

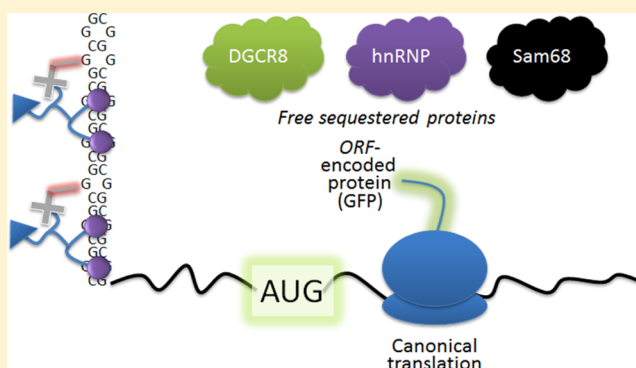
Inhibition of Non-ATG Translational Events in Cells via Covalent Small Molecules Targeting RNA

Wang-Yong Yang,[†] Henry D. Wilson,[†] Sai Pradeep Velagapudi,[†] and Matthew D. Disney^{*,†}

[†]Department of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way, Jupiter, Florida 33458, United States

S Supporting Information

ABSTRACT: One major class of disease-causing RNAs is expanded repeating transcripts. These RNAs cause diseases via multiple mechanisms, including: (i) gain-of-function, in which repeating RNAs bind and sequester proteins involved in RNA biogenesis and (ii) repeat associated non-ATG (RAN) translation, in which repeating transcripts are translated into toxic proteins without use of a canonical, AUG, start codon. Herein, we develop and study chemical probes that bind and react with an expanded r(CGG) repeat (r(CGG)^{exp}) present in a 5' untranslated region that causes fragile X-associated tremor/ataxia syndrome (FXTAS). Reactive compounds bind to r(CGG)^{exp} *in cellulo* as shown with Chem-CLIP-Map, an approach to map small molecule binding sites within RNAs in cells. Compounds also potentially improve FXTAS-associated pre-mRNA splicing and RAN translational defects, while not affecting translation of the downstream open reading frame. In contrast, oligonucleotides affect both RAN and canonical translation when they bind to r(CGG)^{exp}, which is mechanistically traced to a decrease in polysome loading. Thus, designer small molecules that react with RNA targets can be used to profile the RNAs to which they bind in cells, including identification of binding sites, and can modulate several aspects of RNA-mediated disease pathology in a manner that may be more beneficial than oligonucleotides.



INTRODUCTION

Although RNA targets in the transcriptome are numerous, there is a dearth of small molecule chemical probes that can be used to study RNA function and dysfunction. Despite great interest in this area, the development of such compounds is difficult¹ owing to a lack of fundamental information, or design principles, that could enable the development of compounds that selectively target an RNA in a cell.^{2,3} One approach to design compounds that affect function is to study RNA motif–small molecule interactions, thereby identifying small molecule “modules” that bind regions of an RNA of interest. Affinity and selectivity of the modules can be improved by linking them together to bind two or more regions in the desired RNA simultaneously. Indeed, such a bottom-up approach has been used to design compounds that target repeating transcripts and other RNAs.^{4,5}

The use of small molecules to modulate RNA function is of particular interest for studying various aspects of disease pathology. Repeating transcripts cause diseases via a gain-of-function mechanism or by translation into toxic proteins with or without the use of a start codon (Figure 1).⁶ A common defect caused by RNA repeat gain-of-function is dysregulation of alternative pre-mRNA splicing.⁷ For example, fragile X-associated tremor/ataxia syndrome (FXTAS) is caused by an expanded r(CGG) repeat (r(CGG)^{exp}) that binds and sequesters various proteins including DiGeorge syndrome

critical region 8 protein (DGCR8), Src-associated in mitosis 68 kDa (Sam68), and others.^{8,9} Sequestration of these proteins causes dysregulation of microRNA processing and alternative pre-mRNA splicing.^{8,9} As has been demonstrated for the RNAs that cause the myotonic dystrophies (DM), amyotrophic lateral sclerosis and frontal temporal dementia (ALS/FTD), and FXTAS, expanded repeating RNAs are also translated without the use of a start codon, or repeat associated non-ATG (RAN) translation.^{10–12} RAN translation produces toxic polymeric proteins that appear to contribute to disease.

We sought to determine if we could design small molecules that address both modes of toxicity using r(CGG)^{exp} as a model system (Figure 1). Ideally, the designed compound would improve alternative pre-mRNA splicing defects and inhibit RAN translation while having no effect on translation of the downstream open reading frame (ORF). It is particularly important that translation of this downstream ORF is not affected; r(CGG)^{exp} is located in the 5' untranslated region (UTR) of the fragile X mental retardation 1 (FMR1) gene, which encodes fragile X mental retardation protein (FMRP). FMRP is important for regulation of protein synthesis.¹³ In fragile X syndrome (FXS), which is also caused by a r(CGG) expansion, albeit larger than those that cause FXTAS, FMRP is

Received: July 22, 2014

Published: March 31, 2015

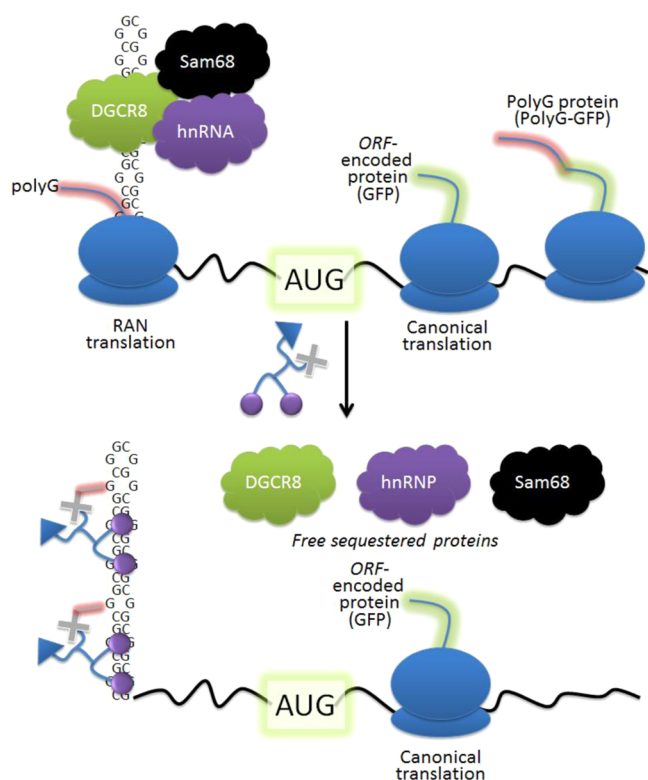


Figure 1. Modes of toxicity associated with $r(\text{CGG})^{\text{exp}}$, the causative agent of FXTAS. (Top) The repeating RNA folds into a hairpin structure that binds and sequesters proteins that regulate RNA processing. Additionally, repeating transcripts are translated without a start codon in a mechanism called RAN translation, producing toxic polymeric proteins that contribute to pathogenesis. (Bottom) Small molecules that bind to and react with $r(\text{CGG})^{\text{exp}}$ and free bound proteins improve defects in RNA processing and inhibit production of RAN, but not normal, translation products.

silenced; mice that do not produce FMRP have learning defects and hyperactivity.¹⁴ There are likely spatial aspects that govern whether RAN and/or canonical translation is inhibited, as has been previously shown in engineered systems.^{15,16} Thus, a focus of the present studies is to examine such aspects in systems in which translational events resemble those found in disease-associated mRNAs and how such events can be advantageously modified.

Previously, we designed the small molecule **2H-5** (Figure 2A) to target $r(\text{CGG})^{\text{exp}}$ by mining interactions in a RNA motif-small molecule database.¹⁷ The compound contains two RNA-binding modules (**H**) that are appropriately spaced to target two 5' $\text{CGG}/3' \text{GGC}$ motifs in $r(\text{CGG})^{\text{exp}}$ simultaneously (Figure 2A). Indeed, **2H-5** improves pre-mRNA splicing defects caused by sequestration of Sam68 at μM concentrations and does not affect translation of a model downstream ORF.¹⁷ In this study, we sought to improve **2H-5**'s potency by appending it to a reactive module akin to our previously reported studies of a designed small molecule that targets $r(\text{CUG})^{\text{exp}}$,³ the causative agent of DM type 1. To engender **2H-5** with the ability to react with cellular RNA targets, it was coupled to a nitrogen mustard (**CA** or **MA**, Figure 2A).³ Since such compounds prefer to react with the N-7 of G,¹⁸ we hypothesized that $r(\text{CGG})^{\text{exp}}$ might be an ideal RNA to which to extend this approach. A biotin tag was also incorporated onto **2H-5** to allow for the facile isolation of the RNA targets it

binds in cells (via Chemical Cross-Linking and Isolation by Pull-down, or Chem-CLIP),³ affording **2H-5-CA-Biotin** and **2H-5-MA-Biotin** (Figure 2A). We also synthesized **2P-5-CA-Biotin** and **2P-5-MA-Biotin**, control compounds that lack the RNA-binding modules, to study the effect on selective targeting of $r(\text{CGG})^{\text{exp}}$. (Please see the Supporting Information for details of chemical syntheses, in particular Schemes S-1–3 and Figure S-1.) Importantly, the reactive compounds bind the desired target in cells and potentially improve FXTAS-associated pre-mRNA splicing and RAN translational defects, while not affecting translation of the downstream ORF.

RESULTS AND DISCUSSION

Studying the Reaction of Designer Small Molecules with RNA *in Vitro*. Compounds **2H-5-CA-Biotin**, **2H-5-MA-Biotin**, **2P-5-CA-Biotin**, and **2P-5-MA-Biotin** were tested for selective reaction with $r(\text{CGG})_{12}$. The compounds contain two RNA-binding modules (**H**), a reactive module (**CA** or **MA**), and a purification tag (**Biotin**) (Figure 2A). Control compounds (**P**) without RNA-binding modules were also tested. The **CA** module on **2H-5-CA-Biotin** can react with two sites (typically an intrastrand cross-link)¹⁹ within a single target RNA because two *N*-chloroethyl groups are present. In contrast, **2H-5-MA-Biotin** only has one reactive *N*-chloroethyl group and is thus incapable of forming intrastrand cross-links.¹⁹ Comparative studies of these two compounds facilitate an understanding of how different reactivity modes can affect biological activity and target recognition.

Initial *in vitro* assays were completed by using ³²P-labeled $r(\text{CGG})_{12}$. In these experiments, the biotin tag installed on the compounds allowed capture of reacted RNAs with streptavidin-coated resin. This allowed us to determine the amount of RNA that reacted with a given small molecule via scintillation counting. These studies showed that the order of reaction was **2H-5-CA-Biotin** > **2H-5-MA-Biotin** > **2P-5-CA-Biotin** > **2P-5-MA-Biotin** (Figure 2B). Thus, addition of the RNA-binding modules enhances the ability of the compounds to react with RNA *in vitro*, and the type of reactive module used influences compound reactivity.

The products of the *in vitro* reaction between $r(\text{CGG})_{12}$ and **2H-5-CA-Biotin** or **2P-5-CA-Biotin** were further studied. After incubation with both **2H-5-CA-Biotin** and **2P-5-CA-Biotin**, $r(\text{CGG})_{12}$ was treated with P1 nuclease, and the reaction products were identified by mass spectral analysis. The observed mass of the reaction products corresponds to a **2H-5-CA-Biotin** adduct that reacted with two guanines (Figure 2C), consistent with the preferred reactivity of chlorambucil (N-7 of guanine).¹⁸ Importantly, no adduct formation between $r(\text{CGG})_{12}$ and **2P-5-CA-Biotin** was detected. Collectively, these studies show that **2H-5-CA-Biotin** reacts with $r(\text{CGG})$ repeats *in vitro* and that the RNA-binding modules (**H**) are required.

Studying the Cellular Activity of Designer Compounds: Inhibition of RAN Translation. Next, we studied if the above compounds inhibit translation events of a $r(\text{CGG})^{\text{exp}}$ -containing transcript, including both RAN and canonical translation. For these studies, we employed a cellular model system in which $r(\text{CGG})_{88}$ was embedded in the 5' UTR of green fluorescent protein (GFP), which was previously designed and validated by the Todd and Paulson groups.¹¹ The construct was developed such that the orientation of $r(\text{CGG})_{88}$ in the 5' UTR relative to the GFP ORF mimics that between $r(\text{CGG})^{\text{exp}}$ and the ORF in *FMR1*. Two translation products

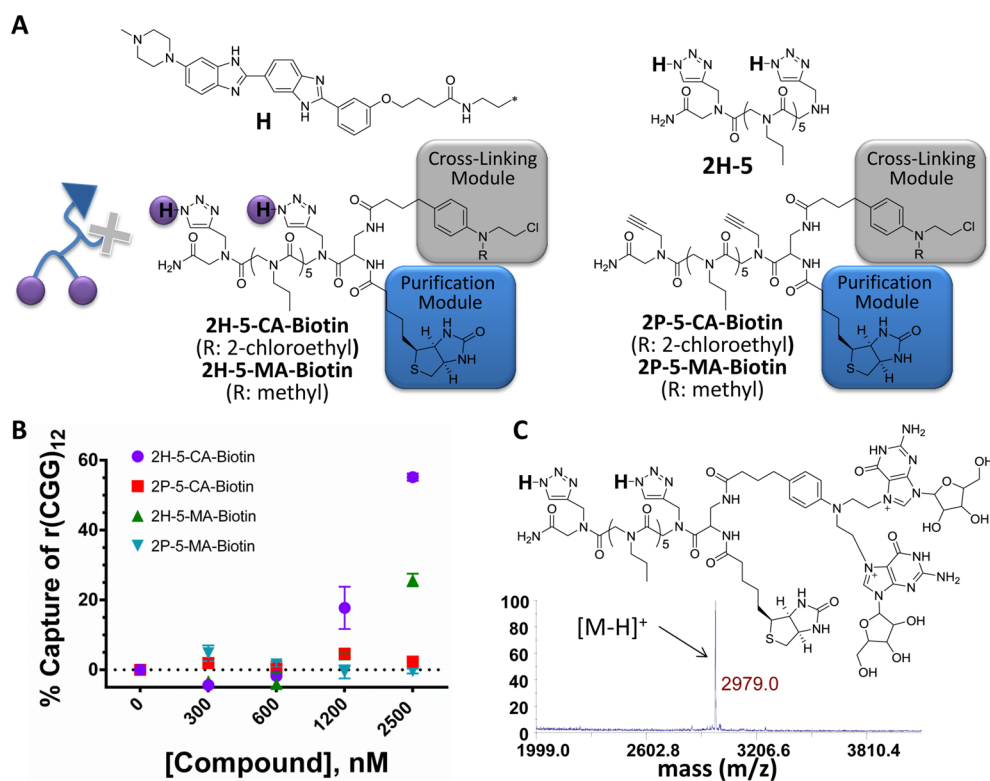


Figure 2. Structures of the compounds used in these studies. (A) **2H-5-CA-Biotin** is comprised of two RNA-binding modules (purple circles), a reactive module (gray box), and a biotin purification module (blue box) to improve potency and enable identification of cellular targets. **2H-5** noncovalently binds $r(\text{CGG})^{\text{exp}}$ and was described previously.¹⁷ **2P-5-CA-Biotin** and **2P-5-MA-Biotin** lack RNA binding modules. **2H-5-MA-Biotin** has a single reactive group. (B) Studying the reaction of designer small molecules with RNA *in vitro*. (C) MS analysis of the **2H-5-CA-Biotin**- $r(\text{CGG})_{12}$ adduct generated *in vitro*.

are observed in cells: the canonical translation product, GFP, and the RAN translation product, polyG-GFP. The two products can be resolved by SDS-PAGE and visualized by Western blot with an anti-GFP antibody as previously described.¹¹

Indeed, **2H-5-CA-Biotin** decreases the RAN translation product by ~40% at 500 nM but does not significantly affect translation of GFP initiated at the canonical translational start site (Figures 3A and S-2). Further, a dose response is observed. Although **2H-5-MA-Biotin** also inhibits RAN translation but not canonical translation, it is less potent (Figure 3A); at 1 μM , **2H-5-MA-Biotin** and **2H-5-CA-Biotin** inhibit 23% and 47% of RAN translation, respectively. Interestingly, the extent of inhibition of RAN translation correlates with *in vitro* reactivity toward ³²P-labeled $r(\text{CGG})_{12}$ for each compound (Figure 2B). We also employed a construct in which the ATG start codon was removed from the $r(\text{CGG})_{88}$ -GFP construct (Figure 4A). As expected, only the RAN translation product was observed, and its abundance was similar to the amount of RAN product observed for the plasmid containing an ATG start codon (Figure S-6). In this system, **2H-5-CA-Biotin** inhibits ~65% of RAN translation at 1 μM compound concentration (Figure 4B).

To further study the features of compounds that allows for inhibition of RAN translation, we studied the control compound that lacks the RNA-binding modules, **2P-5-CA-Biotin**, and the noncovalent binder **2H-5** (Figures 2B and 3A). Neither canonical nor RAN translation were inhibited when cells were treated with 50 μM **2H-5** or 1 μM **2P-5-CA**. Collectively, these studies show that both a reactive module and

RNA-binding modules are required for inhibition of RAN translation in this system (Figure 3A and 3B). We also tested the effect of **2H-5-CA-Biotin** on $r(\text{CGG})_{88}$ -GFP mRNA levels (Figure S-3) and translation of GFP when the 5' UTR lacks $r(\text{CGG})^{\text{exp}}$ (Figure S-4). No statistically significant changes were observed for protein expression or mRNA abundance. The compound exhibited no significant cytotoxicity via a MTT assay at the concentrations used in biological assays (Figure S-5). Collectively, these studies suggest that the effects are due to reaction at specific sites within the mRNA, likely $r(\text{CGG})^{\text{exp}}$ in the 5' UTR, *vide infra*. It also shows that it is possible to inhibit RAN translation of an expanded repeat embedded in a 5' UTR with a small molecule without significantly affecting translation of the downstream ORF.

We also studied if a 2'OMe phosphorothioate oligonucleotide (2'OMe-PS-(CCG)₁₂) that is complementary to $r(\text{CGG})^{\text{exp}}$ modulates RAN and canonical translational events. The oligonucleotide inhibits both RAN and canonical translation (Figure 3C), and selective inhibition of RAN translation could not be obtained despite using a 2-fold alteration in treatment concentrations. We also tested the effect of the oligonucleotide on $r(\text{CGG})_{88}$ -GFP mRNA levels; these data showed that there was no effect on the mRNA's abundance (Figure S-7). The oligonucleotide had no effect on translation of GFP if the 5' UTR lacks $r(\text{CGG})^{\text{exp}}$ (Figure S-8). It has been previously observed that the binding of oligonucleotides to the 5' UTR can affect translation.²⁰ It is likely that the oligonucleotide inhibits both RAN and canonical translation by either preventing ribosome loading or inhibiting ribosomal read-through, *vide infra*.

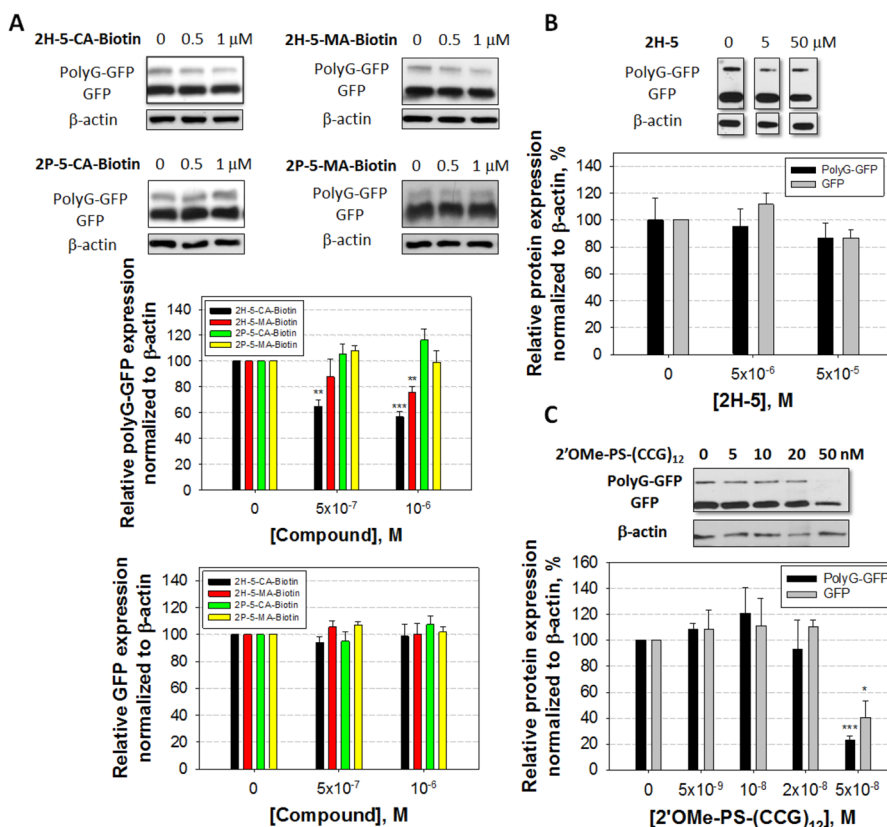


Figure 3. Effects of small molecules and an oligonucleotide on protein translation in cells. (A) Effect of 2H-5-CA-Biotin, 2H-5-MA-Biotin, 2P-5-CA-Biotin, and 2P-5-MA-Biotin on RAN translation of r(CGG)^{exp} and canonical translation of the downstream ORF that encodes GFP, as determined by Western blotting. (B) Effect of 2H-5 on RAN and canonical translation. (C) Effect of a 2'OMe phosphorothioate oligonucleotide, 2'OMe-PS-(CCG)₁₂, on RAN and canonical translation. *, **, ***, and **** denote $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$, respectively, as compared to the untreated sample with a two-tailed Student t test ($n \geq 3$ for compounds and $n = 3$ for 2'OMe-PS-(CCG)₁₂). Values are reported as mean \pm standard error.

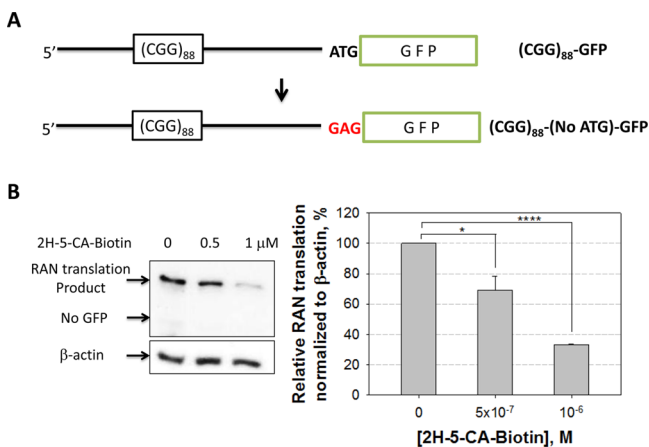


Figure 4. Modification of (CGG)₈₈-GFP to afford a construct without an ATG start codon and the effect of 2H-5-CA-Biotin on RAN translation in the absence of AUG. (A) Schematic of the (CGG)₈₈-(No ATG)-GFP construct. (B) Western blot image and quantification of relative amounts RAN translation products in the presence or absence of 2H-5-CA-Biotin when the AUG start codon is removed. * and **** denote $p < 0.05$ and $p < 0.0001$, respectively, as determined by a two-tailed Student t test ($n = 3$). Values are reported as mean \pm standard error.

Studying the *in Cellulo* Activity of Designer Compounds: Improvement of FXTAS-Associated Pre-mRNA Splicing Defects. Since 2H-5-CA-Biotin improves one

FXTAS-associated defect (RAN translation), we next determined if the compound improves alternative pre-mRNA splicing defects. The ability of 2H-5-CA-Biotin to improve FXTAS-associated pre-mRNA splicing defects was evaluated using a plasmid that encodes r(CGG)₆₀ and a mini-gene that reports on the alternative splicing of survival motor neuron 2 (SMN2) mRNA.²¹ It was previously shown that the alternative splicing of SMN2 exon 7 is dysregulated in FXTAS by sequestration of Sam68 by r(CGG)^{exp}.⁸ In the presence of r(CGG)₆₀, exon 7 is included in $\approx 60\%$ of the mature mRNA, while in the absence of r(CGG)₆₀, exon 7 is included in $\approx 20\%$ (Figure 5A).

Treatment of cells with 1 μM of 2H-5-CA-Biotin improved the SMN2 splicing defect by $\sim 40\%$, similarly to 2H-5-MA-Biotin (Figure 5B). Neither compound affected SMN2 splicing in cells that did not express r(CGG)₆₀ (Figure S-9). Further, no significant effect on SMN2 splicing was observed when cells were treated with 1 μM 2H-5 or 2P-5-CA-Biotin. The observed bioactivity of 2H-5-CA-Biotin was not due to changes in r(CGG)₆₀ abundance as there was no effect on transcription as determined by qRT-PCR (Figure S-10). In agreement with our RAN translation studies (Figure 3), these studies show that both covalent binding and the RNA-binding modules are required for the observed bioactivity (Figure 5B).

Chem-CLIP Evaluation of the Cellular Targets of the Designer Small Molecules. Compounds were evaluated for the RNA targets they reacted with in cells by using a reactive profiling approach named Chem-CLIP (Chemical Cross-

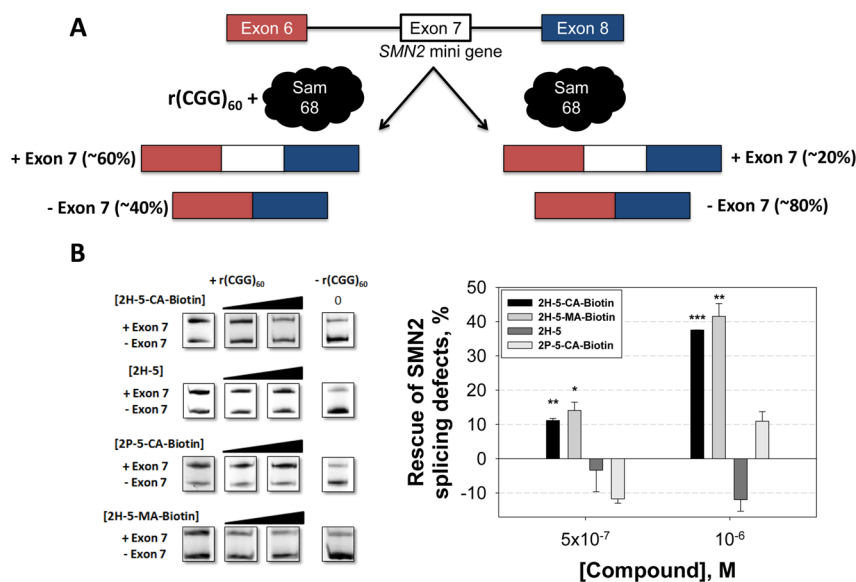


Figure 5. Improvement of FXTAS-associated pre-mRNA splicing defects by **2H-5-CA-Biotin**. (A) Schematic of the alternative pre-mRNA splicing of *SMN2* mini-gene in the presence and absence of $r(\text{CGG})^{\text{exp}}$. (B) Cellular efficacy of **2H-5-CA-Biotin**, **2H-5**, and **2P-5-CA-Biotin** against FXTAS as assessed by improvement in *SMN2* pre-mRNA splicing defects. ** and *** denote $p < 0.01$ and $p < 0.001$, respectively, as compared to the untreated sample by a two-tailed Student t test ($n = 3$). Values are reported as the mean \pm standard error.

Linking and Isolation by Pull-down),³ a small molecule approach similar to cross-linking and immunoprecipitation (CLIP).²² For these studies, we employed the model cellular model system described above for RAN translation.¹¹ Cells were treated with **2H-5-CA-Biotin**, **2H-5-MA-Biotin**, or **2P-5-CA-Biotin**, and cellular targets were isolated using streptavidin-coated resin. Bound targets were released from the resin and analyzed via qRT-PCR (Figure 6A).³ Relative to the starting total RNA, a 24-fold enrichment of $r(\text{CGG})_{88}$ -GFP was

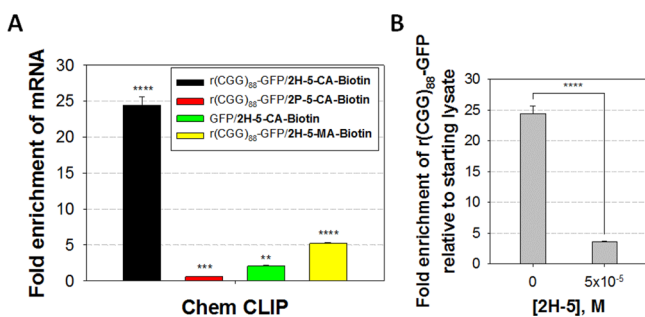


Figure 6. Identification of the cellular RNA targets of **2H-5-CA-Biotin**, **2H-5-MA-Biotin**, **2H-5**, and **2P-5-CA-Biotin**. (A) Results of Chem-CLIP pull-down of RNA targets in cells upon exposure to reactive compounds. An ~ 24 - and ~ 6 -fold enrichment of $r(\text{CGG})_{88}$ -GFP relative to starting lysate was observed in the pulled down fraction from cells treated with **2H-5-CA-Biotin** (500 nM) and **2H-5-MA-Biotin** (1 μM), respectively. No enrichment was observed when cells were treated with **2P-5-CA-Biotin**. Further, $r(\text{CGG})_{88}$ is required for pull down by **2H-5-CA-Biotin** as an ~ 13 -fold decreased enrichment of GFP mRNA was observed in cells that express a construct lacking $r(\text{CGG})^{\text{exp}}$ in the 5' UTR. (B) Results of competitive Chem-CLIP experiment. Cells were cotreated with **2H-5** and **2H-5-CA-Biotin**. The depletion of the $r(\text{CGG})_{88}$ -GFP target confirms that **2H-5** binds to $r(\text{CGG})_{88}$ -GFP. **, ***, and **** denote $p < 0.01$, $p < 0.001$, and $p < 0.0001$, respectively, (compared to the starting lysate for A) as determined by a two-tailed Student t test ($n = 3$). Values are reported as the mean \pm standard error.

observed in the pulled down fractions from cells treated with **2H-5-CA-Biotin**. Compound **2H-5-MA-Biotin** was also able to enrich $r(\text{CGG})_{88}$ -GFP by ~ 6 -fold. There was no enrichment in $r(\text{CGG})_{88}$ -GFP upon treatment with **2P-5-CA-Biotin**, consistent with our *in vitro* reactivity studies (Figure 2B). Only a minor enrichment of the GFP mRNA was observed when cells that express GFP lacking $r(\text{CGG})^{\text{exp}}$ were treated with **2H-5-CA-Biotin** (Figure 6A). Taken together, these studies show that **2H-5-CA-Biotin** reacts with the intended target *in cellulo*.

To assess the extent to which the noncovalent binder binds to $r(\text{CGG})_{88}$ -GFP *in cellulo*, we completed a Competitive Chem-CLIP (C-Chem-CLIP) experiment. In C-Chem-CLIP, cells are cotreated with **2H-5** and **2H-5-CA-Biotin**, and depletion of RNA targets captured by **2H-5-CA-Biotin** is assessed (indicates binding of **2H-5**). The results indicate that **2H-5** can compete with **2H-5-CA-Biotin** for binding to $r(\text{CGG})^{\text{exp}}$ in cells, confirming that both the covalent and the noncovalent compound engage the target (Figure 6B). Thus, noncovalent recognition of 5' $(\text{CGG})^{\text{exp}}$ in cells by **2H-5** is not sufficient to affect a translational event, whereas covalent reaction is. This observation may be due to the degree of thermodynamic stabilization of $r(\text{CGG})^{\text{exp}}$ upon compound binding; that is, a minimum amount of stabilization is required to prevent ribosomal read-through or loading. We therefore measured the thermodynamic stability imparted to $r(\text{CGG})_{12}$ by **2H-5-CA-Biotin** and **2H-5** by optical melting. These experiments showed that reaction of **2H-5-CA-Biotin** with $r(\text{CGG})_{12}$ as compared to the RNA alone increases the T_m by >10 $^{\circ}\text{C}$ (Figure S-11). In contrast, binding of **2H-5** has a negligible effect on T_m . Taken together, **2H-5-CA-Biotin** inhibits either: (i) ribosome loading onto $r(\text{CGG})^{\text{exp}}$; and/or (ii) ribosomal read-through by forming a steric or thermal block within the repeats that does affect ribosome loading onto the downstream ORF. That is, the ribosome can shunt through to the canonical AUG start codon. The latter mechanism is supported by studies in which oligonucleotides that target an ORF only inhibit translation when they are conjugated to

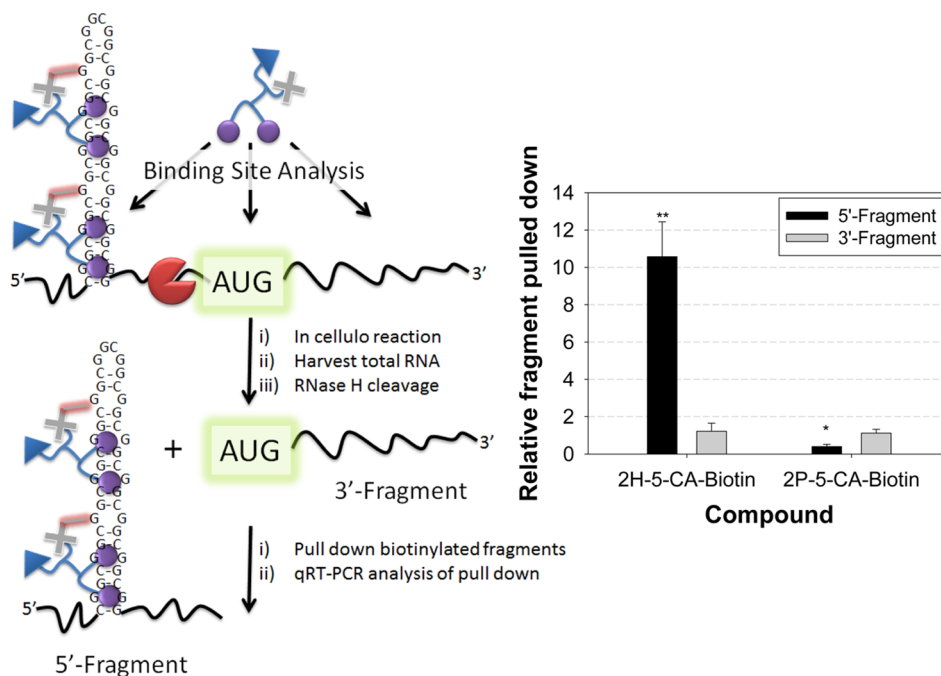


Figure 7. Binding site mapping for reactive compounds. (A) Schematic of Chem-CLIP-Map. (B) Results for **2H-5-CA-Biotin** and **2P-5-CA-Biotin**. The enrichment of the 5' fragment by **2H-5-CA-Biotin** indicates the small molecule binds to the 5' UTR in the mRNA. * and ** denote $p < 0.05$ and $p < 0.01$, respectively, as compared to samples before pull-down by a two-tailed Student t test ($n = 2$). Values are reported as the mean \pm standard error.

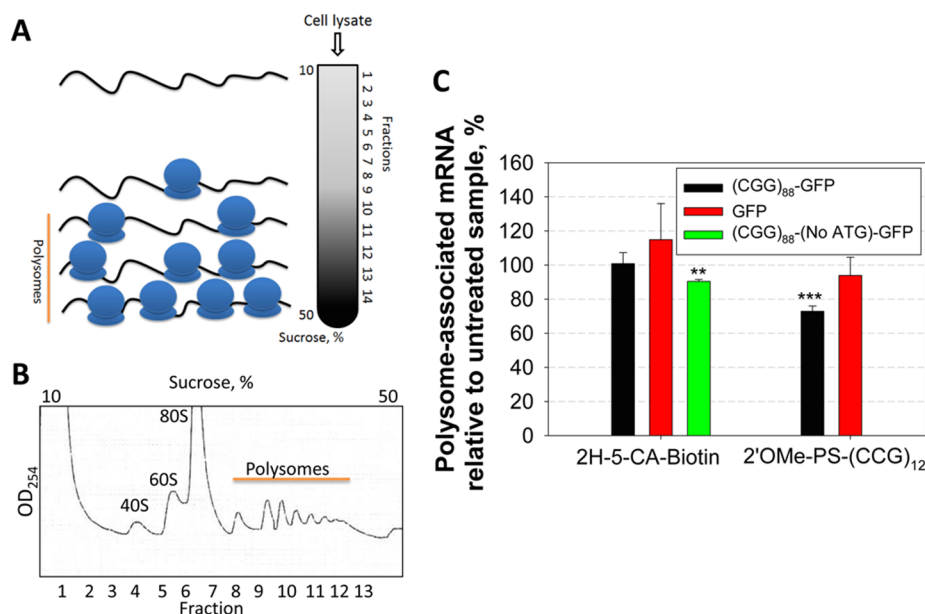


Figure 8. Polysome profiling of cells treated with **2H-5-CA-Biotin** and **2'OMe-PS-(CCG)₁₂**. (A) Schematic of polysome profiling. (B) Representative polysome profile. (C) Plot of the number of polysomes associated with r(CGG)₈₈-GFP upon treatment with **2H-5-CA-Biotin** and **2'OMe-PS-(CCG)₁₂**. While no change in the amount of r(CGG)₈₈-GFP and GFP (control) mRNA associated with translationally active polysomes was observed after **2H-5-CA-Biotin** treatment, the oligonucleotide decreased the number of associated polysomes by ~30%. **2H-5-CA-Biotin** diminished the number of polysomes associated with r(CGG)₈₈-(No ATG)-GFP, which lacks a start codon by ~10%. ** denotes $p < 0.01$ and *** denotes $p < 0.001$, as compared to the untreated sample with a two-tailed Student t test ($n = 2$).

psoralen and hence cross-linked with the mRNA; noncovalent binders had no effect on translation.²⁰ These studies are also consistent with **2H-5**'s inability to inhibit RAN translation despite binding to r(CGG)^{exp} *in cellulo*; binding does not enhance the RNA's thermodynamic stability sufficiently enough to impede ribosomal read-through.

Chem-CLIP-Map Enables Mapping of Small Molecule Binding Sites *in Cellulo*. Given that Chem-CLIP allowed for the targeted pull-down of the r(CGG)₈₈-GFP transcript, we sought to determine the site of reaction. This could be completed by primer extension, as has been described to identify sites of chemical modification by dimethyl sulfate (DMS) or kethoxol to map RNA secondary structure.²³ Such

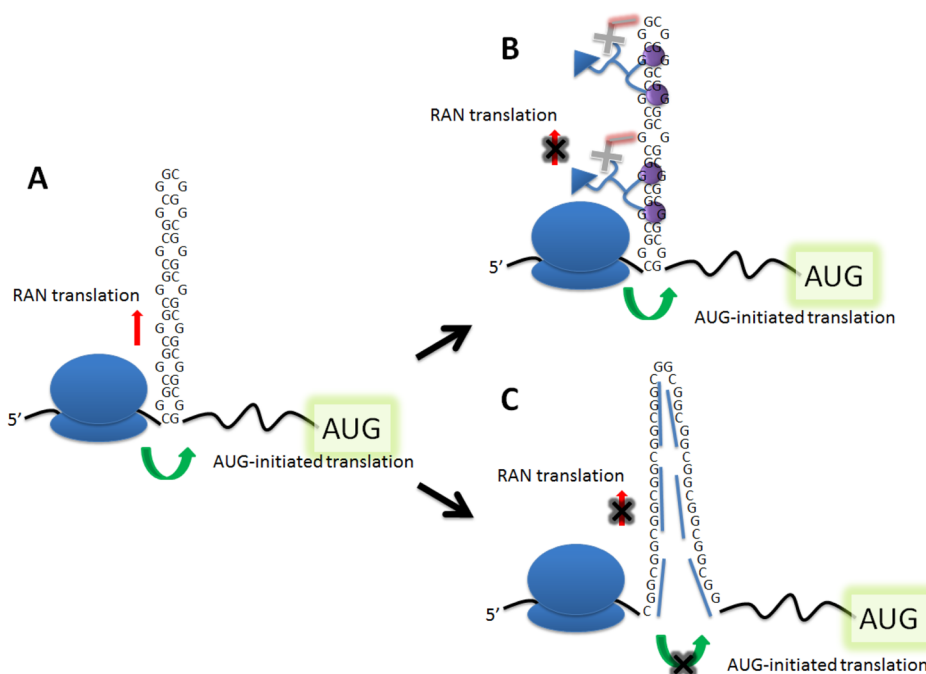


Figure 9. Possible mechanisms of inhibition of RAN translation in FXTAS by **2H-5-CA-Biotin** and **2'OMe-PS-(CCG)₁₂**. (A) Schematic of RAN and canonical translational events in FXTAS. (B) Schematic of a potential mode of action for the selective inhibition of RAN translation by **2H-5-CA-Biotin**. (C) Schematic of a potential mode of action for the inhibition of RAN and canonical translation events by **2'OMe-PS-(CCG)₁₂**.

an approach is intractable for mapping small molecule binding sites in $r(\text{CGG})^{\text{exp}}$, or other repeating transcripts, due to their highly structured, GC-rich nature that complicates read-through by reverse transcriptase. Thus, we mapped the ligand-binding site by using an alternative approach named Chem-CLIP-Map. In this approach, cells are treated with a reactive compound, and total RNA is harvested. The harvested RNA is then cleaved by treatment with an antisense oligonucleotide and RNase H (Figure S-12).²⁴ The cleaved RNA fragments are then captured onto streptavidin resin and quantified by qRT-PCR to identify the region(s) within the mRNA that reacted with compound (Figure 7). In this application, the antisense oligonucleotide binds between $r(\text{CGG})^{\text{exp}}$ and the AUG start codon. If the compound reacts with $r(\text{CGG})^{\text{exp}}$ as expected, then the 5' region of the RNA should be pull-down by streptavidin resin, while the remainder of the RNA (the ORF) should not.

As expected, **2H-5-CA-Biotin** reacted preferentially with the 5' end of the $r(\text{CGG})_{88}$ -GFP mRNA, where the repeat is present (Figure 7). No enrichment was observed for **2P-5-CA-Biotin**. Thus, the RNA-binding modules engender the compound with its reaction preference for the $r(\text{CGG})^{\text{exp}}$ -containing mRNA in cells. These studies support a mechanism for inhibition of FXTAS-associated defects in which the small molecule reacts with $r(\text{CGG})^{\text{exp}}$, displacing bound proteins or preventing their binding altogether (improvement of alternative pre-mRNA splicing defects) and simultaneously inhibiting ribosome recognition/loading of $r(\text{CGG})^{\text{exp}}$ (RAN translation) but not the downstream ORF.

Studying the Mechanism of Compound Inhibition of RAN Translation via Polysome Profiling. Polysome profiling is a powerful tool to study mRNAs bound by the ribosome and undergoing active translation. We therefore used polysome profiling to further investigate the modes of action of **2H-5-CA-Biotin** and **2'OMe-PS-(CCG)₁₂**. COS7 cells were

transfected with the $(\text{CGG})_{88}$ -GFP plasmid and treated with **2H-5-CA-Biotin** or **2'OMe-PS-(CCG)₁₂**. A sucrose gradient was then used to separate RNAs bound to single ribosomes and polysomes (Figure 8A and 8B). The distribution of $r(\text{CGG})_{88}$ -GFP in each fraction was then determined by qRT-PCR. Compared to the untreated control, no significant change in the distribution of $r(\text{CGG})_{88}$ -GFP was observed after treatment with 500 nM **2H-5-CA-Biotin** (Figures 8C and S-13). In contrast, treatment with **2'OMe-PS-(CCG)₁₂** decreased the amount $r(\text{CGG})_{88}$ -GFP mRNA associated with active polysomes by 27% (Figures 8C and S-13), consistent with its inhibition of both RAN and canonical translation events shown in Figure 3C. The control transcript (GFP mRNA lacking $r(\text{CGG})^{\text{exp}}$) showed no significant change in its distribution after **2H-5-CA-Biotin** or **2'OMe-PS-(CCG)₁₂** treatment. We further studied the effect of **2H-5-CA-Biotin** on ribosome loading using a $r(\text{CGG})_{88}$ -GFP that lacks a start codon. As GFP is not translated in this system, we expected a larger effect on polysome loading upon compound treatment as compared to the $r(\text{CGG})_{88}$ -GFP construct containing a start codon. Indeed, **2H-5-CA-Biotin** reduced the amount of the transcript that lacks the start codon associated with polysomes by 10% relative to untreated sample (Figure S-14), indicating the reactive compound inhibits RAN translation in this system and does so by at least in part by reducing the amount of mRNA bound by ribosomes.

Mechanistic Considerations for 2H-5-CA-Biotin and 2'OMe-PS-(CCG)₁₂. Taken together, our data beg the questions: (i) how does **2H-5-CA-Biotin** selectively inhibit RAN translation?; and (ii) why does the oligonucleotide inhibit both RAN and canonical translation? (Figure 9). The lack of selectivity by the oligonucleotide is likely due to its formation of an extraordinarily stable secondary structure 5' to the canonical start codon. In order for a stable secondary structure to inhibit initiation, it must be close to the 5' cap.^{25,26} Stable

manufacturer's standard protocol. The compound of interest was added to the transfection cocktail, which was then applied to the cells. Approximately 5 h post-transfection, the transfection cocktail was removed and replaced with growth medium containing the compound, and the cells were incubated at 37 °C for 18–20 h. Total RNA was harvested as described above. A sample of RNA was subjected to RT-PCR as previously described,¹⁷ and the pre-mRNA splicing products were separated by denaturing polyacrylamide gel electrophoresis. The primer sequences for the SMN2 mini-gene (SMN2-F/R) are provided in Table S-1.

Chem-CLIP and C-Chem-CLIP to Pull-Down the Cellular Targets of Compounds. COS7 cells were grown in growth medium (1× DMEM, 10% FBS, and 1× GlutaMax (Invitrogen)) as monolayers in a 75 cm² flask to ~95% confluency. Cells were transfected with the plasmid of interest ((CGG)₈₈-GFP or GFP) using lipofectamine 2000 (Invitrogen) per the manufacturer's recommended protocol. The compound of interest was added to the transfection cocktail (final concentration of 500 nM for 2H-5-CA-Biotin and 2P-5-CA-Biotin and 1 μM for 2H-5-MA-Biotin), which was then applied to the cells. For C-Chem-CLIP, cells were treated with 500 nM 2H-5-CA-Biotin and 50 μM 2H-5. After 5 h incubation, the transfection cocktail was removed and replaced with growth medium containing compound. The cells were then incubated at 37 °C for 18–20 h. Total RNA was extracted by using Trizol reagent (Ambion) according to the manufacturer's protocol. After RQ1 DNase (Promega) treatment, total RNAs were phenol:chloroform extracted and ethanol precipitated. Approximately 100 μg of total RNA was incubated with streptavidin beads (100 μL, Sigma-Aldrich) in 1× PBS for 2 h at room temperature with gentle shaking. The solution was removed, and the beads were washed with 5 × 200 μL H₂O for 5 min each, until the presence of RNA was no longer detected in the wash solution as determined by absorbance at 260 and 280 nm using a Nanodrop 2000C spectrophotometer (Thermo Scientific). Bound RNA was released from the beads by heating the beads in 1× Elution Buffer (100 μL, 95% formamide, 10 mM EDTA, pH 8.2) at 65 °C for 5 min. The released RNA was purified by a Quick-RNA MiniPrep Kit (Zymo Research), and cDNA was generated from 50 ng of RNA using a qScript cDNA Synthesis Kit per the manufacturer's protocol. qPCR was performed on an ABI 7900 HT Real-Time PCR System. Primer sequences for GFP mRNA are provided in Table S-1.

Mapping the Binding Sites of Small Molecules in Cellulo (Chem-CLIP-Map). COS7 cells were transfected with the (CGG)₈₈-GFP plasmid and treated with compound followed by isolation of total RNA as described above. After RQ1 DNase treatment, total RNAs were phenol:chloroform extracted and ethanol precipitated. Next, 10 μg of total RNA (100 μL) in RNase H Buffer (20 mM Hepes, pH 7.5, 100 mM KCl, 20 mM MgCl₂, and 0.1 mM DTT) was annealed with 500 pmoles of antisense oligonucleotide at 37 °C with slow cooling to room temperature over 20 min. RNase H (Life Sciences) (5 U) was added, and the mixture was incubated at 37 °C for 30 min. After, the mixture was heated at 65 °C for 20 min to deactivate the RNase H, and then the samples were incubated with streptavidin beads (100 μL, Sigma-Aldrich) for 1 h at room temperature with gentle shaking. The solution was removed, and the beads were washed with 5 × 200 μL H₂O for 5 min each, until the presence of RNA was no longer detected in the wash solution as determined by absorbance at 260 and 280 nm. Bound RNA was released from the beads by heating the beads in 1× Elution Buffer at 65 °C for 5 min. The released RNA was purified by ethanol precipitation, and cDNA was generated from 20 ng of RNA using a qScript cDNA Synthesis Kit per the manufacturer's protocol. qPCR was performed on an ABI 7900 HT Real-Time PCR System. The sequences of antisense oligonucleotide and primers (a-d) are provided in Table S-1.

Polysome Profiling. COS7 cells in 100 mm dishes were transfected with the construct of interest ((CGG)₈₈-GFP or (CGG)₈₈(No ATG)-GFP) and treated with compound as described above. About 18 h post-transfection, 100 μg/mL (final concentration) cycloheximide (CHX) was added into the dish, and cells were incubated at 37 °C for 10 min. Cells were washed twice for 5 min with 10 mL of 1× PBS supplemented with 100 μg/mL CHX and then twice

for 5 min with 5 mL of 1× PBS supplemented with 100 μg/mL CHX and 1 mM PMSF. Cells were scraped into 1 mL 1× PBS with 100 μg/mL CHX and 1 mM PMSF, pelleted by centrifugation, and resuspended in 400 μL lysis buffer (20 mM Hepes, pH 7.5, 100 mM KCl, 5 mM MgCl₂, 0.3% NP40, RNasin, protease inhibitor, 100 μg/mL CHX, and 1 mM DTT). Lysed cells were pelleted by centrifugation at 13,200 rpm in a cold microfuge. The resulting supernatant was layered onto 10 mL linear 10–50% sucrose gradients containing 20 mM Hepes pH 7.4, 5 mM MgCl₂, 100 mM KCl, 300 mM DTT, and 100 μg/mL CHX. The samples were centrifuged at 40,000 rpm for 2 h at 4 °C. The gradients were fractionated using a fraction collector (Brandel Inc.), and the absorbance of cytosolic RNA was recorded at 254 nm by an inline UV monitor. Total RNAs were extracted from 100 μL of each fraction by using a Quick-RNA MiniPrep Kit (Zymo Research). A sample of RNA was subjected to qRT-PCR with primers (GFP-F/R) as described above.

Melting Analysis of r(CGG)₁₂. To determine the effect of the reaction of 2H-5-CA-Biotin with r(CGG)₁₂ on the thermodynamic stability of the RNA, optical melting experiments were completed. Briefly, 10 μM of r(CGG)₁₂ (with a trace amount of ³²P-labeled r(CGG)₁₂) in 1× PBS was folded by heating at 60 °C for 5 min and slowly cooling to room temperature. Then, 5 μM of 2H-5-CA-Biotin was added, and the solution was incubated for 4 h at room temperature. The reaction mixture was passed through a Sephadex G-25 column to remove unreacted 2H-5-CA-Biotin. The products of the reaction were then captured onto 50 μL streptavidin resin by incubating for 30 min at room temperature. The resin was then centrifuged, the supernatant was removed, and the resin was washed with 1× PBST. Then, 100 μL of 1× Elution Buffer was added to the resin and was incubated at 95 °C for 5 min to recover 2H-5-CA-Biotin-r(CGG)₁₂ that was bound to the resin. The eluted RNA was ethanol precipitated and resuspended in 100 μL of water. A sample of r(CGG)₁₂ was treated similarly without addition of 2H-5-CA-Biotin for comparison. To determine the stabilization afforded by binding to 2H-5, samples of r(CGG)₁₂ and r(CGG)₁₂ treated with 5 equivalents of 2H-5 were also prepared. Optical melting of the RNA was performed at 260 nm in 1× Melting Buffer (4 mM Na₂HPO₄, pH 7.0, 93 mM NaCl, and 0.5 mM EDTA) on a Beckman DU-800 UV-vis spectrophotometer using a gradient of 1 °C/min.

■ ASSOCIATED CONTENT

📄 Supporting Information

Details for synthesis and characterization of compounds, sequences of oligonucleotides, and supplementary figures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Disney@scripps.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Profs. Peter Todd and Nicolas Charlet-Berguerand for generous gifts of the plasmids used in these studies. This work was funded by the National Institutes of Health (R01 GM097455 to M.D.D.), a FRAXA postdoctoral fellowship to W-Y.Y., and The Scripps Research Institute. M.D.D. is a Camille and Henry Dreyfus Teacher-Scholar.

■ REFERENCES

- (1) Guan, L.; Disney, M. D. *ACS Chem. Biol.* **2012**, *7*, 73.
- (2) Thomas, J. R.; Hergenrother, P. J. *Chem. Rev.* **2008**, *108*, 1171.
- (3) Guan, L.; Disney, M. D. *Angew. Chem., Int. Ed. Engl.* **2013**, *52*, 10010.
- (4) Disney, M. D. *Drug Discov. Today* **2013**, *9*, 00263.

- (5) Velagapudi, S. P.; Gallo, S. M.; Disney, M. D. *Nat. Chem. Biol.* **2014**, *10*, 291.
- (6) Wojciechowska, M.; Krzyzosiak, W. J. *Hum. Mol. Genet.* **2011**, *20*, 3811.
- (7) Day, J. W.; Ranum, L. P. *Neuromuscul. Disord.* **2005**, *15*, 5.
- (8) Sellier, C.; Rau, F.; Liu, Y.; Tassone, F.; Hukema, R. K.; Gattoni, R.; Schneider, A.; Richard, S.; Willemsen, R.; Elliott, D. J.; Hagerman, P. J.; Charlet-Berguerand, N. *EMBO J.* **2010**, *29*, 1248.
- (9) Sellier, C.; Freyermuth, F.; Tabet, R.; Tran, T.; He, F.; Ruffenach, F.; Alunni, V.; Moine, H.; Thibault, C.; Page, A.; Tassone, F.; Willemsen, R.; Disney, M. D.; Hagerman, P. J.; Todd, P. K.; Charlet-Berguerand, N. *Cell Rep.* **2013**, *3*, 869.
- (10) Zu, T.; Gibbens, B.; Doty, N. S.; Gomes-Pereira, M.; Huguet, A.; Stone, M. D.; Margolis, J.; Peterson, M.; Markowski, T. W.; Ingram, M. A. C.; 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. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 260.
- (11) Todd, P. K.; Oh, S. Y.; Krans, A.; He, F.; Sellier, C.; Frazer, M.; Renoux, A. J.; Chen, K. C.; Scaglione, K. M.; Basrur, V.; Elenitoba-Johnson, K.; Vonsattel, J. P.; Louis, E. D.; Sutton, M. A.; Taylor, J. P.; Mills, R. E.; Charlet-Berguerand, N.; Paulson, H. L. *Neuron* **2013**, *78*, 440.
- (12) Ash, P. E. A.; Bieniek, K. F.; Gendron, T. F.; Caulfield, T.; Lin, W.-L.; DeJesus-Hernandez, M.; van Blitterswijk, M. M.; Jansen-West, K.; Paul, J. W.; Rademakers, R.; Boylan, K. B.; Dickson, D. W.; Petrucelli, L. *Neuron* **2013**, *77*, 639.
- (13) Darnell, J. C.; Van Driesche, S. J.; Zhang, C.; Hung, K. Y.; Mele, A.; Fraser, C. E.; Stone, E. F.; Chen, C.; Fak, J. J.; Chi, S. W.; Licatalosi, D. D.; Richter, J. D.; Darnell, R. B. *Cell* **2011**, *146*, 247.
- (14) Bakker, C. E.; Verheij, C.; Willemsen, R.; van der Helm, R.; Oerlemans, F.; Vermey, M.; Bygrave, A.; Hoogeveen, A.; Oostra, B. A.; Reyniers, E.; De Boule, K.; D'Hooge, R.; Cras, P.; van Velzen, D.; Nagels, G.; Martin, J.-J.; De Deyn, P. P.; Darby, J. K.; Willems, P. J. *Cell* **1994**, *78*, 23.
- (15) Sudrik, C.; Arha, M.; Cao, J.; Schaffer, D. V.; Kane, R. S. *Chem. Commun.* **2013**, *49*, 7457.
- (16) Werstuck, G.; Green, M. R. *Science* **1998**, *282*, 296.
- (17) Tran, T.; Childs-Disney, J. L.; Liu, B.; Guan, L.; Rzuczek, S.; Disney, M. D. *ACS Chem. Biol.* **2014**, *9*, 904.
- (18) Florea-Wang, D.; Pawlowicz, A. J.; Sinkkonen, J.; Kronberg, L.; Vilpo, J.; Hovinen, J. *Chem. Biodivers.* **2009**, *6*, 1002.
- (19) Wang, Y. D.; Dziegielewska, J.; Wurtz, N. R.; Dziegielewska, B.; Dervan, P. B.; Beerman, T. A. *Nucleic Acids Res.* **2003**, *31*, 1208.
- (20) Johansson, H. E.; Helsham, G. J.; Sproat, B. S.; Hentze, M. W. *Nucleic Acids Res.* **1994**, *22*, 4591.
- (21) Disney, M. D.; Liu, B.; Yang, W.; Sellier, C.; Tran, T.; Charlet-Berguerand, N.; Childs-Disney, J. L. *ACS Chem. Biol.* **2012**, *7*, 1711.
- (22) Licatalosi, D. D.; Mele, A.; Fak, J. J.; Ule, J.; Kayikci, M.; Chi, S. W.; Clark, T. A.; Schweitzer, A. C.; Blume, J. E.; Wang, X.; Darnell, J. C.; Darnell, R. B. *Nature* **2008**, *456*, 464.
- (23) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7287.
- (24) Arava, Y.; Boas, F. E.; Brown, P. O.; Herschlag, D. *Nucleic Acids Res.* **2005**, *33*, 2421.
- (25) Gray, N.; Hentze, M. *Mol. Biol. Rep.* **1994**, *19*, 195.
- (26) Pickering, B. M.; Willis, A. E. *Sem. Cell. Dev. Biol.* **2005**, *16*, 39.
- (27) Stern-Ginossar, N.; Weisburd, B.; Michalski, A.; Le, V. T.; Hein, M. Y.; Huang, S. X.; Ma, M.; Shen, B.; Qian, S. B.; Hengel, H.; Mann, M.; Ingolia, N. T.; Weissman, J. S. *Science* **2012**, *338*, 1088.
- (28) Slavoff, S. A.; Mitchell, A. J.; Schwaid, A. G.; Cabili, M. N.; Ma, J.; Levin, J. Z.; Karger, A. D.; Budnik, B. A.; Rinn, J. L.; Saghatelian, A. *Nat. Chem. Biol.* **2013**, *9*, 59.