNA in 1× binding buffer containing 40 μg/mL BSA and either 5 μM 2H-K2-Pro or 5 μM 2H-K2-Pro-bleo. Samples were incubated for 30 min before the fluorescence intensity was measured as described above. For all experiments, plots of the concentration of the nucleic acid versus the relative change in fluorescence were used to determine the binding affinity. Curves were plotted in GraphPad Prism and fit using the equation (eq 2):

$$y = (B_{\max} \times x^h) / (EC_{50}^h + x^h) \quad (2)$$

where y is the change in fluorescence, $B_{\max}$ is the extrapolated maximum change in fluorescence, x is the concentration of the nucleic acid, and h is the Hill slope. $n = 3$ replicates for r(CUG)₁₂; $n = 2$ replicates for r(GC)₈; 2 independent experiments.

Radiolabeling r(CUG)₁₀. The r(CUG)₁₀ oligonucleotide was purchased from Dharmacon and deprotected according to the manufacturer's standard protocol. The RNA (500 pmol) was then 5'-end-radiolabeled using [γ -³²P] ATP and T4 polynucleotide kinase as previously described.¹⁴ The labeled RNA was purified on a denaturing 15% polyacrylamide gel, excised from the gel, and extracted as previously described.¹⁴

RNA Cleavage *In Vitro*. The 5'-³²P-r(CUG)₁₀ radiolabeled RNA was dissolved in 40 μL of water. A 4 μL aliquot of this solution was diluted in 150 μL of 5 mM NaH₂PO₄ (pH 7.4) and folded by heating at 95 °C for 30 s. The solution was cooled to room temperature, and compounds were added at varying concentrations. An equimolar amount of freshly prepared (NH₄)₂Fe(SO₄)₂·6H₂O in 5 mM NaH₂PO₄, pH 7.4, was then added. The solutions were incubated at 37 °C and supplemented with additional equimolar aliquots of (NH₄)₂Fe(SO₄)₂·6H₂O in 5 mM NaH₂PO₄, pH 7.4, after 30 min and 1 h. The RNA was incubated with the compounds for a total of 24 h at 37 °C.

A T1 ladder was prepared by mixing 1 μL of radiolabeled RNA with 30 μL of T1 Buffer (20 mM sodium citrate, pH 5, 1 mM EDTA, and 7 M urea) and heating to 95 °C for 30 s. After cooling to room temperature, RNase T1 (3 units/μL final concentration) was added. The samples were incubated at room temperature for 20 min, and the reaction was stopped by adding an equal volume of 2× loading buffer (95% formamide, 20 mM EDTA, pH 8.0). A hydrolysis ladder was prepared by mixing 1 μL of radiolabeled RNA with 10 μL of 1× alkaline hydrolysis buffer (50 mM NaHCO₃, pH 9.2, and 1 mM EDTA) and heating at 95 °C for 5 min.

All reactions were quenched by adding an equal volume of 2× loading buffer, and the fragments were separated on a denaturing 15% polyacrylamide gel (70 W for 3 h in 1× TBE buffer). Gels were imaged using a Typhoon 9410 variable mode imager (GE Healthcare Life Sciences). Percent cleavage was quantified using QuantityOne (BioRad). $n = 2$ replicates; 2 independent experiments.

Cell Culture and Compound Treatment. Conditional MyoD-fibroblast cell lines^{14,34} were grown in 1× DMEM growth medium (Corning), 10% FBS (Sigma), 1% antibiotic/antimycotic solution (Corning), and 1% Glutaqro (Corning) at 37 °C, and 5% CO₂. Once cells reached ~80% confluency, fibroblasts were differentiated in differentiation medium [1× MEM supplemented with 0.01% iron-transferrin, 0.001% insulin (Life Technologies), and 2 μg/mL

doxycycline (Sigma)] for 24 h. For experiments with binding compounds, cells were treated with compounds in differentiation medium for 48 h. For cleaving compounds, after incubating the cells for 24 h in differentiation medium, the medium was replaced with fresh medium lacking iron-transferrin (i.e., 1× DMEM, 10%FBS, 1% antibiotic/antimycotic solution, and 0.01% insulin) containing 2H-K2-Pro-bleo, and the myotubes were incubated for an additional 48 h.

Analysis of Abundance of r(CUG)-Containing Transcripts.²³ DM1 myotubes were grown in 6-well plates and treated as described in *Cell Culture and Compound Treatment*. After 48 h, the cells were lysed and the total RNA was harvested using a Zymo Quick RNA Miniprep Kit as per the manufacturer's protocol. Approximately 1 μg of total RNA was reverse transcribed using a qScript cDNA synthesis kit (20 μL total reaction volume, Quanta BioSciences), and 2 μL of RT reactions were used for each primer pair for qPCR with a SYBR Green Master Mix performed on a QuantStudio 5 real-time PCR system. The relative abundance of each transcript was determined by normalizing to *GAPDH* using the 2^(-ΔΔCt) method. $n = 3$ replicates; 2 independent experiments.

Evaluation of Nuclear Foci Using Fluorescence *In Situ* Hybridization (FISH).²³ RNA FISH was used to determine a small molecule's effect on the number of nuclear foci present as previously described.²³ Briefly, DM1 patient-derived fibroblasts were grown in a Mat-Tek 96-well glass bottom plate, differentiated, and treated as described above with 1 μM compound for 48 h. To fix the cells, the growth medium was removed, and the cells were washed with 1× DPBS followed by addition of 100 μL of 4% paraformaldehyde in 1× DPBS. The cells were incubated at 37 °C for 10 min and then washed five times with 1× DPBS at 37 °C for 2 min each. Myotubes were permeabilized with 100 μL of 1× DPBS containing 0.1% Triton X-100 (v/v) for 5 min at 37 °C followed by incubation with 100 μL of 30% formamide in 2× SSC (saline sodium citrate buffer) for 10 min at room temperature. Then, 100 μL of the FISH probe, DY547-2'OMe-(CAG)₆, was added to each well to a final concentration of 1 ng/μL in 30% formamide, 2× SSC, 2 μg/μL BSA, 1 μg/μL yeast tRNA, and 2 mM vanadyl complex, and the myotubes were incubated at 37 °C overnight. The cells were then washed with 100 μL of 30% formamide in 2× SSC buffer at 37 °C for 30 min followed by washing with 100 μL of 2× SSC buffer at 37 °C for 30 min. MBNL1 was imaged by adding 20 μL of 1:100 anti-MBNL1 (EMD Millipore, #MABE70) in 2× SSC and incubating the cells at 37 °C for 1 h. The cells were washed three times with 100 μL of 0.1% Triton X-100 (v/v) in 1× DPBS for 5 min at 37 °C followed by incubation with a 1:200 dilution of anti-mouse IgG Dylight 488 conjugate (Thermo Scientific) in 2× SSC at 37 °C for 1 h. After washing three times with 100 μL of 1× DPBS containing 0.1% Triton X-100 (v/v) at 37 °C for 5 min, the cells were washed with 1× DPBS for 5 min at 37 °C. Nuclei were stained by incubating with 100 μL of 1 μg/mL DAPI for 5 min at 37 °C. The cells were then washed with 1× DPBS twice and imaged in 100 μL of 1× DPBS using an Olympus Fluoview 1000 confocal microscope at 100× magnification. The number of foci were counted in 40 nuclei/replicate (120 total nuclei counted); $n = 3$ replicates; 1 independent experiment.

Evaluation of MBNL1 Splicing by RT-qPCR. Myotubes were grown in 6-well plates and were treated as described in *Cell Culture and Compound Treatment*. After 48 h, the cells were lysed, and total RNA was harvested using a Zymo Quick RNA Miniprep Kit. Approximately 1 μg of total RNA was reverse transcribed using a qScript cDNA synthesis kit (20 μL total reaction volume, Quanta BioSciences). A 2 μL aliquot of the RT reactions was used for each primer pair for qPCR with SYBR Green Master Mix performed on a QuantStudio 5 real-time PCR System. Relative abundance of *MBNL1* exon 5 was determined by normalizing to *GAPDH*. The percent rescue was calculated according to eq 3. $n = 3$ replicates; 2 independent experiments.

$$\% \text{rescue} = \frac{\text{Exon 5 relative abundance DM1} - \text{Exon 5 relative abundance treated}}{\text{Exon 5 relative abundance DM1} - \text{Exon 5 relative abundance WT}} \times 100 \quad (3)$$

Evaluation of Alternative pre-mRNA Splicing by RT-PCR.

Myotubes were grown in 6-well plates and were treated as described in [Cell Culture and Compound Treatment](#). Alternative pre-mRNA splicing was assessed as previously described.²³ Briefly, after 48 h, the cells were lysed, and the total RNA was harvested using a Zymo Quick RNA Miniprep Kit. Approximately 1 μg of total RNA was reverse transcribed using a qScript cDNA synthesis kit (20 μL total reaction volume, Quanta BioSciences); 2 μL of the RT reaction was used in PCR amplification reactions with GoTaq DNA polymerase (Promega). RT-PCR products for *MAP4K4* RT-PCR products were observed after 35 cycles of 95 $^{\circ}\text{C}$ for 30 s, 58 $^{\circ}\text{C}$ for 30 s, 72 $^{\circ}\text{C}$ for 1 min, and a final extension at 72 $^{\circ}\text{C}$ for 5 min. Products were separated on a 2% agarose gel stained with ethidium bromide (110 V for 1 hour in 1 $\times$ TBE buffer) and visualized using a Typhoon 9410 variable mode imager. Gels were quantified using ImageJ. Primer sequences are provided in [Table S1](#). $n = 3$ replicates; 2 independent experiments.

Evaluation of γ -H2AX Foci.²³ DNA damage induced by small molecules was assessed by imaging γ -H2AX, a known marker of DNA double strand breaks.²³ Briefly, DM1 myotubes were grown, fixed, and washed as described in the [Evaluation of Nuclear Foci Using Fluorescence In Situ Hybridization](#) section. After washing with 2 $\times$ SCC for 30 min at 37 $^{\circ}\text{C}$, cells were incubated with a 1:500 dilution of anti- γ -H2AX (Abcam) at 37 $^{\circ}\text{C}$ for 1 h. The myotubes were then washed three times with 1 $\times$ DPBS containing 0.1% Triton X-100 (v/v) for 5 min at 37 $^{\circ}\text{C}$, followed by incubation with a 1:200 dilution of a goat anti-mouse IgG-DyLight 488 conjugate (Thermo Scientific) at 37 $^{\circ}\text{C}$ for 1 h. After washing three times with 1 $\times$ DPBS containing 0.1% Triton X-100 (v/v) and twice with 1 $\times$ DPBS for 5 min at 37 $^{\circ}\text{C}$, nuclei were stained with DAPI (1 $\mu\text{g}/\text{mL}$). Cells were imaged in 1 $\times$ DPBS using an Olympus Fluoview 1000 confocal microscope at 100 $\times$ magnification. The number of γ -H2AX foci was counted in 40 nuclei/replicate (120 total nuclei counted over three biological replicates). $n = 3$ replicates; 1 independent experiment.

Evaluation of Subcellular Localization by Fluorescence Microscopy. DM1 myotubes were grown in a Mat-Tek 96-well glass bottom plate, differentiated, and treated as described above. Cells were then treated with compounds **2H-K2-Pro** or **2H-K4NMeS** (at 5 μM). After 24 h, the growth medium was removed, and the cells were washed twice with 100 μL of 1 $\times$ DPBS and imaged in 100 μL of Gibco FluoroBrite DMEM using an Olympus Fluoview 1000 confocal microscope at 100 $\times$ magnification. $n = 3$ replicates; 1 independent experiment.

Evaluation of Cellular Permeability. DM1 myotubes were grown in Mat-Tek 96-well glass bottom plates, differentiated, and treated with 5 μM **2H-K2-Pro** or **2H-K4NMeS** as described above for 24 h. Cells were then washed with 100 μL of 1 $\times$ DPBS and lysed in 30 μL of RNA Lysis Buffer from a Zymo Quick RNA Miniprep Kit. Intrinsic fluorescence intensity of the RNA-binding modules was measured using a BioTek FLX-800 fluorescence plate reader (excitation wavelength: 360/340 nm; emission wavelength: 460/440 nm; sensitivity = 90), and concentrations of compound were extrapolated to standard curves of **2H-K2-Pro** or **2H-K4NMeS** spiked into untreated cell lysate. $n = 6$ replicates; 2 independent experiments.

SMILES and Physicochemical Properties. Marvin was used for generating SMILES and for calculating the physicochemical properties, Marvin 20.8.0, ChemAxon (<https://www.chemaxon.com>).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.0c00558>.

Twelve figures, one table, synthetic schemes, HPLC traces, HR-MS, NMR spectra (PDF)

Dataset 1: simplified molecular-input line-entry system (SMILES) notation for each compound and their associated physical properties (CSV)

■ AUTHOR INFORMATION

Corresponding Author

Matthew D. Disney – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States;
 orcid.org/0000-0001-8486-1796; Email: disney@scripps.edu

Authors

Raphael I. Benhamou – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
 Masahito Abe – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
 Shruti Choudhary – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
 Samantha M. Meyer – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
 Alicia J. Angelbello – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jmedchem.0c00558>

Author Contributions

M.D.D. directed the study, conceived the ideas, and designed the experiments. M.A., S.M.M., A.J.A., and S.C. performed the chemical synthesis and *in vitro* experiments. R.I.B. conducted the cellular studies.

Notes

The authors declare the following competing financial interest(s): MDD is a founder of Expansion Therapeutics.

■ ACKNOWLEDGMENTS

We thank Dr. J. L. Childs-Disney for help in writing this manuscript and the agencies that funded this work including the National Institutes of Health (nos. DP1-NS096898 and R35-NS116846-01 to M.D.D. and F31-NS110269 to A.J.A.), the Department of Defense Peer-Reviewed Medical Research Program (no. W81XWH-18-0718 to M.D.D.), the Muscular Dystrophy Association (grant no. 380467 to M.D.D.), the Myotonic U.S. Fellowship Research Grant (to R.I.B. and S.C.), and the National Ataxia Foundation Fellowship Research Grant (to R.I.B.).

■ ABBREVIATIONS

3' UTR, 3' untranslated region; Alloc, allyloxycarbonyl; ANOVA, analysis of variance; BSA, bovine serum albumin; DAPI, 4',6-diamidino-2-phenylindole; DBU, 1,8-diazabicyclo-[5.4.0]undec-7-ene; DIPEA, diisopropylethylamine; DIC, diisopropylcarbodiimide; DM1, myotonic dystrophy type 1; DMPK, dystrophin myotonic protein kinase; DMF, dimethylformamide; DPBS, Dulbecco's phosphate buffered saline; EDCI, *N*-(3-dimethylaminopropyl)-*N*-ethylcarbodiimide hydrochloride; EDTA, ethylenediamine tetraacetic acid; ESI, electrospray ionization; FISH, fluorescence in situ hybridization; FECD, Fuchs endothelial corneal dystrophy; Fmoc, 9-fluorenylmethyloxycarbonyl; γ -H2AX, Gamma H2A histone family member X; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; HATU, 1-[bis(dimethylamino)methyl]-1H-1,2,3-triazolo[4,5-*b*]pyridine-3-oxide hexafluorophosphate;

HCl, hydrochloric acid; HDL-2, Huntington's disease-like-2; HOBt, 1-hydroxybenzotriazole hydrate; HOAt, 1-hydroxy-7-azabenzotriazole; HPLC, high-performance liquid chromatography; IHC, immunohistochemistry; IPA, isopropanol; LCMS, liquid chromatography mass spectrometry; MAP4K4, mitogen-activated protein kinase kinase kinase 4; MBNL1, muscleblind-like 1; Ns, 2-nitrobenzenesulfonyl; oxyma, cyano-(hydroxyimino)acetic acid ethyl ester; r(CUG)^{exp}, r(CUG) repeat expansion; RT-qPCR, reverse transcription-quantitative polymerase chain reaction; SSC, saline sodium citrate; TFA, trifluoroacetic acid; TPSA, topological polar surface area; TR-FRET, time-resolved fluorescence energy transfer.

REFERENCES

- (1) Dang, C. V.; Reddy, E. P.; Shokat, K. M.; Soucek, L. Drugging the 'Undruggable' Cancer Targets. *Nat. Rev. Cancer* **2017**, *17*, 502–508.
- (2) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem. Biol.* **2016**, *23*, 1077–1090.
- (3) Hangauer, M. J.; Vaughn, I. W.; McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genet.* **2013**, *9*, No. e1003569.
- (4) Angelbello, A. J.; Chen, J. L.; Childs-Disney, J. L.; Zhang, P.; Wang, Z.-F.; Disney, M. D. Using Genome Sequence to Enable the Design of Medicines and Chemical Probes. *Chem. Rev.* **2018**, *118*, 1599–1663.
- (5) Bernat, V.; Disney, M. D. RNA Structures as Mediators of Neurological Diseases and as Drug Targets. *Neuron* **2015**, *87*, 28–46.
- (6) Spitale, R. C.; Flynn, R. A.; Zhang, Q. C.; Crisalli, P.; Lee, B.; Jung, J. W.; Kuchelmeister, H. Y.; Batista, P. J.; Torre, E. A.; Kool, E. T.; Chang, H. Y. Structural Imprints In Vivo Decode RNA Regulatory Mechanisms. *Nature* **2015**, *519*, 486–490.
- (7) Wan, Y.; Qu, K.; Zhang, Q. C.; Flynn, R. A.; Manor, O.; Ouyang, Z.; Zhang, J.; Spitale, R. C.; Snyder, M. P.; Segal, E.; Chang, H. Y. Landscape and Variation of RNA Secondary Structure Across the Human Transcriptome. *Nature* **2014**, *505*, 706–709.
- (8) Disney, M. D. Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *J. Am. Chem. Soc.* **2019**, *141*, 6776–6790.
- (9) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777–793.
- (10) Zu, T.; Gibbens, B.; Doty, N. S.; Gomes-Pereira, M.; Huguot, A.; Stone, M. D.; Margolis, J.; Peterson, M.; Markowski, T. W.; Ingram, M. A.; Nan, Z.; Forster, C.; Low, W. C.; Schoser, B.; Somia, N. V.; Clark, H. B.; Schmechel, S.; Bitterman, P. B.; Gourdon, G.; Swanson, M. S.; Moseley, M.; Ranum, L. P. W. Non-ATG-Initiated Translation Directed by Microsatellite Expansions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 260–265.
- (11) Miller, J. W.; Urbinati, C. R.; Teng-Umuay, P.; Stenberg, M. G.; Byrne, B. J.; Thornton, C. A.; Swanson, M. S. Recruitment of Human Muscleblind Proteins to (CUG)(n) Expansions Associated with Myotonic Dystrophy. *EMBO J.* **2000**, *19*, 4439–4448.
- (12) Brook, J. D.; McCurrach, M. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J. P.; Hudson, T.; Sohn, R.; Zelman, B.; Snell, R. G.; Rundle, S. A.; Crow, S.; Davies, J.; Shelbourne, P.; Buxton, J.; Jones, C.; Juvonen, V.; Johnson, K.; Harper, P. S.; Shaw, D. J.; Housman, D. E. Molecular Basis of Myotonic Dystrophy: Expansion of a Trinucleotide (CTG) Repeat at the 3' End of a Transcript Encoding a Protein Kinase Family Member. *Cell* **1992**, *68*, 799–808.
- (13) Konieczny, P.; Stepniak-Konieczna, E.; Sobczak, K. MBNL Proteins and their Target RNAs, Interaction and Splicing Regulation. *Nucleic Acids Res.* **2014**, *42*, 10873–10887.
- (14) Rzuczek, S. G.; Colgan, L. A.; Nakai, Y.; Cameron, M. D.; Furling, D.; Yasuda, R.; Disney, M. D. Precise Small-Molecule Recognition of a Toxic CUG RNA Repeat Expansion. *Nat. Chem. Biol.* **2017**, *13*, 188–193.
- (15) Pushechnikov, A.; Lee, M. M.; Childs-Disney, J. L.; Sobczak, K.; French, J. M.; Thornton, C. A.; Disney, M. D. Rational Design of Ligands Targeting Triplet Repeating Transcripts that Cause RNA Dominant Disease: Application to Myotonic Muscular Dystrophy Type 1 and Spinocerebellar Ataxia Type 3. *J. Am. Chem. Soc.* **2009**, *131*, 9767–9779.
- (16) Lee, M. M.; French, J. M.; Disney, M. D. Influencing Uptake and Localization of Aminoglycoside-Functionalized Peptoids. *Mol. Biosyst.* **2011**, *7*, 2441–2451.
- (17) Childs-Disney, J. L.; Parkesh, R.; Nakamori, M.; Thornton, C. A.; Disney, M. D. Rational Design of Bioactive, Modularly Assembled Aminoglycosides Targeting the RNA that Causes Myotonic Dystrophy Type 1. *ACS Chem. Biol.* **2012**, *7*, 1984–1993.
- (18) Fuchs, S. M.; Raines, R. T. Pathway for Polyarginine Entry into Mammalian Cells. *Biochemistry* **2004**, *43*, 2438–2444.
- (19) Tarcsay, A.; Keserü, G. M. Contributions of Molecular Properties to Drug Promiscuity. *J. Med. Chem.* **2013**, *56*, 1789–1795.
- (20) Charifson, P. S.; Walters, W. P. Acidic and Basic Drugs in Medicinal Chemistry: a Perspective. *J. Med. Chem.* **2014**, *57*, 9701–9717.
- (21) Nguyen, L.; Luu, L. M.; Peng, S.; Serrano, J. F.; Chan, H. Y. E.; Zimmerman, S. C. Rationally Designed Small Molecules That Target Both the DNA and RNA Causing Myotonic Dystrophy Type 1. *J. Am. Chem. Soc.* **2015**, *137*, 14180–14189.
- (22) Siboni, R. B.; Nakamori, M.; Wagner, S. D.; Struck, A. J.; Coonrod, L. A.; Harriott, S. A.; Cass, D. M.; Tanner, M. K.; Berglund, J. A. Actinomycin D Specifically Reduces Expanded CUG Repeat RNA in Myotonic Dystrophy Models. *Cell Rep.* **2015**, *13*, 2386–2394.
- (23) Angelbello, A. J.; Rzuczek, S. G.; McKee, K. K.; Chen, J. L.; Olafson, H.; Cameron, M. D.; Moss, W. N.; Wang, E. T.; Disney, M. D. Precise Small-Molecule Cleavage of an r(CUG) Repeat Expansion in a Myotonic Dystrophy Mouse Model. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 7799–7804.
- (24) Rzuczek, S. G.; Gao, Y.; Tang, Z.-Z.; Thornton, C. A.; Kodadek, T.; Disney, M. D. Features of Modularly Assembled Compounds That Impart Bioactivity Against an RNA Target. *ACS Chem. Biol.* **2013**, *8*, 2312–2321.
- (25) Childs-Disney, J. L.; Hoskins, J.; Rzuczek, S. G.; Thornton, C. A.; Disney, M. D. Rationally Designed Small Molecules Targeting the RNA that Causes Myotonic Dystrophy Type 1 are Potently Bioactive. *ACS Chem. Biol.* **2012**, *7*, 856–862.
- (26) Kumar, A.; Parkesh, R.; Sznajder, L. J.; Childs-Disney, J. L.; Sobczak, K.; Disney, M. D. Chemical Correction of Pre-mRNA Splicing Defects Associated with Sequestration of Muscleblind-like 1 Protein by Expanded r(CAG)-Containing Transcripts. *ACS Chem. Biol.* **2012**, *7*, 496–505.
- (27) Chen, C. Z.; Sobczak, K.; Hoskins, J.; Southall, N.; Maragan, J. J.; Zheng, W.; Thornton, C. A.; Austin, C. P. Two High-Throughput Screening Assays for Aberrant RNA-Protein Interactions in Myotonic Dystrophy Type 1. *Anal. Bioanal. Chem.* **2012**, *402*, 1889–1898.
- (28) MacArthur, M. W.; Thornton, J. M. Influence of Proline Residues on Protein Conformation. *J. Mol. Biol.* **1991**, *218*, 397–412.
- (29) Verhoorck, S. J. M.; Killoran, P. M.; Coxon, C. R. Fluorinated Prolines as Conformational Tools and Reporters for Peptide and Protein Chemistry. *Biochemistry* **2018**, *57*, 6132–6143.
- (30) Dugave, C.; Demange, L. Cis-Trans Isomerization of Organic Molecules and Biomolecules: Implications and Applications. *Chem. Rev.* **2003**, *103*, 2475–2532.
- (31) Taneja, K. L.; McCurrach, M.; Schalling, M.; Housman, D.; Singer, R. H. Foci of Trinucleotide Repeat Transcripts in Nuclei of Myotonic Dystrophy Cells and Tissues. *J. Cell Biol.* **1995**, *128*, 995–1002.
- (32) Jiang, H.; Mankodi, A.; Swanson, M. S.; Moxley, R. T.; Thornton, C. A. Myotonic Dystrophy Type 1 is Associated with Nuclear Foci of Mutant RNA, Sequestration of Muscleblind Proteins and Deregulated Alternative Splicing in Neurons. *Hum. Mol. Genet.* **2004**, *13*, 3079–3088.

- (33) Nakamori, M.; Sobczak, K.; Puwanant, A.; Welle, S.; Eichinger, K.; Pandya, S.; Dekdebrun, J.; Heatwole, C. R.; McDermott, M. P.; Chen, T.; Cline, M.; Tawil, R.; Osborne, R. J.; Wheeler, T. M.; Swanson, M. S.; Moxley, R. T., 3rd; Thornton, C. A. Splicing Biomarkers of Disease Severity in Myotonic Dystrophy. *Ann. Neurol.* **2013**, *74*, 862–872.
- (34) Arandel, L.; Polay Espinoza, M.; Matloka, M.; Bazinet, A.; De Dea Diniz, D.; Naouar, N.; Rau, F.; Jollet, A.; Edom-Vovard, F.; Mamchaoui, K.; Tarnopolsky, M.; Puymirat, J.; Battail, C.; Boland, A.; Deleuze, J. F.; Mouly, V.; Klein, A. F.; Furling, D. Immortalized Human Myotonic Dystrophy Muscle Cell Lines to Assess Therapeutic Compounds. *Dis. Models Mech.* **2017**, *10*, 487–497.
- (35) Lin, X.; Miller, J. W.; Mankodi, A.; Kanadia, R. N.; Yuan, Y.; Moxley, R. T.; Swanson, M. S.; Thornton, C. A. Failure of MBNL1-Dependent Post-Natal Splicing Transitions in Myotonic Dystrophy. *Hum. Mol. Genet.* **2006**, *15*, 2087–2097.
- (36) Benhamou, R. I.; Angelbello, A. J.; Andrews, R. J.; Wang, E. T.; Moss, W. N.; Disney, M. D. Structure-Specific Cleavage of an RNA Repeat Expansion with a Dimeric Small Molecule Is Advantageous over Sequence-Specific Recognition by an Oligonucleotide. *ACS Chem. Biol.* **2020**, *15*, 485–493.
- (37) Li, Y.; Disney, M. D. Precise Small Molecule Degradation of a Noncoding RNA Identifies Cellular Binding Sites and Modulates an Oncogenic Phenotype. *ACS Chem. Biol.* **2018**, *13*, 3065–3071.
- (38) Boger, D. L.; Cai, H. Bleomycin: Synthetic and Mechanistic Studies. *Angew. Chem., Int. Ed.* **1999**, *38*, 448–476.
- (39) Burma, S.; Chen, B. P.; Murphy, M.; Kurimasa, A.; Chen, D. J. ATM Phosphorylates Histone H2AX in Response to DNA Double-Strand Breaks. *J. Biol. Chem.* **2001**, *276*, 42462–42467.
- (40) Wieben, E. D.; Aleff, R. A.; Tosakulwong, N.; Butz, M. L.; Highsmith, W. E.; Edwards, A. O.; Baratz, K. H. A Common Trinucleotide Repeat Expansion within the Transcription Factor 4 (TCF4, E2-2) Gene Predicts Fuchs Corneal Dystrophy. *PLoS One* **2012**, *7*, No. e49083.
- (41) Margolis, R. L.; Rudnicki, D. D. Pathogenic Insights from Huntington's Disease-Like 2 and Other Huntington's Disease Genocopies. *Curr. Opin. Neurol.* **2016**, *29*, 743–748.
- (42) Parkesh, R.; Childs-Disney, J. L.; Nakamori, M.; Kumar, A.; Wang, E.; Wang, T.; Hoskins, J.; Tran, T.; Housman, D. E.; Thornton, C. A.; Disney, M. D. Design of a Bioactive Small Molecule that Targets the Myotonic Dystrophy Type 1 RNA via an RNA Motif-Ligand Database & Chemical Similarity Searching. *J. Am. Chem. Soc.* **2012**, *134*, 4731–4742.



Affecting RNA biology genome-wide by binding small molecules and chemically induced proximity

Jared T. Baisden, Jessica L. Childs-Disney, Lucas S. Ryan and Matthew D. Disney

Abstract

The ENCODE and genome-wide association projects have shown that much of the genome is transcribed into RNA and much less is translated into protein. These and other functional studies suggest that the druggable transcriptome is much larger than the druggable proteome. This review highlights approaches to define druggable RNA targets and structure–activity relationships across genomic RNA. Binding compounds can be identified and optimized into structure-specific ligands by using sequence-based design with various modes of action, for example, inhibiting translation or directing pre-mRNA splicing outcomes. In addition, strategies to direct protein activity against an RNA of interest via chemically induced proximity is a burgeoning area that has been validated both in cells and in preclinical animal models, and we describe that it may allow rapid access to new avenues to affect RNA biology. These approaches and the unique modes of action suggest that more RNAs are potentially amenable to targeting than proteins.

Addresses

Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458 USA

Corresponding author: Disney, Matthew D. (disney@scripps.edu)

Current Opinion in Chemical Biology 2021, **62**:119–129

This review comes from a themed issue on **Next Generation Therapeutics**

Edited by **Alessio Ciulli** and **Ingrid Wertz**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 9 June 2021

<https://doi.org/10.1016/j.cbpa.2021.03.006>

1367-5931/© 2021 Elsevier Ltd. All rights reserved.

Keywords

RNA, Small molecules, Induced proximity, Cancer, Microsatellite disorders, Chemical biology, Transcriptome-wide design.

Introduction

Although ~80% of the human genome is transcribed into RNA, only 1% is translated into protein [1], suggesting that there are more opportunities to affect biology at the transcriptional level (Fig. 1). Furthermore, RNA has diverse functions tied to its structure [2],

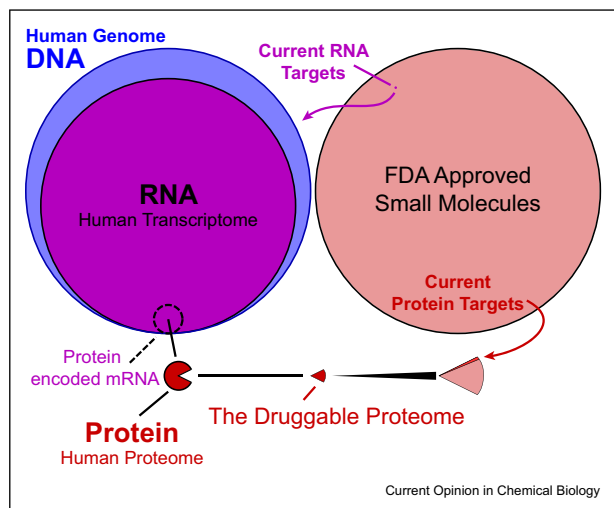
including regulating pre-mRNA splicing outcomes [3], cellular localization [4], phase separation [5], and microRNA (miRNA) biogenesis [6,7]. Indeed, RNA-targeting oligonucleotide-based modalities have garnered Food and Drug Administration (FDA) approval, including those that target RNAs associated with central nervous system disorders such as spinal muscular atrophy (SMA) [8] and liver disease [9]. Although these medicines have transformed patients' lives, their transcriptome-wide implementation has been limited to tissues readily amenable to oligonucleotide targeting, that is, direct delivery to the central nervous system, liver, or kidney.

Highly structured regions in RNA are often functional [10], and the three-dimensional shapes they adopt could provide small-molecule binding pockets [11], occupied by compounds with complementarity of size, shape, stacking, and charge. Although there are challenges associated with small-molecule targeting of RNA, small molecules can be penetrant to a variety of tissues, and often, lead compounds can be medicinally optimized to enhance target engagement [12] and minimize suboptimal features [13]. Herein, we discuss lead drug and chemical probe discovery against RNA targets on a genome-wide scale to elicit effects on the proteome (Table 1): downregulation of 'undruggable' proteins by targeting RNA structures in untranslated regions, upregulation of proteins by targeting miRNAs that repress their expression, and altering the ratio of protein isoforms by influencing pre-mRNA splicing outcomes. Collectively, targeting RNA structures with small molecules can indeed be an effective way to study and control biology.

Drug discovery on a genome-wide scale Defining small-molecule binding landscapes for RNA structures

Targeting RNA with small molecules is not without its challenges, particularly garnering sufficient selectivity and potency to elicit a biological outcome. Selectivity is a composite of (i) the selectivity of the small molecule for the desired RNA structure, (ii) the presence of that structure elsewhere in the transcriptome, (iii) the relative expression levels of all targets with said structure, and (iv) the structure's functionality. We developed a method, two-dimensional combinatorial

Figure 1



Disparity in drug development against protein and RNA targets. The human genome, transcriptome, and proteome are shown to scale, with current clinical targets [1] highlighted as a small slice (bright red) of the druggable proteome. Current RNA therapeutics, including antisense oligonucleotides, currently account for ~1% Food and Drug Administration (FDA)-approved therapies, shown to scale as a purple pixel within all FDA-approved small molecules.

screening (2DCS), that studies the binding capacity and selectivity of small molecules for RNA structures, defining binding landscapes and structure–activity relationships on a transcriptome-wide scale. Coupling 2DCS with the accurate modeling of the RNA's structure from sequence [14–17] has enabled the development of small molecules against RNA targets on a genome-wide scale.

2DCS probes the binding potential of small molecules against a library of discrete RNA structures, wherein compounds from a chemical library are immobilized onto functionalized agarose microarrays. RNA structures that bind a small molecule are identified by high-throughput RNA sequencing (RNA-seq). The resulting RNA-seq data are analyzed to calculate the statistical confidence in the enrichment of RNA structures from a 2DCS selection vs. the starting RNA pool (library) [18]; that is, the analysis accounts for biases that may arise during *in vitro* transcription to generate the RNA library, reverse transcription, polymerase chain reaction amplification, and sequencing. Statistical confidence is a metric of binding fitness, including both affinity and selectivity. Logos [19] and DiffLogos [20] can then be used to visualize privileged sequences and relate structural similarities and differences between molecules, elucidating structure–activity relationships (SAR).

SAR on a transcriptome-wide scale

Fortuitously, RNA secondary structure can be readily and accurately modeled from sequence. RNA's limited

number of building blocks and the large contribution of base pairing energy to its stability enable rapid secondary structure prediction. This ability to model RNA structure facilitates mapping of potential small-molecule binding pockets across the transcriptome and insight into compound selectivity, a function of the number of structures that a compound binds, its relative affinity for each, and the presence of these structural elements across the transcriptome. Notably, the folding landscape for RNA can be dynamic, wherein interconversion between states is necessary for function in some cases [10]. Although labor-intensive, RNA conformational dynamics can be modeled and used to identify lead molecules, as demonstrated with a dynamic ensemble of HIV trans-activation response element (TAR) RNA [11]. Likewise, Hong et al [21] and Dohno and Nakatani [22] have exploited the base pairing propensity of unpaired and dynamic RNA structures by developing ligands to induce higher order interactions within target RNAs.

To integrate RNA structure prediction and small-molecule design on a transcriptome-wide scale, our laboratory has developed a pipeline named Inforna [23,24]. This pipeline starts with computational prediction of the target RNA's secondary structure based on its genomic sequence alone, which can be further refined by analyzing sequence conservation [25] and by experiment [26]. The structures within an RNA target are then compared with the RNA structure–small molecule interactions identified by 2DCS or other methods. In the following section, we describe how a fundamental understanding of the molecular recognition of RNA structures by small molecules enables the development of compounds with various modes of action (MOAs) that influence the content of the proteome (Table 1).

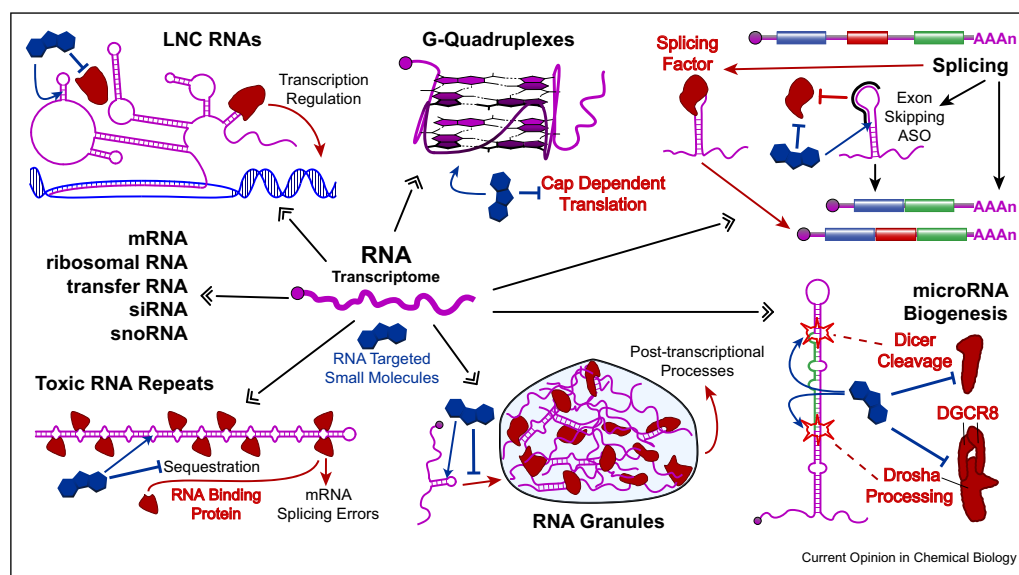
Affecting RNA biology: binding small molecules

As coding and noncoding RNAs adopt 3D folds that influence their biological roles, there are likely many ways to modulate their function. This section describes how simple binding compounds can alter cellular protein content by various mechanisms (Fig. 2).

RNA repeat expansions

Microsatellite disorders are a class of more than 40 inherited diseases [27] caused by RNA repeat expansions, perhaps ideal targets for small molecules. In general, repeat length is correlated with disease onset and/or severity, and disease mechanism is dependent on the location of the repeat within the transcript. Once the repeat expansion reaches a pathogenic length, which is unique to each disease, the RNA folds into a periodic array of loops that form small-molecule binding pockets. These pockets are absent in transcripts lacking the

Figure 2



RNA structure–function relationships. Hubs, or regions, of RNA structure are essential for many cellular processes [10] and provide targetable sites for small-molecule binding. Shown are a subset of important RNA structural motifs and their associated biological processes, both in health and disease. RNA-binding molecules, shown in dark blue, have been developed to target a number of these structures and alter their function. Primary and precursor miRNA structures are shown on the bottom right, which are recognized by the miRNA biogenesis machinery by their specific structural motifs and the distances between them. Inc, long noncoding.

repeat expansion, and the structural differences between pathogenic and short repeats afford the possibility of developing allele-specific small molecules. Furthermore, the repeating nature of these 3D structures enables the design of dimers that bind two adjacent loops simultaneously, which increases affinity, specificity, and potency [28,29]. Notably, repeat expansions, owing to the highly thermodynamically stable structures that they form, are not amenable to antisense oligonucleotide (ASO) targeting, which are not allele specific as they recognize the repeat sequence present in both mutant and wild-type alleles as well as in other transcripts.

The loops formed by RNA repeat expansions can be toxic via gain-of-function mechanisms, such as binding and sequestration of RNA-binding proteins. Such is the case in myotonic dystrophy 1 [DM1; r(CUG)^{exp}] and myotonic dystrophy type 2 [r(CCUG)^{exp}], the two most common forms of adult-onset muscular dystrophy. Indeed, many laboratories have developed ligands that bind r(CUG)^{exp}, although some have mixed MOAs. In early work, the Miller lab used a dynamic combinatorial screening method to identify compounds that bind a model of r(CUG)^{exp} [30], while Nakatani [31] and Zimmerman [32,33] laboratories synthesized dimeric compounds to bind r(CUG)^{exp} and inhibit the sequestration of MBNL1. A peptide has also been developed to bind r(CUG)^{exp} and alleviate defects in a *Drosophila* model of DM1 [34].

A dimer that binds two adjacent loops in r(CUG)^{exp} developed by our laboratory is allele selective, occupying the target in DM1-affected cells while not affecting mRNAs that contained these shorter repeats [35]. This selectivity can be traced to the lack of structure adopted by short repeats in their corresponding transcripts [29]. The dimer was later appended with bleomycin [36,37], affording a chimeric molecule that cleaved the mutant allele selectively over short repeats found in the wild-type allele and other transcripts in patient-derived muscle cells and in a mouse model with no detectable off-targets [38].

This notion that expanded repeats can be selectively targeted by small molecules over short, nonpathogenic repeats has been explored for other repeat expansions [39], for example, r(CAG)^{exp}, which causes Huntington disease [40–43]; r(CGG)^{exp}, associated with fragile X syndrome and fragile X-associated tremor/ataxia syndrome [44–47]; and r(G₄C₂)^{exp}, which causes C9orf72 frontotemporal dementia and amyotrophic lateral sclerosis (c9ALS/FTD). The latter repeat has been shown to adopt various structures [48–50] (an array of internal loops and G-quadruplexes) that have been targeted with small molecules for therapeutic benefit [45,50–52].

MicroRNAs

Small, but highly structured, miRNAs suppress translation of complementary mRNAs, acting as genetic

'dimmer switches.' Expression levels of a subset of proteins can therefore be tuned by modifying miRNA abundance, that is, miRNA biogenesis. Primary and precursor miRNAs form well-defined structures that are recognized by processing enzymes, affording the mature miRNA. Small molecules that bind structures located in processing sites and thus inhibit miRNA biogenesis have been discovered through various methods, including microarrays [53] and a click chemistry approach that assays for Dicer inhibition [54]. The Duca lab developed aminoglycoside conjugates and polyamines to show specific interaction with miR-372 [55,56], while many studies have focused on inhibiting the biogenesis of miR-21 [57–63] as its upregulation has been linked to cardiovascular disease, fibrosis, and many cancers.

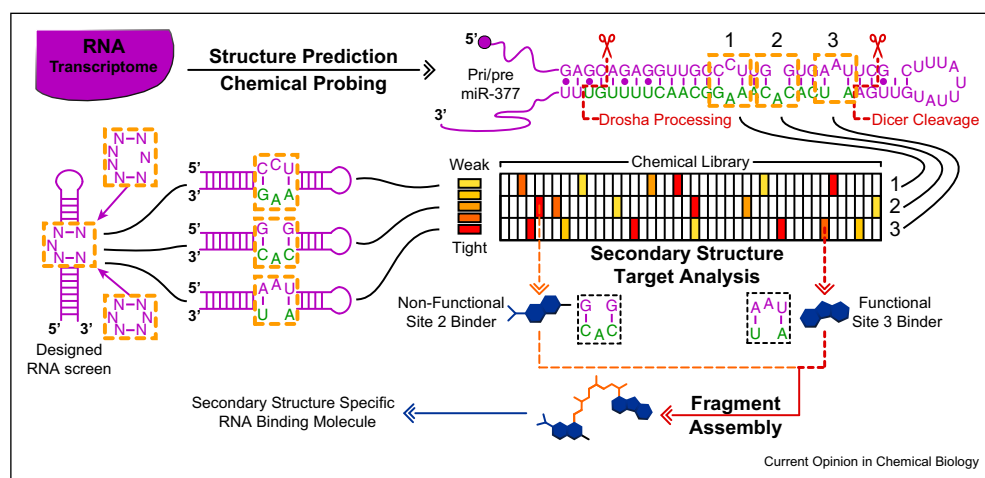
The structures embedded in miRNA processing sites are not always unique in the transcriptome. In the following paragraphs, we describe the lead optimization of a small molecule that binds a structural element found in the Dicer processing sites of two different miRNAs, miR-421 and miR-377, to create a miR-377–selective inhibitor (Fig. 3) [64]. The lead molecule for the two Dicer sites was generated from the RNA-binding preferences of AstraZeneca's small-molecule library, as determined by 2DCS. As expected, the lead molecule occupies both miRNAs in cells similarly and hence inhibits their biogenesis to similar extents. Target occupancy was studied by Chemical Cross-Linking and Isolation by Pull-down (Chem-CLIP), in which the RNA-binding small molecule is functionalized with a cross-linking module. RNAs cross-linked to the Chem-CLIP probe are isolated by cells and analyzed

by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) or RNA-seq.

Thus, the question at hand was if the lead molecule could be optimized in facile fashion to selectively inhibit miR-377. Careful inspection of pre-miR-377 and pre-miR-421 revealed a structural feature unique to pre-miR-377, which fortuitously was nearby the Dicer site and could be exploited to enhance selectivity. Inforna identified a compound that binds the unique site, and tethering the two binding fragments together (a heterodimer) afforded selective inhibition of miR-377 biogenesis, as determined by miRNome-wide profiling and proteomics studies. Importantly, competitive Chem-CLIP studies, in which cells are cotreated with the Chem-CLIP probe and parent compound, showed that the pre-miR-377 target, but not pre-miR-421, was occupied.

Collectively, these studies show that degeneracy in RNA structural motifs in the genome does not preclude the development of selective small molecules that induce a biological effect. The combination of RNA structure modeling and small-molecule binding preferences, in the form of binding landscapes, can inform lead molecules and identify degenerate sites in the transcriptome. Furthermore, this degeneracy can be alleviated by exploiting structural differences in the RNAs with the degenerate site. Although these chimeric molecules lie outside traditional oral bioavailability guidelines, such as the Lipinski Rule of Five [65], there is growing evidence that therapeutic small molecules can indeed deviate from these standards.

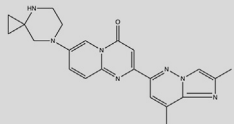
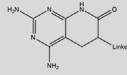
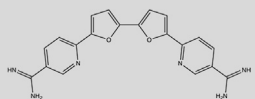
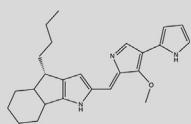
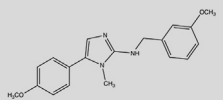
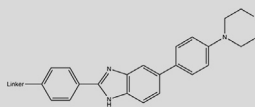
Figure 3



Two-dimensional combinatorial screening and fragment assembly of RNA-binding small molecules. The RNA's sequence can be combined with folding simulations to reveal important structural motifs across the transcriptome. To demonstrate that binding preferences can inform design of a selective small-molecule inhibitor, a degenerate Dicer site in pre-miR-377 and pre-miR-421 was used as a test case. A structural feature unique to miR-377 was exploited, affording a dimer that binds the degenerate and unique site simultaneously, abrogating binding and activity against pre-miR-421.

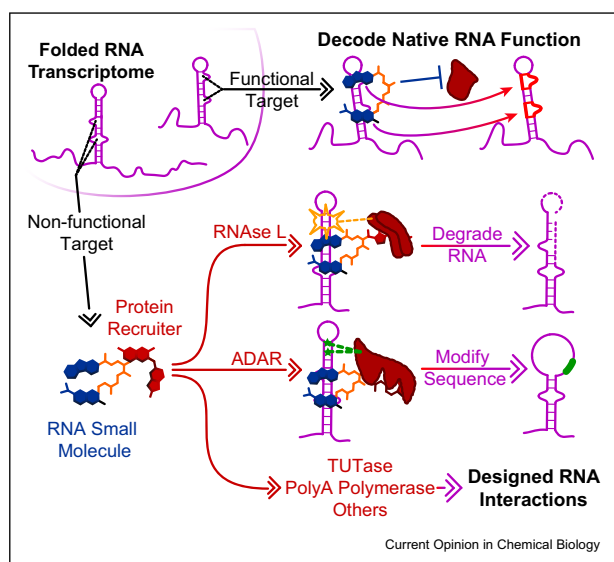
Table 1

Compounds from highlighted references.

Compound name {reference number}	Chemical structure	RNA target {disease}	Potency (EC ₅₀) or change {cell line or model}	Selectivity {off-target}	Structure	Development stage
Risdiplam {8/83/85}		SMN2 splicing {SMA}	4 nM (EC _{1.5x}) {fibroblasts derived from patients with SMA}	16.75 nM {FOXN1 splicing}	Yes	FDA approved
Pyrido[2,3-d]pyrimidine {29}		r(CUG) {DM1}	**Foci reduction {cells of NR patients with DM1}		MD only	NA
DB1246 {49}		r(G ₄ C ₂) ^{exp} {c9ALS/FTD}	**Foci reduction {iPSC cortical neurons}	NR	No	NA
bPGN {70}		Pre-miR-21 {colorectal cancer}	0.035 ± 0.003 μM (GI ₅₀) {HTC-116: colorectal cancer cells}	>10 μM {CRL-1790: normal colon}	No	NR
Compound 5 {95}		MALAT triple helix {metastasis and proliferation}	***Branching morphogenesis {MMTV-PyMT tumors}	No significant differences {Neat1}	MD only	NR
TGP-21-C1-3 {97}		Pre-miR-21 {triple-negative breast cancer}	**Cell invasion {MCF-10A + miR-21 transfection}	No significant differences {miRNA profiling}	MD only	NR

DM1, myotonic dystrophy 1; MALAT, metastasis-associated lung adenocarcinoma transcript; SMA, spinal muscular atrophy; miRNA, microRNA; NA, not applicable; NR, not reported; FDA, Food and Drug Administration; iPSC, induced pluripotent stem cell; MD, molecular dynamics simulation; **, P ≤ 0.001; ***, P ≤ 0.001

Figure 4



Chemically induced proximity expands the mode of action of RNA-targeted small molecules. Using computational prediction and a fully functionalized fragment (FFF) screening strategy, structure-specific dimeric molecules can reveal RNA function and downstream networks (red binding sites), for example, the inhibition of miR-377 biogenesis [64]. In addition, induced proximity can be used to alter RNA lifetime (degradation by RNase L or TUTase) [62,85] or change its sequence using editing enzymes such as ADAR (green), which could change the sequence of the encoded protein or alter splice site selection. Collectively, induced proximity has the potential to impart new modes of action for RNA-targeting small molecules. ADAR, adenosine deaminase that acts on RNA; RNase, ribonuclease; TUTase, terminal U transferase.

Targeting pre-mRNA splicing with RNA-binding compounds, molecular glues that stabilize RNA–protein complexes

Small molecules can also affect the protein isoform generated from an mRNA by affecting alternative pre-mRNA splicing, with important proof-of-concept studies demonstrated with oligonucleotides [66]. The most extensive efforts have centered on alleviation of SMA, culminating in therapies such as risdiplam [8]. In SMA, the survival motor neuron 1 (SMN1) gene is defective, resulting in loss of SMN1 protein. Fortunately, humans have a second SMN gene, SMN2, that varies by a single nucleotide, or single nucleotide polymorphism (SNP). This SNP changes SMN2's splicing pattern, causing exon exclusion and producing a protein isoform with a reduced half-life. Risdiplam directs the splicing of the SMN2 gene to include SMN2 exon 7, affording the stable protein isoform that can compensate for loss of SMN1 [67].

Small molecules that influence splicing have been most commonly discovered by phenotypic screens and generally act by stabilizing RNA-protein interfaces [68]. More recently, drug/probe discovery efforts have focused on RNA-specific screening, as was the case for

the discovery of the '5' splice site bulge repair' compound that directs the inclusion of SMN2 exon 7 [69].

Small molecules can also influence the protein isoforms generated from a transcript by effecting exon exclusion, as demonstrated for microtubule-associated protein tau (*MAPT*) [70–72]. In frontotemporal dementia with Parkinsonism-17, a splice site mutation enhances the interaction of a regulator element in *MAPT* with U1 snRNA, causing inclusion of exon 11 and an aggregation-prone protein isoform [73]. Small molecules that stabilize the mutated structure facilitate exon exclusion, shifting splicing toward the protein isoform that does not aggregate [70].

Targeting internal ribosome entry sites and G-quadruplexes

Small molecules can also alter the proteome by binding and occluding important structures in the encoding transcript that regulate protein expression. In particular, targeting internal ribosome entry sites allows for gene-specific regulation of translation, for example, the internal ribosome entry site in c-MYC [74]. Likewise, inhibition of translation can be achieved by stabilization of other structures in the 5' untranslated region (UTR), preventing cap-dependent translation. Of these UTR targets, G-quadruplex (G4) structures are often targeted as they are incredibly stable and are found in oncogenes KRAS [75], metalloproteinase 10 [76], NRAS [77], telomeric repeat-containing RNA [78,79], and others.

Small molecules that target metastasis-associated lung adenocarcinoma transcript 1, a long noncoding RNA

Long noncoding (lnc) RNAs are an emerging class of targets with a diverse array of functions, including epigenetic regulation, polycomb formation, nuclear localization, and others. A notable lnc RNA target, metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), causes tumorigenesis by forming a highly stable triple helix structure that enables nuclear localization and results in constitutive function. Indeed, small molecules that bind the triple helix reduce oncogenic potential [80].

Genome-wide design: fully functionalized fragments

Fragment-based drug discovery has been a key technology in the protein-targeting field. Unlike traditional screening methods of large small-molecule libraries ($\sim 10^6+$), fragment-based approaches use a much smaller number ($\sim 10^3$) of low-molecular-weight compounds. Indeed, a recent study combined fragment-based ligand discovery with a cross-linking and mass spectrometry strategy that uses fully functionalized fragments (FFFs) equipped with an alkyne and a

photoreactive diazirine [81]. The strategy enabled study of target specificity for each fragment in living cells and allowed for optimization of highly homologous proteins while reducing the resources required for fruitful screening efforts.

This FFF screening strategy was applied to RNA targets using pre-miR-21 as a test case. In brief, the library was screened in the presence of a previously developed Dicer site-binding molecule to identify FFFs that bind other sites on the target RNA [82]. One FFF bound cooperatively with the RNA-binding module, the tethering of which afforded a low-molecular-weight dimer with improved affinity and specificity over its individual counterparts, validated by experiments that measured RNA target occupancy in cells and RNA-seq. Collectively, these studies provide a facile route to lead molecule optimization without substantially increasing the molecule's molecular weight and hence maintaining druglikeness.

Moving forward, the cellular screening of FFFs against RNA could enable rapid targeting of functional sites in the RNA's structure, simplifying the pipeline for transcriptome-wide targeting of RNA; that is, functionalized fragment screening of RNA targets strengthens structure-based targeting platforms by directly mapping small-molecule binding.

Affecting RNA biology: induced proximity

In addition to inhibiting RNA function through structure-specific targeting of tandem binding pockets, fragment assembly has enormous potential to enable design of small molecules with MOAs outside simple binding; that is, a structure-specific RNA-binding fragment can be coupled to a fragment that recruits enzymes or proteins that act on RNA, referred to as induced proximity. Originally developed for protein-targeted ligands by recruiting ubiquitin transferase [83] to target proteins for degradation, 'protein targeting chimeric molecules,' or PROTACs [84], were adopted for targeted RNA-induced degradation and similarly named 'ribonuclease targeting chimeras' or RIBOTACs [62,85,86]. These chimeric molecules comprise an RNA-targeting molecule attached to a molecule that selectively recruits ribonuclease (RNase) L to the desired RNA target. A miR-21-targeting RIBOTAC had enhanced activity and selectivity as compared with the parent binding molecule and selectively altered the proteome, short-circuiting oncogenic molecular pathways in triple-negative breast cancer cells and a mouse model of triple-negative breast cancer. The observed enhancement in selectivity was likely the combined effect of the RNA binder and the inherent substrate

preferences of the RNase. A recent study showed that selectivity could be engineered into a promiscuous protein-binding small molecule by its conversion into a PROTAC [87]. Overall, these proof-of-concept studies of induced proximity directed at RNA targets illustrate the ability to recruit a cellular factor with desired function to a specific RNA using a small molecule. Importantly, induced proximity eliminates the requirement of RNA-targeted small molecules to bind functional sites. Although not induced proximity, a recent study showed that a bifunctional small molecule comprising an RNA binder and an inhibitor of Dicer blocked miR-21 biogenesis [88], an interesting approach that could be applied to other miRNAs.

Induced proximity has immense potential as other cellular factors could be recruited to specific RNA structures with a chimeric small molecule, expanding possible compound MOAs (Fig. 4). For example, induced proximity could be used to introduce site-specific modifications by recruiting adenosine deaminases that act on RNA (ADARs), supported by studies in which a guide RNA, delivered by an adeno-associated virus (AAV) vector, specifically edited an RNA target [89]. RNA lifetime could be altered by recruiting the polyadenylation machinery, deadenylating enzymes, terminal U transferase (TUTase) which targets an RNA for degradation, decapping enzymes, and so on. Induced proximity could unveil exciting new functions for RNA and further our understanding of RNA biology for therapeutic benefit.

Summary and outlook

There is vast potential to alter the proteome by targeting RNA, not only for RNA-related diseases but also for 'undruggable' proteins. RNA's limited number of building blocks and diverse folding landscapes come with unique and challenging problems. However, its hierarchical folding and the accurate modeling thereof can be exploited to design lead molecules. With additional methodology such as Chem-CLIP and Chem-CLIP Fragment Mapping (Chem-CLIP-Frag-Map), small-molecule specificity can be directly measured in cells, validating the target and informing off-target effects. These methodologies can be synergistically combined into a platform for drugging the RNA transcriptome by sequence alone.

As genome-wide association studies provide a stream of RNA targets, many questions remain, such as how relatively small changes in sequence or repeat length cause large changes in structure [90], the uniqueness of RNA structures throughout the transcriptome, if

small molecules can lock dynamic structures in a cellular RNA into a single conformation, and so on. Induced proximity could be an invaluable tool for future RNA research. While initial work demonstrated that removing a specific RNA through the recruitment of an RNase is indeed feasible, induced proximity could provide a more extensive understanding of RNA localization, modification, lifetime, and epigenetic effects. Indeed, recent studies on RNA methylation [91] and sequence modification [89] illustrate its biological diversity. As we anticipate the discovery of new and important RNA functions, we believe that RNA-targeted therapies will become a pillar of small-molecule development.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: MDD is a founder of Expansion Therapeutics.

Acknowledgements

Funding was provided by the National Institutes of Health (R01 CA249180, P01 NS09914, R35 NS116846, and UG3 NS116921 to MDD), the Department of Defense (W81XWH-18-0718 and W81XWH-19-PRMRP-IIRA to MDD), and the Huntington's Disease Society of America (JTB).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, *et al.*: **A comprehensive map of molecular drug targets.** *Nat Rev Drug Discov* 2017, **16**:19–34.
2. Esteller M: **Non-coding RNAs in human disease.** *Nat Rev Genet* 2011, **12**:861–874.
3. Taliaferro JM, Lambert NJ, Sudmant PH, Dominguez D, Merkin JJ, Alexis MS, Bazile C, Burge CB: **RNA sequence context effects measured in vitro predict in vivo protein binding and regulation.** *Mol Cell* 2016, **64**:294–306.
4. Sun L, Fazal FM, Li P, Broughton JP, Lee B, Tang L, Huang W, Kool ET, Chang HY, Zhang QC: **RNA structure maps across mammalian cellular compartments.** *Nat Struct Mol Biol* 2019, **26**:322–330.
5. Langdon EM, Qiu Y, Ghanbari Niaki A, McLaughlin GA, Weidmann CA, Gerbich TM, Smith JA, Crutchley JM, Termini CM, Weeks KM, *et al.*: **mRNA structure determines specificity of a polyQ-driven phase separation.** *Science* 2018, **360**:922–927.
6. Kwon SC, Baek SC, Choi YG, Yang J, Lee YS, Woo JS, Kim VN: **Molecular basis for the single-nucleotide precision of primary microRNA processing.** *Mol Cell* 2019, **73**:505–518 e505.
7. Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P: **Molecular basis for interaction of let-7 microRNAs with Lin28.** *Cell* 2011, **147**:1080–1091.
8. Ratni H, Ebeling M, Baird J, Bendels S, Bylund J, Chen KS, Denk N, Feng Z, Green L, Guerard M, *et al.*: **Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA).** *J Med Chem* 2018, **61**:6501–6517.

The authors discovered risdiplam, a pharmacophore optimized compound, that directs SMN2 splicing to alleviate SMA. Through medicinal chemistry efforts, they achieved better potency, alleviated phototoxicity risk, lowered basicity, and prevented active metabolites. With these modifications, risdiplam successfully underwent clinical trials and garnered FDA-approval.

9. Stein CA, Castanotto D: **FDA-approved oligonucleotide therapies in 2017.** *Mol Ther* 2017, **25**:1069–1075.
10. Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM: **The roles of structural dynamics in the cellular functions of RNAs.** *Nat Rev Mol Cell Biol* 2019, **20**:474–489.
11. Ganser LR, Lee J, Rangadurai A, Merriman DK, Kelly ML, Kansal AD, Sathyamoorthy B, Al-Hashimi HM: **High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble.** *Nat Struct Mol Biol* 2018, **25**:425–434.
12. Rizvi NF, Santa Maria Jr JP, Nahvi A, Klappenbach J, Klein DJ, Curran PJ, Richards MP, Chamberlin C, Saradjian P, Burchard J, *et al.*: **Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space.** *SLAS Discov* 2020, **25**:384–396.
13. Jimenez-Moreno E, Montalvillo-Jimenez L, Santana AG, Gomez AM, Jimenez-Oses G, Corzana F, Bastida A, Jimenez-Barbero J, Canada FJ, Gomez-Pinto I, *et al.*: **Finding the right candidate for the right position: a fast NMR-assisted combinatorial method for optimizing nucleic acids binders.** *J Am Chem Soc* 2016, **138**:6463–6474.
14. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**:3406–3415.
15. Andrews RJ, Roche J, Moss WN: **ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV.** *PeerJ* 2018, **6**, e6136.
16. Mathews DH: **Revolutions in RNA secondary structure prediction.** *J Mol Biol* 2006, **359**:526–532.
17. Parisien M, Major F: **The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data.** *Nature* 2008, **452**:51–55.
18. Velagapudi SP, Luo Y, Tran T, Haniff HS, Nakai Y, Fallahi M, Martinez GJ, Childs-Disney JL, Disney MD: **Defining RNA-small molecule affinity landscapes enables design of a small molecule inhibitor of an oncogenic noncoding RNA.** *ACS Cent Sci* 2017, **3**:205–216.
19. Schneider TD, Stephens RM: **Sequence logos: a new way to display consensus sequences.** *Nucleic Acids Res* 1990, **18**:6097–6100.
20. Nettling M, Treutler H, Grau J, Keilwagen J, Posch S, Grosse I: **Difflogo: a comparative visualization of sequence motifs.** *BMC Bioinf* 2015, **16**:387.
21. Hong C, Otabe T, Matsumoto S, Dohno C, Murata A, Hagihara M, Nakatani K: **Formation of a ligand-assisted complex of two RNA hairpin loops.** *Chemistry* 2014, **20**:5282–5287.
22. Dohno C, Nakatani K: **Molecular glue for RNA: regulating RNA structure and function through synthetic RNA binding molecules.** *ChemBiochem* 2019, **20**:2903–2910.
23. Velagapudi SP, Gallo SM, Disney MD: **Sequence-based design of bioactive small molecules that target precursor micro-RNAs.** *Nat Chem Biol* 2014, **10**:291–297.
24. Disney MD, Winkelsas AM, Velagapudi SP, Southern M, Fallahi M, Childs-Disney JL: **Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs.** *ACS Chem Biol* 2016, **11**:1720–1728.

The authors expanded a database of RNA-small molecule interactions that informs probe design against RNA targets. In addition to data gathered from a library-vs library selection method, or 2DCS, this lead identification platform includes all other information on RNA motif-small molecule binding partners reported in the literature. Together, these selected interactions form an optimized basis set to inform lead small molecules and lead RNA targets.

25. Su H, Peng Z, Yang J: **Recognition of small molecule-RNA binding sites using RNA sequence and structure.** *Bioinformatics* 2021, **37**:36–42.
 26. Mitchell 3rd D, Assmann SM, Bevilacqua PC: **Probing RNA structure in vivo.** *Curr Opin Struct Biol* 2019, **59**:151–158.
 27. Paulson H: **Repeat expansion diseases.** *Handb Clin Neurol* 2018, **147**:105–123.
 28. Ondoño R, Lirio A, Elvira C, Alvarez-Marimon E, Provenzano C, Cardinali B, Perez-Alonso M, Peralvarez-Marín A, Borrell JI, Falcone G, *et al.*: **Design of novel small molecule base-pair recognizers of toxic CUG RNA transcripts characteristics of DM1.** *Comput Struct Biotechnol J* 2021, **19**:51–61.
- The authors used computational modeling to evaluate ligands that established a Janus-Wedge between the U•U bulge in r(CUG) repeats. They identified a pyrido[2,3-d]pyrimidine as a Janus-Wedge binder and optimized scaffolds to create a homodimeric r(CUG)^{exp} binding molecule.
29. Angelbello AJ, Rzuczek SG, McKee KK, Chen JL, Olafson H, Cameron MD, Moss WN, Wang ET, Disney MD: **Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model.** *Proc Natl Acad Sci U S A* 2019, **116**:7799–7804.
 30. Gareiss PC, Sobczak K, McNaughton BR, Palde PB, Thornton CA, Miller BL: **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* 2008, **130**:16254–16261.
 31. Li J, Nakamori M, Matsumoto J, Murata A, Dohno C, Kiliszek A, Taylor K, Sobczak K, Nakatani K: **A dimeric 2,9-diamino-1,10-phenanthroline derivative improves alternative splicing in myotonic dystrophy type 1 cell and mouse models.** *Chemistry* 2018, **24**:18115–18122.
 32. Lee J, Bai Y, Chembazhi UV, Peng S, Yum K, Luu LM, Hagler LD, Serrano JF, Chan HYE, Kalsotra A, *et al.*: **Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1.** *Proc Natl Acad Sci U S A* 2019, **116**:8709–8714.
 33. Hagler LD, Luu LM, Tonelli M, Lee J, Hayes SM, Bonson SE, Vergara JI, Butcher SE, Zimmerman SC: **Expanded DNA and RNA trinucleotide repeats in myotonic dystrophy type 1 select their own multitarget, sequence-selective inhibitors.** *Biochemistry* 2020, **59**:3463–3472.
 34. Garcia-Lopez A, Llamusi B, Orzaez M, Perez-Paya E, Artero RD: **In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models.** *Proc Natl Acad Sci U S A* 2011, **108**:11866–11871.
 35. Rzuczek SG, Colgan LA, Nakai Y, Cameron MD, Furling D, Yasuda R, Disney MD: **Precise small-molecule recognition of a toxic CUG RNA repeat expansion.** *Nat Chem Biol* 2017, **13**:188–193.
 36. Boger DL, Cai H: **Bleomycin: synthetic and mechanistic studies.** *Angew Chem Int Ed Engl* 1999, **38**:448–476.
 37. Carter BJ, de Vroom E, Long EC, van der Marel GA, van Boom JH, Hecht SM: **Site-specific cleavage of RNA by Fe(II). Bleomycin.** *Proc Natl Acad Sci U S A* 1990, **87**:9373–9377.
 38. Angelbello AJ, DeFeo ME, Glinkerman CM, Boger DL, Disney MD: **Precise targeted cleavage of a r(CUG) repeat expansion in cells by using a small-molecule-deglycobleomycin conjugate.** *ACS Chem Biol* 2020, **15**:849–855.
 39. Angelbello AJ, Chen JL, Disney MD: **Small molecule targeting of RNA structures in neurological disorders.** *Ann N Y Acad Sci* 2020, **1471**:57–71.
 40. Khan E, Tawani A, Mishra SK, Verma AK, Upadhyay A, Kumar M, Sandhir R, Mishra A, Kumar A: **Myricetin reduces toxic level of CAG repeats RNA in Huntington's Disease (HD) and spinocerebellar ataxia (SCAS).** *ACS Chem Biol* 2018, **13**:180–188.
 41. Khan E, Mishra SK, Mishra R, Mishra A, Kumar A: **Discovery of a potent small molecule inhibiting Huntington's Disease (HD) pathogenesis via targeting CAG repeats RNA and poly Q protein.** *Sci Rep* 2019, **9**:16872.
 42. Li J, Sakata A, He H, Bai LP, Murata A, Dohno C, Nakatani K: **Naphthyridine-benzoazaquinoline: evaluation of a tricyclic system for the binding to (CAG)_n repeat DNA and RNA.** *Chem Asian J* 2016, **11**:1971–1981.
 43. Matthes F, Massari S, Bochicchio A, Schorpp K, Schilling J, Weber S, Offermann N, Desantis J, Wanker E, Carloni P, *et al.*: **Reducing mutant huntingtin protein expression in living cells by a newly identified RNA CAG binder.** *ACS Chem Neurosci* 2018, **9**:1399–1408.
 44. Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ: **Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X.** *Neurology* 2001, **57**:127–130.
 45. Green KM, Sheth UJ, Flores BN, Wright SE, Sutter AB, Kearse MG, Barmada SJ, Ivanova MI, Todd PK: **High-throughput screening yields several small-molecule inhibitors of repeat-associated non-AUG translation.** *J Biol Chem* 2019, **294**:18624–18638.
- The authors discovered five small molecules that inhibition repeat-associated non-AUG (RAN) but not canonical translation of the RNA repeat expansion that causes FXTAS, r(CGG)^{exp}. The activities of three of the molecules were traced to binding the RNA repeat expansion and interestingly also inhibit RAN translation from the structurally related repeat expansion that causes c9ALS/FTD.
46. Verma AK, Khan E, Mishra SK, Jain N, Kumar A: **Piperine modulates protein mediated toxicity in fragile X-associated tremor/ataxia syndrome through interacting expanded CGG repeat (r(CGG)^{exp}) RNA.** *ACS Chem Neurosci* 2019, **10**:3778–3788.
 47. Verma AK, Khan E, Mishra SK, Mishra A, Charlet-Berguerand N, Kumar A: **Curcumin regulates the r(CGG)^{exp} RNA hairpin structure and ameliorate defects in fragile X-associated tremor ataxia syndrome.** *Front Neurosci* 2020, **14**:295.
 48. Zhou B, Liu C, Geng Y, Zhu G: **Topology of a G-quadruplex DNA formed by C9orf72 hexanucleotide repeats associated with ALS and FTD.** *Sci Rep* 2015, **5**:16673.
 49. Wang ZF, Ursu A, Childs-Disney JL, Guertler R, Yang WY, Bernat V, Rzuczek SG, Fuerst R, Zhang YJ, Gendron TF, *et al.*: **The hairpin form of r(G₄C₂)^{exp} in c9ALS/FTD is repeat-associated non-ATG translated and a target for bioactive small molecules.** *Cell Chem Biol* 2019, **26**:179–190 e112.
 50. Simone R, Balendra R, Moens TG, Preza E, Wilson KM, Heslegrave A, Woodling NS, Niccoli T, Gilbert-Jaramillo J, Abdelkarim S, *et al.*: **G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo.** *EMBO Mol Med* 2018, **10**:22–31.
- The authors identified an selective binder to the G-quadruplex formed by the RNA repeat expansion, r(G₄C₂)^{exp}, that causes c9ALS/FTD. Stabilization of the G-quadruplex structure by the small molecule was confirmed by FRET and circular dichroism studies. These compounds were then shown to reduce RNA foci levels in c9ALS patient-derived iPSCs and a *Drosophila* model and levels of toxic dipeptide repeats in iPSCs.
51. Wortman MJ, Dagdanova AV, Clark AM, Godfrey EW, Pascal SM, Johnson EM, Daniel DC: **A synthetic Pur-based peptide binds and alters G-quadruplex secondary structure present in the expanded RNA repeat of C9orf72 ALS/FTD.** *Biochim Biophys Acta Mol Cell Res* 2020, **1867**:118674.
 52. Zamiri B, Reddy K, Macgregor Jr RB, Pearson CE: **Tmppy4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)_n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins.** *J Biol Chem* 2014, **289**:4653–4659.
- The authors show that the RNA repeat expansion that causes c9ALS/FTD folds into a G-quadruplex structure. The c9ALS/FTD quadruplex binds to the known G-quadruplex binding-ligand TmPyP₄ and inhibits its interactions with various RNA-binding proteins. These studies were one of the first to show that the c9ALS/FTD repeat expansion is ligandable and hence might be a target for therapeutic intervention.
53. Connelly CM, Boer RE, Moon MH, Gareiss P, Schneekloth Jr JS: **Discovery of inhibitors of microRNA-21 processing using**

- small molecule microarrays.** *ACS Chem Biol* 2017, **12**: 435–443.
54. Lorenz DA, Vander Roest S, Larsen MJ, Garner AL: **Development and implementation of an HTS-compatible assay for the discovery of selective small-molecule ligands for pre-microRNAs.** *SLAS Discov* 2018, **23**:47–54.
 55. Staedel C, Tran TPA, Giraud J, Darfeuille F, Di Giorgio A, Tourasse NJ, Salin F, Uriac P, Duca M: **Modulation of oncogenic miRNA biogenesis using functionalized polyamines.** *Sci Rep* 2018, **8**:1667.
 56. Vo DD, Becquart C, Tran TPA, Di Giorgio A, Darfeuille F, Staedel C, Duca M: **Building of neomycin-nucleobase-amino acid conjugates for the inhibition of oncogenic miRNAs biogenesis.** *Org Biomol Chem* 2018, **16**:6262–6274.
 57. Matarlo JS, Krumpe LRH, Heinz WF, Oh D, Shenoy SR, Thomas CL, Goncharova EI, Lockett SJ, O'Keefe BR: **The natural product butylcycloheptyl prodiginine binds pre-miR-21, inhibits Dicer-mediated processing of pre-miR-21, and blocks cellular proliferation.** *Cell Chem Biol* 2019, **26**:1133–1142 e1134.
- The authors used high throughput screening to identify natural products that bind the precursor of miR-21. They described butyl-cycloheptyl prodiginine, which inhibited Dicer processing by binding pre-miR-21 with a K_d of 0.41 μ M. They showed that this compound modulated miR-21 processing, causing up-regulation of downstream miR-21 targets PTEN and PDCD4.
58. Naro Y, Ankenbruck N, Thomas M, Tivon Y, Connelly CM, Gardner L, Deiters A: **Small molecule inhibition of microRNA miR-21 rescues chemosensitivity of renal-cell carcinoma to topotecan.** *J Med Chem* 2018, **61**:5900–5909.
 59. Jiang CS, Wang XM, Zhang SQ, Meng LS, Zhu WH, Xu J, Lu SM: **Discovery of 4-benzoylamino-N-(prop-2-yn-1-yl)benzamides as novel microRNA-21 inhibitors.** *Bioorg Med Chem* 2015, **23**: 6510–6519.
 60. Naro Y, Thomas M, Stephens MD, Connelly CM, Deiters A: **Aryl amide small-molecule inhibitors of microRNA miR-21 function.** *Bioorg Med Chem Lett* 2015, **25**:4793–4796.
 61. Shortridge MD, Walker MJ, Pavelitz T, Chen Y, Yang W, Varani G: **A macrocyclic peptide ligand binds the oncogenic microRNA-21 precursor and suppresses Dicer processing.** *ACS Chem Biol* 2017, **12**:1611–1620.
- Here, the authors showed that peptide-based ligands, in particular a cyclic β -hairpin peptidomimetic, can bind to the Dicer site of oncogenic pre-miR-21 specifically and inhibit its processing by Dicer in vitro and in cells. MiRNA precursors are likely particularly well suited for this type of modality that binds RNA hairpins, which can be customized/ designed to afford selective binding. Indeed, these studies also reported the NMR solution structure, which could inform optimization of the lead molecule.
62. Costales MG, Aikawa H, Li Y, Childs-Disney JL, Abegg D, Hoch DG, Pradeep Velagapudi S, Nakai Y, Khan T, Wang KW, et al.: **Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer.** *Proc Natl Acad Sci U S A* 2020, **117**: 2406–2411.
- The authors showed that ribonuclease targeting chimeras (RIBOTACs) can selectively target and degrade an oncogenic miRNA in cells and in a mouse model of metastatic breast cancer by recruiting endogenous RNase L. Importantly, the chimera comprises a small molecule that binds oncogenic pre-miR-21 and a heterocycle, rather than 2'-5'A₄, to recruit RNase L.
63. Diaz JP, Chirayil R, Chirayil S, Tom M, Head KJ, Luebke KJ: **Association of a peptoid ligand with the apical loop of pri-miR-21 inhibits cleavage by Drosha.** *RNA* 2014, **20**: 528–539.
 64. Haniff HS, Knerr L, Liu X, Crynen G, Bostrom J, Abegg D, Adibekian A, Lekah E, Wang KW, Cameron MD, et al.: **Design of a small molecule that stimulates vascular endothelial growth factor a enabled by screening RNA fold-small molecule interactions.** *Nat Chem* 2020, **12**:952–961.
 65. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.** *Adv Drug Deliv Rev* 2001, **46**:3–26.
 66. Dominski Z, Kole R: **Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides.** *Proc Natl Acad Sci U S A* 1993, **90**:8673–8677.
 67. Naryshkin NA, Weetall M, Dakka A, Narasimhan J, Zhao X, Feng Z, Ling KK, Karp GM, Qi H, Woll MG, et al.: **Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy.** *Science* 2014, **345**:688–693.
 68. Salani M, Urbina F, Brenner A, Morini E, Shetty R, Gallagher CS, Law EA, Sunshine S, Finneran DJ, Johnson G, et al.: **Development of a screening platform to identify small molecules that modify ELP1 pre-mRNA splicing in familial dysautonomia.** *SLAS Discov* 2019, **24**:57–67.
 69. Campagne S, Boigner S, Rudisser S, Moursy A, Gillioz L, * Knorlein A, Hall J, Ratni H, Clery A, Allain FH: **Structural basis of a small molecule targeting RNA for a specific splicing correction.** *Nat Chem Biol* 2019, **15**:1191–1198.
- The authors described a 5' splice site bulge repairing small molecule. Further, a solution structure of the target RNA in complex with the splicing modifier revealed how the drug selectively promotes 5' splicing. This thorough investigation of RNA structure stabilization illustrates the potential for RNA targeted splicing modifiers, enabling specific inclusion of SMN2 exon 7.
70. Chen JL, Zhang P, Abe M, Aikawa H, Zhang L, Frank AJ, Zembryski T, Hubbs C, Park H, Withka J, et al.: **Design, optimization, and study of small molecules that target tau pre-mRNA and affect splicing.** *J Am Chem Soc* 2020, **142**: 8706–8727.
 71. Lopez-Senin P, Gomez-Pinto I, Grandas A, Marchan V: **Identification of ligands for the tau exon 10 splicing regulatory element RNA by using dynamic combinatorial chemistry.** *Chemistry* 2011, **17**:1946–1953.
 72. Zheng S, Chen Y, Donahue CP, Wolfe MS, Varani G: **Structural basis for stabilization of the tau pre-mRNA splicing regulatory element by novantrone (mitoxantrone).** *Chem Biol* 2009, **16**:557–566.
 73. Brunden KR, Trojanowski JQ, Lee VM: **Advances in tau-focused drug discovery for alzheimer's disease and related tauopathies.** *Nat Rev Drug Discov* 2009, **8**:783–793.
 74. Vaklavas C, Meng Z, Choi H, Grizzle WE, Zinn KR, Blume SW: **Small molecule inhibitors of IRES-mediated translation.** *Cancer Biol Ther* 2015, **16**:1471–1485.
 75. Miglietta G, Cogoi S, Marinello J, Capranico G, Tikhomirov AS, Shchekotikhin A, Xodo LE: **RNA G-quadruplexes in kirsten ras (KRAS) oncogene as targets for small molecules inhibiting translation.** *J Med Chem* 2017, **60**:9448–9461.
 76. Dai J, Liu ZQ, Wang XQ, Lin J, Yao PF, Huang SL, Ou TM, Tan JH, Li D, Gu LQ, et al.: **Discovery of small molecules for up-regulating the translation of antiamyloidogenic secretase, a disintegrin and metalloproteinase 10 (ADAM10), by binding to the G-quadruplex-forming sequence in the 5' untranslated region (UTR) of its mRNA.** *J Med Chem* 2015, **58**:3875–3891.
 77. Katsuda Y, Sato S, Asano L, Morimura Y, Furuta T, Sugiyama H, Hagihara M, Uesugi M: **A small molecule that represses translation of G-quadruplex-containing mRNA.** *J Am Chem Soc* 2016, **138**:9037–9040.
 78. Collie GW, Sparapani S, Parkinson GN, Neidle S: **Structural basis of telomeric RNA quadruplex-acridine ligand recognition.** *J Am Chem Soc* 2011, **133**:2721–2728.
 79. Garavis M, Lopez-Mendez B, Somoza A, Oyarzabal J, Dalvit C, Villasante A, Campos-Olivas R, Gonzalez C: **Discovery of selective ligands for telomeric RNA G-quadruplexes (TERRA) through ¹⁹F-NMR based fragment screening.** *ACS Chem Biol* 2014, **9**:1559–1566.
 80. Abulwerdi FA, Xu W, Ageeli AA, Yonkunas MJ, Arun G, Nam H, Schneekloth Jr JS, Dayie TK, Spector D, Baird N, et al.: **Selective small-molecule targeting of a triple helix encoded by the long noncoding RNA, MALAT1.** *ACS Chem Biol* 2019, **14**:223–235.
- The authors showed that small molecules can indeed selectively recognize structures formed by long noncoding RNAs and affect function, in this case the 3' terminal stability element for nuclear expression (ENE) found in oncogenic MALAT1. Indeed, two

compounds that bind the triple helix reduced MALAT1 levels in a mammary tumor model.

81. Parker CG, Galmozzi A, Wang Y, Correia BE, Sasaki K, Joslyn CM, Kim AS, Cavallaro CL, Lawrence RM, Johnson SR, *et al.*: **Ligand and target discovery by fragment-based screening in human cells.** *Cell* 2017, **168**: 527–541 e529.
 82. Suresh BM, Li W, Zhang P, Wang KW, Yildirim I, Parker CG, Disney MD: **A general fragment-based approach to identify and optimize bioactive ligands targeting RNA.** *Proc Natl Acad Sci U S A* 2020, **117**:33197–33203.
- The authors reported the adoption of the fully functionalized fragment (FFF) approach developed for protein targets to identify new ligands for RNA. They showed that FFF could be used to identify ligands that bind cooperatively, which could be tethered together to afford heterodimeric compounds. Assembly of dimers improved affinity and specificity, while still occluding the functional site within miR-21 required for biogenesis. These studies were then validated in cells with RNA-seq experiments, confirming direct engagement of the target RNA.
83. Gosink MM, Vierstra RD: **Redirecting the specificity of ubiquitination by modifying ubiquitin-conjugating enzymes.** *Proc Natl Acad Sci U S A* 1995, **92**:9117–9121.
 84. Sakamoto KM, Kim KB, Kumagai A, Mercurio F, Crews CM, Deshaies RJ: **Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation.** *Proc Natl Acad Sci U S A* 2001, **98**:8554–8559.
 85. Costales MG, Matsumoto Y, Velagapudi SP, Disney MD: **Small molecule targeted recruitment of a nuclease to RNA.** *J Am Chem Soc* 2018, **140**:6741–6744.
 86. Costales MG, Suresh B, Vishnu K, Disney MD: **Targeted degradation of a hypoxia-associated non-coding RNA enhances the selectivity of a small molecule interacting with RNA.** *Cell Chem Biol* 2019, **26**:1180–1186. e1185.
 87. Bondeson DP, Smith BE, Burslem GM, Buhimschi AD, Hines J, Jaime-Figueroa S, Wang J, Hamman BD, Ishchenko A, Crews CM: **Lessons in PROTAC design from selective degradation with a promiscuous warhead.** *Cell Chem Biol* 2018, **25**:78–87. e75.
- These pioneering studies showed that the selectivity of a promiscuous kinase inhibitor (affects >100 kinases) can be vastly improved by its conversion into a proteolysis targeting chimera, or PROTAC. Further, these studies showed that PROTAC activity is traced to stabilization of the ternary complex formed by the PROTAC, target protein, and recruited enzyme.
88. Yan H, Liang FS: **miRNA inhibition by proximity-enabled Dicer inactivation.** *Methods* 2019, **167**:117–123.
 89. Qu L, Yi Z, Zhu S, Wang C, Cao Z, Zhou Z, Yuan P, Yu Y, Tian F, Liu Z, *et al.*: **Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs.** *Nat Biotechnol* 2019, **37**:1059–1069.
 90. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, *et al.*: **Landscape and variation of RNA secondary structure across the human transcriptome.** *Nature* 2014, **505**:706–709.
 91. Liu J, Dou X, Chen C, Chen C, Liu C, Xu MM, Zhao S, Shen B, Gao Y, Han D, *et al.*: **N(6)-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription.** *Science* 2020, **367**:580–586.

Massively Parallel Optimization of the Linker Domain in Small Molecule Dimers Targeting a Toxic r(CUG) Repeat Expansion

Simon Vezina-Dawod, Alicia J. Angelbello, Shruti Choudhary, Kye Won Wang, Ilyas Yildirim, and Matthew D. Disney*



Cite This: *ACS Med. Chem. Lett.* 2021, 12, 907–914



Read Online

ACCESS |



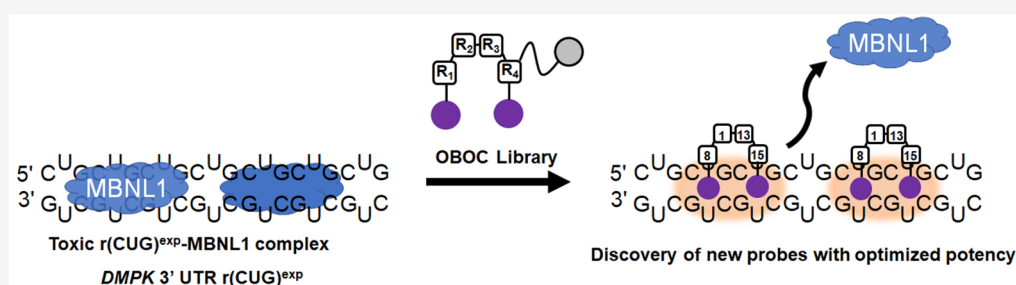
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: RNA contributes to disease pathobiology and is an important therapeutic target. The downstream biology of disease-causing RNAs can be short-circuited with small molecules that recognize structured regions. The discovery and optimization of small molecules interacting with RNA is, however, challenging. Herein, we demonstrate a massively parallel one-bead-one-compound methodology, employed to optimize the linker region of a dimeric compound that binds the toxic r(CUG) repeat expansion [r(CUG)^{exp}] causative of myotonic dystrophy type 1 (DM1). Indeed, affinity selection on a 331,776-member library allowed the discovery of a compound with enhanced potency both in vitro (10-fold) and in DM1-patient-derived myotubes (5-fold). Molecular dynamics simulations revealed additional interactions between the optimized linker and the RNA, resulting in ca. 10 kcal/mol lower binding free energy. The compound was conjugated to a cleavage module, which directly cleaved the transcript harboring the r(CUG)^{exp} and alleviated disease-associated defects.

KEYWORDS: RNA, small molecule targeting, repeat expansion, targeted degradation, one-bead-one-compound, myotonic muscular dystrophy

Targeting RNA with synthetic molecules was long thought to be challenging because of RNA's limited local diversity from only four nucleotides and its flexible and dynamic nature. Many RNAs, however, adopt stable three-dimensional (3D) structures, which are potential binding sites for small molecules.^{1–6} The discovery of selective small molecules targeting disease-causing RNA has historically been accomplished through selection methods⁷ or in various targeted screens for specific RNA structures.^{8,9} Suitable RNA targets for small molecules were identified in a myriad of diseases, such as cancers caused by oncogenic microRNA (miR)-21^{10,11} or spinal muscular atrophy.¹² To identify and optimize novel chemical matter targeting toxic RNA structures, robust combinatorial methods are required. Advanced combinatorial chemistry methods, such as DNA-encoded libraries,^{13–16} one-bead-one-compound (OBOC) libraries,^{17,18} phage display,^{19,20} and affinity selection mass spectrometry,^{21,22} have been successfully applied to proteins and can be adapted for RNA. For example, Patel and co-workers reported a one-bead-two-compounds (OBTC) strategy to discover macrocycles targeting the long noncoding RNA growth arrest specific 5

(GASS), with implications for the treatment of type 2 diabetes mellitus.²³ The application of combinatorial methods to disease-causing RNAs enables the discovery of novel chemical probes and optimization of ligands targeting these RNA structures in a massively parallel fashion.

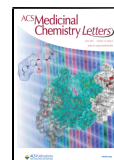
One class of toxic RNAs is repeat expansions that cause greater than 40 genetic diseases including myotonic dystrophy type 1 (DM1). DM1 is caused by a r(CUG) repeat expansion [r(CUG)^{exp}] located in the 3' untranslated region (UTR) of the *dystrophin myotonia protein kinase* (DMPK) mRNA.²⁴ DM1 is an incurable neuromuscular disease with a prevalence of ~1 in 8000 people. Those afflicted with DM1 have symptoms that include muscle weakness and myotonia, heart abnormalities, cataracts, and insulin resistance.^{25,26} The

Special Issue: RNA: Opening New Doors in Medicinal Chemistry

Received: January 13, 2021

Accepted: February 23, 2021

Published: March 2, 2021



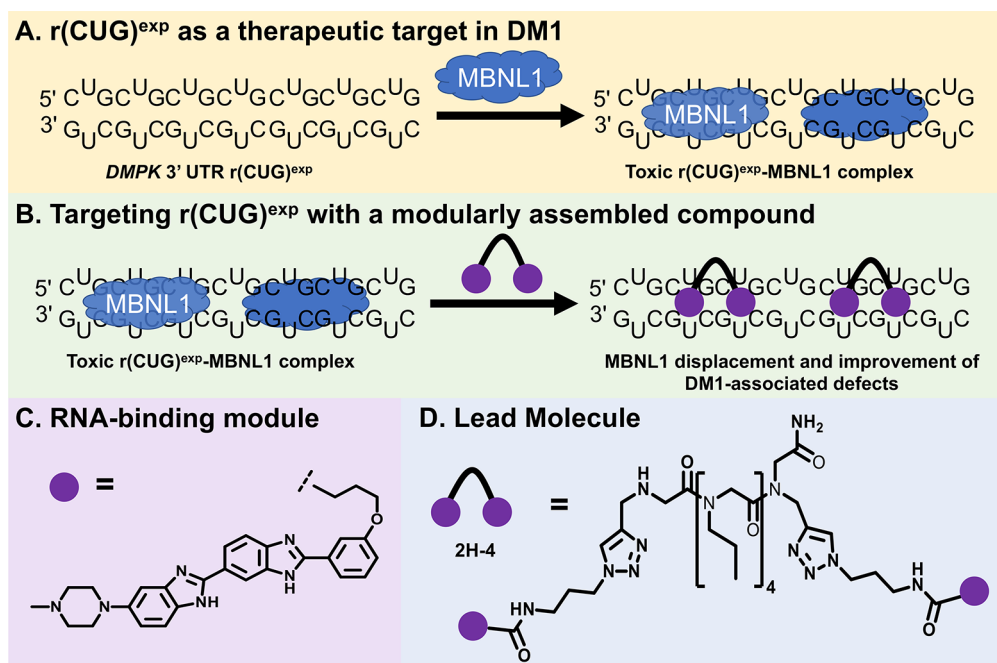


Figure 1. Small molecule targeting approach for the toxic RNA repeat expansion, r(CUG)^{exp}, that causes myotonic dystrophy type 1 (DM1). (A) Toxic r(CUG)^{exp} present in the 3' UTR of the *DMPK* mRNA sequesters MBNL1 protein, a regulator of alternative pre-mRNA splicing, with high affinity, leading to splicing defects. (B) Modularly assembled RNA-binding modules bind with high affinity to r(CUG)^{exp}, liberating MBNL1 and rescuing DM1-associated splicing defects. (C) Chemical structure of the 5' CUG/3' GUC [loops present in r(CUG)^{exp}] RNA-binding module used in this study. (D) Chemical structure of a previously developed dimeric compound that rescues DM1-associated defects, **2H-4**. **2H-4** was used as a starting point to develop the OBOC methodology described herein.

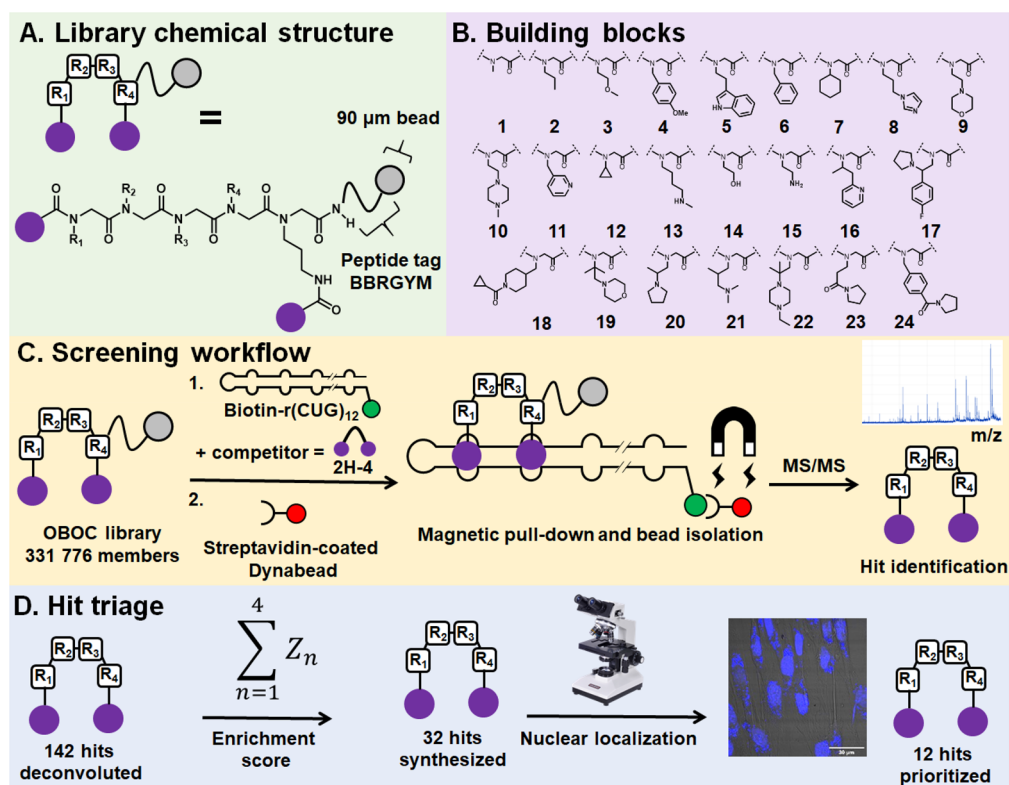


Figure 2. OBOC library and screening method to optimize dimeric compounds targeting r(CUG)^{esp}. (A) Chemical structure of the library, with R₁₋₄ indicating variable residues. (B) Building blocks incorporated to generate the chemical diversity of the 331,776-member library. (C) Screening workflow using compounds supported on 90 μm Tentagel beads. (1) Incubation with biotinylated r(CUG)₁₂; (2) magnetic pull-down using streptavidin-coated Dynabeads; (3) bead isolation and cleavage; (4) hit structure deconvolution by tandem mass spectrometry. (D) Hit triage of the 142 hits by enrichment score and nuclear localization in DM1 myotubes.

r(CUG)^{exp} folds into a highly structured hairpin, the stem of which comprises a periodic array of 5'CUG/GUC 1 × 1 internal loops. These loops bind and sequester the pre-mRNA splicing regulator muscleblind-like 1 (MBNL1) protein as well as other RNA-binding proteins in nuclear foci (Figure 1A). Sequestration of MBNL1 by r(CUG)^{exp} prevents its interaction with its natural substrates, therefore leading to dysregulation of alternative pre-mRNA splicing and manifestation of disease.^{27,28} Different types of modalities can bind to r(CUG)^{exp} and improve disease defects in DM1-affected cells,^{29–31} including monomeric small molecules,^{31–33} modularly assembled dimeric compounds that occupy multiple internal loops simultaneously,^{30,34} pseudopeptides,³⁵ and small molecules that cleave r(CUG)^{exp}.³⁶ These compounds were discovered by various approaches, such as a bead-based screening method called resin-bound dynamic combinatorial chemistry³⁵ and by fragment-based target-guided synthesis.³⁴

Our group has previously reported the modular assembly of RNA-binding modules targeting consecutive 5'CUG/GUC 1 × 1 internal loops to disrupt the toxic MBNL1–r(CUG)^{exp} complex and improve DM1-associated defects (Figure 1B,C).³⁷ A dimeric compound named **2H-4** was built by a *N*-propylglycine peptoid bridge to separate the binding modules at a specific distance to enable simultaneous binding (Figure 1D). Efforts to improve the binding affinity and potency of **2H-4** have involved changing the nature of its backbone including an *N*-methyl alanine linker,³⁸ a proline linker,³⁹ and macrocyclization.⁴⁰ These modifications have improved binding affinity and cellular activity; however, thus far, the optimization of the linker between RNA-binding modules has not been attempted on a large scale.

Peptoids, oligomers of *N*-substituted glycine building blocks, offer access to a large chemical diversity by simple incorporation of primary amine building blocks. Further, peptoids have desirable pharmacological properties, such as cellular permeability and resistance to proteolytic degradation.^{41,42} As their syntheses are fully compatible with OBOC library approaches,⁴³ we envisioned a simple method to introduce a wide variety of functional groups in the linker domain of **2H-4**, which is described herein.

Briefly, the OBOC library was synthesized on 90 μm Tentagel microbeads with four variable positions between the two 5'CUG/GUC internal loop binding modules, as this is the optimal linker length identified from previous studies (Figure 2A).³⁸ A peptide tag, BBRGYM, was incorporated at the C-terminal of each compound, allowing release from the beads by selective cyanogen bromide (CNBr) cleavage of the M residue for facile deconvolution of compound identity by tandem mass spectrometry (MS–MS). Indeed, fragmentation of the amide bonds generates a mass ladder that represents the nature of the different building blocks and their sequential order in the compound. The peptide tag provides a known starting point in the MS–MS spectrum from which unknown building blocks are identified. A set of 24 different building blocks were incorporated by split-and-pool methodology (Figure 2B), generating a theoretical diversity of 331,776 compounds. Building blocks were selected to cover various chemotypes, such as aliphatic, cationic, aromatic, or heteroaromatic moieties. As a quality control step, random beads from the library were selected and processed for compound deconvolution, and indeed MS–MS analysis enabled structure deconvolution unambiguously (Figure S1).

To identify compounds that avidly bind r(CUG)^{exp}, the OBOC library was incubated with 150 nM of an established structural model of r(CUG)^{exp}, r(CUG)₁₂ containing a 5' biotin tag,⁴⁴ in the presence of 15 μM of **2H-4** (100-fold molar excess as compared to the RNA's concentration). The **2H-4** competitor was added to increase the stringency of the screen and to allow identification of compounds that bind more avidly to r(CUG)₁₂ than the competitor. Beads bound to 5'-biotin-r(CUG)₁₂ were isolated with streptavidin-coated Dynabeads (Figure 2C) and identified by MS–MS analysis, affording 142 hit compounds.

To triage the hits to a more manageable number for further investigation, we calculated the enrichment and the statistical confidence thereof (Z-score) of each building block at each variable position (Figure 2D). (A Z-score of 1.96 corresponds to *P* = 0.05 and hence statistical significance.) The Z-score at each position was also summed for each hit compound (Figure S2). Several trends could be observed in selection among the building blocks, including (i) a depletion of tertiary amines, particularly building blocks #19–22, at positions 1–3 (*P* = 0.013 to 0.039) and (ii) enrichment of secondary and primary amines, building blocks #13 and #15, at all positions (*P* < 0.0001) (Figure S3).

We observed that the sum of Z-scores for the building blocks comprising hit compounds correlated with positive charge (Figure S4). As the RNA backbone is negatively charged, this is not surprising. To further investigate this observation as well as to maintain chemical diversity, we binned hit compounds by their linker charge and then selected those in each bin with Z-scores in the top 30%. This first triage provided a reasonable number of compounds (*n* = 32) to synthesize by solid-phase parallel synthesis on Rink amide polystyrene resin, to measure avidity for bind r(CUG)^{exp}, and to study for rescue of DM1-associated defects.

In previous studies of dimeric compounds targeting r(CUG)^{exp}, we observed that cell uptake and localization can change as a function of the linker's structure.^{39,40} Furthermore, charged backbones are known to affect localization.⁴⁵ Because nuclear localization is crucial for biological activity as r(CUG)^{exp} is sequestered in nuclear foci, we studied the compounds' cellular permeability and localization.

DM1 patient-derived myotubes containing 1300 r(CUG) repeats were treated with 5 μM of each compound, and localization was monitored using the inherent fluorescence of the **H** RNA-binding modules. Interestingly, **2H-4** localizes almost completely in the nucleus, as do 12 of the 32 hit OBOC compounds (Figure S5), which were thus prioritized for further study. A closer inspection of building block distribution provided insights into the relationship between compound structure and nuclear localization (Figure S6). Among the nuclear localized compounds, only the polar building block #14 (*N*-hydroxyethyl substituent) was found to be significantly enriched at position 2 (17%, *P* = 0.03). Within the compounds with poor nuclear localization, building blocks #2 (*N*-propyl substituent) and #13 (*N*-methylaninobutyl substituent) were found to be significantly enriched at positions 3 (20%, *P* = 0.03) and 2 (33%, *P* = 0.049), respectively. These data suggest that the position where polar and hydrophobic moieties are incorporated within the dimer linker influence localization.

MBNL1 protein self-regulates the alternative splicing of its exon 5, which is included too frequently in DM1 cells.⁴⁶ Thus, we evaluated which of the 12 compounds rescue the MBNL1 exon 5 splicing defect in DM1-patient-derived myotubes and

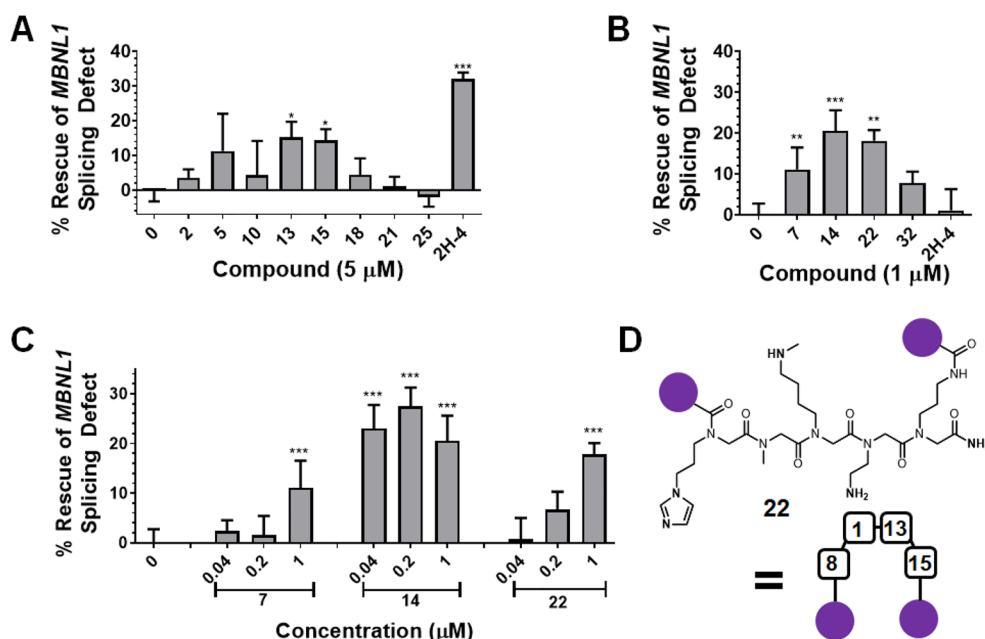


Figure 3. Evaluation of hit compound activity in DM1 patient-derived myotubes. (A) Activity of hit compounds with no toxicity at 5 μM , as assessed by rescue of the *MBNL1* exon 5 splicing defect in DM1 myotubes ($n = 3$). (B) Activity of hit compounds with no toxicity at 1 μM , as assessed by rescue of the *MBNL1* exon 5 splicing defect in DM1 myotubes ($n = 3$). (C) Dose–response analysis of compounds with statistically significant activity at 1 μM ($n = 6$). For all panels: error bars represent standard deviation, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, as determined by one-way ANOVA. (D) Chemical structure of compound 22.

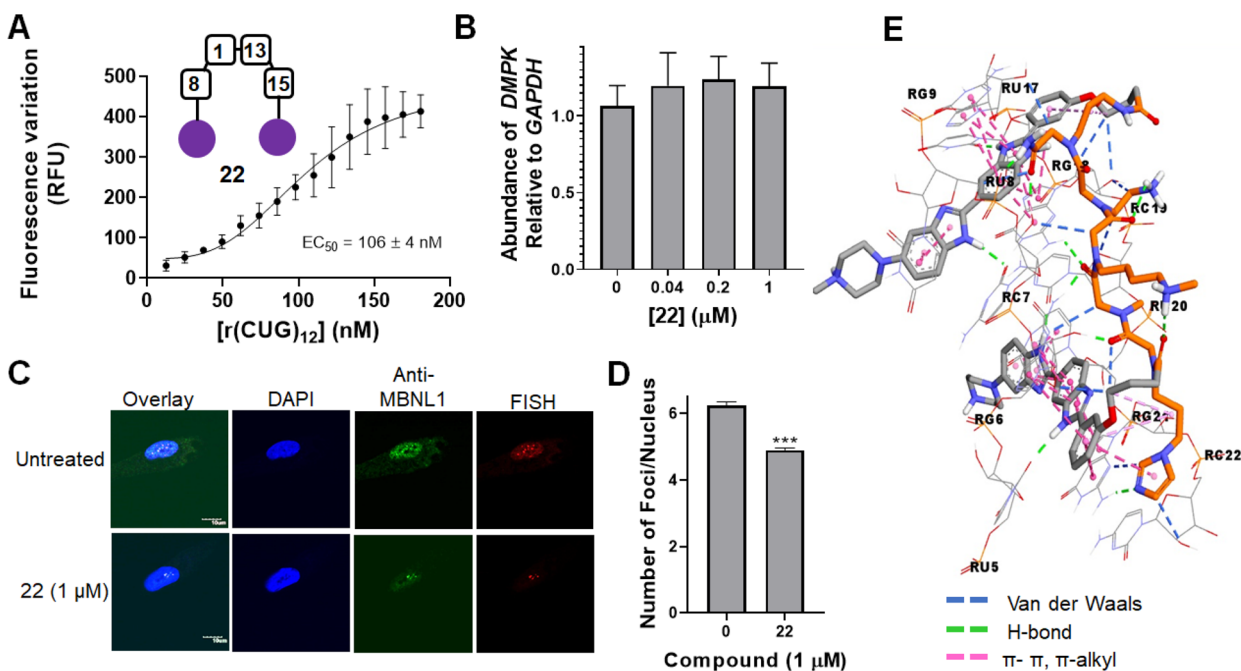


Figure 4. Compound **22** binds avidly to r(CUG)₁₂ in vitro and selectively improves disease-associated defects in DM1 myotubes. (A) Representative direct binding curve of **22** and r(CUG)₁₂ hairpin using the inherent fluorescence of **22** ($n = 3$). (B) Evaluation of DMPK levels in DM1 myotubes treated with **22** via RT-qPCR ($n = 3$). (C) Representative confocal microscopy images of nuclear foci. MBNL1 was imaged by immunofluorescence, whereas r(CUG)^{exp} was imaged by RNA fluorescence in situ hybridization (FISH). (D) Quantification of the number of foci/nucleus in DM1 myotubes treated with **22** ($n = 3$; 40 nuclei counted/replicate). *** $P < 0.001$ Student's *t*-test. (E) Interactions between **22** and two 5' CUG/3' GUC internal loops present in r(CUG)^{exp} from a model generated by molecular dynamics simulation. Dashed lines represent interactions between the linker residues of **22** and the RNA.

compared their activities to 2H-4 (Figure 3). DM1 cells were treated with a 5 μ M dose for 2, 5, 10, 13, 15, 18, 21, and 25 for 48 h. Changes in cellular morphology were observed upon treatment with 5 μ M of 7, 14, 22, and 32, and therefore these

compounds were evaluated at 1 μ M, where no changes in morphology were observed. Of the eight compounds evaluated at 5 μ M, only two slightly improved the *MBNL1* exon 5 splicing defect, by $\sim$ 20%, which is less than the effect observed

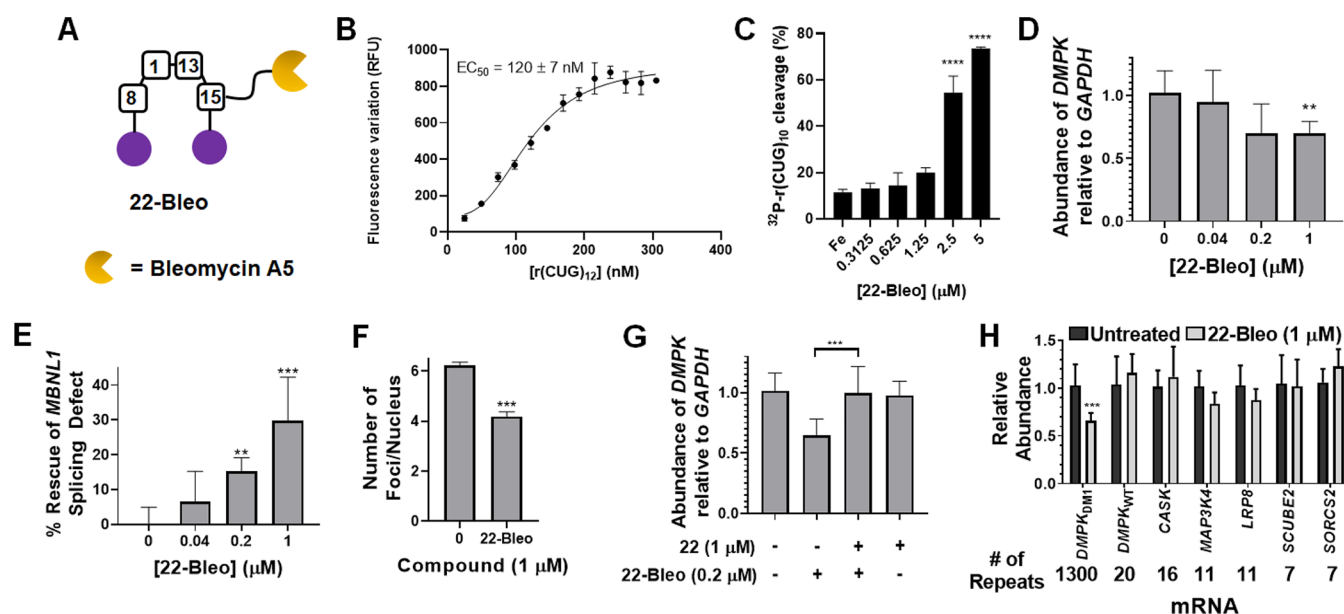


Figure 5. Compound **22-Bleo** cleaves $r(\text{CUG})^{\text{exp}}$ in vitro and in DM1 myotubes, rescuing disease-associated defects. (A) Conjugation of **22** to bleomycin A5 afforded cleaving compound **22-Bleo**. (B) Direct binding assay of **22-Bleo** with $r(\text{CUG})_{12}$ using the inherent fluorescence of **22** ($n = 3$). Note: assays were completed in the absence of Fe^{2+} , required for cleavage. (C) In vitro cleavage activity of **22-Bleo** using radioactively labeled $r(\text{CUG})_{10}$ and analysis of fragments by gel electrophoresis ($n = 3$); **** $P < 0.0001$, as determined by one-way ANOVA. (D) Evaluation of *DMPK* levels upon compound treatment via RT-qPCR ($n = 6$); ** $P < 0.01$, as determined by one-way ANOVA. (E) Ability of **22-Bleo** to improve the *MBNL1* exon 5 splicing defect in DM1 myotubes ($n = 6$); ** $P < 0.01$, *** $P < 0.001$, as determined by one-way ANOVA. (F) Quantification of the number of foci/nucleus in DM1 myotubes treated with **22-Bleo** ($n = 3$; 40 nuclei counted/replicate); *** $P < 0.001$, as determined by a Student's *t*-test. (G) Competitive cleavage experiment between **22** and **22-Bleo** where excess **22** prevents cleavage of *DMPK* by **22-Bleo**. Levels of *DMPK* mRNA were measured by RT-qPCR ($n = 6$); *** $P < 0.001$, as determined by one-way ANOVA. (H) Abundance of $r(\text{CUG})$ repeat-containing transcripts upon treatment with **22-Bleo**, as measured by RT-qPCR ($n = 6$); *** $P < 0.001$, as determined by a Student's *t*-test. For all panels, error bars represent standard deviation.

upon treatment with **2H-4** (~30% improvement) (Figures 3A and S7). The four compounds evaluated at $1 \mu\text{M}$ all improved the *MBNL1* exon 5 splicing defect more potently than did **2H-4**, which is inactive at the $1 \mu\text{M}$ dose (Figures 3B and S7), although the percent rescue by **32** was not statistically significant. Notably, **7** and **22** improved *MBNL1* exon 5 splicing dose-dependently, with **22** appearing to be the more potent of the two (7% rescue for **22** vs 2% rescue for **7** at $0.2 \mu\text{M}$) (Figure 3C). As **14** improved splicing similarly at all concentrations tested, it was eliminated from further study (Figures 3C and S8). Since **22** was the most potent, it was studied in more detail, both in vitro and in DM1-patient-derived myotubes.

We first measured the binding of **22** to $r(\text{CUG})_{12}$ in vitro, affording an EC_{50} of $106 \pm 4 \text{ nM}$ (Figure 4A), ~10-fold more avid than **2H-4** ($\text{EC}_{50} = 1140 \pm 31 \text{ nM}$).⁴⁰ Importantly, no saturable binding was observed for **22** and a base-paired control RNA that does not contain the 1×1 nucleotide UU internal loops (Figure S9). This avid binding of **22** to $r(\text{CUG})$ repeats translated to inhibition of the formation of the $r(\text{CUG})^{\text{exp}}$ -*MBNL1* complex in vitro with an IC_{50} of $2.8 \pm 0.2 \mu\text{M}$ in a TR-FRET assay (Figure S10),⁴⁴ an ~10-fold improvement over **2H-4** (IC_{50} of $32.2 \pm 4.3 \mu\text{M}$; determined by the same TR-FRET assay),⁴⁰ which correlates with its enhanced affinity.

To rationalize the improved in vitro binding of **22** to $r(\text{CUG})_{12}$ when compared to that of **2H-4**, molecular modeling studies were carried out (Tables S2–S5 and Figures S11–S14). Models of dimer **22** bound to a model RNA with two 1×1 nucleotide UU internal loops in $r(\text{CUG})^{\text{exp}}$ [$r(\text{S}'\text{-CCGCUGCUGCGG}/3'\text{GGCGUCGUCGCC})$] were gener-

ated using molecular dynamics simulations and compared with a previously published model of **2H-4**. The best model, as defined by the lowest free binding energy, generated for **22** was 10 kcal/mol lower than that of **2H-4** (−61.7 vs −51.7 kcal/mol).

Inspection of the interactions of **22** with the internal loops shows an increased number of interactions between the dimer's linker region and the RNA backbone, in addition to the interactions of RNA-binding modules with the RNA, which are similar to those of **2H-4** (Figure 4E). The imidazole group of the building block in position 1 hydrogen bonds with G21's NH_2 [the bolded nucleotide in $r(\text{S}'\text{-CCGCUGCUGCGG}/3'\text{GGCGUCGUCGCC})$] and stacks with the RNA-binding module. The carbonyl groups in the backbone undergo a number of hydrogen bonding interactions, including intramolecular hydrogen bonds with the NH-CH_3 side chain of the building block in position 3 and NH_2 of the side chain of the building block in position 4. Intermolecular H-bonding interactions of the dimer backbone with the RNA also contribute significantly to the binding of the dimer, including (i) interaction of building block 2's carbonyl with 2' OH of G6; (ii) H-bond formed between the terminal amide CONH_2 with 2' OH of C7; (iii) interaction of building block 3's carbonyl with G6's and G8's NH_2 ; and (iv) H-bond between the linker's terminal amide NH_2 with the phosphate backbone of U8. These interactions, in addition to the van der Waals and π -stacking interactions formed throughout the dimer–RNA complex, contribute to the observed 10 kcal/mol improvement in the binding energy when compared to that of the bound structure of **2H-4**.

The selectivity and activity of **22** were further investigated in DM1 patient-derived myotubes. Importantly, the observed rescue of *MBNL1* exon 5 pre-mRNA splicing defect was not due to transcriptional inhibition of *DMPK*, as its RNA levels were unaffected (Figure 4B), suggesting direct r(CUG)^{exp} target engagement. Importantly, **22** specifically improved DM1-associated defects as *MAP4K4* exon 22a splicing, a NOVA-regulated splicing event,⁴⁷ was not affected in DM1 myotubes (Figure S15). Furthermore, *MBNL1* exon 5 splicing was not affected in wild-type myotubes (from healthy donors) treated with **22** (Figure S16). Compound **22** also reduced the number of r(CUG)^{exp}–*MBNL1* foci in DM1 myotubes by ~20%, which correlates with the observed improvement in splicing (Figure 4C,D). Thus, through specific binding to r(CUG)^{exp}, **22** potently and specifically improves DM1-associated defects in patient-derived myotubes.

We have previously demonstrated that conjugation of bleomycin A5 to repeat-targeting compounds allows for specific cleavage of the target RNA as well as increases potency.^{36,48} This direct cleavage by bleomycin A5 also enables target identification and analysis of selectivity as cleavage will be observed for all targets that the compound engages. Thus, we appended **22** with bleomycin A5 using its terminal amine as attachment at this site is known to decrease bleomycin's affinity for DNA and ablate off-target effects (Figure 5A).³⁶ The compound **22-Bleo** was first evaluated for its ability to selectively bind and cleave r(CUG)₁₂ in vitro. Indeed **22-Bleo** selectively bound r(CUG)₁₂ with an EC₅₀ of 120 ± 7 nM (similar to the EC₅₀ of **22**, 106 ± 4 nM; note: Fe²⁺, required for cleavage, is absent in binding assays), with no binding observed to a base-paired RNA (Figures 5B and S17). The RNA degrader also cleaved the r(CUG) repeats in vitro, and >70% of the RNA was cleaved at a 5 μM dose (Figures 5C and S18). Thus, **22-Bleo** can selectively recognize r(CUG)^{exp} and cleave it in vitro.

We next studied whether the in vitro cleavage activity of **22-Bleo** translated to cleavage of r(CUG)^{exp} in DM1 myotubes. Indeed, a ~35% reduction in the abundance of *DMPK* was observed after treatment with 1 μM of **22-Bleo** (Figure 5D). Importantly, this cleavage was specific to the disease-causing, *DMPK*-allele-harboring r(CUG)^{exp} as *DMPK* levels were not affected in wild-type cells (Figure S19). A competition experiment between **22-Bleo** and parent compound **22** was then completed to confirm the former's mode of action (cleavage rather than transcriptional inhibition) and the latter's direct engagement of r(CUG)^{exp}. As expected, co-treatment of DM1 myotubes with varying concentration of **22** and a constant concentration of **22-Bleo** afforded a dose-dependent rescue of *DMPK* levels (Figure 5G).

Specific cleavage of r(CUG)^{exp} resulted in improvement of DM1-associated defects including ~30% rescue of the *MBNL1* exon 5 splicing defect and ~30% reduction in the number of r(CUG)^{exp}–*MBNL1* nuclear foci (Figures 5E,F and S20). A small but statistically significant improvement in splicing was observed at 0.2 μM, indicating that **22-Bleo** is more potent than **22** (Figure 5E). Similar to the parent compound, **22-Bleo** did not affect *MBNL1* exon 5 splicing in wild-type myotubes or *MAP4K4* exon 22a splicing in DM1 myotubes, indicating specific effects (Figure S21).

Importantly, the cleavage mode of action of **22-Bleo** allows for direct profiling of potential off-targets. We assessed the levels of all transcripts containing short, nonpathological r(CUG) repeats upon treatment of DM1 myotubes with **22-**

Bleo. None of these genes was significantly affected, indicating that **22-Bleo** can specifically recognize and cleave the disease-causing repeat expansion (Figure 5H). We have previously shown that these RNAs containing shorter r(CUG) repeats do not fold into a structure containing repeating 1 × 1 U/U internal loops, the source of the observed selectivity.³⁶

In conclusion, the OBOC library methodology provides a facile means to optimize the linker domain of a dimeric compound targeting r(CUG)^{exp}. A simple affinity-based selection strategy enabled the screening of >330,000 compounds and subsequent hit identification via MS–MS sequencing. Through subsequent analysis of the bioactivity of hit compounds, we identified compound **22** which bound r(CUG) repeats 10 times more avidly than **2H-4** by forming additional interactions between the target and optimized linker and more potently rescued disease-associated defects in DM1 patient-derived myotubes. A derivative of **22** attached to the natural product bleomycin A5 selectively cleaved r(CUG)^{exp} in cells and improved DM1 defects at concentrations lower than that of the parent binder. Thus, OBOC library synthesis and screening can be used to identify high-affinity binders to r(CUG)^{exp}. Importantly, the methodology developed herein is likely to be general and applicable to numerous other RNA targets to aid in the identification of high-affinity small molecules.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsmchemlett.1c00027>.

Supplemental tables and figures and methods (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Matthew D. Disney – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States;
✉ orcid.org/0000-0001-8486-1796; Email: disney@scripps.edu

Authors

Simon Vezina-Dawod – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
Alicia J. Angelbello – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
Shruti Choudhary – Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States
Kye Won Wang – Department of Chemistry, Florida Atlantic University, Jupiter, Florida 33458, United States
Ilyas Yildirim – Department of Chemistry, Florida Atlantic University, Jupiter, Florida 33458, United States;
✉ orcid.org/0000-0001-8357-1922

Complete contact information is available at:
<https://pubs.acs.org/doi/10.1021/acsmchemlett.1c00027>

Author Contributions

M.D.D. directed the study. S.V.-D., A.J.A., S.C., and K.W.Y. carried out the experiments. I.Y. oversaw the computational experiments. S.V.-D. and A.J.A. contributed equally.

Funding

This work was funded by this work including the National Institutes of Health (R35-NS116846 to M.D.D. and F31-

NS110269 to A.J.A.), the Department of Defense Peer-Reviewed Medical Research Program (W81XWH-18-0718 to M.D.D.), Myotonic U.S. Fellowship Research grant (to S.C.), and the Fonds de Recherche du Québec, Nature et Technologies (B3X scholarship to S.V.D.).

Notes

The authors declare the following competing financial interest(s): M.D.D. is a founder of Expansion Therapeutics. A.J.A. is a current employee of Expansion Therapeutics.

ACKNOWLEDGMENTS

We thank Prof. Denis Furling [Centre de Recherche en Myologie (UPMC/Inserm/CNRS), Institut de Myologie] for his generous gift of cell lines used in this paper.

ABBREVIATIONS

CNBr, cyanogen bromide; DM1, myotonic dystrophy type 1; DMPK, dystrophin myotonia protein kinase; EC₅₀, concentration of compound that affords half-maximal response; MAP4K4, mitogen-activated protein kinase kinase kinase 4; MBNL1, muscleblind-like 1; miR, microRNA; MS, mass spectrometry; OBOC, one-bead-one compound; UTR, untranslated region

REFERENCES

- (1) Wan, Y.; Qu, K.; Zhang, Q. C.; Flynn, R. A.; Manor, O.; Ouyang, Z.; Zhang, J.; Spitale, R. C.; Snyder, M. P.; Segal, E.; Chang, H. Y. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **2014**, *505* (7485), 706–709.
- (2) Lu, Z.; Chang, H. Y. Decoding the RNA structurome. *Curr. Opin. Struct. Biol.* **2016**, *36*, 142–148.
- (3) Andrews, R. J.; Baber, L.; Moss, W. N. RNAstructuromeDB: A genome-wide database for RNA structural inference. *Sci. Rep.* **2017**, *7* (1), 17269.
- (4) Disney, M. D. Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J. Am. Chem. Soc.* **2019**, *141* (17), 6776–6790.
- (5) Donlic, A.; Hargrove, A. E. Targeting RNA in mammalian systems with small molecules. *Wiley Interdiscip. Rev.: RNA* **2018**, *9* (4), No. e1477.
- (6) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem. Biol.* **2016**, *23* (9), 1077–1090.
- (7) Childs-Disney, J. L.; Wu, M.; Pushechnikov, A.; Aminova, O.; Disney, M. D. A small molecule microarray platform to select RNA internal loop-ligand interactions. *ACS Chem. Biol.* **2007**, *2* (11), 745–754.
- (8) Ursu, A.; Vézina-Dawod, S.; Disney, M. D. Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discovery Today* **2019**, *24* (10), 2002–2016.
- (9) Haniff, H. S.; Knerr, L.; Chen, J. L.; Disney, M. D.; Lightfoot, H. L. Target-directed approaches for screening small molecules against RNA targets. *SLAS Discovery* **2020**, *25* (8), 869–894.
- (10) Costales, M. G.; Aikawa, H.; Li, Y.; Childs-Disney, J. L.; Abegg, D.; Hoch, D. G.; Pradeep Velagapudi, S.; Nakai, Y.; Khan, T.; Wang, K. W.; Yildirim, I.; Adibekian, A.; Wang, E. T.; Disney, M. D. Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (5), 2406–2411.
- (11) Velagapudi, S. P.; Costales, M. G.; Vummidi, B. R.; Nakai, Y.; Angelbello, A. J.; Tran, T.; Haniff, H. S.; Matsumoto, Y.; Wang, Z. F.; Chatterjee, A. K.; Childs-Disney, J. L.; Disney, M. D. Approved anti-cancer drugs target oncogenic non-coding RNAs. *Cell Chem. Biol.* **2018**, *25* (9), 1086–1094.
- (12) Groen, E. J. N.; Talbot, K.; Gillingwater, T. H. Advances in therapy for spinal muscular atrophy: promises and challenges. *Nat. Rev. Neurol.* **2018**, *14* (4), 214–224.
- (13) Kodadek, T.; Paciaroni, N. G.; Balzarini, M.; Dickson, P. Beyond protein binding: recent advances in screening DNA-encoded libraries. *Chem. Commun.* **2019**, *55* (89), 13330–13341.
- (14) Madsen, D.; Azevedo, C.; Micco, I.; Petersen, L. K.; Hansen, N. J. V. An overview of DNA-encoded libraries: A versatile tool for drug discovery. In *Progress in Medicinal Chemistry*; Witty, D. R., Cox, B., Eds.; Elsevier, 2020; Vol. 59, Chapter 4, pp 181–249.
- (15) Litovchick, A.; Tian, X.; Monteiro, M. I.; Kennedy, K. M.; Guie, M.-A.; Centrella, P.; Zhang, Y.; Clark, M. A.; Keefe, A. D. Novel nucleic acid binding small molecules discovered using DNA-encoded chemistry. *Molecules* **2019**, *24* (10), 2026.
- (16) Ottl, J.; Leder, L.; Schaefer, J. V.; Dumelin, C. E. Encoded library technologies as integrated lead finding platforms for drug discovery. *Molecules* **2019**, *24* (8), 1629.
- (17) Komnatny, V. V.; Nielsen, T. E.; Qvortrup, K. Bead-based screening in chemical biology and drug discovery. *Chem. Commun.* **2018**, *54* (50), 6759–6771.
- (18) Liu, T.; Qian, Z.; Xiao, Q.; Pei, D. High-throughput screening of one-bead-one-compound libraries: identification of cyclic peptidyl inhibitors against Calcineurin/NFAT interaction. *ACS Comb. Sci.* **2011**, *13* (5), 537–546.
- (19) Sohrabi, C.; Foster, A.; Tavassoli, A. Methods for generating and screening libraries of genetically encoded cyclic peptides in drug discovery. *Nat. Rev. Chem.* **2020**, *4* (2), 90–101.
- (20) Simonetti, L.; Ivarsson, Y. Genetically encoded cyclic peptide phage display libraries. *ACS Cent. Sci.* **2020**, *6* (3), 336–338.
- (21) Quartararo, A. J.; Gates, Z. P.; Somsen, B. A.; Hartrampf, N.; Ye, X.; Shimada, A.; Kajihara, Y.; Ottmann, C.; Pentelute, B. L. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. *Nat. Commun.* **2020**, *11* (1), 3183.
- (22) Annis, D. A.; Nickbarg, E.; Yang, X.; Ziebell, M. R.; Whitehurst, C. E. Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Curr. Opin. Chem. Biol.* **2007**, *11* (5), 518–526.
- (23) Shi, Y.; Parag, S.; Patel, R.; Lui, A.; Murr, M.; Cai, J.; Patel, N. A. Stabilization of lncRNA GAS5 by a small molecule and its implications in diabetic adipocytes. *Cell Chem. Biol.* **2019**, *26* (3), 319–330.
- (24) Brook, J. D.; McCurrach, M. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J.-P.; Hudson, T.; Sohn, R.; Zemelman, B.; Snell, R. G.; Rundle, S. A.; Crow, S.; Davies, J.; Shelbourne, P.; Buxton, J.; Jones, C.; Juvonen, V.; Johnson, K.; Harper, P. S.; Shaw, D. J.; Housman, D. E. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **1992**, *68* (4), 799–808.
- (25) Mathieu, J.; Prévost, C. Epidemiological surveillance of myotonic dystrophy type 1: A 25-year population-based study. *Neuromuscular Disord.* **2012**, *22* (11), 974–979.
- (26) Machuca-Tzili, L.; Brook, D.; Hilton-Jones, D. Clinical and molecular aspects of the myotonic dystrophies: A review. *Muscle Nerve* **2005**, *32* (1), 1–18.
- (27) Taneja, K. L.; McCurrach, M.; Schalling, M.; Housman, D.; Singer, R. H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* **1995**, *128* (6), 995–1002.
- (28) Jiang, H.; Mankodi, A.; Swanson, M. S.; Moxley, R. T.; Thornton, C. A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **2004**, *13* (24), 3079–3088.
- (29) Li, J.; Nakamori, M.; Matsumoto, J.; Murata, A.; Dohno, C.; Kiliszek, A.; Taylor, K.; Sobczak, K.; Nakatani, K. A dimeric 2,9-diamino-1,10-phenanthroline derivative improves alternative splicing in myotonic dystrophy type 1 cell and mouse models. *Chem. Eur. J.* **2018**, *24* (68), 18115–18122.

- (30) Rzuczek, S. G.; Colgan, L. A.; Nakai, Y.; Cameron, M. D.; Furling, D.; Yasuda, R.; Disney, M. D. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **2017**, *13*, 188.
- (31) Jenquin, J. R.; Coonrod, L. A.; Silverglate, Q. A.; Pellitier, N. A.; Hale, M. A.; Xia, G.; Nakamori, M.; Berglund, J. A. Furamidine rescues myotonic dystrophy type I associated mis-splicing through multiple mechanisms. *ACS Chem. Biol.* **2018**, *13* (9), 2708–2718.
- (32) Rzuczek, S. G.; Southern, M. R.; Disney, M. D. Studying a drug-like, RNA-focused small molecule library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. *ACS Chem. Biol.* **2015**, *10* (12), 2706–2715.
- (33) Warf, M. B.; Nakamori, M.; Matthys, C. M.; Thornton, C. A.; Berglund, J. A. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (44), 18551–18556.
- (34) Hagler, L. D.; Luu, L. M.; Tonelli, M.; Lee, J.; Hayes, S. M.; Bonson, S. E.; Vergara, J. I.; Butcher, S. E.; Zimmerman, S. C. Expanded DNA and RNA trinucleotide repeats in myotonic dystrophy type 1 select their own multitarget, sequence-selective inhibitors. *Biochemistry* **2020**, *59* (37), 3463–3472.
- (35) Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. 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.* **2008**, *130* (48), 16254–16261.
- (36) Angelbello, A. J.; Rzuczek, S. G.; McKee, K. K.; Chen, J. L.; Olafson, H.; Cameron, M. D.; Moss, W. N.; Wang, E. T.; Disney, M. D. Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (16), 7799–7804.
- (37) Pushechnikov, A.; Lee, M. M.; Childs-Disney, J. L.; Sobczak, K.; French, J. M.; Thornton, C. A.; Disney, M. D. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J. Am. Chem. Soc.* **2009**, *131* (28), 9767–9779.
- (38) Rzuczek, S. G.; Gao, Y.; Tang, Z.-Z.; Thornton, C. A.; Kodadek, T.; Disney, M. D. Features of modularly assembled compounds that impart bioactivity against an RNA target. *ACS Chem. Biol.* **2013**, *8* (10), 2312–2321.
- (39) Benhamou, R. I.; Abe, M.; Choudhary, S.; Meyer, S. M.; Angelbello, A. J.; Disney, M. D. Optimization of the linker domain in a dimeric compound that degrades an r(CUG) repeat expansion in cells. *J. Med. Chem.* **2020**, *63* (14), 7827–7839.
- (40) Benhamou, R. I.; Vezina-Dawod, S.; Choudhary, S.; Won Wang, K.; Meyer, S. M.; Yildirim, I.; Disney, M. D. Macrocyclization of a ligand targeting a toxic RNA dramatically improves potency. *ChemBioChem* **2020**, *21*, 3229–3233.
- (41) Fowler, S. A.; Blackwell, H. E. Structure-function relationships in peptoids: Recent advances toward deciphering the structural requirements for biological function. *Org. Biomol. Chem.* **2009**, *7* (8), 1508–1524.
- (42) Miller, S. M.; Simon, R. J.; Ng, S.; Zuckermann, R. N.; Kerr, J. M.; Moos, W. H. Proteolytic studies of homologous peptide and N-substituted glycine peptoid oligomers. *Bioorg. Med. Chem. Lett.* **1994**, *4* (22), 2657–2662.
- (43) Gao, Y.; Amar, S.; Pahwa, S.; Fields, G.; Kodadek, T. Rapid Lead discovery through iterative screening of one bead one compound libraries. *ACS Comb. Sci.* **2015**, *17* (1), 49–59.
- (44) Chen, C. Z.; Sobczak, K.; Hoskins, J.; Southall, N.; Marugan, J. J.; Zheng, W.; Thornton, C. A.; Austin, C. P. Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* **2012**, *402* (5), 1889–98.
- (45) Lee, M. M.; French, J. M.; Disney, M. D. Influencing uptake and localization of aminoglycoside-functionalized peptoids. *Mol. BioSyst.* **2011**, *7* (8), 2441–51.
- (46) Gates, D. P.; Coonrod, L. A.; Berglund, J. A. Autoregulated splicing of muscleblind-like 1 (MBNL1) pre-mRNA. *J. Biol. Chem.* **2011**, *286* (39), 34224–33.
- (47) Ule, J.; Ule, A.; Spencer, J.; Williams, A.; Hu, J. S.; Cline, M.; Wang, H.; Clark, T.; Fraser, C.; Ruggiu, M.; Zeeberg, B. R.; Kane, D.; Weinstein, J. N.; Blume, J.; Darnell, R. B. Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **2005**, *37* (8), 844–52.
- (48) Benhamou, R. I.; Angelbello, A. J.; Andrews, R. J.; Wang, E. T.; Moss, W. N.; Disney, M. D. Structure-specific cleavage of an RNA repeat expansion with a dimeric small molecule is advantageous over sequence-specific recognition by an oligonucleotide. *ACS Chem. Biol.* **2020**, *15* (2), 485–493.