

RNA Targeting

Covalent Small-Molecule–RNA Complex Formation Enables Cellular Profiling of Small-Molecule–RNA Interactions**

Lirui Guan and Matthew D. Disney*

Covalent adduct formation between small molecules and DNA has been an effective therapeutic strategy for cancer.^[1] However, analogous approaches for disease-associated RNAs have been only sparsely reported.^[2] Herein, we demonstrate a programmable and potentially general approach to react small molecules with RNA in cells. Specifically, a designed small molecule that binds r(CUG)^{exp} (the causative agent of myotonic dystrophy type 1 (DM1)) and improves DM1-associated defects in cell culture models, was appended with an electrophilic, nucleic acid-reactive module. Using this compound, we sought to answer the following: 1) is compound potency affected by covalent adduct formation; 2) does adduct formation affect the lifetime of damaged (i.e., alkylated) mRNAs; and 3) can covalent adduct formation enable profiling of the cellular RNA targets of small molecules? We determined that: 1) covalent binding engenders a circa 2500-fold improvement in potency; 2) damaged mRNAs are not degraded more rapidly than undamaged ones; and 3) RNA–small molecule interactions can be isolated and identified in a cell-wide manner, unequivocally demonstrating that designer small molecules target r(CUG)^{exp}. These studies provide a potentially transformative foundation to manipulate and study RNA cellular function with designer small molecules.

RNA is an interesting and increasingly important drug target owing to its essential functions and association with various diseases. However, there are relatively few small molecules that target RNAs in living cells and affect function.^[3] The bacterial ribosome is the most well-studied target of small molecules, which have served as therapeutics and probes of ribosomal function.^[4] Compounds targeting other RNAs are needed to enable similar studies yet few have been reported.

In an effort to develop generalizable methods to design small molecules that affect RNA function, we initiated a program to identify the RNA motifs, or folds, that are privileged for binding small molecules and chemotypes that are privileged for binding RNA. These interactions comprise

a database of RNA motif (fold)–small molecule interactions that can inform the design of compounds that bind a target of interest.^[5] The database is compared to the secondary structure of an RNA target, and the overlap between them establishes a lead targeting modality. This approach has been validated with various disease-causing RNAs including those that cause myotonic dystrophy types 1 and 2 (DM1 and DM2, respectively).^[6] DM1 and DM2 are caused by non-coding, repeating transcripts that fold into hairpins that bind proteins, including muscleblind-like 1 protein (MBNL1), a pre-mRNA splicing regulator (Figure 1A). Sequestration of MBNL1 by the repeating transcripts (RNA gain-of-function) causes its

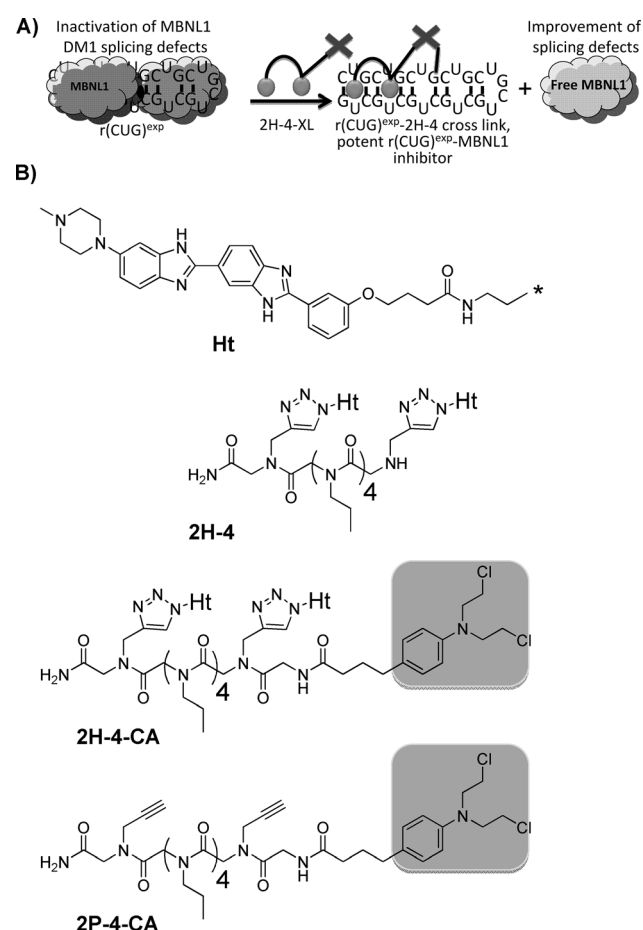


Figure 1. The r(CUG)^{exp}-MBNL1 interaction that causes DM1 and the structures of designed compounds. A) the r(CUG)^{exp}-MBNL1 complex that causes DM1. Modularly assembled small molecules targeting r(CUG)^{exp} were appended with a reactive group (XL) to cross link them to the toxic transcript. B) Structures of the compounds designed and tested in this study. The reactive module, a derivative of chlorambucil, is highlighted with gray boxes.

[*] Dr. L. Guan, Prof. Dr. M. D. Disney
Department of Chemistry
The Scripps Research Institute, Scripps Florida
130 Scripps Way, 3A1, Jupiter, FL 33458 (USA)
E-mail: Disney@scripps.edu

[**] This work was funded by the National Institutes of Health (RO1-GM079235) and The Scripps Research Institute. We thank S. G. Rzuczek for the synthesis of Ht azide used to procure 2H-4 derivatives. M.D.D. is a Camille & Henry Dreyfus Teacher–Scholar.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201301639>.

inactivation and hence dysregulation of alternative pre-mRNA splicing.^[7]

Previous studies have shown that reactive small molecules target RNA in vivo (*Escherichia coli* and *Saccharomyces cerevisiae* ribosomes)^[2a,b] and in vitro (HIV rev-responsive element (RRE) RNA).^[2c] The work described herein utilizes those studies as a launching point for a programmable approach to design small molecules that react with specific transcripts in living cells. That is, information about RNA motif–small-molecule interactions informs design of specific small molecules for a target of interest; the small molecules are then coupled with a reactive module that forms a covalent adduct with the transcript in vivo.

We previously designed a compound that improves DM1-associated defects in cellular models of disease using our RNA motif–small-molecule database.^[6b] The compound, 2H-4, is comprised of a peptoid scaffold that displays two copies of the RNA-binding module Ht (Figure 1 B). Each Ht module binds to one 5' CUG/3' GUC motif displayed in r(CUG)^{exp}^[6d] and affords modest potency in vivo.^[6b] 2H-4 was engendered with nucleic acid reactivity by functionalization with chlorambucil (CA),^[8] affording 2H-4-CA (Figure 1 B). Thus, the cellular RNAs that react with CA are dictated by the nature of the RNA-binding modules, which bring the reactive group into close proximity and facilitate adduct formation.

The in vitro potency of 2H-4-CA was compared to the parent compound, 2H-4, using a previously reported assay that measures r(CUG)₁₀–MBNL1 complex formation.^[6c,9] As summarized in Table 1, 2H-4-CA is about 14-fold more potent

Table 1: Potency of 2H-4 and related compounds for inhibition of r(CUG)₁₀–MBNL1 complex formation.^[a]

Compound	IC ₅₀ [μM]
2H-4	71 ± 1
2H-4-CA	5 ± 1
2P-4-CA	47 ± 2
Chlorambucil	> 150

[a] Experiments were completed by using a previously described time-resolved FRET assay^[6c,9] with minor modifications.

than 2H-4. Importantly, 2P-4-CA, which does not contain RNA-binding modules (Figure 1 B), is a weak inhibitor of the r(CUG)₁₀–MBNL1 complex as is chlorambucil itself (Table 1). Gel electrophoresis experiments confirm that 2H-4-CA reacts with r(CUG)₁₀ in vitro (Supporting Information, Figure S3). The products of the reaction of 2H-4-CA with r(CUG)₁₀ were digested with nuclease P1 and analyzed by MALDI MS. Analysis indicates that an intrastrand cross-linked adduct is formed between 2H-4-CA and r(GC) dinucleotides (Supporting Information, Figure S4), which is likely due to alkylation of guanine N7 and cytosine N3.^[10] Furthermore, RNase T1 digestion of r(CUG)₁₀ after incubation with 2H-4-CA confirms that adduct formation occurs at guanosine residues (Supporting Information, Figure S5). Protection from T1 cleavage is observed with as little as 1 μM compound; complete protection is observed with 50 μM 2H-4-CA (Supporting Information, Figure S5). 2H-4-CA was also incubated with MBNL1 and the reaction products

analyzed by Western blotting using anti-MBNL1^[11] and anti-biotin antibodies (Supporting Information, Figure S6). These experiments reveal that 2H-4-CA does not react with MBNL1, suggesting that the function of MBNL1 is unaffected and potency is solely derived from the reaction of 2H-4-CA with r(CUG)₁₀.

Owing to these promising results, 2H-4-CA was evaluated in a model cellular system for improving DM1-associated alternative pre-mRNA splicing defects. In particular, we monitored improvement of the dysregulation of *MBNL1* alternative splicing using a previously reported mini-gene (Figure 2 A).^[6c] The splicing of *MBNL1* pre-mRNA is con-

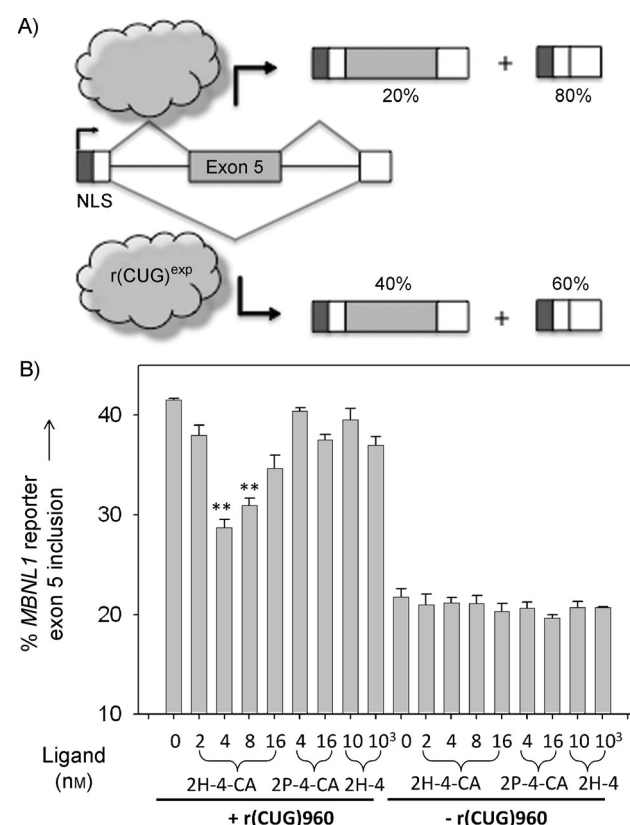


Figure 2. A designer small molecule that covalently cross-links to r(CUG)₉₆₀ is more potent than a related compound that only binds the target. A) Representation of the *MBNL1* mini-gene used to study DM1-associated pre-mRNA splicing defects that was previously described.^[6c] MBNL1 protein is indicated by the cloud. B) Data for the series of compounds studied showing that 2H-4-CA improves pre-mRNA splicing defects to a greater extent than 2H-4 and 2P-4-CA. “**” indicates a statistical significance of more than 99% confidence.

trolled by MBNL1 itself by binding and repression close to the 3' splice site of exon 5.^[12] In the absence of r(CUG)₉₆₀, *MBNL1* exon 5 has an inclusion rate of about 20%; expression of r(CUG)₉₆₀ induces exon inclusion, leading to an inclusion rate of about 40% (Figure 2 A). Thus, compound potency in vivo can be measured by the improvement of *MBNL1* splicing dysregulation. Ideally, a compound will completely restore splicing patterns to wild type, that is, splicing patterns in cells that express r(CUG)^{exp} are indistinguishable from those that do not.

As shown in Figure 2B, the IC_{50} of 2H-4-CA is 4 nM, where the IC_{50} is the concentration at which splicing patterns are restored to about 30% inclusion of exon 5 (Figure 2B; Supporting Information, Figure S7). The effect of 2H-4-CA plateaus at doses up to 8 nM and then diminishes at higher concentrations, perhaps reflecting non-specific interactions. It should be noted that 2P-4-CA shows some activity at 16 nM (Figure 2B); as this compound lacks the RNA-binding modules, it likely also interacts with other biomolecules that could give rise to non-specific effects. Neither the control compound 2P-4-CA nor 2H-4 has any effect on alternative pre-mRNA splicing when cells are dosed with 4 nM compound (Figure 2B; Supporting Information, Figure S7).^[13] In fact, 2H-4-CA is the most potent compound known to date that improves DM1-associated pre-mRNA splicing defects.^[14] Moreover, it is about 2500-fold more potent than the parent compound, 2H-4, which has an IC_{50} of circa 10 μ M (Figure 2B). Importantly, 2H-4-CA does not affect the alternative splicing of a transcript that is not regulated by MBNL1, illustrating selectivity for targeting r(CUG)₉₆₀ (Supporting Information, Figure S9). Therefore, our investigations show that 2H-4-CA improves pre-mRNA splicing defects caused by r(CUG)^{exp} and that compound potency in vivo is drastically improved by covalent adduct formation.

There is a significant difference in the in vitro IC_{50} value of 2H-4-CA for inhibiting r(CUG)₁₀-MBNL1 complex formation and the in vivo IC_{50} for improving splicing defects (Table 1 and Figure 2B). These observations could be due to: 1) differences in repeat length in vitro (r(CUG)₁₀) and in vivo (r(CUG)₉₆₀); 2) differences in incubation time and temperature (4 h/room temperature in vitro vs. 20–24 h/37°C in vivo). In vitro potency is improved by about 2.5-fold by increasing the incubation time and temperature (Supporting Information, Table S2); or 3) the two assays measure very different phenomena: in vitro assays measure inhibition of complex formation (all or none), while in vivo assays measure MBNL1 activity (degree of restoration). In support of the latter explanation, a recent report showed that the amount of active MBNL1 has a graded effect on the severity of splicing dysregulation.^[15] That is, only partial inhibition of the r(CUG)^{exp}-MBNL1 can cause significant improvement of splicing defects.

The cellular targets of 2H-4-CA were identified using a biotinylated derivative, 2H-4-CA-biotin (Figure 3A). There is no difference in the in vivo potencies of 2H-4-CA-biotin and 2H-4-CA (Supporting Information, Figure S10). In these studies, cells were treated with 4 nM 2H-4-CA-biotin, allowing formation of 2H-4-CA-biotin-biomolecule adducts in vivo. Adducts were isolated from the cells using TRIzol reagent and streptavidin beads. The beads were washed exhaustively with 1X PBST then water. 2H-4-CA-Biotin adducts were eluted by heating the beads at 65°C in 95% formamide. Gel electrophoresis and northern blotting of the eluted fraction showed that 2H-4-CA-biotin reacts with r(CUG)₉₆₀ (Figure 3B). Next, these pull-down experiments were completed with 4 nM 2H-4-CA and varying concentrations of 2H-4; 2H-4 should compete with 2H-4-CA for binding, reducing adduct formation between targets that are specifically bound by 2H-4. Indeed, 2H-4 inhibits adduct formation between 2H-4-CA

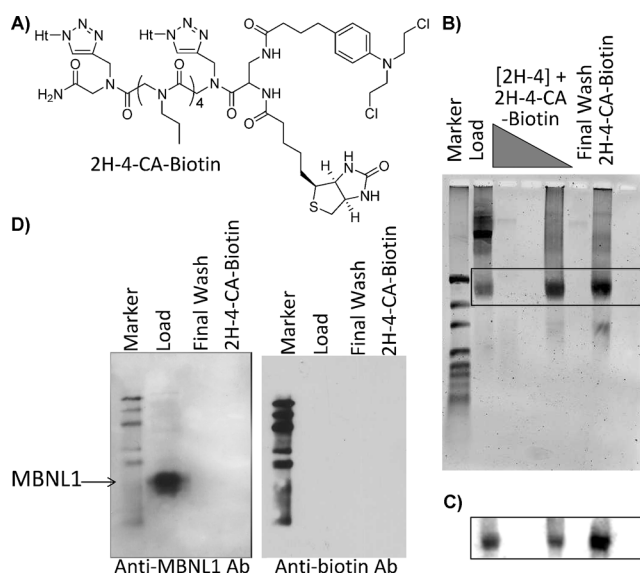


Figure 3. Structure of 2H-4-CA-biotin and identification of its cellular targets. A) structure of 2H-4-CA-biotin. B) Gel electrophoresis of nucleic acids captured by 2H-4-CA-biotin (4 nM) in cells in the presence and absence of competing 2H-4 (10 nM, 1 μ M, or 10 μ M). The gel was imaged after staining with SYBR Gold. C) Northern blot analysis of the gel in (B) shows that the major target of 2H-4-CA-biotin is r(CUG)₉₆₀. D) Western blot analysis to determine if MBNL1^[11] or other proteins are captured by 2H-4-CA-biotin (4 nM) in cells. Neither MBNL1 nor other proteins could be detected.

and r(CUG)₉₆₀ at micromolar concentrations (Figure 3B,C), demonstrating that both compounds recognize r(CUG)₉₆₀ in vivo. These results were further investigated by using qRT-PCR (Supporting Information, Figure S11). The amounts of 18S rRNA and r(CUG)₉₆₀ pulled down from cells treated with 4 nM 2H-4-CA and 10 μ M 2H-4, the amount of r(CUG)^{exp} is reduced by about 20-fold while the amount of 18S rRNA is only reduced by about 2-fold. Such competition experiments can be used to control for targets non-selectively pulled down by reaction with alkylators. Importantly, western blotting shows 2H-4-CA-biotin does not react with MBNL1 (Figure 3D, left; probed with anti-MBNL1^[11]) or other proteins (Figure 3D, right; probed with anti-biotin) in cellular pull-down experiments.

To investigate the source of the improved potency of 2H-4-CA, we studied whether adduct formation induces degradation of r(CUG)^{exp}, as its degradation would also lead to improvement of DM1-associated defects.^[16] Real-time RT-PCR experiments of cellular RNA harvested after reaction with 2H-4-CA show that r(CUG)^{exp}-mRNAs containing covalent adducts are not degraded more than those that do not (Supporting Information, Figure S12). These studies suggest that covalent adduct formation alone is responsible for the circa 2500-fold enhancement in the activity of 2H-4-CA, not reaction-induced mRNA degradation.

In summary, we have demonstrated that the potency of designer small molecules that solely bind RNA targets can be improved by engineering them with the ability to react with their cellular targets. Moreover, adduct formation provides a potentially general method to identify the cellular targets of

RNA-directed small molecules in living cells and, perhaps, animal models of disease. Transcriptome-wide probing could identify bystander (unintended) targets, and such information could be used to design and identify compounds with improved selectivity akin to activity-based profiling approaches used for proteins.^[17] This approach could also have uses in developing experimental tools to identify RNAs that bind to and are regulated by metabolites, or riboswitches, in cells.

Received: February 25, 2013

Revised: May 2, 2013

Published online: August 1, 2013

Keywords: bioorganic chemistry · chemical biology · nucleic acids · RNA · RNA recognition

- [1] S. E. Wolkenberg, D. L. Boger, *Chem. Rev.* **2002**, *102*, 2477–2495.
- [2] a) K. Rijal, C. S. Chow, *Chem. Commun.* **2009**, 107–109; b) A. A. Hostetter, M. F. Osborn, V. J. DeRose, *ACS Chem. Biol.* **2012**, *7*, 218–225; c) J. Boer, K. F. Blount, N. W. Luedtke, L. Elson-Schwab, Y. Tor, *Angew. Chem.* **2005**, *117*, 949–954; *Angew. Chem. Int. Ed.* **2005**, *44*, 927–932.
- [3] a) L. Guan, M. D. Disney, *ACS Chem. Biol.* **2012**, *7*, 73–86; b) J. R. Thomas, P. J. Hergenrother, *Chem. Rev.* **2008**, *108*, 1171–1224.
- [4] J. Poehlsgaard, S. Douthwaite, *Nat. Rev. Microbiol.* **2005**, *3*, 870–881.
- [5] a) J. L. Childs-Disney, M. Wu, A. Pushechnikov, O. Aminova, M. D. Disney, *ACS Chem. Biol.* **2007**, *2*, 745–754; b) M. D. Disney, L. P. Labuda, D. J. Paul, S. G. Poplawski, A. Pushechnikov, T. Tran, S. P. Velagapudi, M. Wu, J. L. Childs-Disney, *J. Am. Chem. Soc.* **2008**, *130*, 11185–11194.
- [6] a) M. M. Lee, A. Pushechnikov, M. D. Disney, *ACS Chem. Biol.* **2009**, *4*, 345–355; b) J. L. Childs-Disney, J. Hoskins, S. G. Rzuczek, C. A. Thornton, M. D. Disney, *ACS Chem. Biol.* **2012**, *7*, 856–862; c) R. Parkesh, J. L. Childs-Disney, M. Nakamori, A. Kumar, E. Wang, T. Wang, J. Hoskins, T. Tran, D. E. Housman, C. A. Thornton, M. D. Disney, *J. Am. Chem. Soc.* **2012**, *134*, 4731–4742; d) A. Pushechnikov, M. M. Lee, J. L. Childs-Disney, K. Sobczak, J. M. French, C. A. Thornton, M. D. Disney, *J. Am. Chem. Soc.* **2009**, *131*, 9767–9779.
- [7] a) H. Jiang, A. Mankodi, M. S. Swanson, R. T. Moxley, C. A. Thornton, *Hum. Mol. Genet.* **2004**, *13*, 3079–3088; b) N. A. Faustino, T. A. Cooper, *Genes Dev.* **2003**, *17*, 419–437.
- [8] A. Gellhorn, G. A. Hyman, J. E. Ulmann, *J. Am. Med. Assoc.* **1956**, *162*, 178–183.
- [9] C. Z. Chen, K. Sobczak, J. Hoskins, N. Southall, J. J. Marugan, W. Zheng, C. A. Thornton, C. P. Austin, *Anal. Bioanal. Chem.* **2012**, *402*, 1889–1898.
- [10] a) D. Mohamed, S. Mowaka, J. Thomale, M. W. Linscheid, *Chem. Res. Toxicol.* **2009**, *22*, 1435–1446; b) S. Kallama, K. Hemminki, *Acta Pharmacol. Toxicol.* **1984**, *54*, 214–220; c) J. N. Lampe, I. V. Kutayavin, R. Rhinehart, M. W. Reed, R. B. Meyer, H. B. Gamper, Jr., *Nucleic Acids Res.* **1997**, *25*, 4123–4131.
- [11] I. Holt, S. Mittal, D. Furling, G. S. Butler-Browne, J. D. Brook, G. E. Morris, *Genes Cells* **2007**, *12*, 1035–1048.
- [12] D. P. Gates, L. A. Coonrod, J. A. Berglund, *J. Biol. Chem.* **2011**, *286*, 34224–34233.
- [13] 2H-4-Pt, a 2H-4 conjugate containing a cisplatin reactive module, was synthesized and tested for improving DM1-associated pre-mRNA splicing defects. The compound is inactive (Supporting Information, Figure S8).
- [14] a) L. O. Ofori, J. Hoskins, M. Nakamori, C. A. Thornton, B. L. Miller, *Nucleic Acids Res.* **2012**, *40*, 6380–6390; b) M. B. Warf, M. Nakamori, C. M. Matthys, C. A. Thornton, J. A. Berglund, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18551–18556.
- [15] S. P. Jog, S. Paul, W. Dansithong, S. Tring, L. Comai, S. Reddy, *PLoS One* **2012**, *7*, e48825.
- [16] L. Guan, M. D. Disney, *Angew. Chem.* **2013**, *125*, 1502–1505; *Angew. Chem. Int. Ed.* **2013**, *52*, 1462–1465.
- [17] B. F. Cravatt, A. T. Wright, J. W. Kozarich, *Annu. Rev. Biochem.* **2008**, *77*, 383–414.