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
DIVISION OF SPONSORED RESEARCH
1523 UNION RD RM 207
GAINESVILLE FL 32611-1941

REPORT DATE: October 2020

TYPE OF REPORT: ANNUAL

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2020	ANNUAL	1SEPT2019 - 31AUG2020

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 dystrophia myotonica 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	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	162	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

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 dystrophia myotonica 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(CUG)^{exp}, r(CUG)$_{12}$. **100% complete**
Subtask 1: Development of FACS r(CUG)$_{12}$ binding assay
Months 3-4: 11/01/19-12/31/19

Subtask 2: Multiplex FACS screen of DEL using r(CUG)$_{12}$ and two competing RNAs (r(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 **In progress (may not be 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.
Subtask 1: Hit validation: inhibit r(CUG)$_{12}$-MBNL1 complex.
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
Subtask 1: Assess cell permeability & cytotoxicity of compounds
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.
Months 18-24: 02/01/21-08/31/21

Subtask 4: Study transcriptome-wide effects of compound treatment by RNA-seq
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.

Previously, small molecules have been designed to selectively target r(CUG)^{exp}, liberating MBNL1 and thereby improving DM1-associated disease defects.[1, 2] 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.[2] 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)[1, 2] or by attaching cellular uptake tags such as lysine and arginine,[3] thus enhancing the uptake and localization of the compounds. Subsequently, uptake tags[9] have also been used as spacing modules.[4] The enhanced permeability of these previously developed uptake tags has been traced to polycation's interactions with cell surface heparin sulfate.[5] However, the increased basicity of these compounds has also been correlated with promiscuity and off-target binding.[6, 7] 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,[8, 9] and degrading the RNA by appending cleaving modules (such as bleomycin A5) to structure-binding small molecules, which enhances both potency and selectivity. [1, 10]

Below, we describe the design and synthesis of 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 affinity for r(CUG)^{exp} and cellular potency, without increasing the cationic nature of the compound. The optimized binder for r(CUG)^{exp} was then appended with bleomycin A5, further enhancing potency, as compared to the parent compound in DM1 patient-derived myotubes (Figure 1).

Previously, dimeric compounds comprising two copies of an RNA-binding module and a linker moiety have been shown to be biologically active against r(CUG)^{exp}. These molecules contain a peptoid backbone,[2] which was optimized with an *N*-methyl peptide as a spacer between the two RNA-binding modules (**2H-K4NMeS**, Figure 2A).[11] Both the first generation compound and **2H-K4NMeS** improve various DM1-associated disease processes in cells; a bleomycin conjugate of **2H-K4NMeS** provided a selective, bioactive compound in a pre-clinical animal model.[10] Given the success for improving compound properties by changing the spacer module and of other work that has shown the identity of the spacer can affect affinity, uptake, and compound localization in various disease cell models,[11, 12] we further optimized the linker moiety by changing its length and geometry.

Design and in vitro evaluation of dimer 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 4-spacer module of previously reported **2H-K4NMeS** (Figure 2A & B).[1] 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* IC_{50}s 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.[13, 14] Two dimers were modestly more potent than **2H-K4NMeS**, **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 part of the backbone.[15, 16] The implications of this *cis-trans* isomerization in biological systems translates to distinct conformational flexibility, hydrogen bonding interactions, hydrophobic interactions and solvation,[17] 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 IC_{50}s in the TR-FRET assay (Figure 3A). Notably, **2H-K2-Pro**, containing two prolines in the linker, was the most potent with an IC_{50} of 5.3 ± 0.8 μM, 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 area (TPSA), were calculated and no significant differences were observed amongst the proline derivatives.

The binding affinity and selectivity of **2H-K2-Pro** was next evaluated using a direct binding assay, measuring the change in the inherent fluorescence of the RNA-binding modules as a function of RNA concentration, whether 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 EC_{50} of 150 ± 5 nM, while binding was not observed to the GC base-paired control RNA (Figures 3B). Additionally, **2H-K2-Pro** showed an ~2-fold greater affinity for r(CUG)$_{12}$ than the parent compound **2H-K4NMeS** (EC_{50} = 280 ± 90 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 is capable of disruption of a pre-formed complex between r(CUG)^{exp} and 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.[18-20] We therefore studied the rescue of these DM1-associated defects by **2H-K2-Pro**. DM1 patient-derived fibroblasts harboring 1300 repeats were differentiated into myotubes[21] as a robust model of human disease. We first assessed improvement of the MBNL1-dependent splicing of its own exon 5;[22] in DM1-affected cells exon 5 is included too frequently, 50% vs. 15% in healthy cells, as previously observed.[10] Notably, **2H-K2-Pro** rescued splicing of *MBNL1* exon 5 in a dose-dependent fashion (Figure S2). At the 5 μM dose, both **2H-K2-Pro** and **2H-K4NMeS** improved the *MBNL1* exon 5 splicing defect similarly, by ~30% (Figure 3D). 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 μM compound (Figures 3E and 3F).

Potential reasons that the ~3-fold enhancement in **2H-K2-Pro**'s *in vitro* potency was not recapitulated in cells could be due to 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 compound into lysate from untreated myotubes. Interestingly, uptake of the two compounds was similar. 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, 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 μM and was not investigated further (Figure S3A). Interestingly, **2H-K4-OH-Pro** was taken up into cells at a higher concentration than **2H-K4NMeS**, while the permeability of **2H-K4-D-Ala** was similar to **2H-K4NMeS**. 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. Neither of these compounds improved *MBNL1* exon 5 splicing in DM1 myotubes, likely due to the combination of the less favorable subcellular localization or linker length. 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.[1, 10, 23, 24] That is, **2H-K2-Pro** was conjugated to bleomycin A5 via its amine (known to contribute to DNA binding[25]) to yield **2H-K2-Pro-bleo** (Figures 4A & 4B).[10] The affinity of **2H-K2-Pro-bleo** was similar to **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).

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 μM (Figure 4D). No significant RNA cleavage was observed with **Ac-K2-Pro-bleo**, as expected (Figure 4D).[1, 10]

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, thus ablating its ability to cleave DNA.[10] 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.[26] Indeed, **2H-K2-Pro-bleo** did not cause a significant increase in γ–H2AX in DM1 myotubes, whereas bleomycin A5 alone showed a ~6-fold increase in the number of foci observed per cell (Figure 5), as expected based on its ability to cause DNA damage in cancer cells.[26] Thus, DNA recognition and cleavage by bleomycin A5 is 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-assiociated 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**).[10] Notably, **2H-K2-Pro-bleo** reduced *DMPK* levels by 45 ± 4% at 5 μM, whereas only ~30 ± 3% decrease was observed with Cugamycin at 5 μM (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], 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 binder, both bleomycin A5 conjugates are taken up ~4-fold less (Figure S8).

Next, the ability of **2H-K2-Pro-bleo** to rescue DM1-associated splicing defects and the formation of nuclear foci was investigated. **2H-K2-Pro-bleo** rescued the *MBNL1* exon 5 splicing defect by 50 ± 8% at 5 μM and 30 ± 7% at 1 μM (Figure 6B), similar to 5 μM **2H-K2-Pro** (Figure 3D). Thus, in this assay, **2H-K2-Pro-bleo** is about 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 kinase 4 (*MAP4K4*) exon 22a, which is NOVA-, not MBNL1-, dependent. Likewise, **2H-K2-Pro-bleo** did not change *MBNL1* exon 5 splicing patterns in wild-type myotubes. Altogether, these control experiments indicate specificity for r(CUG)^{exp}. Next, we evaluated the ability of **2H-K2-Pro-bleo** (1 μM) 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 & 6D).

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 μM **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 μM) inhibited *DMPK* cleavage more efficiently than **2H-K4NMeS** (EC_{25} = 1 μM) and restored levels to that observed in untreated cells (Figure 6E).

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 that of r(CUG)^{exp}, as shown in our previous folding analysis.[10] 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. 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.[10]

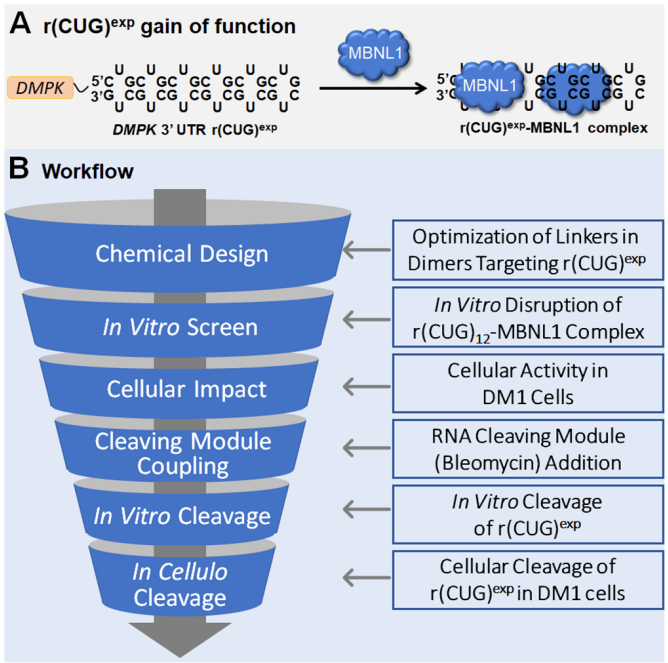


Figure 1: The design of potent dimers targeting r(CUG)^{exp} in DM1. (*A*) r(CUG)^{exp} in the 3' UTR of *DMPK* folds into a hairpin displaying a periodic array of internal loops that bind and sequester MBNL1, resulting in pre-mRNA splicing defects and formation of nuclear foci. (*B*) Step-by-step methodology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

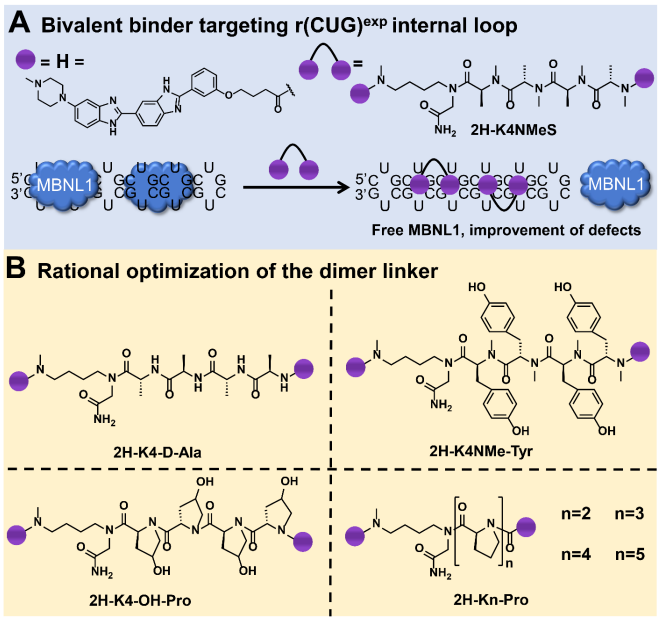


Figure 2: Design of small molecules that selectively target r(CUG)^{exp}. (*A*) Bivalent small molecules linked together using an *N*-methyl peptide scaffold (**2H-K4NMeS**) with two **H** RNA-binding modules (purple spheres). The dimer binds two r(CUG)^{exp} loops simultaneously and releases MBNL1,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

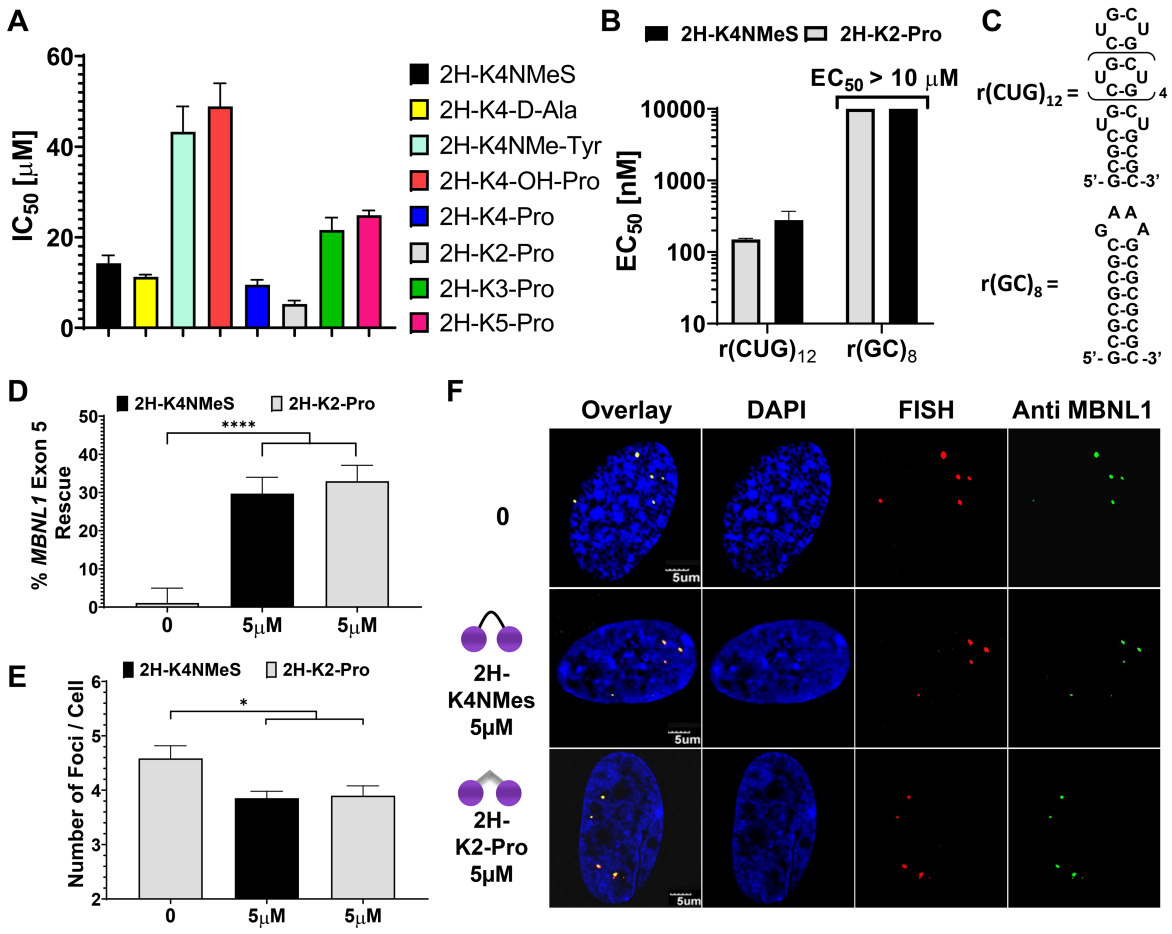


Figure 3: Studying the ability of designer compounds to inhibit the r(CUG)^{exp}-MBNL1 complex in vitro and assessment of compound activity in DM1 myotubes. (*A*) *In vitro* IC_{50}s for disruption of the r(CUG)^{exp}-MBNL1 complex by new dimer derivatives, compared to parent compound (n = 3 replicates; 2 independent experiments). (*B*) Binding affinities (EC_{50}s) 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, E, 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 r(CUG)^{exp}-MBNL1 foci/nucleus. Error bars represent SD, n = 3 biological replicates, 40 nuclei counted per 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 **2K-K2-Pro** or **2H-K4NMeS**.

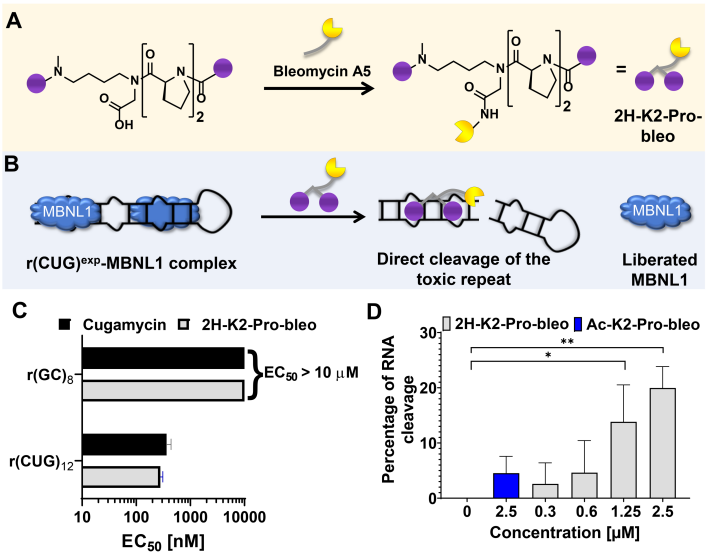


Figure 4: Design and *in vitro* evaluation of small molecules that cleave r(CUG)^{exp}. (*A*) The binder **2H-K2-Pro** was conjugated to bleomycin to afford the cleaver **2H-K2-Pro-bleo**. (*B*) **2H-K2-Pro-bleo** binds to r(CUG)^{exp}, cleaves the toxic RNA repeat, and displaces MBNL1, thereby improving 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 replicates); 2 independent experiments 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.

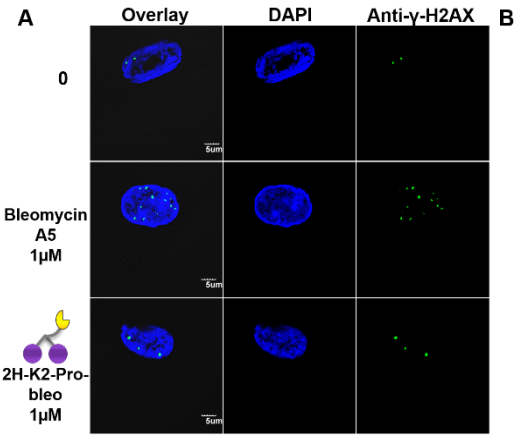


Figure 5: Effects of small molecules on the DNA damage response pathway. *(A)* Representative images from γ-H2AX immunofluorescence to assess DNA damage in DM1 myotubes upon treatment with **2H-K2-Pro-bleo** and Bleomycin A5. (*B*) Quantification of the number of γ -H2AX foci/nucleus. Error bars represent SD; n = 3 biological replicates, with 40 nuclei counted per replicate. ****$P < 0.0001$, as determined by one-way ANOVA.

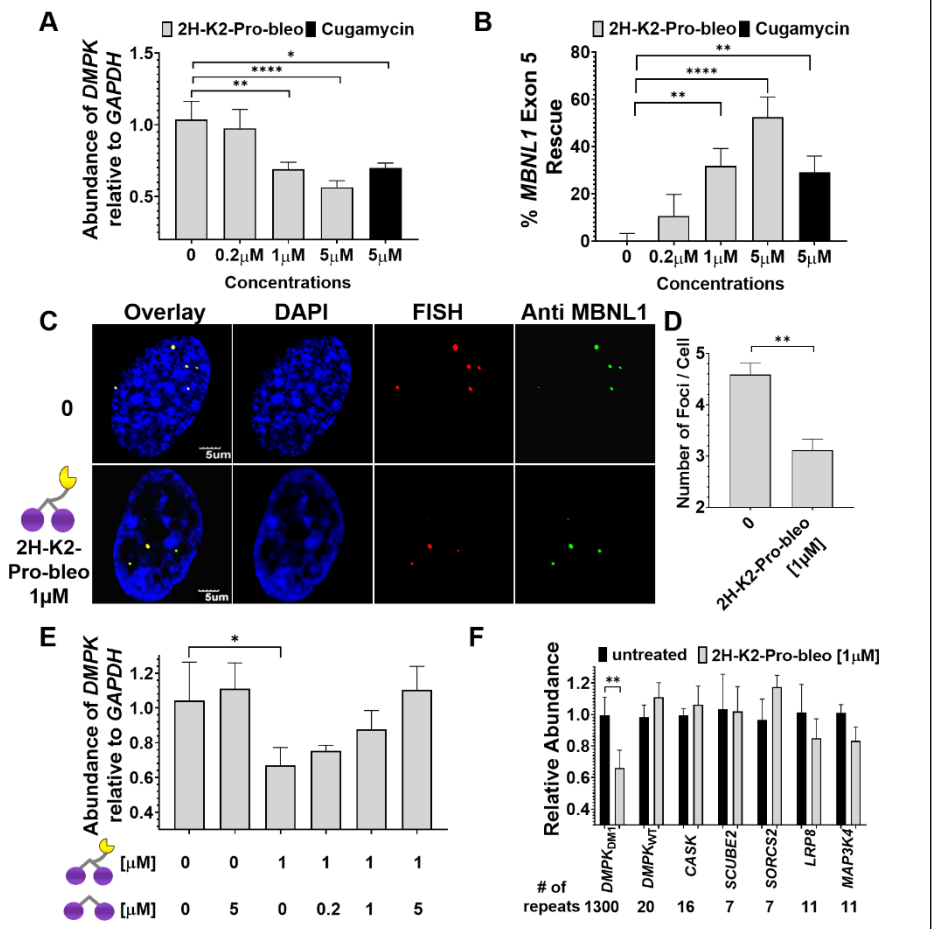


Figure 6: Biological activity of 2H-K2-Pro-bleo in DM1 patient-derived myotubes. (*A*) Effect of **2H-K2-Pro-bleo** on r(CUG)^{exp}-containing *DMPK* levels in DM1 myotubes compared to Cugamycin, as determined by RT-qPCR. Error bars represent SD, n = 3 biological replicates. **P* < 0.05, ***P* < 0.01, *****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 r(CUG)^{exp}-MBNL1 foci/nucleus. Error bars represent SD, n = 3 biological replicates, 40 nuclei counted per 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*) The effect of **2H-K2-Pro-bleo** (1 μM) on the abundance of mRNAs containing more than six, 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).

On-going studies: DNA-encoded library (DEL)

Note: The section below summarizes activities completed by the Paegel Laboratory at University of Californa, Irvine

During the reporting period, the Paegel Laboratory relocated to the University of California, Irvine. We have successfully rebuilt our DNA-encoded library (DEL) synthesis platform and brought all necessary instrumentation back on-line. Efforts quickly shifted toward curation of our DEL building block (BB) stocks, computational DEL design, DEL synthesis, and quality control (QC). DEL Synthesis and QC procedures were followed according to our previous publication (10.1021/acscombsci.5b00106).

Our goal during this period was to design, synthesize, and produce QC data for a "druglike" DEL. This first project DEL followed a previously published and highly successful DEL reaction sequence that entailed coupling an amino acid (AA) with a carboxylic acid (CA).

In the design phase, we conducted a full analysis of our BB collection to purge those containing previously identified nuisances while procuring new BBs to maximize structural diversity. Nuisances included ambiguous stereochemistry, metabolic liabilities, spuriously reactive functional groups, and protecting groups. BBs exhibiting these qualities tended to complicate validation studies. We also wished to expand our collection to include new BBs that maximized structural diversity within the accepted space of Lipinski/Veber druglikeness. We produced new code for analyzing our BB collection by Tanimoto chemical similarity clustering (> 0.7). New BBs (42 AAs, 29 CAs) were procured. Finally, we included a minor tweak in our encoding scheme; the coding for the new library assigned multiple codes to each BB during synthesis. This approach allowed us to use a simpler and shorter DNA encoding tag, which should enhance sequencing confidence and quality. New coding oligonucleotides were thus procured. Total library chemical complexity was 192 AAs x 288 CAs, or 55,296 unique compounds. An analysis of chemical similarity to PubChem and ChemBL databases is underway.

We successfully executed the DEL synthesis. New BBs and coding procedures were implemented uneventfully. We had previous modified our solid-phase synthesis procedures to enable greater scale of DEL production. The first project DEL synthesis was executed at the 1 g (2E9 beads) scale, corresponding to ~36k equivalents of DEL. Synthesis procedures were still conducted manually using multi-channel pipettors. We are exploring options for robotic automation.

QC methods have remained unchanged for the moment. We continue to conduct DEL synthesis on mixed-scale 10- and 160-µm beads. The 160-µm QC beads are harvested post-synthesis, manually segregated into single microplate wells for qPCR analysis and Sanger DNA sequencing, and the analyzed beads are then treated w/ acid and analyzed by MALDI-TOF MS. UCI funded the procurement of a new Thermo LC-MS system, and we are now in the process of developing analytical methods for automated synthesis yield calculation using the QC beads from this DEL as a testbed.

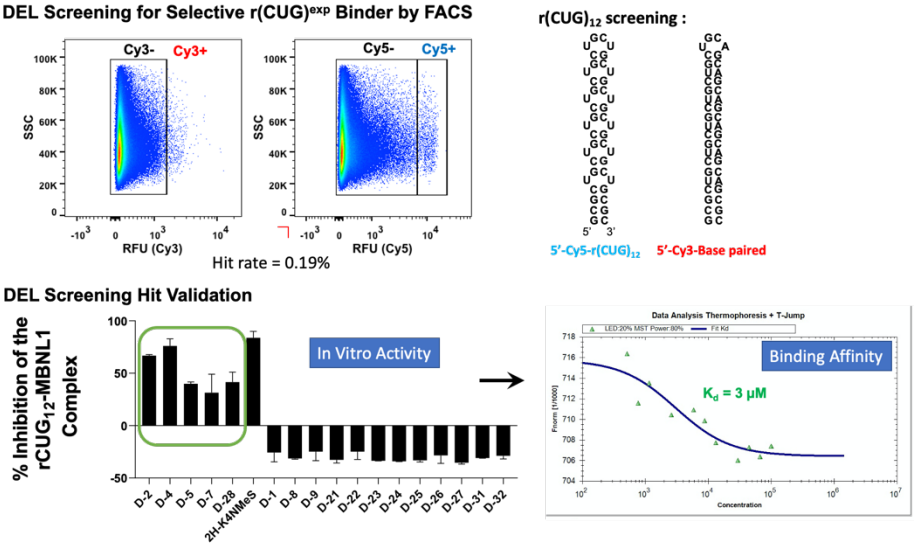
Finally, early in our time at UCI, we learned that the DNA sequencing facility does not support the ION Torrent platform. The Torrent was previously our preferred sequencing platform for cost

and turnaround. As a result, we both curated primers for Torrent library preparation and had only developed code for reducing Torrent data files to chemical structure. While rebuilding synthesis capabilities, we developed a sequencing library preparation protocol that would be compatible with UCI's Illumina sequencers. Using a 3 ~1000- and ~100 bead aliquots of a previous library, we showed that we could successfully detect bead-derived sequences from a solid-phase DEL.

Thus far, we have assayed 32 QC beads from the newly synthesized library. Of the 32 beads, 30 produced sequence that predicted a chemical synthesis product and/or truncate that was the predominant feature in the corresponding MALDI-TOF mass spectrum. The two beads that did not validate did not produce legible sequence. There were no QC beads indicating a chemical structure that did not match with the DNA codes. These metrics indicate that the DEL was successfully synthesized. An aliquot (250 mg, 9000 DEL equivalents) are to be transferred to Disney Lab for screening.

Note: The section below summarizes activities completed by a collaborative effort between the Paegel (UCI) and Disney Laboratories (TSRI) related to DEL screening.

Per our SOW, 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 Cy-3, 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 0.19%. We completed two secondary validations of these hits: (i) *in vitro* activity as assessed by inhibition of the r(CUG)12-MBNL1 complex by using a previously established TR-FRET assay;[27] and (ii) affinity measurement of those with inhibitory activity by microscale thermophoresis (MST). These data are summarized in the figure below. We are currently evaluating these lead compounds in cellular models of DM1.



Screening of a DNA-encoded library (DEL) to identify small molecules with enhanced affinity for and activity against the DM1 RNA. Top, representative FACS analysis for selective binders of r(CUG)$_{12}$ vs. a fully paired RNA. The secondary structures of the RNAs screened are shown to the right. **Bottom**, secondary validation of the DEL screening hits by using an in vitro activity assay that measured inhibition of the r(CUG)$_{12}$-MBNL1 complex. Affinity of the inhibitors were then measured by microscale thermophoresis (MST).

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).

One approach to alleviate DM-associated defects is to utilize small molecules that recognize the structure of r(CUG)^{exp}, thereby liberating bound MBNL1 or preventing its binding.[28-30] Alternatively, expression of r(CUG)^{exp} has been reduced or eliminated by using RNA targeted-Cas9 editing of r(CUG)^{exp},[31] ASOs,[32, 33] DNA-binding small molecules that inhibit transcription,[34, 35] and small molecules that bind and directly cleave r(CUG)^{exp}.[28, 36, 37] The latter approach (Fig. 7) has been accomplished with Cugamycin, a small molecule that selectively binds r(CUG)^{exp}'s structure conjugated to the natural product bleomycin A5 (BLM) (Fig. 8). Indeed, Cugamycin broadly improved DM1-associated defects with no off-target effects in a mouse model of DM1.[36]

BLM is an anti-cancer natural product that cleaves DNA and RNA through H-atom abstraction and the production of a radical species by the metal binding core.[25, 38] Extensive structure-activity relationship (SAR) evaluations of BLM derivatives[25, 39] and structural data for DNA-bound BLM[40, 41] have revealed structural components that are essential for metal coordination, oxygen activation, DNA binding, and subsequent cleavage. This information has been used to guide attachment of RNA-binding small molecules at the C-terminal amine of BLM, eliminating a charge critical to DNA binding and producing BLM-conjugated compounds that specifically cleave a target RNA.[36, 42] SAR studies of BLM can guide the selection of analogs to further enhance RNA selectivity by eliminating DNA-binding interactions.[25] One such analog is deglycobleomycin (DeglycoBLM; Fig. 8) in which the disaccharide moiety of BLM is removed. The carbohydrate domain can contribute to DNA binding affinity by participating in hydrogen bonding interactions with the DNA backbone, and DeglycoBLM cleaves DNA between 2-5 fold less efficiently than BLM.[25, 41] This disaccharide also contributes to the cellular permeability of BLM.[43] Collectively, attachment of DeglycoBLM to small molecules targeting r(CUG)^{exp} can further reduce its affinity for DNA to enhance RNA selectivity in cells, provided the compound retains cellular permeability. The examination of such features is summarized below.

DeglycoBLM was synthesized via HF-pyridine cleavage of the carbohydrate of BLM[44] and conjugated to a dimeric compound that recognizes r(CUG)^{exp} (2H-K4NMeS, **3**)[28] to afford compound **1** (DeglycoCugamycin; Fig. 8). A control compound that does not contain the RNA-binding modules in 1 and thus has no affinity for the RNA target was also synthesized (compound **2**; Fig. 2). To assess the molecular recognition of **1**, its affinity for r(CUG)$_{12}$, r(GC)$_8$, and DNA was measured in the absence of Fe(II). Compound **1** only bound avidly to r(CUG)$_{12}$ (K_d = 610 ± 150 nM) (Fig. 9A), which is comparable to the affinity of Cugamycin (K_d = 365 ± 75 nM).[28] Thus, removal of the carbohydrate domain does not affect the ability to bind r(CUG)^{exp} *in vitro*.

Next, the ability of **1** to cleave r(CUG)$_{10}$ and DNA was assessed *in vitro*. Cugamycin and 1 cleaved r(CUG)$_{10}$ to a similar extent at the same concentrations (~35% cleavage at 1 μM), while BLM only cleaved r(CUG)$_{10}$ by 15% at 2 μM (Fig. 9B). In contrast, DeglycoBLM and **2** (lacks RNA-binding modules) were unable to cleave r(CUG)$_{10}$ at the concentrations tested (up to 2 μM; Fig. 9B), as expected since DeglycoBLM alone is 5-fold less efficient at cleaving nucleic acids than BLM.[19] Thus, functional RNA cleavage by **1** is not affected through removal of the disaccharide. The selectivity of the observed cleavage was assessed by measuring DNA cleavage (Fig. 9C). While BLM efficiently cleaved DNA *in vitro* with >50% cleavage observed at all concentrations (250 nM – 2 μM), DeglycoBLM cleaved DNA about 5-fold less efficiently, with >50% cleavage only observed at 2 μM (Fig. 9C), consistent with previous studies.[25] We previously showed that Cugamycin does not cleave DNA when r(CUG)$_{12}$ is present,[36] and thus is selective for cleaving the RNA target. However, when incubated in the absence of r(CUG)$_{12}$, Cugamycin cleaved DNA at concentrations >500 nM (Fig. 9C). In contrast, **1** and **2** did not significantly cleave DNA at any of the concentrations tested (250 nM – 2 μM; Fig. 9C). Thus, by eliminating two key DNA-binding interactions through removal of the disaccharide and attachment of the r(CUG)-binding compound at the C-terminal amine, DNA cleavage is further ablated and selectivity for r(CUG)^{exp} is enhanced.

To study potential off-target binding to DNA in cells, we measured the amount of phosphorylated histone H2A variant H2AX (γ-H2AX), which forms foci in response to DNA double strand breaks, induced by compound treatment in the rapidly growing mouse myoblast cell line C2C12 and in DM1 patient-derived myotubes. In C2C12 cells, we used a fluorescence resonance energy transfer (FRET) assay to quantify γ-H2AX foci after treating with the compound of interest for 24 h. BLM caused a significant increase in γ-H2AX at all concentrations tested (1.65 – 25 μM), as expected from previous studies[45] (Fig. 9D). In contrast, Cugamycin only induced DNA damage at 25 μM, a concentration that is ~10 fold higher than its bioactive concentration in DM1 myotubes,[36] while no increase in γ-H2AX foci was observed for 1 or DeglycoBLM upon treatment with up to 25 μM compound (Fig. 9D). Importantly, and consistent with these studies in C2C12 cells, neither 1 nor DeglycoBLM induced DNA damage in DM1 patient-derived myotubes, as determined from immunostaining and imaging by fluorescence microscopy. [Note: the signal:noise observed in the FRET assay described above for C2C12 cells was not sufficient for quantification in DM1 myotubes.] Thus, **1** further diminished off-target DNA cleavage in cells as compared to Cugamycin, in agreement with *in vitro* DNA cleavage analysis.

To probe if the difference in DNA damage in cells is due to changes in cellular uptake, as the disaccharide has previously been implicated in cell permeability,[43] the concentration of Cugamycin, **1**, and the dimer from which they are derived (**3**[28]) taken up into DM1 myotubes was determined by measuring the fluorescence of the RNA-binding modules after washing and lysing treated cells. Cugamycin and **1** had similar cell permeabilities, and both compounds were only ~3-fold less permeable than **3**. To confirm these results, permeability and localization were compared by using live-cell fluorescence microscopy. Both Cugamycin and **1** localized in the nucleus where r(CUG)^{exp} is sequestered in foci to a similar extent. Thus, although the carbohydrate domain has been shown to affect the permeability of DeglycoBLM itself and may account for its lack of DNA damage in cells (Fig. 9D), the disaccharide did not affect permeability of conjugate compounds, as determined by comparing Cugamcyin and **1**.

Since **1** ablated DNA damage observed for Cugamycin without reducing cell permeability, its ability to cleave r(CUG)^{exp} and improve DM1-associated defects in cells was measured. In DM1 patient-derived myotubes,[46] **1** cleaved ~30% of r(CUG)^{exp}-containing *DMPK* at low micromolar concentrations (Fig. 10A), which is comparable to the cleaving activity of Cugamycin.[36] Importantly, 2, which lacks RNA-binding modules, did not affect *DMPK* levels. To demonstrate that reduction in *DMPK* levels was due to direct cleavage of the RNA, a competition experiment was performed in which cells were co-treated with **1** and **3**; **3** binds r(CUG)^{exp} but does not affect *DMPK* mRNA levels (Fig. 10B). Indeed, upon co-treatment, cleavage by **1** was inhibited by **3**, and *DMPK* levels were restored to levels similar to untreated samples or samples treated with **3** alone (Fig. 10B). Notably, **1** was selective for cleaving r(CUG)^{exp}, as *DMPK* levels were not affected in wild-type cells expressing r(CUG)$_{20}$ (Figs. 10C), and mRNAs containing short non-pathogenic r(CUG) repeats were also unaffected (Fig. 10C). We have previously shown that this selectivity is due to structural differences in transcripts containing short r(CUG) repeats vs. r(CUG)^{exp}, as the small molecule recognizes the structure formed by the repeat expansion.[36] Indeed, compounds that recognize the structure of r(CUG)^{exp} can be selective for the toxic disease-driving repeat expansion; that is, structure-targeting ligands can be allele-selective.[36] In contrast, an ASO complementary to the r(CUG) repeats is not able to discriminate between short and long repeats and thus has off-target effects.[36]

After confirming that **1** cleaved r(CUG)^{exp} with similar selectivity and potency as Cugamycin, the ability of 1 to rescue DM1-defects, including formation of r(CUG)^{exp}-MBNL1 nuclear foci[47] and MBNL1-regulated splicing defects,[48] was assessed. At 2 μM, **1** reduced the number of r(CUG)^{exp}-MBNL1 nuclear foci by ~40% (Figs. 10D & E), similar to Cugamycin,[36] while 2, which lacks RNA-binding modules, had no effect. In DM1 myotubes, *MBNL1* exon 5 splicing is dysregulated (Fig. 7B), as MBNL1 regulates the alternative splicing of its own pre-mRNA.[49] Cleavage of r(CUG)^{exp} by 1 resulted in an ~30% improvement in the *MBNL1* exon 5 splicing defect (Fig. 10F), a similar improvement to that observed for Cugamycin.[36] Compound 2 had no effect on *MBNL1* exon 5 splicing, as expected. Importantly, 1 did not affect *MBNL1* exon 5 splicing in wild-type myotubes nor the NOVA-dependent splicing of *MAP4K4* exon 22a[50]. Thus, rescue of the *MBNL1* exon 5 splicing defect can be traced to the specific to cleavage of r(CUG)^{exp}. Collectively, these studies show that removal of the carbohydrate domain of BLM allows for enhanced selectivity by further ablating DNA damage without affecting cellular permeability or activity.

Small molecules that selectively cleave a target RNA are attractive chemical probes as they can more potently improve disease-associated defects than simple binding compounds.[36, 51] Furthermore, RNA cleavage, either through direct cleavage as demonstrated herein or through recruitment of a cellular nuclease,[51] can be used to profile molecular recognition of RNA-binding small molecules. The use of BLM analogs to specifically cleave r(CUG)^{exp} offers an attractive method to enhance RNA cleavage selectivity by further diminishing off-target DNA cleavage. Although the carbohydrate domain of BLM is necessary for its efficient cleavage of DNA[25] and cellular permeability,[43] the disaccharide is not essential for permeability or cleavage of r(CUG)^{exp} when attached to r(CUG)^{exp}-binding small molecules. Thus, by using BLM analogs, RNA cleavage and the ability to improve DM1-associated defects is retained while further enhancing selectivity by reducing DNA damage that occurs with high concentrations of BLM-conjugated small molecules.

The most common way to target RNA for destruction is by using oligonucleotide-based target recognition elements. These approaches recognize unstructured regions in RNA. The ability to design ligands that target structured regions in an RNA and cleave them selectively provides an alternative approach to probe the biology of RNA in general and RNA structure in particular. Although bleomycin-small molecules conjugates have higher molecular weights than orally bioactive drugs, they are still lower molecular weight than oligonucleotides and significantly lower molecular weight than CRISPR approaches that are packed into viruses.[31] Additionally, medicinal chemistry approaches may be more broadly applicable to these compound sets as the RNA-binding modules and linkers that tether them can be therapeutically optimized. It is likely that as more information is accumulated on the RNA folds that bind small molecules and on the small molecules that bind RNA folds that the deglycobleomycin cleavage module described herein could be broadly deployed. Furthermore, the ability to effect cleavage of an RNA target can allow for more diverse modes of action. Small molecules can now target an RNA for destruction in cells via three mechanism: (i) direct cleavage by using bleomycin conjugates[28, 36]; (ii) nuclease recruitment by using ribonuclease targeting chimeras (RIBOTACs)[51, 52]; and (iii) shunting introns with toxic expanded repeats to decay pathways.[53] Some targets may be more amendable to one strategy than the others. The ability to minimize off-target effects by using the deglycobleomycin cleavage module described here could have broad implications in this emerging area.

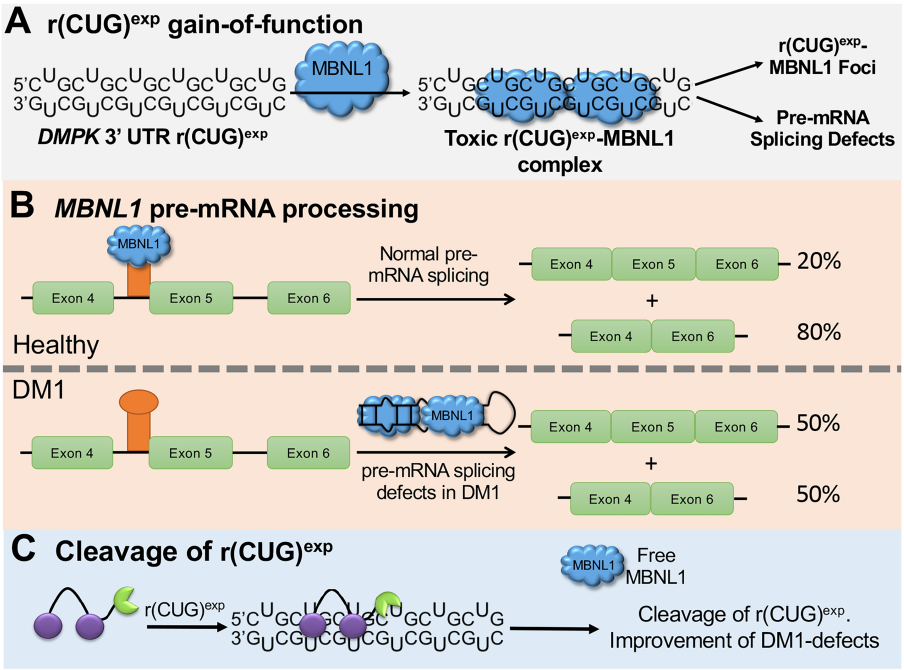


Figure 7. Small molecule cleavage of r(CUG)^{exp}. (A) DM1 is caused by r(CUG)^{exp}, located in the 3' UTR of the *DMPK* gene, which forms a structure with repeating 1 x 1 U/U internal loops. The loops bind and sequester MBNL1, resulting in nuclear foci and pre-mRNA splicing defects. (B) MBNL1 protein regulates the splicing of its own pre-mRNA. When MBNL1 is sequestered by r(CUG)^{exp}, *MBNL1* exon 5 is included too frequently. (C) Scheme of small molecule cleavage of r(CUG)^{exp}, resulting in improvement of DM1-associated defects.

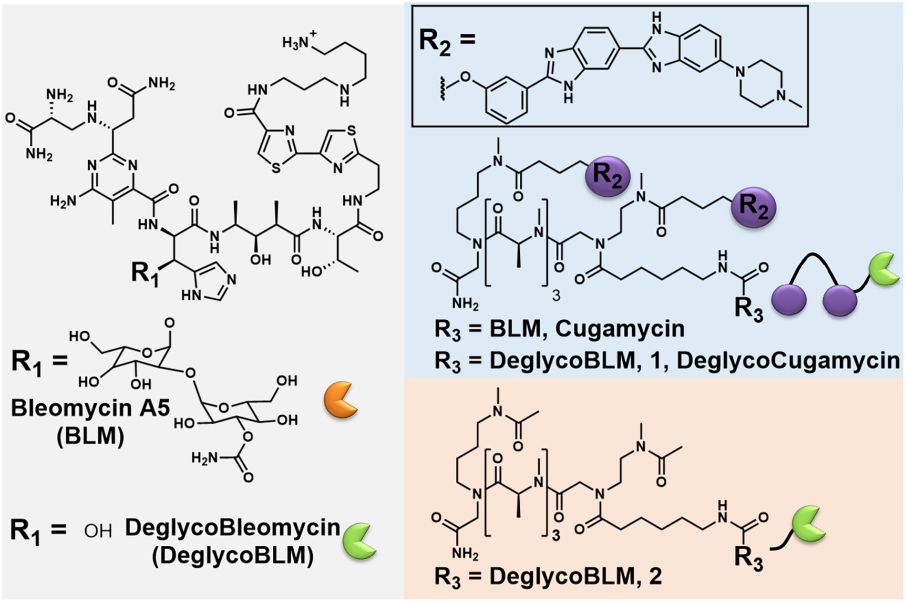


Figure 8. Chemical structures of bleomycin A5 (BLM), deglycobleomycin (DeglycoBLM), Cugamycin, **1** and **2**.

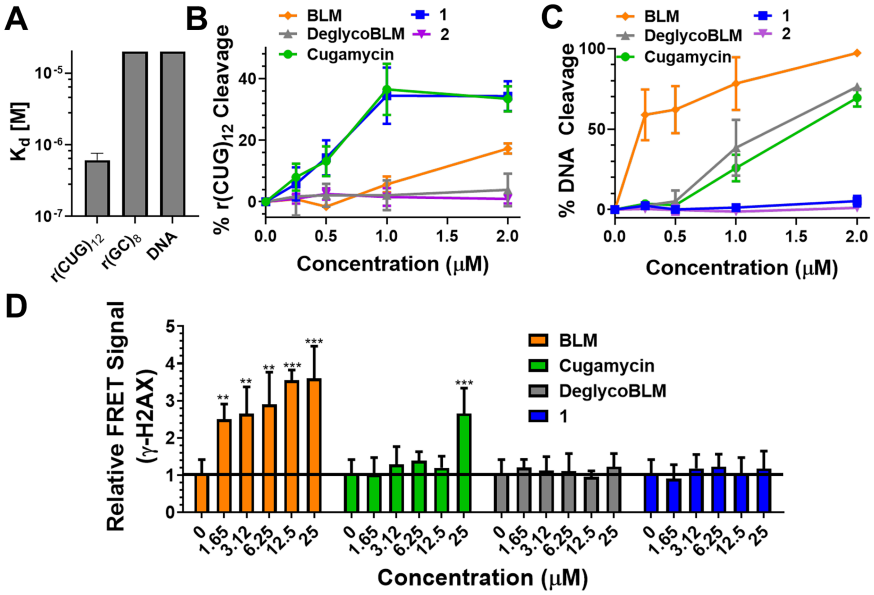


Figure 9. Cleaving capacity and selectivity of small molecule cleavers. (A) Binding affinity of **1** for $r(CUG)_{12}$ (K_d = 610 ± 150 nM), $r(GC)_8$ (K_d > 20 μM) and DNA (K_d > 20 μM); n = 3. (B) Quantification of cleavage of $r(CUG)_{10}$ by **1**, **2**, Cugamycin, BLM, and DeglycoBLM; n = 3 (C) Quantification of cleavage of DNA by **1**, **2**, Cugamycin, BLM, and DeglycoBLM; n = 3 (D) Effects of **1**, Cugamycin, BLM, and DeglycoBLM on γ-H2AX, a marker of DNA damage, in C2C12 cells. n = 8, ** $P < 0.01$, *** $P < 0.001$, as determined by comparison to untreated cells by a one-way ANOVA. Error bars indicate SD for all panels.

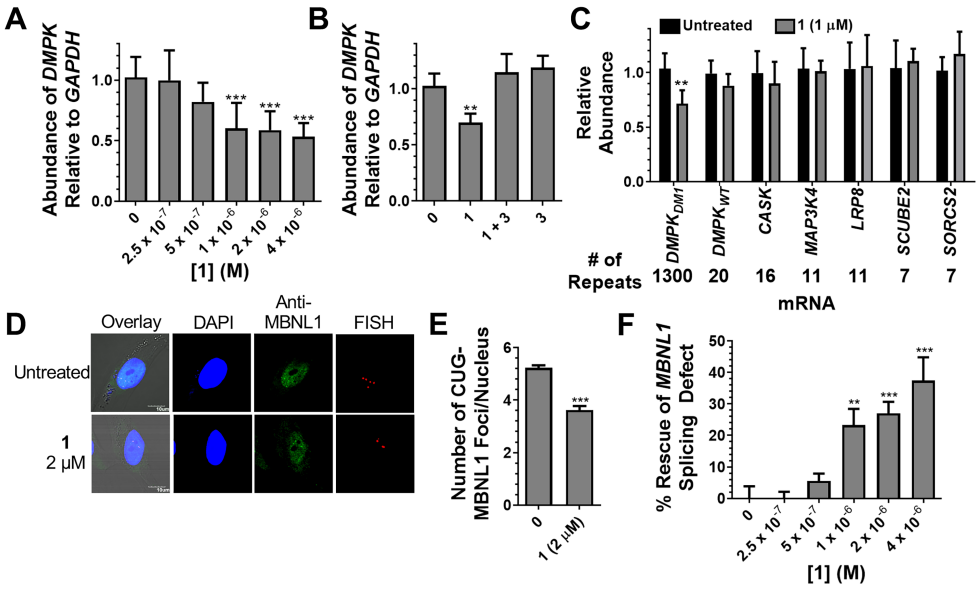


Figure 10. Activity of **1** in DM1 myotubes. (A) Cleavage of r(CUG)^{exp}-containing *DMPK* by **1** as determined by RT-qPCR. n = 6, *** $P < 0.001$, as compared to untreated cells; determined by a one-way ANOVA. (B) A competitive cleavage experiment between **1** (1 μM) and **3** (5 μM) in which **3** prevents cleavage of *DMPK*. n = 3, ** $P < 0.01$, as compared to untreated cells; determined by a one-way ANOVA. (C) Effect of **1** on r(CUG)$_n$-containing transcripts. n = 3, ** $P < 0.01$, as compared to untreated cells; determined by a Student *t*-test. (D) Representative images of r(CUG)^{exp}-MBNL1 foci imaged by RNA FISH and anti-MBNL1 immunostaining. (E) Quantification of nuclear foci. n = 3, 40 nuclei quantified/replicate, *** $P < 0.001$, as compared to untreated cells; determined by a Student *t*-test. (F) Improvement of the *MBNL1* splicing defect in DM1 myotubes upon treatment with **1**. n = 6, ** $P < 0.01$, *** $P < 0.001$, as compared to untreated cells; determined by a one-way ANOVA. Error bars indicate SD for all panels.

Wang Lab Progress Report: University of Florida

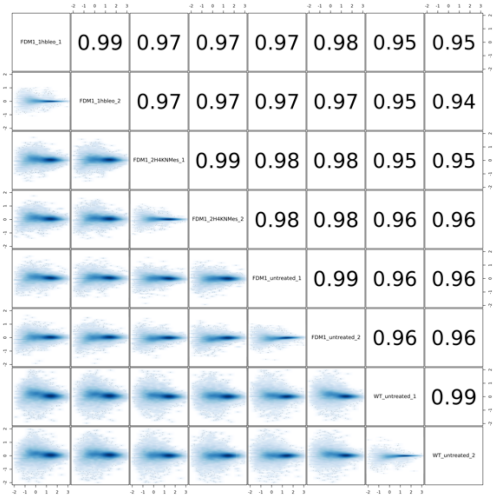
For this project period, we have focused on tasks outlined in the statement of work from months 1-12. For our site, we proposed to:

- **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 myoblasts treated with a bleomycin conjugate (1H-Bleomycin) or another tool compound (2H4KNMes). We compared these samples to WT unaffected myoblasts. We generated two RNA-seq libraries for each condition (8 libraries total) and obtained the following read coverage for each library:

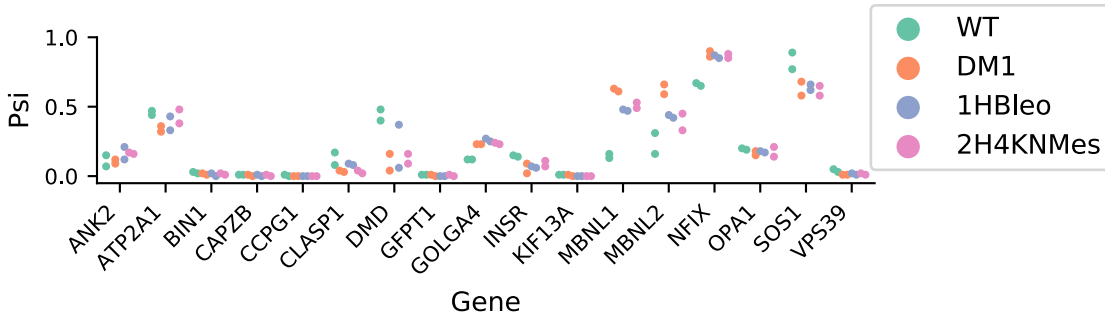
Sample	No. reads (Millions)
DM1 + 2H4KNMes	133
DM1 + 2H4KNMes	174
DM1 + 1HBleo	143
DM1 + 1HBleo	112
DM1	152
DM1	151
Unaffected	174
Unaffected	142

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.



Because a major question regarding these small molecules is whether they can rescue MBNL-dependent mis-splicing in DM1 cells, we focused on a subset of DM1-relevant exons known to be mis-regulated in human DM1 muscle. Some of these show mis-regulation in DM1 myoblasts as well, and some show rescue

in response to small molecule treatment. Plotted below are percent spliced in (PSI) values for each of these exons, in each sample:



As visualized in this plot, MBNL1 and MBNL2 contain exons that are mis-regulated between WT and DM1 myoblasts, and these show rescue in response to both 1HBleo and 2H4KNMes.

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).

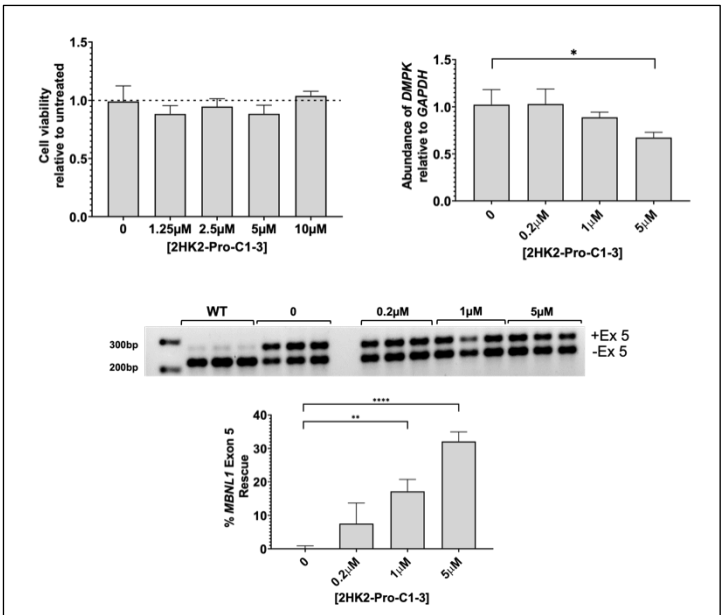
A potential approach to direct malfunctioning RNAs down endogenous decay pathways is to exploit ribonucleases (RNases) and recruit them to specific transcripts with a small molecule. RNase L, an integral part of the viral immune response, is present in minute quantities in all cells as an inactive monomer. 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.[54] 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.[55] 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 **1** and 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*.

We designed and synthesized a RIBOTAC to cleave r(CUG)^{exp} selectively by conjugating a heterocyclic small molecule that recruits RNase L[56] to our lead molecule, 2H-K4NMeS. We also synthesized a control couple to a less efficient RNase L recruiting small molecule. We first studied the ability of the RIBOTAC to cleave r(CUG)^{exp} *in vitro*. Notably, it dose-dependently cleaved r(CUG)$_{12}$, cleaving ~90% of the RNA at the highest concentration tested, 10 μM. No significant RNA cleavage was observed with the control RIBOTAC, as expected.

We next studied the cytotoxicity of the r(CUG)^{exp}-targeting RIBOTAC in DM1-patient derived cells as well as its ability to cleave the *DMPK* transcript harboring the repeat expansion. Although, a reduction in *DMPK* levels were observed at a 5 μM dose, cell viability was also affected. Interestingly, as these studies were being completed, we had optimized the linker in 2H-K4NMeS, as reported above in our summary for Specific Aim 1. We therefore created a RIBOTAC using this optimized molecule, **2H-K2-Pro.**

Importantly, no cytotoxicity was observed up to the highest concentration tested, 10 μM. Therefore, the RIBOTAC was evaluated in DM1-patient derived cells to evaluate cleavage of *DMPK* and for rescue of dysregulation of MBNL1-dependent alternative pre-mRNA splicing. The RIBOTAC reduced dose dependently reduced *DMPK* levels, significantly at 5 μM and rescued of MBNL1-dependent splicing at doses of 1 and 5 μM. Unfortunately, this potency is not significantly improved over the parent, simple binding compound. A summary of these data can be found in the figure below.

We are currently optimizing the RIBOTAC for cleaving activity against r(CUG)^{exp} by identifying *nuclear* RNase that we can recruit to the RNA. As aforementioned, r(CUG)^{exp} is sequestered in nuclear foci; RNase L is mainly localized to the cytoplasm, likely limiting the effectiveness of the RIBOTAC.



Design of a RIBOTAC that degrades r(CUG)^{exp}. Top, an optimized RIBOTAC has no effect on the viability of DM1 fibroblasts (left) and degrades the *DMPK* transcript harboring r(CUG)^{exp} (right). **Middle,** representative gel image of rescue of MBNL1 exon 5 misplacing by the optimized RIBOTAC. **Bottom,** quantification of MBNL1 exon 5 rescue.

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 1: Hit validation: inhibit r(CUG)$_{12}$-MBNL1 complex; **Months 12-15: 09/01/20-11/30/20; Disney Laboratory**

- Subtask 2: In vitro evaluation of most potent compounds from Subtask 1: affinity, k_{on}, k_{off}, residence time by biolayer interferometry; **Months 15-18: 11/01/20-02/28/21; Disney Laboratory**

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

- Subtask 1: Assess cell permeability & cytotoxicity of compounds**; Months 15-18: 11/01/20-02/28/21; Disney Laboratory**
- Subtask 2: Study non-toxic, cell permeable compounds for improving splicing defects and foci in patient-derived cells; **Months 18-24: 02/01/21-08/31/21; Disney Laboratory**

- Subtask 4: Study transcriptome-wide effects of compound treatment by RNA-seq**; Months 18-24: 02/01/21-08/31/21; Wang 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 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; ; Disney Laboratory**

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; ; Disney Laboratory**

- Subtask 2: Rigorously evaluate the ability of small molecule-RNase L conjugates to recruit RNase L and cleave r(CUG)$_{12}$; **Months 6-18: 02/01/20-02/28/21; Disney Laboratory**

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; Disney Laboratory**

- 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; ; Disney Laboratory**

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; 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 mRNAs[57] and >700 novel conserved RNA structures, almost all with unknown function, in the mammalian transcriptome.[58] 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

1. 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).

2. 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).

3. 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).

4. 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), in press. doi: 10.1039/d0cs00560f. acknowledges federal support (yes).

5. 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), in press. doi: 10.1039/d0cs00455c. acknowledges federal support (yes).

6. Fitzgerald PR & Paegel BM; *DNA-Encoded Chemistry: Drug Discovery From a Few Good Reactions*; ACS Chemical Reviews, (2020), accepted and in press; 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:*3 calendar months
Contribution to Project: Professor Disney oversees all aspects of the Initiating project.
Funding Support: in addition to this award, active awards include W81XWH2010727; UG3 NS116921; R35 NS116846; R01 CA249180; and P01 NS099114. R33 NS096032 and R01 GM097455 are both in one-year no cost extension ending in 2020 and 2021, respectively.

Name: Raphael I. Benhamou
Project Role: Postdoctoral Associate, Disney Lab
Researcher Identifier: 0000-0003-1743-0886
Nearest Person Month Worked: 1 calendar month
Contribution to Project: Completed experimental work associated with Disney Laboratory milestones/tasks
Funding Support: None at time of project participation; Myotonic Dystrophy Foundation and National Ataxia Foundation awarded in Spring 2020.

Name: Brian M. Paegel
Project Role: Co-Investigator
Nearest Person Month Worked: 1 calendar month
Contribution to Project: The Paegel lab is designing and synthesizing DNA-encoded libraries for RNA ligand discovery and providing computational support to deconvolute screening hits.
Funding Support: in addition to this award, active awards include R01 GM120491.

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(CUG)^{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(CUG)^{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?

Matthew D. Disney
New Awards Received Throughout Budget Period (2019-2020):

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

UG3 NS116921 (Mouradian/Disney) 06/01/20-05/31/21 0.76 calendar (6.33%)
NIH/NINDS
Disney lab share: (total direct costs for project) | (total costs for project) *Project period beyond Year 1 contingent upon promising first annual progress report.* **RNA Targeted Drug Discovery and Development for Parkinson's Disease**
6001 Executive Boulevard Suite 3309
Bethesda, MD 20892- 9531
Program Official: Joel A. Saydoff, joel.saydoff@nih.gov
This project aims to identify, optimize, and test drug-like compounds that modulate the levels of a key protein in Parkinson's disease and protect the brain from degeneration.
Specific Aim 1: Advance a number of biologically active small molecules of diverse chemotypes that target α-synuclein 5' UTR through *in vitro* and cellular assays.
Specific Aim 2: Pre-development *in vivo* proof of mechanism.
Specific Aim 3: Pre-development activities: Enabling Therapeutic Index (TI) assessment.
Specific Aim 4: IND-enabling work.
Specific Aim 5: Single Ascending Dose (SAD) Phase 1 trial.
Overlap: None

R35NS116846 (Disney) 05/01/20-04/30/28 6.0 calendar (50%)
NIH/NINDS
(total direct costs for project) | (total costs for project)
Design of Precision Small Molecules Targeting RNA Repeating Transcripts to Manipulate and Study Disease Biology
6001 Executive Boulevard, Suite 3290, MSC 9537
Bethesda, MD 20892-9537
Program Official: James Washington, washingj@ninds.nih.gov
We will develop generalizable and potentially transformative methods to selectively target the RNA repeat expansions that cause >30 incurable neuromuscular diseases and genetically-defined dementia.
Specific Aim 1: Study molecular recognition *in cellulis*.
Specific Aim 2: Use on-site drug synthesis to improve compound activity and selectivity.
Specific Aim 3: Interfacing SMIRNAs with natural decay processes.
Specific Aim 4: Recruit nuclease with small molecules to cleave expanded repeats selectively
Specific Aim 5: Evaluate therapeutic efficacy in mouse models.
Overlap: None

R01 CA249180 (Disney) 04/01/20-03/31/25. 1.9 calendar (15.83%)
NIH/NCI
Disney lab share: (total direct costs for project) (total costs for project)
Targeted Degradation of RNAs by Using Small Molecules
9000 Rockville Pike
Bethesda, MD 20892-9086
Program Official: Jeffrey Smiley, smileyja@csr.nih.gov

Our goal is to develop a new strategy to destroy toxic RNAs that cause the metastatic phenotype using small molecules. We will use a transformative approach to drug RNA, by generating chimeric small molecules that selectively bind to and recruit nucleases to cleave a desired RNA target.
Specific Aim 1: Comprehensively characterize our lead RIBOTAC (5) against miR-21 *in vitro* and *in situ*, a benchmark for lead optimization.
Specific Aim 2: Lead optimize 5 to deliver a proof-of-concept compound with properties amenable for *in vivo* testing.
Specific Aim 3: Assess efficacy of optimized RIBOTACS against a panel of TNBC lines and TNBC PDX tumor cells.
Overlap: None

Changes to Information Previously Listed on Prof. Disney's Post-Submission PCPS:
DP1 NS096898: formerly active; completed effective 07/31/2020.
R21/33 NS096032: currently in 1-year no cost extension at decreased effort.
R01 GM097455: currently in 1-year no cost extension at decreased effort.

Changes to Information Previously Listed on Prof. Wang's Post-Submission PCPS:
Target ALS Consortium Grant: "Therapeutic Potential of CRISPR-Cas9 in C9ORF72 - Repeat Deletion and Transcriptional Repression Strategies in C9ORF72 BAC Transgenic Mice": formerly active; completed effective 04/30/2020.

Changes to Information Previously Listed on Prof. Paegel's Post-Submission PCPS:
R42 GM115130 and **R21 EB024116**: formerly active; completed effective 03/31/2019 and 02/29/2020, respectively.

What other organizations were involved as partners?

Organization Name:	The University of Florida
Location of Organization:	Center for NeuroGenetics Dept. of Molecular Genetics & Microbiology 2033 Mowry Road Gainesville, FL 32610
Partner's Contribution:	Collaboration (academic worksite of Partnering PI Eric Wang)
Organization Name:	The University of California, Irvine
Location of Organization:	101 Theory, Suite 154, Mail Code 3958 Irvine, CA 92617
Partner's Contribution:	Collaboration (academic worksite of Co-I Brian Paegel)

8. SPECIAL REPORTING REQUIREMENTS

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

QUAD CHARTS: Please see attached appendices with quad chart.

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

REFERENCES

1. Rzuczek, S.G. et al. Precise Small-Molecule Recognition of a Toxic CUG RNA Repeat Expansion. *Nat. Chem. Biol.* **13**, 188-193 (2017).
2. 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).
3. Lee, M.M., French, J.M. & Disney, M.D. Influencing Uptake and Localization of Aminoglycoside-Functionalized Peptoids. *Mol. Biosyst.* **7**, 2441-2451 (2011).
4. 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.* **7**, 1984-1993 (2012).
5. Fuchs, S.M. & Raines, R.T. Pathway for Polyarginine Entry into Mammalian Cells. *Biochemistry* **43**, 2438-2444 (2004).
6. Tarcsay, A. & Keseru, G.M. Contributions of Molecular Properties to Drug Promiscuity. *J. Med. Chem.* **56**, 1789-1795 (2013).
7. Charifson, P.S. & Walters, W.P. Acidic and Basic Drugs in Medicinal Chemistry: a Perspective. *J. Med. Chem.* **57**, 9701-9717 (2014).
8. Nguyen, L. et al. Rationally Designed Small Molecules That Target Both the DNA and RNA Causing Myotonic Dystrophy Type 1. *J. Am. Chem. Soc.* **137**, 14180-14189 (2015).
9. Siboni, Ruth B. et al. Actinomycin D Specifically Reduces Expanded CUG Repeat RNA in Myotonic Dystrophy Models. *Cell Reports* **13**, 2386-2394 (2015).
10. 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).
11. Rzuczek, S.G. et al. Features of Modularly Assembled Compounds That Impart Bioactivity Against an RNA Target. *ACS Chem. Biol.* **8**, 2312-2321 (2013).
12. 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).
13. 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).
14. 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).
15. MacArthur, M.W. & Thornton, J.M. Influence of Proline Residues on Protein Conformation. *J. Mol. Biol.* **218**, 397-412 (1991).
16. Verhoork, S.J.M., Killoran, P.M. & Coxon, C.R. Fluorinated Prolines as Conformational Tools and Reporters for Peptide and Protein Chemistry. *Biochemistry* **57**, 6132-6143 (2018).
17. Dugave, C. & Demange, L. Cis-Trans Isomerization of Organic Molecules and Biomolecules: Implications and Applications. *Chem. Rev.* **103**, 2475-2532 (2003).

18. 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).
19. 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).
20. Nakamori, M. et al. Splicing Biomarkers of Disease Severity in Myotonic Dystrophy. *Ann. Neurol.* **74**, 862-872 (2013).
21. Arandel, L. et al. Immortalized Human Myotonic Dystrophy Muscle Cell Lines to Assess Therapeutic Compounds. *Dis. Models Mech.* **10**, 487-497 (2017).
22. Lin, X. et al. Failure of MBNL1-Dependent Post-Natal Splicing Transitions in Myotonic Dystrophy. *Hum. Mol. Genet.* **15**, 2087-2097 (2006).
23. Benhamou, R.I. et al. 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.* **15**, 485–493 (2020).
24. 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.* **13**, 3065-3071 (2018).
25. Boger, D.L.a.C., H. Bleomycin: Synthetic and Mechanistic Studies. *Angewandte Chemie International Edition* **38**, 448-476 (1999).
26. 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.* **276**, 42462-42467 (2001).
27. Chen, C.Z. et al. Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Analytical and bioanalytical chemistry* **402**, 1889-1898 (2012).
28. Rzuczek, S.G. et al. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat Chem Biol* **13**, 188-193 (2017).
29. 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 chemical biology* **7**, 856-862 (2012).
30. Nakamori, M., Taylor, K., Mochizuki, H., Sobczak, K. & Takahashi, M.P. Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy. *Annals of Clinical and Translational Neurology* **3**, 42-54 (2016).
31. Batra, R. et al. Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9. *Cell* **170**, 899-912.e810 (2017).
32. Wheeler, T.M. et al. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature* **488**, 111-115 (2012).
33. Jauvin, D. et al. Targeting DMPK with Antisense Oligonucleotide Improves Muscle Strength in Myotonic Dystrophy Type 1 Mice. *Molecular Therapy. Nucleic Acids* **7**, 465-474 (2017).
34. Siboni, R.B. et al. Actinomycin D Specifically Reduces Expanded CUG Repeat RNA in Myotonic Dystrophy Models. *Cell reports* **13**, 2386-2394 (2015).

35. Lee, J. et al. Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1. *Proceedings of the National Academy of Sciences* **116**, 8709-8714 (2019).
36. Angelbello, A.J. et al. Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proceedings of the National Academy of Sciences* **116**, 7799-7804 (2019).
37. Guan, L. & Disney, M.D. Small-molecule-mediated cleavage of RNA in living cells. *Angew Chem Int Ed Engl* **52**, 1462-1465 (2013).
38. Abraham, A.T., Lin, J.-J., Newton, D.L., Rybak, S. & Hecht, S.M. RNA Cleavage and Inhibition of Protein Synthesis by Bleomycin. *Chemistry & Biology* **10**, 45-52 (2003).
39. Madathil, M.M. et al. Modified bleomycin disaccharides exhibiting improved tumor cell targeting. *Biochemistry* **53**, 6800-6810 (2014).
40. Wu, W., Vanderwall, D.E., Turner, C.J., Kozarich, J.W. & Stubbe, J. Solution Structure of Co·Bleomycin A2 Green Complexed with d(CCAGGCCTGG). *Journal of the American Chemical Society* **118**, 1281-1294 (1996).
41. Goodwin, K.D., Lewis, M.A., Long, E.C. & Georgiadis, M.M. Crystal structure of DNA-bound Co(III)·bleomycin B₂: Insights on intercalation and minor groove binding. *Proceedings of the National Academy of Sciences* **105**, 5052-5056 (2008).
42. Li, Y. & Disney, M.D. Precise Small Molecule Degradation of a Noncoding RNA Identifies Cellular Binding Sites and Modulates an Oncogenic Phenotype. *ACS chemical biology* **13**, 3065-3071 (2018).
43. Schroeder, B.R. et al. The Disaccharide Moiety of Bleomycin Facilitates Uptake by Cancer Cells. *Journal of the American Chemical Society* **136**, 13641-13656 (2014).
44. Wanner, J. et al. A new and improved method for deglycosidation of glycopeptide antibiotics exemplified with vancomycin, ristocetin, and ramoplanin. *Bioorganic & medicinal chemistry letters* **13**, 1169-1173 (2003).
45. Burma, S., Chen, B.P., Murphy, M., Kurimasa, A. & Chen, D.J. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *The Journal of biological chemistry* **276**, 42462-42467 (2001).
46. Arandel, L. et al. Immortalized human myotonic dystrophy muscle cell lines to assess therapeutic compounds. *Disease Models & Mechanisms* **10**, 487-497 (2017).
47. 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).
48. 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. *Human molecular genetics* **13**, 3079-3088 (2004).
49. Gates, D.P., Coonrod, L.A. & Berglund, J.A. Autoregulated splicing of muscleblind-like 1 (MBNL1) Pre-mRNA. *The Journal of biological chemistry* **286**, 34224-34233 (2011).
50. Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nature genetics* **37**, 844-852 (2005).
51. Costales, M.G., Matsumoto, Y., Velagapudi, S.P. & Disney, M.D. Small Molecule Targeted Recruitment of a Nuclease to RNA. *Journal of the American Chemical Society* **140**, 6741-6744 (2018).

52. Costales, M.G., Suresh, B., Vishnu, K. & Disney, M.D. Targeted degradation of a hypoxia-associated non-coding RNA enhances the selectivity of a small molecule interacting with RNA. *Cell chemical biology* **26**, 1180-1186.e1185 (2019).
53. Benhamou, R.I., Angelbello, A.J., Wang, E.T. & Disney, M.D. A Toxic RNA Catalyzes the Cellular Synthesis of Its Own Inhibitor, Shunting It to Endogenous Decay Pathways. *bioRxiv*, 741926 (2019).
54. Silverman, R.H. Viral Encounters with 2′,5′-Oligoadenylate Synthetase and RNase L during the Interferon Antiviral Response. *Journal of Virology* **81**, 12720-12729 (2007).
55. Torrence, P.F. et al. Recruiting the 2-5A system for antisense therapeutics. *Antisense & nucleic acid drug development* **7**, 203-206 (1997).
56. 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).
57. 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).
58. 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).

Journal of Medicinal Chemistry

pubs.acs.org/jmc

Article

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

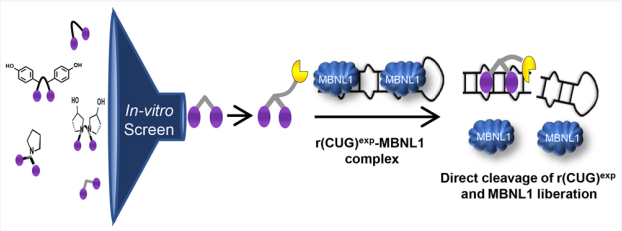
Raphael I. Benhamou, Masahito Abe, Shruti Choudhary, Samantha M. Meyer, Alicia J. Angelbello, and Matthew D. Disney*

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 dystrophia myotonica 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′ untranslated region (UTR) of the dystrophia myotonica 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 tags[9] 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

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

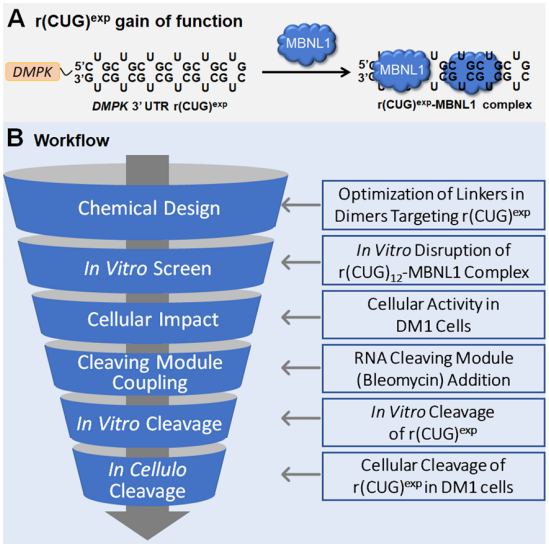


Figure 1. Design of potent dimers targeting r(CUG)^{exp} in DM1. (A) r(CUG)^{exp} in the 3′ UTR of *DMPK* folds into a hairpin displaying a periodic array of internal loops that bind and sequester MBNL1, resulting in pre-mRNA splicing defects and the formation of nuclear foci. (B) Step-by-step methodology for the design of compounds targeting r(CUG)^{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(CUG)^{exp} and cellular potency without increasing the cationic nature of the compound. The optimized binder for r(CUG)^{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(CUG)^{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

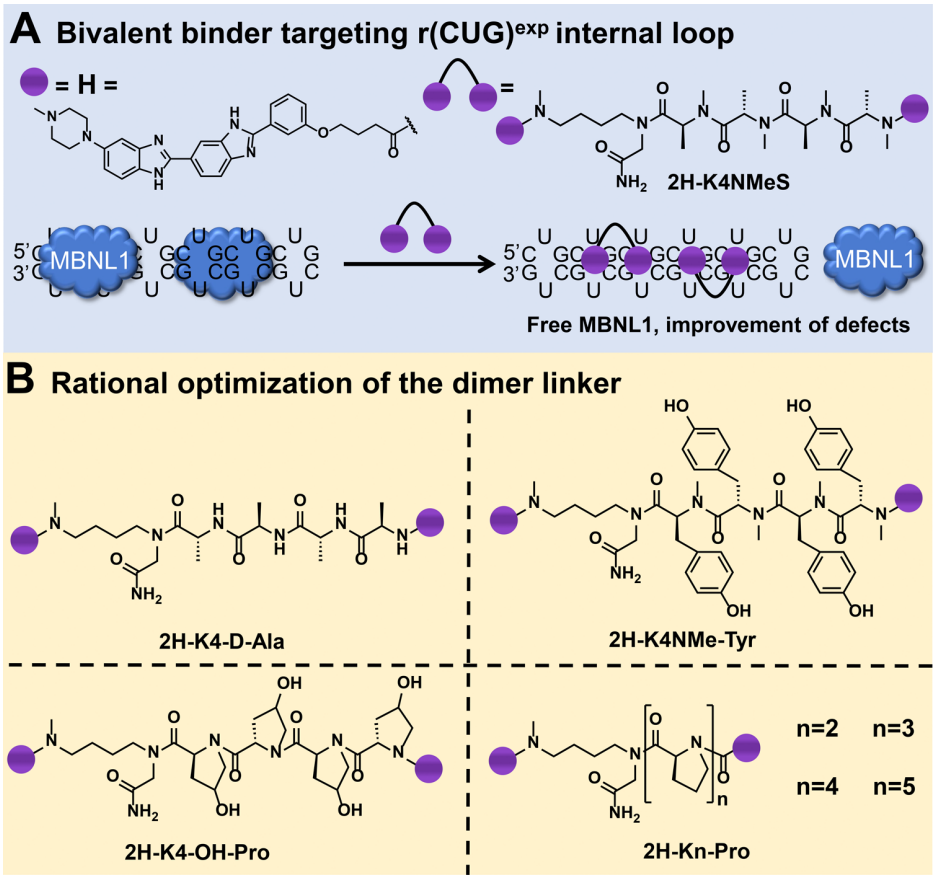


Figure 2. Design of small molecules that selectively target r(CUG)^{exp}. (A) Bivalent small molecules linked together using an *N*-methyl peptide scaffold (**2H-K4NMeS**) with two **H** RNA-binding modules (purple spheres). The dimer binds two r(CUG)^{exp} loops simultaneously and releases MBNL1, improving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

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

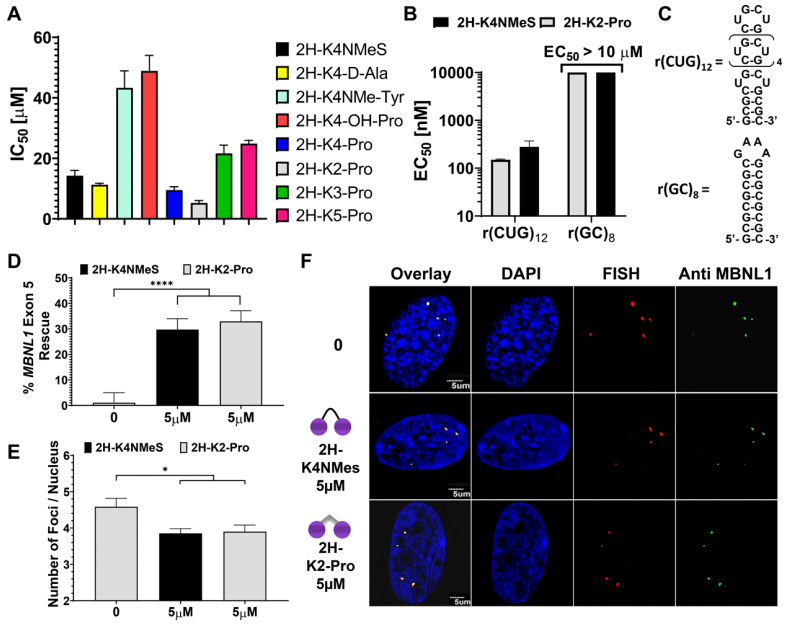


Figure 3. Studying the ability of designer compounds to inhibit the r(CUG)^{exp}–MBNL1 complex *in vitro* and assessment of compound activity in DM1 myotubes. (A) *In vitro* IC_{50}s for disruption of the r(CUG)^{exp}–MBNL1 complex by new dimer derivatives, compared to the parent compound (n = 3 replicates; 2 independent experiments). (B) Binding affinities (EC_{50}s) 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* IC_{50}s 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 interactions, 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 IC_{50}s in the TR-FRET assay (Figure 3A). Notably, **2H-K2-Pro** containing two prolines in the linker was the most potent with an IC_{50} of 5.3 ± 0.8 μM, 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 log*P*, 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 EC_{50} of 150 ± 5 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** (EC_{50} = 280 ± 90 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

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827–7839

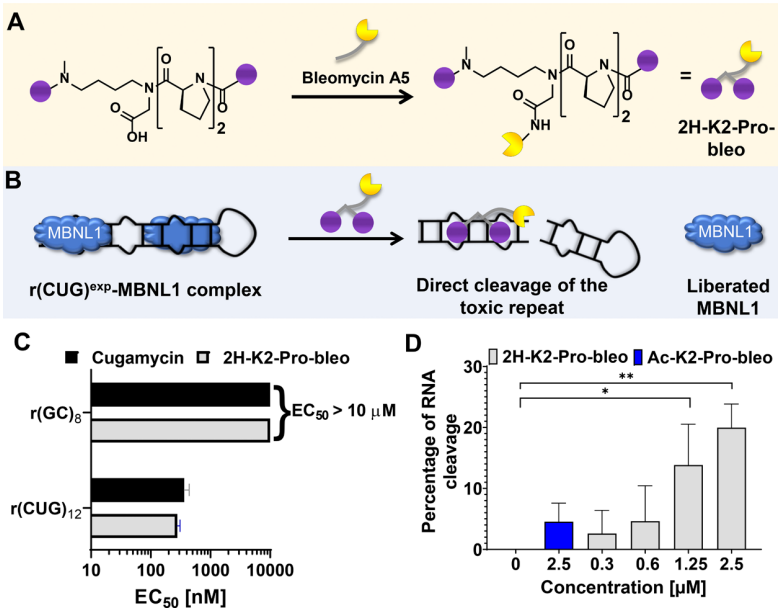


Figure 4. Design and *in vitro* evaluation of small molecules that cleave r(CUG)^{exp}. (A) The binder **2H-K2-Pro** was conjugated to bleomycin to afford the cleaver **2H-K2-Pro-bleo**. (B) **2H-K2-Pro-bleo** binds to r(CUG)^{exp}, cleaves the toxic RNA repeat, and displaces MBNL1, thereby improving 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 myotubes[34] 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 μM 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 μM 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 μM 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 binding[38]) 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).

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

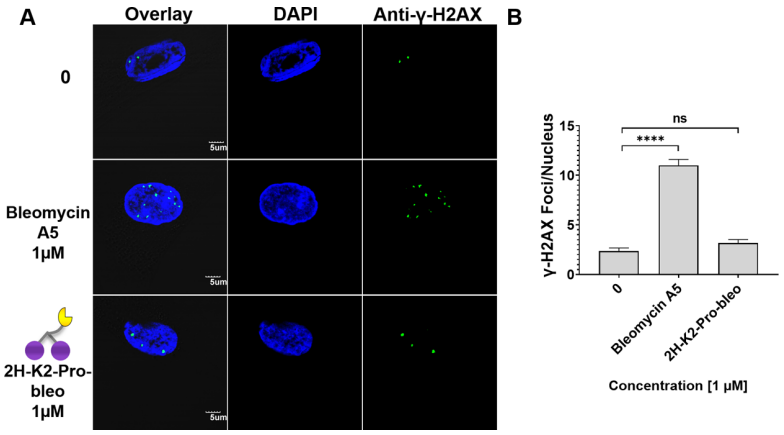


Figure 5. Effects of small molecules on the DNA damage response pathway. (A) Representative images from γ-H2AX immunofluorescence to assess DNA damage in DM1 myotubes upon treatment with **2H-K2-Pro-bleo** and Bleomycin A5. (B) Quantification of the number of γ-H2AX foci/nucleus. Error bars represent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

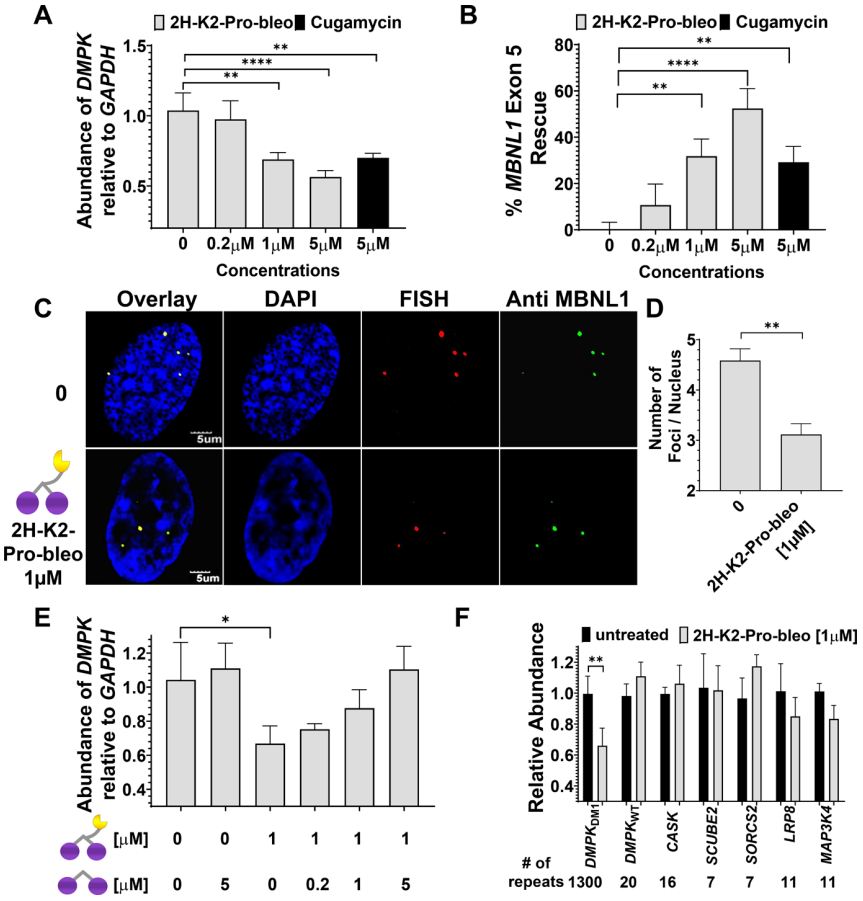


Figure 6. Biological activity of **2H-K2-Pro-bleo** in DM1 patient-derived myotubes. (A) Effect of **2H-K2-Pro-bleo** on r(CUG)^{exp}-containing *DMPK* levels in DM1 myotubes compared to Cugamycin, as determined by RT-qPCR. Error bars represent SD, n = 3 biological replicates. *P < 0.05, **P < 0.01, ****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 μM) 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).

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

***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 μM (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-assiociated 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 ± 4% at 5 μM, whereas only an ∼30 ± 3% decrease was observed with Cugamycin at 5 μM (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 ± 8% at 5 μM and 30 ± 7% at 1 μM (Figure 6B), the latter of which is similar to 5 μM **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 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 μM) 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 μM **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 μM) inhibited *DMPK* cleavage more efficiently than **2H-K4NMeS** (EC_{25} = 1 μM) 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)-1*H*,3′*H*-[2,5′-bibenzo[*d*]imidazol]-2′-yl)phenoxy)butanoic acid (**6**) (Ht-CO_2H), intermediate (**10**), and **2H-K4NMeS** were synthesized as

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

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 μm 19 × 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 μm 4.6 × 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 ≥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 × 4.6 mm, 2.7 μm).

Synthesis of 2. A mixture of *tert*-butyl (4-(methylamino)butyl)-carbamate (3.00 g, 14.8 mmol), NEt_3 (4.13 mL, 29.7 mmol), and AllocCl (1.89 mL, 17.8 mmol) in CH_2Cl_2 (60 mL) was stirred at room temperature overnight. After completion of the reaction, the reaction was diluted with CH_2Cl_2 (100 mL), extracted with brine, and washed with 5% HCl aq (v/v, 15 mL) twice and aqueous $NaHCO_3$ (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:CH_2Cl_2 was added, and the reaction was stirred at room temperature overnight. The reaction mixture was concentrated, and then $NaHCO_3$ (aq; 15 mL) was added. The mixture was then extracted with CH_2Cl_2 (5 × 15 mL) to give **2** (759 mg, 27% yield over 2 steps). This compound was used for the next reaction without further purification. ^{1}H NMR (600 MHz, $CDCl_3$) 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) ^{13}C NMR (150 MHz, $CDCl_3$) 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 $[C_9H_{19}N_2O_2]^+$, 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 × 20 min). The resin was washed with DMF (3 × 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 CH_2Cl_2 (3 × 8 mL) and DMF (3 × 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 CH_2Cl_2 (3 × 8 mL) and DMF (3 × 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 CH_2Cl_2 (3 × 8 mL) and DMF (3 × 8 mL) to afford intermediate **3**.

Synthesis of 2H-K4-D-Ala. Resin **3** (20 mg, 10 μmol) was swollen in DMF (500 μL) for 10 min, and DMF was removed. To the resin, a pre-mixed solution of Fmoc-D-Ala-OH (27.0 mg, 50 μmol), HATU (19.0 mg, 50 μmol), HOAt (6.8 mg, 50 μmol), and DIPEA (16.5 μL, 100 μmol) in DMF (500 μL) was added, and the reaction was shaken at room temperature for 30 min. The resin was then washed with DMF (5 × 1 mL). To the resin, 20% piperidine in DMF (500 μL) was added, and the resin was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 × 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 μmol), HATU (19.0 mg, 50 μmol), HOAt (6.8 mg, 50 μmol), and DIPEA (16.5 μL, 100 μmol) was added to the resin, which was then shaken at room temperature for 30 min. The resin was washed with DMF (5 × 1 mL). To this resin, $Pd(PPh_3)_4$ (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 step was repeated once more. After that, the resin was washed with CH_2Cl_2 (5 × 1 mL, 30 s), 0.5% DIPEA (v/v) in DMF (2 × 1 mL, 30 s) and 0.5% sodium diethyl dithiocarbamate (w/v) in DMF (2 × 1 mL, 30 s), and DMF (5 × 1 mL, 30 s). To the resin, 20% piperidine in DMF (500 μL) was added, and the reaction was shaken at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 × 1 mL). These two steps were repeated. To the resin, 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 was added, and it was shaken at 45 °C for 40 min. The resin was washed with DMF (5 × 1 mL). To this resin, 1 mL of TFA/H_2O (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 Et_2O, and the resulting precipitate was collected. The precipitate was washed with Et_2O (5× volume), and the supernatant was removed by decantation. This crude product was purified by HPLC (41−81% MeOH+0.1% TFA vs H_2O + 0.1% TFA, over 20 min) to give **2H-K4-D-Ala** (372 nmol, 1.86% yield); HR-MS (MALDI): Calculated for $[C_{77}H_{94}N_{19}O_9]^+$, 1428.7476; found, 1428.7456.

Synthesis of 2H-K4NMe-Tyr. Resin **3** (20 mg, 10 μmol) was swollen in DMF (500 μL) 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 μmol), DIC (7.7 μL, 50 μmol), and oxyma (7.1 mg, 50 μmol) in DMF (500 μL) was added, and the resin was shaken at room temperature for 2 h. After washing the resin with DMF (5 × 1 mL), 20% piperidine in DMF (500 μL) was added, followed by shaking at room temperature for 10 min. This step was repeated. The resin was washed with DMF (5 × 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 μmol), DIC (7.7 μL, 50 μmol), and oxyma (7.1 mg, 50 μmol) in DMF (500 μL) was added, and it was shaken at room temperature for 2 h. The resin was washed with DMF (5 × 1 mL). To this resin, $Pd(PPh_3)_4$ (5.78 mg, 5.0 μmol) in CH_2Cl_2 (500 μL), $PhSiH_3$ (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 that, the resin was washed with CH_2Cl_2 (5 × 1 mL, 30 s), 0.5% DIPEA (v/v) in DMF (2 × 1 mL, 30 s) and 0.5% sodium diethyl dithiocarbamate (m/v) in DMF (2 × 1 mL, 30 s), and DMF (5 × 1 mL, 30 s). To the resin, 20% piperidine/DMF (500 μL) was added, and it was shaken at room temperature for 10 min twice. The resin was washed with DMF (5 × 1 mL). To the resin, 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 was added, and it was shaken at room temperature overnight. The resin was washed with DMF (5 × 1 mL) followed by addition of 1 mL of 1:1 TFA: H_2O. The mixture was shaken at room temperature for 1 h. After that, the solution was collected and concentrated. To the mixture was added 10× the volume of Et_2O, and the resultant precipitate was collected. The precipitate was washed with Et_2O (5× volume), and the supernatant was removed by decantation. This crude product was purified by HPLC (50−100% MeOH + 0.1% TFA/H_2O + 0.1% TFA, over 20 min) to give **2H-4N-Me-Tyr** (60 nmol, 0.60% yield) HR-MS (ESI): Calculated for $[C_{105}H_{119}N_{19}O_{13}]^{2+}$, 926.9612; found, 926.9590.

Synthesis of 2H-K4-OH-Pro. Resin **3** (20 mg, 10 μmol) was swollen in DMF (500 μL) 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 μmol), DIC (7.7 μL, 50 μmol), and oxyma (7.1 mg, 50 μmol) in DMF (500 μL) was added followed by shaking at room temperature for 2 h. The resin was washed with DMF (5 × 1 mL), and then 20% piperidine in DMF (500 μL) was added. The resin was shaken at room temperature for 10 min and then washed with DMF (5 × 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 μmol), DIC (7.7 μL, 50 μmol), and oxyma (7.1 mg, 50 μmol) was added, and it was shaken at room temperature for 2 h. The resin was washed with DMF (5 × 1 mL), and then $Pd(PPh_3)_4$ (5.78 mg, 5.0 μmol) in CH_2Cl_2 (500 μL), $PhSiH_3$ (24.7 μL, 200 μmol), and CH_2Cl_2 (1 mL) were added. The mixture was shaken at room temperature for 40 min.

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

This deprotection step was repeated once more. After that, the resin was washed with CH_2Cl_2 (5 × 1 mL, 30 s each), 0.5% DIPEA (v/v) in DMF (2 × 1 mL, 30 s each) and 0.5% sodium diethyl dithiocarbamate (m/v) in DMF (2 × 1 mL, 30 s each), and DMF (5 × 1 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 × 1 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 × 1 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× the volume of Et_2O, and the resultant precipitate was collected. The precipitate was washed with 5× 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 $[C_{85}H_{103}N_{19}O_{13}]^{2+}$, 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 × 5 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× 10 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× the volume of Et_2O. The precipitate was washed with Et_2O (5× 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 °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 $[C_{75}H_{88}N_{17}O_7]^+$, 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 × 1 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 × 1 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 × 1 mL). To this resin, $Pd(PPh_3)_4$ (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 × 1 mL, 30 s each), 0.5% DIPEA (v/v) in DMF (2 × 1 mL, 30 s each) and 0.5% sodium diethyl dithiocarbamate (w/v) in DMF (2 × 1 mL, 30 s each), and DMF (5 × 1 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 × 1 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 °C for 40 min. The resin was washed with DMF (5 × 1 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× the volume of Et_2O, and the resultant precipitate was collected. The precipitate was washed with Et_2O (5× 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 $[C_{80}H_{95}N_{18}O_8]^+$, 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 × 5 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× 10 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× the volume of Et_2O. The resulting precipitate was collected and washed with 5× 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 °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 $[C_{85}H_{102}N_{19}O_9]^+$, 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 × 1 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 × 1 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× the volume of Et_2O. The resultant precipitate was collected and washed with Et_2O (5× 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 °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 $[C_{85}H_{102}N_{19}O_9]^+$, 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 NsCl (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-

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

ture of this material in 1:1 TFA/CH_2Cl_2 (98 mL) was stirred at room temperature for 3 h. The mixture was concentrated followed by addition of aq $NaHCO_3$ (15 mL) and extraction with CH_2Cl_2 (5 × 15 mL) to give **13** (1.521 g, 54%, 2 steps). This was used for the next reaction without further purification. ^{1}H NMR (400 MHz, $CDCl_3$) 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). ^{13}C NMR (100 MHz, $CDCl_3$) 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 $[C_{11}H_{18}N_3O_4S]^+$, 288.1013; found, 288.1015.

Synthesis of 2H-K2-Pro-bleo. The 2-chlorotrityl resin (**12**) (500 mg, 730 μmol) was washed with CH_2Cl_2, which was then removed. To this resin, CH_2Cl_2 (6.6 mL) and HCl in dioxane (4 M, 3.0 mL) were added. After shaking for 30 min, the resin was washed with CH_2Cl_2 (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 CH_2Cl_2 (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 CH_2Cl_2 (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 CH_2Cl_2 (5× 10 mL) and DMF (5× 10 mL). To the resin, a pre-mixed solution of (Ht-CO_2H (**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 CH_2Cl_2 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% CH_3CN + 0.1% TFA/$H_2$0 + 0.1% TFA, over 60 min) to afford **17** (83.0 mg, ca. 80% purity).

Compound **17** (13.0 mg) was washed with Et_2O (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 H_2O + 0.1% TFA, over 90 min; and step 4: after concentrating the product in step 3, 8−58% MeOH + 0.1% TFA vs H_2O + 0.1% TFA, over 60 min] to give **2H-K2-Pro-bleo** (1.02 μmol, 0.89% yield); HR-MS (MALDI): Calculated for $[C_{132}H_{174}N_{35}O_{28}S_2]^+$, 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 CH_2Cl_2 followed by treatment with a mixture of 6 mL of CH_2Cl_2 and 3.0 mL of 4 M HCl in dioxane. After shaking for 30 min at room temperature, the resin was washed with CH_2Cl_2 (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 mmol) in CH_2Cl_2 (3.65 mL) were added. The mixture was shaken at room temperature for 3 h, and then the resin was washed with CH_2Cl_2 (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 CH_2Cl_2 (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 CH_2Cl_2 (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 CH_2Cl_2 and DMF, DBU (546 μL, 3.65 mmol) and $HO(CH_2)_2SH$ (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 CH_2Cl_2 (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 CH_2Cl_2 (5 × 10 mL) and DMF (5 × 10 mL), and then 10 mL of 30% TFA in CH_2Cl_2 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 $[C_{78}H_{122}N_{23}O_{26}S_2]^+$, 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* IC_{50} 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 $MgCl_2$, 2 mM $CaCl_2$, 5 mM DTT, 0.1% BSA, and 0.5% Tween- 20). Next, MBNL1-His_6 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-His_6-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 IC_{50} 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 IC_{50} is the concentration of a compound where half of the protein is displaced by a compound. n = 3 replicates; 2 independent experiments.

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

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 Na_2HPO_4, 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-^{32}P] 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′-^{32}P-$r(CUG)_{10}$ radiolabeled RNA was dissolved in 40 μL of water. A 4 μL aliquot of this solution was diluted in 150 μL of 5 mM NaH_2PO_4 (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 $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ in 5 mM NaH_2PO_4, pH 7.4, was then added. The solutions were incubated at 37 °C and supplemented with additional equimolar aliquots of $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$ in 5 mM NaH_2PO_4, 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 $NaHCO_3$, 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 aTyphoon 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% Glutagro (Corning) at 37 °C, and 5% CO_2. 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 $2^{(-\Delta\Delta Ct)}$ 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.

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

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

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 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min, and a final extension at 72 °C for 5 min. Products were separated on a 2% agarose gel stained with ethidium bromide (110 V for 1 hour in 1× 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× SCC for 30 min at 37 °C, cells were incubated with a 1:500 dilution of anti-γ-H2AX (Abcam) at 37 °C for 1 h. The myotubes were then washed three times with 1× DPBS containing 0.1% Triton X-100 (v/v) for 5 min at 37 °C, followed by incubation with a 1:200 dilution of a goat anti-mouse IgG-DyLight 488 conjugate (Thermo Scientific) at 37 °C for 1 h. After washing three times with 1× DPBS containing 0.1% Triton X-100 (v/v) and twice with 1× DPBS for 5 min at 37 °C, nuclei were stained with DAPI (1 μg/mL). Cells were imaged in 1× DPBS using an Olympus Fluoview 1000 confocal microscope at 100× 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 × DPBS and imaged in 100 μL of Gibco FluoroBrite DMEM using an Olympus Fluoview 1000 confocal microscope at 100× 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× 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*, dystrophia 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)methyliumyl]-1*H*-1,2,3-triazolo[4,5-*b*]pyridine-3-oxide hexafluorophosphate;

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

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 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.; Huguet, 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-Umnuay, 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.; 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*, 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.; Keserű, 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.; 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*, 1889−1898.

(28) MacArthur, M. W.; Thornton, J. M. Influence of Proline Residues on Protein Conformation. *J. Mol. Biol.* **1991**, *218*, 397−412.

(29) Verhoork, 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.

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

(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.

https://dx.doi.org/10.1021/acs.jmedchem.0c00558
J. Med. Chem. 2020, 63, 7827−7839

pubs.acs.org/acschemicalbiology

Letters

Precise Targeted Cleavage of a r(CUG) Repeat Expansion in Cells by Using a Small-Molecule−Deglycobleomycin Conjugate

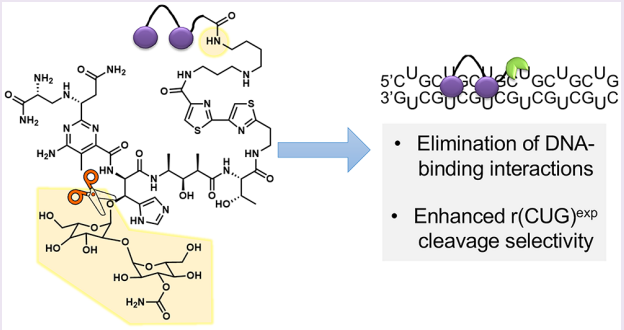
Alicia J. Angelbello, Mary E. DeFeo, Christopher M. Glinkerman, Dale L. Boger, and Matthew D. Disney*

Cite This: *ACS Chem. Biol.* 2020, 15, 849−855

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: RNA repeat expansions cause more than 30 neurological and neuromuscular diseases with no known cures. Since repeat expansions operate via diverse pathomechanisms, one potential therapeutic strategy is to rid them from disease-affected cells, using bifunctional small molecules that cleave the aberrant RNA. Such an approach has been previously implemented for the RNA repeat that causes myotonic dystrophy type 1 [DM1, r(CUG)^{exp}] with Cugamycin, which is a small molecule that selectively binds r(CUG)^{exp} conjugated to a bleomycin A5 cleaving module. Herein, we demonstrate that, by replacing bleomycin A5 with deglycobleomycin, an analogue in which the carbohydrate domain of bleomycin A5 is removed, the selectivity of the resulting small-molecule conjugate (DeglycoCugamycin) was enhanced, while maintaining potent and allele-selective cleavage of r(CUG)^{exp} and rescue of DM1-associated defects. In particular, DeglycoCugamycin did not induce the DNA damage that is observed with high concentrations (25 μM) of Cugamycin, while selectively cleaving the disease-causing allele and improving DM1 defects at 1 μM.



Developing small-molecule chemical probes that modulate RNA function is increasingly important, because of the numerous mechanisms by which RNA can cause disease. One way to target RNAs is the recognition of unstructured regions by antisense (ASOs) and other oligonucleotide-base modalities that bind via base pairing interactions. However, many RNAs have structured regions that directly influence biological function. Small molecules can target these biologically important structures, by matching the RNA's three-dimensional binding pocket, in terms of size, shape, and complementarity in the display of functional groups and/or surfaces.[1,2]

One class of disease-causing RNAs that form stable structures is repeat expansions that cause more than 30 microsatellite disorders, including Huntington's disease[3] [HD, r(CAG)^{exp}, where the repeating nucleotides are given in parentheses and "exp" denotes expansion] and myotonic dystrophy types 1[4] [DM1, r(CUG)^{exp}] and 2[5] [DM2, r(CCUG)^{exp}]. DM1 is the most common form of adult onset muscular dystrophy, which presently has no cure. This neuromuscular disorder is caused by an expanded CTG repeat, ranging in size from 75 to thousands, harbored in the 3′ untranslated region (UTR) of the dystrophia myotonica protein kinase (*DMPK*) gene.[4] When transcribed into RNA, r(CUG)^{exp} forms a hairpin structure with repeating 1 × 1 U/U internal loops. These loops provide high-affinity binding sites for RNA-binding proteins such as muscleblind-like 1 (MBNL1), which are sequestered in nuclear foci (Figure 1A).[6] Thus, r(CUG)^{exp} operates by a gain-of-function mechanism. MBNL1 regulates the alternative splicing of a subset of pre-mRNAs, and its sequestration by r(CUG)^{exp} in nuclear foci[6] results in pre-mRNA splicing defects that contribute to the phenotypes found in DM1 (Figure 1B).[7,8]

One approach to alleviate DM-associated defects is to utilize small molecules that recognize the structure of r(CUG)^{exp}, thereby liberating bound MBNL1 or preventing its binding.[9−11] Alternatively, the expression of r(CUG)^{exp} has been reduced or eliminated by using RNA targeted-Cas9 editing of r(CUG)^{exp},[12] ASOs,[13,14] DNA-binding small molecules that inhibit transcription,[15,16] and small molecules that bind and directly cleave r(CUG)^{exp}.[9,17,18] The latter approach (Figure 1C) has been accomplished with Cugamycin, which is a small molecule that selectively binds the r(CUG)^{exp} structure conjugated to the natural product bleomycin A5 (BLM) (see Figure 2, as well as Figure S1 in the Supporting Information).

Received: January 16, 2020
Accepted: March 10, 2020
Published: March 18, 2020

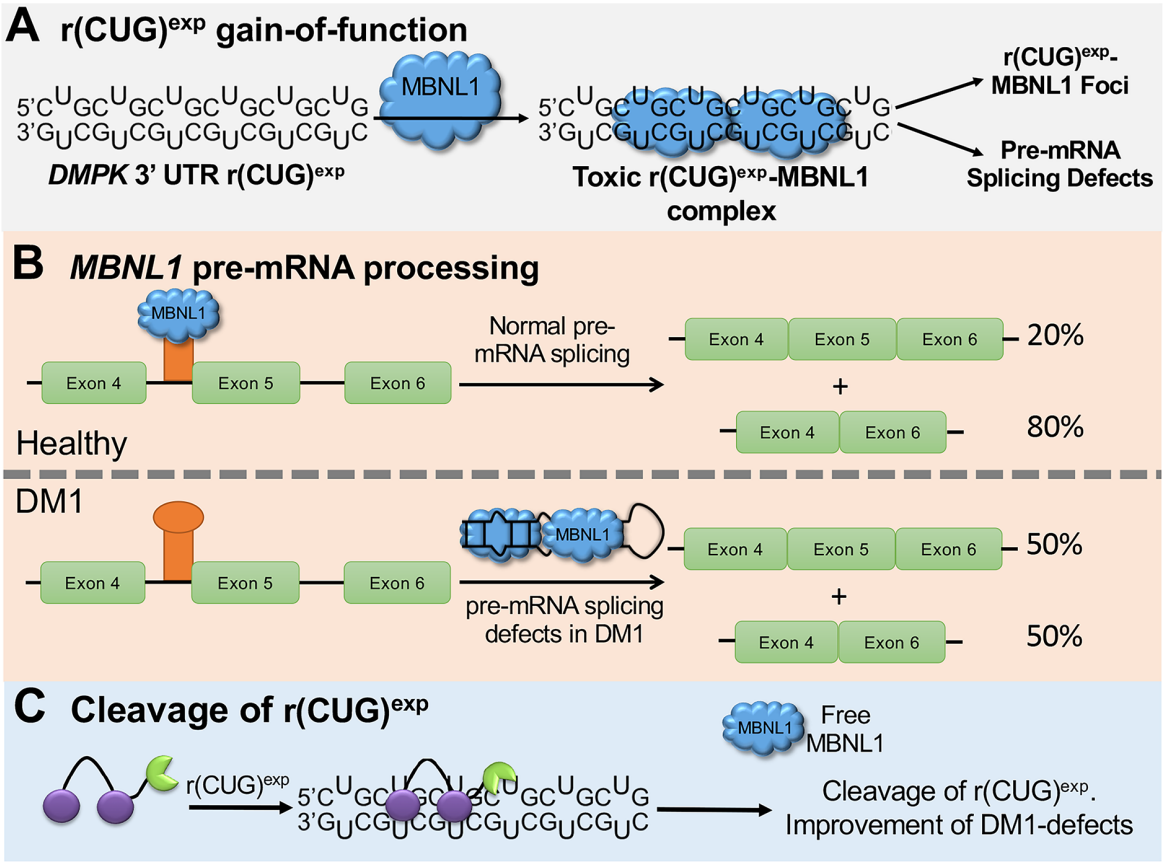


Figure 1. Small-molecule cleavage of r(CUG)^{exp}. (A) DM1 is caused by r(CUG)^{exp}, located in the 3′ UTR of the *DMPK* gene, which forms a structure with repeating 1 × 1 U/U internal loops. The loops bind and sequester MBNL1, resulting in nuclear foci and pre-mRNA splicing defects. (B) MBNL1 protein regulates the splicing of its own pre-mRNA. When MBNL1 is sequestered by r(CUG)^{exp}, its exon 5 is included too frequently. (C) Scheme of small-molecule cleavage of r(CUG)^{exp}, resulting in improvement of DM1-associated defects.

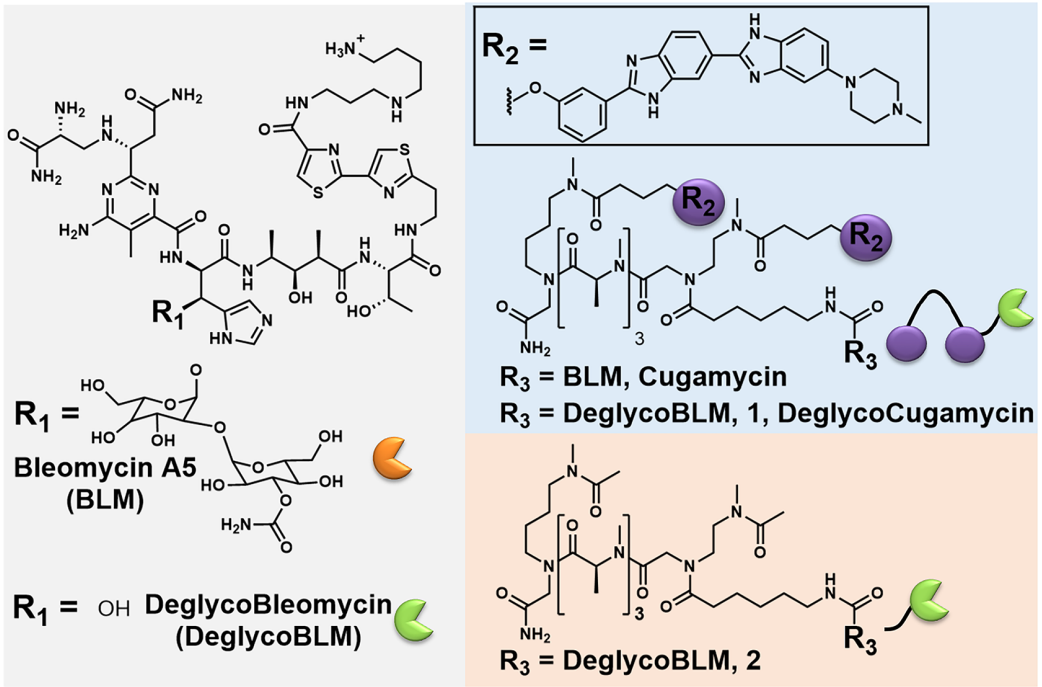


Figure 2. Chemical structures of bleomycin A5 (BLM), deglycobleomycin (DeglycoBLM), Cugamycin, **1**, and **2**.

https://dx.doi.org/10.1021/acschembio.0c00036
ACS Chem. Biol. 2020, 15, 849−855

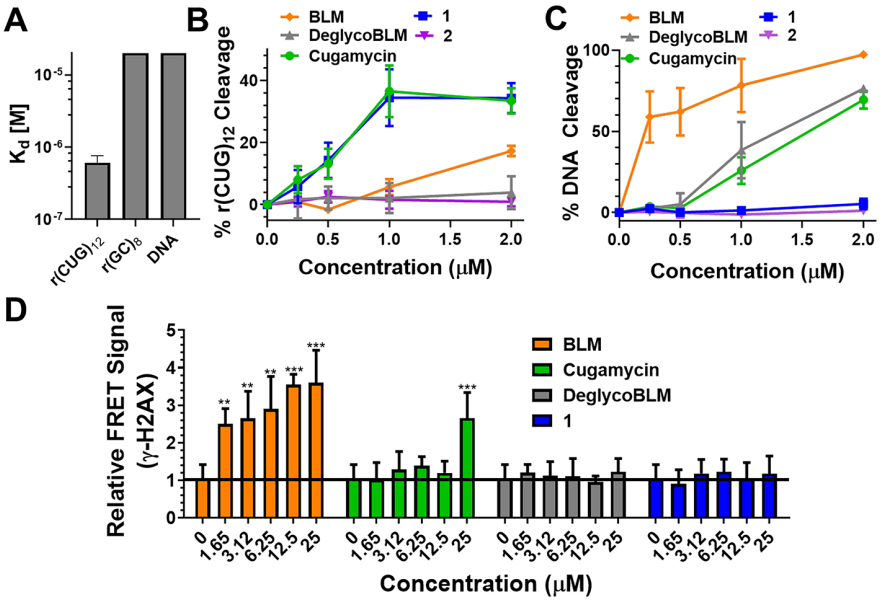


Figure 3. Cleaving capacity and selectivity of small-molecule cleavers. (A) Binding affinity of **1** for $r(CUG)_{12}$ (K_d = 610 ± 150 nM), $r(GC)_8$ (K_d > 20 μM), and DNA (K_d > 20 μM); n = 3. (B) Quantification of cleavage of $r(CUG)_{10}$ by **1**, **2**, Cugamycin, BLM, and DeglycoBLM; n = 3. (C) Quantification of cleavage of DNA by **1**, **2**, Cugamycin, BLM, and DeglycoBLM; n = 3. (D) Effects of **1**, Cugamycin, BLM, and DeglycoBLM on γ-H2AX, a marker of DNA damage, in C2C12 cells. n = 8, (**) P < 0.01, (***) P < 0.001, as determined by comparison to untreated cells by a one-way analysis of variance (ANOVA). Error bars indicate the standard deviation (SD) for all panels.

Indeed, Cugamycin broadly improved disease associated defects with no off-target effects in a mouse model of DM1.[17]

BLM is an anticancer natural product that cleaves DNA and RNA through H atom abstraction and the production of a radical species by the metal binding core.[19,20] Extensive structure−activity relationship (SAR) evaluations of BLM derivatives[19,21] and structural data for DNA-bound BLM[22,23] have revealed structural components that are essential for metal coordination, oxygen activation, DNA binding, and subsequent cleavage. This information has been used to guide attachment of RNA-binding small molecules at the C-terminal amine of BLM, eliminating a charge critical to DNA binding and producing BLM-conjugated compounds that specifically cleave a target RNA.[17,24] SAR studies of BLM can guide the selection of analogues to further enhance RNA selectivity by eliminating DNA-binding interactions.[19] One such analogue is deglycobleomycin (DeglycoBLM; see Figure 2, as well as Figure S1), in which the disaccharide moiety of BLM is removed. The carbohydrate domain can contribute to DNA binding affinity by participating in hydrogen bonding interactions with the DNA backbone, and DeglycoBLM cleaves DNA between 2- and 5-fold less efficiently than BLM.[19,23] This disaccharide also contributes to the cellular permeability of BLM.[25] Collectively, the attachment of DeglycoBLM to small molecules targeting $r(CUG)^{exp}$ could further reduce its affinity for DNA to enhance RNA selectivity in cells, provided the compound retains cellular permeability. The examination of such features is the subject of this report.

DeglycoBLM was synthesized via HF-pyridine cleavage of the carbohydrate of BLM[26] and conjugated to a dimeric compound that recognizes $r(CUG)^{exp}$ (2H−K4NMeS, **3**; see Figure S1)[9] to afford compound **1** (DeglycoCugamycin; see Figure 2, as well as Figure S1). A control compound that does not contain the RNA-binding modules in **1** and, thus, has no affinity for the RNA target, was also synthesized (compound **2**; see Figure 2, as well as Figure S1). To assess the molecular recognition of **1**, its affinity for $r(CUG)_{12}$, $r(GC)_8$, and DNA was measured in the absence of Fe(II). Compound **1** only bound avidly to $r(CUG)_{12}$ (K_d = 610 ± 150 nM) (see Figure 3A, as well as Figure S2 in the Supporting Information), which is comparable to the affinity of Cugamycin (K_d = 365 ± 75 nM).[9] Thus, removal of the carbohydrate domain does not affect the ability to bind $r(CUG)^{exp}$ *in vitro*.

Next, the ability of **1** to cleave $r(CUG)_{10}$ and DNA was assessed *in vitro*. Cugamycin and **1** cleaved $r(CUG)_{10}$ to a similar extent at the same concentrations (~35% cleavage at 1 μM), while BLM only cleaved $r(CUG)_{10}$ by 15% at 2 μM (see Figure 3B, as well as Figure S3 in the Supporting Information). In contrast, DeglycoBLM and **2** (lacks RNA-binding modules) were unable to cleave $r(CUG)_{10}$ at the concentrations tested (up to 2 μM; see Figure 3B, as well as Figure S3), as expected, since DeglycoBLM alone is 5-fold less efficient at cleaving nucleic acids than BLM.[19] Thus, functional RNA cleavage by **1** is not affected through removal of the disaccharide. The selectivity of the observed cleavage was assessed by measuring DNA cleavage (see Figure 3C, as well as Figure S4 in the Supporting Information). While BLM efficiently cleaved DNA *in vitro* with >50% cleavage observed at all concentrations (from 250 nM to 2 μM), DeglycoBLM cleaved DNA ~5-fold less efficiently, with >50% cleavage only observed at 2 μM (Figure 3C), consistent with previous studies.[19] We previously showed that Cugamycin does not cleave DNA when $r(CUG)_{12}$ is present[17] and, thus, is selective for cleaving the RNA target. However, when incubated in the absence of $r(CUG)_{12}$, Cugamycin cleaved DNA at concentrations >500 nM (Figure 3C). In contrast, **1** and **2** did not significantly cleave DNA at any of the concentrations tested (from 250 nM to 2 μM; see Figure 3C). Thus, by eliminating two key DNA-binding interactions through removal of the disaccharide and attachment of the r(CUG)-binding compound at the C-terminal

https://dx.doi.org/10.1021/acschembio.0c00036
ACS Chem. Biol. 2020, 15, 849−855

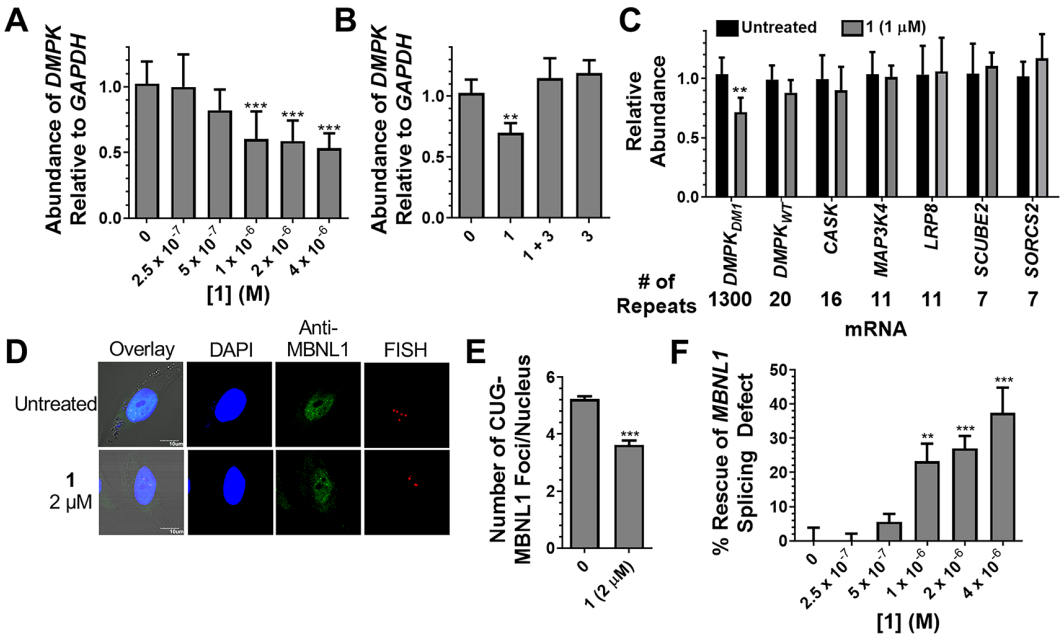


Figure 4. Activity of **1** in DM1 myotubes. (A) Cleavage of r(CUG)^{exp}-containing *DMPK* by **1** as determined by RT-qPCR. $n = 6$, (***) $P < 0.001$, as compared to untreated cells; determined by a one-way ANOVA. (B) Competitive cleavage experiment between **1** (1 μM) and **3** (5 μM) in which **3** prevents the cleavage of *DMPK*. $n = 3$, (**) $P < 0.01$, as compared to untreated cells; determined by a one-way ANOVA. (C) Effect of **1** on r(CUG)$_n$-containing transcripts. $n = 3$, (**) $P < 0.01$, as compared to untreated cells; determined by a Student's *t*-test. (D) Representative images of r(CUG)^{exp}–MBNL1 foci imaged by RNA fluorescence *in situ* hybridization (FISH) and anti-MBNL1 immunostaining. (E) Quantification of nuclear foci. $n = 3$, 40 nuclei quantified/replicate, (***) $P < 0.001$, as compared to untreated cells; determined by a Student's *t*-test. (F) Improvement of the *MBNL1* exon 5 splicing defect in DM1 myotubes upon treatment with **1**. $n = 6$, (**) $P < 0.01$, (***) $P < 0.001$, as compared to untreated cells; determined by a one-way ANOVA. Error bars indicate the SD for all panels.

amine, DNA cleavage is further ablated and selectivity for r(CUG)^{exp} is enhanced.

To study potential off-target binding to DNA in cells, we measured the amount of phosphorylated histone H2A variant H2AX (γ-H2AX), which forms foci in response to DNA double strand breaks, induced by compound treatment in the rapidly growing mouse myoblast cell line C2C12 and in DM1 patient-derived myotubes. In C2C12 cells, we used a fluorescence resonance energy transfer (FRET) assay to quantify γ-H2AX foci after treating with the compound of interest for 24 h. BLM caused a significant increase in γ-H2AX at all concentrations tested (1.65–25 μM), as expected from previous studies[27] (Figure 3D). In contrast, Cugamycin only induced DNA damage at 25 μM, which is a concentration that is ~10-fold higher than its bioactive concentration in DM1 myotubes,[17] while no increase in γ-H2AX foci was observed for **1** or DeglycoBLM upon treatment with up to 25 μM compound (Figure 3D). Importantly, and consistent with these studies in C2C12 cells, neither **1** nor DeglycoBLM induced DNA damage in DM1 patient-derived myotubes, as determined from immunostaining and imaging by fluorescence microscopy (see Figure S5 in the Supporting Information). [Note: the signal:noise observed in the FRET assay described above for C2C12 cells was not sufficient for quantification in DM1 myotubes.] Thus, **1** further diminished off-target DNA cleavage in cells, compared to Cugamycin, in agreement with *in vitro* DNA cleavage analysis.

To probe if the difference in DNA damage in cells is due to changes in cellular uptake, as the disaccharide has previously been implicated in cell permeability,[25] the concentration of Cugamycin, **1**, and the dimer from which they are derived (**3**;[9] see Figure S1) taken up into DM1 myotubes was determined by measuring the fluorescence of the RNA-binding modules after washing and lysing treated cells. Cugamycin and **1** had similar cell permeabilities, and both compounds were only ~3-fold less permeable than **3** (see Figure S6 in the Supporting Information). To confirm these results, permeability and localization were studied by using live-cell fluorescence microscopy. Both Cugamycin and **1** localized in the nucleus where r(CUG)^{exp} is sequestered in foci to a similar extent (see Figure S6). Thus, although the carbohydrate domain has been shown to affect the permeability of DeglycoBLM itself and may account for its lack of DNA damage in cells (Figure 3D), the disaccharide did not affect permeability of conjugate compounds, as determined by comparing Cugamcyin and **1**.

Since **1** ablated DNA damage observed for Cugamycin without reducing cell permeability, its ability to cleave r(CUG)^{exp} and improve DM1-associated defects in cells was measured. In DM1 patient-derived myotubes,[28] **1** cleaved ~30% of r(CUG)^{exp}-containing *DMPK* at low micromolar concentrations (Figure 4A), which is comparable to the cleaving activity of Cugamycin.[17] Importantly, **2**, which lacks RNA-binding modules, did not affect *DMPK* levels (see Figure S7A in the Supporting Information). To demonstrate that the reduction in *DMPK* levels was due to direct cleavage of the RNA, a competition experiment was performed in which cells were co-treated with **1** and **3**; **3** binds r(CUG)^{exp} but does not affect *DMPK* mRNA levels (Figure 4B). Indeed, upon co-treatment, cleavage by **1** was inhibited by **3**, and *DMPK* levels were restored to levels similar to untreated samples or samples treated with **3** alone (Figure 4B). Notably, **1** was selective for cleaving r(CUG)^{exp}, as *DMPK* levels were not affected in wild-type cells expressing r(CUG)$_{20}$ (see Figure 4C, as well as Figure S8A in the Supporting Information), nor were mRNAs

https://dx.doi.org/10.1021/acschembio.0c00036
ACS Chem. Biol. 2020, 15, 849–855

containing short nonpathogenic r(CUG) repeats (Figure 4C). We have previously shown that this selectivity is due to structural differences in transcripts containing short r(CUG) repeats versus r(CUG)^{exp}, as the small molecule recognizes the structure formed by the repeat expansion.[17] Indeed, compounds that recognize the structure of r(CUG)^{exp} can be selective for the toxic disease-driving repeat expansion; that is, structure-targeting ligands can be allele-selective.[17] In contrast, an ASO complementary to the r(CUG) repeats is not able to discriminate between short and long repeats and, thus, has off-target effects.[17]

After confirming that **1** cleaved r(CUG)^{exp} with similar selectivity and potency as Cugamycin, the ability of **1** to rescue DM1 defects, including the formation of r(CUG)^{exp}−MBNL1 nuclear foci[6] and MBNL1-regulated splicing defects,[7] was assessed. At 2 μM, **1** reduced the number of r(CUG)^{exp}−MBNL1 nuclear foci by ∼40% (see Figures 4D and 4E), similar to Cugamycin,[17] while **2**, which lacks RNA-binding modules, had no effect (see Figure S7 in the Supporting Information). In DM1 myotubes, *MBNL1* exon 5 splicing is dysregulated (Figure 1B), as MBNL1 regulates the alternative splicing of its own pre-mRNA.[29] Cleavage of r(CUG)^{exp} by **1** resulted in an ∼30% improvement in this splicing defect (see Figure 4F, as well as Figure S9 in the Supporting Information), which is an improvement that is similar to that observed for Cugamycin.[17] Compound **2** had no effect on *MBNL1* exon 5 splicing, as expected (see Figure S7). Importantly, **1** did not affect *MBNL1* exon 5 splicing in wild-type myotubes (see Figure S8 in the Supporting Information) nor the NOVA-dependent splicing of *MAP4K4* exon 22a[30] (see Figure S9). Thus, rescue of the *MBNL1* exon 5 splicing defect can be traced specifically to cleavage of the r(CUG)^{exp}. Collectively, these studies show that the removal of the carbohydrate domain of BLM allows for enhanced selectivity by further ablating DNA damage without affecting cellular permeability or activity.

Small molecules that selectively cleave a target RNA are attractive chemical probes, because they can more potently improve disease-associated defects than simple binding compounds.[17,31] Furthermore, RNA cleavage, either through direct cleavage as demonstrated herein or through recruitment of a cellular nuclease,[31] can be used to profile molecular recognition of RNA-binding small molecules. The use of BLM analogues to specifically cleave r(CUG)^{exp} offers an attractive method to enhance RNA cleavage selectivity by further diminishing off-target DNA cleavage. Although the carbohydrate domain of BLM is necessary for its efficient cleavage of DNA[19] and cellular permeability,[25] the disaccharide is not essential for permeability or cleavage of r(CUG)^{exp} when attached to r(CUG)^{exp}-binding small molecules. Thus, by using BLM analogues, RNA cleavage and the ability to improve DM1-associated defects is retained while further enhancing selectivity by reducing DNA damage that occurs with high concentrations of BLM-conjugated small molecules.

The most common way to target RNA for destruction is by using oligonucleotide-based target recognition elements. These approaches recognize unstructured regions in RNA. The ability to design ligands that target structured regions in an RNA and cleave them selectively provides an alternative approach to probe the biology of RNA in general and RNA structure in particular. Although bleomycin−small-molecule conjugates have higher molecular weights than orally bioactive drugs, they still have lower molecular weight than oligonucleotides and significantly lower molecular weight than CRISPR approaches that are packed into viruses.[12] In addition, medicinal chemistry approaches may be more broadly applicable to these compound sets as the RNA-binding modules and linkers that tether them can be therapeutically optimized. It is likely that, as more information is accumulated on the RNA folds that bind small molecules and on the small molecules that bind RNA folds, the deglycobleomycin cleavage module described herein could be broadly deployed. Furthermore, the ability to effect cleavage of an RNA target can allow for more diverse modes of action. Small molecules can now target an RNA for destruction in cells via three mechanisms: (i) direct cleavage by using bleomycin conjugates;[9,17] (ii) nuclease recruitment by using ribonuclease targeting chimeras (RIBOTACs);[31,32] and (iii) shunting introns with toxic expanded repeats to decay pathways.[33] Some targets may be more amendable to one strategy than the others. The ability to minimize off-target effects by using the deglycobleomycin cleavage module described here could have broad implications in this emerging area.

METHODS

A detailed description of methods can be found in the Supporting Information.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.0c00036.

Table S1; Figures S1−S9; experimental methods; synthetic methods and characterization (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; Phone: 561-228-2203; Email: Disney@scripps.edu; Fax: 561-228-2147

Authors

Alicia J. Angelbello − *Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States*

Mary E. DeFeo − *Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States*

Christopher M. Glinkerman − *Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, United States*

Dale L. Boger − *Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.0c00036

Notes

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

ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Defense Peer Reviewed Medical Research Program [No. W81XWH-18-0718 (to M.D.D.)], the National Institutes of Health F31 NS110269 (to A.J.A.), and R01 CA042056 (to D.L.B.)], the National Science Foundation [No. NSF/DGE-1346837 (to C.M.G.)],

and a Shelton Endowment Scholarship (C.M.G.). We thank D. Furling at the Centre de Recherche en Myologie for the generous gift of the cell lines used in this study. We thank P. Dawson for experimental advice with the synthesis of deglycobleomycin.

REFERENCES

(1) Ursu, A., Vezina-Dawod, S., and Disney, M. D. (2019) Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discov. Today 24*, 2002−2016.
(2) Disney, M. D. (2019) Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J. Am. Chem. Soc. 141*, 6776−6790.
(3) MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., and Lin, C. et al. (The Huntington's Disease Collaborative Research Group). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993, *72*, 971−983.
(4) 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., and Housman, D. E. (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell 68*, 799−808.
(5) Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. W. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of *ZNF9*. *Science 293*, 864−867.
(6) Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., and Singer, R. H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol. 128*, 995−1002.
(7) Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T., and Thornton, C. A. (2004) 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.
(8) 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., Moxley, R. T., and Thornton, C. A. (2013) Splicing biomarkers of disease severity in myotonic dystrophy. *Ann. Neurol. 74*, 862−872.
(9) Rzuczek, S. G., Colgan, L. A., Nakai, Y., Cameron, M. D., Furling, D., Yasuda, R., and Disney, M. D. (2017) Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol. 13*, 188−193.
(10) Childs-Disney, J. L., Hoskins, J., Rzuczek, S. G., Thornton, C. A., and Disney, M. D. (2012) Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. *ACS Chem. Biol. 7*, 856−862.
(11) Nakamori, M., Taylor, K., Mochizuki, H., Sobczak, K., and Takahashi, M. P. (2016) Oral administration of erythromycin decreases RNA toxicity in myotonic dystrophy. *Ann. Clin. Transl. Neurol. 3*, 42−54.
(12) Batra, R., Nelles, D. A., Pirie, E., Blue, S. M., Marina, R. J., Wang, H., Chaim, I. A., Thomas, J. D., Zhang, N., Nguyen, V., Aigner, S., Markmiller, S., Xia, G., Corbett, K. D., Swanson, M. S., and Yeo, G. W. (2017) Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. *Cell 170*, 899−912.
(13) Wheeler, T. M., Leger, A. J., Pandey, S. K., MacLeod, A. R., Nakamori, M., Cheng, S. H., Wentworth, B. M., Bennett, C. F., and Thornton, C. A. (2012) Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature 488*, 111−115.
(14) Jauvin, D., Chrétien, J., Pandey, S. K., Martineau, L., Revillod, L., Bassez, G., Lachon, A., MacLeod, A. R., Gourdon, G., Wheeler, T. M., Thornton, C. A., Bennett, C. F., and Puymirat, J. (2017) Targeting *DMPK* with antisense oligonucleotide improves muscle strength in myotonic dystrophy type 1 mice. *Mol. Ther.—Nucleic Acids 7*, 465−474.
(15) Siboni, R. B., Nakamori, M., Wagner, S. D., Struck, A. J., Coonrod, L. A., Harriott, S. A., Cass, D. M., Tanner, M. K., and Berglund, J. A. (2015) Actinomycin D specifically reduces expanded CUG repeat RNA in myotonic dystrophy models. *Cell Rep. 13*, 2386−2394.
(16) Lee, J., Bai, Y., Chembazhi, U. V., Peng, S., Yum, K., Luu, L. M., Hagler, L. D., Serrano, J. F., Chan, H. Y. E., Kalsotra, A., and Zimmerman, S. C. (2019) Intrinsically cell-penetrating multivalent and multitargeting ligands for myotonic dystrophy type 1. *Proc. Natl. Acad. Sci. U. S. A. 116*, 8709−8714.
(17) Angelbello, A. J., Rzuczek, S. G., Mckee, K. K., Chen, J. L., Olafson, H., Cameron, M. D., Moss, W. N., Wang, E. T., and Disney, M. D. (2019) 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.
(18) Guan, L., and Disney, M. D. (2013) Small-molecule-mediated cleavage of RNA in living cells. *Angew. Chem., Int. Ed. Engl. 52*, 1462−1465.
(19) Boger, D. L., and Cai, H. (1999) Bleomycin: synthetic and mechanistic studies. *Angew. Chem., Int. Ed. Engl. 38*, 448−476.
(20) Abraham, A. T., Lin, J.-J., Newton, D. L., Rybak, S., and Hecht, S. M. (2003) RNA cleavage and inhibition of protein synthesis by bleomycin. *Chem. Biol. 10*, 45−52.
(21) Madathil, M. M., Bhattacharya, C., Yu, Z., Paul, R., Rishel, M. J., and Hecht, S. M. (2014) Modified bleomycin disaccharides exhibiting improved tumor cell targeting. *Biochemistry 53*, 6800−6810.
(22) Wu, W., Vanderwall, D. E., Turner, C. J., Kozarich, J. W., and Stubbe, J. (1996) Solution structure of Co·bleomycin A2 green complexed with d(CCAGGCCTGG). *J. Am. Chem. Soc. 118*, 1281−1294.
(23) Goodwin, K. D., Lewis, M. A., Long, E. C., and Georgiadis, M. M. (2008) Crystal structure of DNA-bound Co(III)·bleomycin B2: Insights on intercalation and minor groove binding. *Proc. Natl. Acad. Sci. U. S. A. 105*, 5052−5056.
(24) Li, Y., and Disney, M. D. (2018) Precise small molecule degradation of a noncoding RNA identifies cellular binding sites and modulates an oncogenic phenotype. *ACS Chem. Biol. 13*, 3065−3071.
(25) Schroeder, B. R., Ghare, M. I., Bhattacharya, C., Paul, R., Yu, Z., Zaleski, P. A., Bozeman, T. C., Rishel, M. J., and Hecht, S. M. (2014) The disaccharide moiety of bleomycin facilitates uptake by cancer cells. *J. Am. Chem. Soc. 136*, 13641−13656.
(26) Wanner, J., Tang, D., McComas, C. C., Crowley, B. M., Jiang, W., Moss, J., and Boger, D. L. (2003) A new and improved method for deglycosidation of glycopeptide antibiotics exemplified with vancomycin, ristocetin, and ramoplanin. *Bioorg. Med. Chem. Lett. 13*, 1169−1173.
(27) Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., and Chen, D. J. (2001) ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J. Biol. Chem. 276*, 42462−42467.
(28) 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., and Furling, D. (2017) Immortalized human myotonic dystrophy muscle cell lines to assess therapeutic compounds. *Dis. Model. Mech. 10*, 487−497.
(29) Gates, D. P., Coonrod, L. A., and Berglund, J. A. (2011) Autoregulated splicing of muscleblind-like 1 (MBNL1) pre-mRNA. *J. Biol. Chem. 286*, 34224−34233.
(30) 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., and Darnell, R. B. (2005) Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet. 37*, 844−852.
(31) Costales, M. G., Matsumoto, Y., Velagapudi, S. P., and Disney, M. D. (2018) Small molecule targeted recruitment of a nuclease to RNA. *J. Am. Chem. Soc. 140*, 6741−6744.

https://dx.doi.org/10.1021/acschembio.0c00036
ACS Chem. Biol. 2020, 15, 849−855

(32) Costales, M. G., Suresh, B., Vishnu, K., and Disney, M. D. (2019) Targeted degradation of a hypoxia-associated non-coding RNA enhances the selectivity of a small molecule interacting with RNA. *Cell Chem. Biol. 26*, 1180−1186.
(33) Benhamou, R. I., Angelbello, A. J., Wang, E. T., and Disney, M. D. (2020) A Toxic RNA catalyzes the cellular synthesis of its own inhibitor, shunting it to endogenous decay pathways. *Cell Chem. Biol. 27*, 223−231.

https://dx.doi.org/10.1021/acschembio.0c00036
ACS Chem. Biol. 2020, 15, 849−855

Journal of Medicinal Chemistry

pubs.acs.org/jmc

Perspective

How We Think about Targeting RNA with Small Molecules

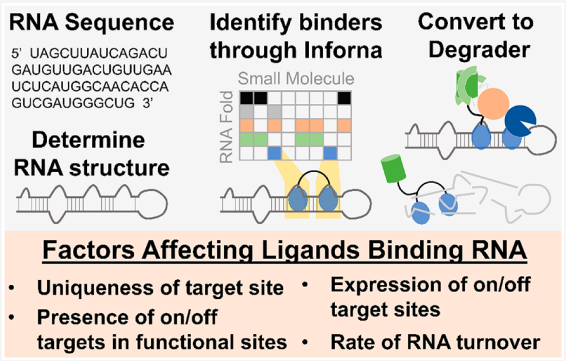
Matthew G. Costales, Jessica L. Childs-Disney, Hafeez S. Haniff, and Matthew D. Disney*

Cite This: https://dx.doi.org/10.1021/acs.jmedchem.9b01927

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: RNA offers nearly unlimited potential as a target for small molecule chemical probes and lead medicines. Many RNAs fold into structures that can be selectively targeted with small molecules. This Perspective discusses molecular recognition of RNA by small molecules and highlights key enabling technologies and properties of bioactive interactions. Sequence-based design of ligands targeting RNA has established rules for affecting RNA targets and provided a potentially general platform for the discovery of bioactive small molecules. The RNA targets that contain preferred small molecule binding sites can be identified from sequence, allowing identification of off-targets and prediction of bioactive interactions by nature of ligand recognition of functional sites. Small molecule targeted degradation of RNA targets (ribonuclease-targeted chimeras, RIBOTACs) and direct cleavage by small molecules have also been developed. These growing technologies suggest that the time is right to provide small molecule chemical probes to target functionally relevant RNAs throughout the human transcriptome.



DRUG DISCOVERY IN THE "-OMICS" ERA

In 2001, the dissemination of the draft sequence of the human genome was noted as a crowning scientific achievement.[1] Genomics inspired "omics" for nearly every biomolecule such as the proteome, the glycome, and the transcriptome.[2−5] Many of these studies have been aimed at associating differential biomolecule expression to disease states.[6] Ideally, being armed with an encyclopedia of genome sequence, one could readily identify the basis for understudied diseases and rapidly develop effective, orally bioavailable small molecule treatments.

Because of the focus of using genomics to inform medicinal development, one of the first genome-wide analyses determined the number of open reading frames (ORFs). Surprisingly, these annotations, culminating in the Encyclopedia of DNA Elements (ENCODE) project,[7] revealed that canonical ORFs are present in only $^{1}/_{3}$ of the genes that humans were previously predicted to have. Additionally, only 2% of human DNA is translated into protein, the most well studied small molecule drug target (Figure 1).[8] Interestingly, it was found that 90% of the genome is transcribed into RNA, with the vast majority of these having non-protein-coding functions.[8,9] Noncoding RNAs have been subsequently found to have many different functions, and analysis of the noncoding RNA between organisms follows their complexity, in contrast to differences in their respective protein-coding genes (Figure 1).[10,11] Historically, protein is the predominant biomolecule considered for small molecule drug targeting. Thus, druggability, or whether a target protein is a member of a family that has been previously targeted with a small molecule, was initially assessed across the genome for protein-coding genes. This analysis revealed that 15% of proteins are considered "druggable", i.e., within genes that have been targeted with a small molecule (Figure 1A).[12,13] Recent studies suggest that this notion may need reassessment, however. Perhaps the language should be changed to "undrugged", as a wide variety of difficult to target proteins have been targeted by small molecules, including the mutant Kirsten rat sarcoma (KRAS) proto-oncogene.[14,15]

A dogma for the development of small molecules against this fraction of the proteome has been centered on some tenants that perhaps need re-evaluation. First, binding sites for a bioactive ligand must be in an active site or an allosteric site to affect function. Second, the occupancy-driven view of pharmacology dictates that high target occupancy is necessary for effective target inhibition. Thus, maintained exposure to the target is the major driver of pharmacodynamic profiles. Third, the proteome is targetable at defined three-dimensional clefts that allow for high surface area recognition of ligands.

Received: November 21, 2019

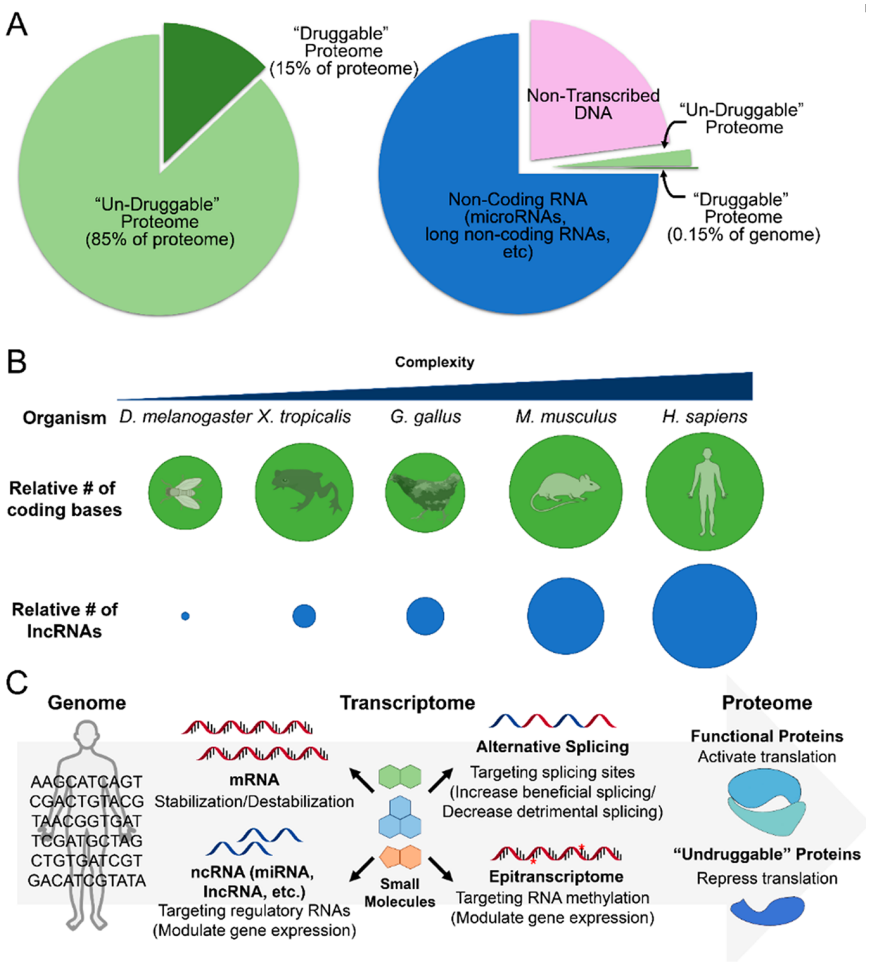


Figure 1. RNA as a viable drug target. (A) The conventional binary approach to small molecule drugs is their molecular recognition of proteins. Among the ~20 000 proteins that comprise the proteome, only about 15% are in traditional "druggable" protein families. In turn, this only represents a fraction of the genome that is transcribed, leaving much of the transcriptome underexploited as therapeutic targets. (B) Noncoding genes relate to the complexity of the organism, as the relative number of coding bases remains similar, while the relative number of long noncoding RNAs (lncRNA) significantly increases, indicating that much of the intricacies of human biology and disease are represented among noncoding regions. (C) Due to the importance of coding and noncoding RNA to biology, small molecules interacting with RNA can act on the transcriptome, resulting in varied downstream effects. Importantly, validated activities for small molecules that target human RNA include: (i) changing gene expression by modulating the stability of mRNA by direct binding; (ii) affecting its noncoding RNA effectors; (iii) affecting the epitranscriptome; or (iv) influencing alternative splicing. Affecting the transcriptome with small molecule drugs can rescue disease by modulating the translation of beneficial or detrimental proteins.

Many groups have worked tirelessly to challenge these assertions. While there are no hard and fast rules in drug discovery, pragmatic solutions to important biomedical problems always win out. For example, recent developments on targeting KRAS have shown that ligands can be developed to target important proteins that do not have the traditionally desired ligand binding pockets.[16−18] There are a variety of new ligands in the clinic that affect these previously perceived as impossible-to-drug targets. Additionally, work in the targeted protein degradation, or proteolysis targeting chimeras (PROTACs), area has challenged notions of occupancy-driven pharmacology, as well as the necessity for bioactive ligands targeting allosteric or active, i.e., functional, sites.[19] For example, the ability to recruit E3-ubiquitin ligases to selectively degrade target proteins bound by selective ligands suggests that small molecules do not require occupancy to enable a pharmacological effect. Additionally, ligand occupancy or "ligandability" has been studied across the proteome and has provided an encyclopedia of information on proteome-wide ligand occupancy.[20,21]

RNA could be advantageous as a small molecule drug target, following the paradigm of occupancy-driven inhibition of functional sites with defined three-dimensional structure employed for enzymes and receptors. RNA is also an attractive target as it is causative and/or upstream of pathological mechanisms related to disease states (Figure 1).[22,23] Importantly, RNAs can be analyzed from tissue or liquid biopsies more easily than proteins because nucleic acids can be amplified and sequenced. This Perspective focuses on our current understanding of the molecular recognition of RNA by small molecules, the design of *bioactive* small molecules, and tools used to study RNA target validation, engagement, and selectivity. There are many excellent reviews on RNA chemical biology and drug discovery that cover topics outside the scope of this Perspective.[24−29]

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

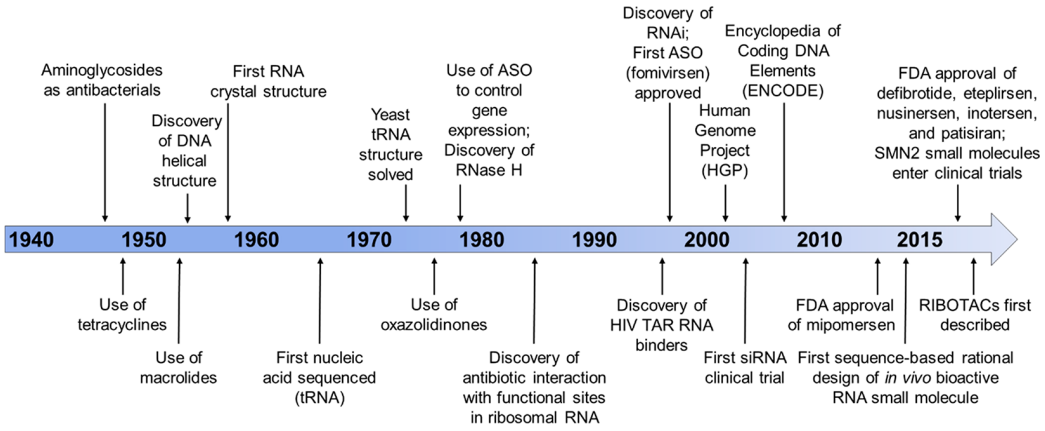


Figure 2. Timeline of major developments in the RNA-targeting field. The history of drugging RNA is tied closely with molecular biology discoveries (DNA/RNA structural determination). Antibacterials that targeted RNA preceded the first investigations into antisense oligonucleotides. However, FDA approvals of antisense oligonucleotides increased upon completion of the human genome project. Recent developments, such as the use of rational design-based approaches, the initiation of clinical trials for small molecule drugs treating spinal muscular atrophy (SMA), and the first report of ribonuclease targeting chimeras (RIBOTACs), demonstrate the rapid development of small molecules targeting RNA. These advancements provide a convincing argument to turn our focus to the druggable transcriptome.

■ ASSESSING RNA AS A DRUG TARGET

Since the beginning of modern medicine, RNA played an essential role as a small molecule drug target (Figure 2). Streptomycin was found to target the bacterial ribosome and shortly after its discovery, became a first-in-line treatment for *Mycobacterium tuberculosis*. Prior to this breakthrough discovery, tuberculosis was a devastating disease, and those affected were housed in sanitariums until they succumbed to the infection. Other natural products isolated from species of *Streptomyces* bacteria are also tried and true antibacterials such as other aminoglycoside antibiotics like neomycin B, the active ingredient in Neosporin. Studies of these compounds and their derivatives quickly expanded to various other aminoglycosides-based antibiotics, which are still in wide use today.[30] While these aminoglycosides are effective as broad-spectrum antibiotics, they tend to be promiscuous binders, leading to their limited use unless they are modified.[31] However, between their discovery during WWII and now, there have been few ligands that target RNA that have progressed to the clinic (Figure 2).

Phenotypic screens have been broadly used in the drug discovery industry beyond screening bacterial lysates to find active antibacterials. These approaches provide compounds that affect a wide variety of pathways. The challenge with phenotypic screens is identifying the engaged targets to understand the mechanism of action, compounded by the fact that target validation tools for these assays were generally designed to probe protein-centric pathways. For example, pooled CRISPR and shRNA hairpin libraries knock down ORFs (protein) or forced expression of the ORF can be used to track down a target for a small molecule.[32,33]

Although these phenotypic screens do not traditionally consider RNA as a potential target, they have indeed discovered compounds that modulate RNA. Unfortunately, these screens do not necessitate interaction with the target RNA of interest to modulate downstream RNA biology. Discovered serendipitously through phenotypic screening of small molecules interfering with the bacterial riboflavin biosynthetic pathway, Merck reported ribocil, which acts to inhibit the flavin mononucleotide (FMN) riboswitch and subsequently disrupt translation of the downstream mRNA.[34] After demonstrating that riboflavin genes (*ribA* and *ribB*) are essential for *Escherichia coli* growth, ∼57 000 compounds were screened for inhibition of bacterial growth that could be reversed by riboflavin supplementation. This screen identified that ribocil inhibited cellular production of riboflavin and its metabolites. Compound-resistant *E. coli* strains were then sequenced, revealing that all mutants contained base pair mutations affecting the aptamer region of the FMN riboswitch, which controls *ribB* expression. Additional binding and reporter gene assays indicated that ribocil was a competitive inhibitor of FMN and that inhibition was ablated in FMN mutant constructs, validating this RNA as the compound's target. Although ribocil selectively binds the FMN riboswitch and is structurally distinct from the endogenous FMN riboswitch ligand, the rapid development of bacterial resistance in various species precluded it from progression as a clinical candidate. Overall, the development of ribocil has proven that phenotypic screens can provide small molecules that bind RNA, encouraging investigators to consider RNA modulation as a mechanism of action.

An early adaptor of these phenotypic studies was PTC Therapeutics. Their programs include compounds that perturb readthrough of stop codons and affect pre-mRNA splicing outcomes. Phenotypic screening identified ataluren, a clinical candidate for treatment of Duchenne's muscular dystrophy (DMD). DMD is caused by a premature stop codon in the dystrophin mRNA, and ataluren is thought to stimulate its readthrough. Although ataluren was conditionally approved for treatment of DMD in Europe, the FDA declined to accept it as a drug based on a clinical trial in which ataluren missed its primary end point.[35,36] While exact details of its mechanism are not known, studies revealed that ataluren directly binds and stabilizes firefly luciferase, thus allowing for its identification in the luciferase-based nonsense codon suppression assay.[37,38] Therefore, in order to avoid false positives, target activity observed from phenotypic screens must also consider potential interference with downstream assay signals.

Risdaplam is a second compound that emerged from these screens and affects pre-mRNA splicing as a treatment for spinomuscular atrophy (SMA). This autosomal recessive neurodegenerative disease is caused by deficiencies in survival motor neuron (SMN) protein, which is caused by deletion or

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

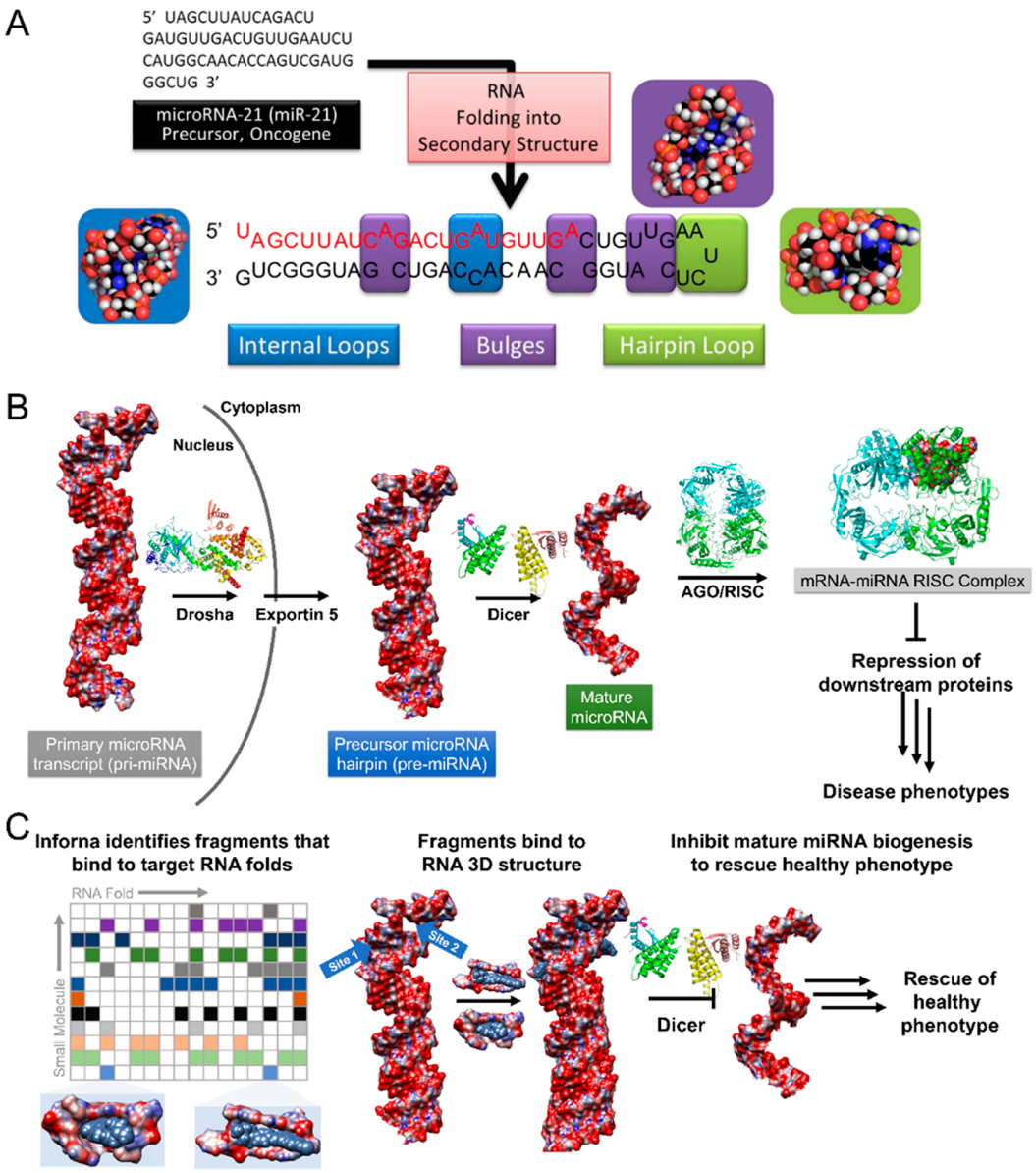


Figure 3. Modular RNA secondary structure motifs form three-dimensional structures. (A) Hierarchical assembly of RNA structure from sequence to secondary structure. Many of these secondary structures can form modular RNA motifs that can allow for small molecule recognition. (B) Structural schematic of microRNA processing. Primary transcripts (pri-miRNAs) are processed by the Drosha endonuclease to yield precursor hairpins (pre-miRNAs), which are exported to the cytoplasm and subsequently processed by the Dicer endonuclease to liberate a mature miRNA. One of the mature strands is then loaded into the argonaute/RNA-induced silencing complex (AGO/RISC), whereupon it acts on RNAs to modulate gene expression. Aberrant miRNA expression can be causative of disease phenotypes. (C) One rational design approach to target RNA is to understand the molecular recognition of structural elements by small molecules, those elements that are preferred by the small molecule and those that are discriminated against. Inforna compares structural elements within an RNA target to a database of these preferred interactions to afford lead small molecules. For example, binding to the pri- or pre-miRNAs at functional Drosha or Dicer sites can prevent their processing to the active, mature strand, thus allowing the rescue of disease-associated phenotypes through the inhibition of biogenesis.

loss in function of the *SMN1* gene.[39] SMA is classified by disease severity, which scales with the reduction of SMN protein levels. The loss of functional SMN protein can be compensated by its expression through the *SMN2* gene. However, *SMN2* typically produces shorter and nonfunctional SMN protein due to a C to T single nucleotide change that causes biased exon 7 skipping, decreasing the half-life of the resulting protein. Phenotypic screens led by Roche/PTC identified compounds that increased inclusion of exon 7 in a luciferase-based cellular assay. Following efficacy studies in patient-derived cells and mouse models and RNA-Seq analysis to verify selectivity, the orally active Risdaplam was identified to convert the SMN2-encoding pre-mRNA into a longer lived version by modifying its splicing outcome.[40] In a similar report, phenotypic screening at Novartis identified a compound, Branaplam, that similarly increased production of functional SMN protein through splicing modulation, which is now progressing through clinical development.[41]

Due to their identification through phenotypic screens, the mechanisms of action of these splicing modulators have not been completely defined.[42] Multiple studies have proposed that compounds in the Risdiplam series can achieve such

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

exceptional molecular recognition and specificity by formation of a higher order complex comprising the exonic splicing enhancer (ESE) sequence, the 5′-splice site (5′-ss), and other regulatory splicing proteins.[43] Wang and others used several approaches, including chemical cross-linking and isolation by pull-down (Chem-CLIP), in order to determine direct target engagement and interactions of these compounds with RNA.[42,44] These Chem-CLIP studies, in addition to a series of genomic and proteomic experiments of this class of compounds, revealed their direct binding to the AGGAAG motif present in exon 7 of the SMN2 pre-mRNA. Small molecule binding promotes conformational changes that result in increased binding of splicing modulators (FUBP1, KHSRP) to enhance SMN2 splicing.[42] These studies suggest that compounds can directly bind and alter *SMN2* RNA structure, thereby affecting interactions with other splicing modifiers.[42,43] Unfortunately, the specificity afforded by these compounds in the context of interacting with specific splicing proteins and RNA motifs in SMA may not be available for all RNA targets.

Collectively, none of these efforts purposefully targeted RNA. While they validate that small molecules can affect several different RNA functions, purposeful targeting of RNA is altogether different. Tools such as Chem-CLIP will therefore by key for exploiting RNA as a drug target, providing direct evidence of target engagement and hence compound mode of action.

RICH STRUCTURE OF RNA

Mammalian RNA encodes both genetic and structural information, exemplified by noncoding transfer (t)RNA, which was the first nucleic acid sequenced (Figure 2).[45] The discovery that the stable cloverleaf structure of tRNA interacts with the codon of messenger (m)RNA represents that RNA−RNA interactions function as a physical template for protein synthesis, setting the stage for the central dogma of biology.[46] Beyond their fundamental roles in protein synthesis, noncoding RNAs also play essential regulatory roles. Recent research even points to the fact that organismal complexity is directly associated with the number of noncoding RNAs, rather than the size of the genome (Figure 1).[10] As an early example, the Steitz group proved in seminal work that noncoding U1 small nuclear RNA (U1 snRNA) recognizes 5′ splice sites in pre-mRNA,[47] demonstrating that RNA structural recognition is necessary for accurate splicing and that RNA adopts intricate structures that influence genetic processes.

RNA folds into elaborate structures that enable its essential functions in diverse applications in biology, encompassing regulation of gene expression, ligand sensing, catalysis, and others.[48] In order to minimize its free energy, single stranded RNA forms fully base-paired as well as noncanonically paired regions (hairpins, internal loops, bulges, etc.) that determine higher order (tertiary) folding patterns (Figure 3).[48] These secondary and tertiary RNA structural elements are highly dynamic and dependent on their environment (protein, salt concentration, small molecules, etc.), which can greatly affect function. Within structured RNAs, over 50% of residues are confined in Watson−Crick base-paired helices, and emerging studies show that subtle changes to secondary structure, or variation in Watson−Crick base pair interactions within RNA helices, significantly impact the formation of tertiary structures.[49,50] Notably, not all human RNAs contain long-range interactions between secondary structural elements, or tertiary structure (unlike RNase P RNA, (t)RNAs, and ribosomal (r)RNAs that display well-defined tertiary structures), including coaxial stacking and pseudoknots, among others. Recent investigations into RNA tertiary folding energetics have also suggested that tertiary structure can be defined through a dynamic thermodynamic ensemble of assembled secondary structural elements based on RNA sequence.[51] That is, secondary structure limits the number of tertiary interactions available and hence constrains tertiary structure.

As understanding RNA folding is foundational to deducing its overall structure, much effort has been dedicated toward accurately predicting RNA secondary structure from sequence, including phylogenetic comparison, free energy minimization, and combinations thereof.[52−54] For example, entire kingdoms of life have been determined through the conservation of RNA secondary structure and phylogenetic comparison.[55] In cases where limited RNA sequences are available, secondary structure prediction through free energy minimization is commonly used.[56] These free energy calculations are able to provide both optimal and suboptimal structures.[57] This technique is more powerful when incorporating experimental constraints from structural probing, allowing accurate determination of more complicated RNA structures (rRNA, viral RNAs, etc.).[58] Use of dimethyl sulfate (DMS) chemical probing has also allowed global monitoring of RNA structure *in vivo* with single nucleotide resolution, showing that cellular mRNAs are predominantly unfolded.[59,60] Optimal integration of chemical probing experimental data and computational tools remains an ongoing challenge in the field of RNA structural determination, as recent quantitative tools to measure covariations in RNA alignments have disputed evidence of proposed secondary structures of various long noncoding RNAs.[61] Recently, advanced bioinformatics scanning window models (ScanFold) have allowed generation of high likelihood functional RNA structures.[62] These resultant RNA structuromes provide evidence for hubs of structured regions within viral and mammalian RNA.[52]

RNAs play essential roles in cellular processes; thus their dysregulated expression or misfolding can be causative of disease pathologies.[22,23] For example, overexpression of regulatory noncoding RNAs, such as microRNAs (miRNAs) that have defined secondary structural elements but lack tertiary structure, can result in oncogenic phenotypes.[63] Improper folding of RNA can also be responsible for disease pathology, as is commonly seen with short tandem repeat, or microsatellite, RNA expansions.[64] Due to their misfolding, several pathological mechanisms can cause a wide variety of disease mechanisms. Examples include (i) dysregulation of RNA splicing due to single nucleotide polymorphisms (SNPs), as seen in Tau neurodegenerative disorders; (ii) sequestration of essential proteins that disrupts their normal function, as observed in the sequestration of splicing factors by expanded r(CUG) or r(CCUG) repeats in myotonic dystrophy type 1 (DM1) and type 2 (DM2), respectively; and (iii) the production of toxic or nonfunctional proteins, as observed in r(G_4C_2) repeat expansions observed in frontotemporal dementia and amyotrophic lateral sclerosis (FTD/ALS).[64] Overall, the molecular basis for many diseases is rooted in deregulated RNA function, which is intimately tied to its proper structural folds.

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

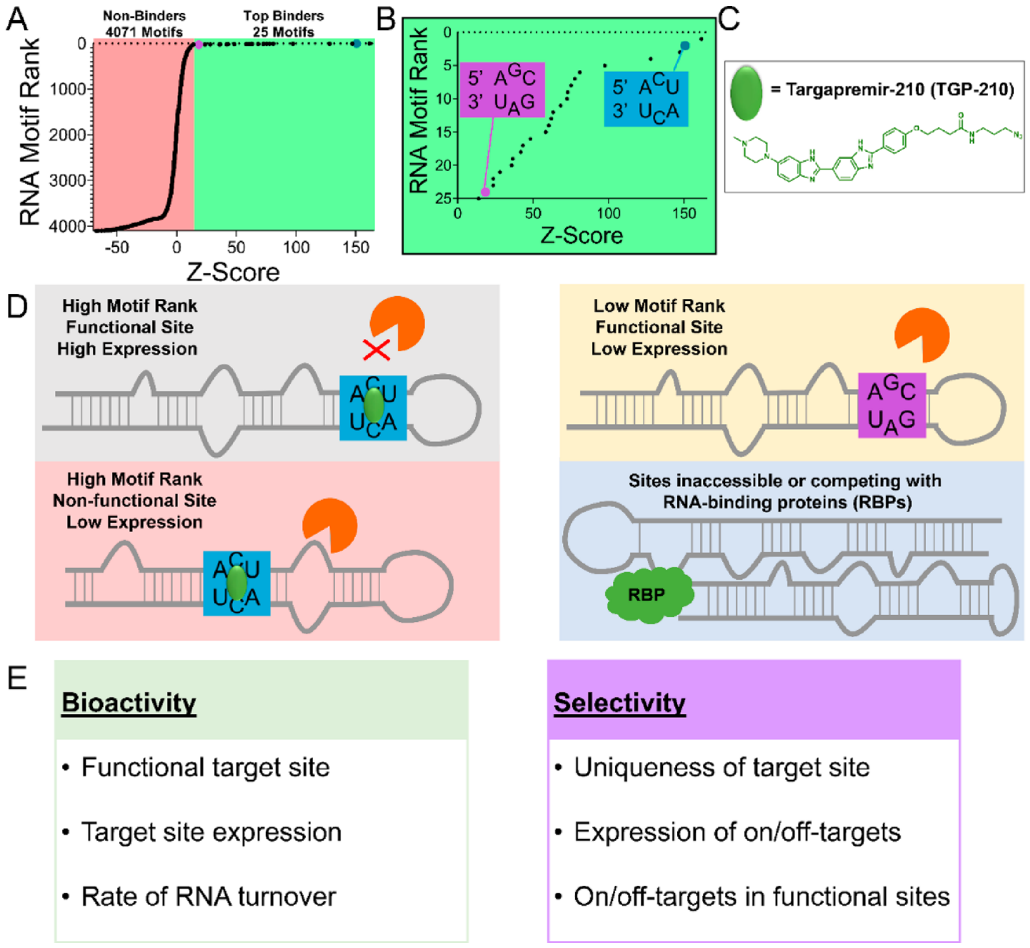


Figure 4. Factors that influence bioactivity and selectivity of RNA-binding ligands. (A) 2DCS selection and HiT-StARTS analysis identifies the top binding RNA motifs and nonbinders to targapremir-210 (TGP-210). Z-Score is a calculated value that represents fitness of the RNA-small molecule interaction. (B) Fitness plot of the top 25 binding motifs are shown. One of the highest fitness RNA motif interactions with TGP-210 is the 5′ ACU/3′ UCA 3 × 3 internal loop (blue box) found in the functional Dicer site of miR-210. The 5′ AGC/3′ UAG RNA motif is a lower affinity interaction (purple box). (C) Structure of TGP-210. (D, E) The bioactivity of TGP-210 to selectively inhibit miR-210 biogenesis is a function of multiple factors, including the fitness of the RNA motif-small molecule interaction, the expression and turnover of the target RNA(s), whether or not occupancy of the target site will result in a functional interaction (i.e., occupying a Dicer/Drosha site), and the accessibility of TGP-210 to off-target sites.

BASICS OF TARGETING RNA

As aberrantly folded RNA structures contribute to disease, it is of utmost importance to provide compounds to correct their dysfunction or to facilitate their study in healthy and disease states. Beyond simple molecular recognition, however, targeting RNA is also a function of a variety of interrelated factors (Figure 4). First of all, drugs to target RNA must be able to access the structural elements present in the target RNA, which may be in prohibitive structural interactions with other biomolecules [i.e., competition with RNA-binding proteins (RBPs)], or the RNA itself may be difficult to access (e.g., blood−brain penetrance). Additionally, selective compounds must be able to disrupt the on-target RNA, while not binding to off-target transcripts, within acceptable limits. Binding to a specific transcript is also related to the expression levels of competing RNA motifs, which is intimately connected to tissue-specific or cellular compartment-specific expression. Furthermore, compounds must be able to elicit observable biological modulations; that is, compounds must interact with RNA at a functional site (Figure 4).

A common approach to affect RNA function is sequence complementarity through antisense oligonucleotide (ASO) hybridization with a target strand. The resultant antiparallel base-paired duplex can either affect the natural folding of the target RNA, thus disrupting its interactions with other biomolecules, or recruit endogenous cellular machinery to cleave the RNA.[65] These interactions, however, are dependent on the thermodynamic and kinetic energy barriers necessary for the folding/unfolding/hybridization of the native conformations of both the target RNA and ASO, which can limit their effectiveness against highly structured RNAs.[64] Thus, the antisense-based strategy is best applied to weakly structured RNAs. Beyond these difficulties to provide meaningful molecular interactions, oligonucleotides have been known to have inadequate *in vivo* properties (limited delivery strategies, biodistribution, and tissue penetrance) and exhibit various side effects that include eliciting of an immune response, thrombocytopenia, and hepatotoxicity in patients that have caused clinical trials to be halted.[65−68]

As the previous FDA standards indicate, ASO drugs are assumed to be selective. Intriguingly, the selectivity of past

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

FDA-approved ASOs have not been thoroughly studied transcriptome-wide or proteome-wide, despite the increasing stringency to provide mode-of-action centric, selective drugs at the molecular level.[66,69] Indeed, investigations into the specificity and selectivity of ASOs are more nuanced than previously thought.[66] Naively, it might be assumed that selectivity can be enhanced by simply lengthening the oligonucleotide. Herschlag and co-workers found, however, that there is an optimal length of oligonucleotide to achieve selectivity.[70] Oligonucleotides that are too long can form thermodynamically stable duplexes containing mismatches with off-targets. Indeed, the sources of hybridization of oligonucleotides with unintended targets are well-studied, and developments in analytical bioinformatics and algorithms have allowed computational screening to determine the maximal on-target transcript effects with an acceptable tolerance of off-targets.[66,71] Still, the experimental determination of on-/off-targets through transcriptome-wide profiling of ASOs is still necessary to measure selectivity. Interestingly, there are multiple demonstrations that small molecules that target RNA can rival the selectivity of ASOs or even exhibit enhanced selectivity in certain cases.[72−74] Thus, small molecule targeting of RNA structure may overcome the various liabilities of ASOs.

An emerging strategy to selectively target RNA with small molecules is to focus on the molecular recognition of structural elements.[75] One approach to rapidly define selective small molecule binders of RNA structural motifs is through a selection-based method that screens a library of RNA motifs vs a library of small molecules embedded in an agarose microarray (Figure 3C).[76] Dubbed two-dimensional combinatorial screening (2DCS), the small molecule microarray is incubated with a labeled RNA library that displays discrete structural elements in a randomized region. Weak binders and interactions with elements common to all library members (i.e., outside the randomized region) are competed off with an excess of oligonucleotide competitors (including DNA and regions that mimic the cassette displaying the randomized motif). Upon excision, amplification, and RNA-seq analysis of the bound RNA, the motifs preferred by each small molecule are defined by a statistical method named "high throughput structure−activity relationships through sequencing" (HiT-StARTS).[77] HiT-StARTS determines the statistical significance of the enrichment of a given motif in the 2DCS selection vs the starting library. The higher the statistical confidence in this enrichment, the more privileged, or fit, the interaction is. Normalization of the statistical confidence affords a fitness score for all members of an RNA library for a given small molecule. By creating a database of the results from 2DCS and HiT-StARTS, we developed a lead identification strategy for RNA targets, dubbed Inforna (Figure 3C).[78,79] That is, Inforna searches RNA targets for structural elements preferred by small molecules, which are chemical leads. Similar to the Watson−Crick base-pairing rules that govern oligonucleotide selectivity to RNA sequence, Inforna provides the small molecule equivalent of experimentally derived RNA fold-small molecule interactions for selective recognition of RNA structure.

While structure-based targeting shows great promise as detailed below, it does have limitations. For example, not all structural elements are functional; thus, even if a small molecule binder is identified, it likely will have no biological consequence. Alternatively, a ligand may not yet have been discovered for a particular functional structure. We have recently developed a strategy dubbed ribonuclease targeting chimeras (RIBOTACs) to overcome these limitations (discussed below). Furthermore, not all structures may be accessible for ligand binding, for example, due to the presence of long-range tertiary interactions or its interaction with RBPs.

SELECTIVE SMALL MOLECULE APPROACHES TO PURPOSEFULLY AFFECT microRNA BIOLOGY

Much of the transcriptome exists in defined hubs of structured RNA folds, and Inforna's foundation is in querying these highly probable regions of structured RNA elements against experimentally identified and characterized RNA 3D fold-small molecule interactions. Built into Inforna are several strategies to rationally design compounds to purposefully target RNA and affect downstream biology. As selectivity remains the largest concern for small molecule targeting of RNA, Inforna has the ability to predict on- and off-target transcripts (Figure 4) (that is, RNAs with 3D-folds that are optimal (high fitness) or suboptimal for small molecule targeting).[80] Additionally, these RNA 3D folds can be limited to parts of RNA that are important for bioactivity, such as the Dicer and Drosha processing sites in miRNAs. By use of the highest fitness and most selective binders from Inforna, small molecules that target oncogenic miRNA precursors have been designed and have shown promising preclinical data *in vivo*.[75,77]

Studies have determined that Inforna-defined RNA-small molecule interactions can inform selective targeting of miRNA precursors. These studies have formulated guidelines for targeting RNA with small molecules. The oncogenic miR-210 provides a prototypical example. Aberrant expression of miR-210 is observed in cancer cells that are in low oxygen, hypoxic environments, such as in solid breast cancer tumors.[81,82] Inforna identified a compound, Targapremir-210 (TGP-210), that selectively inhibits miR-210 biogenesis by binding to the C/C internal loop displayed in its Dicer processing site.[80] Inhibition of miR-210 disrupted the hypoxic circuit, resulting in stimulation of apoptosis in cellular and *in vivo* models.[80,83]

Inhibition of miR-210 as a result of selective on-target engagement was confirmed with a Chem-CLIP probe in which TGP-210 was appended to a cross-linking (chlorambucil) module and a purification (biotin) module. Chem-CLIP studies revealed that miR-210 was the most enriched transcript among highly abundant RNAs (rRNAs, tRNAs, mRNAs, etc.) but also among hypoxia-associated miRNAs and miRNAs that contain suboptimal binding sites (Figure 4). As the less expressed miR-497 contained the same C/C internal loop as miR-210, it was also pulled down in the enriched Chem-CLIP fraction. Transcripts like miR-497, or other RNAs that contain predicted interactions with the lead compound, are termed RNA isoforms. Interestingly, TGP-210 bound to pre-miR-497 in a nonfunctional site; thus miR-497 biogenesis remained unaffected.

Competitive Chem-CLIP (C-Chem-CLIP) was also applied by co-treating cells with the TGP-210 Chem-CLIP probe and the parent compound. Here, levels of miR-210 were depleted in the pulled down fraction due to competition for occupancy of target binding sites with TGP-210. These thorough studies elucidated various rules necessary for small molecules targeting RNA structure to selectively affect RNA biology. Importantly, the presence of the C/C internal loop in the functionally relevant Dicer processing site enabled TGP-210 to modulate

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

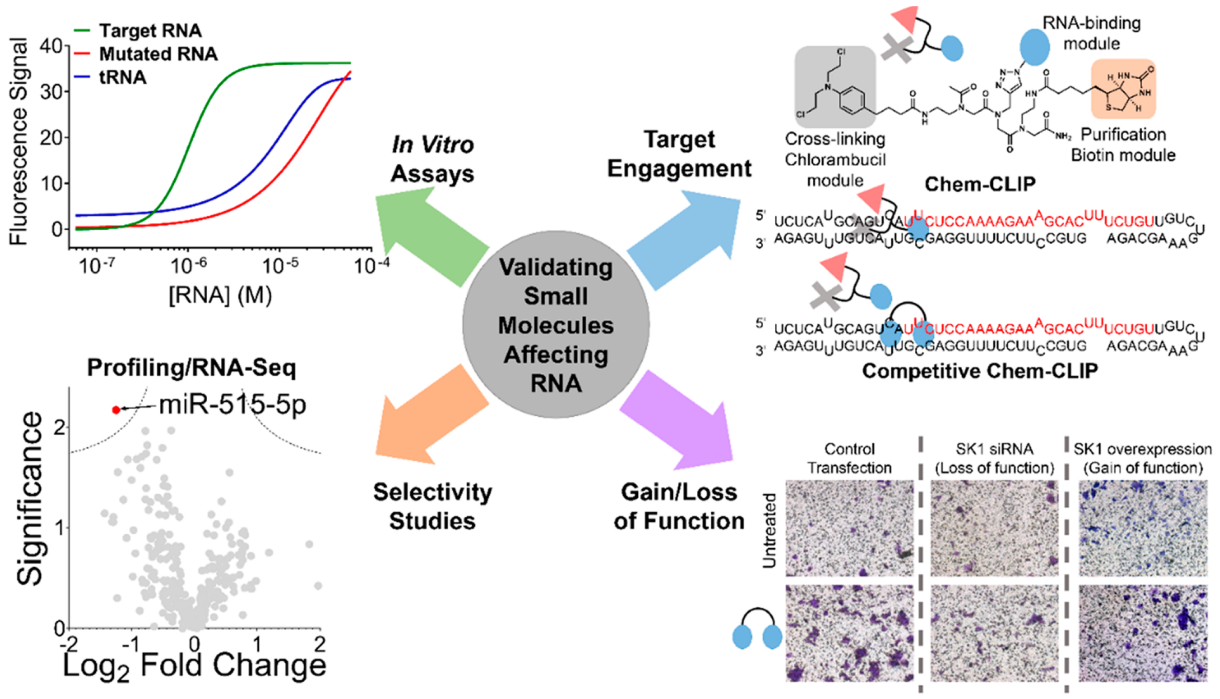


Figure 5. Tools to assess target engagement and selectivity of small molecules targeting RNA. Developing small molecules against RNA starts with identification of a hit, whether through Inforna (Figure 3) or screening approaches (target-based, phenotypic, fragment-based, DNA-encoded, etc.). Considering the factors from Figure 4, the hit must then be validated and optimized, including for *in vitro* binding affinity to the RNA structural element over RNAs that do not contain the motif and other abundant RNA/DNAs. Further validation *in vitro* and in cells can be accomplished with target engagement approaches that use chemical probing methods that measure RNA enrichment (Chem-CLIP) or RNA depletion (Competitive (C)-Chem-CLIP), among others. Comprehensively evaluating cellular selectivity on a transcriptome- and proteome-wide scale is also part of the workflow to validate a small molecule RNA target. After demonstration of selective on-target effects, the compound's functional effect must then be validated in more advanced models, including the effect of a gain or loss in expression of the target.

miR-210's downstream biology. Despite Chem-CLIP studies indicating target engagement with off-target miR-497, due to its interaction with TGP-210 in a nonfunctional site, its binding remained "silent". Additionally, abundance of the target plays an important role, as miR-497 expression is 10-fold lower than the overexpressed miR-210 in hypoxia, potentially explaining its lower occupancy by TGP-210 (Figure 4). Giving confidence to target expression being necessary for compound bioactivity is the fact that apoptosis was not triggered in normoxic cells, i.e., in a regular oxygen environment, that do not overexpress miR-210. Overall, target expression levels and occupancy of a biologically relevant site are necessary variables to consider for small molecules that target structured RNAs. Furthermore, target engagement techniques, such as Chem-CLIP, can be used to verify on- and off-targets.

In the case of miR-210, Inforna identified lead RNA motif-small molecule interactions that are uniquely displayed in highly expressed, functional sites. Given the high concentration of bystander RNAs (rRNA, tRNA) and RNA isoforms, there are small molecules that bind motifs found in multiple RNAs.[84,85] As an example, Inforna identified a small molecule with overlapping affinity for the functional Drosha sites present in both pri-miR-515 and pri-miR-885.[78,86] Selective inhibition of pri-miR-515 was further compounded by the ~2.5-fold increased expression of miR-885 relative to miR-515 (Figure 4). In order to discriminate between these two miRNAs, Inforna queried adjacent structured regions for a lead small molecule, enabling optimization through a modular approach in which multiple motifs can be targeted with the same molecule.[73] While many transcripts may contain the specific RNA motif targeted by a small molecule, not all of them will be presented in functional sites, while even fewer RNAs will contain two targetable sites separated by defined, specific distances (Figure 4). By screening a library of RNA-binding modules separated by different linker lengths,[87] a dimeric compound, Targaprimir-515 (TGP-515), was developed to selectively inhibit production of miR-515 vs miR-885.[73] Treatment with TGP-515 in a nonmigratory breast cancer cell line resulted in the increased biosynthesis of sphingosine 1-phosphate (S1P) through the derepression of sphingosine kinase 1 (SK1), triggering a migratory phenotype. This multivalent approach has been also been applied to target other specific RNAs successfully.[88,89]

The selectivity of TGP-515 to inhibit miR-515 biogenesis and modulate downstream biology was validated through several approaches. In a similar manner to the miR-210 studies, a Chem-CLIP probe was synthesized, and competition with TGP-515 indicated selective occupancy of only pri-miR-515, and not pri-miR-885. The selectivity of the dimer compound was rigorously confirmed through quantitative polymerase chain reaction (qPCR) profiling, RNA-seq, and global neo-protein studies, demonstrating on-target pathway activation and limited off-target effects in MCF-7 cells (Figure 5). Interestingly, proteomic studies indicated that human epidermal growth factor receptor 2 (HER2) was the most upregulated protein, which is normally not highly abundant in basal MCF-7 cells. Pretreatment with TGP-515 in various HER2-negative cell lines sensitized cells to the anti-HER2 therapies, Herceptin and Kadcyla, proving the hypothesis that increased HER2 production upon TGP-515 treatment can

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

render insensitive cells sensitive to targeted therapies. Generally, the selectivity of a small molecule for its target is one of the first considerations when designing chemical probes. In this study, potential off-targets were identified computationally through Inforna, and suboptimal selectivity was overcome through the use of multivalency. Ultimately, validation and selectivity studies were able to identify a potential precision medicine approach for difficult-to-treat cancers.

In a similar case to miR-515, a monomeric compound inhibited the biogenesis of both miR-377 and miR-421 by binding to their common Dicer functional sites.[90] Fortuitously, Inforna identified a lead compound that bound to an adjacent RNA motif in the pre-miR-377 hairpin that was not present in pre-miR-421. Upon optimization of the linker spacing between the RNA-binding modules that bound the two sites in pre-miR-377, the resulting dimeric compound selectively inhibited pre-miR-377, effecting angiogenesis through modulation of VEGFA protein.

Thus, Inforna has proven to be broadly applicable for the purposeful targeting of functional sites in human RNAs with selective, bioactive small molecules, while also providing comprehensive data sets defining RNA landscapes that are key for selective targeting and for identifying chemotypes with RNA-binding capacities. In addition, Inforna can be applied to develop multivalent ligands with improved selectivity and potency for RNAs that display unique targetable RNA motifs.

EXPLOITING RNA STRUCTURE FOR CHEMICAL PROBE DESIGN

Many proteins are difficult to drug directly since they lack traditionally targeted small molecule binding pockets; that is, they have intrinsically disordered domains. Therefore, affecting the coding mRNA upstream of protein production may present a more viable therapeutic option to decrease aberrant protein levels. One aspect of protein folding only recently receiving attention for its possible functional roles is intrinsically disordered regions.[91,92] Indeed, only a small fraction of protein crystal structures recorded in the Protein Data Bank (PDB) do not possess disordered regions. These regions could represent a protein−protein interface or potential allosteric site, the folding of which is induced by binding of another protein, metabolite, or small molecule drug. Indeed, the study of molecular recognition of these protein structures by small molecules has provided well-informed paths toward drug design.[93]

One clinically relevant intrinsically disordered protein is α-synuclein, which is causative of neurodegeneration in Parkinson's disease due to its aggregation in Lewy bodies.[94] While α-synuclein protein is intrinsically disordered, the *SNCA* mRNA that encodes this pathogenic protein has a highly structured iron responsive element (IRE) in its 5′ untranslated region (UTR) that controls its translation.[95] Inforna identified a small molecule, Synucleozid, that directly binds an A bulge in the IRE, inhibiting translation in cells and providing a cytoprotective effect.[96] Direct molecular recognition of the RNA by Synucleozid was validated using antisense oligonucleotide ligand binding site mapping (ASO-Bind-Map), a small molecule application of a previously developed technique.[97] ASO-Bind-Map profiles binding sites of small molecules *in vitro* and in cells as the small molecule binds and stabilizes the target RNA's structural element, thus interfering with ASO binding and subsequent RNase H-mediated cleavage. Targeting α-synuclein upstream of its pathogenic protein demonstrates that targeting RNA elements with extensive structure can inhibit canonical translation. Importantly, affecting disease-causing intrinsically disordered proteins encoded by structured RNAs may not be limited to mRNAs with IRE and could be broadly applied to other proteins that lack binding sites for traditional "drug-like" small molecules.

The biology of RNA repeat expansion disorders presents another notable target for small molecules that bind to structured regions. From a small molecule targeting standpoint, these repeat expansions form stable repeating structural units that can be exploited for targeting, especially with on-site compound synthesis using the disease-causing RNA as a catalyst for inhibitor synthesis.[89,98] These repeats are potentially difficult to treat with ASOs that have to disrupt these stable RNA structures. Repeating transcripts are causative of >30 incurable diseases including ALS/FTD, fragile X-associated tremor ataxia syndrome (FXTAS), and DM1 and DM2.[99] Disease-causing RNA repeats can be present within coding or noncoding regions, which can affect their downstream biology. In FXTAS and DM, repeats are present in the 5′ and 3′ UTRs, respectively. ASOs sterically blocking protein interactions within the RNA repeat expansions can therefore modulate downstream biology. In the case of FXTAS, however, ASOs can also decrease downstream translation, suggesting that targeting r(CGG) repeat expansions with ASOs could worsen, rather than alleviate, disease. The repeating units, however, are amenable for small molecule targeting, as they can allow for increased specificity through the use of multivalent ligands.

In the case of myotonic dystrophy type 1, the r(CUG) repeats trigger disease by sequestering muscle-blind like protein 1 (MBNL1), which regulates the alternative splicing of various transcripts, including the insulin receptor (*IR*) and muscle-specific chloride ion channel (*Clnc1*). This gain of function by the repeats thus results in aberrant splicing products and hence DM1 disease biology. In addition to spliceopathy, other phenotypic complications arise from the expanded repeat such as aggregation of mutant transcripts with RBPs in nuclear foci, activation of stress pathways, haploinsufficiency, alteration of neighboring genes, and dysregulation of translation.[100−102] The binding affinity of MBNL1 for r(CUG) repeats has been measured in the nM range (∼3−200 nM).[103] Therefore, multivalent small molecules that bind the repeats with binding affinities also in the nM range were developed that can compete for binding with MBNL1, freeing it to resume its normal splicing functions. Thus, directly binding RNA to compete with RBPs presents a viable strategy to treat repeat expansion disorders. This strategy is in contrast to small molecules that form a ternary complex with the RNA and protein to treat SMA, as described above.

TOOLS TO ASSESS TARGET ENGAGEMENT AND SELECTIVITY OF RNA-BINDING SMALL MOLECULES

A significant amount of weight in the drug discovery process for any biomolecule is placed on *in vitro* binding affinity. While important, depending too heavily on binding studies can also be problematic, especially for dynamic RNA targets whose structure can vary *in vitro* vs *in situ*. Indeed, studies have generated ligands that interact with RNA with high affinity but display little to no bioactivity. This mismatch between

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

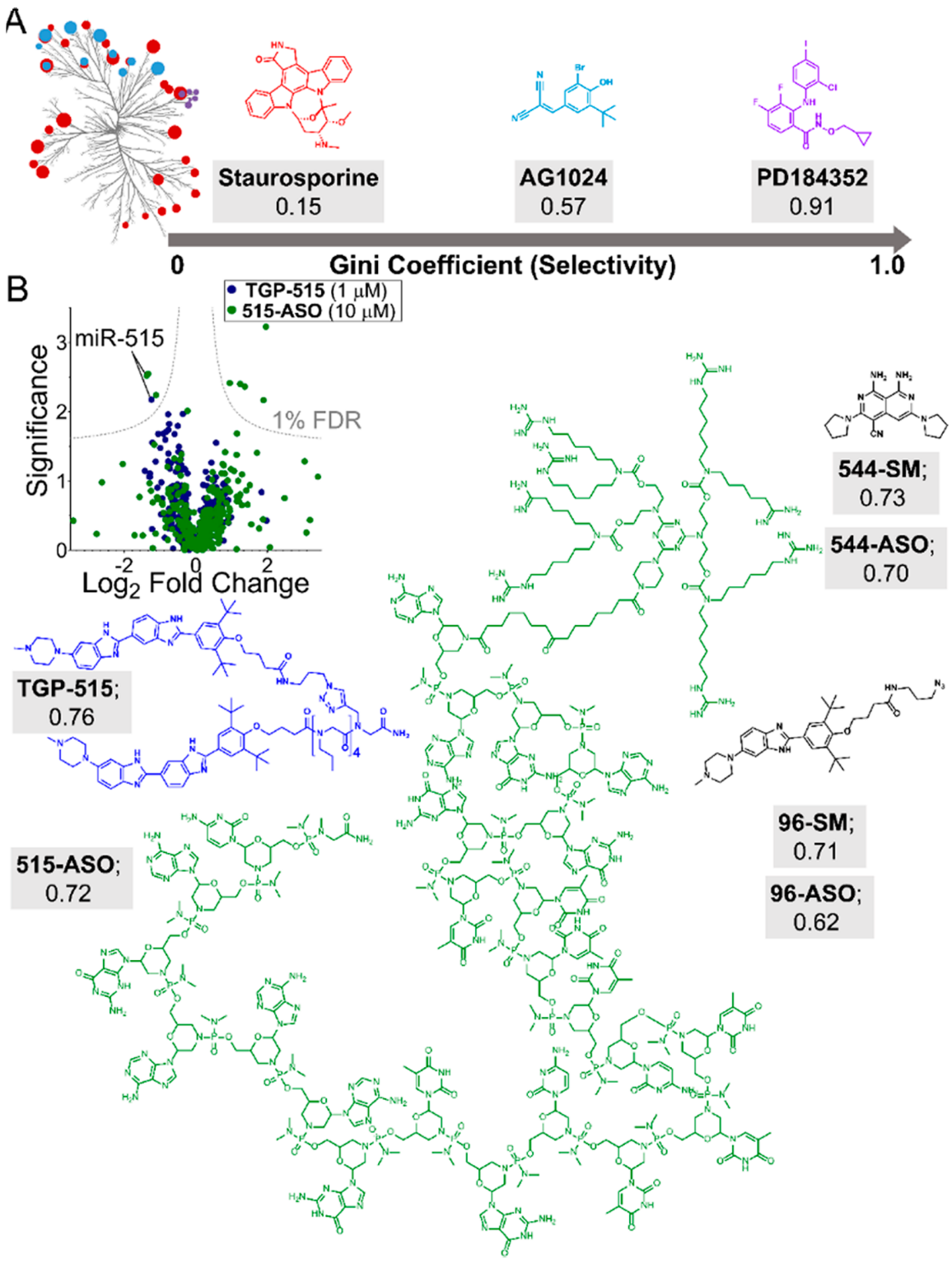


Figure 6. Quantitatively evaluating the selectivity of small molecules targeting RNA. (A) Data from profiling experiments can be used to quantify compound selectivity by calculating a Gini coefficient (GC). GC analysis of kinase inhibitors showed that promiscuous compounds (staurosporine) are characterized by values close to 0, while highly selective compounds exhibit Gini coefficient values close to 1 (PD184352; 0.91), with selective compounds being defined as >0.6. (B) GC analyses can be applied to profiling data (left), such as miRNA qPCR profiling data between small molecules (TGP-515; blue) and ASOs (515-ASO; green). Applying this analysis to various small molecule ligands targeting RNA *structure* indicates that they demonstrate high selectivity for their targets. Antisense oligonucleotides targeting RNA *sequence* are also selective for their targets. Shown are the structures and GC values of TGP-515 (blue) and Vivo-Morpholino ASO targeting miR-515 (green), in addition to other small molecule/ASO GC analyses. Overall, Gini coefficients provide a metric to quantitatively define compound selectivity. When applied to RNA targeting, GC analyses demonstrate that small molecules that recognize RNA structure can rival or exceed the selectivity of ASOs designed to bind via Watson−Crick base pairing.

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

optimized binding affinity and bioactivity could be due to many factors, including *in vitro* conditions significantly differing from *in situ* conditions (concentration/expression level, presence of other biomolecules, metal ions, etc.), *in vitro* models adopting different structures than the cellular RNA, the binding method itself, poor cellular permeability, localization of the small molecule to a cellular compartment where the target is not located, etc. Thus additional methods of analysis are required to validate target engagement (Figures 4 and 5).

Studying the *in vitro* selectivity of small molecules has been more or less standardized. Typically, binding affinities of the compound to the target RNA are quantified using various biophysical techniques (biolayer interferometry [BLI], surface plasmon resonance [SPR], NMR spectroscopy, microscale thermophoresis [MST], ESI-LC/MS, electrophoretic mobility shift assays [EMSA], etc.), that have been previously used for other biomolecules.[104−106] In addition, studying RNA-small molecule interactions has also made the use of ligand or RNA-based fluorescence and fluorescence-based displacement assays.[106,107] In these *in vitro* studies, the selectivity of ligands to the target RNA are measured relative to several controls, including point mutants of the target RNA that change its three-dimensional structure (mutating an internal loop into a base pair, for example), highly expressed RNAs (yeast tRNAs), or bulk DNA (fish sperm DNA). These initial studies provide a starting point to inform the on- and off-rate of the small molecule to bind to a specific region of the RNA in a general sense; however these assays may simplify the complexity of RNA structural dynamics in the cell, which typically exist across a population of states rather than a single explicit structure and where stability of RNA structures can be influenced by RBPs or post-transcriptional modifications.[48,54]

If we are indeed to develop guidelines for the selective recognition of RNA by small molecules, it is imperative to trace bioactivity to direct engagement of the target. Indeed, even among well-developed protein targeting modalities that have reached clinical trials, the compound action and mechanism of action remains not well-characterized, which can result in off-target effects.[108] Until recently, target engagement methods had not been developed for RNA targets. These new methods, which employ cross-linking or cleavage, have been reviewed previously.[75] In addition to target engagement, many profiling techniques are available to assess selectivity of small molecules targeting RNA, each with their own experimental advantages and liabilities.[109,110] Below, we summarize three such methods.

Quantifying expression differences of RNA transcripts has traditionally been performed through real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR). Indeed, new instrumentation and technologies have allowed for increased throughput while maintaining high sensitivity and resolution, even at the single-cell level.[111,112] The use of RT-qPCR has been commonly used to measure differential miRNA expression (Figure 5). RT-qPCR miRNA profiling revealed that small molecule hits, defined by Inforna, that target pri-miR-96 or pri-miR-515 had comparable selectivity to ASOs targeting mature miR-96-5p or mature miR-515-5p, respectively (Figures 5 and 6). Potential liabilities with measuring RNA levels with RT-qPCR include the potential biases in melting temperatures and amplification efficiency, and the labor-intensive process can be prone to user error. RT-qPCR profiling as a tool to measure the transcriptome is still useful to measure a large quantity of genes, but it is used more importantly for validation of known pathways. The method's reliance on gene specific primers potentially precludes its use as a discovery tool.

Whole transcriptome, RNA sequencing (RNA-Seq) is the emerging technology to measure transcript abundance in gene expression studies. As RNA-Seq is based on counting reads aligned to reference sequences and does not require specific probe sequences for expression measurements, it overcomes limitations encountered by RT-qPCR profiling. Additionally, the dynamic range of RNA-Seq is only limited by the read count and can be performed with a limited amount of transcript material. A potential limitation with RNA-Seq has been standards for data processing and methods for normalization and statistical analysis, where best practices are still being developed and must be completed on a case-by-case basis. Differential expression analysis from RNA-Seq can also be used to quantitatively define on- and off-targets, as well as to study the downstream effects of compound treatment, i.e., downstream pathway analysis. A recent example includes utilizing RNA-Seq at an early time point in breast cancer cells treated with a compound that inhibited miR-515 biogenesis. This analysis showed the compound's limited off-target effects on the transcriptome but also the upregulation of downstream proliferative pathways.[73]

Studying the effects on the downstream proteins takes investigations into small molecule selectivity beyond the transcriptome and can define how the modulation of RNA biology affects the proteome. Although compounds are affecting the transcriptome, proteomics can be used as a readout to measure the change in the protein landscape as a result of affecting RNA. Global proteomics, commonly used to study compounds selectively targeting proteins, or neo-protein synthesis, measuring proteins that are up- or down-regulated after compound treatment, can reveal novel biology. Indeed, the downstream neo-protein synthesis upon inhibition of miR-515 with TGP-515 revealed compound enhancement of HER2 protein expression in a normally HER2 negative cell line. Upon this realization, the levels of HER2 were able to be tuned in order to sensitize their treatment with anti-HER2 precision medicines.

Functional assays are important to verify that the compound is functioning through the predicted RNA-centric mode of action. With TGP-515, important loss- and gain-of-function assays were performed to validate the sensitivity of HER2 negative cell lines affecting miR-515 and SK1 pathways. For example, an siRNA or chemical knockdown of SK1 ablated the migratory phenotype (Figure 5) and sensitivity to HER2 therapies. Similarly, overexpression of SK1 resulted in a similar sensitivity, indicating that derepression of SK1 by miR-515 inhibition was responsible for increased sensitivity to HER2-targeted therapies. In addition to confirmation of target engagement and selectivity through techniques such as Chem-CLIP and profiling, functional assays are necessary to ascertain that the compound is operating through the hypothesized mode of action.

QUANTIFICATION OF SELECTIVITY

Quantification of compound selectivity has previously been measured with kinase inhibitors.[113] One particular metric is called a Gini coefficient, which was originally developed as a measure of statistical dispersion of wealth distribution to measure inequality.[114] More recently, the Gini coefficient has been used to measure selectivity of kinase inhibitors as it expresses a frequency distribution of differential inhibition

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

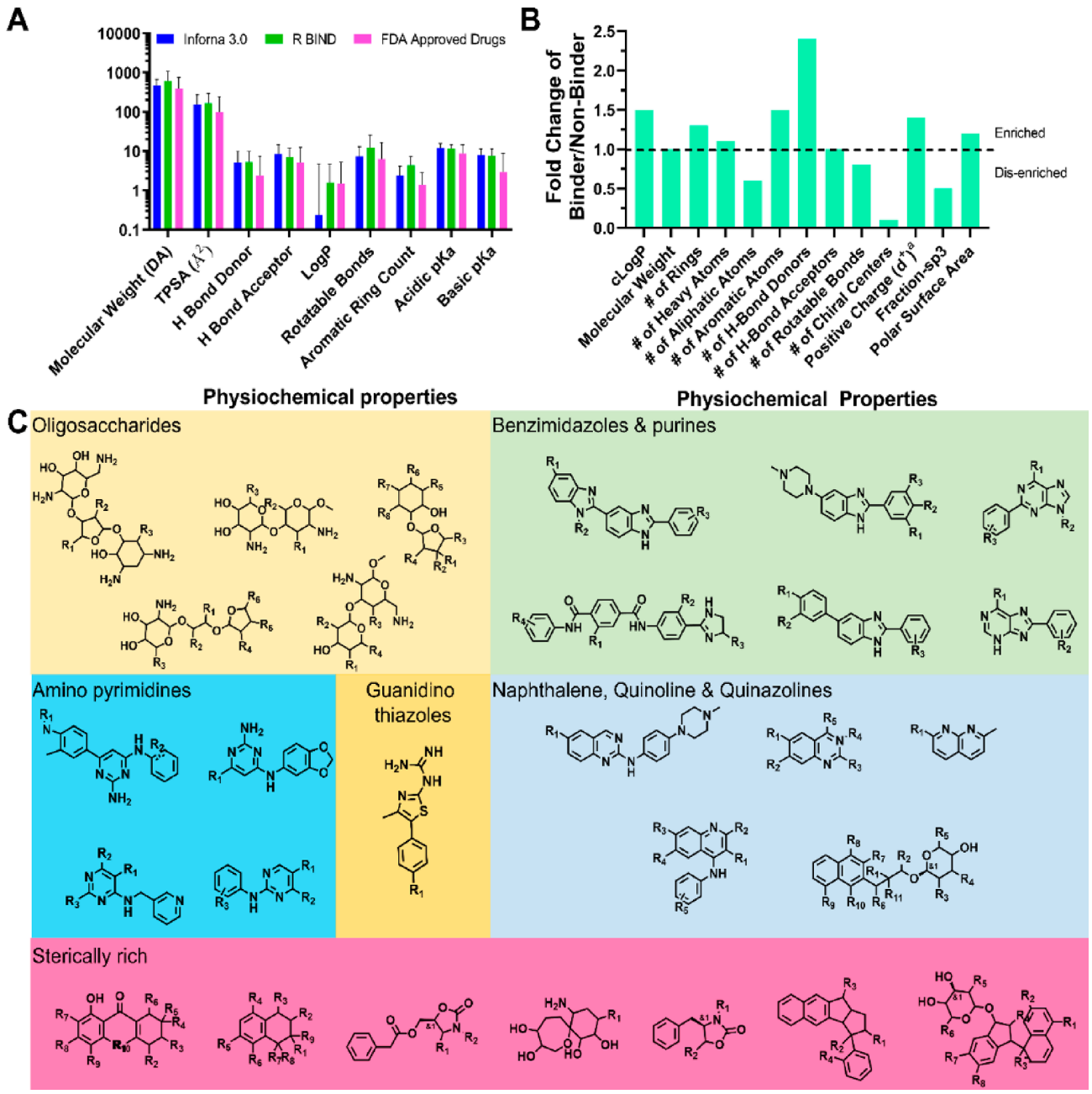


Figure 7. Physiochemical properties of RNA binders and common RNA-binding scaffolds. (A) Physiochemical properties of cataloged RNA binders contained within Inforna and R-BIND show convergence and correlate with properties of FDA approved drugs. FDA approved drugs were taken from DrugBank.[122] (B) Properties that are enriched within RNA binders include greater positive charge at pH 7.4, number of H-bond donor and acceptor counts, and total polar surface area (TPSA). RNA binders also exhibit fewer chiral centers, aliphatic atoms, and rotational bonds compared to nonbinders. (C) Scaffolds contained within Inforna and R-BIND that exhibit RNA binding. These include oligosaccharides, benzimidazoles, purines, naphthalenes, quinolines, quinazolines, aminopyrimidines, and quinidine thiazoles. Interestingly, while most data generally point to planar molecules as RNA binders, some sterically rich compounds have been found to bind RNA.

among a population of kinases. In brief, Gini coefficients range from 0, indicating a nonselective compound, to 1, indicating selectivity for a single target. For example, a kinase inhibitor that lacks selectivity (inhibits 78/85 kinases tested) had a Gini coefficient of 0.15 while inhibitors considered selective had Gini coefficients ranging from 0.69 to 0.91 (Figure 6).[114,115] Applying such an analysis to RT-qPCR profiling data from miRNAs expressed in MCF-7 cells indicated similar selectivity of a small molecule that affects pri-miR-515 and a Vivo-Morpholino modified ASO that targets miR-515-5p (Gini coefficients of 0.75 and 0.72, respectively) (Figure 6). In another analysis, the small molecule and ASO targeting miR-96 displayed comparable levels of selectivity with Gini coefficient values of 0.71 and 0.62, respectively (Figure 6). Besides qPCR profiling data, Gini coefficient analysis can be applied to larger data sets. For example, an analysis on a transcriptome-wide microarray experiment highlighted the broad selectivity of both a small molecule targeting miR-544 (Gini coefficient = 0.73) and an ASO targeting miR-544 (Gini coefficient = 0.70).[116] Interestingly, the monomeric miR-544 small molecule has

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

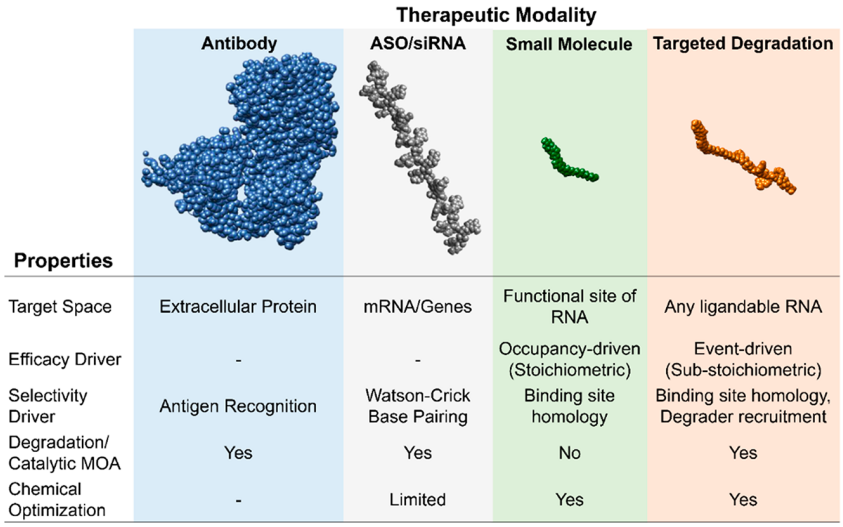


Figure 8. Properties of bioactive therapeutic modalities. Current and emerging strategies to affect downstream biology include antibodies (rituximab, PDB code 4KAQ), ASOs (Nusinersen), small molecules (TGP-210), and targeted degradation approaches (TGP-210 RIBOTAC). Each targets a unique space, but RIBOTACs can affect bioactivity of RNA without binding to a functional site. RIBOTACs also degrade their targets in a catalytic and substoichiometric fashion, thus allowing greater potency. Small-molecule-based modalities are advantageous as their physicochemical properties can be potentially medicinally optimized. ASO, small molecule, and targeted degradation models were made using the Online SMILES Generator (National Cancer Institute).

highly drug-like properties, supporting the idea that selective small molecules targeting structured RNA motifs also occupy traditional "drug-like" chemical space. Overall, in order to be able to objectively measure the selectivity of small molecules interacting with RNA, metrics such as the Gini coefficient are useful tools to compare compound selectivity. Importantly, Gini coefficient analyses have shown that the selectivity of small molecules can rival or exceed that of ASOs.

DRUGLIKENESS AND BEYOND THE RULE OF 5

Drug discovery has long used a set of physicochemical parameters to define the chemical space that is most likely to become an orally bioavailable drug. As RNA has previously been considered an "undruggable" target, then perhaps the physicochemical profiles of orally bioactive drugs may not necessarily be the best measure for RNA-binding ligands. Indeed, not all drugs are created for the same purpose, which may inaccurately bias trends toward an average that is not necessarily true for the specific applications of lead compounds.[117] For example, many small molecules differ in clearance rates and exposure to the tissue type of interest; i.e., blood−brain barrier penetrant compounds are different from systemic broad spectrum antibiotics. Therefore, the commonly used parameters that make drugs particularly attractive for oral delivery in protein targeting applications may not be appropriate for targeting RNA. In fact, there is an increasing trend of compounds that exist in a space that is extended from the Rule of 5 (eRo5) or even beyond the Rule of 5 (bRo5).[118] Indeed, as the methods to analyze the selectivity of compounds has progressively become more rigorous (full proteomics and transcriptomics profiles) and with the greater rigor for compounds with defined modes of action, the landscape of what is considered a "drug" that selectively hits a defined drug target has significantly changed.[108] Traditional ASOs, and other biologics such as antibody-drug conjugates, have also seen a rapid rise in approvals as FDA-approved drugs, suggesting that compounds that do not fit into traditional medicinal chemistry parameters are also viable as effective therapeutics.[119]

In pursuit of more selective and potent compounds, a recent trend has become known as "compound bloat". That is, drugs and potential drug candidates have been increasing in molecular weight and changing the classic idea of drug-like compounds. As mentioned by others, this "compound bloat" has been in response to decreasing off-target effects that were not previously measurable with technologies available at the time the drug was developed (CRISPR screens, RNA-Seq, proteomics).[108] Indeed, targeting multiple "hot sites" of the same target or using binding sites between two different surfaces (i.e., protein−protein interactions (PPI)) has now been implemented.[120] However, these innovative techniques have required increased molecular weight and polar surface area, along with an increased number of hydrogen bond donors and acceptors.[119]

While many drug-like compounds are heavily tested *in vitro* before *in cellulis* evaluation, binding contacts are not the only factor that influences bioactivity when concerned with drug-like properties. Furthermore, bioactivity does not necessarily indicate on-target effects, as without proper target engagement and validation, the compound may work in a different mechanism than hypothesized. Indeed, focusing on optimal physicochemical properties is important toward discovering privileged RNA space, but more innovative and novel approaches to the problem of targeting RNA with small molecules may result in compounds that look more "undrug-like" than traditional small molecule therapeutics. More provocatively, chemical matter that affects RNA space most effectively may differ significantly from traditional guidelines used to identify, design, and optimize ligands that affect other biomolecules.

Insight into the physicochemical properties of RNA-binding small molecules can begin to be gleaned from interactions housed in Inforna[79] and a recently created catalog named R-BIND.[121] Analysis of the compounds in both databases revealed highly similar chemical properties to FDA approved

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

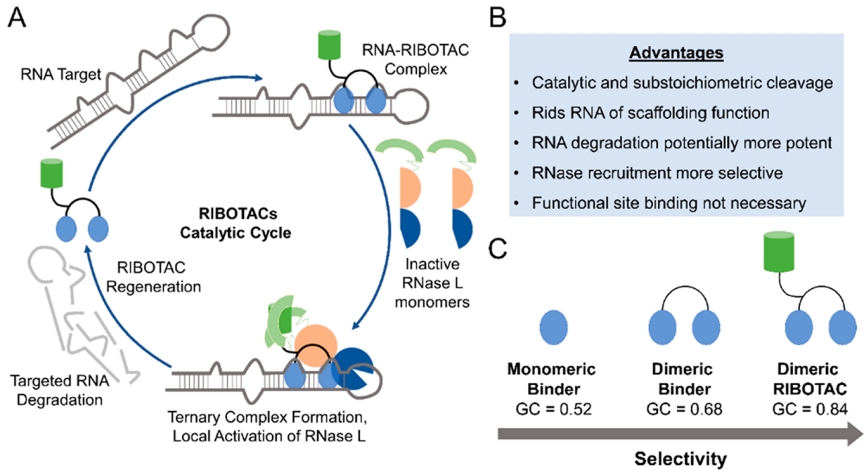


Figure 9. Ribonuclease targeting chimeras (RIBOTACs) as heterobifunctional degraders of RNA. (A) Taking cues from PROTACs and RNase H-based antisense oligonucleotide approaches, RIBOTACs are heterobifunctional compounds that recruit endogenous nucleases to degrade a targeted transcript. RIBOTACs can potentially increase potency of small molecules as they can catalytically and substoichiometrically degrade an RNA target. These RIBOTACs simply need to bind the target (not necessarily at a functional site) and use endogenous ribonuclease pathways to remove the RNA via targeted degradation, which also rids the RNA of any potential scaffolding functions with RBPs. Formation of the ternary complex may also increase selectivity as only meaningful interactions between the RNA:RIBOTAC:RNase L will result in cleavage. While this approach is potentially broadly applicable, development and optimization of both RNA binders and RNase-recruiting modules remain time-consuming, especially as there are a limited number of known RNase activators. Additionally, RIBOTACs that function through RNase L can have less pronounced effects on nuclear RNA, as RNase L is primarily cytoplasmic. (B) Advantages provided by the RIBOTAC approach. (C) Demonstration of increased selectivity of different RNA-binding modules, as indicated by GC analysis. The monomeric RNA-binding module that binds a single functional site on an RNA is less selective than the multivalent ligand targeting the same RNA. Adding an RNase L recruitment module to convert the dimeric compound into a RIBOTAC allows for increased selectivity, potentially due to the requirement of effective ternary complex formation between the RNA, RIBOTAC, and RNase L.

drugs in DrugBank (Figure 7A).[122] However, differences are indeed observed. For example, RNA binders on average have more aromatic rings (3.9 ± 1.9 vs 1.4 ± 1.4 in approved drugs) and more H bond donor (5.2 ± 4.5 vs 2.4 ± 0.5.0) and acceptor groups (8.5 ± 6.2 vs 5.1 ± 7.4).

In one large screen to identify novel compounds that bind RNA, a comparison of binding and nonbinding molecules showed that binders had on average 1.5-fold higher cLogP values, 1.3-fold more ring systems, 1.7-fold fewer aliphatic atoms, 1.5-fold more H-bond donors, 2.4-fold more H-bond acceptors, 10-fold fewer chiral centers, and slightly greater positive charge (δ^+ of 0.4 vs 0.3 at pH 7.4) (Figure 7B).[90] These data indicate a propensity for RNA binders to be structurally rigid and planar, likely driving binding of the RNA via $\pi-\pi$ stacking and hydrogen bonding interactions with the nucleobases and/or the phosphate backbone.[90]

Scaffolds commonly associated with RNA binding that have yielded bioactive ligands include benzimidazoles, *bis*-benzimidazoles and phenylbenzimidazoles, aminopyrimidines, and quinazolines (Figure 7C). Other scaffolds identified from Inforna and R-BIND include oligosaccharides, naphthalenes, and purines. While the above physiochemical data and previous works suggest a preference for planar molecules, sterically rich compounds have also been identified such as oligosaccharides and those found in Figure 7C including spirocycles and fused ring spirocycles.

EMERGING MODALITIES FOR AFFECTING RNA BIOLOGY

As RNA presents a novel and challenging target, new approaches are needed to affect its downstream biology through its direct targeting. Recent work has demonstrated the use of other modalities to enable the targeted degradation of RNA, such as bleomycin conjugates (reviewed more in-depth previously),[69,99,119,123] outside the traditional "drug-like" physicochemical properties.[124] Many compounds (described above) have rescued disease-associated phenotypes through a binding mode of action; however, it may be advantageous to expand the mode of action to cleave the RNA target, whether directly or by recruiting endogenous nucleases (Figure 8). These "RNA degraders" can potentially ameliorate disease pathologies in a more potent and selective manner than simple binding compounds.[72,89]

RNA degraders have been applied extensively to RNA repeat expansion disorders, as many of these highly stable structures can function as scaffolds to sequester proteins or to trigger repeat associated non-ATG (RAN) translation. Recently, Angelbello et al. demonstrated that RNA-degraders rescue DM1 disease pathology through intraperitoneal delivery of Cugamycin (a dimeric r(CUG)-binding compound appended to bleomycin), which selectively cleaves the repeat expansion in a preclinical mouse model.[72] In addition to its therapeutic potential, these degrader molecules can also function as effective chemical probes, as described through the development of small molecule nucleic acid profiling by cleavage applied to RNA (RiboSNAP). RiboSNAP uses the cleavage activity of bleomycin to deplete the RNA target's levels in patient-derived cells, thus demonstrating target engagement. RiboSNAP has also been applied with other noncoding RNAs to map small molecule binding sites,[75] thus establishing RNA degraders and RiboSNAP as a new potential class of targeted therapeutic and chemical probing technique.

Recently, our lab has designed small molecules that can recruit a nuclease to a specific transcript, triggering its degradation. RIBOTACs, akin to PROTACs,[125] recruit nucleases for the targeted degradation of structured RNA

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

sites (Figure 8). Specifically, RIBOTACs locally recruit endogenous latent ribonuclease (RNase L) to a specific transcript, allowing for assembly of dimeric, active RNase L to selectively degrade the RNA target. This strategy is akin to a small molecule form of CRISPR, without the need for transfection of foreign guide RNA transcripts or proteins. In contrast, RIBOTACs exploit a cell's endogenous machinery for nucleic acid disposal by bringing together all the "players" in a ternary complex (target RNA:RIBOTAC:RNase L). By locally eliminating aberrantly functioning RNA, rather than simply binding and inhibiting it, RIBOTACs offer specific advantages:

Occupancy vs Event Driven Processes. Current pharmacological paradigms require small molecules to modulate downstream function by occupying enzyme active sites or blocking receptors, which requires high concentrations of drug at the correct position. Not only can maintaining high concentrations of drug lead to off-target effects, this occupancy-driven model requires RNA with accessible, "druggable" active sites, which may be difficult to target selectively. By eliminating the aberrant RNA, rather than binding and inhibiting them, RIBOTACs circumvent the occupancy-driven model and instead are event-driven: binding enables activation of ribonucleases to initiate degradation (Figures 8 and 9).

Increased Potency and Catalytic Activity. Low levels of occupancy of RIBOTACs may be able to maintain a rate of RNA degradation that can provide the desired pharmacological effect. Previous small molecule catalytic activity was seen with Ru(bypy) photocatalysts as singlet oxygen generators,[126] and RIBOTACs have demonstrated a similar level of substoichiometric degradation of their targets.[127] This catalytic effect also suggests that RIBOTAC concentrations required for degradation may be much lower than those required for levels of inhibition by simple binding. As a catalytic degrader, RIBOTAC action also allows for greater RNA depletion over time; that is, low exposures can lead to meaningful knockdown, reducing potential off-target toxicity.

Avoiding Target Accumulation. Drug binding can potentially stabilize RNA, thereby increasing its half-life. Once drug exposure drops below a certain inhibitory level, the disease-causing RNA will then persist. Degradation avoids this dilemma entirely, as it eliminates the target RNA completely. Additionally, degradation rather than binding also rids the targeted RNA of their potential disease-causing interactions with other biomolecules, such as RBPs like MBNL1 with r(CUG) expanded repeats. Accumulation of mutations in the drug target can also occur under selective pressure, which may negatively affect drug binding and occupancy, thereby leading to a decrease in efficacy, as observed with ribocil. Interestingly, RIBOTACs do not necessarily have to target an active site to form a ternary complex and degrade the target, as even an allosteric binder can be used. Degradation of the RNA also leads to a lower risk of the target complex evolving a resistant form, but upon formation of resistance, the modularity of RIBOTACs can potentially lead to the simple solution of using an alternative target binder. As ribonucleases are highly conserved and ubiquitous among cells, resistance to RIBOTAC-induced degradation may be much lower than observed for simple binding compounds.

Gain of Specificity. Achieving good selectivity over off-targets still presents a significant challenge among small molecule inhibitors, especially in RNA targets that contain highly homologous binding sites. RIBOTACs achieve targeted degradation in two steps, first from binding of the RIBOTAC to the target and, second, upon nuclease recruitment to the target RNA (Figure 9). While the first step is mainly controlled by the affinity of the RIBOTAC to form binary complexes with its target RNA, the selectivity of the latter step can be appropriately adjusted for the RNA of interest via linker length or even by modulating the RNase recruiter. Additionally, as described above, compounds typically require binding to active sites within an RNA target to be effective; however RIBOTACs can be designed to bind structural motifs unique to a specific target RNA, thus enabling greater selectivity. Therefore, selectivity can be engineered into each component of the RIBOTAC to modulate the intrinsic binding affinity of the lone RNA-binding module to off-target RNAs. One important factor to consider for future RIBOTACs may be the role of binding kinetics (k_{on}/k_{off} rates). Indeed, with the advent of newer biophysical methods to analyze kinetics, recent studies have shown greater appreciation for on-/off-rates for drugs binding to their targets.[128] These properties are further complicated upon accounting for the optimal presentation/rate of recruitment of the RNase recruitment module and the complexities of ternary complex formation with the RNA target and RNA-binding module in the cell, which are required for RIBOTACs to propagate cleavage.

Collectively, our inaugural studies suggest (i) RIBOTACs could be a general strategy broadly applicable to imbue RNA-targeted small molecules with cleavage capability, extending mode of action beyond simple binding and (ii) selectivity, as compared to the RNA-binding compound alone, can be enhanced by conjugation of a degradation module, as also observed with nonspecific kinase inhibitors converted to a PROTAC.[129] Additionally, RIBOTACs can use guiding principles from studies into ternary complex formation and the role of kinetics for optimal activity in the PROTACs space when considering RNA-targeted degradation.[130,131] Bearing this in mind, future studies to apply the RIBOTAC technology for the druggable transcriptome may require (i) medicinal chemistry of RNA-binding small molecule, linker, and RNase L recruitment modules, especially regarding on-/off-rates; (ii) studies toward the recruitment of other ribonuclease modalities, preferably with favorable physicochemical properties; (iii) new linker chemistries to optimize presentation of the bifunctional ends; and (iv) application of RIBOTACs to *in vivo* models. Indeed, a recent study has shown that a RIBOTAC using a heterocyclic recruiter of RNase L was able to substoichiometrically degrade pre-miR-21 to impede metastasis of breast cancer to the lung in a mouse model.[132] Interestingly, selectivity and Gini coefficient analysis indicated that the RIBOTAC targeting miR-21 was more selective than the parent binding compounds, rivaling or exceeding that of protein-targeted drugs. Further downstream transcriptome and proteome-wide studies demonstrated broad inhibition of oncogenic pathways, as expected upon miR-21 depletion.[132]

CONCLUSION

As controlled RNA expression manages nearly every function in the human body, modulation of these phenomena represents an integral goal in biomedical research. As a result of chemical probes and small molecule drugs being developed for a tiny fraction of the proteome, a call to develop chemical probes (small molecules, degraders, biologics, etc.) for the entirety of the human proteome was recently announced

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

(Target 2035).[133] Indeed, this ambitious aspiration will require many groups willing to work on proteins in the "dark proteome" and the advancement and optimization of new and current technologies (DNA-encoded libraries, massively parallel high throughput screening, functional assays for understudied proteins, etc.).[134,135] Another approach to increase coverage of chemical probes would be to study compounds that modulate protein levels by affecting them at the RNA level. Not only could compounds that target RNA effectively inhibit protein expression by binding to their precursor mRNA transcript, but they also could potentially activate proteins by modifying their associated regulatory RNAs (miRNAs, lncRNAs, etc.), thus providing a bilateral approach to affect protein biology.

While effective control of RNA expression using antisense technologies has been demonstrated, these modalities are still limited in their own unique ways, leading researchers to examine if the same level of control and selectivity can be achieved using small molecules. Highlights from this Perspective have illustrated that the selective molecular recognition of RNA with small molecules is possible through the interaction of ligands with structured RNA motif regions. Bioactivity of small molecule ligands requires more than just optimizing for high affinity *in vitro* binding interactions. Furthermore, the analysis of transcriptome- and downstream proteome-wide selectivity can determine how structural recognition can be exploited for specific RNA modulation. Importantly, target engagement and validation studies are essential to verify RNA-centric modes of action and to rescue disease-associated phenotypes through an expanding toolbox of techniques. Finally, RNA is indeed druggable, although the compounds used may not look like traditional drugs, such as heterobifunctional chimeric compounds that have demonstrated efficacy in selective RNA degradation.

As opportunities arise and challenges are met for RNA-targeting small molecules, it is an exciting time to propose a druggable transcriptome project in order to provide chemical probes for functional RNAs on a transcriptome-wide scale. Developing technologies, such as RIBOTACs, and currently existing methodologies, such as Inforna, can be used in conjunction to study and optimize small molecules targeting all functionally relevant RNAs broadly. As advancements and innovations in the design and discovery small molecules interacting with RNA continue, the use of these techniques can bring about a new paradigm in chemical biology and potentially launch research into new therapeutic modalities and viable medicines.

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

Matthew G. Costales − *Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States*

Jessica L. Childs-Disney − *Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States*

Hafeez S. Haniff − *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.9b01927

Notes

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

Biographies

Matthew G. Costales completed his B.S. in Chemistry at the University of Maryland, College Park. He spent the following years as an Oak Ridge Institute for Science and Education Fellow in the Immunobiology Group in the Center for Food Safety and Applied Nutrition at the Food and Drug Administration. He then received his Ph.D. in Chemical Biology under the supervision of Prof. Matthew D. Disney at the Scripps Research Institute in Florida for his dissertation research on strategies to selectively target RNA with small molecules. Currently, he is working in the Neuro Discovery group at Dicerna Pharmaceuticals.

Jessica L. Childs-Disney attended Messiah College where she received her B.S. in Chemistry. She then completed her Ph.D. in Biophysical Chemistry, studying how to induce RNA misfolding under the tutelage of Prof. Douglas H. Turner, and postdoctoral studies at the Swiss Federal Institute of Technology (ETH) Zurich. After spending 4 years as an Assistant Professor at Canisius College, she moved to The Scripps Research Institute in Florida where she is a senior staff scientist.

Hafeez S. Haniff finished his B.S. in Chemistry at the University of Central Florida. He is currently completing his Ph.D. in the lab of Prof. Matthew D. Disney at the Scripps Research Institute in Florida. His dissertation research is focused on the development and application of RNA-focused small molecules libraries to modulate the function of RNA-based diseases for regenerative medicine.

Matthew D. Disney received his B.S. in Chemistry from the University of Maryland, College Park. He then completed his M.S. and Ph.D. in Biophysical Chemistry with Prof. Douglas H. Turner. He completed postdoctoral studies in the lab of Prof. Peter H. Seeberger at the Massachusetts Institute of Technology and the Swiss Federal Institute of Technology (ETH) Zurich. He started his independent research career and eventually joined the faculty in the bicoastal Department of Chemistry at The Scripps Research Institute in Florida, where he is Full Professor. His group focuses on developing small molecules to selectively drug human RNAs. His research has also led to the founding of Expansion Therapeutics.

ACKNOWLEDGMENTS

We thank the members of our laboratory that have helped advance the science of small molecules interacting with RNA, especially human RNAs, over the past 15 years. We thank the funding agencies (National Institutes of Health Grants R01 GM97455, R01 CA249180, P01 NS099114, and DP1 NS096898 to M.D.D.; the Department of Defense Peer Reviewed Medical Research Program Grant W81XWH-18-0718 to M.D.D.; ACS Medicinal Chemistry Predoctoral Fellowship to M.G.C.) that have supported our work in this area.

ABBREVIATIONS USED

2′-5′ A, 2′-5′ linked oligoadenylate; 2DCS, two-dimensional combinatorial screening; AGO/RISC, argonaute/RNA-induced silencing complex; ASO, antisense oligonucleotide; ASO-Bind-Map, antisense oligonucleotide ligand binding site mapping; BLI, biolayer interferometry; bRo5, beyond rule of 5; C-Chem-CLIP, competitive cross-linking and isolation by

pulldown; Chem-CLIP, chemical cross-linking and isolation by pulldown; CRISPR, clustered regularly interspaced short palindromic repeats; DM1/DM2, myotonic dystrophy type I/II; DMD, Duchenne's muscular dystrophy; EMSA, electrophoretic mobility shift assay; ENCODE, Encyclopedia of DNA Elements; eRo5, extended rule of 5; ESE, exonic splicing enhancer; ESI-LC/MS, electrospray ionization liquid chromatography/mass spectrometry; FMN, flavin mononucleotide; FTD/ALS, frontotemporal dementia/amyotrophic lateral sclerosis; FXTAS, fragile X-associated tremor ataxia syndrome; GC, Gini coefficient; HER2, human epidermal growth factor receptor 2; HiT-StARTS, high throughput structure−activity relationships through sequencing; IDP, intrinsically disordered protein; IRE, iron responsive element; IRP, iron responsive protein; KRAS, Kirsten rat sarcoma proto-oncogene; lncRNA, long noncoding RNA; MBNL1, muscle-blind like protein 1; miR, microRNA; MST, microscale thermophoresis; PPI, protein−protein interactions; pre-miR, precursor microRNA hairpin; pre-mRNA, precursor messenger RNA; pri-miR, primary microRNA transcript; PROTAC, proteolysis targeting chimera; RiboSNAP, small-molecule nucleic acid profiling by cleavage applied to RNA; RIBOTAC, ribonuclease targeting chimera; RNase, ribonuclease; RNA-Seq, RNA sequencing; RT-qPCR, real time reverse transcription quantitative polymerase chain reaction; S1P, sphingosine 1-phosphate; shRNA, short hairpin RNA; siRNA, short interfering RNA; SK1, sphingosine kinase 1; SMA, spinomuscular atrophy; SMN, survival motor neuron protein; SPR, surface plasmon resonance; TGP-210, targapremir-210; TGP-515, targaprimir-515; U1 snRNA, U1 small nuclear RNA; UTR, untranslated region; VEGFA, vascular endothelial growth factor A

REFERENCES

(1) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, Y.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F.; Stupka, E.; Szustakowki, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; de Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y. J.; Szustakowki, J. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860−921.

(2) Anderson, N. L.; Anderson, N. G. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **1998**, *19*, 1853−1861.

(3) Bertozzi, C. R.; Sasisekharan, R. Glycomics. In *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E., Eds.; Cold Spring Harbor Laboratory Press: New York, 2009.

(4) Ejsing, C. S.; Sampaio, J. L.; Surendranath, V.; Duchoslav, E.; Ekroos, K.; Klemm, R. W.; Simons, K.; Shevchenko, A. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2136−2141.

(5) McGettigan, P. A. Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* **2013**, *17*, 4−11.

(6) Yang, Y.; Adelstein, S. J.; Kassis, A. I. Target discovery from data mining approaches. *Drug Discovery Today* **2012**, *17*, S16−S23.

(7) 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.

(8) Djebali, S.; Davis, C. A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; Xue, C.; Marinov, G. K.; Khatun, J.; Williams, B. A.; Zaleski, C.; Rozowsky, J.; Röder, M.; Kokocinski, F.; Abdelhamid, R. F.; Alioto, T.; Antoshechkin, I.; Baer, M. T.; Bar, N. S.; Batut, P.; Bell, K.; Bell, I.; Chakrabortty, S.; Chen, X.; Chrast, J.; Curado, J.; Derrien, T.; Drenkow, J.; Dumais, E.; Dumais, J.; Duttagupta, R.; Falconnet, E.; Fastuca, M.; Fejes-Toth, K.; Ferreira, P.; Foissac, S.; Fullwood, M. J.; Gao, H.; Gonzalez, D.; Gordon, A.; Gunawardena, H.; Howald, C.; Jha, S.; Johnson, R.; Kapranov, P.; King, B.; Kingswood, C.; Luo, O. J.; Park, E.; Persaud, K.; Preall, J. B.; Ribeca, P.; Risk, B.; Robyr, D.; Sammeth, M.; Schaffer, L.; See, L.-H.; Shahab, A.; Skancke, J.; Suzuki, A. M.; Takahashi, H.; Tilgner, H.; Trout, D.; Walters, N.; Wang, H.; Wrobel, J.; Yu, Y.; Ruan, X.; Hayashizaki, Y.; Harrow, J.; Gerstein, M.; Hubbard, T.; Reymond, A.; Antonarakis, S. E.; Hannon, G.; Giddings, M. C.; Ruan, Y.; Wold, B.; Carninci, P.; Guigó, R.; Gingeras, T. R. Landscape of transcription in human cells. *Nature* **2012**, *489*, 101−108.

(9) Wade, J. T.; Grainger, D. C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.* **2014**, *12*, 647−653.

(10) Liu, G.; Mattick, J. S.; Taft, R. J. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle* **2013**, *12*, 2061−2072.

(11) Necsulea, A.; Soumillon, M.; Warnefors, M.; Liechti, A.; Daish, T.; Zeller, U.; Baker, J. C.; Grützner, F.; Kaessmann, H. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **2014**, *505*, 635−640.

(12) Makley, L. N.; Gestwicki, J. E. Expanding the number of "druggable" targets: non-enzymes and protein-protein interactions. *Chem. Biol. Drug Des.* **2013**, *81*, 22−32.

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

(13) Limbird, L. E. The receptor concept: a continuing evolution. *Mol. Interventions* **2004**, *4*, 326−336.
(14) Lazo, J. S.; Sharlow, E. R. Drugging undruggable molecular cancer targets. *Annu. Rev. Pharmacol. Toxicol.* **2016**, *56*, 23−40.
(15) Dang, C. V.; Reddy, E. P.; Shokat, K. M.; Soucek, L. Drugging the 'undruggable' cancer targets. *Nat. Rev. Cancer* **2017**, *17*, 502−508.
(16) Ostrem, J. M. L.; Shokat, K. M. Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. *Nat. Rev. Drug Discovery* **2016**, *15*, 771−785.
(17) Kessler, D.; Gmachl, M.; Mantoulidis, A.; Martin, L. J.; Zoephel, A.; Mayer, M.; Gollner, A.; Covini, D.; Fischer, S.; Gerstberger, T.; Gmaschitz, T.; Goodwin, C.; Greb, P.; Häring, D.; Hela, W.; Hoffmann, J.; Karolyi-Oezguer, J.; Knesl, P.; Kornigg, S.; Koegl, M.; Kousek, R.; Lamarre, L.; Moser, F.; Munico-Martinez, S.; Peinsipp, C.; Phan, J.; Rinnenthal, J.; Sai, J.; Salamon, C.; Scherbantin, Y.; Schipany, K.; Schnitzer, R.; Schrenk, A.; Sharps, B.; Siszler, G.; Sun, Q.; Waterson, A.; Wolkerstorfer, B.; Zeeb, M.; Pearson, M.; Fesik, S. W.; McConnell, D. B. Drugging an undruggable pocket on KRAS. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 15823−15829.
(18) Lou, K.; Steri, V.; Ge, A. Y.; Hwang, Y. C.; Yogodzinski, C. H.; Shkedi, A. R.; Choi, A. L. M.; Mitchell, D. C.; Swaney, D. L.; Hann, B.; Gordan, J. D.; Shokat, K. M.; Gilbert, L. A. KRASG12C inhibition produces a driver-limited state revealing collateral dependencies. *Sci. Signaling* **2019**, *12*, No. eaaw9450.
(19) Pettersson, M.; Crews, C. M. PROteolysis TArgeting Chimeras (PROTACs)—past, present and future. *Drug Discovery Today: Technol.* **2019**, *31*, 15−27.
(20) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; Gonzalez-Paez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. Proteome-wide covalent ligand discovery in native biological systems. *Nature* **2016**, *534*, 570−574.
(21) Parker, C. G.; Galmozzi, A.; Wang, Y.; Correia, B. E.; Sasaki, K.; Joslyn, C. M.; Kim, A. S.; Cavallaro, C. L.; Lawrence, R. M.; Johnson, S. R.; Narvaiza, I.; Saez, E.; Cravatt, B. F. Ligand and target discovery by fragment-based screening in human cells. *Cell* **2017**, *168*, 527−541.
(22) O'Rourke, J. R.; Swanson, M. S. Mechanisms of RNA-mediated disease. *J. Biol. Chem.* **2009**, *284*, 7419−7423.
(23) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and disease. *Cell* **2009**, *136*, 777−793.
(24) Sztuba-Solinska, J.; Chavez-Calvillo, G.; Cline, S. E. Unveiling the druggable RNA targets and small molecule therapeutics. *Bioorg. Med. Chem.* **2019**, *27*, 2149−2165.
(25) Ursu, A.; Vezina-Dawod, S.; Disney, M. D. Methods to identify and optimize small molecules interacting with RNA (SMIRNAs). *Drug Discovery Today* **2019**, *24*, 2002−2016.
(26) Rizvi, N. F.; Smith, G. F. RNA as a small molecule druggable target. *Bioorg. Med. Chem. Lett.* **2017**, *27*, 5083−5088.
(27) Di Giorgio, A.; Duca, M. Synthetic small-molecule RNA ligands: future prospects as therapeutic agents. *MedChemComm* **2019**, *10*, 1242−1255.
(28) Morgan, B. S.; Forte, J. E.; Hargrove, A. E. Insights into the development of chemical probes for RNA. *Nucleic Acids Res.* **2018**, *46*, 8025−8037.
(29) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discovery* **2018**, *17*, 547−558.
(30) Begg, E. J.; Barclay, M. L. Aminoglycosides - 50 years on. *Br. J. Clin. Pharmacol.* **1995**, *39*, 597−603.
(31) Disney, M. D.; Labuda, L. P.; Paul, D. J.; Poplawski, S. G.; Pushechnikov, A.; Tran, T.; Velagapudi, S. P.; Wu, M.; Childs-Disney, J. L. Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. *J. Am. Chem. Soc.* **2008**, *130*, 11185−11194.
(32) Datlinger, P.; Rendeiro, A. F.; Schmidl, C.; Krausgruber, T.; Traxler, P.; Klughammer, J.; Schuster, L. C.; Kuchler, A.; Alpar, D.; Bock, C. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **2017**, *14*, 297−301.
(33) Moffat, J.; Sabatini, D. M. Building mammalian signalling pathways with RNAi screens. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 177−187.
(34) Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; de Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T. Selective small-molecule inhibition of an RNA structural element. *Nature* **2015**, *526*, 672−677.
(35) DeFrancesco, L. Drug pipeline: 1Q17. *Nat. Biotechnol.* **2017**, *35*, 400.
(36) McDonald, C. M.; Campbell, C.; Torricelli, R. E.; Finkel, R. S.; Flanigan, K. M.; Goemans, N.; Heydemann, P.; Kaminska, A.; Kirschner, J.; Muntoni, F.; Osorio, A. N.; Schara, U.; Sejersen, T.; Shieh, P. B.; Sweeney, H. L.; Topaloglu, H.; Tulinius, M.; Vilchez, J. J.; Voit, T.; Wong, B.; Elfring, G.; Kroger, H.; Luo, X.; McIntosh, J.; Ong, T.; Riebling, P.; Souza, M.; Spiegel, R. J.; Peltz, S. W.; Mercuri, E. Ataluren in patients with nonsense mutation Duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **2017**, *390*, 1489−1498.
(37) Auld, D. S.; Thorne, N.; Maguire, W. F.; Inglese, J. Mechanism of PTC124 activity in cell-based luciferase assays of nonsense codon suppression. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3585−3590.
(38) Auld, D. S.; Lovell, S.; Thorne, N.; Lea, W. A.; Maloney, D. J.; Shen, M.; Rai, G.; Battaile, K. P.; Thomas, C. J.; Simeonov, A.; Hanzlik, R. P.; Inglese, J. Molecular basis for the high-affinity binding and stabilization of firefly luciferase by PTC124. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4878−4883.
(39) Rao, V. K.; Kapp, D.; Schroth, M. Gene therapy for spinal muscular atrophy: an emerging treatment option for a devastating disease. *J. Manag. Care Spec. Pharm.* **2018**, *24*, S3−S16.
(40) Cheung, A. K.; Hurley, B.; Kerrigan, R.; Shu, L.; Chin, D. N.; Shen, Y.; O'Brien, G.; Sung, M. J.; Hou, Y.; Axford, J.; Cody, E.; Sun, R.; Fazal, A.; Fridrich, C.; Sanchez, C. C.; Tomlinson, R. C.; Jain, M.; Deng, L.; Hoffmaster, K.; Song, C.; Van Hoosear, M.; Shin, Y.; Servais, R.; Towler, C.; Hild, M.; Curtis, D.; Dietrich, W. F.; Hamann, L. G.; Briner, K.; Chen, K. S.; Kobayashi, D.; Sivasankaran, R.; Dales, N. A. Discovery of small molecule splicing modulators of survival motor neuron-2 (SMN2) for the treatment of spinal muscular atrophy (SMA). *J. Med. Chem.* **2018**, *61*, 11021−11036.
(41) Charnas, L.; Voltz, E.; Pfister, C.; Peters, T.; Hartmann, A.; Berghs-Clairmont, C.; Praestgaard, J.; de Raspide, M.; Deconinck, N.; Born, A.; Baranello, G.; Bertini, E.; Schara, U.; Goemans, N.; Roubenoff, R. Safety and efficacy findings in the first-in-human trial (FIH) of the oral splice modulator branaplam in type 1 spinal muscular atrophy (SMA): interim results. *Neuromuscul. Disord.* **2017**, *27*, S207−S208.
(42) Wang, J.; Schultz, P. G.; Johnson, K. A. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4604−E4612.
(43) Sivaramakrishnan, M.; McCarthy, K. D.; Campagne, S.; Huber, S.; Meier, S.; Augustin, A.; Heckel, T.; Meistermann, H.; Hug, M. N.; Birrer, P.; Moursy, A.; Khawaja, S.; Schmucki, R.; Berntenis, N.; Giroud, N.; Golling, S.; Tzouros, M.; Banfai, B.; Duran-Pacheco, G.; Lamerz, J.; Hsiu Liu, Y.; Luebbers, T.; Ratni, H.; Ebeling, M.; Cléry, A.; Paushkin, S.; Krainer, A. R.; Allain, F. H. T.; Metzger, F. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun.* **2017**, *8*, 1476.
(44) Yang, W.-Y.; Wilson, H. D.; Velagapudi, S. P.; Disney, M. D. Inhibition of non-ATG translational events in cells via covalent small molecules targeting RNA. *J. Am. Chem. Soc.* **2015**, *137*, 5336−5345.
(45) Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquisee, M.; Merrill, S. H.; Penswick, J. R.; Zamir, A. Structure of a ribonucleic acid. *Science* **1965**, *147*, 1462−1465.
(46) Rich, A.; RajBhandary, U. L. Transfer RNA: molecular structure, sequence, and properties. *Annu. Rev. Biochem.* **1976**, *45*, 805−860.

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

(47) Lerner, M. R.; Boyle, J. A.; Mount, S. M.; Wolin, S. L.; Steitz, J. A. Are snRNPs involved in splicing? *Nature* **1980**, *283*, 220−224.
(48) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 474−489.
(49) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* **2009**, *37*, 2294−2312.
(50) Yesselman, J. D.; Denny, S. K.; Bisaria, N.; Herschlag, D.; Greenleaf, W. J.; Das, R. Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 16847−16855.
(51) Denny, S. K.; Bisaria, N.; Yesselman, J. D.; Das, R.; Herschlag, D.; Greenleaf, W. J. High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell* **2018**, *174*, 377−390.
(52) Kwok, C. K.; Tang, Y.; Assmann, S. M.; Bevilacqua, P. C. The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem. Sci.* **2015**, *40*, 221−232.
(53) Fallmann, J.; Will, S.; Engelhardt, J.; Grüning, B.; Backofen, R.; Stadler, P. F. Recent advances in RNA folding. *J. Biotechnol.* **2017**, *261*, 97−104.
(54) Bevilacqua, P. C.; Ritchey, L. E.; Su, Z.; Assmann, S. M. Genome-wide analysis of RNA secondary structure. *Annu. Rev. Genet.* **2016**, *50*, 235−266.
(55) Hasegawa, M.; Iida, Y.; Yano, T.-a.; Takaiwa, F.; Iwabuchi, M. Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. *J. Mol. Evol.* **1985**, *22*, 32−38.
(56) Turner, D. H.; Mathews, D. H. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.* **2010**, *38*, D280−D282.
(57) Zuker, M. On finding all suboptimal foldings of an RNA molecule. *Science* **1989**, *244*, 48−52.
(58) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7287−7292.
(59) Rouskin, S.; Zubradt, M.; Washietl, S.; Kellis, M.; Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **2014**, *505*, 701−705.
(60) Zubradt, M.; Gupta, P.; Persad, S.; Lambowitz, A. M.; Weissman, J. S.; Rouskin, S. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **2017**, *14*, 75−82.
(61) Rivas, E.; Clements, J.; Eddy, S. R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **2017**, *14*, 45−48.
(62) Andrews, R. J.; Roche, J.; Moss, W. N. ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV. *PeerJ* **2018**, *6*, No. e6136.
(63) He, L.; Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **2004**, *5*, 522−531.
(64) Bernat, V.; Disney, M. D. RNA structures as mediators of neurological diseases and as drug targets. *Neuron* **2015**, *87*, 28−46.
(65) Bennett, C. F. Therapeutic antisense oligonucleotides are coming of age. *Annu. Rev. Med.* **2019**, *70*, 307−321.
(66) Hagedorn, P. H.; Hansen, B. R.; Koch, T.; Lindow, M. Managing the sequence-specificity of antisense oligonucleotides in drug discovery. *Nucleic Acids Res.* **2017**, *45*, 2262−2282.
(67) Chi, X.; Gatti, P.; Papoian, T. Safety of antisense oligonucleotide and siRNA-based therapeutics. *Drug Discovery Today* **2017**, *22*, 823−833.
(68) Frazier, K. S. Antisense oligonucleotide therapies: the promise and the challenges from a toxicologic pathologist's perspective. *Toxicol. Pathol.* **2015**, *43*, 78−89.
(69) Valeur, E.; Jimonet, P. New modalities, technologies, and partnerships in probe and lead generation: Enabling a mode-of-action centric paradigm. *J. Med. Chem.* **2018**, *61*, 9004−9029.
(70) Herschlag, D. Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules in vivo: more isn't always better. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 6921−6925.
(71) Pedersen, L.; Hagedorn, P. H.; Lindholm, M. W.; Lindow, M. A kinetic model explains why shorter and less affine enzyme-recruiting oligonucleotides can be more potent. *Mol. Ther.–Nucleic Acids* **2014**, *3*, No. e149.
(72) 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 a r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 7799−7804.
(73) Costales, M. G.; Hoch, D. G.; Abegg, D.; Childs-Disney, J. L.; Velagapudi, S. P.; Adibekian, A.; Disney, M. D. A designed small molecule inhibitor of a non-coding RNA sensitizes HER2 negative cancer to Herceptin. *J. Am. Chem. Soc.* **2019**, *141*, 2960−2974.
(74) Costales, M. G.; Rzuczek, S. G.; Disney, M. D. Comparison of small molecules and oligonucleotides that target a toxic, non-coding RNA. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 2605−2609.
(75) Disney, M. D.; Velagapudi, S. P.; Li, Y.; Costales, M. G.; Childs-Disney, J. L. Identifying and validating small molecules interacting with RNA (SMIRNAs). *Methods Enzymol.* **2019**, *623*, 45−66.
(76) 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 anticancer drugs target oncogenic non-coding RNAs. *Cell Chem. Biol.* **2018**, *25*, 1086−1094.
(77) Velagapudi, S. P.; Luo, Y.; Tran, T.; Haniff, H. S.; Nakai, Y.; Fallahi, M.; Martinez, G. J.; Childs-Disney, J. L.; Disney, M. D. 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.
(78) Velagapudi, S. P.; Gallo, S. M.; Disney, M. D. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol.* **2014**, *10*, 291−297.
(79) Disney, M. D.; Winkelsas, A. M.; Velagapudi, S. P.; Southern, M.; Fallahi, M.; Childs-Disney, J. L. Inforna 2.0: A platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol.* **2016**, *11*, 1720−1728.
(80) Costales, M. G.; Haga, C. L.; Velagapudi, S. P.; Childs-Disney, J. L.; Phinney, D. G.; Disney, M. D. Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit. *J. Am. Chem. Soc.* **2017**, *139*, 3446−3455.
(81) Huang, X.; Ding, L.; Bennewith, K. L.; Tong, R. T.; Welford, S. M.; Ang, K. K.; Story, M.; Le, Q.-T.; Giaccia, A. J. Hypoxia-inducible mir-210 regulates normoxic gene expression involved in tumor initiation. *Mol. Cell* **2009**, *35*, 856−867.
(82) Qin, Q.; Furong, W.; Baosheng, L. Multiple functions of hypoxia-regulated miR-210 in cancer. *J. Exp. Clin. Cancer Res.* **2014**, *33*, 50.
(83) Kelly, T. J.; Souza, A. L.; Clish, C. B.; Puigserver, P. A hypoxia-induced positive feedback loop promotes hypoxia-inducible factor 1alpha stability through miR-210 suppression of glycerol-3-phosphate dehydrogenase 1-like. *Mol. Cell. Biol.* **2011**, *31*, 2696−2706.
(84) Johnson, L. F.; Abelson, H. T.; Penman, S.; Green, H. The relative amounts of the cytoplasmic RNA species in normal, transformed and senescent cultured cell lines. *J. Cell. Physiol.* **1977**, *90*, 465−470.
(85) Liu, B.; Childs-Disney, J. L.; Znosko, B. M.; Wang, D.; Fallahi, M.; Gallo, S. M.; Disney, M. D. Analysis of secondary structural elements in human microRNA hairpin precursors. *BMC Bioinf.* **2016**, *17*, 112.
(86) Velagapudi, S. P.; Seedhouse, S. J.; French, J.; Disney, M. D. Defining the RNA internal loops preferred by benzimidazole derivatives via 2D combinatorial screening and computational analysis. *J. Am. Chem. Soc.* **2011**, *133*, 10111−10118.
(87) Disney, M. D.; Lee, M. M.; Pushechnikov, A.; Childs-Disney, J. L. The role of flexibility in the rational design of modularly assembled

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

ligands targeting the RNAs that cause the myotonic dystrophies. *ChemBioChem* **2010**, *11*, 375−382.
(88) Velagapudi, S. P.; Cameron, M. D.; Haga, C. L.; Rosenberg, L. H.; Lafitte, M.; Duckett, D. R.; Phinney, D. G.; Disney, M. D. Design of a small molecule against an oncogenic noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 5898−5903.
(89) 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.
(90) Haniff, H. S.; Knerr, L.; Liu, X.; Crynen, G.; Boström, J.; Abegg, D.; Adibekian, A.; Lekah, E.; Wang, K.-W.; Camerson, M. D.; Yildirm, I.; Lemurell, M.; Disney, M. D. Design of a small molecule that stimulates VEGFA informed from an expanded encyclopedia of RNA fold-small molecule interactions. *ChemRxiv* **2020**, DOI: 10.26434/chemrxiv.7967291.v1.
(91) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114*, 6589−6631.
(92) Uversky, V. N. Intrinsically disordered proteins and their "mysterious" (meta)physics. *Front. Phys.* **2019**, *7*, 10.
(93) Ruan, H.; Sun, Q.; Zhang, W.; Liu, Y.; Lai, L. Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discovery Today* **2019**, *24*, 217−227.
(94) Kim, W. S.; Kågedal, K.; Halliday, G. M. Alpha-synuclein biology in Lewy body diseases. *Alzheimer's Res. Ther.* **2014**, *6*, 73.
(95) Cahill, C. M.; Lahiri, D. K.; Huang, X.; Rogers, J. T. Amyloid precursor protein and alpha synuclein translation, implications for iron and inflammation in neurodegenerative diseases. *Biochim. Biophys. Acta, Gen. Subj.* **2009**, *1790*, 615−628.
(96) Zhang, P.; Park, H.-J.; Zhang, J.; Junn, E.; Andrews, R. J.; Velagapudi, S. P.; Abegg, D.; Vishnu, K.; Costales, M. G.; Childs-Disney, J. L.; Adibekian, A.; Moss, W. N.; Mouradian, M. M.; Disney, M. D. Translation of the intrinsically disordered protein α-synuclein is inhibited by a small molecule targeting its structured mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 1457−1467.
(97) Zarrinkar, P. P.; Williamson, J. R. Kinetic intermediates in RNA folding. *Science* **1994**, *265*, 918−924.
(98) Rzuczek, S. G.; Park, H.; Disney, M. D. A toxic RNA catalyzes the in cellulo synthesis of its own inhibitor. *Angew. Chem., Int. Ed. Engl.* **2014**, *53*, 10956−10959.
(99) 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.
(100) Mateos-Aierdi, A. J.; Goicoechea, M.; Aiastui, A.; Fernández-Torrón, R.; Garcia-Puga, M.; Matheu, A.; López de Munain, A. Muscle wasting in myotonic dystrophies: a model of premature aging. *Front. Aging Neurosci.* **2015**, *7*, 125.
(101) Meola, G.; Cardani, R. Myotonic dystrophies: An update on clinical aspects, genetic, pathology, and molecular pathomechanisms. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2015**, *1852*, 594−606.
(102) Udd, B.; Krahe, R. The myotonic dystrophies: molecular, clinical, and therapeutic challenges. *Lancet Neurol.* **2012**, *11*, 891−905.
(103) 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.
(104) Moon, M. H.; Hilimire, T. A.; Sanders, A. M.; Schneekloth, J. S. Measuring RNA-ligand interactions with microscale thermophoresis. *Biochemistry* **2018**, *57*, 4638−4643.
(105) Umuhire Juru, A.; Patwardhan, N. N.; Hargrove, A. E. Understanding the contributions of conformational changes, thermodynamics, and kinetics of RNA-small molecule interactions. *ACS Chem. Biol.* **2019**, *14*, 824−838.
(106) Blakeley, B. D.; DePorter, S. M.; Mohan, U.; Burai, R.; Tolbert, B. S.; McNaughton, B. R. Methods for identifying and characterizing interactions involving RNA. *Tetrahedron* **2012**, *68*, 8837−8855.
(107) Wicks, S. L.; Hargrove, A. E. Fluorescent indicator displacement assays to identify and characterize small molecule interactions with RNA. *Methods* **2019**, *167*, 3−14.
(108) Lin, A.; Giuliano, C. J.; Palladino, A.; John, K. M.; Abramowicz, C.; Yuan, M. L.; Sausville, E. L.; Lukow, D. A.; Liu, L.; Chait, A. R.; Galluzzo, Z. C.; Tucker, C.; Sheltzer, J. M. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **2019**, *11*, No. eaaw8412.
(109) Lowe, R.; Shirley, N.; Bleackley, M.; Dolan, S.; Shafee, T. Transcriptomics technologies. *PLoS Comput. Biol.* **2017**, *13*, No. e1005457.
(110) Git, A.; Dvinge, H.; Salmon-Divon, M.; Osborne, M.; Kutter, C.; Hadfield, J.; Bertone, P.; Caldas, C. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* **2010**, *16*, 991−1006.
(111) Bittker, J. A. High-throughput RT-PCR for small-molecule screening assays. *Curr. Protoc. Chem. Biol.* **2012**, *4*, 49−63.
(112) Citri, A.; Pang, Z. P.; Südhof, T. C.; Wernig, M.; Malenka, R. C. Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat. Protoc.* **2012**, *7*, 118−127.
(113) Bosc, N.; Meyer, C.; Bonnet, P. The use of novel selectivity metrics in kinase research. *BMC Bioinf.* **2017**, *18*, 17.
(114) Graczyk, P. P. Gini coefficient: a new way to express selectivity of kinase inhibitors against a family of kinases. *J. Med. Chem.* **2007**, *50*, 5773−5779.
(115) Drewry, D. H.; Wells, C. I.; Andrews, D. M.; Angell, R.; Al-Ali, H.; Axtman, A. D.; Capuzzi, S. J.; Elkins, J. M.; Ettmayer, P.; Frederiksen, M.; Gileadi, O.; Gray, N.; Hooper, A.; Knapp, S.; Laufer, S.; Luecking, U.; Michaelides, M.; Muller, S.; Muratov, E.; Denny, R. A.; Saikatendu, K. S.; Treiber, D. K.; Zuercher, W. J.; Willson, T. M. Progress towards a public chemogenomic set for protein kinases and a call for contributions. *PLoS One* **2017**, *12*, No. e0181585.
(116) Haga, C. L.; Velagapudi, S. P.; Strivelli, J. R.; Yang, W. Y.; Disney, M. D.; Phinney, D. G. Small molecule inhibition of miR-544 biogenesis disrupts adaptive responses to hypoxia by modulating ATM-mTOR signaling. *ACS Chem. Biol.* **2015**, *10*, 2267−2276.
(117) Kenny, P. W.; Montanari, C. A. Inflation of correlation in the pursuit of drug-likeness. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 1−13.
(118) Shultz, M. D. Two decades under the influence of the Rule of Five and the changing properties of approved oral drugs. *J. Med. Chem.* **2019**, *62*, 1701−1714.
(119) Valeur, E.; Gueret, S. M.; Adihou, H.; Gopalakrishnan, R.; Lemurell, M.; Waldmann, H.; Grossmann, T. N.; Plowright, A. T. New modalities for challenging targets in drug discovery. *Angew. Chem., Int. Ed. Engl.* **2017**, *56*, 10294−10323.
(120) Lagorce, D.; Douguet, D.; Miteva, M. A.; Villoutreix, B. O. Computational analysis of calculated physicochemical and ADMET properties of protein-protein interaction inhibitors. *Sci. Rep.* **2017**, *7*, 46277.
(121) Morgan, B. S.; Sanaba, B. G.; Donlic, A.; Karloff, D. B.; Forte, J. E.; Zhang, Y.; Hargrove, A. E. R-BIND: an interactive database for exploring and developing RNA-targeted chemical probes. *ACS Chem. Biol.* **2019**, *14*, 2691−2700.
(122) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, *36*, D901−D906.
(123) Valeur, E.; Narjes, F.; Ottmann, C.; Plowright, A. T. Emerging modes-of-action in drug discovery. *MedChemComm* **2019**, *10*, 1550−1568.
(124) Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technol.* **2004**, *1*, 337−341.

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

(125) Churcher, I. Protac-induced protein degradation in drug discovery: Breaking the rules or just making new ones? *J. Med. Chem.* **2018**, *61*, 444−452.
(126) Lee, J.; Udugamasooriya, D. G.; Lim, H.-S.; Kodadek, T. Potent and selective photo-inactivation of proteins with peptoid-ruthenium conjugates. *Nat. Chem. Biol.* **2010**, *6*, 258−260.
(127) Costales, M. G.; Matsumoto, Y.; Velagapudi, S. P.; Disney, M. D. Small molecule targeted recruitment of a nuclease to RNA. *J. Am. Chem. Soc.* **2018**, *140*, 6741−6744.
(128) Vauquelin, G. Effects of target binding kinetics on in vivo drug efficacy: k_{off}, k_{on} and rebinding. *Br. J. Pharmacol.* **2016**, *173*, 2319−2334.
(129) Bondeson, D. P.; Smith, B. E.; Burslem, G. M.; Buhimschi, A. D.; Hines, J.; Jaime-Figueroa, S.; Wang, J.; Hamman, B. D.; Ishchenko, A.; Crews, C. M. Lessons in PROTAC design from selective degradation with a promiscuous warhead. *Cell Chem. Biol.* **2018**, *25*, 78−87.
(130) Roy, M. J.; Winkler, S.; Hughes, S. J.; Whitworth, C.; Galant, M.; Farnaby, W.; Rumpel, K.; Ciulli, A. SPR-measured dissociation kinetics of PROTAC ternary complexes influence target degradation rate. *ACS Chem. Biol.* **2019**, *14*, 361−368.
(131) Copeland, R. A. The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discovery* **2016**, *15*, 87−95.
(132) 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*, 2406−2411.
(133) Carter, A. J.; Kraemer, O.; Zwick, M.; Mueller-Fahrnow, A.; Arrowsmith, C. H.; Edwards, A. M. Target 2035: probing the human proteome. *Drug Discovery Today* **2019**, *24* (11), 2111−2115.
(134) Kehe, J.; Kulesa, A.; Ortiz, A.; Ackerman, C. M.; Thakku, S. G.; Sellers, D.; Kuehn, S.; Gore, J.; Friedman, J.; Blainey, P. C. Massively parallel screening of synthetic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 12804−12809.
(135) Neri, D.; Lerner, R. A. DNA-Encoded Chemical Libraries: A selection system based on endowing organic compounds with amplifiable information. *Annu. Rev. Biochem.* **2018**, *87*, 479−502.

https://dx.doi.org/10.1021/acs.jmedchem.9b01927
J. Med. Chem. XXXX, XXX, XXX−XXX

Chem Soc Rev

REVIEW ARTICLE

View Article Online
View Journal

Cite this: DOI: 10.1039/d0cs00560f

Small molecule recognition of disease-relevant RNA structures†

Samantha M. Meyer, Christopher C. Williams, Yoshihiro Akahori, Toru Tanaka, Haruo Aikawa, Yuquan Tong, Jessica L. Childs-Disney and Matthew D. Disney *

Targeting RNAs with small molecules represents a new frontier in drug discovery and development. The rich structural diversity of folded RNAs offers a nearly unlimited reservoir of targets for small molecules to bind, similar to small molecule occupancy of protein binding pockets, thus creating the potential to modulate human biology. Although the bacterial ribosome has historically been the most well exploited RNA target, advances in RNA sequencing technologies and a growing understanding of RNA structure have led to an explosion of interest in the direct targeting of human pathological RNAs. This review highlights recent advances in this area, with a focus on the design of small molecule probes that selectively engage structures within disease-causing RNAs, with micromolar to nanomolar affinity. Additionally, we explore emerging RNA-target strategies, such as bleomycin A5 conjugates and ribonuclease targeting chimeras (RIBOTACs), that allow for the targeted degradation of RNAs with impressive potency and selectivity. The compounds discussed in this review have proven efficacious in human cell lines, patient-derived cells, and pre-clinical animal models, with one compound currently undergoing a Phase II clinical trial and another that recently garnerd FDA-approval, indicating a bright future for targeted small molecule therapeutics that affect RNA function.

Received 17th May 2020

DOI: 10.1039/d0cs00560f

rsc.li/chem-soc-rev

1. Introduction

RNA is a critical component of the Central Dogma, best known for its roles in transcription and translation. However, non-coding (nc) RNAs play important functions critical for the regulation of cell homeostasis and normal biology.[1] These ncRNAs, such as microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), transfer RNAs (tRNAs), long non-coding RNAs (lncRNAs), *etc.* (Fig. 1) are highly structured[1] and offered the first clues that RNA structures may play vital roles in human biology beyond the encoding and synthesis of protein. Indeed, this hypothesis has been proven true as RNA structures have been linked to both normal biology and disease pathology.[2,3]

The variability and complexity of RNA structures has been widely explored, leading to the appreciation that RNAs range from being largely disordered (dynamic) to adopting simple

Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL, 33458, USA. E-mail: disney@scripps.edu

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cs00560f

Samantha M. Meyer

Samantha M. Meyer received her BS in Biochemistry and Molecular Biology from the University of Wisconsin – Eau Claire in 2019. The following fall she began doctoral studies under the guidance of Prof. Matthew D. Disney at The Scripps Research Institute in Jupiter, Florida. Her current research focuses on targeting disease-causing RNAs with small molecules, with an emphasis on expanding the versatility of RIBOTACs.

Christopher C. Williams

Christopher C. Williams received his BA in Chemistry from Hamilton College (2017). He conducts his graduate studies under the guidance of Prof. Matthew D. Disney at The Scripps Research Institute in Florida, where he explores RNA-small molecule binding interactions in the context of infectious disease.

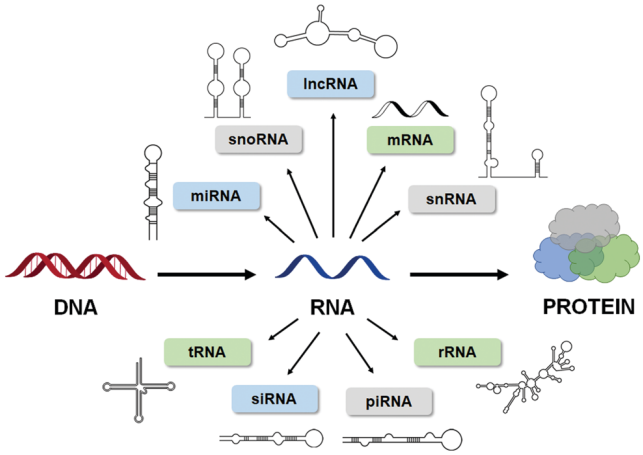


Fig. 1 RNA is highly structured. The Central Dogma of biology, showcasing the numerous types of structured RNAs that have been identified to date.

structures such as loops and bulges (secondary structure) to creating highly intricate pseudoknots, G-quadruplexes, and coaxial stacking (tertiary structure). The influence of these structures has been explored in the context of bacterial gene expression and riboswitches[4] and in viral replication and infection.[5] In the context of human biology, structured RNAs influence translational regulation,[6–8] alternative splicing,[9,10] and even enzymatic catalysis,[11–14] further demonstrating their intimate involvement in maintaining healthy biology. As these topics will not be reviewed in depth here, we direct the reader to the references cited above for additional detail.

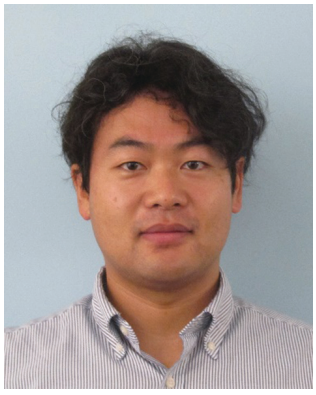
Predictably, disruption of RNA structure *via* mutation, formation of unnatural RNA structures, *e.g.*, by insertions or expansions, or aberrant expression, leads to dysregulation of cellular processes, resulting in disease. For example, dysregulation of miRNAs, short regulatory RNAs that modulate gene expression *via* the RNA-induced silencing complex (RISC),[7] have

Yoshihiro Akahori

Yoshihiro Akahori received his PhD in Synthetic Organic Chemistry (2014) from the Nagoya City University, supervised by Prof. Seiichi Nakamura. While there he worked on synthesizing oxygenated natural terpenoids. He then joined Daiichi Sankyo Co., Ltd where he engaged in drug discovery research. Since 2020, he has worked at The Scripps Research Institute with Prof. Matthew D. Disney. His current research focuses on developing small molecules that target RNA.

Toru Tanaka

Toru Tanaka received his PhD in Organic Chemistry (2016) from Kyoto Pharmaceutical University under the guidance of Prof. Masayuki Yamashita, where he worked on using sulfoxonium methylide skeletal transformation reactions to open cyclopropane and cyclobutene rings. Since 2020, he has worked as a postdoctoral fellow with Prof. Matthew D. Disney. His current research focuses on the design and synthesis of small molecules that bind RNA.



Haruo Aikawa

Haruo Aikawa was awarded his PhD in Chemistry from Tohoku University, Japan, in 2009, under Drs. Yoshinori Yamamoto and Naoki Asao. Dr. Aikawa also worked on peptide chemistry in the lab of Dr. Hirokazu Tamamura at Tokyo Medical and Dental University. His research on nucleic acid-binding molecules began in 2013 in the lab of Dr. Kazuhiko Nakatani at Osaka University. In 2018, he joined Dr. Matthew D. Disney's lab at The Scripps Research Institute in Florida as a postdoctoral fellow. His current research themes are regulation of RNA biology by small molecules targeting repeat expansion disorders and miRNAs.

Yuquan Tong

Yuquan Tong received his BS in Chemical and Biomolecular Engineering from the University of Illinois at Urbana – Champaign in 2019. Under the supervision of Prof. Matthew D. Disney, he started his doctoral studies at The Scripps Research Institute in Florida, working on small molecules targeting RNA. His current research focuses on targeting mRNAs of "undruggable" disease-causing proteins, including alpha synuclein and tau.

been associated with, among others, cardiovascular disease, inflammatory disorders, and cancer.[7,15,16] Additionally, structured RNAs have been implicated in several neurological disorders, as reviewed in Bernat *et al.*,[17] a well-known example being RNA repeat expansion/microsatellite disorders. This class of disorders is responsible for over 30 human diseases including Huntington's disease (HD), amyotrophic lateral sclerosis (ALS), fragile X-associated tremor and ataxia syndrome (FXTAS), and myotonic dystrophies type 1 and 2 (DM1 and DM2).[17] The biological consequences of these repeat expansions will be reviewed in detail below.

To date, two main therapeutic strategies have been employed to target disease-causing RNAs: antisense oligonucleotides (ASOs) and small molecules.[18] ASOs are single-stranded nucleotide sequences designed to complementarily base pair a target RNA's primary sequence. ASOs, which often contain modified phosphate backbones and sugar motifs to protect against cellular degradation, either repress translation by sterically blocking ribosomal loading onto the RNA or induce degradation of the target RNA *via* Ribonuclease H (RNase H).[19] RNase H recognizes the RNA–DNA heteroduplex and hydrolyzes the phosphodiester bonds of the RNA strand, cleaving it.[19] Conversely, small molecules are designed to target RNA structure instead of sequence, much like how small molecules are designed to target proteins *via* structure-based recognition. Small molecule binding of an RNA target can modulate disease biology, thus creating avenues to further explore RNA-disease biology and potential therapeutics against RNA-associated disorders.[18]

This review provides a general overview of recently developed RNA-targeting small molecules, highlighting advances in the field that continue to push towards the development of potent and selective small molecule lead therapeutics. A focus is placed on small molecules targeting miRNA biogenesis, lncRNAs, mRNAs encoding intrinsically disordered proteins (IDPs), and repeat expansion disorders. This review details both the pathomechansims caused by the RNA's structure and how small molecules can alleviate this pathology. Additionally, emergent modalities such as RNA-targeted cleaver and degrader compounds, including ribonuclease targeting chimeras (RIBOTACs), are reviewed in detail, highlighting the selectivity and potency of these compounds. There is still much to be learned about small molecules targeting RNA before these probes can be converted into lead medicines, but a solid foundation has been laid to enable clinical advancement across multiple indications. (See Table 1 for a complete list of diseases mentioned in this review and the abbreviations used to define them.) A tutorial on targeting RNA structures derived from sequence with small molecules can be found in ref. 20.

Table 1 Commonly mentioned diseases and abbreviations

Disease	Abbreviation
Triple negative breast cancer	TNBC
Fragile X-associated tremor and ataxia syndrome	FXTAS
Spinal muscular atrophy	SMA
Frontotemporal dementia	FTD
Parkinsonism linked to chromosome 17	FTDP-17
Myotonic dystrophy type 1	DM1
Myotonic dystrophy type 2	DM2
C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis	c9FTD/ALS
Parkinson's disease	PD

2. Small molecule targeting of miRNAs

MiRNAs are short 20–25 nucleotide (nt) sequences of RNA that negatively regulate gene expression through translational repression of their mRNA targets, dictated by sequence complementarity to the 3′ untranslated region (UTR) of the mRNA target. These RNAs are actively involved in regulation of cellular processes including proliferation, development, differentiation, and apoptosis. Like other types of RNAs, miRNAs are transcribed as primary transcripts (pri-miRs) that are processed into precursor miRNAs (pre-miRs), both of which fold into hairpin structures. These structures are cleaved sequentially by the nucleases Drosha and Dicer to produce the final, single-stranded mature miRNA (Fig. 2A).[7]

Biogenesis begins with the Drosha:DiGeorge syndrome critical region 8 (DGCR8) microprocessor complex that excises a portion of the pri-miR at the open stranded end of the hairpin, yielding a pre-miR of ∼70 nucleotides in length. The pre-miR is then exported from the nucleus *via* exportin 5.[21] In the cytoplasm, pre-miR is further processed at the hairpin loop by the enzyme Dicer, which acts as a molecular ruler, yielding a double stranded miRNA.[22] The miRNA is then loaded into the argonaute (AGO)/RISC complex where the guide strand stays successfully loaded and the complementary strand is degraded.[23] After biogenesis, the RISC complex regulates gene expression either through translational inhibition *via* steric hinderance of ribosomal loading or *via* stimulation of complete mRNA decay.[24–26]

Due to the complexity of miRNA interaction networks (*i.e.*, multiple miRNAs often act upon one mRNA, and one miRNA can regulate multiple mRNAs),[27,28] dysregulation of miRNA expression has been associated with a variety of human diseases, especially cancer.[29–31] Examples of how RNA-binding small molecules have been designed and optimized to bring about potent and selective regulators of miRNA function are discussed in detail below.

2.1 Neomycin–nucleobase conjugates targeting oncogenic miRNAs

Neomycin–nucleobase conjugates are small molecules that target disease-causing miR-372 and -373 (Fig. 2C and Table S1, ESI†).[32] These bifunctional conjugates consist of (i) an artificial nucleobase designed to specifically recognize an RNA base pair of the double-stranded region of pre-miRNA and (ii) an aminoglycoside shown to have strong binding affinity to stem-loop RNA motifs. Artificial nucleobases engage in the formation of Hoogsteen-type triplex DNA helices,[33] and when conjugated to basic amino acids, form compounds with high affinity and selectivity for the stem loop structure of human immunodeficiency virus type 1 (HIV-1) transactivation response element (TAR) RNA.[34] Aminoglycoside antibiotics, which alone constitute

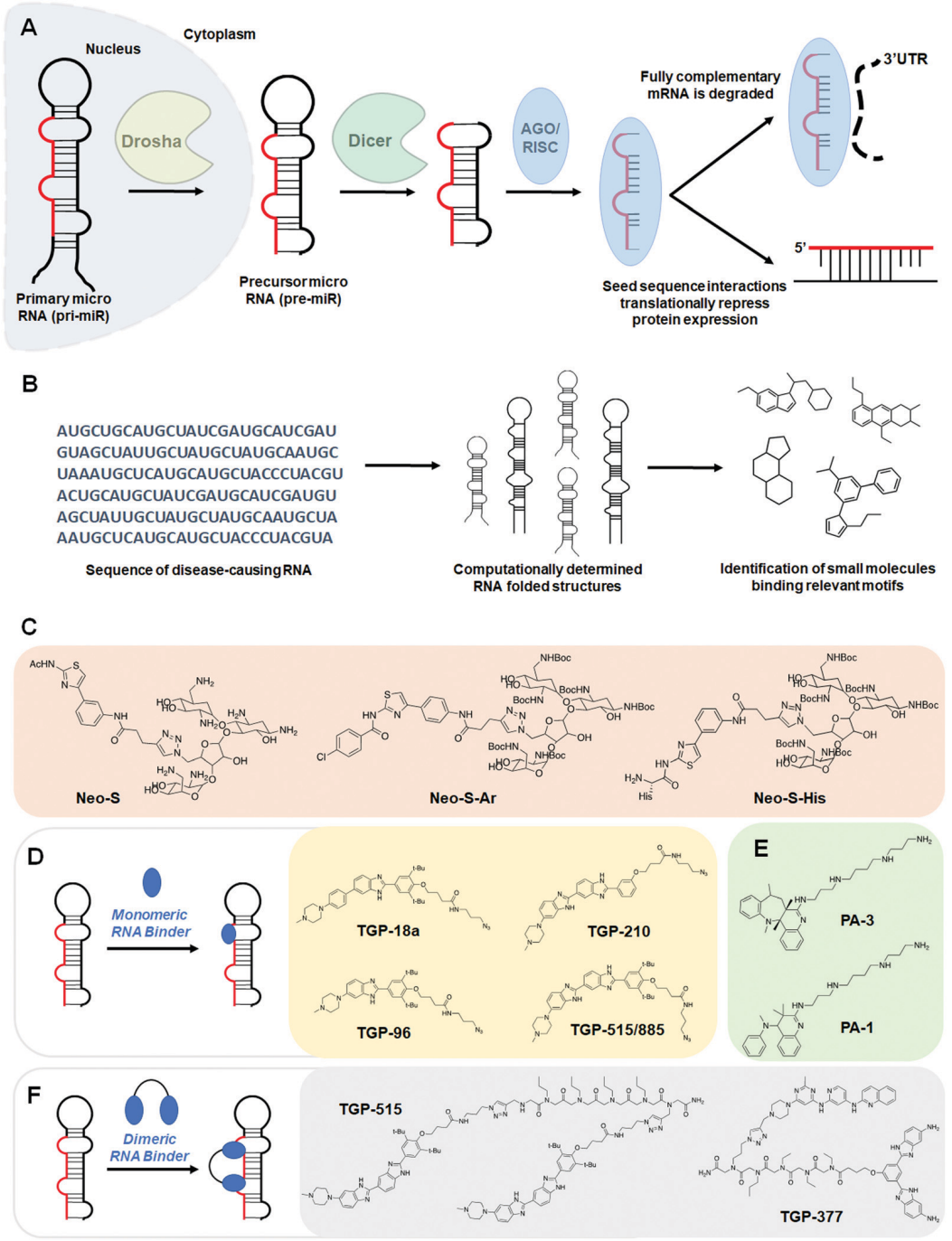


Fig. 2 Small molecule targeting of miRNAs. (A) Schematic of the biogenesis of miRNAs. Primary miRNAs (pri-miR) are processed by the nuclear RNase III Drosha and exported to the cytoplasm, affording precursor miRNAs (pre-miR), which are then processed by the RNase III endonuclease Dicer. The miRNA duplex is loaded into the AGO/RISC complex, where the duplex is dissociated and acts through either translational repression or mRNA degradation to downregulate target proteins. (B) Workflow schematic of the Inforna hit identification process. (C) Structures of neomycin conjugates that inhibit miRNA biogenesis. (D) Schematic of monomeric RNA binder mode of action, blocking functional processing sites on miRNA. Representative chemical structures of these monomers are also shown. (E) Structures of polyamines that inhibit miRNA biogenesis. (F) Dimeric RNA binders have improved potency and selectivity by binding to a functional site and nearby druggable motif simultaneously. Structures of representative miRNA-targeting dimers are shown.

a class of widely prescribed medicines targeting the decoding A-site in prokaryotic rRNA to inhibit protein translation, bind stem-loop structured RNAs along the major groove of the RNA duplex.

On the basis of these findings, the Duca lab rationalized conjugation of the aminoglycoside neomycin with an artificial nucleobase would yield chemical matter capable of binding the stem-loop sequences of miR-372 and -373.[32] These first-generation conjugate compounds fortuitously bound the Dicer processing sites of pre-miR-373 and pre-miR-372, inhibiting their biogenesis *in vitro*, as determined by a cell-free Förster resonance energy

transfer (FRET) based assay, and reduced oncogenic burden in cells. The **Neo-S** conjugate inhibited Dicer cleavage *in vitro* and rescued expression of the miRNA-regulated protein Large Tumor Suppressor homologue 2 (LATS2). However, **Neo-S** was not entirely selective and affected expression of miR-17-5p, -21, and -200b in a dose-dependent manner, albeit to a lesser extent than miR-372 and -373 (Table S1, ESI†).

Through medicinal chemistry efforts, Vo *et al.*[35] synthesized and evaluated the properties of second generation compounds with the aim of improving potency and selectivity. Using a cell-free FRET based assay, they learned that select modifications of the artificial nucleobase motif yielded little to no improvement in inhibitory activity. Extended linker length proved deleterious and between a selection of other aminoglycosides, neomycin still remained the best at inhibiting Dicer processing. Preliminary evaluation of compounds offered two new structures for examination in further studies, the first of which quickly fell out of favor due to unspecific binding to both the stem and loop regions of pre-miR-372 and evidence of binding to double-stranded DNA and tRNA. The second structure, **Neo-S-Ar**, with an improved IC_{50} relative to **Neo-S** (1.0 μM *versus* 2.4 μM for **Neo-S**), decreases miR-372 and -373 levels in cells in a dose-dependent manner, but much like **Neo-S**, inhibits Dicer processing of pre-miR-17 and -21 and affects levels of miR-200b in AGS (human Caucasian gastric adenocarcinoma) cells (Table S1, ESI†). The authors noted that **Neo-S** and **Neo-S-Ar** only elicit a phenotypic response in AGS cells, which overexpress miR-372 and -373, and not in MKN74 (human gastric tubular adenocarcinoma) cells, which do not overexpress these oncogenic miRNAs. Despite imperfect selectivity, these efforts provided a lead for further drug optimization.

With the aim of improving potency and selectivity for the miR-372 and -373 targets, Vo *et al.*[36] reasoned that because amino acids are natural ligands of RNA and easily interact with negatively charged RNA structures/sequences, appending one such amino acid could improve potency and selectivity.[37] The lab had also shown that basic amino acids, including arginine, lysine, and histidine are particularly effective in the design of selective RNA ligands,[34,38] inspiring **Neo-S-His**, which had improved selectivity for pre-miR-372 over previous generations (**Neo-S** and **Neo-S-Ar**) (Table S1, ESI†).[36] Conjugation of different amino acids appended at various positions on the neomycin–nucleobase scaffold were synthesized, but **Neo-S-His** was the only compound selective for pre-miR-372 over DNA. Treatment of AGS cells with **Neo-S-His** showed the compound inhibited cell growth by ~40% and even though the compound also affected expression levels of oncogenic miR-21, aberrant expression of miR-21 has been shown to have no effect on the proliferation of AGS cells, indicating this off-target did not contribute to the observed *anti*-proliferative effects. Furthermore, expression levels of other miRNAs, including miR-371, -373, -17, and -200b, were not affected by **Neo-S-His**, unlike previous generations of the compound (Table S1, ESI†).

2.2 Polyamines targeting oncogenic miRNAs

In addition to neomycin–nucleobase conjugates, Staedel *et al.*[39] screened a 640-member library for inhibition of Dicer-mediated pre-miR-372 processing to identify novel scaffolds with enhanced potency and selectivity. The top hits were all polyamines, the most potent of which, **PA-1**, inhibited growth of AGS cells, but not MKN74 cells (Fig. 2E and Table S1, ESI†). Treatment of AGS cells with **PA-1** also resulted in a dose-dependent accumulation of the downstream protein LATS2, much like the first-generation neomycin–nucleobase conjugate series. **PA-1**, however, binds and affects expression levels of miRs other than miR-372 in AGS cells (Table S1, ESI†). Binding studies of **PA-1** revealed that RNA binding was enhanced most significantly by interactions between the polyamine chain and the RNA phosphate backbone, and less significantly by π–π interactions between the dihydroquinoline motif and specific nucleotides.[40] Therefore, a strained analog of **PA-1**, **PA-3**, featured a fused benzazepine-dihydroquinoline motif appended to the polyamine chain. Using the previously employed cell-free FRET based assay, it was shown that **PA-3** inhibited Dicer processing of pre-miR-372 twice as well as **PA-1**. Compared to **PA-1** and other newly synthesized analogs, **PA-3** showed (i) the most selective inhibition of Dicer-mediated processing of miR-372, (ii) the most selective binding of pre-miR-372 in the presence of a large excess of tRNA and DNA, and (iii) the greatest selectivity for pre-miR-372 and pre-miR-373 over other pre-miRNAs in terms of activity and affinity (Table S1, ESI†). Furthermore, thermodynamic binding profiles of the polyamine/pre-miR-372 complex revealed that **PA-3** bears the highest enthalpic contribution.

2.3 Design of monomeric small molecules targeting disease-causing miRNAs

Additional small molecules targeting miRNAs have been identified by the lead identification strategy dubbed Inforna.[41] (For a more in-depth, tutorial review of Inforna and its utilization please see ref. 20.) Inforna comprises a database of experimentally selected RNA motif-small molecule interactions and mines the structural motifs in a chosen disease-related RNA target, deduced from its sequence, for overlap with the database (Fig. 2B). Inforna allows for transcriptome-wide probing of bioactive small molecules that target RNA without target bias (a target agnostic approach). This ''bottom-up'' strategy has enabled the design of modularly assembled small molecules that bind RNAs linked to human disease, proving particularly successful in the targeting of disease-causing miRNAs. One such example includes the design of Targapremir-18a (**TGP-18a**), named for its targeting of pre–miR–18a (Fig. 2D and Table S1, ESI†).[42] *In vitro* studies showed that **TGP-18a** inhibits Dicer processing of multiple members of the miR-17-92 cluster, namely pre-miR-17, pre-miR-18a, and pre-miR-20a, which share a common bulge at the Dicer site and adjacent structural similarity. Using RT-qPCR, these *in vitro* results were corroborated in DU145 prostate cancer cells, in which miR-18a is overexpressed. Importantly, inhibition de-represses serine/threonine protein kinase 4 (STK4) and rescues phenotype, *i.e.*, triggers apoptosis. Interestingly, these studies used Inforna to identify potential miRNA off-targets, that is other miRNAs with binding sites for **TGP-18a**, albeit with less avidity. The potential off-targets are expressed at much lower levels than the miRNAs in the miR-17-92 cluster, on average about 10-fold

less (Table S1, ESI†). Not only were these miRNAs unaffected by **TGP-18a**, target engagement studies show that they were not bound by the small molecule, demonstrating that differences in target expression level can be exploited to enhance the observed selectivity.

Another example of a miRNA target proven druggable through the use of Inforna is miR-96 (Table S1, ESI†).[43] Velagapudi *et al.*[43] showed the compound **Targaprimir-96** (**TGP-96**) reduces miR-96 levels (*via* inhibition of Drosha processing) at least as selectively as a locked nucleic acid (LNA). In one example, when dosed at concentrations high enough to silence approximately 90% of miR-96 expression, the miR-96 LNA also silenced ~50% of miR-183 expression, owing to the overlapping seed sequences of the two miRNAs (only the first nucleotide differs). In contrast, **TGP-96** only silenced miR-182 expression by ~15% when dosed at concentrations that silenced miR-96 expression by ~90% (Table S1, ESI†). Inhibition of miR-96 biogenesis by **TGP-96** de-repressed downstream protein expression of Forkhead box O1 (FOXO1), a putative tumor suppressor regulated by miR-96,[44] and stimulated apoptosis in MCF-7 breast cancer cells. Importantly, these studies confirmed that **TGP-96** acts along the miR-96-FOXO1 circuit by knocking down FOXO1 expression with an siRNA. Indeed, knockdown of FOXO1 reduces **TGP-96** activity.

In complementary studies, Costales *et al.*[45] designed **TGP-210**, a miR-210 binding small molecule that inhibits Dicer processing (Table S1, ESI†). MiR-210 controls hypoxia inducible factors (HIFs) through the negative regulation of glycerol-3-phosphate dehydrogenase 1-like (GPDL1).[46] *In vitro* and in triple negative breast cancer (TNBC) MDA-MB-231 cells, cultured under hypoxic conditions, **TGP-210** dose-dependently inhibited Dicer processing of pre-miR-210. In cells, this inhibition resulted in rescue of GPDL1 expression, reduction of levels of HIF1-α, and triggering of apoptosis. **TGP-210** was selective across a panel of hypoxia-associated miRNAs, as determined by RT-qPCR of treated MDA-MB-231 cells (Table S1, ESI†).

A technique termed Chemical-Cross Linking and Isolation by Pull Down (Chem-CLIP)[47] was then used to confirm direct target engagement of pre-miR-210 by **TGP-210**. In this technique, the small molecule of interest (in this case **TGP-210**) was appended with cross-linking (ex. chlorambucil) and purification (ex. biotin) modules. Upon compound binding to the target RNA, proximity-induced cross-linking occurs, which results in a complex that can be purified *via* pull-down with streptavidin beads. The RNAs enriched in the pull-down fraction, relative to the starting lysate, can be determined either through RT-qPCR or RNA-seq to confirm the compound's cellular target. Although expression levels of other mature miRNAs had been shown to be unaffected by **TGP-210**, Chem-CLIP experiments demonstrated binding does occur to other miRNAs. These studies showed that binding to a functional site is required for bioactivity and confirmed the observation by Velagapudi *et al.*[42] that expression level influences the degree of target occupancy. *In vivo* studies in NOD/SCID mice showed that treatment with **TGP-210** effectively reduces tumor growth *via* inhibition of miR-210 levels, de-repression of GPDL1, and reduction of HIF1-α levels.

2.4 Design of dimeric small molecules targeting disease-causing miRNAs

The observed selectivity for **TGP-210** was fortuitous, as off-targets were bound significantly less avidly and/or at non-functional sites and their expression levels were significantly lower than the desired target. However, such factors are unlikely to align for most targets. Therefore, facile methods to enhance small molecule potency and selectivity would be beneficial. As a test case, Costales *et al.*[48] explored **TGP-515/885**, a monomeric compound designed using Inforna that binds with dual selectively to the Drosha processing sites of both miR-515 and -885 (Fig. 2D and Table S1, ESI†). While both hairpin miRNA structures bear similar sequences at the Drosha processing sites, miR-515 folds with an additional internal loop adjacent to the Drosha processing site that also binds **TGP-515/885**. Dimerization of **TGP-515/885** yields **TGP-515**, which binds both the Drosha processing site and the adjacent internal loop to confer selectivity for pri-miR-515 over pri-miR-885 (Fig. 2F). Indeed, **TGP-515** bound miR-515 avidly while discriminating against pri-miR-885 *in vitro* and in cells (Table S1, ESI†). Its >200-fold enhancement in affinity compared to **TGP-515/885** translated into a >10-fold boost in potency in cells. Experiments in MCF-7 cells revealed that across all miRNAs, the entire transcriptome, and the proteome, **TGP-515** was selective for its RNA target.

Interestingly, cellular inhibition of miR-515 biogenesis de-repressed sphingosine kinase 1 (SK1), responsible for the synthesis of the second messenger sphingosine 1-phosphate (S1P), both of which were upregulated by **TGP-515** treatment. Activation of this circuit triggers migratory and proliferative characteristics of MCF-7 cells. However, it also enhances levels of human epidermal growth factor receptor 2 (HER2) levels, sensitizing HER2 negative cells (MCF-7 cells) to HER2-targeting precision medicines. Taken all together, this study shows that Inforna can inform how to design a specific compound from a dual-selective monomeric fragment.

In addition to the design of homodimeric molecules, Inforna can be used to design heterodimeric compounds which bind avidly to miRNAs.[49] Vascular endothelial growth factor A (VEGFA) stimulates angiogenesis in human endothelial cells and is a sought after target in the treatment of heart failure.[50–52] MiR-377 regulates VEGFA expression, and repression of miR-377 by an antisense oligonucleotide has been shown to rescue VEGFA expression and stimulate angiogenesis.[53,54] Inforna-based design afforded **TGP-377**, which binds pre-miR-377 at the Dicer site and another bulge directly adjacent (Fig. 2F and Table S1, ESI†).[49] Expression levels of miR-377 from human umbilical vein endothelial cells (HUVEC) treated with **TGP-377** were knocked down with an IC_{50} of ~500 nM, 10-fold more potently than the lead small molecule monomer (Table S1, ESI†). Accumulation of pre-miR-377 was also observed, demonstrating **TGP-377** acts through inhibition of Dicer processing and correspondingly rescues VEGFA expression. A miRNA profiling experiment showed that **TGP-377** targets miR-377 selectively, including

among miR-377 isoforms (Table S1, ESI†). Global proteomics analysis revealed that **TGP-377** affects only 160 of over 4000 unique proteins. A bioinformatic STRING analysis uncovering protein association networks showed, unsurprisingly, cell proliferative pathways including FGFR, Hedgehog, MAP kinase, and ERK were upregulated. Furthermore, **TGP-377** induced a pro-angiogenic phenotype in HUVEC cells as evidenced by increased tubule branching density by ~50% relative to control. As gene therapy is the only known treatment strategy to increase VEGFA expression, **TGP-377** represents the first small molecule to do so.[50–52,55]

3. Small molecule recognition of lncRNAs

LncRNAs are eukaryotic transcripts >200 nt in length that do not encode a protein.[56] These RNAs play key regulatory roles in cellular processes such as proliferation, differentiation, and development, the aberrant expression of which can lead to cancer,[57] neurodegenerative[58] and neuromuscular[59] disorders, and immune disorders.[60,61] LncRNAs are promising therapeutic targets because of their differential expression between cancerous and normal tissues and their important roles in carcinogenesis.[62] Not surprisingly, small molecule screening against lncRNAs has been attracting attention.[63–65] In this section, we describe examples of small molecule regulation of lncRNAs.

3.1 Small molecule recognition of the lncRNA HOTAIR

The lncRNA HOX transcript antisense RNA (HOTAIR) is involved in several cellular processes associated with carcinogenesis, such as those affecting cell mobility, proliferation, apoptosis, invasion, aggression, and metastasis.[66] Additionally, HOTAIR recruits chromatin-modifying complexes, such as polycomb repressive complex 2 (PRC2) and lysine-specific histone demethylase 1 (LSD1) to modulate the cancer epigenome and suppress tumor suppressor genes.[67]

Ren *et al.*[68] used *in silico* high-throughput screening to identify **ADQ** as a potent small molecule binder of HOTAIR (Fig. 3A and Table S1, ESI†). In multiple cancer cell lines, **ADQ** increased expression of nemo like kinase (NLK), a transcriptional target of HOTAIR, in a luciferase assay. Electrophoretic mobility shift assay (EMSA) confirmed **ADQ** directly binds HOTAIR. To further confirm the functional domains of **ADQ**, full-length HOTAIR, the 5′ domain, or a mutant 5′ domain construct were stably transfected into U87 and MDA-MB-231 cells. The **ADQ**-mediated dissociation of HOTAIR and enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2) was confirmed using full-length HOTAIR but was not observed with the mutant 5′ domain in which the **ADQ** binding site was ablated (Fig. 3B).

3.2 Small molecule recognition of the lncRNA MALAT1

The lncRNA metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) has recently been identified to be upregulated and coupled to tumorigenesis in several cancers.[69] MALAT1 has been linked to several physiological processes, including alternative splicing, nuclear organization, and epigenetic modulation of gene expression.[70] A study in colorectal cancer cells showed that an ~1500 nt segment at the evolutionarily conserved 3′ end of MALAT1 was sufficient to increase invasion and proliferation, implying that this region enables its oncogenic function.[71] The recent structural characterization of a 74 nt region at the 3′ end of MALAT1 by X-ray diffraction confirmed a unique, bipartite triple helix where the U-rich stem-loop sequesters the A-rich tail, a phenomenon proposed to prevent exonucleolytic degradation (Fig. 3B).[72,73] Notably, the deletion of this segment decreased accumulation of the MALAT1 transcript. A comparable decrease in accumulation was also observed upon mutation of a Hoogsteen-positioned uridine, thought to disrupt the triple-helix structure, indicating that subtle alterations in the stability of this structure can lead to significant changes in transcript level.

Donlic *et al.*[74] have identified small molecule binders of MALAT1 through *in vitro* assays. They used furamidine, the tunable diphenylfuran (DPF)-based scaffold, as a starting point because furamidine is known to bind to triple helix structures

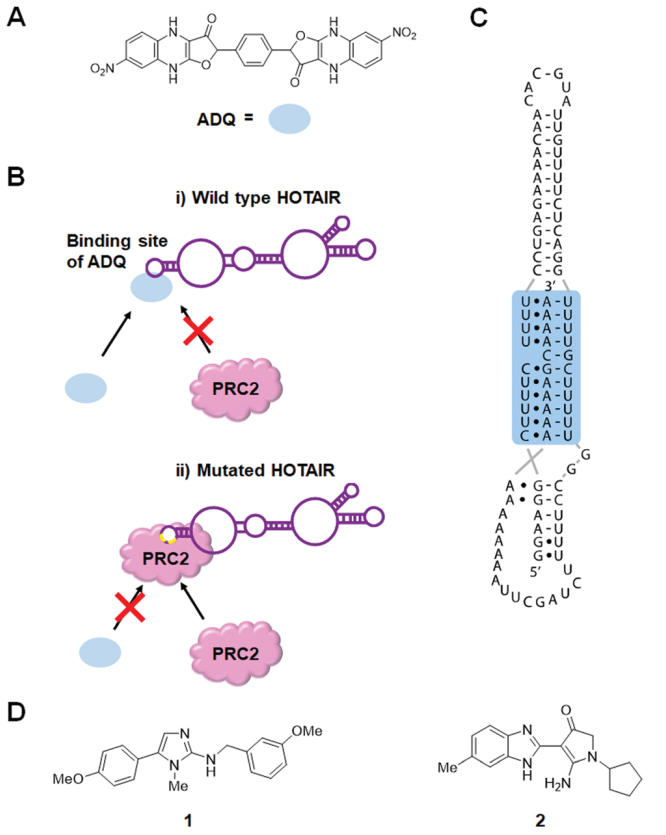


Fig. 3 Small molecule inhibition of lncRNAs. (A) Chemical structure of **ADQ**. (B) **ADQ** binds to the 5′ domain of HOTAIR and suppresses trimethylation of histone H3 lysine 27 (H3K27) in the promoter region of nemo like kinase (NLK) by weakening HOTAIR's ability to recruit and bind enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2), the enzymatic component of the PRC2 complex, thus restoring expression of NLK. (C) MALAT1 triple helix structure. (D) Chemical structures of MALAT1 small molecule binders.

of various DNAs.[75,76] They synthesized a DPF scaffold-based small molecule library, diversified in subunit composition and positioning, to explore the recognition of MALAT1.

Using a small molecule microarray (SMM) strategy, Abulwerdi *et al.*[64] reported the discovery of two structurally unrelated derivatives (**1** and **2**) that target the triplex region of MALAT1 (Fig. 3C and Table S1, ESI†). Compound **1** was selective for MALAT1 and nuclear paraspeckle assembly transcript 1 (NEAT1), which has a similar structure to MALAT1 (Table S1, ESI†). FRET, isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR) spectroscopic experiments confirmed **1** binds to MALAT1 *in vitro*. However, understanding of the inhibitory mechanism of **1** is limited by the lack of knowledge surrounding the actual mechanism of triplex-mediated protection. Additional research in this area will prove advantageous for the design of therapeutics targeting oncogenic lncRNAs and provide further support for target engagement.

4. Small molecule rescue of repeat-associated transcriptional repression in fragile X syndrome

Currently without a cure, fragile X syndrome (FXS) is the most common hereditary disorder that causes mental retardation, resulting from >200 CGG triplet repeats [full mutation allele; r(CGG)^{exp}] in the 5′ UTR of the fragile X mental retardation 1 (*FMR1*) gene on the X chromosome.[77] The *FMR1* promoter is epigenetically silenced through elevated levels of DNA CpG methylation and repressive histone marks H3K9me2, H3K9me3, H3K27me3, and H4K20me3, as well as lower levels of active histone marks H3K9ac, H3K4me2, and H4K16ac. Silencing progresses during embryonic development with the consequent loss of fragile X mental retardation protein (FMRP) encoded by the *FMR1* gene.[77] Although the mechanism of disease progression of FXS is not fully understood at present, a small molecule targeting the FXS RNA has been discovered that rescued repeat-associated epigenetic silencing.

4.1 Small molecule prevents the formation of RNA:DNA hybrids

Colak *et al.*[78] reported that treatment of human embryonic stem cells (hESCs) from FXS patients with **1a**, which was discovered using Inforna,[79] can prevent epigenetic silencing during neuronal differentiation (Table S1, ESI†). Knockdown of *FMR1* mRNA in hESCs decreased silencing histone marks, suggesting the *FMR1* transcript is involved in the gene silencing process of its own gene. In the presence of **1a**, repressive histone marks induced by differentiation also decreased. Since the compound thermodynamically stabilizes the r(CGG)^{exp} hairpin by binding to its 1 × 1 GG internal loops, it was speculated that the unfolded *FMR1* mRNA is responsible for epigenetic silencing. To support this hypothesis, they performed chromatin isolation by RNA purification, a technique used to identify DNA sequences which bind to a specific RNA sequence. These studies showed the *FMR1* DNA adjacent to the genomic CGG repeat is highly enriched only in the absence of **1a**. Furthermore, treatment with RNase H, which selectively digests RNA:DNA duplexes, significantly reduced the enrichment of the *FMR1* promoter. Based on these results, a mechanism was proposed by which *FMR1* mRNA containing extended CGG repeats binds to complementary DNA to form the RNA:DNA duplex that induces epigenetic silencing of the *FMR1* promoter (Fig. 4A). It was also suggested that **1a** promotes CGG stem-loop formation of the *FMR1* transcript and thus prevents the formation of the RNA:DNA duplex. In addition, because silencing decreases *FMR1* mRNA expression, RNA:DNA duplex formation would be engaged only at the initiation of silencing. In fact, **1a** has no effect on silenced cells, as subsequent gene silencing is maintained by other factors.

4.2 Small molecule targeting of r(CGG)^{exp} in combination with 5-azadeoxycytidine

In 2016, Kumari *et al.*[80] proposed that **1a** also has an inhibitory effect on histone methyltransferase polycomb repressive complexes 2 (PRC2) recruitment. Treatment of FXS patient cells with 5-azadeoxycytidine (AZA), an inhibitor of DNA methyltransferase 1, has been reported to demethylate the *FMR1* promoter and reactivate the *FMR1* gene.[81,82] Although AZA withdrawal causes re-silencing of the *FMR1* gene, this can be greatly delayed in the presence of **1a**, but not by inhibiting the RNA:DNA hybrid. Rather, **1a** inhibits association of r(CGG)^{exp} with PRC2, interfering with its recruitment to unmethylated CpG motifs and thus slowing *FMR1* resilencing in FXS patient cells (Table S1, ESI†).[79,83] It is assumed that H3K27 in the *FMR1* promoter is methylated by the aberrantly recruited histone methyltransferase. Indeed, it was observed that inhibitors of EZH2, the enzymatic component of PRC2, affect the maintenance of the reactivated state similar to how **1a** does. Knockdown of *FMR1* mRNA also reduced EZH2 levels associated with the *FMR1* gene. Taken together, these data support that **1a** is a dual functioning compound, preventing DNA:RNA hybrid formation and the recruitment of PRC2 by binding to the r(CGG)^{exp} stem-loop (Fig. 4B).

5. Small molecules modulate alternative splicing

Alternative splicing is a complex, elegant cellular process that allows for variation in protein isoforms to modulate protein function.[84] During the splicing process, exons can be included or excluded, giving rise to a variety of splicing isoforms afforded from a single pre-mRNA.[84] Mutations that change splicing patterns unsurprisingly cause human diseases including muscular atrophy,[85,86] tauopathies,[87] β-thalassemia,[88] progeria,[89] and Pompe disease.[90]

5.1 Small molecules modulate *SMN2* splicing

Spinal muscular atrophy (SMA) is caused by mutations in the *SMN1* gene that decrease levels of survival motor neuron (SMN) protein produced in the spinal cord.[86] In humans, *SMN1* and *SMN2* are the two genes that encode for SMN, but the majority

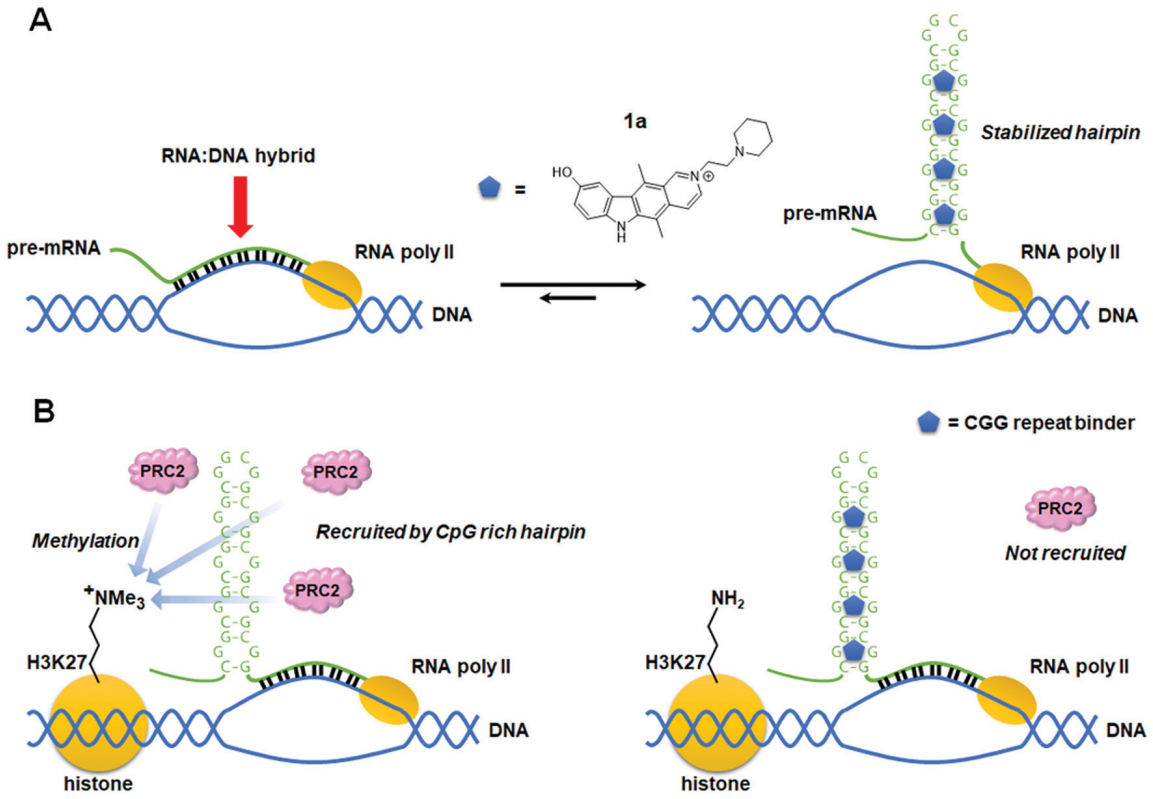


Fig. 4 Proposed mode of action of an r(CGG)^{exp} repeat binder that prevents epigenetic silencing of the *FMR1* promoter in fragile X syndrome (FXS). (A) Schematic mechanism showing stabilized r(CGG)^{exp} hairpins restrict formation of the RNA:DNA hybrids responsible for epigenetic silencing of *FMR1*. (B) Schematic mechanism showing binding of a small molecule to the r(CGG)^{exp} hairpin preventing recruitment of polycomb repressive complex 2 (PRC2), a H3K27 methylation enzyme complex.

of SMN protein is translated from the full-length mRNA produced from the *SMN1* pre-mRNA.[91] Due to a C-to-U transition at position 6 on exon 7, exon 7 skipping is dominant in the splicing of *SMN2* pre-mRNA,[92] producing a truncated SMN protein with a reduced half-life.[93] Currently, there are three treatment options for SMA: **nusinersen**, an ASO that regulates *SMN2* splicing to produce the full-length SMN protein;[94] onasemnogene abeparvovec, an adeno-associated virus (AAV) carrying the normal *SMN1* gene;[95] and **risdiplam** (PTC/Roche), an orally avaliable small molecule that was recently FDA-approved.[96,97] Another small molecule therapeutic candidate, **branaplam** (Novartis), is also currently undergoing clinical trials (Fig. 5A and Table S1, ESI†).[96] **Risdiplam** and **branaplam** generate the SMN protein *via* regulation of *SMN2* splicing. Since both compounds were discovered from phenotypic screening, a series of studies on their modes of action (MOAs) were reported and are discussed below.

5.1.1 Small molecule stabilization of exon 7 5′ splice site–U1 snRNP complex. Palacino *et al.*[98] investigated **branaplam's** MOA by using the active derivative, **NVS-SM2**, as a proxy (Table S1, ESI†). Since it is known that mutations at the end of *SMN2* exon 7[86,99] and in breast cancer 1 (*BRCA1*) exon 18[100] induce exon skipping, the authors tested the ability of **NVS-SM2** to modulate the splicing of these two exons, the latter as a counter screen of small molecule selectivity. While **NVS-SM2** rescued *SMN2* exon 7 splicing, it failed to rescue *BCRA1* exon 18 splicing. To define the *SMN2* RNA sequence that interacts with **NVS-SM2**, the authors utilized a set of *SMN2-BCRA1* chimeric genes to pinpoint the sequence necessary for **NVS-SM2** interaction. Only one chimeric gene, containing 21 nucleotides of the 5′ splice site in *SMN2* exon 7 fused to *BRCA1*, showed exon inclusion activity when treated with **NVS-SM2**, suggesting **NVS-SM2** interacts with the 5′ splice site of *SMN2* exon 7. Interestingly, the GA sequence at the end of exon 7 was found to be critical for the activity of **NVS-SM2**, as determined by base mutation experiments of the 5′ splice site. A RefSeq comparison revealed nGA sequences at the 3′ ends of exons are rare, suggesting the U1 small nuclear ribonucleoprotein (snRNP), a splice site-recognizing RNP, is involved in the mode of action of **NVS-SM2**. Size-exclusion chromatography confirmed that exon 7′s 5′ splice site bound to U1 snRNP only when **NVS-SM2** was present. In addition, total correlated spectroscopy (TOCSY) NMR showed that chemical-shift perturbations were observed at residues proximal to the nGA motif. Combining all these data with the crystal structure of U1 snRNP,[101,102] it was proposed **NVS-SM2** has a novel mode of action by which it stabilizes a ternary complex between the small molecule, U1 snRNP, and the 5′ splice site, particularly at the major groove of the –1A RNA bulge (Fig. 5B and Table S1, ESI†).

Risdiplam's MOA was defined using a derivative dubbed **SMN-C5** and a duplex model of the 5′ splice site/U1 snRNP complex (Table S1, ESI†).[103] The model consisted of 11 nt of the 5′ splice site hybridized to 11 nt of the U1 snRNA. The three-dimensional

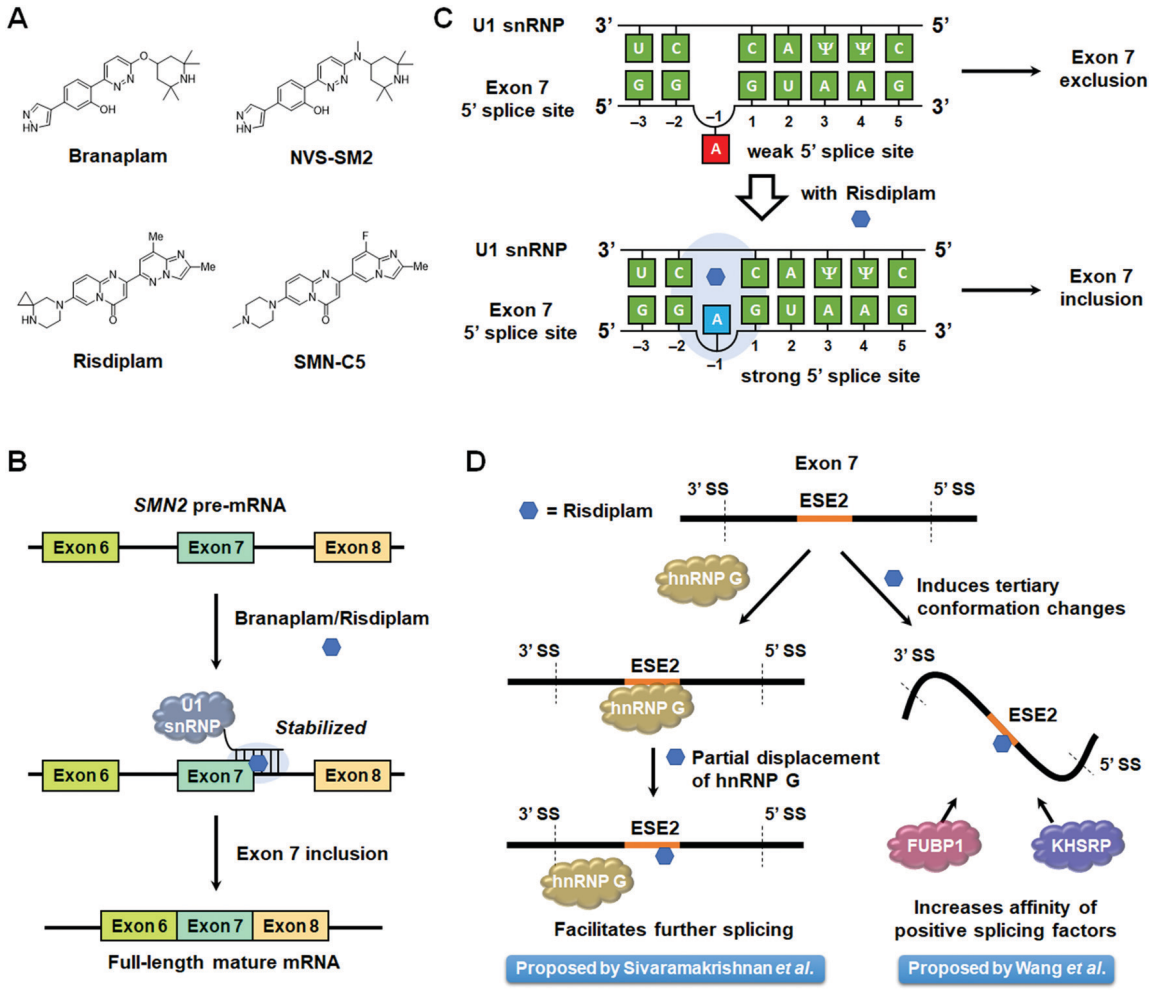


Fig. 5 Mode of action of small molecule splicing modulators targeting *SMN2 pre-mRNA*. (A) Structures of small molecule splicing modulators targeting *SMN2* pre-mRNA and the derivatives used to study their mechanisms of action. (B) Schematic mechanism of small molecules facilitating *SMN2* exon 7 inclusion by stabilizing the complex between *SMN2* exon 7, the 5′ splice site (SS), and the U1 snRNP. (C) Schematic representation of 5′ splice site bulge repair mediated by **risdiplam**. (D) Competing modes of action proposed for **risdiplam**.

structure of this model with and without **SMN-C5** was defined by NMR spectroscopy, constrained by nuclear Overhauser effects (NOEs). In the binding model of the apo form, the unpaired adenine in the 5′ splice site is located in the minor groove. Upon **SMN-C5** binding to the RNA's major groove, the bulged adenine is pushed back into the duplex, stabilized by the hydrogen bond formed between the carbonyl group of **SMN-C5** and the amino group of the adenine. Previous structural studies have shown that the U1 snRNP zinc finger stabilizes the minor groove at exon–intron junction of RNA duplexes.[101,102] Modeling the apo duplex in the zinc finger produces an obvious steric clash between the bulged adenine and the zinc finger. In contrast, the **SMN-C5**-bound duplex alleviates this clash, improving the accessibility of the minor groove. Collectively, these studies suggest that **SMN-C5** improves splice site recognition by U1 snRNP, facilitating exon 7 inclusion and expression of functional, full length SMN protein (Table S1, ESI†).

5.1.2 Small molecule interaction with exonic splicing enhancer 2 (ESE2) in *SMN2* exon 7 recruits splicing factors. **Risdiplam** modulates the alternative splicing of other exons such as striatin 3 (*STRN3*) exon 8, among others. Sivaramakrishnan *et al.*[104] searched for common sequence motifs around these exons (*STRN3* exon 8 and *SMN2* exon 7) and found the sequences of their 5' splice site are an exact match, while they share similar exonic splicing enhancer (ESE) sequences juxtaposed to a purine rich sequence. ESE sequences are known to recruit positive splicing factors and thus aid in the splicing process.[105] These results suggested that **SMN-C5** may have an additional mode of action besides ternary complex formation with the 5' splice site and U1 snRNP. Surface plasmon resonance (SPR) studies indicated binding of **SMN-C5**, but not **NVS-SM1**, to the ESE2 in *SMN2* pre-mRNA. In addition, NMR spectroscopy showed large chemical-shift perturbations of the ESE2 RNA were induced by addition of **SMN-C5**, resulting in the formation of broad imine signals, indicative of a small molecule-induced conformational change.

The authors then sought to identify potential protein components that may be contributing to *SMN2* exon 7 skipping using a pull-down experiment. Ten proteins were enriched only in the presence of **SMN-C5**, among them heterogenous nuclear

ribonucleoprotien (hnRNP) G, a known positive splicing factor that interacts with ESE2.[106] Unexpectedly, **SMN-C5** partially competes with hnRNP G for ESE2 binding, and small molecule binding alters the RNA structure of the region to which hnRNP G normally binds. Thus, one hypothesis is that partial displacement of hnRNP G by **SMN-C5** facilitates the progression of the splicing process (Fig. 5D).

Wang *et al.*[107] also reported that **SMN-C2** and **SMN-C3**, derivatives of **risdiplam**, act on ESE2. *SMN2* exon 7 is known to form two stem-loops, terminal stem-loop 1 (TSL1) and terminal stem-loop 2 (TSL2), that have an inhibitory effect on splicing.[108] Both cell-free and cell-based selective 2′ hydroxyl acylation analyzed by primer extension (SHAPE) analysis showed that the addition of **SMN-C2** altered the reactivity of some bases in TSL1, suggesting this compound induces conformational changes of this inhibitory loop. Further, proteomics analysis using a photo-cross-linking probe revealed enrichment of far upstream element binding protein 1 (FUBP1)[109] and far upstream element binding protein 2 (KHSRP).[110] Fluorescence polarization assays with **SMN-C2** and recombinant FUBP1 induced higher polarization in the presence of ESE2. These results indicated that **SMN-C2**, FUBP1, and exon 7 form a ternary complex. Furthermore, EMSA showed the formation of FUBP1–exon 7 complexes are enhanced in a dose-dependent manner by **SMN-C3**. Based on these results, it was concluded that derivatives of **risdiplam** interact with ESE2 to induce conformational changes in exon 7 and improve the binding affinity of positive splicing factors (Fig. 5D).

In summary, **risdiplam** has been proposed to have two modes of action: (i) stabilizing the RNA duplex of exon 7 5′ splice site and U1 snRNP and (ii) inducing conformational changes of exon 7 ESE2 to facilitate the formation of a complex with positive splicing factors. These two modes of action may contribute to **risdiplam**'s high selectivity.

5.2 Small molecule modulation of *MAPT* pre-mRNA splicing

The small molecules described above direct *SMN2* splicing towards exon 7 inclusion. However, many diseases are caused by aberrant exon inclusion and therapeutic benefit is achieved by exclusion of exons. One such example is tauopathies, caused by aggregation of the protein tau, a regulator of microtubule stability that is highly expressed in neurons.[111] The microtubule-associated protein tau (*MAPT*) gene encoding tau is composed of 16 exons and is known to produce six isoforms by alternative splicing.[87] Exclusion of exon 10 produces the 3R isoform, with three microtubule binding domains (MBDs), while inclusion produces the aggregation-prone 4R isoform, with four MBDs.[112] The ratio of 3R tau to 4R tau is nearly equal in healthy adults (Fig. 6A).[113] However, in frontotemporal dementia (FTD) and Parkinsonism linked to chromosome 17 (FTDP-17), genetic mutations of the *MAPT* gene increase the rate of exon 10 inclusion, and hence the ratio of 4R/3R tau.[114] This causes aggregation of tau proteins and ultimately neuronal death.[114]

The 5′ splice site of *MAPT* exon 10 forms a stem-loop, known as a splicing regulatory element (SRE).[115,116] Genetic mutations in the SRE destabilize its structure, increasing the rate of exon 10 inclusion in the mature transcript.[115–117] For example, DDPAC is an intronic mutation in which the 14th C downstream from the 5′ splice site is mutated to U, therefore mutating a GC base pair to a GU base pair, thermodynamically destabilizing the SRE by 1.2 kcal mol^{-1}.[117] This results in an ~30 : 1 ratio of 4R : 3R tau isoforms. Thus, thermodynamic stabilization of the tau SRE *via* small-molecule targeting could be a viable therapeutic option.

One of the first studies to demonstrate the ligandability of the SRE in tau exon 10 showed the anticancer drug, **mitoxantrone** (MTX) binds and stabilizes the SRE, resulting in decreased production of the tau 4R isoform (Table S1, ESI†).[118] Zheng *et al.*[118] reported the NMR structure of the tau pre-mRNA-MTX complex, showing MTX interacts with the bulged region of the SRE stem-loop. The elucidation of this structure highlighted the importance of structure-based recognition between RNA and small molecule ligands as it showed the three-dimensional shape of the RNA was necessary for binding to MTX.[118] Additional structure–activity relationships (SAR) were used to optimize MTX's ability to decrease exon 10 inclusion, leading to compounds with enhanced affinity for tau pre-mRNA and increased potency for reducing the levels of 4R tau (Table S1, ESI†).[119]

More recently, Chen *et al.*[120] reported that stabilizing the SRE by small molecule binding to the A bulge present in the SRE structure could inhibit recognition by U1 snRNP and promote exon 10 exclusion in wild-type (WT) and DDPAC tau. Tanimoto score-based similarity searching using a previously reported Inforna hit[121] as a query identified **A-1** as a modulator of the 4R/3R tau ratio (Fig. 6B and Table S1, ESI†). To improve physical properties of **A-1**, *in silico*-based hit expansions were conducted. As a result, **A-2**, **A-3**, and **A-4** were obtained from the pharmacophore modeling of **A-1**, and **A-5** was obtained from structure-based design using the three-dimensional structure of the SRE (Table S1, ESI†).

All five compounds not only decrease the 4R/3R ratio by 50% at 10–25 μM, but also had improved physicochemical properties, including potential for blood–brain barrier (BBB) penetrance, compared to **A-1** (Table S1, ESI†). In particular, the average central nervous system multiparameter optimization (CNS-MPO) score for **A-3**, **A-4**, and **A-5** was 4.8; CNS-MPO scores >4 indicate high potential for brain pentrance.[122]

Target engagement studies using Chem-CLIP confirmed **A-5** binds directly to the *MAPT* pre-mRNA SRE. Furthermore, melting curve analysis showed that hit compounds specifically increased the melting temperature (T_m) of tau SRE, providing experimental evidence that small molecule binding to the A bulge indeed thermodynamically stabilizes the tau SRE and prevents recognition by U1 snRNP (Fig. 6C). To further elucidate the binding mode, three-dimensional structures of the apo-SRE and the compound bound SRE (**A-1**, **A-2**, and **A-5**) were characterized by NMR spectroscopy. In both cases, the RNA duplex was consistent with an A-form helical structure, and all compounds bound to a cavity around the bulged adenine, despite having different binding modes. Altogether, these studies demonstrated that compounds identified using Inforna can be converted to more

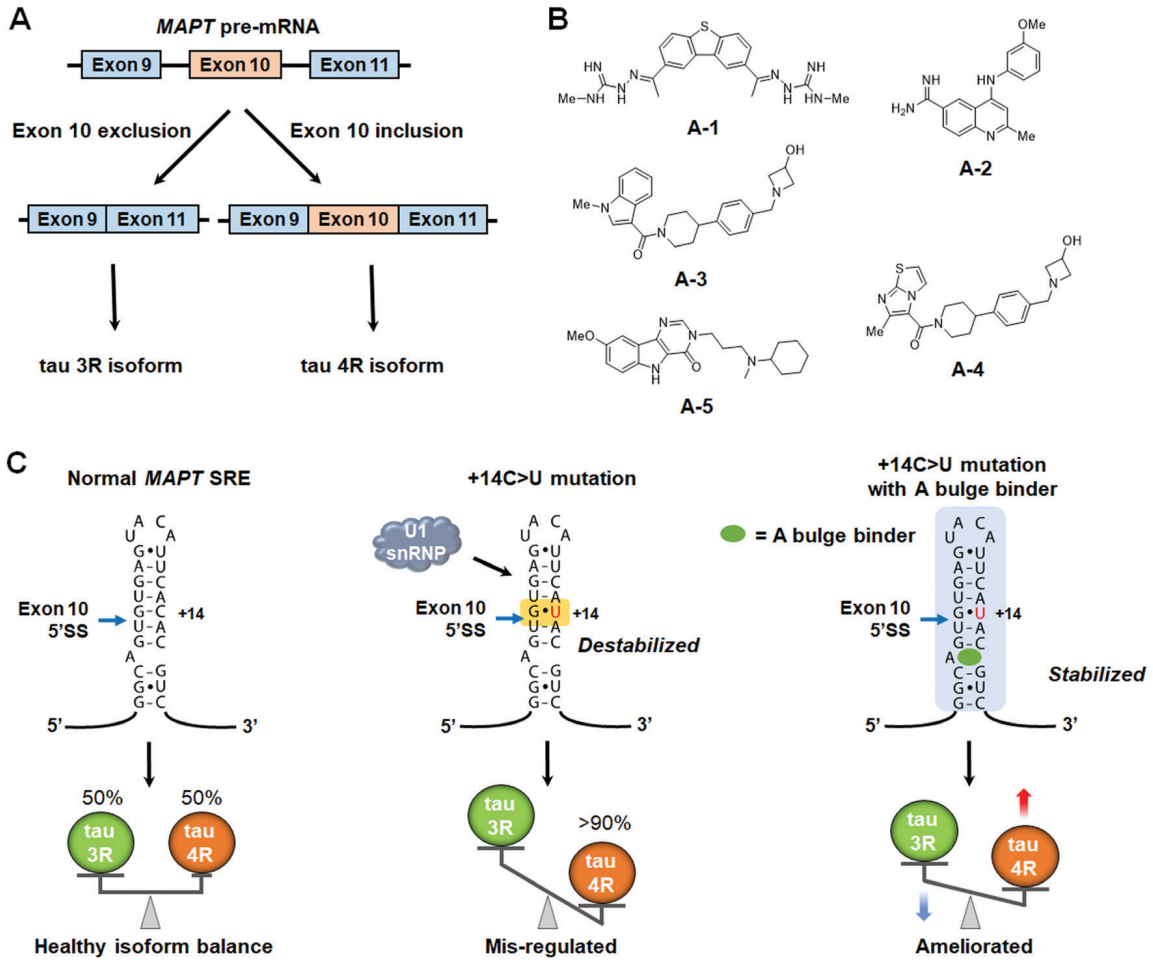


Fig. 6 Small molecule modulation of *MAPT* pre-mRNA splicing. (A) Alternative splicing of *MAPT* exon 10 yields tau 3R and 4R isoforms. (B) Structures of small molecule splicing modulators that bind to the A-bulge of the *MAPT* splicing regulatory element (SRE). (C) Schematic representations showing the effect of U1 snRNP accessibility to the *MAPT* SRE on tau 3R/4R isoform balance.

potent and drug-like compounds possessing the designed RNA-centric mechanism of action.

As is presented here, small molecules can modulate the alternative splicing of pre-mRNAs selectively, either by binding to RNA structural motifs or stabilizing complexes between pre-mRNA and splicing factors. Since aberrant alternative splicing has been associated with various diseases, including rare hereditary diseases,[123] central nervous system disorders,[124,125] and cancers,[126,127] further studies in this field could lead to the development of potent and selective small molecules that can direct splicing outcomes.

6. Small molecules targeting RNA repeat expansion disorders

RNA repeat expansion, or microsatellite, disorders are characterized by long abnormal stretches of repeating RNA nucleotides that can be harbored in intronic, coding, or untranslated regions of pre-mRNAs. These expanded repeats often fold into hairpin structures that interfere with normal RNA processing, leading to disease. Indeed, RNA repeat expansions are responsible for over 30 human diseases, with a large majority being neurodegenerative and neuromuscular in nature.[17] Repeats contribute to disease *via* various mechanisms, including: (i) RNA gain-of-function in which RNA-binding proteins (RBPs) are sequestered and inactivated; (ii) formation of nuclear foci; and (iii) production of toxic proteins, either as a result of canonical translation of an open reading frame (ORF) or as a result of repeat-associated non-ATG (RAN) translation (discussed in Section 8, "Small Molecules Targeting RNA Repeat Expansions Inhibit RAN Translation").

A common RNA gain-of-function pathomechanism in microsatellite disorders is the formation of RNA–protein complexes between the hairpin structures of repeating RNA and RBPs. However, there are various ways by which these complexes lead to disease (Fig. 7). For example, the sequestration of endogenous splicing factors by RNA repeats leads to deregulation of alternative pre-mRNA splicing that affects overall cellular protein levels and homeostasis. Additionally, RNA–protein complexes aggregate in the nucleus in toxic RNA foci, affecting nucleocytoplasmic transport. Thus, the driving idea behind small molecule therapeutics for these disorders is that binding of small molecules competes with RBPs for the disease-causing RNA target, liberating them to fulfill their normal function.

In this section, we focus on four neurodegenerative diseases and their associated RNA–protein complexes: (i) the r(CUG)^{exp}–muscleblind-like 1 (MBNL1) complex in myotonic dystrophy type 1 (DM1) where the repeating nucleotides are indicated in parentheses and "exp" denotes "expanded"; (ii) the r(CCUG)^{exp}–MBNL1 complex causative of myotonic dystrophy type 2 (DM2); (iii) the r(CGG)^{exp}–DGCR8 complex that forms a scaffold for splicing regulators Src-associated in mitosis 68 kDa protein (Sam68) and hnRNP in FXTAS; and (iv) the r(G_4C_2)^{exp}–hnRNP H complex in *C9orf72*-associated frontotemporal dementia and amyotrophic lateral sclerosis (c9FTD/ALS).

6.1 Small molecule inhibition of the r(CUG)^{exp}–MBNL1 complex in DM1

DM1 is an adult-onset neuromuscular disorder caused by an expanded repeating CUG sequence [r(CUG)^{exp}] in the 3′ UTR of dystrophia myotonica protein kinase (*DMPK*) mRNA (Fig. 8A). The expanded RNA affects disease biology by folding into a hairpin structure with a periodic array of internal loops that sequester proteins (RNA gain-of-function), such as the splicing factor MBNL1. Sequestration of MBNL1 deregulates the alternative splicing of the protein's natural substrates, which are directly correlated with disease symptoms. For example, MBNL1 regulates the alternative splicing of the muscle-specific chloride ion channel (CLCN1). In DM1-affected cells, CLCN1's aberrant splicing causes loss of the chloride ion channel from the surface of muscle cells, altering conductance and resulting in myotonia. MBNL1 also self-regulates splicing of its own exon 5. In normal cells, exon 5 is included in the mature mRNA sequence ~45% of the time (Fig. 8B). In DM1-affected cells however, exon 5 is included 85% of the time. In addition, r(CUG)^{exp}–MBNL1 complexes aggregate and form RNA foci in the nucleus that lead to reduction of nucleocytoplasmic transport and result in cytotoxicity.[128,129] Being that the r(CUG)^{exp} hairpin plays a key role in DM1 pathology, this structure has become a promising target for small molecule therapeutics.

In 2013, Rzuczek *et al.*[130] identified a *bis*-benzimidazole derivative (**H**) as a 1 × 1 UU internal loop RNA binder. To target the repeating chain of UU internal loops present in r(CUG)^{exp}, they synthesized a series of **H**-dimers with various linkers. After assessing rescue of DM1-associated splicing, cellular permeability, cytotoxicity, and proteolytic stability of

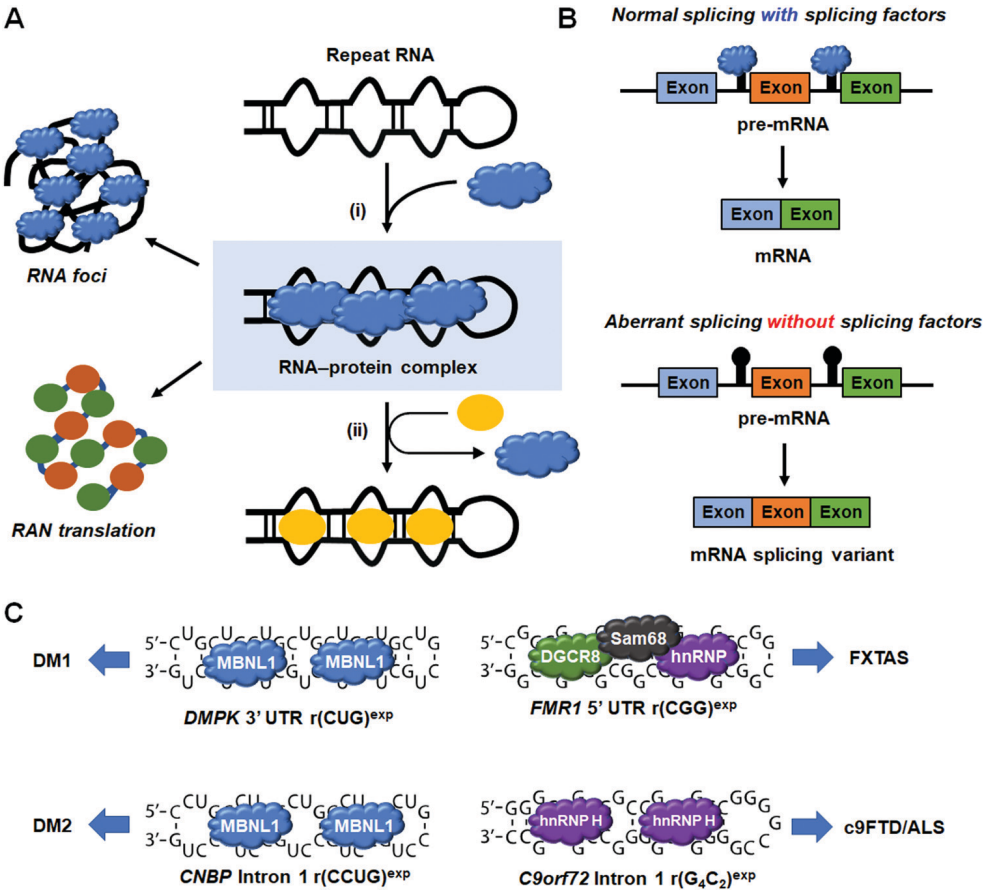


Fig. 7 Small molecule binding of RNA repeat expansions releases sequestered RNA-binding proteins (RBPs). (A) Schematic of RBP sequestration by RNA repeat expansions. (i) RBPs, such as splicing factors, are sequestered by RNA repeat expansions, contributing to disease pathology. (ii) Small molecules competitively bind to RNA repeats and release sequestered proteins, resulting in rescue of splicing defects, reduction in RNA foci, and repression of repeat-associated non-ATG (RAN) translation. (B) Schematic of alternative splicing resulting from the presence or absence of endogenous splicing factors. (C) The RNA:protein complexes that contribute to myotonic dystrophy type 1 (DM1), myotonic dystrophy type 2 (DM2), fragile X-associated tremor and ataxia syndrome (FXTAS), and *C9orf72*-associated frontotemporal dementia and amyotrophic lateral sclerosis (c9FTD/ALS).

the compounds, **2H-K4NMe** was identified as the most promising ligand (Table S1, ESI†). **2H-K4NMe** showed >30-fold binding selectivity to r(CUG)$_{12}$ over other RNA sequences, with a K_d of 13 nM (Table S1, ESI†). Further, **2H-K4NMe** rescued the cardiac troponin T (*cTNT)* splicing defect at a 5 μM dose.

Based on these findings they developed the dimer **2H-K4NMeS**, which displayed enhanced metabolic stability over **2H-K4NMe** (Table S1, ESI†).[131] **2H-K4NMeS** has K_d's of 280 nM and 12 nM for r(CUG)$_{12}$ and r(CUG)$_{109}$, respectively, indicating cooperative binding (Table S1, ESI†). Treatment of DM1-patient-derived cells with as little as 100 nM of **2H-K4NMeS** improved the *MBNL1* exon 5 pre-mRNA splicing defects (Fig. 8D). **2H-K4NMeS** also rescued splicing defects of other MBNL1-regulated splicing events, such as calcium/calmodulin dependent protein kinase II gamma (*CAMK2G)* exon 14 and nuclear receptor corepressor 2 (*NCOR2)* exon 45a splicing, and to a similar extent as *MBNL1* exon 5. This study clearly showed that RNA-binding small molecules can free MBNL1 from RNA–protein complexes at reasonable concentrations for therapeutic use, thereby normalizing splicing events.

Another mechanism by which RNA–protein complexes contribute to DM1 pathology is by aggregating into RNA foci in the nucleus (Fig. 8C). RNA-binding small molecules are expected to disrupt RNA foci by competitively binding to the RNA, preventing protein binding or releasing bound proteins from the complex. Indeed, **2H-K4NMeS** decreased the number of foci present in cells by ~50% when treated at 1 μM.[131] As expected, the activity of **2H-K4NMeS** for improving nucleocytoplasmic transport defects was also observed using a firefly *luciferase* reporter with r(CUG)$_{800}$ in the 3′ UTR. Disruption of RNA foci was also reported after treatment of cells with compound **3**.

To study target engagement, **2H-K4NMeS** was converted into a Chem-CLIP probe, **2H-K4NMeS-CA-Biotin**.[131] This molecule potently rescued splicing defects and reduced the number of

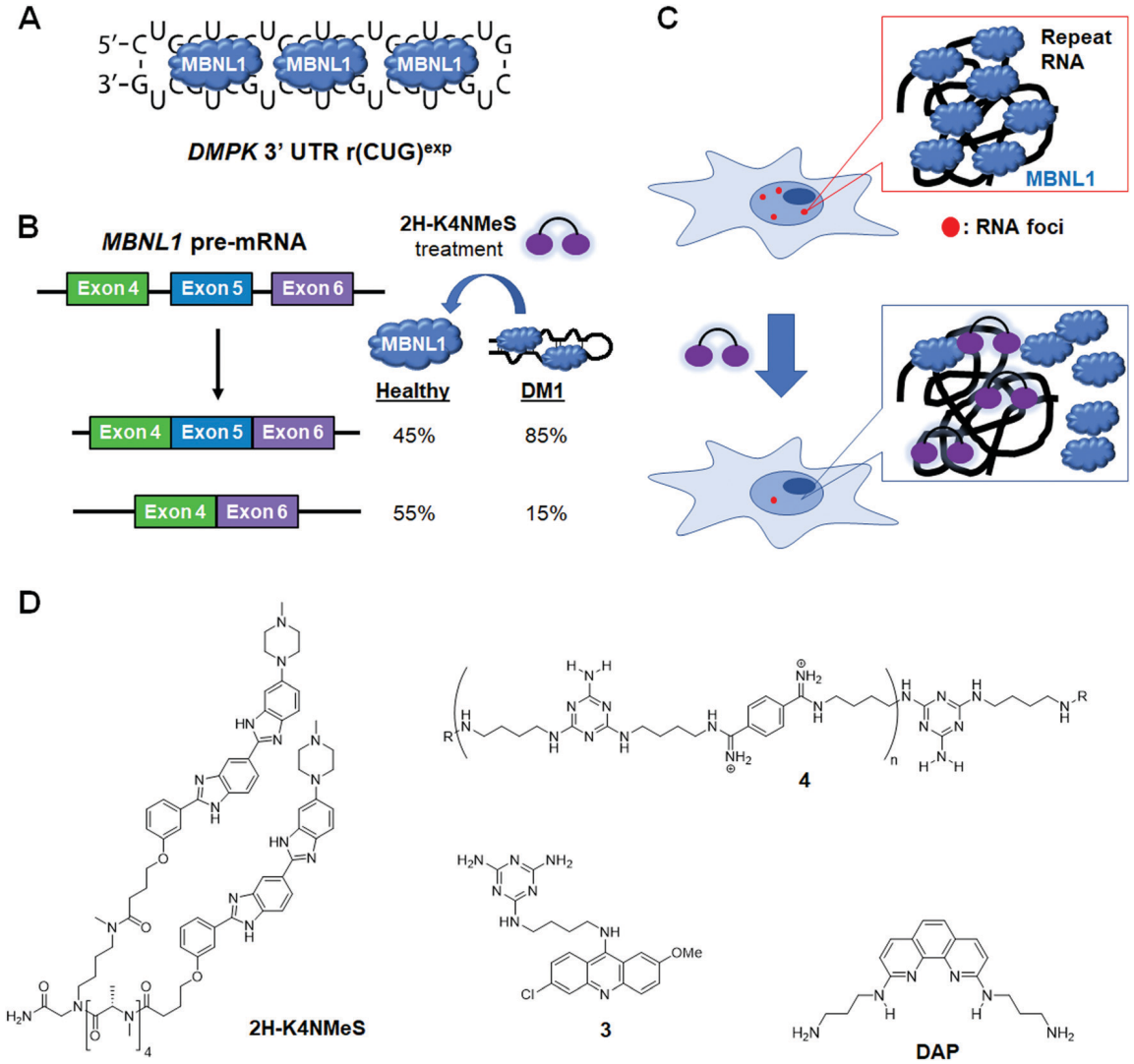


Fig. 8 Small molecule targeting of r(CUG)^{exp}, the RNA causative of myotonic dystrophy type 1 (DM1). (A) r(CUG)^{exp} sequesters MBNL1 protein, which regulates alternative pre-mRNA splicing. (B) MBNL1 regulates self-splicing of its own exon 5. Sequestration of MBNL1 by r(CUG)^{exp} results in exon 5 being included in the mature *MBNL1* transcript too frequently, contributing to DM1 pathology. (C) Schematic representation of RNA foci formation and disruption by small molecule binding. (D) Structures of compounds that bind r(CUG)^{exp}.

nuclear foci when DM1 patient-derived cells were dosed at a 10 nM concentration. In pulled down factions, an ~13 000-fold enrichment of *DMPK* mRNA was observed, as compared to the starting cell lysate. Using the competitive version of Chem-CLIP, C-Chem-CLIP, in which increasing concentrations of **2H-K4NMeS** were co-treated with a constant concentration of **2H-K4NMeS-CA-Biotin**, confirmed **2H-K4NMeS** and **2H-K4NMeS-CA-Biotin** share the same binding site in cells. The specific binding site was further defined by Chem-CLIP-Map,[132] confirming binding of the UU internal loops of r(CUG)^{exp} in the *DMPK* mRNA.

More interestingly, Rzuczek *et al.*[131] demonstrated target-templated oligomerization of an **H**-dimer in cells (Fig. 9). The designed **H**-dimer was modified with azide and alkyne moieties at opposite ends of the molecule, allowing oligomerization upon binding r(CUG)^{exp} through click chemistry. This oligomerization only occurred in DM1 cells, as healthy cells lack the repeating RNA necessary to template the reaction. This *in situ*-produced oligomer rescued splicing defects at concentrations as low as 100 pM in DM1 patient-derived cells.

Arambula *et al.*[133] developed acridine–triaminotriazine conjugate **3** targeting the r(CUG)^{exp} (Table S1, ESI†). They designed **3** based on the complementary Janus-wedge hydrogen bonding between triaminotriazine and the UU internal loops of r(CUG)^{exp}. This bonding is further stabilized by stacking interactions of the acridine moiety (Fig. 8D). ITC, using a model RNA hairpin, r(CUG)$_4$, revealed **3** has a K_d of 430 nM and 1 : 1 binding stoichiometry (Table S1, ESI†). However, **3** also binds with similar avidity to d(CTG)$_2$ duplex with a K_d of 390 nM. *In vitro*, **3** inhibits r(CUG)$_4$–MBNL1 complex formation with an IC_{50} of 52 μM and a K_i of 6 μM to r(CUG)$_4$, similar to values observed for r(CUG)$_{12}$. To capitalize on the multiple binding sites (UU internal loops) of the target, bivalent derivatives of **3** were developed.[134] A bivalent ligand containing an oligoamino linker was deemed the most superior with improved properties such as aqueous solubility and cell permeability compared to monomeric **3**. The dimer inhibited formation of RNA foci in a transfected cellular model of DM1 at 20 μM, and almost complete disruption at 50 μM.

In 2016, Luu *et al.*[135] demonstrated that dimerization of a dimeric compound which has two triaminotriazines linked with bisimidate produced a potent inhibitor of the r(CUG)^{exp}–MBNL1 complex. This intricate ''dimer of dimers'', has 1000-fold improved potency *in vitro* (K_i of 25 nM) compared to the original dimer. This molecule reduced RNA foci by ~20% when treated at 1 μM in cells, significantly improved splicing defects of insulin receptor (*IR*) exon 11 (10 μM dose in cells), and alleviated disease phenotypes in a *Drosophila* model of DM1. However, due to the compounds high molecular weight, it had issues with cellular uptake. To overcome this weakness, Lee *et al.*[136] developed oligomeric ligand **4**, composed of triaminotriazine units (targeting the UU internal loops of r(CUG)^{exp}) and bisimidate units (targeting the major groove of RNA) (Fig. 8D and Table S1, ESI†). Although **4** is still too large to permeate the cell membrane, its poly-cationic nature makes it membrane penetrant by endocytosis. Using 200 nM of **4**, they showed full rescue of *IR* mis-splicing and a decrease in foci number in a transfected model of DM1. However, **4** also inhibits transcription of d(CTG)^{exp}, indicating the compound is not specific for the RNA repeat (Table S1, ESI†). They used adult DM1 *Drosophila* (CTG_{480}) to investigate the *in vivo* effects of **4** by measuring the improvement of climbing defects observed after treatment with the compound. Approximately 80% of untreated files show significant defects in their ability to climb, but this was rescued by treatment with **4** (80 μM; 37% fail to climb). In addition, in a liver-specific DM1 mouse model containing 960 interrupted CUG repeats, **4** decreased the levels of the transgene, likely due to the compound's inhibitory effect on d(CTG)^{exp} transcription, improved pre-mRNA splicing defects, and reduced RNA foci formation, highlighting the compound's potential in preclinical animal models.

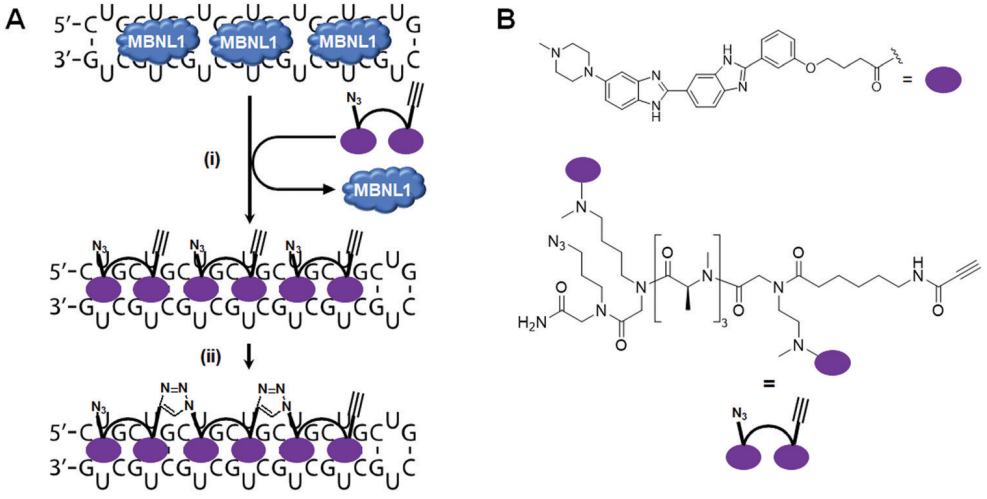


Fig. 9 RNA-templated ligand oligomerization catalyzed by r(CUG)^{exp}. (A) *In cellulis* click chemistry, templated by the RNA repeat expansion, forms an oligomeric compound on-site, that is bound to the r(CUG)^{exp} RNA target. (i) MBNL1 sequestered by r(CUG)^{exp} is released upon binding of the dimeric click compound. (ii) The azide terminus of one dimer reacts with the alkyne terminus of another dimer in close proximity to synthesize an oligomer *in cellulis*. (B) Structures of the RNA binding motif and dimeric click compound that oligomerizes *in cellulis*.

As another example of an r(CUG)^{exp} binding molecule, Li *et al.*[137] designed a 1,10-phenanthroline derivative (**DAP**) and studied its effect by *in vitro* translation (Fig. 8D and Table S1, ESI†). Using a transfected template RNA with r(CUG)$_{20}$ inserted between *Renilla* luciferase (*R*luc) and *firefly* luciferase (*F*luc) showed treatment with **DAP** suppressed translation of *F*luc downstream of the repeat sequence in a concentration-dependent manner. The translation of *R*luc was also moderately affected by **DAP** treatment. The selectivity of **DAP** was assessed by SPR and melting temperature, revealing **DAP** shows preferential binding to r(CUG)$_9$ and r(CCG)$_9$ among r(CXG)$_9$ sequences (X = A, U, G, or C) (Table S1, ESI†). Furthermore, electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) analysis showed **DAP** binds to r(CUG)$_9$ with an RNA:compound ratio of 1 : 4.

6.2 Small molecule inhibition of the r(CCUG)^{exp}–MBNL1 complex in DM2

DM2 is caused by r(CCUG)^{exp} in intron 1 of CCHC-type zinc finger nucleic acid binding protein (*CNBP*) pre-mRNA (Fig. 10A). As observed in DM1, r(CCUG)^{exp} sequesters MBNL1, causing MBNL1-dependent splicing defects and RNA foci formation, but also causes aberrant splicing of *CNBP* intron 1 (intron retention). To target DM2, Lee *et al.*[138] developed a dimeric kanamycin compound (**5**) that inhibits formation of r(CCUG)$_{12}$–MBNL1 complexes *in vitro* with an IC$_{50}$ of ∼90 nM (∼2500-fold more potent than the monomer; IC$_{50}$ = ∼220 μM) (Table S1, ESI†). Compound **5** demonstrated good cellular permeability and localized to both the nucleus and cytoplasm. In DM2 fibroblasts, **5** (10 μM) successfully rescued *IR* splicing defects and significantly reduced the number of RNA foci (Fig. 10B and C).[139] These activities were further improved by the incorporation of a cleavage module on the ligand (discussed in Section 10.4, "Targeted Cleavage of r(CCUG)^{exp} by a Small Molecule–Bleomycin A5 Conjugate").

Similar to the case shown with DM1 (Fig. 9), incorporation of azide and alkyne moieties into the kanamycin RNA-binding module afforded target-templated oligomerization in DM2 patient-derived cells.[140] When DM2 fibroblasts were treated with this clickable molecule (1 μM), the number of foci observed was reduced by ∼45% and *IR* exon 11 splicing defects were rescued by approximately the same percentage. These results clearly indicated the activity of the compound was far improved by on-site, *in situ* oligomerization. This oligomeric molecule also affected aberrant splicing of *CNBP* mRNA by inhibiting binding of MBNL1 to intron 1, thus allowing the intron to be properly spliced out of *CNBP* pre-mRNA (discussed in Section 7, "Small Molecules Shunt Toxic RNA to Endogenous Decay Pathways").

6.3 Small molecule inhibition of the r(CGG)^{exp}–protein complexes in FXTAS

In FXTAS, expanded r(CGG) repeats of lengths >55 but <200 (premutation allele) in the 5′ UTR of *FMR1* mRNA cause disease (Fig. 11A). The repeat folds into a hairpin structure with repeating 1 × 1 GG internal loops that sequester several proteins, such as DGCR8, Sam68, and hnRNP. Because these proteins have important roles in RNA biogenesis, their sequestration alters pre-mRNA splicing, thus resulting in disease.

Disney *et al.*[79] previously identified compound **1a** as a binder to r(CGG)^{exp} by screening compounds using a time-resolved fluorescence resonance energy transfer (TR-FRET) assay that monitors r(CGG)$_{12}$–DGCR8Δ complex formation and SAR (Fig. 11C and Table S1, ESI†). In particular, **1a** disrupted the r(CGG)$_{12}$–DGCR8Δ complex with an IC$_{50}$ of 12 μM, in the presence of competitor tRNA. Sequestration of Sam68 by r(CGG)^{exp} dysregulates splicing of *SMN2* mRNA, therefore the ability of **1a** to improve Sam68-regulated splicing defects was assessed (Fig. 11B). In transfected COS7 cells, r(CGG)^{exp} causes *SMN2* exon 7 to be included too frequently (∼70% compared to 30% in healthy cells). Upon

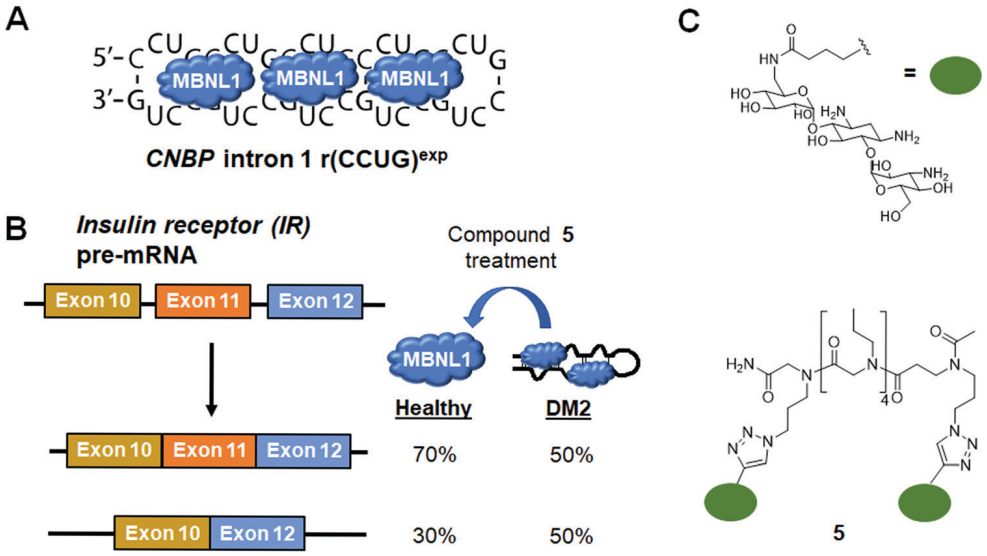


Fig. 10 Small molecule targeting of r(CCUG)^{exp}, the causative agent of myotonic dystrophy type 2 (DM2). (A) r(CCUG)^{exp} sequesters MBNL1. (B) MBNL1 regulates splicing of insulin receptor (*IR*) exon 11. Aberrant splicing results in exon 11 being excluded from the mature *IR* transcript, contributing to DM2 pathology. (C) Structures of the kanamycin RNA-binding motif and dimeric compound that bind r(CCUG)^{exp}.

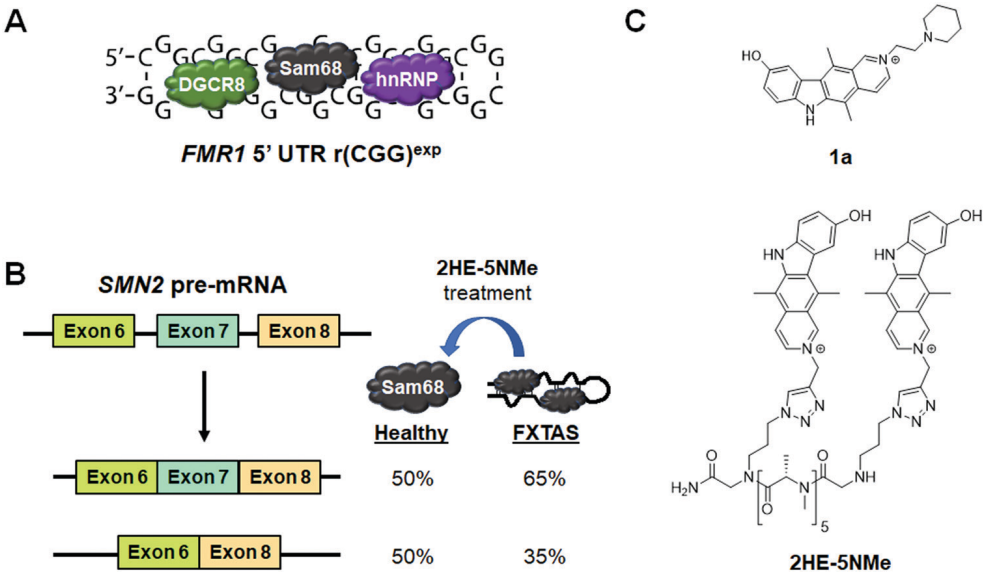


Fig. 11 Small molecule targeting of r(CGG)^{exp}, the RNA causative of fragile X-associated tremor ataxia syndrome (FXTAS). (A) r(CGG)^{exp} sequesters proteins such as DGCR8, Sam68, and hnRNP. (B) Sam68 regulates splicing of *SMN2* exon 7. Thus, its sequestration results in increased exon 7 inclusion in the mature *SMN2* transcript, contributing to FXTAS pathology. (C) Structures of compounds that bind to r(CGG)^{exp}.

treatment of these cells with **1a** (20 μM), improvement of the *SMN2* splicing defect was observed while improvement of another Sam68-regulated splicing event, apoptosis regulator Bcl-X (*Bcl-x*) exon 2, was observed upon treatment with 100 μM of **1a** (Table S1, ESI†). **1a** (10 μM) also reduced the number of RNA foci, as studied by RNA fluorescence *in situ* hybridization (FISH).[79]

As discussed previously, dimerization of RNA-binding modules is a powerful and easy method to obtain highly potent and selective compounds. Thus, a dimeric derivative of **1a**, **2HE-5NMe**, was designed and studied (Fig. 11C and Table S1, ESI†).[141] Inhibition of the r(CGG)$_{12}$–DGCR8Δ complex by **2HE-5NMe** was assessed by TR-FRET and revealed the compound inhibits complex formation with 6-fold greater activity than monomeric **1a** (IC_{50} = 3.5 μM in the presence of tRNA). Further, **2HE-5NMe** has 16-fold higher affinity for r(CGG)$_{12}$ (K_d = 50 ± 0.6 nM) than **1a**, which translates into a 3-fold higher occupancy *in cellulis*, as revealed by Chem-CLIP studies (Table S1, ESI†).[141]

The activity of **2HE-5NMe** for rescuing splicing defects was assessed by exon 7 inclusion in *SMN2* mRNA. Treatment of **2HE-5NMe** at 50 μM rescued exon 7 inclusion levels back to those observed in wild type cells, demonstrating a >10-fold increase in activity over **1a**. While **1a** can inhibit foci formation but not disrupt existing foci, **2HE-5NMe** has the ability to do both (~70% reduction at 50 μM). It should be noted that most foci in this study were dissolved within 1 h of treatment and fully dissolved after 4 h. Recovery of *SMN2* splicing defects were observed in parallel to this time course. The effect of these compounds on RAN translation is discussed in Section 8.2, "Small Molecules Targeting the r(CGG)^{exp} in *FMR1* Inhibit RAN Translation".

6.4 Small molecule inhibition of the r(G$_4$C$_2$)^{exp}–hnRNP H complex in c9FTD/ALS

An expanded repeat of G_4C_2 [r(G$_4$C$_2$)^{exp}] in intron 1 of *C9orf72* mRNA is the most common genetic cause of the neurodegenerative disease c9FTD/ALS (Fig. 12A). The structure of r(G$_4$C$_2$)^{exp} has been well-studied, revealing the repeating RNA can adopt two main structures, a hairpin with an array of internal loops and a G-quadruplex. Because the hairpin form of r(G$_4$C$_2$)^{exp} forms the same 1 × 1 GG internal loops as r(CGG)^{exp}, Su *et al.*[142] hypothesized that **1a** might also bind to r(G$_4$C$_2$)^{exp}. Using **1a** as a lead, a library of chemically similar compounds was created and screened for binding r(G$_4$C$_2$)$_8$ using a dye displacement assay. The screen revealed **1a** and two additional compounds, **6** and **7**, bind r(G$_4$C$_2$)^{exp} (Fig. 12B and Table S1, ESI†), with K_ds of 9.7, 10, and 16 μM, respectively (Table S1, ESI†). To assess the biological activities of each compound, foci formation was evaluated in r(G$_4$C$_2$)$_{66}$-expressing COS7 cells. Compounds **1a** and **6**, but not **7**, showed a 3-fold reduction of foci-positive cells. This reduction in foci by **1a** can be traced to its direct engagement of r(G$_4$C$_2$)^{exp}, as determined by Chem-CLIP, which revealed an 80-fold enrichment of r(G$_4$C$_2$)$_{66}$ in the pulled down fractions, as compared to 18S rRNA.[142] C-Chem-CLIP studies where r(G$_4$C$_2$)$_{66}$-expressing COS7 cells were co-treated with **1a** and its Chem-CLIP probe verified target engagement by the parent compound.[142]

Interestingly, *C9orf72* mRNA is bidirectionally transcribed and thus repeats from the sense [r(G$_4$C$_2$)^{exp}] and antisense [r(G$_2$C$_4$)^{exp}] strand are produced. Like the sense strand, r(G$_2$C$_4$)^{exp} also forms nuclear inclusions. However, the antisense foci were not reduced by the treatment of **1a**, confirming its selectivity for the sense strand. Furthermore, **1a** showed significant reduction of RNA foci-positive cells in three *C9ORF72*+ induced neuron (iNeuron) lines.

In a subsequent study,[143] **1a** was further optimized, affording **8** (Fig. 8 and Table S1, ESI†). Compound **8** binds to r(G$_4$C$_2$)$_8$ with a K_d of 0.26 μM, while showing ~300-fold weaker binding to antisense r(G$_2$C$_4$)$_8$ and ~540-fold weaker binding to base-pair control r(G$_2$C$_2$)$_8$ (Table S1, ESI†). With the remarkable binding

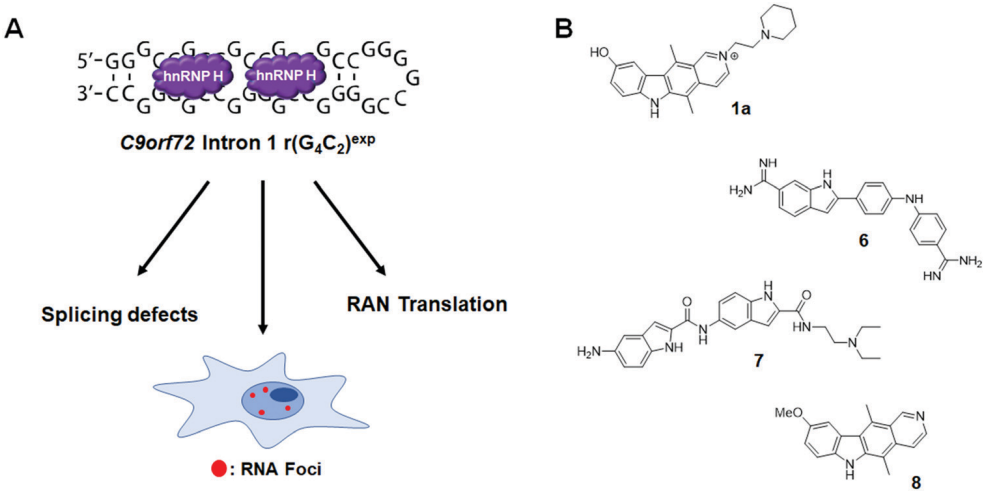


Fig. 12 Small molecule targeting of $r(G_4C_2)^{exp}$, the most common genetic cause of *C9orf72*-associated frontotemporal dementia and amyotrophic lateral sclerosis (c9FTD/ALS). (A) $r(G_4C_2)^{exp}$ sequesters hnRNP H, resulting in splicing defects. The repeat expansion also undergoes RAN translation and forms RNA foci. (B) Structures of compounds that bind to $r(G_4C_2)^{exp}$. Compound **1a** also binds to $r(CGG)^{exp}$ due to the 5′-CGG/3′-GGC binding site shared between the two repeats.

affinity of **8**, the binding mechanism was further investigated by NMR spectroscopy. In brief, **8** stacks between GG internal loops and closing GC base pairs to stabilize the closing base pairs with π–π interactions. *In vitro*, **8** inhibited the $r(G_4C_2)_8$-hnRNP H complex with an IC_{50} of 19 μM and reduced both the number of foci-positive cells and the number of foci present per cell by half in HEK293T cells transfected with a plasmid encoding $r(G_4C_2)_{66}$ (5 μM dose). Therefore, **8** is a potent and selective small molecule capable of alleviating disease-associated phenotypes in cellular models of c9ALS/FTD. The inhibition of RAN translation by **8** is discussed in Section 8.3, "Small Molecules Targeting the $r(G_4C_2)^{exp}$ in *C9orf72* Inhibit RAN Translation". Most importantly, these studies with **8** revealed that the hairpin form of $r(G_4C_2)^{exp}$, not the G-quadruplex, is RAN translated.

7. Small molecules shunt toxic RNAs to endogenous decay pathways

As discussed in Section 6.2 ("Small Molecule Inhibition of the $r(CCUG)^{exp}$-MBNL1 Complex in DM2"), $r(CCUG)^{exp}$ causes *CNBP* intron 1 retention. Although formation of nuclear foci and splicing defects have been well studied in DM2, intron retention was only recently discovered by the Swanson group (~40% retained in DM2-affected cells *vs.* ~10% in healthy cells) (Fig. 13A).[144] Intronic regions of pre-mRNAs are normally subjected to endogenous decay upon liberation, but in DM2 the intron containing the repeat expansion remains present in the mature mRNA transcript.[145] Shortly after this discovery, **5**, previously reported to target $r(CCUG)^{exp}$ and alleviate DM2-associated defects, was employed as a chemical probe to investigate the mechanism of intron retention (Table S1, ESI†).[140] These studies showed that binding of MBNL1 causes intron retention and that small molecules can alleviate this retention by shunting the intron to endogenous decay pathways.

Treatment of DM2 patient-derived cells with **5** (1–10 μM) led to ~15–20% of the retained intron being eliminated, while no effect was observed on *CNBP* mature mRNA levels (Fig. 13B),[140] suggesting a mechanism by which small molecule binding of the $r(CCUG)^{exp}$ shunts pathogenic RNAs to endogenous quality control pathways. Notably, there are a variety of disease-causing RNA repeats harbored in introns that lead to intron retention, such as c9FTD/ALS caused by $r(G_4C_2)^{exp}$ and Fuchs endothelial corneal dystrophy (FED) caused by $r(CUG)^{exp}$. Small molecule intervention in these cases may have similar cooperative effects with endogenous RNA decay mechanisms to be therapeutically advantageous.

8. Small molecules targeting RNA repeat expansions inhibit RAN translation

8.1 RAN translation in microsatellite diseases

An additional pathomechanism found in some neurodegenerative RNA repeat expansion disorders, such at $r(CGG)^{exp}$ and $r(G_4C_2)^{exp}$, is RAN translation.[146–150] In this phenomenon, repeat expansions serve as non-canonical translation initiation sites, thus giving rise to homopolymeric, as in the case of $r(CGG)^{exp}$,[149,150] or dipeptide repeat (DPR) proteins, as in the case of $r(G_4C_2)^{exp}$.[146,148] These proteins are intrinsically disordered and form neurotoxic aggregates that contribute to disease pathology.[151] Therefore, small molecules that inhibit RAN translation are of high therapeutic importance.

8.2 Small molecules targeting the $r(CGG)^{exp}$ in *FMR1* inhibit RAN translation

In FXTAS, RAN translation produces the homopolymeric protein poly(G) (Fig. 14A).[149,150] As **1a** and **2HE-5NMe** (Fig. 14B and Table S1, ESI†) were shown to bind $r(CGG)^{exp}$ selectively both

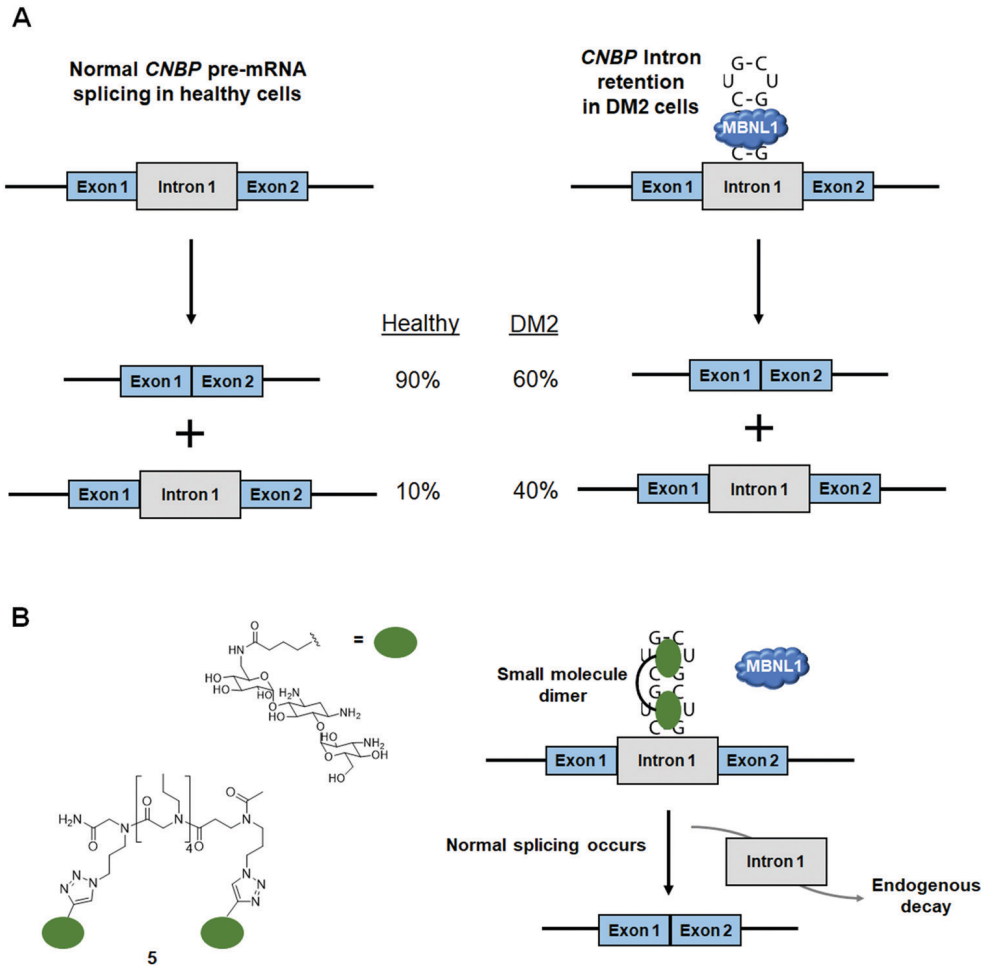


Fig. 13 Small molecule binding causes toxic RNAs to be shunted to endogenous decay pathways. (A) MBNL1 sequestration by r(CCUG)^{exp} results in *CNBP* intron 1 retention. (B) Small molecule binding frees MBNL1 and allows for proper intron splicing to occur. The excised intron is shunted to endogenous decay pathways.

in vitro and *in cellulis* (as determined by Chem-CLIP),[79,141] their ability to inhibit RAN translation was also assessed. Interestingly, both **1a** and **2HE-5NMe** thermally stabilize r(CGG)$_{12}$ (by 1.4 and 0.9 kcal mol^{-1} respectively),[141] suggesting they may prevent ribosomal readthrough or loading and thereby inhibit RAN translation. In agreement with their similar degree of stabilization of r(CGG)$_{12}$, **1a** and **2HE-5NMe** inhibited RAN translation to a similar extent (~80% inhibition at 50 μM) as well as reduced the number of poly(G) nuclear inclusions.[141] Notably, both compounds reduced polysome loading onto r(CGG)$_{88}$, as hypothesized.[79,141] Importantly, neither compound affects mRNA levels or canonical translation of the downstream ORF.[79,141]

To date, the most potent inhibitor of r(CGG)^{exp} RAN translation is the covalent cross-linker **2H-5-CA-Biotin** (Table S1, ESI†).[152] **2H-5-CA-Biotin** selectively engaged the RNA target in cells and inhibited RAN translation at a dose of only 500 nM (~40% decrease in poly(G) levels) while not affecting canonical translation of the downstream ORF.[152] Additionally, polysome profiling indicated that **2H-5-CA-Biotin** disrupts polysome loading onto r(CGG)^{exp}-containing transcripts.

8.3 Small molecules targeting the r(G$_4$C$_2$)^{exp} in *C9orf72* inhibit RAN translation

In c9FTD/ALS, RAN translation gives rise to five DPRs.[151] Poly(GA) and poly(GR) are translated from the sense strand [r(G$_4$C$_2$)^{exp}], while poly(PA) and poly(PR) are translated from the antisense strand [r(G$_2$C$_4$)^{exp}].[153] Poly(GP) is RAN translated from both strands and is highly expressed in the central nervous system. Additionally, it is the most soluble of the DPRs, making its detection facile (Fig. 14A).[146,151,154] In agreement with its ability to alleviate another c9FTD/ALS-associated defect (nuclear inclusions), **1a** dose-dependently reduced levels of poly(GP) by 10%, 18%, and 47% in iNeurons treated at 25, 50, or 100 μM, respectively.[142] Likewise, the **1a** derivative **8** dose-dependently inhibited RAN translation in HEK293T cells expressing r(G$_4$C$_2$)$_{66}$ (IC$_{50}$ = 1.6 ± 0.20 μM), while having no effect on canonical translation (Fig. 14B).[143] Polysome profiling upon treatment with **8** showed the amount of r(G$_4$C$_2$)$_{66}$ transcripts loaded into polysomes was significantly decreased for high and low molecular weight polysomes and monosomes, indicating that **8** works by sterically blocking the assembly of ribosomes onto r(G$_4$C$_2$)^{exp}, thus reducing the levels of toxic DPRs produced.[143]

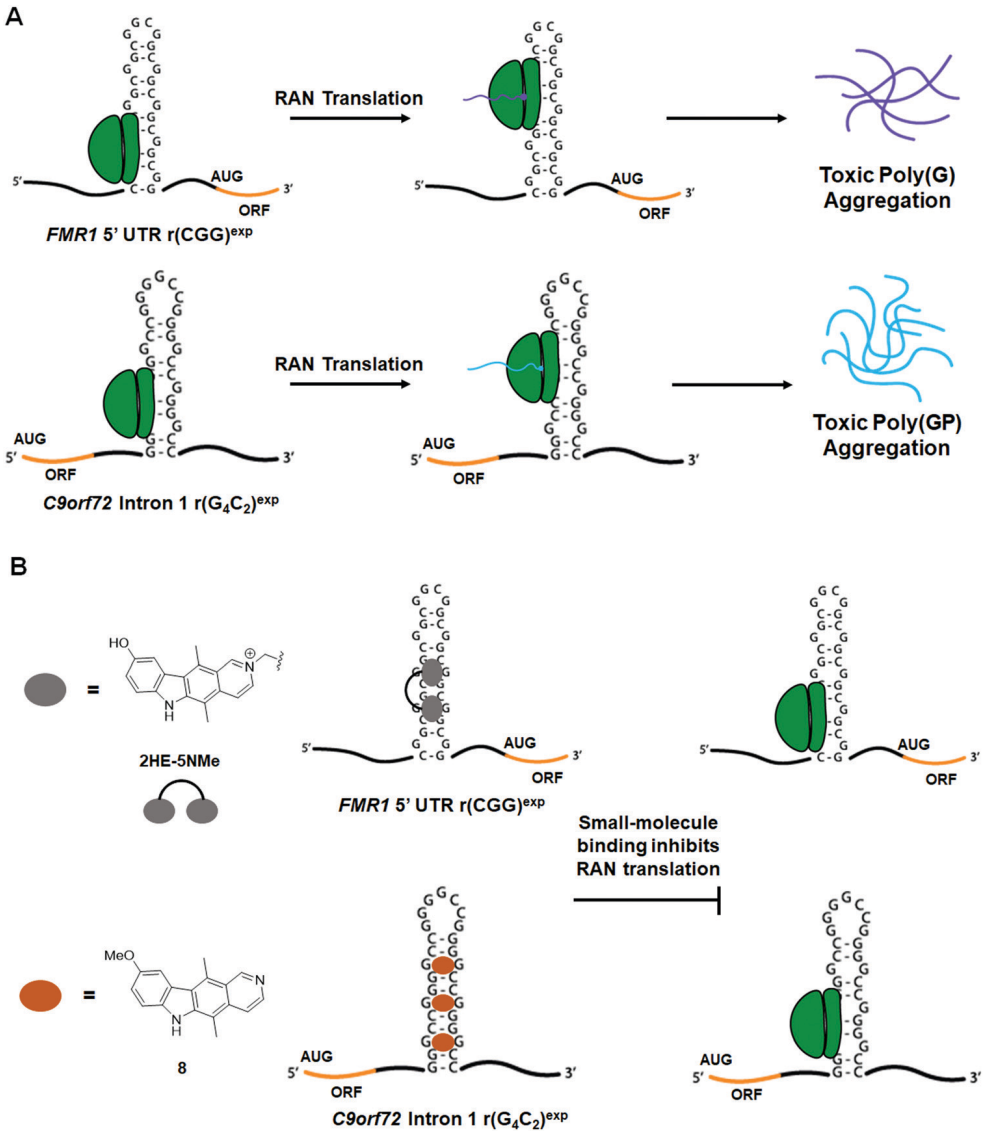


Fig. 14 Small molecule targeting of RNA repeat expansions reduces aberrarnt repeat-associated non-ATG (RAN) translation. (A) Schematic of RAN translation of *FMR1* due to r(CGG)^{exp} in the 5′ UTR and *C9orf72* due to r(G$_4$C$_2$)^{exp} in intron 1. (B) Small molecules targeting r(CGG)^{exp} and r(G$_4$C$_2$)^{exp} inhibit RAN translation.

9. Small molecules inhibit translation of traditionally undruggable proteins

Over the past decades, tremendous efforts have been invested in the development of small molecules targeting disease-causing proteins, and yet only 15% of proteins are considered "druggable" from genome-wide analysis.[155,156] One major roadblock in drugging the remaining 85% is their lack of defined structures that can serve as potential small molecule binding pockets.[157,158] To overcome this challenge, an alternative strategy, especially useful for proteins with aberrantly high expression levels, is to target their encoding mRNA specifically with small molecules and hence inhibit downstream translation. A recent example of this is the development of a small molecule targeting the α-synuclein mRNA,[159] which encodes an intrinsically disordered protein (IDP) key to the pathogenesis of Parkinson's disease (PD).[160]

The α-synuclein protein, encoded by the *SNCA* gene, can oligomerize to form fibrils across neurons in the brain as well as accumulate in Lewy bodies and Lewy neurites, contributing to the risk of developing PD (Fig. 15A).[161,162] Since individuals with multiplication of the *SNCA* gene locus develop dominantly inherited PD with a gene-dosage effect,[163] reducing the level of α-synuclein expression could be a promising disease-alleviating strategy.[164,165] As an IDP, α-synuclein is challenging to target. The *SNCA* mRNA, however, displays a functionally important and structurally defined 5′ UTR with an iron responsive element (IRE) that provides opportunities for small molecule targeting.[166,167] Indeed, employment of the sequence-based design and lead identification strategy, Inforna (discussed in Section 2.3, "Design of Monomeric Small Molecules Targeting Disease-Causing miRNAs"),[41] yielded a set of small molecules that bind the IRE region of *SNCA* mRNA. These initial hits were

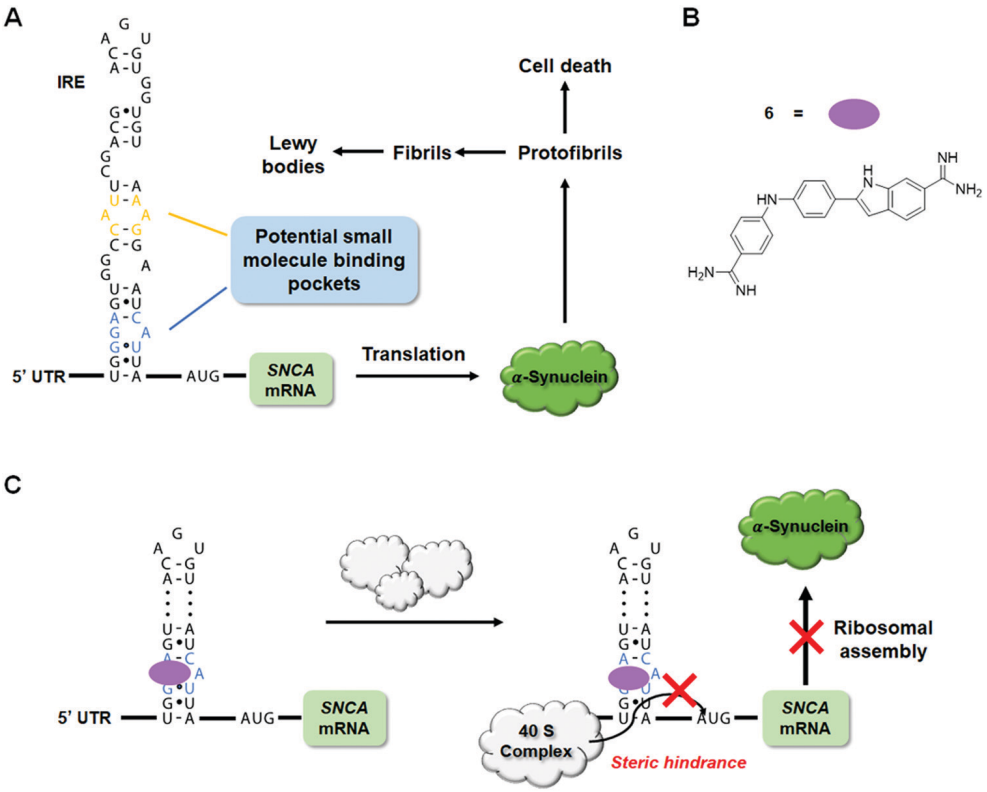


Fig. 15 RNA-targeted small molecules inhibit translation of traditionally undruggable proteins. (A) Schematic depiction of α-synuclein-mediated disease pathway. (B) Structure of **Synucleozid**, an Inforna-designed small molecule that binds the A-bulge of the *SNCA* IRE that regulates translation of the mRNA. (C) **Synucleozid** targets the IRE structure of *SNCA* mRNA and represses α-synuclein protein expression by inhibiting ribosomal assembly onto the *SNCA* transcript.

subjected to a western blot screen of α-synuclein inhibition potency in neuroblastoma cells, with the most potent compound, **Synucleozid**, exhibiting an IC_{50} ~500 nM (Fig. 15B and Table S1, ESI†).[159] To ensure that inhibition occurs at the translational and not transcriptional level, RT-qPCR was used to confirm the level of *SNCA* mRNA remained constant upon treatment of **Synucleozid**.

It should be noted, however, that observation of the expected biological effects can only support, not validate, the putative binding mode of a small molecule. To validate the A bulge of the IRE of *SNCA* mRNA as the binding site of **Synucleozid**, competitive binding assays and mutational analyses were performed. Indeed, mutations of the A bulge to either a U/G bulge or a base pair reduced the binding affinity of **Synucleozid** by 10-fold, while mutations of other non-canonical base pairs had no effect on **Synucleozid** binding avidity. Furthermore, ASO-Bind-Map[18,168] was used to confirm the binding of **Synucleozid** to the IRE both *in vitro* and *in cellulis*. Briefly, ASO-Bind-Map relies on ASO-mediated RNA cleavage, *via* RNase H, which can be inhibited by small molecule binding of the RNA target. Small molecule binding thermodynamically stabilizes the RNA and impedes ASO binding at the binding site. *In vitro*, protection from RNase H cleavage can be read out by gel electrophoresis while RT-qPCR or RNA-seq can be used to read out protection by the binding small molecule *in cellulis*. In this case, treatment of **Synucleozid** impeded cleavage of the IRE, indicating that **Synucleozid** indeed binds to the IRE and stabilizes its structure.

In addition to its intrinsic specificity for the binding pocket on RNA, the overall specificity of a small molecule depends on the prevalence of the structured pocket across the entire human transcriptome. In other words, a small molecule that is specific to its target binding pocket would still suffer from off-target effects if this binding pocket is shared by other non-target RNAs. The selectivity of **Synucleozid** was first assessed by studying its effect on mRNAs expressed in the nervous system that contain known IREs in their UTRs, including amyloid precursor protein (APP), prion protein (PrP), ferritin and the transferritin receptor (TfR). Upon treatment of **Synucleozid** (1 μM), no effect was observed upon APP, PrP, or TfR levels, but ferritin levels were reduced by ~50%. This reduction could be the result of an off-target effect of **Synucleozid** or could be due to compound-mediated rescue of autophagic and lysosomal dysfunction caused by an accumulation of α-synuclein in PD.[169] While future studies are necessary to elucidate **Synucleozid**'s effect on ferritin levels, the compound demonstrated overall high selectivity for *SNCA* mRNA due to its unique structure in the 5′ UTR. Moreover, the targeted 5′-G_G/3′-CAU region was searched across a database of structural elements expressed in the human transcriptome, including miRNA hairpin precursors (7436 motifs) and 2459 other known motifs from rRNA, RNase P RNA, U4/U6 snRNA, and nonredundant tRNAs.[159,170]

Remarkably, the bulge targeted by **Synucleozid** only occurs five times among these motifs (0.051%) and fortuitously, the potential miRNA off-targets are not expressed at appreciable levels.[170] Not surprisingly, **Synucleozid** had no effect on their expression. A transcriptome-wide assessment of **Synucleozid** treatment using RNA-seq revealed very few changes (55/20 034 genes changed; 0.3%).[159,171] Similarly, a proteome-wide selectivity analysis also showed limited changes (283/3300 proteins changed; 8%). Collectively, these data support the fundamental claim that RNAs can indeed form unique 3D structures suitable for targeting with small molecules, therefore expanding the druggability of proteins broadly. Notably, this assertion is bolstered by studies that direct the splicing outcome of *MAPT* exon 10 (tau), another IDP.[120]

10. Targeted cleavage of disease-causing RNAs using bleomycin A5-conjugates

10.1 Bleomycin A5 cleavage of miRNAs

Bleomycin A5 is a well-known, DNA-cleaving natural product[172] that also cleaves RNA (Fig. 16A).[173,174] Building on the foundational studies of Hecht,[173,174] it was recently determined that bleomycin A5 has two preferred RNA cleavage sites, AU rich regions, with longer stretches of AU base pairs correlating with more efficient RNA cleavage and purine-rich sequences.[175] Angelbello *et al.*[175] identified a compilation of 13 human miRNAs that contain AU-rich regions, seven of which have been tied to disease. Of these, pri-miR-10b has a 5′AUAUAU/3′UAUAUA

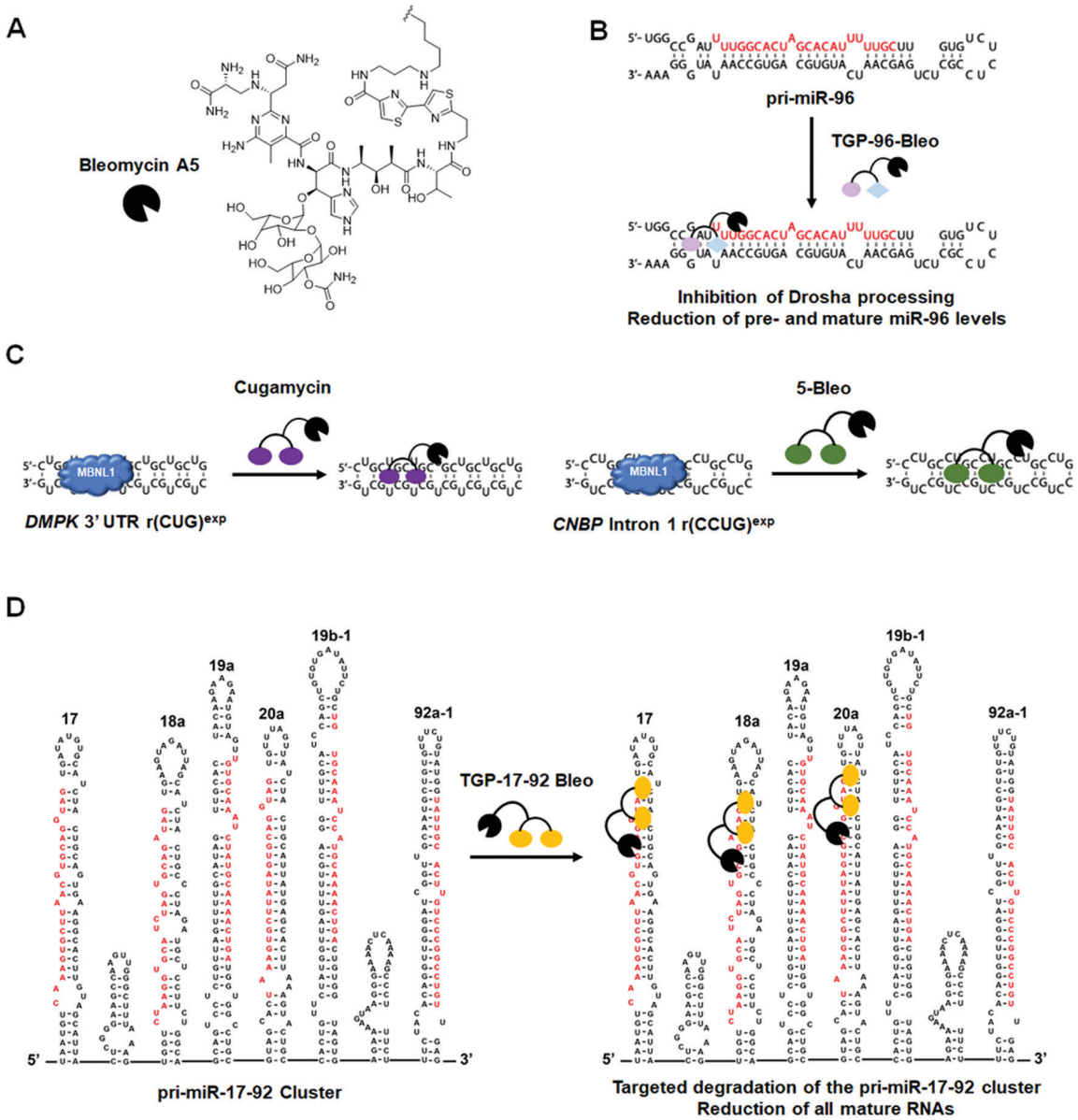


Fig. 16 Small molecule-bleomycin A5 conjugates cleave disease-causing RNAs in a targeted manner. (A) Structure of Bleomycin A5. (B) **TGP-96-Bleo** targets and cleaves oncogenic pri-miR-96. (C) Targeted degradation of r(CUG)^{exp} by **Cugamycin** and r(CCUG)^{exp} by **5-Bleo**. (D) Targeted degradation of the pri-miR-17-92 cluster by **TGP-17-92 Bleo**.

sequence, creating a potential recognition site for bleomycin A5 cleavage. Indeed, bleomycin A5 cleaved pri-miR-10b at two locations, the predicted AU-rich region and a 5′GUG/3′CAC site near the Dicer processing site. This finding was not surprising as bleomycin A5 also prefers purine-rich sequences.[175]

Bleomycin A5 was then studied for cleavage of pri-miR-10b in two cellular models: (i) HeLa cells transfected with a plasmid encoding pri-miR-10b and (ii) the TNBC cell line MDA-MB-231 in which miR-10b is overexpressed.[176] Aberrant levels of miR-10b have been linked to both tumor invasion and metastasis in TNBC.[176] At nM concentrations, the compound cleaved pri-miR-10b in both cell types thereby reducing levels of mature miR-10b, as determined by RT-qPCR.

This study highlighted the ability of bleomycin A5 to cleave RNA in cells, opening the door for the development of small molecule-bleomycin conjugates that direct the natural product to cleave a specific RNA target. This work also emphasized that ncRNAs can be targets of known drugs and should therefore be considered in drug discovery screens.

10.2 Targaprimir-96-Bleo (TGP-96-Bleo): Targeted cleavage of pri-miR-96 by a small molecule–bleomycin A5 conjugate

The first example of using small molecule–bleomycin conjugates to cleave miRNAs came from Li *et al.*,[177] in which they used a heterodimeric–bleomycin A5 conjugate to target oncogenic pri-miR-96. Both the bleomycin derivative and its site of conjugation to the small molecule were carefully selected. Bleomycin contains four domains: (i) a metal-binding nucleic acid cleavage domain; (ii) a C-terminal DNA-binding domain; (iii) a linker connecting the cleavage and DNA-binding domains; and (iv) a carbohydrate domain important for cellular uptake of the molecule.[172] The derivative bleomycin A5 was chosen for conjugation to RNA-binding small molecules because the DNA-binding domain contains a butyl-1,4-diamine side chain that allows for easy conjugation of small molecules. Additionally, it has been shown that conjugation through bleomycin's C-terminal free amine reduces affinity for DNA *via* ablation of the amine's positive charge.[178,179] These studies suggest that small molecule–bleomycin A5 conjugates have the potential to selectively cleave RNA targets.

As discussed previously (in Section 2.3. ''Design of Monomeric Small Molecules Targeting Disease-Causing miRNAs''), **TGP-96**[47] was designed using Inforna and is a potent binder of the pri-miR-96 Drosha processing site and adjacent 1 × 1 GG loop. Binding of **TGP-96** to pri-miR-96 inhibited the biogenesis of mature miR-96, derepressed FOXO1, and triggered apoptosis in MDA-MB-231 cells.[47] To further improve bioactivity, a small molecule cleaver was synthesized by conjugating **TGP-96** to bleomycin A5 (**TGP-96-Bleo**) *via* the C-terminal amine in the traditional DNA-binding domain of bleomycin A5, thus disrupting key interactions necessary for DNA recognition (Table S1, ESI†).[177] As bleomycin A5 has been shown to cleave AU base pairs,[175] and pri-miR-96 has AU base pairs in close proximity to **TGP-96**'s binding site, this conjugation strategy had a high potential for success.

Indeed, **TGP-96-Bleo** bound pri-miR-96 with a K_d of 64 ± 11 nM and cleaved the hairpin at the predicted AU base pairs, while no binding to DNA was observed (Fig. 16B and Table S1, ESI†).[177] A control compound lacking the RNA-binding modules cleaved plasmid DNA at levels 5-fold greater than those seen with **TGP-96-Bleo**,[177] indicating conjugation of bleomycin A5 to RNA binding modules reduces its affinity for DNA, lowering the potential for off-targets. This was further supported by visualizing γ-H2AX foci formation in cells, a marker for DNA double stranded breaks. Cells treated with the control compound lacking RNA-binding modules displayed ∼2.3-fold more foci than cells treated with **TGP-96-Bleo**.[177] Notably, the concentrations of **TGP-96-Bleo** that cleaved DNA and induced double stranded DNA breaks are 20-fold higher than the concentrations that reduced mature miR-96 levels, *vide infra*.

Based on these promising results, **TGP-96-Bleo** was compared to **TGP-96** in MDA-MB-231 TNBC cells for reducing levels of mature miR-96. RT-qPCR analysis confirmed treatment with **TGP-96-Bleo** decreased the levels of both pri-miR-96 and mature miR-96.[177] As mentioned above, **TGP-96** decreased mature miR-96 levels but increased pri-miR-96 levels. These data are consistent with the mechanisms of action for the two compounds; **TGP-96-Bleo** as an RNA cleaver and **TGP-96** as an RNA binder. Target occupancy of **TGP-96-Bleo** for the predicted pri-miR-96 binding sites was confirmed *via* a competitive cleavage assay in which cells were co-treated with **TGP-96-Bleo** and **TGP-96**, with **TGP-96** added in increasing concentrations to compete off a constant concentration of **TGP-96-Bleo**. Treatment with **TGP-96-Bleo** also resulted in rescue of FOXO1 expression and subsequent activation of apoptotic pathways in MDA-MB-231 cells, demonstrating rescue of disease phenotypes by **TGP-96-Bleo**.[177] **TGP-96-Bleo** did not have an effect on any other miRNAs predicted to target FOXO1.[180]

To further profile the selectivity of **TGP-96-Bleo**, small molecule nucleic acid profiling by cleavage applied to RNA (RiboSNAP) was utilized against the 349 miRNAs expressed in MDA-MB-231 cells.[177] The results of this profiling showed miR-96 levels were the most drastically and significantly affected by **TGP-96-Bleo** treatment, highlighting the selectivity of this small molecule RNA cleaver (Table S1, ESI†). Additionally, this experiment showed DNA off-targets of bleomycin A5 can be ablated by conjugation to an RNA-binder and that small molecule cleaver compounds can be successfully used for cellular profiling. A variation of RiboSNAP, RiboSNAP-Map, in which cleavage fragments are analyzed to determine the exact binding site of a small molecule, was also debuted in this paper.[177] RiboSNAP-Map uses a gene specific forward primer and universal reverse primer to amplify cleavage products, which are then sequenced to determine the site of cleavage.[177] This method confirmed the **TGP-96-Bleo** cleavage site is in close proximity to the known binding sites of **TGP-96**, positioning bleomycin A5 to cleave the proximal AU base pairs of pri-miR-96.

The data presented in this paper demonstrate the utility of conjugating bleomycin A5 to selective RNA-binding small molecules for the targeted degradation of disease-relevant RNAs and introduces novel methods for cellular profiling and target engagement validation using these compounds.

10.3 Cugamycin: Targeted cleavage of r(CUG)^{exp} by a small molecule–bleomycin A5 conjugate

Small molecule–bleomycin A5 conjugates can be used to target structured RNAs other than miRNAs. For example, another class of important RNA targets is the hairpin structures characteristic to microsatellite/repeat expansion disorders. The dimeric compound **2H-K4NMeS**, described above, that reverses DM1-associated defects was appended with bleomycin A5 to yield the small molecule cleaver, **Cugamycin** (Fig. 16C and Table S1, ESI†).[181] *In vitro* binding studies demonstrated **Cugamycin**'s selectivity for r(CUG)^{exp} (EC_{50} = 365 nM) over DNA, and cleavage studies confirmed bleomycin A5′s reduction in affinity for DNA when conjugated to an RNA-binding small molecule, as DNA cleavage was reduced by 4-fold as compared to bleomycin A5 (Table S1, ESI†).[181]

In DM1 patient-derived myotubes, **Cugamycin** localized to the nucleus and cleaved ∼70% of r(CUG)^{exp}-containing *DMPK* transcripts (dosed at 1 μM), while having no effect on wild-type *DMPK* transcripts that contain only a few r(CUG) repeats (non-pathogenic).[181] Further, **Cugamycin** demonstrated allele selectivity; that is, it only cleaved the mutant *DMPK* allele (1300 repeats). In conjunction with this cleavage, **Cugamycin** rescued MBNL1-dependent alternative splicing by ∼40%, leading to a ∼30% reduction in MBNL1-r(CUG)^{exp} nuclear foci when treated at 1 μM. **Cugamycin** did not have an effect on NOVA-mediated splicing, indicating selectivity for the r(CUG)^{exp} target.[181] Additionally, selectivity was profiled by measuring cleavage of five additional mRNAs that contain short r(CUG) repeats (<20 repeats). All five transcripts were unaffected by **Cugamycin** treatment.[181] Of note, an antisense LNA gap-mer complementary to r(CUG)^{exp} reduced the levels of all five r(CUG) repeat-containing mRNAs, as well as wild-type *DMPK* levels.[181] Modeling of RNA structures present in these mRNAs, as well as r(CUG)^{exp}, indicated the hairpin structure adopted by r(CUG)^{exp} is not recapitulated by shorter repeat lengths, bolstering the notion that structure-binding small molecules can indeed be selective in patient-derived cells and that selectivity translates *in vivo* (discussed below).

Off-target DNA cleavage in DM1 myotubes was assessed by visualizing γ-H2AX foci after treatment with **Cugamycin**, a control compound lacking the RNA-binding modules, or bleomycin A5, all tested at concentrations at which **Cugamycin** cleaved r(CUG)^{exp} and improved DM1-associated defects. Both the control compound and bleomycin A5 caused formation of γ-H2AX foci, while **Cugamycin** had no effect.[181] This again demonstrates that conjugation of bleomycin A5 through its C-terminal amine ablated affinity for DNA.

Cugamycin was also tested *in vivo* using the *HSA*^{LR} mouse model of DM1.[181] This model contains 250 CTG repeats driven by the human skeletal actin (*HSA*) promoter and recapitulates DM1 disease phenotypes such as dysregulation of MBNL1-dependent splicing, loss of the muscle-specific chloride ion channel (CLCN1), and myotonia.[182,183] **Cugamycin** was *i.p.* injected every other day, at a dose of 10 mg kg^{-1}, for a total of 8 days. After treatment, an ∼40% reduction in the levels of the *HSA* transgene [r(CUG)$_{250}$] was observed in tibialis anterior (TA) and gastrocnemius muscles, indicating that **Cugamycin** engaged its RNA target *in vivo*.[181] Lung fibrosis, a common side effect of bleomycin,[184] was not observed with **Cugamycin** treatment.

Rescue of aberrant alternative splicing in the TA and gastrocnemius muscles were also studied upon treatment with **Cugamycin**, showing that MBNL1-dependent splicing events *Mbnl1* exon 7 and *Clcn1* exon 7A were rescued by ∼50%, while alternative splicing of integrin β-1 precursor (*Itgb1*) exon 17 and capping actin protein of muscle *z*-line subunit β (*Capzb*) exon 8, non MBNL1-dependent events, were unaffected.[181] Loss of the CLCN1 protein, due to aberrant alternative splicing and exon 7A inclusion contributes directly to myotonia.[183] Therefore, recuse of MBNL1-dependent splicing should increase CLCN1 protein expression on the cell surface, leading to a rescue of disease phenotype. Indeed, upon **Cugamycin** treatment, the levels of CLCN1 in TA muscle plasma membranes increased and an ∼40% reduction in myotonia was observed.[181] These results were consistent across TA, gastrocnemius, and quadriceps muscles, indicating **Cugamycin** reaches DM1-affected tissues and rescues disease-associated phenotypes broadly.

Interestingly, **Cugamycin**'s parent compound, **2H-K4NMeS**, when delivered at the same dose and route of administration was unable to rescue MBNL1-dependent splicing or myotonia, indicating the cleavage capacity of **Cugamycin** is essential for *in vivo* efficacy.[181] These data also suggest that a cleavage-driven mechanism of action can provide a more prolonged and potent effect *in vivo* than a simple binding mode of action.

The ability of **Cugamycin** to rescue MBNL1-associated alternative splicing defects broadly and selectively was assessed by transcriptome-wide analysis of splicing events. By comparison to wild-type mice, the extent of the dysregulation of each splicing event in *HSA*^{LR} was measured. Angelbello *et al.*[181] identified 138 exons that are deregulated, reported as percent spliced in (Ψ), using mixture of isoforms (MISO)[185] analysis. Of these 138 exons, 134 of them showed Ψ values shifted back towards wild-type upon treatment with **Cugamycin**.[181] These data indicate that through **Cugamycin**'s selective recognition of r(CUG)^{exp}, the compound was able to globally rescue aberrant MBNL1-dependent alternative splicing in a mouse model of DM1. In addition to changes in alternative splicing, transcriptomic changes are also observed in DM1 mice. In particular, 326 genes are significantly dysregulated in *HSA*^{LR} mice.[181] Treatment with **Cugamycin** resulted in rescue of expression of 177 of these genes, highlighting the ability of the compound to normalize the transcriptome.[181] **Cugamycin** had no effect on the transcriptome of wild-type mice, as measured by RNA-seq, again highlighting the selectivity of this small molecule cleaver compound.

This study validated the strategy of using a small molecule-bleomycin A5 conjugate, **Cugamycin**, to target and cleave RNA repeats selectively in microsatellite/repeat expansion disorders, including in preclinical mouse models. **Cugamycin** showed remarkable potency *in vitro* and rescued DM1-associated phenotypes both in cells and *in vivo*. Additionally, the compound showed high selectivity for the *DMPK* mutant allele harboring r(CUG)^{exp} and demonstrated the ability to rescue MBNL1-dependent alternative

splicing transcriptome-wide with no significant off-targets. The data presented in this study indicate the cleavage-mediated mechanism of action of **Cugamycin** is highly effective at rescuing DM1-associated disease phenotypes in a mouse model, suggesting **Cugamycin** is a strong lead candidate for further optimization into a preclinical compound.

10.4 Targeted cleavage of r(CCUG)^{exp} by a small molecule-bleomycin A5 conjugate

After the success observed with **Cugamycin**, a small molecule-bleomycin A5 conjugate was created to target the r(CCUG)^{exp} in intron 1 of *CNBP*, causative of DM2.[139] Building off a previously designer dimer (**5**)[139] (discussed in Sections 6.2 and 7), Benhamou *et al.*[139] conjugated bleomycin A5 through the natural product's C-terminal amine to afford **5-Bleo** (Table S1, ESI†). *In vitro* studies showed **5** and **5-Bleo** bind r(CCUG)$_{10}$ with similar affinities (~100 nM) and that **5-Bleo** cleaved RNA between 5′-GC steps in base paired regions adjacent to the compound's binding site (Fig. 16C and Table S1, ESI†).[139] This compound also demonstrated selectivity for the RNA target over DNA, consistent with the results from previous bleomycin A5 conjugates.[139,181]

While the dimer and cleaver compounds display similar binding affinities for r(CCUG)^{exp} (~100 nM) *in vitro*, in DM2 fibroblasts the cleaver reduced levels of intron 1-containing r(CCUG)^{exp} transcripts by an ~2.5-fold greater extent than the dimer, the binding of which shunts the intron down endogenous decay pathways (*i.e.*, has a different mode of action).[139] Mature *CNBP* mRNA levels were also reduced upon treatment of **5-Bleo**. Thus, the bleomycin A5 conjugate was able to more effectively reduce levels of mutant *CNBP* mRNA, compared to the binder, and cleaves the r(CCUG)^{exp} target. Off-target DNA cleavage was again assessed by visualizing the formation of γ-H2AX foci in DM2 fibroblasts. Treatment with **5-Bleo** did not result in a significant increase in foci at the active concentration, demonstrating the compound's selectivity *in cellulis*.[139]

Target engagement of **5-Bleo** was confirmed in cells using a competitive cleavage assay in which increasing concentrations of the dimer were co-treated with a constant concentration of **5-Bleo** and the levels of mature *CNBP* mRNA were measured. Increasing the concentration of the simple binding dimer led to an increase in *CNBP* mRNA levels (a reduction in cleavage), indicating **5-Bleo** and the dimer share the same RNA target and that the mechanism of **5-Bleo** is through direct cleavage of r(CCUG)^{exp}.[139]

Further cellular studies demonstrated **5-Bleo**'s enhanced ability to rescue MBNL1-dependent *IR* pre-mRNA splicing defects compared to the dimer. **5-Bleo** rescued splicing by ~50% at 5 μM, while the dimer only rescued splicing defects by ~20% at 10 μM, a 2-fold higher concentration.[139] Additionally, an ~50% reduction in r(CCUG)^{exp}-containing foci was observed upon treatment of **5-Bleo**. Evaluation of mature *CNBP* mRNA levels and *IR* splicing in healthy fibroblasts after treatment with **5-Bleo** showed no changes, confirming **5-Bleo**'s allele selectivity for the disease-causing r(CCUG)^{exp} as the shorter r(CCUG) repeats present in healthy fibroblasts were unaffected.[139]

11. Targeted degradation of disease-causing RNAs using RIBOTACs

The advent of proteolysis targeting chimeras (PROTACs)[186] demonstrated the ability to trigger protein degradation with small molecules. This concept has been broadened to other biomolecules, such as RNAs, as is the case of RIBOTACs.[187] RIBOTACs mediate RNA decay by recruiting endogenous RNases to degrade specific transcripts. In particular, RIBOTACs have been developed that recruit RNase L, a component of the antiviral immune response. RNase L is present in minute quantities in all cells as an inactive monomer. Upon viral infection, it is upregulated, dimerized and activated by 2′–5′ oligoadenylate [2′–5′ poly(A)].[188] RNase L is thus an intriguing enzyme for small molecule recruitment and targeted RNA destruction. That is, an RNA-binding small molecule coupled to an RNase L-recruiting module could locally recruit and activate RNase L to cleave the target selectively, without activation of the immune system.

11.1 Targaprimir-96 RIBOTAC (TGP-96 RIBOTAC): Targeting pri-miR-96 for degradation

As previously described in Section 2.3 (''Design of Monomeric Small Molecules Targeting Disease-Causing miRNAs''), **TGP-96** is a dimeric small molecule that binds pri-miR-96 and inhibits its biogenesis, thereby derepressing the transcription factor FOXO1 and triggering apoptosis in TNBC cells.[47] **TGP-96** was converted into a RIBOTAC (**TGP-96 RIBOTAC**) by appending a short 2′–5′ A_4 oligonucleotide as the RNase L recruiting module onto the compound (Fig. 17 and Table S1, ESI†).[187] *In vitro* binding assays confirmed the recruiting module does not affect the avidity of the compound for pri-miR-96's Drosha processing site (K_d = 20 nM) (Table S1, ESI†).[187] *In vitro* cleavage assays demonstrated the ability of **TGP-96 RIBOTAC** to recruit and dimerize RNase L, leading to the selective cleavage of pri-miR-96.[187] This cleavage was inhibited when **TGP-96** was added as a competitor, validating the binding sites of **TGP-96** and **TGP-96 RIBOTAC** are the same.

In MDA-MB-231 cells, despite having ~2-fold reduced cellular permeability compared to **TGP-96**, **TGP-96 RIBOTAC** reduced the levels of pri-miR-96 and mature miR-96, confirming compound mode of action.[187] RNase L-dependent cleavage was verified in multiple experiments: (i) immunoprecipitation with an RNase L antibody confirmed ternary complex formation between pri-miR-96, RNase L, and **TGP-96 RIBOTAC**; (ii) competitive cleavage between **TGP-96 RIBOTAC** and a derivative lacking the RNase L-recruiting module showed a dose-dependent decrease in cleavage of pri-miR-96, validating the ability of **TGP-96 RIBOTAC** to locally dimerize RNase L; and (iii) siRNA knockdown of RNase L ablated **TGP-96 RIBOTAC**'s ability to degrade pri-miR-96.[187]

Treatment of **TGP-96 RIBOTAC** in MDA-MB-231 cells resulted in modulation of the invasive phenotype associated with miR-96 in cancer. An ~2-fold increase in FOXO1 expression was observed upon treatment, consistent with inhibition of miR-96 biogenesis, and thus resulted in apoptosis.[187] This effect can be reversed upon over expression of pri-miR-96.

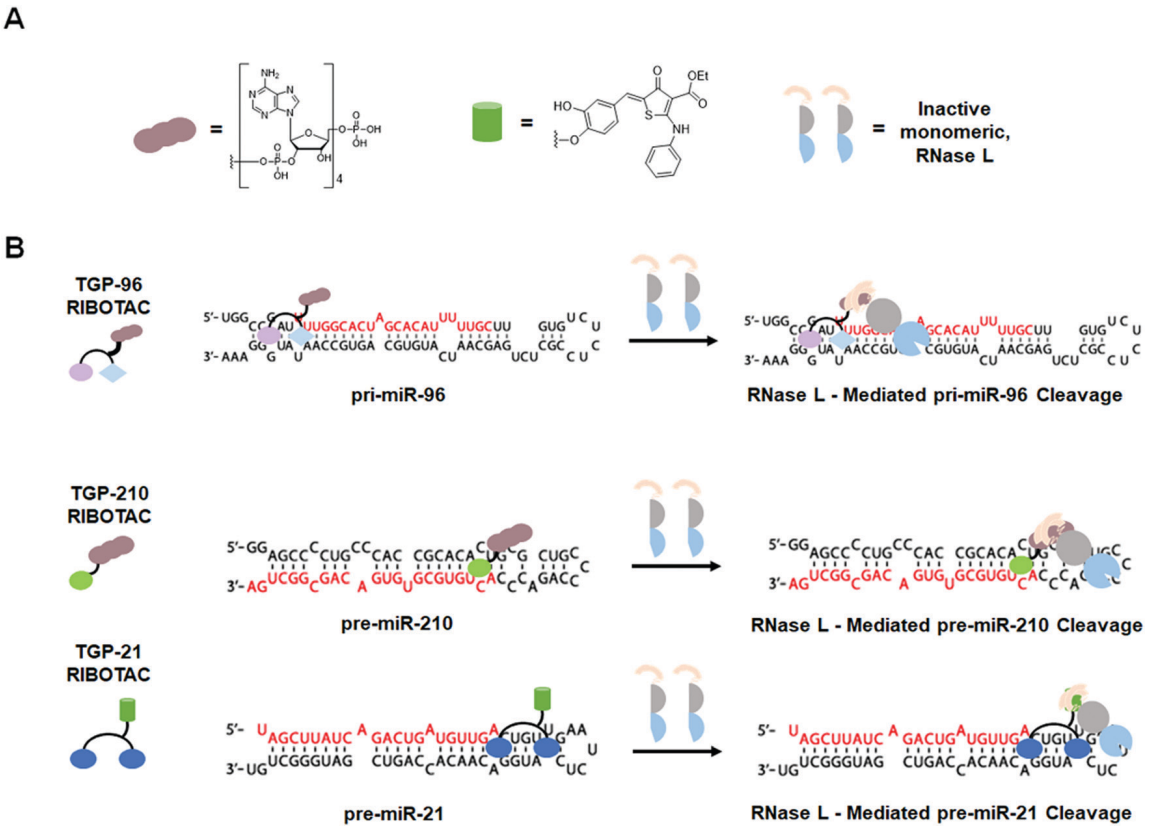


Fig. 17 Ribonuclease targeting chimeras (RIBOTACs) degrade disease-causing RNAs in a targeted manner. (A) Structures of first- and second-generation RNase L recruiting modules and schematized monomeric RNase L. (B) RIBOTAC-mediated degradation of oncogenic pri-miR-96, pre-miR-210, and pre-miR-21.

TGP-96 RIBOTAC has no effect on apoptosis in MCF-10a cells (healthy breast epithelial cells).[187] Further, **TGP-96 RIBOTAC** stimulated apoptosis to the same extent as its parent binding compound at a 2.5-fold lower dose. Combined with the decreased uptake of the compound, RNase L recruitment by **TGP-96 RIBOTAC** enhances the compound's activity by ~5-fold compared to **TGP-96.**[187] Most importantly, **TGP-96 RIBOTAC** acts catalytically and in a substoichiometric fashion to recruit RNase L for targeted RNA degradation, cleaving 3.1 pri-miR-96 molecules per molecule of RIBOTAC.[187]

11.2 Targapremir-210 RIBOTAC (TGP-210 RIBOTAC): Targeting pre-miR-210 for degradation

As discussed in Section 2.3 ("Design of Monomeric Small Molecules Targeting Disease-Causing miRNAs"), **TGP-210**, designed by Inforna, binds the Dicer processing site of pre-miR-210, inhibits its biogenesis and normalizes proteins in this circuit, ultimately inducing apoptosis in hypoxic cancer cells.[45] Costales *et al.*[189] optimized **TGP-210** by appending a 2′–5′ A_4 RNase L recruiting module to yield **TGP-210 RIBOTAC** (Fig. 17 and Table S1, ESI†). *In vitro* cleavage assays showed **TGP-210 RIBOTAC** cleaved pre-miR-210 and binding assays demonstrated **TGP-210 RIBOTAC** is more selective than **TGP-210**; an ~10-fold difference in affinity is observed between the Dicer processing site of pre-miR-210 and DNA while only an ~5-fold difference was observed for **TGP-210** (Table S1, ESI†).[189] Thus RNA degraders can show enhanced selectivity over their simple binding counterparts.

In hypoxic MDA-MB-231 cells, **TGP-210 RIBOTAC** decreased the levels of both pre-miR-210 and mature miR-210, consistent with its mode of action.[189] Upon treatment with **TGP-210 RIBOTAC**, *GPD1L* mRNA levels were significantly increased and *HIF1α* mRNA levels were decreased.[189] As a result of deactivation of the oncogenic circuit, **TGP-210 RIBOTAC** triggered apoptosis in hypoxic cancer cells, to a similar extent as **TGP-210.** However, because **TGP-210 RIBOTAC** is half as cell permeable as **TGP-210**, these results demonstrate the RIBOTAC has ~2-fold enhanced activity over **TGP-210.**[189]

Specific RNase L recruitment was confirmed *via*: (i) competitive cleavage assays, in which co-treatment of increasing amounts of **TGP-210** with **TGP-210 RIBOTAC** resulted in a dose-dependent decrease in pre-miR-210 cleavage; (ii) overexpression of RNase L, resulting in increased cleavage; (iii) overexpression of pre-miR-210, resulting in decreased cleavage; and (iv) siRNA ablation of RNase L, resulting in no **TGP-210 RIBOTAC**-mediated cleavage of pre-miR-210.[189] Additionally, immunoprecipitation of RNase L showed enrichment for pre-miR-210 only in cells treated with **TGP-210 RIBOTAC.**

RNA-seq and RT-qPCR profiling experiments revealed **TGP-210 RIBOTAC** has no significant off-targets trancriptome-wide (Table S1, ESI†).[189] Combined with its catalytic and substoichiometric mode of action, **TGP-210 RIBOTAC** demonstrates

that enhanced activity and selectivity can be achieved through the targeted recruitment of nucleases *via* RIBOTAC compounds.

11.3 Targapremir-21 RIBOTAC (TGP-21 RIBOTAC): Targeting pre-miR-21 for degradation

A RIBOTAC targeting pre-miR-21, dubbed **TGP-21 RIBOTAC**, was recently reported that is based on the dimeric binding compound **TGP-21** (Table S1, ESI†).[190] **TGP-21** was first validated in MDA-MB-231 TNBC cells, reducing levels of mature miR-21 and increasing levels of pre-miR-21, in accordance with its mechanism of inhibiting Dicer processing (Table S1, ESI†).[190] The binding dimer also increased expression of programmed cell death protein 4 (PDCD4) and phosphatase and tension homolog (PTEN), two proteins that are translationally repressed by miR-21.[190] Additionally, invasion assays confirmed **TGP-21**'s ability to inhibit the invasive phenotype of MDA-MB-231 cells.[190]

TGP-21 was optimized by appending a heterocyclic small molecule recruiter of RNase L to create **TGP-21 RIBOTAC** (Fig. 17 and Table S1, ESI†). Previous studies by Thakur *et al.*[191] demonstrated the ability of a heterocyclic small molecule to recruit and activate RNase L in place of the traditional 2′–5′ poly(A) substrate. Extensive optimization of this structure by Costales *et al.*[190] yielded the small molecule RNase L-recruiting heterocycle that was incorporated in **TGP-21 RIBOTAC**.

In MDA-MB-231 cells, **TGP-21 RIBOTAC** showed a 20-fold enhancement in activity for reducing mature miR-21 levels compared to **TGP-21** and reduced pre-miR-21 levels in a sub-stoichiometric manner (Table S1, ESI†).[190] RNase L recruitment was confirmed using the same assays as described for **TGP-210 RIBOTAC**, further supporting the mechanism of small molecule-targeted degradation.[190] A time course experiment in which reduction of mature miR-21 levels were monitored up to 96 h post-treatment revealed **TGP-21 RIBOTAC** has a more potent and prolonged effect than **TGP-21**.[190]

Importantly, **TGP-21 RIBOTAC** did not trigger an innate immune response, as monitored by mRNA and protein levels of innate immunity-associated biomarkers.[190] In contrast, transfection of 2′–5′ A_4 into MDA-MB-231 cells resulted in upregulation of several innate immunity biomarkers, such as interferon gamma (IFN-γ), 2′–5′-oligoadenylate synthase 1 (OAS1), retinoic acid-inducible gene 1 (RIG-1), and melanoma differentiation-associated protein 5 (MDA5).[190] These results support the hypothesis that **TGP-21 RIBOTAC** locally recruits and activates RNase L, instead of triggering a global antiviral innate immune response.[190]

Selectivity was also assessed miRNome-wide and quantified by calculating Gini coefficients. Gini coefficients were first introduced as a metric of biological selectivity by *Graczyk*[192] as a measure of kinase inhibitor selectivity, however the metric can be broadly applied to any biomolecule and small molecule modulator. A Gini coefficient of 0 represents a nonselective compound while a Gini coefficient of 1 represents an exquisitely selective compound. Gini coefficients for the monomer that comprises **TGP-21**, **TGP-21**, and **TGP-21 RIBOTAC**, by measuring their effects on the miRNome, are 0.52, 0.68, and 0.84 respectively, highlighting the increase in selectivity that is achieved by dimerization and by RNase L-mediated targeted degradation of RNA.[190] Proteome-wide analysis of the effects of **TGP-21 RIBOTAC** in MDA-MB-231 cells confirmed the selectivity of the compound with only 47 of 4181 proteins being significantly affected.[190]

TGP-21 RIBOTAC was also evaluated *in vivo* in a mouse model of metastatic breast cancer (NOD/SCID mice *i.v.* injected with MDA-MB-231-Luc cells). Treatment with **TGP-21 RIBOTAC** (10 mg kg^{-1}, every other day for 6 weeks) inhibited metastasis to lung, as evident by a significant decrease in the number of lung nodules present in **TGP-21 RIBOTAC**-treated mice.[190] Additionally, tissues from mice in the RIBOTAC treatment group displayed decreased hematoxylin and eosin (H&E) staining, decreased pre-miR-21 and mature miR-21 levels (as assayed by FISH and RT-qPCR), and increased levels of PDCD4 (as assayed by immunohistochemistry).[190] Thus, **TGP-21 RIBOTAC** selectivity and potently modulates the miR-21 pathway in a preclinical mouse model, resulting in inhibition of breast cancer metastasis.

12. Case study: direct comparison of bleomycin A5-mediated cleavage *versus* RNase L-mediated degradation of the pri-miR-17-92 cluster

The pri-miR-17-92 cluster is a direct target of the transcription factor c-MYC[193] and is upregulated in human diseases ranging from cancers[194–196] to fibrosis.[197] For each disease, the mature miRNA deregulated from the cluster and its downstream effects can be unique or overlap,[198] and the mature miRNAs can act either individually or synergistically to affect disease biology.[193,199] Through the use of Inforna and subsequent optimization steps, Liu *et al.*[200] developed a dimeric small molecule that binds to three miRNAs in the pri-miR-17-92 cluster (miR-17, miR-18a, and miR-20a) that share structural commonalities at and adjacent to their Dicer processing sites. They then appended the small molecule with either bleomycin A5, yielding **Targaprimir-17-92 Bleo** (**TGP-17-92 Bleo**) or the heterocyclic RNase L-recruiting module, yielding **TGP-17-92 RIBOTAC** (Fig. 16D and Table S1, ESI†).[200] The ability of these two compounds to reduce levels of pri-miR-17-92 was evaluated in MDA-MB-231 TNBC cells and DU-145 prostate cancer cells.[200] **TGP-17-92 Bleo** reduced pri-miR-17-92 levels and hence functionally inhibited all six miRNAs it encodes (Table S1, ESI†). Further, pre-miR17, pre-miR-18a, and pre-miR-20a levels were reduced. Cleavage of the cluster and the three pre-miRNAs that **TGP-17-92 Bleo** binds is consistent with the compound's cellular localization. In contrast, **TGP-17-92 RIBOTAC** only reduced levels of pre-miR17, pre-miR-18a, and pre-miR-20a and their mature miRNAs while not affecting the primary transcript (Table S1, ESI†).[200] That is, the RIBOTAC only cleaved the pre-miRNAs that bind **TGP-17-92** because of the co-localization of the RIBOTAC, RNase L, and the pre-miRNA, the three components required for cleavage, in the cytoplasm.[200] Thus, these

studies devised a facile design strategy to remove an entire pri-miRNA cluster, of importance since ~25% of all miRNAs are transcribed in clusters, or individual members of the cluster simply by careful selection of the cleavage module and by exploiting differences in cellular localization.

13. Conclusion

As this review showcases, various types of RNAs have been successfully targeted with small molecules including miRNAs, lncRNAs, splicing modifiers, repeat expansion disorders, and structured elements found within disease-causing RNAs. Since only 1–2% of the genome is translated into protein but ~80% is transcribed into RNA, it is not surprising that RNA is rapidly emerging as a promising target of small molecule therapeutics.[3,18,201] As illustrated by the examples discussed herein, RNA-targeting small molecules can display impressive potency (nM) and selectivity for their RNA targets, often rivaling those seen with traditional protein-targeting small molecules.

Current chemotherapeutic options come with a myriad of side-effects due to off-target effects of the drug. However, compounds such as **TGP-96 RIBOTAC**, **TGP-210 RIBOTAC**, **TGP-21 RIBOTAC** and **TGP-17-92 Bleo** demonstrate proteome- and transcriptome-wide selectivity while decreasing levels of oncogenic miRs. Thus, these compounds could offer a starting place for the development of novel anti-cancer treatments with fewer side effects.

In addition, this review highlights several RNA-targeting small molecules that affect neurodegenerative disease biology, further highlighting the potential of these small molecules to act as novel therapeutics. More importantly, RNA-targeting small molecules have unlocked therapeutic avenues against proteins of neurological relevance that were traditionally viewed as "undruggable," such as the case of **Synucleozid** for α-synuclein and **A-5** for tau. Examples such as these highlight the importance of targeting RNA over traditional protein targets as a means of investigating the vast space of disease biology that has remained elusive to drug development. Additionally, RNA cleavers and degraders, such as bleomycin A5-conjugates and RIBOTACs, offer novel therapeutic modalities to reduce levels of disease-causing RNAs, whether oncogenic or neurodegenerative, further expanding the types of diseases that can be targeted with small molecules.

In comparison to antisense oligonucleotide (ASO) technology, the current leading therapeutic option for targeting RNA, small molecules are an attractive alternative due to their synthetic accessibility, on- and off-target optimization *via* structure–activity relationships, mode of action, and ease of administration. ASOs recognize RNA sequence, while small molecules recognize RNA structure; this essential difference gives small molecules many advantages over ASOs. For example, ASOs targeting repeat expansion disorders are often designed to target the coding region of the gene harboring the expansion, not the expansion itself due to off-target effects on transcripts that contain short repeats.[43,139,181] Thus, ASOs have to be specifically designed for each disease, even if the diseases are caused by the same RNA repeat expansion. In contrast, a single small molecule recognizing a disease-causing structure could be a potential treatment for all diseases that harbor the same repeat expansion. For example, DM1, Fuchs endothelial corneal dystrophy (FECD),[202] and Huntington's disease-like 2 (HDL-2)[203] are all caused by r(CUG)^{exp}. Therefore, small molecules targeting r(CUG)^{exp} could be applied broadly across these diseases as shorter, non-pathogenic repeats found in other transcripts typically lack structure. Another advantage of small molecules over ASOs is the route of therapeutic administration. ASOs are injected intrathecally, a painful and sometimes dangerous procedure for patients to endure. However, RNA-targeting small molecules have shown efficacy in mouse models with intraperitoneal injection[181] and success with oral bioavailability in clinical trials.[96]

Looking at the various classification of RNA-targeting small molecules that have emerged it is interesting to note the differences in affinity and bioactivity that result from appending bleomycin A5 or an RNase L-recruiting module to simple RNA-binding small molecules. While the conjugates tend to decrease the compounds' binding affinities for the target RNA, the overall biological effect observed upon treatment with the cleaving/degrading compounds far exceeds that of the simple RNA binder (Table S1, ESI†). This is demonstrated in the case of **TGP-96** (K_d = 39 nM) and **TGP-96 Bleo** (K_d = 64 nM), where **TGP-96 Bleo** demonstrates a greater biological effect than **TGP-96**. This trend was also observed for **TGP-210** (K_d = 160 nM) *versus* **TGP-210 RIBOTAC** (K_d = 340 nM) and **2H-K4NMeS** (K_d = 12 nM) *versus* **Cugamycin**, the bleomycin conjugate of **2H-K4NMeS** (EC_{50} = 365 nM). In all of these cases, the small decrease in affinity that results from appending a cleaving or degrading module to the compound is made up for by the large increase in biological activity (up to 20-fold) due to target RNA ablation, rather than simple binding.

Selectivity is also improved when an RNA-binding small molecule is converted to either a RIBOTAC or bleomycin A5 conjugate. Again, looking at the case of **TGP-210** and **TGP-210 RIBOTAC**, converting the small molecule to a RIBOTAC resulted in an increase in the selectivity window over DNA from 5-fold with the RNA binder to 10-fold with the degrader, despite **TGP-210 RIBOTAC** having decreased affinity for pre-miR-210.[189] Additionally, appending bleomycin A5 to RNA-binding small molecules decreases both the affinity of bleomycin A5 and the RNA binder itself for DNA, further enhancing the compound's selectivity. Thus, RNA-targeting small molecules demonstrate a novel avenue for achieving highly selective RNA-centric therapeutics.

The field of RNA-targeting small molecules is still in its infancy, as such the only FDA-approved small molecules on the market are anti-bacterials (targeting bacterial RNA), noting that **risdiplam**, which targets an RNA–protein interface, recently received FDA approval for the treatment of SMA (approved August 7th, 2020).[97] Additionally, **branaplam** (Novartis)[204] is also undergoing a Phase II clinical trial for the treatment of SMA. Both **risdiplam** and **branaplam** are administered orally, further supporting the advantages of using small molecules to target RNA. Despite the lack of branded "RNA-targeting"

FDA-approved small molecules it is important to note that many protein-targeting FDA-approved drugs, such as kinase and topoisomerase inhibitors, have been found to also bind RNA as off-targets. Thus, the future is exciting for the field of RNA-targeted small molecule therapeutics and will undoubtedly contribute to the advancement of modern, precision medicines.

Conflicts of interest

Matthew D. Disney is a founder of Expansion Therapeutics.

References

1 X.-D. Fu, *Natl. Sci. Rev.*, 2014, **1**, 190–204.
2 C. M. Connelly, M. H. Moon and J. S. Schneekloth, *Cell Chem. Biol.*, 2016, **23**, 1077–1090.
3 T. A. Cooper, L. Wan and G. Dreyfuss, *Cell*, 2009, **136**, 777–793.
4 R. R. Breaker, *Cold Spring Harbor Perspect. Biol.*, 2018, **10**, a032797.
5 Z. A. Jaafar and J. S. Kieft, *Nat. Rev. Microbiol.*, 2019, **17**, 110–123.
6 A. Bugaut and S. Balasubramanian, *Nucleic Acids Res.*, 2012, **40**, 4727–4741.
7 L. A. Macfarlane and P. R. Murphy, *Curr. Genomics*, 2010, **11**, 537–561.
8 K. Leppek, R. Das and M. Barna, *Nat. Rev. Mol. Cell Biol.*, 2018, **19**, 158–174.
9 M. B. Warf and J. A. Berglund, *Trends Biochem. Sci.*, 2010, **35**, 169–178.
10 E. Buratti and F. E. Baralle, *Mol. Cell. Biol.*, 2004, **24**, 10505–10514.
11 A. J. Zaug and T. R. Cech, *Science*, 1986, **231**, 470–475.
12 B. C. Stark, R. Kole, E. J. Bowman and S. Altman, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 3717–3721.
13 E. A. Doherty and J. A. Doudna, *Annu. Rev. Biophys. Biomol. Struct.*, 2001, **30**, 457–475.
14 S. Müller, B. Appel, D. Balke, R. Hieronymus and C. Nübel, *F1000Research*, 2016, **5**, 1511.
15 Z. Huang, J. Shi, Y. Gao, C. Cui, S. Zhang, J. Li, Y. Zhou and Q. Cui, *Nucleic Acids Res.*, 2019, **47**, 1013–1017.
16 A. M. Ardekani and M. M. Naeini, *Avicenna J. Med. Biotechnol.*, 2010, **2**, 161–179.
17 V. Bernat and M. D. Disney, *Neuron*, 2015, **87**, 28–46.
18 M. D. Disney, *J. Am. Chem. Soc.*, 2019, **141**, 6776–6790.
19 X.-H. Liang, H. Sun, J. G. Nichols and S. T. Crooke, *Mol. Ther.*, 2017, **25**, 2075–2092.
20 A. Ursu, J. L. Childs-Disney, R. J. Andrews, C. A. O'Leary, S. M. Meyer, A. J. Angelbello, W. N. Moss and M. D. Disney, *Chem. Soc. Rev.*, 2020, DOI: 10.1039/D0CS00455C.
21 M. M. Chong, G. Zhang, S. Cheloufi, T. A. Neubert, G. J. Hannon and D. R. Littman, *Genes Dev.*, 2010, **24**, 1951–1960.
22 L. F. R. Gebert and I. J. MacRae, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 21–37.
23 T. A. Rand, S. Petersen, F. Du and X. Wang, *Cell*, 2005, **123**, 621–629.
24 H. A. Meijer, Y. W. Kong, W. T. Lu, A. Wilczynska, R. V. Spriggs, S. W. Robinson, J. D. Godfrey, A. E. Willis and M. Bushell, *Science*, 2013, **340**, 82–85.
25 T. Fukaya, H. O. Iwakawa and Y. Tomari, *Mol. Cell*, 2014, **56**, 67–78.
26 A. Fukao, Y. Mishima, N. Takizawa, S. Oka, H. Imataka, J. Pelletier, N. Sonenberg, C. Thoma and T. Fujiwara, *Mol. Cell*, 2014, **56**, 79–89.
27 M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin and N. Rajewsky, *Nature*, 2008, **455**, 58–63.
28 S. Uhlmann, H. Mannsperger, J. D. Zhang, E. A. Horvat, C. Schmidt, M. Kublbeck, F. Henjes, A. Ward, U. Tschulena, K. Zweig, U. Korf, S. Wiemann and O. Sahin, *Mol. Syst. Biol.*, 2012, **8**, 570.
29 M. Esteller, *Nat. Rev. Genet.*, 2011, **12**, 861–874.
30 S. Lin and R. I. Gregory, *Nat. Rev. Cancer*, 2015, **15**, 321–333.
31 C. P. Bracken, H. S. Scott and G. J. Goodall, *Nat. Rev. Genet.*, 2016, **17**, 719–732. 31.
32 D. D. Vo, C. Staedel, L. Zehnacker, R. Benhida, F. Darfeuille and M. Duca, *ACS Chem. Biol.*, 2014, **9**, 711–721.
33 V. Malnuit, M. Duca and R. Benhida, *Org. Biomol. Chem.*, 2011, **9**, 326–336.
34 M. Duca, V. Malnuit, F. Barbault and R. Benhida, *Chem. Commun.*, 2010, **46**, 6162–6164.
35 D. D. Vo, T. P. Tran, C. Staedel, R. Benhida, F. Darfeuille, A. Di Giorgio and M. Duca, *Chem. – Eur. J.*, 2016, **22**, 5350–5362.
36 D. D. Vo, C. Becquart, T. P. A. Tran, A. Di Giorgio, F. Darfeuille, C. Staedel and M. Duca, *Org. Biomol. Chem.*, 2018, **16**, 6262–6274.
37 S. Kumar and S. Maiti, *Biochimie*, 2013, **95**, 1422–1431.
38 J. P. Joly, G. Mata, P. Eldin, L. Briant, F. Fontaine-Vive, M. Duca and R. Benhida, *Chem. – Eur. J.*, 2014, **20**, 2071–2079.
39 C. Staedel, T. P. A. Tran, J. Giraud, F. Darfeuille, A. Di Giorgio, N. J. Tourasse, F. Salin, P. Uriac and M. Duca, *Sci. Rep.*, 2018, **8**, 1667.
40 C. Becquart, M. Le Roch, S. Azoulay, P. Uriac, A. Di Giorgio and M. Duca, *ACS Omega*, 2018, **3**, 16500–16508.
41 M. D. Disney, A. M. Winkelsas, S. P. Velagapudi, M. Southern, M. Fallahi and J. L. Childs-Disney, *ACS Chem. Biol.*, 2016, **11**, 1720–1728.
42 S. P. Velagapudi, Y. Luo, T. Tran, H. S. Haniff, Y. Nakai, M. Fallahi, G. J. Martinez, J. L. Childs-Disney and M. D. Disney, *ACS Cent. Sci.*, 2017, **3**, 205–216.
43 S. P. Velagapudi, S. M. Gallo and M. D. Disney, *Nat. Chem. Biol.*, 2014, **10**, 291–297.
44 I. K. Guttilla and B. A. White, *J. Biol. Chem.*, 2009, **284**, 23204–23216.
45 M. G. Costales, C. L. Haga, S. P. Velagapudi, J. L. Childs-Disney, D. G. Phinney and M. D. Disney, *J. Am. Chem. Soc.*, 2017, **139**, 3446–3455.
46 T. J. Kelly, A. L. Souza, C. B. Clish and P. Puigserver, *Mol. Cell. Biol.*, 2011, **31**, 2696–2706.
47 S. P. Velagapudi, M. D. Cameron, C. L. Haga, L. H. Rosenberg, M. Lafitte, D. R. Duckett, D. G. Phinney and

M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 5898–5903.
48 M. G. Costales, D. G. Hoch, D. Abegg, J. L. Childs-Disney, S. P. Velagapudi, A. Adibekian and M. D. Disney, *J. Am. Chem. Soc.*, 2019, **141**, 2960–2974.
49 H. Haniff, L. Knerr, X. Liu, G. Crynen, J. Boström, D. Abegg, A. Adibekian, M. Lemurell and M. Disney, *Nat. Chem.*, 2020, DOI: 10.1038/s41557-020-0514-4.
50 N. Ferrara, H. P. Gerber and J. LeCouter, *Nat. Med.*, 2003, **9**, 669–676.
51 N. Sun, B. Ning, K. M. Hansson, A. C. Bruce, S. A. Seaman, C. Zhang, M. Rikard, C. A. DeRosa, C. L. Fraser, M. Wagberg, R. Fritsche-Danielson, J. Wikstrom, K. R. Chien, A. Lundahl, M. Holtta, L. G. Carlsson, S. M. Peirce and S. Hu, *Sci. Rep.*, 2018, **8**, 17509.
52 S. Yla-Herttuala, T. T. Rissanen, I. Vajanto and J. Hartikainen, *J. Am. Coll. Cardiol.*, 2007, **49**, 1015–1026.
53 Z. Wen, W. Huang, Y. Feng, W. Cai, Y. Wang, X. Wang, J. Liang, M. Wani, J. Chen, P. Zhu, J. M. Chen, R. W. Millard, G. C. Fan and Y. Wang, *PLoS One*, 2014, **9**, 104666.
54 D. Joladarashi, V. N. S. Garikipati, R. A. Thandavarayan, S. K. Verma, A. R. Mackie, M. Khan, A. M. Gumpert, A. Bhimaraj, K. A. Youker, C. Uribe, S. Suresh Babu, P. Jeyabal, R. Kishore and P. Krishnamurthy, *J. Am. Coll. Cardiol.*, 2015, **66**, 2214–2226.
55 Z. Taimeh, J. Loughran, E. J. Birks and R. Bolli, *Nat. Rev. Cardiol.*, 2013, **10**, 519–530.
56 X. Guo, L. Gao, Y. Wang, D. K. Y. Chiu, T. Wang and Y. Deng, *Briefings Funct. Genomics*, 2015, **15**, 38–46.
57 F. Zhang, L. Zhang and C. Zhang, *Tumor Biol.*, 2016, **37**, 163–175.
58 C.-W. Wei, T. Luo, S.-S. Zou and A.-S. Wu, *Front. Behav. Neurosci.*, 2018, **12**, 348–358.
59 M. V. Neguembor, M. Jothi and D. Gabellini, *Skeletal Muscle*, 2014, **4**, 8.
60 S. Dahariya, I. Paddibhatla, S. Kumar, S. Raghuwanshi, A. Pallepati and R. K. Gutti, *Mol. Immunol.*, 2019, **112**, 82–92.
61 J. Chen, L. Ao and J. Yang, *Ann. Transl. Med.*, 2019, **7**, 494.
62 M. J. Delás and G. J. Hannon, *Open Biol.*, 2017, **7**, 170121.
63 R. Pedram Fatemi, S. Salah-Uddin, F. Modarresi, N. Khoury, C. Wahlestedt and M. A. Faghihi, *J. Biomol. Screening*, 2015, **20**, 1132–1141.
64 F. A. Abulwerdi, W. Xu, A. A. Ageeli, M. J. Yonkunas, G. Arun, H. Nam, J. S. Schneekloth, T. K. Dayie, D. Spector, N. Baird and S. F. J. Le Grice, *ACS Chem. Biol.*, 2019, **14**, 223–235.
65 N. F. Rizvi and E. B. Nickbarg, *Methods*, 2019, **167**, 28–38.
66 M. Hajjari and A. Salavaty, *Cancer Biol. Med.*, 2015, **12**, 1–9.
67 K. Zhang, X. Sun, X. Zhou, L. Han, L. Chen, Z. Shi, A. Zhang, M. Ye, Q. Wang, C. Liu, J. Wei, Y. Ren, J. Yang, J. Zhang, P. Pu, M. Li and C. Kang, *OncoTargets Ther.*, 2015, **6**, 537–546.
68 Y. Ren, Y.-F. Wang, J. Zhang, Q.-X. Wang, L. Han, M. Mei and C.-S. Kang, *Clin. Epigenet*, 2019, **11**, 29.
69 T. Gutschner, M. Hämmerle and S. Diederichs, *J. Mol. Med.*, 2013, **91**, 791–801.
70 R. Yoshimoto, A. Mayeda, M. Yoshida and S. Nakagawa, *Biochim. Biophys. Acta, Gene Regul. Mech.*, 2016, **1859**, 192–199.
71 C. Xu, M. Yang, J. Tian, X. Wang and Z. Li, *Int. J. Oncol.*, 2011, **39**, 169–175.
72 J. A. Brown, D. Bulkley, J. Wang, M. L. Valenstein, T. A. Yario, T. A. Steitz and J. A. Steitz, *Nat. Struct. Mol. Biol.*, 2014, **21**, 633–640.
73 J. E. Wilusz, C. K. JnBaptiste, L. Y. Lu, C.-D. Kuhn, L. Joshua-Tor and P. A. Sharp, *Genes Dev.*, 2012, **26**, 2392–2407.
74 A. Donlic, B. S. Morgan, J. L. Xu, A. Liu, C. Roble Jr. and A. E. Hargrove, *Angew. Chem., Int. Ed.*, 2018, **57**, 13242–13247.
75 W.-Y. Yang, R. Gao, M. Southern, P. S. Sarkar and M. D. Disney, *Nat. Commun.*, 2016, **7**, 11647.
76 J. B. Chaires, J. Ren, D. Hamelberg, A. Kumar, V. Pandya, D. W. Boykin and W. D. Wilson, *J. Med. Chem.*, 2004, **47**, 5729–5742.
77 D. Kumari, I. Gazy and K. Usdin, *Brain Sci.*, 2019, **9**, 39–57.
78 D. Colak, N. Zaninovic, M. S. Cohen, Z. Rosenwaks, W. Y. Yang, J. Gerhardt, M. D. Disney and S. R. Jaffrey, *Science*, 2014, **343**, 1002–1005.
79 M. D. Disney, B. Liu, W. Y. Yang, C. Sellier, T. Tran, N. Charlet-Berguerand and J. L. Childs-Disney, *ACS Chem. Biol.*, 2012, **7**, 1711–1718.
80 D. Kumari and K. Usdin, *Hum. Mol. Genet.*, 2016, **25**, 3689–3698.
81 P. Chiurazzi, M. G. Pomponi, R. Willemsen, B. A. Oostra and G. Neri, *Hum. Mol. Genet.*, 1998, **7**, 109–113.
82 B. Coffee, F. Zhang, S. T. Warren and D. Reines, *Nat. Genet.*, 1999, **22**, 98–101.
83 D. Kumari and K. Usdin, *Hum. Mol. Genet.*, 2014, **23**, 6575–6583.
84 Y. Lee and D. C. Rio, *Annu. Rev. Biochem.*, 2015, **84**, 291–323.
85 M. R. Lunn and C. H. Wang, *Lancet*, 2008, **371**, 2120–2133.
86 S. Lefebvre, L. Bürglen, S. Reboullet, O. Clermont, P. Burlet, L. Viollet, B. Benichou, C. Cruaud, P. Millasseau, M. Zeviani, D. Le Paslier, J. Frézal, D. Cohen, J. Weissenbach, A. Munnich and J. Melki, *Cell*, 1995, **80**, 155–165.
87 K. R. Brunden, J. Q. Trojanowski and V. M. Lee, *Nat. Rev. Drug Discovery*, 2009, **8**, 783–793.
88 M. Busslinger, N. Moschonas and R. A. Flavell, *Cell*, 1981, **27**, 289–298.
89 A. Shilo, F. A. Tosto, J. W. Rausch, S. F. J. Le Grice and T. Misteli, *Nucleic Acids Res.*, 2019, **47**, 5922–5935.
90 A. J. Bergsma, S. L. In't Groen, F. W. Verheijen, A. T. van der Ploeg and W. W. M. P. Pijnappel, *Mol. Ther. – Nucleic Acids*, 2016, **5**, 361.
91 C. F. Rochette, N. Gilbert and L. R. Simard, *Hum. Genet.*, 2001, **108**, 255–266.
92 C. L. Lorson, E. Hahnen, E. J. Androphy and B. Wirth, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6307–6311.
93 B. G. Burnett, E. Muñoz, A. Tandon, D. Y. Kwon, C. J. Sumner and K. H. Fischbeck, *Mol. Cell. Biol.*, 2009, **29**, 1107–1115.

94 T. Gidaro and L. Servais, *Dev. Med. Child Neurol.*, 2019, **61**, 19–24.
95 S. M. Hoy, *Drugs*, 2019, **79**, 1255–1262.
96 A. N. Calder, E. J. Androphy and K. J. Hodgetts, *J. Med. Chem.*, 2016, **59**, 10067–10083.
97 U.S. Food & Drug Administration, https://www.fda.gov/news-events/press-announcements/fda-approves-oral-treatment-spinal-muscular-atrophy, (accessed August 2020).
98 J. Palacino, S. E. Swalley, C. Song, A. K. Cheung, L. Shu, X. Zhang, M. Van Hoosear, Y. Shin, D. N. Chin, C. G. Keller, M. Beibel, N. A. Renaud, T. M. Smith, M. Salcius, X. Shi, M. Hild, R. Servais, M. Jain, L. Deng, C. Bullock, M. McLellan, S. Schuierer, L. Murphy, M. J. Blommers, C. Blaustein, F. Berenshteyn, A. Lacoste, J. R. Thomas, G. Roma, G. A. Michaud, B. S. Tseng, J. A. Porter, V. E. Myer, J. A. Tallarico, L. G. Hamann, D. Curtis, M. C. Fishman, W. F. Dietrich, N. A. Dales and R. Sivasankaran, *Nat. Chem. Biol.*, 2015, **11**, 511–517.
99 S. J. Kolb and J. T. Kissel, *Arch. Neurol.*, 2011, **68**, 979–984.
100 E. Goina, N. Skoko and F. Pagani, *Mol. Cell. Biol.*, 2008, **28**, 3850–3860.
101 D. A. Pomeranz Krummel, C. Oubridge, A. K. Leung, J. Li and K. Nagai, *Nature*, 2009, **458**, 475–480.
102 Y. Kondo, C. Oubridge, A. M. van Roon and K. Nagai, *eLife*, 2015, **4**, e04986.
103 S. Campagne, S. Boigner, S. Rudisser, A. Moursy, L. Gillioz, A. Knorlein, J. Hall, H. Ratni, A. Clery and F. H. Allain, *Nat. Chem. Biol.*, 2019, **15**, 1191–1198.
104 M. Sivaramakrishnan, K. D. McCarthy, S. Campagne, S. Huber, S. Meier, A. Augustin, T. Heckel, H. Meistermann, M. N. Hug, P. Birrer, A. Moursy, S. Khawaja, R. Schmucki, N. Berntenis, N. Giroud, S. Golling, M. Tzouros, B. Banfai, G. Duran-Pacheco, J. Lamerz, Y. Hsiu Liu, T. Luebbers, H. Ratni, M. Ebeling, A. Cléry, S. Paushkin, A. R. Krainer, F. H. Allain and F. Metzger, *Nat. Commun.*, 2017, **8**, 1476.
105 B. J. Blencowe, *Trends Biochem. Sci.*, 2000, **25**, 106–110.
106 A. Moursy, F. H. Allain and A. Cléry, *Nucleic Acids Res.*, 2014, **42**, 6659–6672.
107 J. Wang, P. G. Schultz and K. A. Johnson, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, e4604–e4612.
108 N. N. Singh, R. N. Singh and E. J. Androphy, *Nucleic Acids Res.*, 2007, **35**, 371–389.
109 L. Debaize and M. B. Troadec, *Cell. Mol. Life Sci.*, 2019, **76**, 259–281.
110 P. Briata, D. Bordo, M. Puppo, F. Gorlero, M. Rossi, N. Perrone-Bizzozero and R. Gherzi, *Wiley Interdiscip. Rev.: RNA*, 2016, **7**, 227–240.
111 D. W. Cleveland, S. Y. Hwo and M. W. Kirschner, *J. Mol. Biol.*, 1977, **116**, 207–225.
112 M. Goedert, M. G. Spillantini, M. C. Potier, J. Ulrich and R. A. Crowther, *EMBO J.*, 1989, **8**, 393–399.
113 Q. Gao, J. Memmott, R. Lafyatis, S. Stamm, G. Screaton and A. Andreadis, *J. Neurochem.*, 2000, **74**, 490–500.
114 M. Hong, V. Zhukareva, V. Vogelsberg-Ragaglia, Z. Wszolek, L. Reed, B. I. Miller, D. H. Geschwind, T. D. Bird, D. McKeel, A. Goate, J. C. Morris, K. C. Wilhelmsen, G. D. Schellenberg, J. Q. Trojanowski and V. M. Lee, *Science*, 1998, **282**, 1914–1917.
115 A. Grover, H. Houlden, M. Baker, J. Adamson, J. Lewis, G. Prihar, S. Pickering-Brown, K. Duff and M. Hutton, *J. Biol. Chem.*, 1999, **274**, 15134–15143.
116 L. Varani, M. Hasegawa, M. G. Spillantini, M. J. Smith, J. R. Murrell, B. Ghetti, A. Klug, M. Goedert and G. Varani, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 8229–8234.
117 C. P. Donahue, C. Muratore, J. Y. Wu, K. S. Kosik and M. S. Wolfe, *J. Biol. Chem.*, 2006, **281**, 23302–23306.
118 S. Zheng, Y. Chen, C. P. Donahue, M. S. Wolfe and G. Varani, *Chem. Biol.*, 2009, **16**, 557–566.
119 Y. Liu, E. Peacey, J. Dickson, C. P. Donahue, S. Zheng, G. Varani and M. S. Wolfe, *J. Med. Chem.*, 2009, **52**, 6523–6526.
120 J. L. Chen, P. Zhang, M. Abe, H. Aikawa, L. Zhang, A. J. Frank, T. Zembryski, C. Hubbs, H. Park, J. Withka, C. Steppan, L. Rogers, S. Cabral, M. Pettersson, T. T. Wager, M. A. Fountain, G. Rumbaugh, J. L. Childs-Disney and M. D. Disney, *J. Am. Chem. Soc.*, 2020, **142**, 8706–8727.
121 Y. Luo and M. D. Disney, *ChemBioChem*, 2014, **15**, 2041–2044.
122 T. T. Wager, X. Hou, P. R. Verhoest and A. Villalobos, *ACS Chem. Neurosci.*, 2010, **1**, 435–449.
123 M. M. Scotti and M. S. Swanson, *Nat. Rev. Genet.*, 2016, **17**, 19–32.
124 G. Biamonti, A. Amato, E. Belloni, A. Di Matteo, L. Infantino, D. Pradella and C. Ghigna, *Aging: Clin. Exp. Res.*, 2019, DOI: 10.1007/s40520-019-01360-x.
125 M. Hecker, A. Rüge, E. Putscher, N. Boxberger, P. S. Rommer, B. Fitzner and U. K. Zettl, *Autoimmun. Rev.*, 2019, **18**, 721–732.
126 S. C. Lee and O. Abdel-Wahab, *Nat. Med.*, 2016, **22**, 976–986.
127 L. M. Urbanski, N. Leclair and O. Anczuków, *Wiley Interdiscip. Rev.: RNA*, 2018, **9**, e1476.
128 N. P. Mastroyiannopoulos, M. L. Feldman, J. B. Uney, M. S. Mahadevan and L. A. Phylactou, *EMBO Rep.*, 2005, **6**, 458–463.
129 J. L. Childs-Disney, J. Hoskins, S. G. Rzuczek, C. A. Thornton and M. D. Disney, *ACS Chem. Biol.*, 2012, **7**, 856–862.
130 S. G. Rzuczek, Y. Gao, Z. Z. Tang, C. A. Thornton, T. Kodadek and M. D. Disney, *ACS Chem. Biol.*, 2013, **8**, 2312–2321.
131 S. G. Rzuczek, L. A. Colgan, Y. Nakai, M. D. Cameron, D. Furling, R. Yasuda and M. D. Disney, *Nat. Chem. Biol.*, 2017, **13**, 188–193.
132 W. Y. Yang, H. D. Wilson, S. P. Velagapudi and M. D. Disney, *J. Am. Chem. Soc.*, 2015, **137**, 5336–5345.
133 J. F. Arambula, S. R. Ramisetty, A. M. Baranger and S. C. Zimmerman, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 16068–16073.
134 A. H. Jahromi, Y. Fu, K. A. Miller, L. Nguyen, L. M. Luu, A. M. Baranger and S. C. Zimmerman, *J. Med. Chem.*, 2013, **56**, 9471–9481.
135 L. M. Luu, L. Nguyen, S. Peng, J. Lee, H. Y. Lee, C. H. Wong, P. J. Hergenrother, H. Y. Chan and S. C. Zimmerman, *ChemMedChem*, 2016, **11**, 1428–1435.

136 J. Lee, Y. Bai, U. V. Chembazhi, S. Peng, K. Yum, L. M. Luu, L. D. Hagler, J. F. Serrano, H. Y. E. Chan, A. Kalsotra and S. C. Zimmerman, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 8709–8714.
137 J. Li, J. Matsumoto, L.-P. Bai, A. Murata, C. Dohno and K. Nakatani, *Chem. – Eur. J.*, 2016, **22**, 14881–14889.
138 M. M. Lee, A. Pushechnikov and M. D. Disney, *ACS Chem. Biol.*, 2009, **4**, 345–355.
139 R. I. Benhamou, A. J. Angelbello, R. J. Andrews, E. T. Wang, W. N. Moss and M. D. Disney, *ACS Chem. Biol.*, 2020, **15**, 485–493.
140 R. I. Benhamou, A. J. Angelbello, E. T. Wang and M. D. Disney, *Cell Chem. Biol.*, 2020, **27**, 223–231.
141 W. Y. Yang, F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney, *ACS Chem. Biol.*, 2016, **11**, 2456–2465.
142 Z. Su, Y. J. Zhang, T. F. Gendron, P. O. Bauer, J. Chew, W. Y. Yang, E. Fostvedt, K. Jansen-West, V. V. Belzil, P. Desaro, A. Johnston, K. Overstreet, S. Y. Oh, P. K. Todd, J. D. Berry, M. E. Cudkowicz, B. F. Boeve, D. Dickson, M. K. Floeter, B. J. Traynor, C. Morelli, A. Ratti, V. Silani, R. Rademakers, R. H. Brown, J. D. Rothstein, K. B. Boylan, L. Petrucelli and M. D. Disney, *Neuron*, 2014, **83**, 1043–1050.
143 Z. F. Wang, A. Ursu, J. L. Childs-Disney, R. Guertler, W. Y. Yang, V. Bernat, S. G. Rzuczek, R. Fuerst, Y. J. Zhang, T. F. Gendron, I. Yildirim, B. G. Dwyer, J. E. Rice, L. Petrucelli and M. D. Disney, *Cell Chem. Biol.*, 2019, **26**, 179–190.
144 Ł. J. Sznajder, J. D. Thomas, E. M. Carrell, T. Reid, K. N. McFarland, J. D. Cleary, R. Oliveira, C. A. Nutter, K. Bhatt, K. Sobczak, T. Ashizawa, C. A. Thornton, L. P. W. Ranum and M. S. Swanson, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 4234–4239.
145 J. R. Hesselberth, *Wiley Interdiscip. Rev.: RNA*, 2013, **4**, 677–691.
146 P. E. Ash, K. F. Bieniek, T. F. Gendron, T. Caulfield, W. L. Lin, M. Dejesus-Hernandez, M. M. van Blitterswijk, K. Jansen-West, J. W. Paul, 3rd, R. Rademakers, K. B. Boylan, D. W. Dickson and L. Petrucelli, *Neuron*, 2013, **77**, 639–646.
147 T. F. Gendron, K. F. Bieniek, Y. J. Zhang, K. Jansen-West, P. E. Ash, T. Caulfield, L. Daughrity, J. H. Dunmore, M. Castanedes-Casey, J. Chew, D. M. Cosio, M. van Blitterswijk, W. C. Lee, R. Rademakers, K. B. Boylan, D. W. Dickson and L. Petrucelli, *Acta Neuropathol.*, 2013, **126**, 829–844.
148 K. Mori, T. Arzberger, F. A. Grasser, I. Gijselinck, S. May, K. Rentzsch, S. M. Weng, M. H. Schludi, J. van der Zee, M. Cruts, C. Van Broeckhoven, E. Kremmer, H. A. Kretzschmar, C. Haass and D. Edbauer, *Acta Neuropathol.*, 2013, **126**, 881–893.
149 T. Zu, B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. C. Ingram, Z. Nan, C. Forster, W. C. Low, B. Schoser, N. V. Somia, H. B. Clark, S. Schmechel, P. B. Bitterman, G. Gourdon, M. S. Swanson, M. Moseley and L. P. W. Ranum, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 260–265.
150 P. K. Todd, S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur, K. Elenitoba-Johnson, J. P. Vonsattel, E. D. Louis, M. A. Sutton, J. P. Taylor, R. E. Mills, N. Charlet-Berguerand and H. L. Paulson, *Neuron*, 2013, **78**, 440–455.
151 T. F. Gendron, V. V. Belzil, Y. J. Zhang and L. Petrucelli, *Acta Neuropathol.*, 2014, **127**, 359–376.
152 W.-Y. Yang, H. D. Wilson, S. P. Velagapudi and M. D. Disney, *J. Am. Chem. Soc.*, 2015, **137**, 5336–5345.
153 T. F. Gendron, J. Chew, J. N. Stankowski, L. R. Hayes, Y. J. Zhang, M. Prudencio, Y. Carlomagno, L. M. Daughrity, K. Jansen-West, E. A. Perkerson, A. O'Raw, C. Cook, L. Pregent, V. Belzil, M. van Blitterswijk, L. J. Tabassian, C. W. Lee, M. Yue, J. Tong, Y. Song, M. Castanedes-Casey, L. Rousseau, V. Phillips, D. W. Dickson, R. Rademakers, J. D. Fryer, B. K. Rush, O. Pedraza, A. M. Caputo, P. Desaro, C. Palmucci, A. Robertson, M. G. Heckman, N. N. Diehl, E. Wiggs, M. Tierney, L. Braun, J. Farren, D. Lacomis, S. Ladha, C. N. Fournier, L. F. McCluskey, L. B. Elman, J. B. Toledo, J. D. McBride, C. Tiloca, C. Morelli, B. Poletti, F. Solca, A. Prelle, J. Wuu, J. Jockel-Balsarotti, F. Rigo, C. Ambrose, A. Datta, W. Yang, D. Raitcheva, G. Antognetti, A. McCampbell, J. C. Van Swieten, B. L. Miller, A. L. Boxer, R. H. Brown, R. Bowser, T. M. Miller, J. Q. Trojanowski, M. Grossman, J. D. Berry, W. T. Hu, A. Ratti, B. J. Traynor, M. D. Disney, M. Benatar, V. Silani, J. D. Glass, M. K. Floeter, J. D. Rothstein, K. B. Boylan and L. Petrucelli, *Sci. Transl. Med.*, 2017, **9**, eaai7866.
154 T. F. Gendron, M. van Blitterswijk, K. F. Bieniek, L. M. Daughrity, J. Jiang, B. K. Rush, O. Pedraza, J. A. Lucas, M. E. Murray, P. Desaro, A. Robertson, K. Overstreet, C. S. Thomas, J. E. Crook, M. Castanedes-Casey, L. Rousseau, K. A. Josephs, J. E. Parisi, D. S. Knopman, R. C. Petersen, B. F. Boeve, N. R. Graff-Radford, R. Rademakers, C. Lagier-Tourenne, D. Edbauer, D. W. Cleveland, D. W. Dickson, L. Petrucelli and K. B. Boylan, *Acta Neuropathol.*, 2015, **130**, 559–573.
155 M. Clamp, B. Fry, M. Kamal, X. Xie, J. Cuff, M. F. Lin, M. Kellis, K. Lindblad-Toh and E. S. Lander, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 19428–19433.
156 A. Hopkins and C. Groom, *Nat. Rev. Drug Discovery*, 2002, **1**, 727–730.
157 C. V. Dang, E. P. Reddy, K. M. Shokat and L. Soucek, *Nat. Rev. Cancer*, 2017, **17**, 502–508.
158 J. Spiegel, P. M. Cromm, G. Zimmermann, T. N. Grossmann and H. Waldmann, *Nat. Chem. Biol.*, 2014, **10**, 613–622.
159 P. Zhang, H. J. Park, J. Zhang, E. Junn, R. J. Andrews, S. P. Velagapudi, D. Abegg, K. Vishnu, M. G. Costales, J. L. Childs-Disney, A. Adibekian, W. N. Moss, M. M. Mouradian and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 1457–1467.
160 V. M. Lee and J. Q. Trojanowski, *Neuron*, 2006, **51**, 33–38.
161 M. G. Spillantini, M. L. Schmidt, V. M.-Y. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert, *Nature*, 1997, **388**, 839–840.

162 K. C. Luk, V. Kehm, J. Carroll, B. Zhang, P. O'Brien, J. Q. Trojanowski and V. M.-Y. Lee, *Science*, 2012, **338**, 949–953.
163 A. B. Singleton, M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum, S. Lincoln, A. Crawley, M. Hanson, D. Maraganore, C. Adler, M. R. Cookson, M. Muenter, M. Baptista, D. Miller, J. Blancato, J. Hardy and K. Gwinn-Hardy, *Science*, 2003, **302**, 841.
164 E. Junn, K. W. Lee, B. S. Jeong, T. W. Chan, J. Y. Im and M. M. Mouradian, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13052–13057.
165 D. M. Maraganore, *J. Mov. Disord.*, 2011, **4**, 1–7.
166 A. L. Friedlich, R. E. Tanzi and J. T. Rogers, *Mol. Psychiatry*, 2007, **12**, 222–223.
167 F. Febbraro, M. Giorgi, S. Caldarola, F. Loreni and M. Romero-Ramos, *NeuroReport*, 2012, **23**, 576–580.
168 J. L. Childs-Disney, T. Tran, B. R. Vummidi, S. P. Velagapudi, H. S. Haniff, Y. Matsumoto, G. Crynen, M. R. Southern, A. Biswas, Z. F. Wang, T. L. Tellinghuisen and M. D. Disney, *Chem*, 2018, **4**, 2384–2404.
169 J. R. Mazzulli, F. Zunke, O. Isacson, L. Studer and D. Krainc, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 1931–1936.
170 B. Liu, J. L. Childs-Disney, B. M. Znosko, D. Wang, M. Fallahi, S. M. Gallo and M. D. Disney, *BMC Bioinf.*, 2016, **17**, 112.
171 H. Pimentel, N. L. Bray, S. Puente, P. Melsted and L. Pachter, *Nat. Methods*, 2017, **14**, 687–690.
172 D. L. Boger and H. Cai, *Angew. Chem., Int. Ed.*, 1999, **38**, 448–476.
173 B. J. Carter, E. de Vroom, E. C. Long, G. A. van der Marel, J. H. van Boom and S. M. Hecht, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 9373–9377.
174 A. T. Abraham, J. J. Lin, D. L. Newton, S. Rybak and S. M. Hecht, *Chem. Biol.*, 2003, **10**, 45–52.
175 A. J. Angelbello and M. D. Disney, *ChemBioChem*, 2018, **19**, 43–47.
176 L. Ma, J. Teruya-Feldstein and R. A. Weinberg, *Nature*, 2007, **449**, 682–688.
177 Y. Li and M. D. Disney, *ACS Chem. Biol.*, 2018, **13**, 3065–3071.
178 Q. Ma, Z. Xu, B. R. Schroeder, W. Sun, F. Wei, S. Hashimoto, K. Konishi, C. J. Leitheiser and S. M. Hecht, *J. Am. Chem. Soc.*, 2007, **129**, 12439–12452.
179 Z.-D. Xu, M. Wang, S.-L. Xiao, C.-L. Liu and M. Yang, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 2595–2599.
180 V. Agarwal, G. W. Bell, J.-W. Nam and D. P. Bartel, *eLife*, 2015, **4**, e05005.
181 A. J. Angelbello, S. G. Rzuczek, K. K. Mckee, J. L. Chen, H. Olafson, M. D. Cameron, W. N. Moss, E. T. Wang and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 7799–7804.
182 A. Mankodi, E. Logigian, L. Callahan, C. McClain, R. White, D. Henderson, M. Krym and C. A. Thornton, *Science*, 2000, **289**, 1769–1773.
183 A. Mankodi, M. P. Takahashi, H. Jiang, C. L. Beck, W. J. Bowers, R. T. Moxley, S. C. Cannon and C. A. Thornton, *Mol. Cell*, 2002, **10**, 35–44.
184 R. H. Blum, S. K. Carter and K. Agre, *Cancer*, 1973, **31**, 903–914.
185 Y. Katz, E. T. Wang, E. M. Airoldi and C. B. Burge, *Nat. Methods*, 2010, **7**, 1009–1015.
186 K. M. Sakamoto, K. B. Kim, A. Kumagai, F. Mercurio, C. M. Crews and R. J. Deshaies, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8554–8559.
187 M. G. Costales, Y. Matsumoto, S. P. Velagapudi and M. D. Disney, *J. Am. Chem. Soc.*, 2018, **140**, 6741–6744.
188 R. H. Silverman, *J. Virol.*, 2007, **81**, 12720–12729.
189 M. G. Costales, B. Suresh, K. Vishnu and M. D. Disney, *Cell Chem. Biol.*, 2019, **26**, 1180–1186.
190 M. G. Costales, H. Aikawa, Y. Li, J. L. Childs-Disney, D. Abegg, D. G. Hoch, S. Pradeep Velagapudi, Y. Nakai, T. Khan, K. W. Wang, I. Yildirim, A. Adibekian, E. T. Wang and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 2406–2411.
191 C. S. Thakur, B. K. Jha, B. Dong, J. Das Gupta, K. M. Silverman, H. Mao, H. Sawai, A. O. Nakamura, A. K. Banerjee, A. Gudkov and R. H. Silverman, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9585–9590.
192 P. P. Graczyk, *J. Med. Chem.*, 2007, **50**, 5773–5779.
193 P. Mu, Y. C. Han, D. Betel, E. Yao, M. Squatrito, P. Ogrodowski, E. de Stanchina, A. D'Andrea, C. Sander and A. Ventura, *Genes Dev.*, 2009, **23**, 2806–2811.
194 K. Conkrite, M. Sundby, S. Mukai, J. M. Thomson, D. Mu, S. M. Hammond and D. MacPherson, *Genes Dev.*, 2011, **25**, 1734–1745.
195 G. Huang, K. Nishimoto, Z. Zhou, D. Hughes and E. S. Kleinerman, *Cancer Res.*, 2012, **72**, 908–916.
196 K. Kim, G. Chadalapaka, S. O. Lee, D. Yamada, X. Sastre-Garau, P. A. Defossez, Y. Y. Park, J. S. Lee and S. Safe, *Oncogene*, 2012, **31**, 1034–1044.
197 V. K. Topkara and D. L. Mann, *Cardiovasc. Drugs Ther.*, 2011, **25**, 171–182.
198 E. Mogilyansky and I. Rigoutsos, *Cell Death Differ.*, 2013, **20**, 1603–1614.
199 S. P. Kabekkodu, V. Shukla, V. K. Varghese, J. D'Souza, S. Chakrabarty and K. Satyamoorthy, *Biol. Rev.*, 2018, **93**, 1955–1986.
200 X. Liu, H. S. Haniff, J. L. Childs-Disney, A. Shuster, H. Aikawa, A. Adibekian and M. D. Disney, *J. Am. Chem. Soc.*, 2020, **142**, 6970–6982.
201 M. G. Costales, J. L. Childs-Disney, H. S. Haniff and M. D. Disney, *J. Med. Chem.*, 2020, DOI: 10.1021/acs.jmedchem.9b01927.
202 E. D. Wieben, R. A. Aleff, X. Tang, M. L. Butz, K. R. Kalari, E. W. Highsmith, J. Jen, G. Vasmatzis, S. V. Patel, L. J. Maguire, K. H. Baratz and M. P. Fautsch, *Invest. Ophthalmol. Visual Sci.*, 2017, **58**, 343–352.
203 R. L. Margolis and D. D. Rudnicki, *Curr. Opin. Neurol.*, 2016, **29**, 743–748.
204 ClinicalTrials.gov, https://clinicaltrials.gov/ct2/show/NCT02268552, accessed July 2020.

Chem Soc Rev

TUTORIAL REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/d0cs00455c

Design of small molecules targeting RNA structure from sequence†

Andrei Ursu,[a] Jessica L. Childs-Disney,[a] Ryan J. Andrews,[b] Collin A. O'Leary,[b] Samantha M. Meyer,[a] Alicia J. Angelbello,[a] Walter N. Moss*[b] and Matthew D. Disney*[a]

Received 29th April 2020

DOI: 10.1039/d0cs00455c

rsc.li/chem-soc-rev

The design and discovery of small molecule medicines has largely been focused on a small number of druggable protein families. A new paradigm is emerging, however, in which small molecules exert a biological effect by interacting with RNA, both to study human disease biology and provide lead therapeutic modalities. Due to this potential for expanding target pipelines and treating a larger number of human diseases, robust platforms for the rational design and optimization of small molecules interacting with RNAs (SMIRNAs) are in high demand. This review highlights three major pillars in this area. First, the transcriptome-wide identification and validation of structured RNA elements, or motifs, within disease-causing RNAs directly from sequence is presented. Second, we provide an overview of high-throughput screening approaches to identify SMIRNAs as well as discuss the lead identification strategy, Inforna, which decodes the three-dimensional (3D) conformation of RNA motifs with small molecule binding partners, directly from sequence. An emphasis is placed on target validation methods to study the causality between modulating the RNA motif *in vitro* and the phenotypic outcome in cells. Third, emergent modalities that convert occupancy-driven mode of action SMIRNAs into event-driven small molecule chemical probes, such as RNA cleavers and degraders, are presented. Finally, the future of the small molecule RNA therapeutics field is discussed, as well as hurdles to overcome to develop potent and selective RNA-centric chemical probes.

Key learning points

- Aberrant RNA structure contributes to the pathology of numerous human diseases.
- Structured, evolutionarily conserved RNA motifs can be predicted directly from sequence with the state-of-the-art computational tool, ScanFold.
- Inforna decodes these evolutionarily conserved RNA 3D folds with small molecules to provide high-quality chemical probes.
- Robust target engagement techniques are necessary to validate RNA-centric modes of action.
- Emergent therapeutic modalities include RNA-targeted degraders and cleavers that destroy disease-causing RNAs.

1. Introduction

Most drug discovery campaigns, both past and present, are focused on protein targets. Decades of technological advancements and scientific discoveries have been dedicated to exploring the proteome and modulating protein activity for therapeutic benefit. These efforts yielded chemical probes to test mechanistic hypotheses, uncover new biology, and manipulate biological processes. Ultimately, this knowledge has been translated into novel and safe medicines for a plethora of human diseases. However, druggable proteins are confined to a small set of families. To expand druggability and increase our understanding of disease biology, many have turned to RNA targets. RNA is best known for its role in translation, where messenger RNAs (mRNAs) are translated into proteins *via* the ribosome, a complex macromolecular machine composed of ribosomal RNAs (rRNAs) and proteins, in conjunction with transfer RNAs (tRNAs). The functions of RNA, however, go well beyond this critical role in biology. For example, RNA molecules encode unique secondary and tertiary structures that have biological functions on their own (acting in *cis*) or can recruit other factors (RNAs, proteins) to assist in their function (acting in *trans*).

[a] *Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL, 33458, USA. E-mail: disney@scripps.edu*
[b] *Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, Iowa, USA. E-mail: wmoss@iastate.edu*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cs00455c

As many disease phenotypes can be traced back to dysregulation of RNA function, various approaches have been employed to target disease-causing RNAs for therapeutic benefit. The two most studied modalities are antisense oligonucleotides (ASOs) and small molecules, *i.e.*, small molecules interacting with RNAs (SMIRNAs), which fundamentally differ in their modes of action.[1] ASOs, in general, consist of modified nucleotides, either *via* the backbone or sugar moiety, and are designed by sequence complementarity. That is, ASOs recognize RNA primary sequence (Fig. 1A) and hybridize to cognate disease-causing RNAs to: (i) sterically block the assembly of RNA–protein or RNA–RNA interactions; or (ii) promote degradation of the disease-causing RNAs *via* Ribonuclease H (RNase H), an endoribonuclease that hydrolyzes the phosphodiester bonds of the RNA strand in RNA–DNA heteroduplexes. Although the design and generation of complementary ASOs for any given disease-causing RNA is rapid and straightforward, their binding sites must be accessible, *i.e.*, unstructured. Both RNA's intramolecular (secondary and tertiary) structures and intermolecular structures with other biomolecules can affect ASO binding in cellular context.

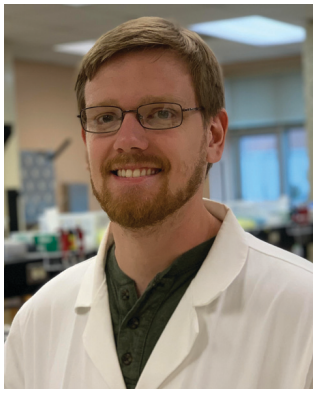
In contrast to ASOs, SMIRNAs recognize unique three-dimensional (3D) RNA conformations, or structure. RNA secondary structure is dictated by its sequence, which restricts and directs the formation of intramolecular base pairing, generating helical regions interspersed with loops, bulges, and hairpins (Fig. 1B) (see ref. 2 and citations therein for a detailed description of structured RNA motifs). That is, the overall secondary structure of an RNA can be viewed as modules of structured elements, or motifs, strung together. Though built only on four nucleotide building blocks, RNA sequence encodes dynamic and sufficiently unique ensembles of 3D folds that can be targeted and/or

Andrei Ursu

Andrei Ursu earned his PhD (2015) in Chemical Biology under the supervision of Prof. Herbert Waldmann at the Max Planck Institute of Molecular Physiology, Dortmund, Germany, focusing on chemical reprogramming of stem cells with small molecules. Since 2016, Dr. Ursu has been working as a Postdoctoral Fellow in the lab of Prof. Matthew D. Disney at The Scripps Research Institute in Jupiter, Florida. His focus is on assessing the ligandability of GGGGCC repeat expansions, that cause genetically defined amyotrophic lateral sclerosis (ALS), with small molecule chemical probes.

Ryan J. Andrews

Ryan J. Andrews is a PhD student in the lab of Prof. Walter Moss at Iowa State University. His research has been dedicated to the computational prediction of RNA structures in large RNA sequences. His initial research project involved scanning the whole human genome for unusually stable RNA secondary structures (compiled in the RNAStructuromeDB). More recently his research led to the development of the ScanFold method, which has since been used to analyze human messenger RNAs (mRNAs) and viral genomes in order to discover functional RNA structures as well as structured RNA drug targets.



Collin A. O'Leary

Collin A. O'Leary is currently pursuing his PhD at Iowa State University in the lab of Prof. Walter Moss. His research is focused on understanding the structure and function relationships of RNAs in human and pathogenic transcriptomes. Some of Collin's initial work included computational prediction of functional regions in the human mRNA MYC, followed by experimental validation of a predicted functional region. His current work is focused on combining a dimethyl sulphate (DMS)-based RNA structure sequencing protocol with the ScanFold method to yield biochemically informed, transcriptome-wide structure models.

Samantha M. Meyer

Samantha M. Meyer received her BS in Biochemistry and Molecular Biology from the University of Wisconsin – Eau Claire in 2019. The following fall she began doctoral studies under the guidance of Prof. Matthew D. Disney at The Scripps Research Institute in Jupiter, Florida. Her current research focuses on targeting disease-causing RNAs with small molecules, with an emphasis on expanding the versatility of Ribonuclease Targeting Chimeras (RIBOTACs).

stabilized selectively by small molecules (Fig. 1C). Importantly, RNA secondary structure can be predicted or determined accurately from RNA sequence. Secondary structure then constrains available tertiary interactions and thus tertiary structure (Fig. 1C). As tertiary structures are generally weak, they can be disrupted by small molecule binding, affecting the RNA's function.

Small molecules offer several advantages that support their use as a viable modality to target 3D folds of structured motifs within RNA. For example, structurally related analogs can be used to define structure–activity relationships (SAR), informing lead optimization for biological activity and selectivity. Moreover, SMIRNAs targeting adjacent structured RNA motifs can be covalently linked together, yielding dimeric molecules with increased binding affinity and selectivity compared to the individual compounds from which they were derived.[1] Finally, SMIRNAs can be functionalized with various modules to affect direct cleavage, to induce degradation *via* recruitment of endogenous nucleases,[3] or to image disease-causing RNAs through on-site synthesis of a Förster resonance energy transfer (FRET) pair. These features expand the mode of action of SMIRNAs to explore RNA biology and to provide therapeutic opportunities for many human diseases mediated by RNA structures.

This review highlights three key components required to design high-quality SMIRNAs with defined RNA-centric modes of action: (1) state-of-the-art approaches to identify ligandable 3D structured motifs within RNA that are evolutionarily conserved and hence likely to be functional; (2) methods to target structured motifs within RNA; and (3) RNA target validation methods. We also highlight novel modalities developed by converting occupancy-driven SMIRNAs into event-driven chemical probes (RNA cleavers and degraders) that ablate disease-causing RNAs. Finally, we offer an overview of the future challenges that need to be overcome to facilitate the design and optimization of potent and selective small molecule RNA therapeutics in a robust and rational fashion. A comprehensive review of targeting disease-causing RNAs extending beyond this tutorial can be found in ref. 4.

2. The role of RNA structure in biology & disease

RNA structure is intimately linked to both normal biology and disease pathology.[5] RNA structures range from simple loops and bulges to more complex structures such as coaxial stacking, pseudoknots, and other tertiary structures. Indeed, these structures influence and dictate human biology, ranging from the regulation of translation, to splice site selection, and catalysis. RNA structure also controls viral replication and infection as well as bacterial gene expression (riboswitches). As these topics have been extensively reviewed, we direct the reader to the excellent references below for additional details.

Not surprisingly, RNA mutation and aberrant expression can trigger disease by causing deregulation of normal cellular processes. For example, transcriptomic studies have revealed that microRNAs (miRNAs), small regulatory RNAs that modulate gene expression by binding to complementary mRNAs, are commonly dysregulated in tumor tissue, suggesting a mechanism by which cancer cells downregulate tumor suppressor genes or enhance expression of oncogenes. Aberrant expression of miRNAs, whether up- or down-regulated, has been linked to many other diseases, including cardiovascular disease, inflammatory and neurodevelopmental disorders and liver disease.

RNA structure has also been implicated in many neurological disorders. RNA repeat expansions cause over 30 human diseases, including Huntington's disease (HD) [r(CAG)^{exp}], amyotrophic lateral sclerosis (ALS) [r(G_4C_2)^{exp}] and myotonic dystrophy type 1 (DM1) [r(CUG)^{exp}]. In these disorders, the repeating RNA, often found in intronic or untranslated regions (UTRs), forms hairpin structures containing repeating structured RNA motifs that interfere with normal RNA processing and function. These structures can sequester RNA-binding proteins, lead to the formation of nuclear foci, and undergo repeat-associated non-ATG (RAN) translation. This disruption in normal biology has substantial consequences, leading to

Alicia J. Angelbello

Alicia J. Angelbello received her BS in Chemistry from Villanova University. She joined the Disney laboratory in 2015 as a graduate student where she works on developing small molecules to target RNA repeat expansions.

Walter N. Moss

Walter N. Moss is an Assistant Professor at Iowa State University whose research focuses on identifying and characterizing functional RNA structures using computational and experimental approaches. His long-term goal is to establish methodological pipelines that facilitate the discovery of structural motifs with significance to human health and disease using innovative in silico tools, biochemical and cell/molecular biological approaches.

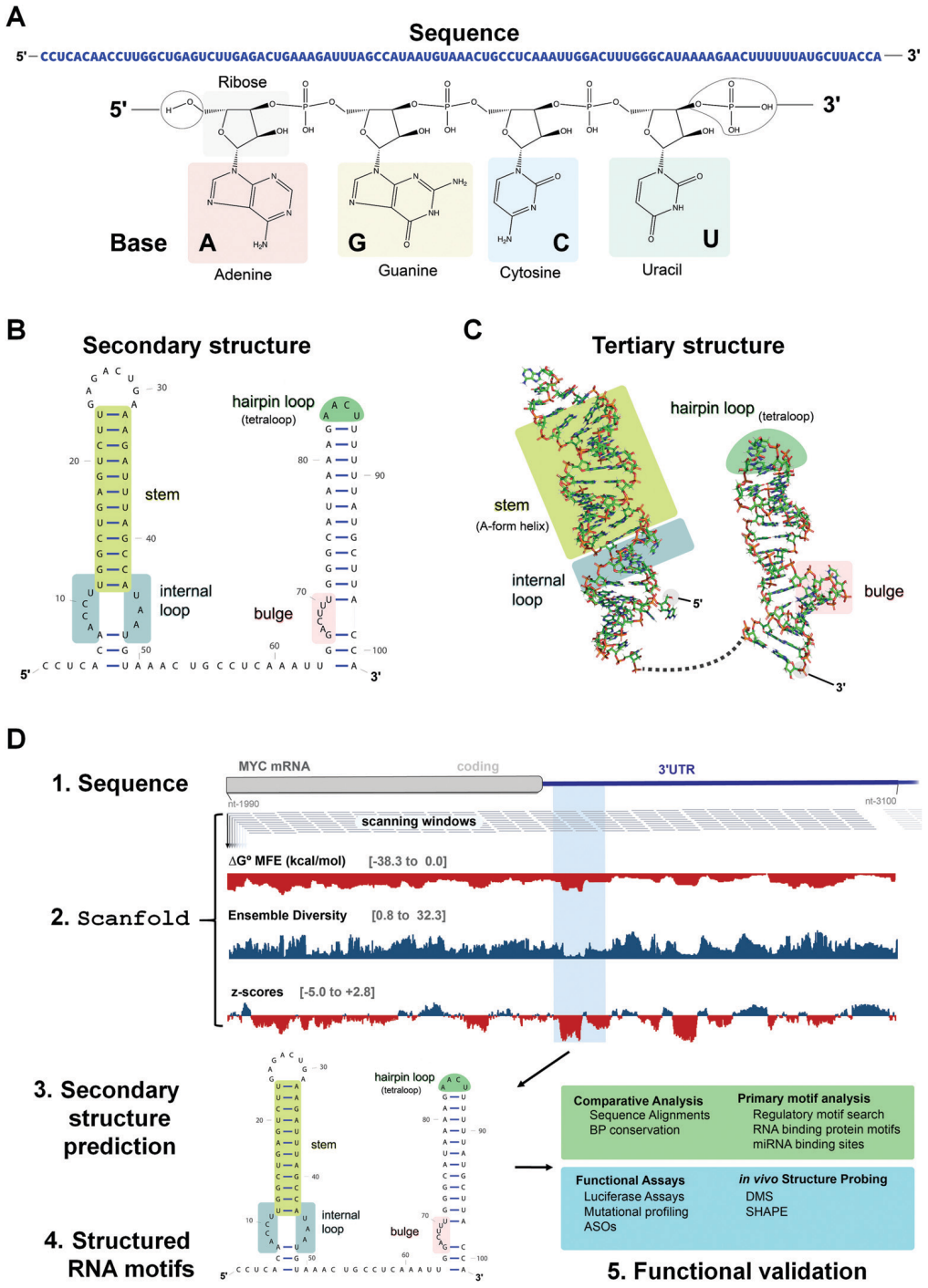


Fig. 1 Overview of RNA structure and its prediction directly from sequence using ScanFold. (A) The primary structure of RNA, *i.e.*, sequence, consists of four bases; two purines, adenine (A) and guanine (G), and two pyrimidines, cytosine (C) and uracil (U). (B) The secondary structure of an RNA consists of the non-covalent bonds that form between A and U, G and C, or G and U, bases. These pairings consist of hydrogen bonds and base stacking interactions which form stems (light green) and are often punctuated with internal loops (blue), bulges (pink), and hairpin loops (dark green). (C) The tertiary structure of RNA is largely dictated by the base pairs that form the secondary structure. Stems (light green) will form structured A-form helices and internal loops (blue), bulges (pink), and hairpin loops (dark green) will be less structured, more accessible regions that distort the more rigid helix and offer sites for *trans*-acting factors to bind in a sequence specific manner. Here, the dotted black line represents the single strand between the two more structured hairpins. (D) Identification of structured RNA motifs within the mRNA sequence of *MYC via* ScanFold. Portions of the *MYC* mRNA coding region and 3′ untranslated region (UTR) are depicted with overlapping ScanFold analysis windows below. In each scanning analysis window, ScanFold calculates numerous folding metrics including the minimum free energy (MFE), ensemble diversity, and *z*-scores which are depicted as bar graphs. It is important to note that a window will be represented by a single bar, but the downstream nucleotides (nt) (corresponding to the window size) are used to predict the metrics. ScanFold then determines the most stable and significant base pairs and uses them to generate a consensus structure (displayed as an arc diagram). Regions with highly negative *z*-scores and low ensemble diversity indicate regions of presumed function, with one (or few) dominating structures and that may merit further, in-depth analyses: *e.g.*, comparative analysis, additional bioinformatics analyses, functional assays, and structure probing assays. These techniques can further characterize and validate the biological function of the structured RNA motif.

disease pathologies that are both common and unique to different microsatellite disorders.

Collectively, regulation and maintenance of RNA structure is critically important to sustain normal biology, and identification of novel functional RNA structures (discussed below) featuring motifs that can be targeted with SMIRNAs will be critically important to study RNA's role in disease for therapeutic benefit.

3. Methods to identify functional RNA structures *via* evolutionary conservation

3.1 Overview of RNA Structure Prediction

An RNA structure is defined by the intramolecular base pairs which form as the RNA molecule folds back on itself, *i.e.*, by the helices formed between complementary stretches of RNA. The composite strength of base pairs in a secondary structure are relatively strong compared to the weaker interactions that form an RNA's tertiary structure. Generally, the formation of RNA tertiary structure does not alter the underlying secondary structure and is instead guided by it in a hierarchical manner. Therefore, the accurate prediction of RNA secondary structure is highly valuable when defining a RNA's structured landscape in order to: (i) generate biological hypotheses about RNA structure–function relationships; and (ii) identify structured 3D folds within RNA for modulation with SMIRNAs.

When predicting a single secondary structure model for a given RNA sequence, the most frequently used method is free energy minimization. This method calculates the most stable secondary structure (*i.e.*, the structure with the most negative ΔG_{37}°) as evaluated from an underlying set of experimentally-derived thermodynamic parameters. The key assumption is the base pairing pattern that yields the most stable minimum free energy (MFE) secondary structure is also the best representation of the native fold. The reality of RNA folding is of course much more complicated in the cellular milieu, where a multitude of 3D conformations can not only exist, but also interconvert, depending on environmental factors and external stimuli. Therefore, the predictions made *via* free energy minimization methods serve only as a valuable guide for building hypotheses as to the structured RNA motifs responsible for the phenotype(s) of interest.

The accuracy of secondary structure prediction by free energy minimization, however, decreases with sequences >700 nucleotides (such as mRNAs or viral genomes).[6] For example, RNA folding algorithms performed best when the analyzed sequence length was restricted to between 100 and 150 nucleotides, thus limiting the analysis to locally stable RNA regions rather than calculating the most globally stable structure. Further, free energy minimization alone cannot clearly define whether a structured RNA motif is functional.

Recently, tools have been developed to predict structured RNA motifs throughout the transcriptome.[7] These tools consider two hallmarks of functional RNAs: (i) unusual structural stability; and (ii) evolutionarily conserved base pairs. These approaches focus on finding not only well-defined, *i.e.*, stable RNA structures, but also structured elements that are more stable than expected for their nucleotide composition (as characterized by the thermodynamic *z*-score eqn (1)). Further, if a specific RNA structure is likely to be functional, conservation across homologous sequences, as indicated by mutations which retain the secondary structure, should be observed.

$$z\text{-score} = \frac{\mathrm{MFE}_{\mathrm{native}} - \overline{\mathrm{MFE}_{\mathrm{random}}}}{\sigma} \quad (1)$$

As shown in eqn (1), the *z*-score compares the MFE of a sequence within an RNA of interest ($\mathrm{MFE}_{\mathrm{native}}$) to the average MFE of a set of randomized RNA sequences ($\mathrm{MFE}_{\mathrm{random}}$), normalized by the standard deviation (SD; σ) of the MFE. That is, a native RNA sequence that is more thermodynamically stable (lower MFE) than a set of randomized sequences will yield a negative *z*-score and be considered to form a stable structure. The *z*-score reports the number of SDs the native MFE is away from the average MFE from random sequences with similar nucleotide composition.

Indeed, the most reliable tools to date for computational prediction of functional RNA secondary structures from sequence[7] incorporate these strategies. The Moss Lab has recently developed a computational method which prioritizes RNA structural characterization and analysis followed by conservation analysis. This method, named ScanFold,[8] characterizes the structured landscape of any large RNA sequence (Fig. 1D and Table S1, ESI†). In brief, ScanFold analyzes RNA sequences using a scanning window approach and reports the results of MFE and ensemble-based predictions across the entire sequence.

Whenever available, the predicted secondary structures are further validated with RNA structural data obtained from chemical probing experiments in cells, for example using dimethyl sulfate (DMS) or selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE).[9,10] These chemical probing reagents react with non-canonically paired or single stranded nucleotides, modifying the bases in the case of DMS or the sugar moieties of dynamic nucleotides in the case of SHAPE. After modification, the RNAs are then analyzed by RNA sequencing (RNA-seq), which requires reverse transcription (RT) and polymerase chain reaction (PCR) amplification. Reaction of a nucleotide with a mapping reagent creates a unique signature during reverse transcription, either by preventing readthrough resulting in a ''stop'' or creating a mutation. The reactivity of each nucleotide with the chemical modifying reagent, or the extent of mutation or termination of the RT-PCR step, is calculated as normalized to untreated RNA. Increased reactivity indicates that the nucleotide is not canonically base paired. These data can then be used as checks on existing structure models or used directly during MFE calculations as a constraint on secondary structure predictions.

Base reactivities from structure probing are calculated and can be incorporated as constraints during MFE calculations in programs such as RNAfold[11] (which ScanFold utilizes) or RNAstructure (Table S1),[6] and are then cross-referenced

with ScanFold results. Incorporating such data helps to yield biologically relevant models of RNA secondary structure(s). Notably, results from chemical probing experiments must be carefully controlled and the statistical confidence of the resulting data must be calculated, as various artifacts arising from transcriptional noise, limitations of high-throughput experimentation, and computational analysis errors can generate erroneous RNA structures.

To date, ScanFold has been applied to several genomes, including human,[12] human immunodeficiency virus type 1 (HIV-1) and Zika virus (ZIKV),[8] as well as mRNA sequences encoding microtubule-associated protein tau (tau)[13] and α-synuclein (SNCA),[14] the results of which are summarized below.

3.2 Validation of ScanFold: structured RNA Motifs within HIV-1 and ZIKV

The genomes of ZIKV and HIV-1 are composed of positive (+) sense (protein-coding) single-stranded RNA molecules: 10 807 and 9175 nucleotides (nt) in length, respectively. Their small genomes are translation-competent and, much like mRNAs, are composed of coding and non-coding regions, the former flanked by 5′ and 3′ UTRs. Because of their relevance to human health, these viruses have been studied extensively and were each found to utilize several structured RNA motifs to carry out aspects of their viral life cycles including replication, packaging, and translation. Due to the thorough structural and functional characterization of these viral RNA structures, the HIV-1 and ZIKV genomes served as ideal RNA sequences to test ScanFold's ability to detect structured RNA motifs.

The ScanFold platform, introduced in Andrews *et al.*,[8] accurately identified all known functional structures from the HIV-1 and ZIKV genomes and revealed additional potentially structured RNA motifs throughout each. The ideal settings for detecting known structures in HIV-1 and ZIKV were optimized in this report (where a window size of 120 nt was found to best recapitulate known functional models). In a follow-up report, a detailed description of these settings were described to advise researchers using ScanFold on how to adjust settings for any RNA sequence.[15] An emphasis was placed on practical usage, for quick and accurate characterization of an RNA's overall landscape of structured motifs. In this follow-up study, it was also revealed that ScanFold's characterizations of HIV-1 and ZIKV agreed with available SHAPE probing data, accurately characterizing RNA regions as either housing a uniquely structured RNA motif (where low *z*-score structures correlated with unambiguous experimental results and high prediction accuracy) or a more dynamic/loose structure (where more positive *z*-score motifs correlated to experimental results which allow more than one structural interpretation and suggest an overall unstructured nature). These results showed that while ScanFold excels at highlighting potential (and known) structured RNA motifs, it can also accurately characterize an RNA's structural landscape. Importantly, such results can be obtained quickly, easily, and using only a single sequence to point investigators towards potentially structured RNA motifs, which are likely to be biologically relevant.

4. Methods to target functional, evolutionarily conserved structures

The next challenge is to exploit the discovery of evolutionarily conserved structures to design small molecules that selectively recognize them and modulate RNA function. There are at least three critical factors for the development of SMIRNAs: (i) exploration of diverse chemical space to identify privileged chemotypes that selectively bind structured 3D folds within RNA(s); (ii) complementary exploration of structured RNA 3D folds within disease-causing RNAs that form well-defined pockets for small molecule ligands; and (iii) development of a bioinformatic platform that links (i) and (ii) and ultimately yields bioactive SMIRNAs against disease-causing RNAs.

In order to gather data on the first two key factors, we developed the selection-based strategy termed two-dimensional combinatorial screening (2DCS) (Fig. 2).[1,16–18] 2DCS is a massively parallel screening method that probes the interaction of small molecule libraries against libraries of structured RNA motifs found within cellular RNAs. The library-*vs*-library screen is performed by covalently immobilizing or absorbing (dubbed AbsorbArray[16]) small molecules onto agarose-coated microarrays, followed by incubation with a labeled library of RNA motifs (Fig. 2). These RNA libraries contain thousands of structured RNA motifs in discrete patterns, featuring bulges, hairpins, internal loops, *etc.* The screen is performed in the presence of excess competitor RNAs that mimic regions common to all RNA library members, DNA and RNA base pairs, and/or tRNAs (Fig. 2). That is, the screen is completed under conditions of high oligonucleotide stringency. This screening format can be performed with structurally related small molecules such that SAR can be derived or with diverse chemical matter to expand our understanding of chemotypes that confer avidity and selectivity for RNA. This experimental approach is highly advantageous when compared to other small molecule microarray (SMM) approaches, which typically screen a single RNA target at a time.[19]

RNAs that bind each small molecule are isolated from the surface of the 2DCS microarray, amplified, and subjected to RNA-seq analysis (Fig. 2).[16] Simultaneously, an aliquot of the RNA library that was not incubated with the array is also amplified and analyzed by RNA-seq. The RNA-seq data undergo a rigorous statistical analysis, named High Throughput Structure–Activity Relationships Through Sequencing (HiT-StARTS), where the frequency of each structured RNA 3D fold bound to the small molecule is compared to the frequency of each structured motif in the starting library.[17] A pooled population comparison calculates the statistical significance of the enrichment, reported as a *Z*-score (Z_{obs}) (Fig. 2). We have shown that a $Z_{obs} > 8$ represents an avid RNA motif-small molecule interaction and that the relative affinity of the interactions for a given SMIRNA directly correlates with Z_{obs}.[17] Importantly, the output of 2DCS and HiT-StARTS are privileged RNA 3D fold-small molecule interactions, *i.e.*, the RNA affinity landscape for each small molecule, which informs ligand design and potential off-targets (Fig. 2).

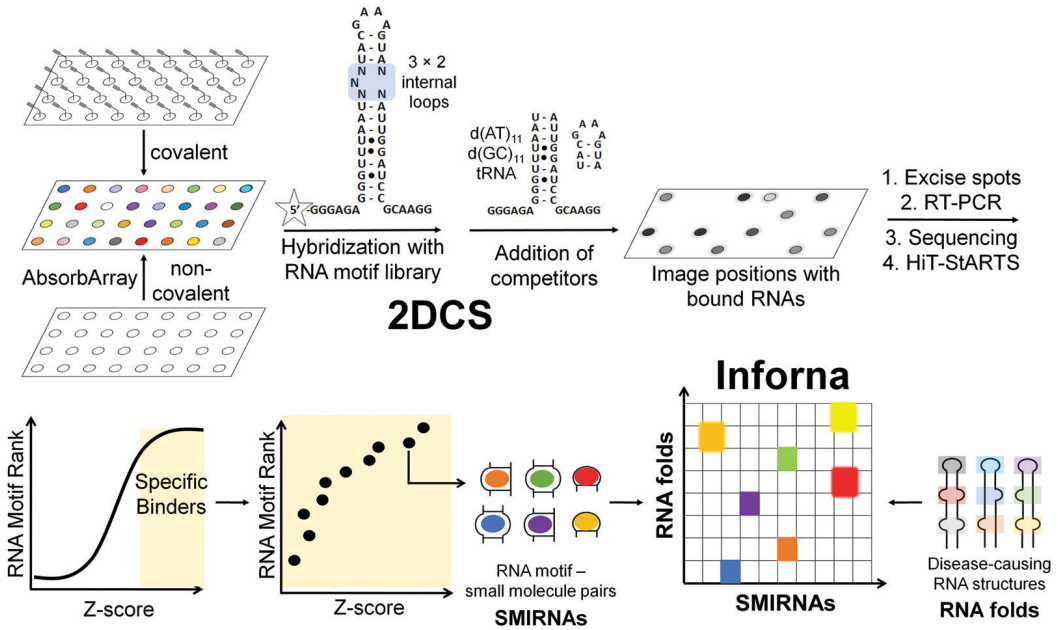


Fig. 2 Overview of two-dimensional combinatorial screening (2DCS) and Inforna. In 2DCS, a small molecule library is spatially arrayed onto a microarray, either through covalent attachment or absorption (AbsorbArray). Compounds are then incubated with a labeled RNA motif library, *e.g.*, 3 × 2 internal loops, containing randomized regions that form structured RNA motifs found in disease-causing RNAs. Unlabeled competitor oligonucleotides that mimic regions common to all library members, *r*(AU) and *r*(GC) base pairs, DNA oligonucleotides, and other RNAs are added to eliminate non-specific binding. Small molecules that bind RNAs are excised, amplified by RT-PCR, sequenced by RNA-seq, and analyzed by High Throughput Structure–Activity Relationships Through Sequencing (HiT-StARTS). HiT-StARTS calculates the statistical significance of the enrichment of an RNA in the selection, reported as a *Z*-score (Z_{obs}). Selective small molecule-RNA motif interactions generally exhibit $Z_{obs} > 8$. These small molecule-RNA motif interactions and their corresponding *Z*-scores comprise Inforna. Using Inforna, privileged SMIRNAs can be identified for functionally relevant RNA 3D folds within disease-causing RNAs, such as miRNAs. In addition to mining for SMIRNAs with favorable affinity landscapes for the RNA target of interest, Inforna can also predict potential off-target RNAs.

The third key factor to enable the rational design of SMIRNAs is a bioinformatic pipeline to link these privileged interactions to structured 3D folds found in evolutionarily conserved regions of cellular RNAs. Indeed, our lead identification strategy, Inforna (Table S1, ESI†),[20] is this pipeline and has enabled the design of many bioactive small molecules that target disease-causing RNAs, as described in Sections 7 and 8.

5. Other methods to identify small molecules that bind RNA

Various other approaches have been developed to identify small molecules that bind structured RNA motifs, particularly structure-based design. As a starting point, structure-based design uses NMR spectroscopy or X-ray crystallography to generate an ensemble of 3D structures for an RNA motif either in its free form or in complex with a small molecule. Then, virtual small molecule libraries can be docked into this ensemble and predicted hit molecules can be ranked according to the free energy of binding to the structured RNA motif. The accuracy of these predictions, however, must subsequently be validated *in vitro* using various biophysical techniques. Indeed, such combined approaches have been successfully implemented for viral RNAs including the hepatitis C virus (HCV) internal ribosomal entry site (IRES) and HIV transactivation response element (TAR) RNA.[21]

Besides structure-based design, a variety of other high-throughput screening methods have been employed to identify small molecules that bind structured RNA motifs. However, in many cases, such approaches focus on a single RNA target. That is, a library of small molecules is screened against a single structured RNA motif at a time, rather than the thousands of RNA motifs probed in a target agnostic fashion as in 2DCS. These target-centric methods include SMMs,[25] which have been used to identify ligands that bind the HIV TAR RNA, among others; fluorescent dye displacement assays or the use of a small molecule's intrinsic fluorescence, which was used to identify small molecules that bind the long noncoding RNA (lncRNA) metastasis associated lung adenocarcinoma transcript 1 (MALAT1);[22] or monitoring the change in fluorescence of RNAs containing fluorescent nucleosides[23] or end-labeled RNA constructs, which identified small molecule binders to a self-splicing group II intron.

Other emerging high-throughput screening methods for the identification of small molecules binding structured RNA motifs include automated ligand identification system (ALIS), which identifies RNA motif-small molecule binding partners through affinity-selection mass spectrometry (AS-MS), pattern recognition,[24] SMM,[25] and catalytic enzyme-linked click chemistry assay (cat-ELCCA), which can be used to screen for small molecule inhibitors of miRNA processing *in vitro* through the use of a system that amplifies chemiluminescence if processing

is inhibited.[26] Rational design and a variety of other screening methods have also been utilized to identify small molecules that bind RNA repeat expansions.[27–29] Extensive reviews of these methods and the small molecules they identified can be found in the following ref. 30–32.

6. Target validation methods

Two challenges in identifying high-quality SMIRNAs is confirming target engagement and quantifying selectivity for the desired target relative to other RNAs featuring identical, similar, or disparate structured RNA motifs.[1] Therefore, robust target validation approaches are of key importance to: (i) confirm that phenotype modulation is a direct cause of the RNA-centric mechanism of action of the SMIRNA, *i.e.*, confirming cellular occupancy of the RNA target; and (ii) broadly profile the selectivity of SMIRNAs in a transcriptome-wide manner. Indeed, various target engagement methods have been developed and validated to assess RNA target occupancy *in vitro* and in cells including cross-linking and cleavage-based approaches as well as competition experiments between a SMIRNA and an ASO.[1] These methods, discussed in detail below, are imperative to implement as high-quality SMIRNAs are being developed for current, as well as emerging, RNA targets.

Chemical Cross-linking and Isolation by Pull-Down (Chem-CLIP) is a target validation method in which a SMIRNA is appended with nucleic acid cross-linking (*e.g.*, chlorambucil, diazirine) and purification (*e.g.*, biotin) modules at positions that do not affect molecular recognition (Fig. 3).[1,33] In cells, the Chem-CLIP probe undergoes a proximity-induced cross-linking reaction upon binding a structured RNA motif. Total RNA is extracted and cross-linked RNAs are isolated and purified by using the purification module, enriching them in the pulled-down fraction. The RNA targets of the Chem-CLIP probe are then identified *via* RNA-seq or quantitative (q)RT-PCR (Fig. 3). This method can also be used in a competitive fashion (C-Chem-CLIP) to confirm the target occupancy of an unmodified SMIRNA.[1] That is, in C-Chem-CLIP, the SMIRNA competes for binding to the same RNA target as the Chem-CLIP probe, which prevents crosslinking and therefore decreases enrichment of the RNA target. Additionally, the Chem-CLIP probe can be used to map binding sites of SMIRNAs in cells *via* Chem-CLIP-Map-Seq (Fig. 3).[1,33] Here, after cross-linking, the bound RNAs isolated from cells are reverse transcribed, PCR amplified, and sequenced. The binding sites of SMIRNAs on RNA targets can then be identified by deconvolution of RT ''stops'', which are proximal to the cross-linking sites.

Complementary to Chem-CLIP is the cleavage-based approach named small-molecule nucleic acid profiling by cleavage applied to RNA (RiboSNAP; Fig. 3), which has been used to confirm target engagement, map binding sites, and profile off-targets of SMIRNAs *in vitro* and in cells.[1,33] In RiboSNAP, a SMIRNA is appended to a nucleic acid cleaving module, such as bleomycin A5,[34] at a position that does not contribute to the binding of the SMIRNA to the target (Fig. 3). Attachment of bleomycin A5 *via* its primary amino group has been shown to eliminate off-target DNA cleavage upon amide bond formation.[1] Thus, the bleomycin-SMIRNA conjugate selectively cleaves sequences proximal to the structured RNA motifs engaged by the SMIRNA. Cellular targets of SMIRNAs are then identified through RNA-seq or RT-qPCR, where the abundance of targeted RNAs are reduced as a result of the RiboSNAP probe. Similarly to C-Chem-CLIP, the competitive version of RiboSNAP, coined C-RiboSNAP, can also be employed to study the parent compound (Fig. 3). SMIRNAs that compete with the RiboSNAP probe for the same RNA binding site will reduce the amount of cleavage.[1] Cellular mapping of binding sites can also be accomplished with RiboSNAP probes, or RiboSNAP-Map, using RNA target-specific RT primers to identify the cleavage site.[1]

Although both Chem-CLIP and RiboSNAP have been robustly applied to validate engagement of SMIRNAs with various RNA targets, both require chemical functionalization of the SMIRNA, which can involve laborious, multi-step

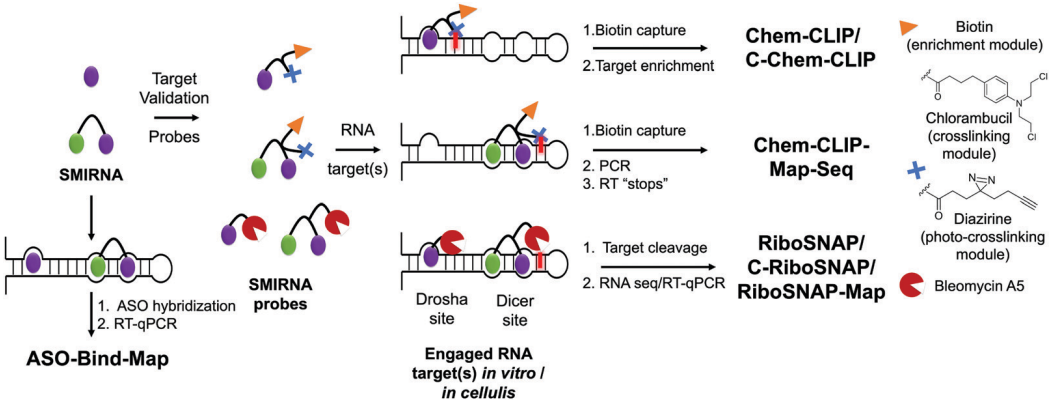


Fig. 3 Methods to validate the targets of SMIRNAs, to study cellular selectivity, and to map SMIRNA binding sites within an RNA target. Schematics of target validation techniques for SMIRNAs. In ASO-Bind-Map, unmodified SMIRNAs are used to prevent hybridization of complementary ASOs, thus preventing cleavage. In Chemical Cross-Linking and Isolation by Pull-Down (Chem-CLIP) and related methods (competitive Chem-CLIP (C-Chem-CLIP) and Chem-CLIP-Map-Seq), SMIRNAs are functionalized with cross-linking (chlorambucil or diazirine) modules and a purification module (biotin) at positions that do not affect binding to the intended RNA target. In small molecule nucleic acid profiling by cleavage applied to RNA (RiboSNAP) and its competitive variant, the SMIRNA is appended with the natural product bleomycin A5.

synthetic procedures. Therefore, the development of label free target validation methods that avoid chemical derivatization of SMIRNAs are highly desirable. As an example, ASO-Bind-Map[18] exploits the endogenous activity of RNase H to cleave RNA–DNA heteroduplexes instead of derivatizing the SMIRNA (Fig. 3). To validate target engagement and map the binding site of a SMIRNA using ASO-Bind-Map, ASOs are designed to span the target RNA binding site such that upon RNA–DNA heteroduplex formation, RNase H efficiently cleaves the RNA target. If binding of a SMIRNA, however, thermally stabilizes the RNA binding site or triggers a conformational change that hinders the hybridization of an ASO, cleavage will be inhibited, which can be read out using RT-qPCR or RNA-seq (Fig. 3). ASO-Bind-Map is advantageous over other reagents that are used to map RNA structure and determine binding sites, such as DMS and SHAPE, which require highly resident small molecule interactions that may not be able to inhibit an irreversible reaction with the chemical modifier. Additionally, the sites that react with mapping reagents may not overlap with small molecule binding sites. Collectively, ASO-Bind-Map can confirm the binding site(s) and selectivity of SMIRNAs, both *in vitro* and in cells. However, unlike Chem-CLIP and RiboSNAP, this method is not target agnostic and cannot be applied across the transcriptome.

Collectively, the target validation methods presented in this section offer unparalleled accessibility to assess RNA target occupancy, profile off-targets, and map binding sites of SMIRNAs *in vitro* and in cells. Application of these methods early in the development of SMIRNAs is key to developing high-quality chemical probes that modulate disease biology with a defined, RNA-centric mode of action.

7. Targeting disease-causing RNAs with SMIRNAs, enabled by Inforna

Structured RNAs have long been linked to disease,[5] making them ideal targets for novel SMIRNAs. For example, dysregulation of miRNA expression has been linked to cancers of the lung, prostate, and breast, cardiovascular disease, inflammatory disorders, and liver disease.[35] Additionally, neurotoxic proteins such as SNCA[14] and tau[13] are encoded by pre-mRNAs featuring unique 3D structured RNA motifs, further substantiating the therapeutic potential of targeting disease-causing RNAs with small molecules. Our lead identification strategy, Inforna,[20] can be utilized to assess the ligandability of these disease-relevant RNA 3D folds and rapidly identify privileged SMIRNAs that target these structures and affect disease biology.

7.1 The RNA structurome of human miRNA precursors

As mentioned above, miRNAs regulate a myriad of biological processes and their dysregulation triggers a wide variety of human diseases.[35] Thus, they are an important class of emergent therapeutic targets. Fortuitously, miRNA precursors fold into accurately predicted structures, forming well-defined structured 3D folds that can be recognized by small molecules. Indeed, blocking miRNA processing sites could directly inhibit miRNA biogenesis, *i.e.*, reduce mature miRNA levels, and consequently deactivate signaling pathways modulated by mature miRNAs.

Liu *et al.*,[36] cataloged all structured motifs formed by human miRNA precursor hairpins in an effort to enable lead design by Inforna (Table S1, ESI†). Over 7000 motifs were cataloged, among which small loops, such as 1-nucleotide bulges and 1×1 internal loops, were highly represented. These bulges and loops featured various closing base pairs, increasing the overall diversity of structured RNA motifs within the miRNome and hence the ensemble of 3D folds amenable to SMIRNA targeting.

Further, 752 unique functional RNA motifs within Dicer (n = 451) and Drosha (n = 301) processing sites were reported. Among these, only 10 were identified in other highly expressed human RNAs (potential off-targets), rendering the remaining motifs highly valuable as SMIRNA binding sites. That is, there are a plethora of well-defined structured RNA motifs present within the Drosha and Dicer processing sites of miRNAs that could be selectively targeted with SMIRNAs. Access to this database of motifs present within human miRNA hairpin precursors is accessible upon request (Table S1, ESI†).

7.1.1 Small molecules that recognize the 3D fold of oncogenic pri-miR-96. In the inaugural study to validate Inforna as a lead identification strategy, we compared the structured 3D folds in all human miRNA hairpin precursors to our database of privileged RNA fold-small molecule interactions.[20] The hits were further refined by disease-association and requiring the small molecule to occupy the Drosha or Dicer processing site. We studied each potential interaction in cells by measuring reduction of mature miRNA levels, affording a hit rate of 44%.

The optimal interaction from this query, as defined by inspection of affinity landscapes, was between the Drosha site of pri-miR-96, 5′UU̲U/3′AU̲A (1×1 UU internal loop), and monomeric compound **96-SM1** (Fig. 4A). We therefore studied the effects of **96-SM1** in more detail, confirming compound mode of action (inhibition of Drosha processing), de-repression of the downstream pro-apoptotic transcription factor Forkhead box protein O1 (FOXO1), and induction of apoptosis. Importantly, knock down of FOXO1 by an siRNA reduced **96-SM1**'s activity, providing further evidence that the observed rescue of phenotype is through the miR-96-FOXO1 circuit. Additional *in cellulis* selectivity studies *via* miRNA profiling by RT-qPCR of detectable miRNAs showed that **96-SM1** significantly affected only miR-96 levels and was as selective as an ASO antagomiR.

Although **96-SM1** inhibited miR-96 levels in cells, its cellular potency (IC_{50} of $\sim$20 μM) was not sufficient for *in vivo* studies. Numerous examples, including this study, have shown that covalently linking monomeric units targeting adjacent structured RNA motifs increases binding affinity and potency.[1] We therefore used Inforna to identify SMIRNAs that engage motifs adjacent to pri-miR-96's Drosha site. This search yielded a small molecule binder **96-SM2** (Fig. 4A) of a nearby 5′CG̲A/3′UG̲G (1×1 GG) internal loop.[37] Linking **96-SM1** and **96-SM2** *via* a peptoid linker afforded dimeric compound **Targaprimir-96** (in which ''Targa'' indicates targets and ''primiR-96''

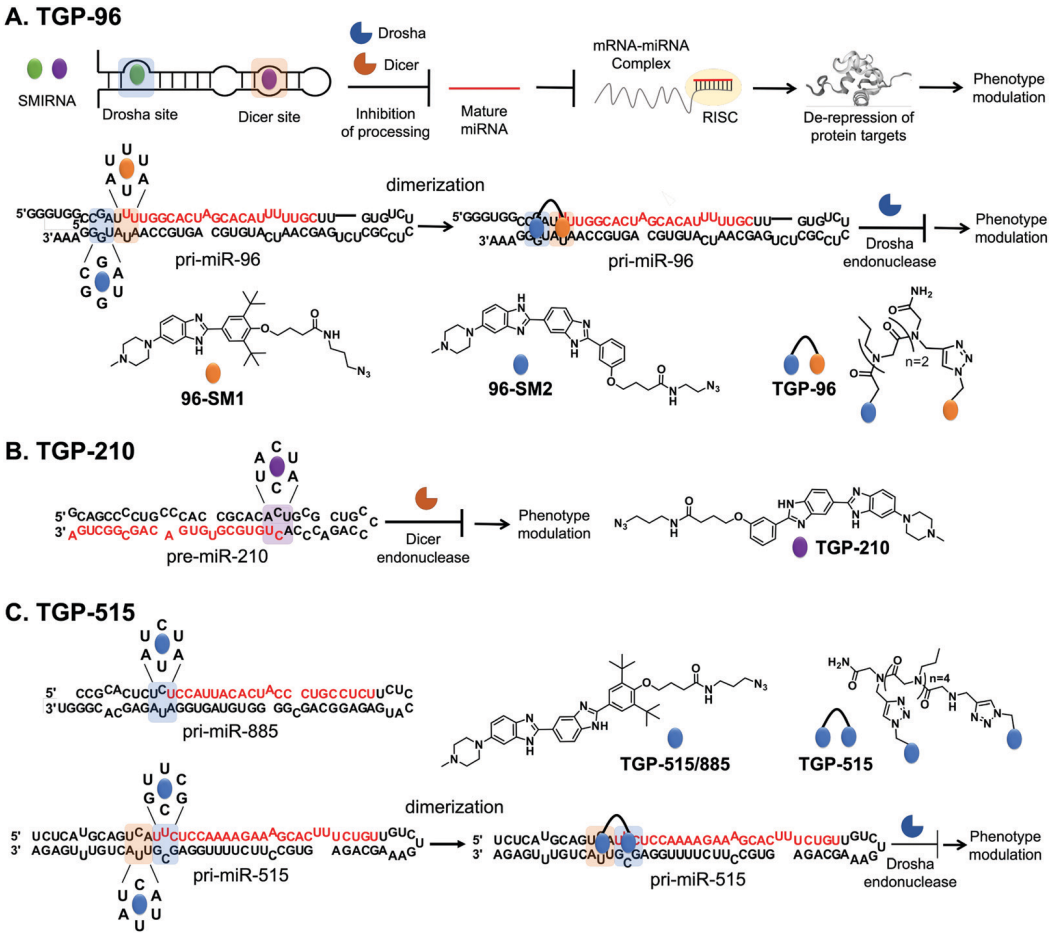


Fig. 4 Using Inforna to identify SMIRNAs targeting disease-causing miRNAs. (A) Schematic representation of miRNA biogenesis, where SMIRNA binding can inhibit processing by binding to either Drosha or Dicer sites and thereby reduce the levels of the mature miRNA. Reduction of mature miRNA levels results in decreased translational inhibition of target mRNAs by the RNA-induced silencing complex (RISC). Thus, SMIRNA inhibition of miRNA biogenesis derepresses the miRNA's protein targets, resulting in phenotype modulation. Structure of pri-miR-96 and chemical structures of monomeric compounds **96-SM1** and **96-SM2** that target 1 × 1 GG and UU internal loops (blue and orange, respectively) in the Drosha processing site. Covalent attachment of **96-SM1** to **96-SM2** *via* a peptoid linker yields dimeric compound **TGP-96**, a more potent and selective SMIRNA compared to the monomeric units. Indeed, **TGP-96** decreases tumor burden in a mouse xenograft model. (B) Secondary structure of pre-miR-210 and chemical structure of **TGP-210**, which targets a 1 × 1 CC internal loop in the Dicer processing site (highlighted in purple). (C) Secondary structure of pri-miR-885 and pri-miR-515, with the Drosha processing sites highlighted in blue and the adjacent 5′UCA/3′AUU motif present in pri-miR-515 highlighted in orange. The chemical structures of monomeric **TGP-515/885** and dimeric compound **TGP-515** are also shown. **TGP-515** is an example of a potent and selective SMIRNA, generated by simultaneously targeting two 1 × 1 CU internal loops near the Drosha processing site.

indicates pri-miR-96; **TGP-96**) (Fig. 4A). Notably, the optimal length of the peptoid linker was experimentally determined to mimic the precise distance between the Drosha site and the 1 × 1 GG internal loop.[37] Indeed, **TGP-96** bound ∼40-fold more tightly to pri-miR-96 than **96-SM1** and ∼30-fold more avidly than **96-SM2**. In a triple negative breast cancer (TNBC) cell line, MDA-MB-231, **TGP-96** decreased mature miR-96 levels and increased levels of pri-miR-96, as a result of inhibiting Drosha processing at a dose of 50 nM.[37] As expected, TGP-96 also boosted levels of FOXO1 and triggered apoptosis, but at an 800-fold lower concentration. Importantly, in this study direct target engagement of pri-miR-96 by **TGP-96** in cells was demonstrated using both Chem-CLIP and C-Chem-CLIP. The **TGP-96** Chem-CLIP probe was used in a follow-up study to map the exact binding site of **TGP-96** within pri-miR-96, the Drosha binding site, which was further validated by RiboSNAP-Map.

Fortuitously, **TGP-96** has a favorable drug metabolism and pharmacokinetic profile. *In vivo* studies using NOD/SCID mice injected with MDA-MB-231 cells to form breast tumors showed that **TGP-96** (10 mg kg^{-1}) reduced tumor growth by inhibiting miR-96 biogenesis and increasing FOXO1. Collectively, these studies validated Inforna as a lead identification strategy, utilizing primary RNA sequence to mine small molecules targeting structured 3D folds within disease-causing miRNA. This approach allows for the subsequent modular assembly of identified small molecules to improve the potency and selectivity of SMIRNAs. Ultimately, Inforna provides the means of directly connecting structured 3D folds with privileged small molecule interactions. Moreover, Inforna's SMIRNA predictions readily translate into biological activity in disease-relevant cell lines as a result of the RNA-centric mode of action.

7.1.2 A small molecule that recognizes the 3D fold of oncogenic pre-miR-210. One clinical feature of difficult to treat and aggressive cancers is hypoxia, a reduction in normal levels of tissue oxygenation. Tumors with hypoxia exhibit increased resistance to radiation and chemotherapy and are associated with increased invasion. Thus, modulation of hypoxia-associated pathways is an important therapeutic target. MiR-210 is key to adaption to a low oxygen environment, and its expression under hypoxia is upregulated by hypoxia inducible factors (HIFs), as has been demonstrated in oxygen depleted solid tumors. At the molecular level, miR-210 represses glycerol-3-phosphate dehydrogenase 1-like (GPDL1), which leads to suppression of prolyl hydroxylase (PHD).[38] Under normoxic conditions, PHD hydroxylates proline residues in HIF-1α, preventing its interaction with HIF-1β, therefore blocking the formation of the HIF-1α/HIF-1β heterodimer, which functions as a turn-on switch for genes that contribute to metastasis. Thus, suppression of PHD due to elevated levels of miR-210 enables adaptation and metastasis of cancer cells.[38] In MDA-MB-231 TNBC cells, miR-210 levels are upregulated significantly under hypoxic conditions compared to normoxic conditions.

Inforna identified a SMIRNA, **Targapremir-210** (**TGP-210**; $K_d \sim 200$ nM), that targets the Dicer processing site of pre-miR-210, which features a 5′AC̲U/3′AC̲U (1 × 1 CC) internal loop (Fig. 4B).[39] **TGP-210** inhibited pre-miR-210 processing by Dicer *in vitro* and in MDA-MB-231 TNBC cells ($IC_{50} \sim 200$ nM), as demonstrated by decreased levels of mature miR-210 and increased levels of pre-miR-210 and upon compound treatment.[39] As a result of inhibiting miR-210 biogenesis, levels of *GPDL1* mRNA were increased, *HIF-1α* mRNA levels were decreased, and apoptosis was triggered selectively in hypoxic MDA-MB-231 cells.[38] That is, **TGP-210** modulated the hypoxic miR-210-HIF-1α axis *via* GPDL1. Microarray analysis of all human miRNAs revealed that **TGP-210** was selective, similar to a miR-210-targeted antagomiR. Chem-CLIP and C-Chem-CLIP studies showed direct target engagement of both the **TGP-210** Chem-CLIP probe and **TGP-210** itself.[39] In particular, the Chem-CLIP probe selectively enriched miR-210, and this enrichment was depleted by addition of **TGP-210**. As a further measure of selectivity, the enrichment of other miRNAs that have motifs recognized by **TGP-210** as predicted by Inforna, or RNA isoforms, was also measured. Of these 15 RNA isoforms, only miR-497 contained the same 1 × 1 CC internal loop as miR-210, while the other 14 isoforms featured motifs with weaker affinity for **TGP-210**. Of these 15 miRNAs, the **TGP-210**-Chem-CLIP-probe only enriched four, including miR-497; however, they were enriched to a lesser extent than miR-210 as they bind **TGP-210** less avidly or were expressed less abundantly.[39] Importantly, **TGP-210** did not inhibit the biogenesis of these enriched miRNAs despite engaging them in cells because binding did not occur in a functional, *i.e.*, Dicer or Drosha processing, site and/or these miRNAs were less abundant and contained weaker affinity motifs. Further, **TGP-210** treatment decreased tumor burden *in vivo* using a NOD/SCID mouse model of hypoxic breast cancer.

Taken together, the study elucidated important insights into SMIRNAs targeting structured RNA motifs. For example, a SMIRNA must engage a functional RNA motif (Dicer site in the case of **TGP-210**; or Drosha site in the case of **TGP-96**) within the disease-causing miRNA, and selectivity can be obtained if the target miRNA is expressed at sufficiently higher levels than potential off-targets.

7.1.3 Small molecules that recognize the 3D fold of pri-miR-515, a miRNA with a pivotal role in cell signaling. Fortuitously for miR-210, potential off-target liabilities were ameliorated because their small molecule binding sites occurred outside of processing sites, however, this is unlikely to be the case for other RNA targeting endeavors, which begs the question of how to selectively target one RNA over another if they harbor the same motif in a functional site. A case study of pri-miR-515 and pri-miR-885 sought to provide a general solution to this problem. The two miRNAs have similar loops in their Drosha binding sites, 5′UC̲A/3′AUU (miR-515) and 5′UC̲U/3′AU̲A (miR-885), that bind with similar affinity to a small molecule identified by Inforna, **Targaprimir-515/885** (**TGP-515/885**) (Fig. 4C). Further, the processing of both is inhibited to a similar extent in MCF-7 cells.[20]

In order to selectively target miR-515 over miR-885, Costales *et al.*,[40] employed a modular approach to exploit the differences in the two miRNAs' 3D folds. In particular, pri-miR-515 features an adjacent 5′UU̲C/3′GC̲G loop not present in pri-miR-885 (Fig. 4C). We therefore used Inforna to identify a small molecule lead for this loop. Tethering the two RNA-binding modules *via* a linker of precise length afforded **Targaprimir-515** (**TGP-515**) (Fig. 4C). As compared to **TGP-515/885**, **TGP-515** was ~250-fold more avid and >3200-fold more selective *in vitro*, validating the modular assembly strategy to bolster binding affinity and selectivity.[40] Interestingly, **TGP-515** did not bind an RNA with only a singular binding site. This effect can be traced in part to **TGP-515**'s self-structure, acting as a stringency clamp. The increased avidity and selectivity observed *in vitro* translated *in cellulis*, where **TGP-515** inhibited biogenesis of miR-515, reducing mature levels and boosting pri-miRNA levels, while not affecting miR-885.[40] This selectivity was widespread across the miRNome, as determined by RT-qPCR profiling of all miRNAs detectable in MCF-7 cells.[40]

A key downstream target of miR-515 is sphingosine kinase 1 (SK1) protein that synthesizes sphingosine 1-phosphate (S1P), a second messenger involved in migration. As expected, inhibition of pri-miR-515 by **TGP-515** increased levels of both SK1 and S1P. Further, the compound's effect was reduced by both an siRNA directed at *SK1* mRNA and a small molecule inhibitor of SK1, validating the compound's mode of action. A proteome-wide study upon **TGP-515** treatment revealed that human epidermal growth factor receptor 2 (HER2) was significantly upregulated. Interestingly, MCF-7 cells are HER2-negative, and these results suggest that they may be sensitized to treatment with anti-HER2 precision medicines. Indeed, **TGP-515** sensitized MCF-7 cells to Herceptin. In conclusion, this study provided a general strategy to lead optimize a dual-targeting SMIRNA into a single-target, selective compound.

7.2 Targeting the IRE within the 5′ UTR of *SNCA* mRNA with Synucleozid

7.2.1 Prediction of RNA structural motifs within *SNCA* mRNA that encodes an IDP. SNCA, or α-synuclein, is a critical component in the pathogenesis of Parkinson's disease and belongs to a class of genes defined as producing intrinsically disordered proteins (IDPs), meaning the proteins do not form well-defined tertiary structures. Therefore, SNCA, as well as other IDPs, do not feature defined pockets that can accommodate small molecules. The RNAs encoding these proteins, however, could contain structured RNA motifs more amenable to small molecule targeting, therefore providing a viable therapeutic alternative. Indeed, the *SNCA* mRNA contains various structured RNA motifs within its 5′ UTR, known as an iron-responsive element (IRE), whose secondary structure was found to be targetable by a small molecule, thus altering protein translation.[14]

ScanFold was used by Zhang *et al.*,[14] to define the structured motif landscape of all human mRNAs encoding IDPs, including *SNCA*. In this case, ScanFold results were used to determine if these mRNAs were particularly enriched for unusually stable structures (leading to lower average *z*-scores across the entire mRNA sequence). While IDP-encoding mRNAs overall did not appear to be any more enriched with unusually stable structures than the average mRNA, for each IDP mRNA that was scanned, there was at least one region which contained well-defined, structured RNA motifs. The important finding of ScanFold's results was that structure-less IDPs are produced from intrinsically structured mRNAs, opening up new therapeutic modalities for diseases caused by IDPs. In the *SNCA* mRNA, for example, 36% of its 3,167 nt contribute to structures that generate significantly low *z*-scores. These nts are organized into many new structured motifs, beyond the known IRE structure that was recently targeted by Zhang *et al.*[14]

7.2.2 Targeting the *SNCA* IRE to selectively inhibit translation. SNCA is an IDP involved in Parkinson's disease (PD) that contributes to neurotoxicity by accumulating in Lewy bodies. Thus, lowering SNCA protein levels, by preventing its translation, could mitigate neurotoxicity in PD. Since SNCA protein is difficult to target, an alternative method to reduce protein levels is to target *SNCA* mRNA and inhibit translation. Fortuitously, the 5′ UTR of *SNCA* mRNA contains a structured IRE that directly modulates protein production as a function of iron levels *via* iron regulatory protein (IRP). Using Inforna, a small molecule named **Synucleozid** (Fig. 5) was identified to bind the 5′GG/3′CAU A bulge in the IRE, along with other SMIRNAs.[14] Only **Synucleozid** reduced SNCA protein levels in cells without affecting *SNCA* mRNA expression, and this reduction conferred cytoprotection against cell death caused by aggregation of preformed α-synuclein fibrils. Furthermore, selective inhibition of translation was observed as the compound did not affect the translation of other mRNA sequences featuring IREs with different structures in their 5′ UTRs, such as amyloid precursor protein (APP) and prion protein (PrP).

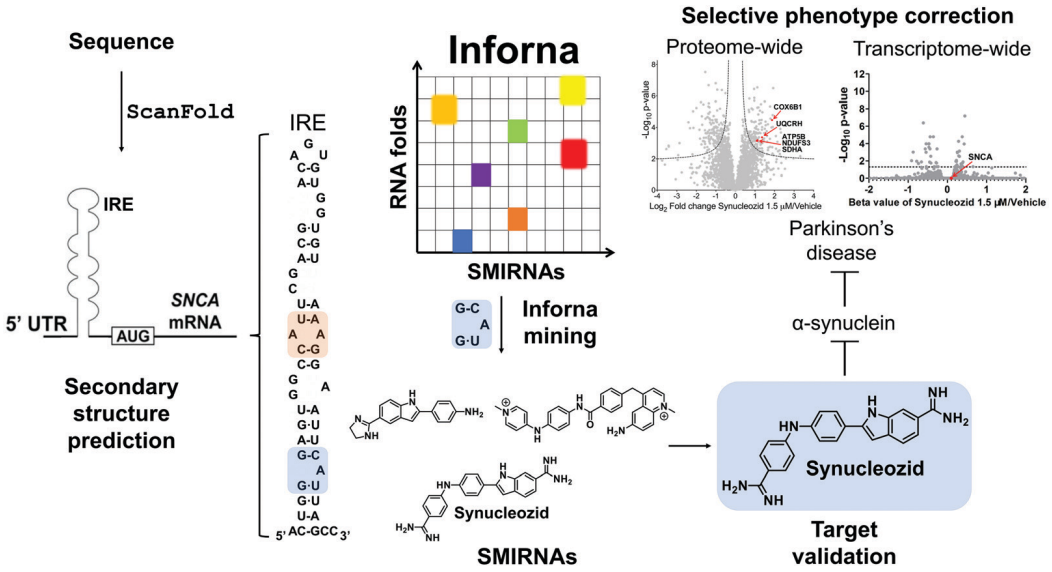


Fig. 5 Mining Inforna to identify **Synucleozid**, which targets the iron responsive element (IRE) within α-synuclein's (*SNCA*) mRNA and inhibits translation *in cellulis*. The 5′ UTR of *SNCA* mRNA sequence contains ligandable structured RNA motifs within the IRE (highlighted in orange and blue). Mining Inforna for small molecules targeting these RNA motifs yielded potential candidates, the most potent of which named **Synucleozid** binds to the 5′G_G/3′CAU A-bulge of the IRE. Among the 3300 proteins detectable in the proteome-wide analysis, only ~8% were significantly down- or upregulated (p-value < 0.01) upon treatment with **Synucleozid** (1.5 μM). Various proteins involved in the oxidative phosphorylation pathway, such as the mitochondrial ATP synthase subunit beta (ATP5B), NADH dehydrogenase [ubiquinone] iron–sulfur protein 3 (NDUFS3), cytochrome *c* oxidase subunit 6B1 (COX6B1), succinate dehydrogenase [ubiquinone] flavoprotein subunit (SDHA), and cytochrome b-*c*1 complex subunit 6 (UQCRH), were down-regulated upon **Synucleozid** treatment. RNA-sequencing (RNA-seq) analysis revealed limited off-target effects transcriptome-wide (99.7% of the differentially expressed genes were unchanged) following treatment with **Synucleozid**. Note: **Synucleozid** has no effect on *SNCA* RNA levels as its mode of action is binding the RNA and inhibiting its translation.

Target engagement was demonstrated and the exact binding site of **Synucleozid** was defined both *in vitro* and in cells using ASO-Bind-Map.[14] Careful design of ASOs spanning *SNCA's* IRE confirmed that **Synucleozid** targets the 5′GG/3′CAU structural motif both *in vitro* and in cells. Optical melting experiments showed that **Synucleozid** thermally stabilizes the IRE. Cellular mechanistic studies demonstrated that **Synucleozid** selectively inhibited *SNCA's* translation *via* this stabilization, which alters ribosomal loading. Furthermore, proteome- and transcriptome-wide studies showed that **Synucleozid** exhibited favorable selectivity at both the protein and RNA levels (Fig. 5).

Importantly, transcriptome-wide analysis of mRNAs that encode IDPs revealed that each has structured RNA motifs that could be targeted with small molecules.[14] Collectively, these studies demonstrate the potential for targeting proteins with poorly defined tertiary structure at the level of their structured coding mRNAs.

7.3 Targeting *MAPT* pre-mRNA with SMIRNAs

7.3.1 Prediction and validation of structured RNA motifs within tau's pre-mRNA. Tau protein, encoded by *MAPT*, stimulates microtubule assembly and stability, with different isoforms playing roles in cytoskeletal plasticity and stability. Differential expression of tau isoforms in the nervous system is involved in the establishment and maintenance of neuronal polarity. Not surprisingly, dysregulation of tau levels, as a result of mutations, leads to several neurodegenerative disorders, collectively termed tauopathies, including Alzheimer's and Parkinson's diseases. They are marked by the aberrant deposition of protein into tau inclusion bodies leading to deleterious phenotypes such as neurodegeneration. As there are currently no effective therapeutics for tauopathies, likely due to the fact that tau is an IDP, the selective reduction of tau levels provides a viable therapeutic option.

Chen *et al.*,[13] applied ScanFold to tau's pre-mRNA sequence to explore the existence of structured RNA motifs that may be functionally relevant, and potentially targetable with SMIRNAs (Fig. 6A). Novel structured RNA motifs were discovered, especially at exon–intron junctions and within the 5′ and 3′ UTRs. Twenty structured RNA regions were predicted at the exon–intron junctions. The 5′ UTR contained a single predicted region that overlaps a known IRES, while the 3′ UTR contained eight structured regions. Additional analyses of these structured RNA motifs *via* luciferase reporters showed their ability to affect stability and splicing of the tau pre-mRNA. In conclusion, ScanFold successfully identified previously validated structured RNA motifs within tau's pre-mRNA and predicted additional motifs that could be targeted with SMIRNAs.

7.3.2 Targeting the tau exon 10-intron junction. RNA structures at exon–intron junctions can direct the alternative splicing of the *MAPT* (tau) gene. For example, a mutation at the exon 10-intron junction (+14C > U) causes frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17) by destabilizing the RNA's structure (Fig. 6B).[41] This destabilization enables increased binding of U1 small nuclear RNA (snRNA) and increases exon 10 inclusion (Fig. 6B). This inclusion leads to overproduction of an mRNA encoding four microtubule domains, or 4R tau, which is aggregation-prone and contributes to the progression of neurodegenerative diseases (Fig. 6B). Thus, one therapeutic strategy is to stabilize the RNA structure at the exon–intron junction with a SMIRNA by targeting a structured RNA motif. Although various studies identified small molecules that indeed stabilize the junction, *in vitro* binding did not translate to rescue of aberrant tau splicing observed in FTDP-17.[42]

Recently, drug-like small molecules were identified that bind an A bulge, 5′CAG/3′CG, present in the exon 10-intron junction, that rescued endogenous tau splicing in the human neuroblastoma cell line Lan5 and in primary neurons from an hTau transgenic mouse model (Fig. 6B).[43] These small molecules were designed from a previously Inforna-derived compound and by analysis of chemotypes that confer RNA-binding capacity as determined from the Inforna database.[43]

Particularly, these studies were initiated by searching for chemically similar small molecules related to the substituted 2-phenyl-1*H*-indole-derived compound discovered *via* Inforna. We were able to determine the structure of a potent compound, **SMIRNA1**, that bound to the exon 10-intron junction and reduced exon 10 inclusion in a cell-based reporter of exon 10 splicing (Fig. 5B). The free and bound RNA structures revealed that the A bulge was dynamic, and its conformation changed upon **SMIRNA1** binding. These observations enabled a facile, high-throughput binding assay in which the A bulge was replaced with the nucleobase 2-aminopurine (2-AP), the fluorescence emission of which changes with its microenvironment, *i.e.*, stacked or unstacked in a helix. We used this assay as well as a cell-based reporter and docking to identify three new scaffolds from chemical libraries.

As **SMIRNA1** was unlikely to be blood–brain barrier (BBB) penetrant, two different hit expansion strategies were employed to identify potent SMIRNAs with favorable physiochemical properties for BBB penetrance, as determined from Central Nervous System Multiparameter Optimization (CNS-MPO) scores.[44] CNS-MPO scores quantify favorable physicochemical properties for BBB penetrance, each on a scale from 0–1. These properties include: lipophilicity (clog *P*), distribution coefficient at pH 7.4 (clog *D*), molecular weight (MW), topological polar surface area (TPSA), number of hydrogen bond donors (HBD), and pK_a values. The scores for each parameter are then summed; a CNS-MPO score $\geq$4.0 is considered promising for BBB penetrance.[44] Applying this CNS-MPO score criterion early in the lead identification and optimization process increases chances of success for developing CNS clinical candidates.

In one method, a pharmacophore model was generated from **SMIRNA1** and chemically similar compounds that rescued splicing in a cellular model. In the second hit expansion method, >500 analogs of the three new scaffolds were studied, selected based on their structural similarity and CNS-MPO scores. Of these, **SMIRNA2** (Fig. 6B) was the most optimal with enhanced cellular potency and improved physiochemical properties. Indeed, **SMIRNA2** rescued aberrant endogenous exon 10 splicing in Lan5 cells and in primary neurons from an hTau transgenic

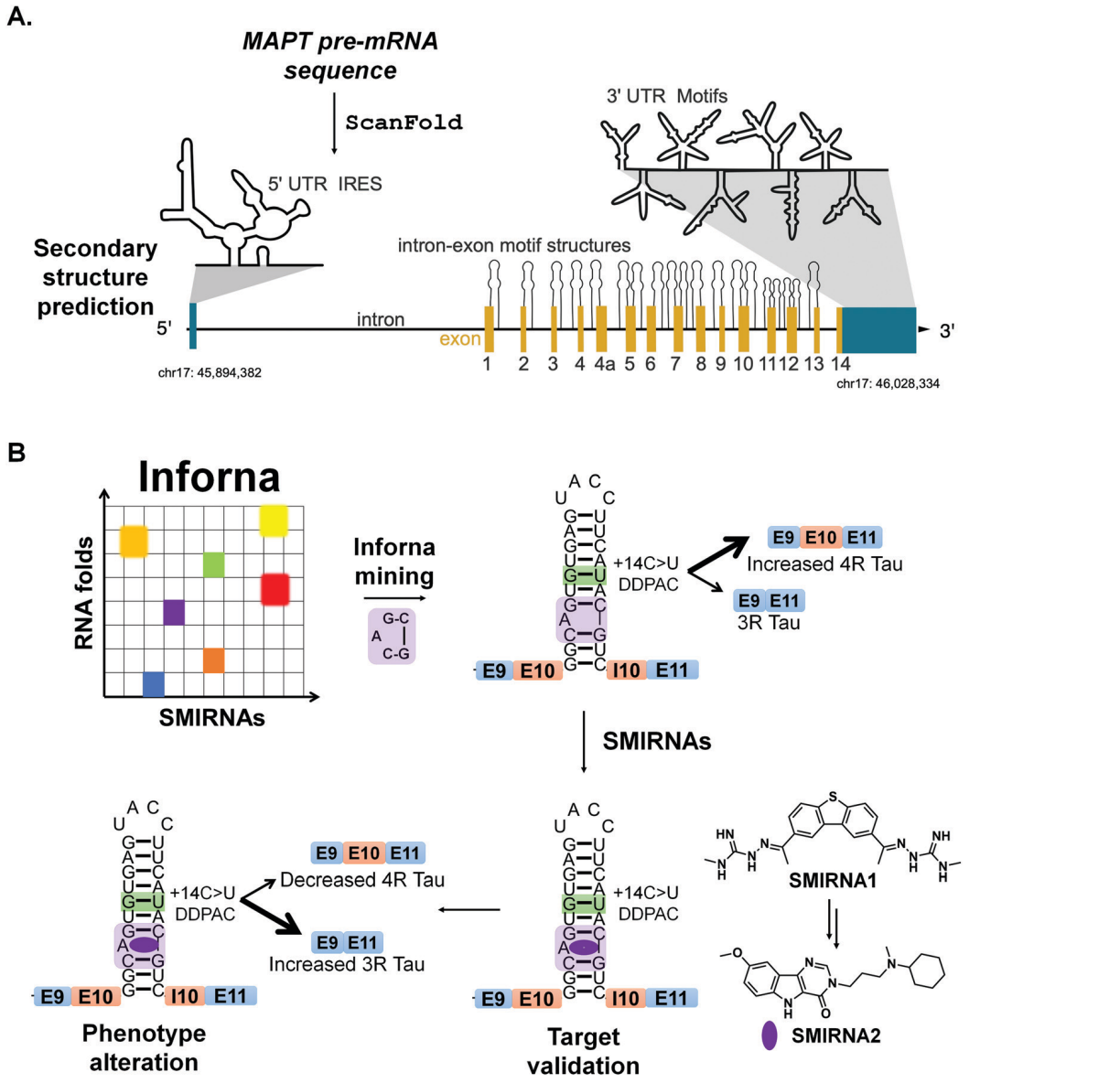


Fig. 6 RNA structure prediction and design of SMIRNAs that target structured RNA motifs within tau's pre-mRNA. (A) Secondary structure prediction *via* Scanfold of microtubule associated protein tau's (*MAPT*) pre-mRNA sequence. The *MAPT* pre-mRNA is depicted with 5′ and 3′ UTRs (blue regions), introns (solid, black lines), and exons (yellow regions), along with its chromosomal location (chr17: 45,894,382-46,028,334). The 5′ UTR contains a single, large, structured region that encompasses a known internal ribosome entry site (IRES). ScanFold predicted structured RNA motifs, depicted as hairpins, at exon–intron junctions throughout the *MAPT* pre-mRNA. These structures are expected to affect which regions are effectively spliced out of the final mRNA product. In the 3′ UTR, eight structured regions were predicted and presumed to confer regulatory effects on mRNA processing. (B) A mutation in *MAPT* exon 10 (+14C > U, green box around GU pair) destabilizes a splicing regulatory element (SRE) at the exon 10-intron junction, resulting in increased inclusion of exon 10 and increased production of 4R tau. This form of tau is prone to aggregation, triggering neurotoxicity. Chemical similarity searching identified **SMIRNA1** that binds the A bulge of the exon 10-intron hairpin (highlighted in purple). Further optimization of **SMIRNA1** yielded compound **SMIRNA2** with improved properties. Both compounds stabilize the SRE's RNA structure at the exon 10-intron junction, consequently increasing production of 3R tau and reducing production of the aggregation-prone 4R form.

mouse model. Importantly, target engagement studies of **SMIRNA2** *via* Chem-CLIP demonstrated that it directly and selectively engaged tau pre-mRNA, as RNAs containing other bulge motifs, such as mRNAs with IREs that regulate translation and miRNAs with the same A bulge, were not enriched. Thus, Inforna can be integrated with traditional medicinal chemistry strategies for the facile lead optimization of drug-like SMIRNAs with improved physiochemical properties.

8. Emerging modalities for targeted degradation of disease-causing RNAs

The studies described above demonstrate the power of Inforna, in concert with computational methods to predict evolutionarily conserved and structured RNA motifs, to design small molecules that modulate RNA function, provided the SMIRNA binds to a functional site. There is ongoing interest, however, to

develop new strategies to target any RNA, particularly if a functional site has not yet been identified or validated. Here, we describe two such strategies that employ degradation and cleavage, rather than simple binding, of the RNA target: (i) nuclease recruitment; and (ii) direct cleavage by conjugation of bleomycin A5 to SMIRNAs.[1] These cleavage modes of action remove the requirement of the SMIRNA to occupy a functional site as they rid the cell of the RNA altogether. In many cases, discussed below, these small molecule degraders and cleavers are more potent and selective than the occupancy-driven compounds from which they are derived.[3]

8.1 Targeted degradation *via* recruitment of RNase L

Ribonuclease targeting chimeras (RIBOTACs) hijack the cell's endogenous machinery of quality control and degradation pathways to degrade RNA targets selectively.[3,45] RIBOTACs are bifunctional, *i.e.*, chimeric compounds, in which one component binds a structured RNA 3D fold and the other locally recruits endogenous RNase L to the RNA target.[3,45] In inaugural studies, the RNase L recruiting module was based on RNase L endogenous activator, 2′-5′poly(A),[45] but more recently has been replaced with a small molecule heterocycle.[3]

A RIBOTAC was recently developed to target oncogenic miR-21 in cells and *in vivo* (Fig. 7A). MiR-21 is overexpressed in various types of cancers, and its expression negatively correlates with survival rate in triple negative breast cancer. The RIBOTAC is built on **Targapremir-21** (**TGP-21**), a dimer that binds pre-miR-21's Dicer site and an adjacent U bulge simultaneously (Fig. 7A).[3] **TGP-21** bound pre-miR-21 with ~20-fold greater affinity than the monomer from which it was derived **21-SM** (K_d = 20 μM for **21-SM** and 1 μM for **TGP-21**). Treatment of MDA-MB-231 TNBC cells with **TGP-21** reduced mature miR-21 levels and did so selectively across the miRNome, as assessed by miRNA profiling.[3] Moreover, the expression levels of phosphatase and tensin homolog (PTEN) and programmed cell death protein 4 (PDCD4), downstream targets of miR-21, increased by ~50% upon **TGP-21** treatment, ultimately leading to reduced invasion of MDA-MB-231 cells.[3]

To increase potency, **TGP-021 RIBOTAC** was synthesized by conjugating **TGP-21** to a heterocyclic small molecule that recruits RNase L (Fig. 7A).[3] This RIBOTAC was more potent than **TGP-21** *in cellulis*, as assessed by three different metrics: the IC_{50} for reducing levels of mature miR-21 (IC_{50} ~ 0.05 μM for **TGP-21 RIBOTAC** *vs.* 1 μM for **TGP-21**),[3] boosting PTEN and PDCD4 levels, and rescuing phenotype (invasion). This improved potency can be traced at least partially to **TGP-21 RIBOTAC**'s substoichiometric cleavage, degrading 26 molecules of pre-miR-21 per RIBOTAC molecule. Notably, cleavage was RNase L-dependent as indicated by both gain- and loss-of-function studies. Both miRNome- and proteome-wide studies showed that **TGP-21 RIBOTAC** is indeed selective.

Comparing the biological activity of **TGP-21** and **TGP-21 RIBOTAC** allowed for direct evaluation between the two modes of action, event-driven RNA degradation of RIBOTACs *vs.* occupancy-driven binding of SMIRNAs. Treatment with **TGP-21 RIBOTAC** exhibited a faster, more active and prolonged reduction of miR-21 levels as compared to **TGP-21**. The selectivity of **TGP-21** (dimer binder), **21-SM** (monomeric ligand), and **TGP-21 RIBOTAC** were compared by calculating Gini coefficients from miRNome-wide profiling studies. Gini coefficients range in value from 0 to 1, indicating a non-selective and an exquisitely selective compound, respectively.[46] A Gini coefficient considers the percent inhibition of each target analyzed by a small molecule, ranking the targets by the corresponding percent inhibition; that is selectivity is not scored relative to a particular target, rather over the entire target population. We point the reader to ref. 46 for details about how Gini coefficients are calculated. Generally, a compound is considered selective if the Gini Coefficient >0.75. Our studies showed that selectivity can be improved by multivalency as the Gini Coefficients for **21-SM** and **TGP-21** are 0.52 and 0.68, respectively. Selectivity can be further improved by converting a simple binding compound into a nuclease-recruiting probe, as the Gini Coefficient for **TGP-21 RIBOTAC** is 0.84.

Importantly, in a mouse model of breast cancer metastasis, **TGP-21 RIBOTAC** inhibited metastasis to lung, quantified by reduction of lung nodules. This reduction was due to diminished levels of pre- and mature miR-21 and increased expression of PDCD4, validating the RNA-centric mode of action of **TGP-21 RIBOTAC** *in vivo*.

This study highlighted the comparison of two modes of action that affect cellular levels of mature miR-21. On one hand, occupancy-driven pharmacology exhibited by **21-SM** (monomer) and **TGP-21** (dimer) reduced mature miR-21 levels by interfering with the Dicer processing of pre-miR-21. On the other hand, a more potent and selective biological activity was achieved *via* event-driven pharmacology exhibited by RNA degrader **TGP-21 RIBOTAC**, as a result of degradation of pre-miR-21. Therefore, converting SMIRNAs to RIBOTACs increases potency and selectivity in cells, resulting in a more rapid, effective, and prolonged pharmacological effect in cells and *in vivo*. Interestingly, the catalytic nature of RIBOTACs and its prolonged effect suggest that ideal, or even perhaps very good pharmacokinetic (PK) properties might not be required to observe a therapeutic effect.

8.2 Direct cleavage of RNA targets by SMIRNAs conjugated to bleomycin A5

Another method to ablate RNA is direct cleavage through conjugation of bleomycin A5, a natural product known to cleave nucleic acids, to a SMIRNA. Through attachment of a SMIRNA at the C-terminal primary amine of bleomycin A5, DNA cleavage is reduced such that off-target DNA cleavage does not occur at concentrations required to cleave the desired RNA target.[47] This bleomycin-SMIRNA conjugation strategy has been used to cleave RNA repeats in cells[33] and *in vivo*[47] and various miRNAs in cells.

In one recent example, a bleomycin-conjugated SMIRNA was used to affect the biology of an entire oncogenic miRNA cluster through cleavage.[48] The pri-miR-17-92 cluster is upregulated in various cancers and polycystic kidney disease with the mature miRNAs acting synergistically in some diseases.[49] Thus, a method to simultaneously affect all six miRNAs within the 17-92 cluster could be advantageous. Interestingly, three of the miRNAs share a

common Dicer site, 5′GU/3′CUA: pre-miR-17, pre-miR-18a, and pre-miR-20a (Fig. 7B). Pre-miR-17 and pre-miR-20a also share an adjacent G bulge, while pre-miR-18a contains an A bulge (Fig. 7B). Inforna identified a small molecule, **SMIRNA3**, that binds all three bulges with 30 μM affinity (Fig. 6B). A homodimer, **SMIRNA4**, was created to target the two bulges simultaneously (Fig. 7B).[48] As a simple binding compound, **SMIRNA4**, inhibited the biogenesis of the three miRNAs in TNBC, prostate cancer, and polycystic kidney disease cells. Interestingly, cellular target engagement studies, revealed that **SMIRNA4** bound both pri-miR-17-92 and pre-miR-17, pre-miR-18a, and pre-miR-20a, in agreement with its cellular localization. The dimer de-repressed the corresponding downstream protein in each disease model and rescued phenotype in the two systems in which it was studied (breast and prostate cancer).

Since the occupancy-driven compound demonstrated on-target activity and rescued disease-associated molecular defects in an RNA-centric manner, it was an excellent candidate to employ the direct cleavage approach by conjugation to bleomycin A5, which would allow for the ablation of the entire cluster (Fig. 7B). Indeed, not only did the SMIRNA-bleomycin A5 conjugate, **SMIRNA4-bleo**, reduce levels of all six mature miRNA in the pri-miR-

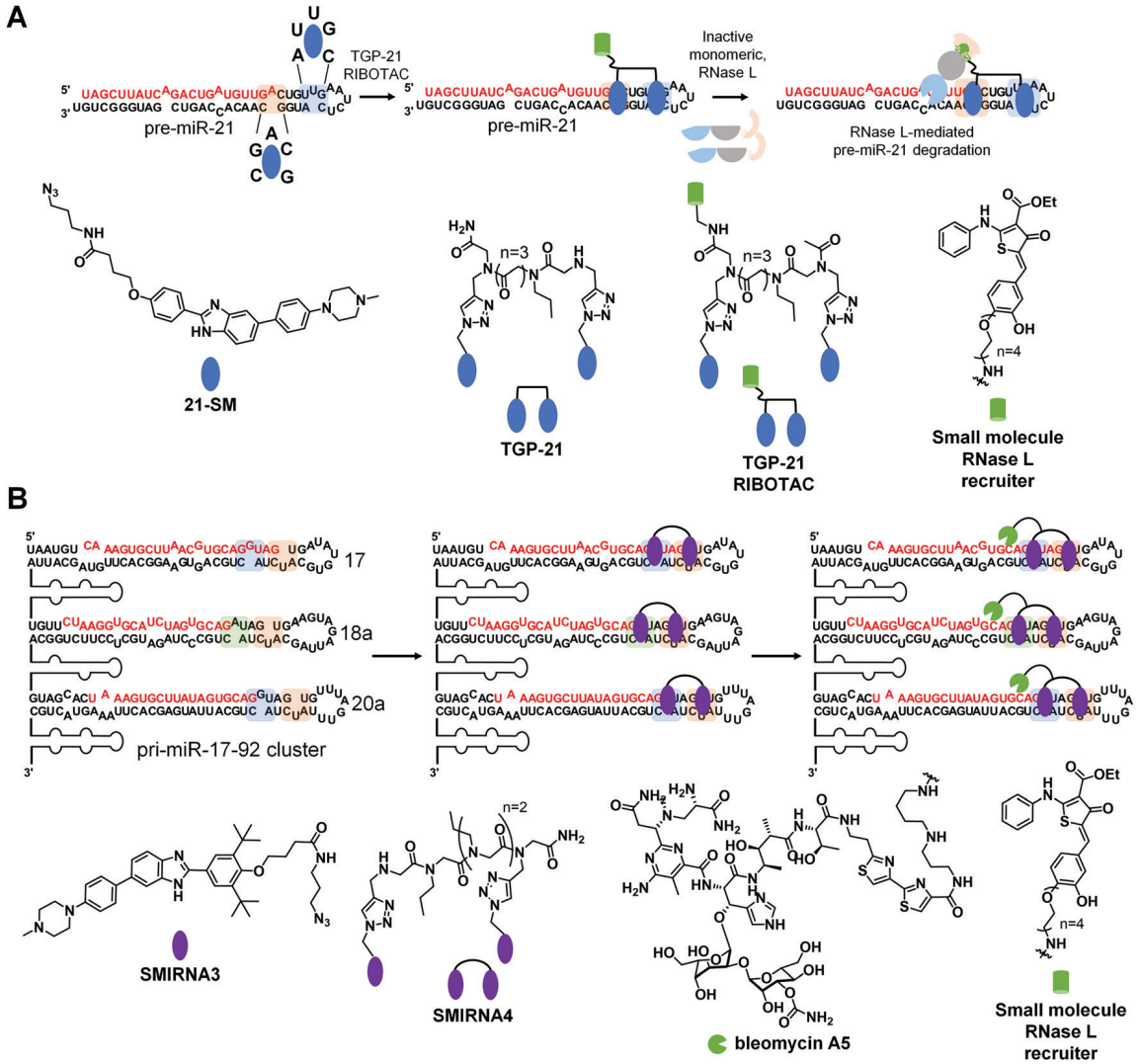


Fig. 7 Developing SMIRNAs into chimeric probes that degrade and cleave disease-causing miRNAs. (A) Inforna identified **21-SM** that targets the Dicer processing site within pre-miR-21 (highlighted in blue). From monomeric **21-SM**, the dimeric compound **TGP-21** was generated to target the Dicer processing site and an adjacent bulge (highlighted in orange). A RIBOTAC probe (**TGP-21 RIBOTAC**) was then synthesized by appending dimeric compound **TGP-21** with a small molecule that recruits endogenous RNase L. **TGP-21 RIBOTAC** more potently and selectively inhibits mature miR-21 levels as a result of the selective RNase L-mediated degradation of pre-miR-21. (B) Inforna identified **SMIRNA3** that binds structured RNA motifs (highlighted in green, blue and orange), within the Dicer sites of pre-miR-17, -18a, and -20a in the miR-17-92 cluster. Dimeric compound **SMIRNA4** was generated from monomeric **SMIRNA3** units connected *via* a peptoid linker. **SMIRNA4** simultaneously targets adjacent bulges present in pre-miR-17, -18a, and -20a, respectively. **SMIRNA4** was appended with bleomycin A5 as a cleaving module, yielding **SMIRNA4-bleo**, and with an RNase L recruiting module, generating **SMIRNA4 RIBOTAC**. **SMIRNA4-bleo** selectively ablated the pri-miR-17-92 cluster resulting in a reduction of all mature miRNAs from this cluster. In contrast, **SMIRNA4 RIBOTAC** only degraded pre-miR-17, pre-miR-18a, and pre-miR-20, as RNase L is cytoplasmic and its interaction is restricted to RNAs present in the cytoplasm.

17-92 cluster, but it also did so more potently than **SMIRNA4** while rescuing downstream circuits in three disease models. As many miRNAs are embedded in clusters, a strategy to cleave a cluster in its entirety could have far reaching implications.

Interestingly, this study also converted **SMIRNA4** into a nuclease recruiting **SMIRNA4 RIBOTAC** (Fig. 6B). In contrast, to the SMIRNA-bleomycin A5 conjugate, **SMIRNA4 RIBOTAC** was only able to cleave pre-miR-17, pre-miR-18a, and pre-miR-20. This is because RNase L is localized to the cytoplasm, meaning **SMIRNA4 RIBOTAC** can only cleave pre-miRs of the pri-miR-17-92 cluster that are present outside the nucleus. Thus, these studies showed that cellular localization can be used to tune compound activity.

9. Conclusions

Although the pharmaceutical industry remains focused on drugging protein targets, many companies have begun investigating the tractability of drugging RNA targets. Indeed, this movement towards RNA has been bolstered by the success of the splicing modifier risdiplam, and derivatives thereof, that treat spinal muscular atrophy.[50] As a result, we are currently experiencing a boom in the identification and validation of druggable human disease-causing RNAs, made possible by advances in sequencing, computation, bioinformatics, chemical probing of RNA structure *in vitro* and *in vivo*, biophysical techniques, structural determination by X-ray crystallography and NMR spectroscopy, *etc.* These interdisciplinary approaches also validate the intimate connection between RNA's 3D structure and its importance in the regulation of biological processes. The ligandability of structured RNA motifs can be achieved by increasingly accurate computational prediction tools, which are easier to implement than various RNA secondary structure chemical probing methods that can be experimentally costly and laborious. For example, ScanFold[8] can rapidly identify biologically relevant structured RNA motifs with high probability to form, especially when coupled with chemical probing of RNA structure in cells and in functional biological experiments. The improvement of such tools will offer the scientific community a more accessible visual perspective of RNA structure and its associated 3D folds, which will ultimately translate into establishing more rational approaches to develop SMIRNAs directly from sequence. However, continued research into the fundamental nature of RNA's 3D structure and the ensemble of conformations featured by structured RNA motifs is sure to inform even more advanced target discovery methods.

In addition to fully understanding RNA structure and dynamics, an equally important aspect is the identification of chemical matter that potently and selectively interacts with structured RNA motifs, *i.e.*, efficient charting of the chemical space for SMIRNAs. Currently available compound libraries are enriched with small molecules designed and optimized for protein targets and the fraction targeting RNA, in a selective manner, is currently unknown. Therefore, screening technologies such as 2DCS along with other methods mentioned above, will aid in identifying chemical matter that potently and selectively bind structured 3D RNA motifs within disease-causing RNAs.

Performing such campaigns by iteratively integrating chemoinformatic/machine learning/statistical approaches will help populate existing databases, such as Inforna, to: (i) improve understanding of the physicochemical properties, parameters and chemical features of small molecules that mediate RNA binding; and (ii) better design tailored-chemical libraries that are more prone to interact with structured RNA motifs.

As previously observed with small molecule chemical probes of protein targets, high potency and selectivity *in vitro* does not always translate into on-target activity in cells or *in vivo*, highlighting the fact that not all chemical matter will be biologically or therapeutically relevant. Therefore, applying target engagement techniques to probe RNA target occupancy by SMIRNAs in cells will help better prioritize chemical scaffolds to be pursued at various stages of chemical probe development. Collectively, these studies will yield the identification of potent and selective SMIRNAs. An array of techniques to assess target engagement to probe RNA-centric modes of action of SMIRNAs have been developed, including Chem-CLIP and Chem-CLIP-Map-Seq,[1] RiboSNAP and RiboSNAP-Map,[1] RIBOTACs,[3] ASO-Bind-Map,[1] and SHAPE.[9,10]

Notably, Chem-CLIP and its competitive version, C-Chem-CLIP, allow for direct assessment of target occupancy *via* covalent crosslinking reactions that either enrich or deplete, respectively, crosslinked SMIRNA-RNA motifs in pull-down fractions. This technique can be used to simultaneously conduct cellular profiling and binding studies and is advantageous over: (i) non-covalent pull down, which lacks precision in which targets are bound in the purification process; and (ii) competitive profiling with SHAPE or DMS, which leaves many sites unreactive and can generate false negatives as the labeling reaction does not occur under equilibrium.

Taken together, the use of target engagement techniques during early stages of the discovery and development process could mitigate off-target effects of SMIRNAs sooner. Although optimization of potency and selectivity *in vitro* is important, more relevant for the development of high-quality SMIRNAs is rescue of phenotype *via* an RNA-centric mode of action, *i.e.*, potent and selective engagement of a biologically relevant structured RNA motif with minimum off-targets proteome- and transcriptome-wide.

An ongoing discussion in the field of small molecule RNA therapeutics is the drug-likeness of SMIRNAs. These semi-empirical rules were historically generated from a pool of approved drugs over a certain interval of time. However, new molecular entities (NME) that were approved since 2002 are deviating from the traditionally considered drug space. Moreover, a recent survey of the approved oral drug space indicated that parameters such as MW and hydrogen-bond acceptors (HBA) have significantly increased over the last 20 year period. Contrarily, over interpretation of ligand and/or drug-likeness metrics might filter out promising chemical candidates. ''Drugging'' RNA with small molecules is still in its infancy, and using parameters derived from protein-targeted drug campaigns to filter out SMIRNAs featuring ''undruglike'' properties might hinder the exploratory research that is necessary to advance the field of small molecule RNA therapeutics.

As previously noted, drug targets are unique; thus, the compounds that successfully target them are also unique. RNA-targeted lead and drug discovery campaigns need to be careful not to lose potential candidates due to selection guidelines that are too narrow, particularly for a field that is rapidly evolving. For example, protein–protein interactions (PPIs), featuring relatively large and flat polar surface areas, are traditionally addressed with macrocyclic compounds, that typically reside outside the ''Rule of Five'' (Ro5), *i.e.*, they are ''Beyond Rule of Five'' (bRo5). The same principle might very well apply to RNAs, where most potent and selective SMIRNAs with *in vivo* activity to date are chimeric compounds, *e.g.*, homo- and/or heterodimers. Interestingly, a survey for active ingredients in recently approved bRo5 drugs revealed several examples of chimeric compounds, including HCV NS5A homodimeric inhibitors such as Pibentrasvir, Ledispasvir, Ombitasvir, Daclatasvir, Elbasvir and Velpatasvir. Although these derivatives exhibit poor drug metabolism and pharmacokinetic (DMPK) properties, including low permeability and solubility and high plasma protein binding capacity that limit oral absorption, these liabilities are overcome by delivery to target organs by human serum proteins and their high affinity binding to the target HCV NS5A protein.

Conversely, other bRo5 approved drugs act locally, thus avoiding systemic exposure. The most recent example is Tenapanor, a sodium-proton exchange sodium/hydrogen exchanger 3 (NHE3) inhibitor, approved in 2019 for irritable bowel syndrome with constipation. Tenapanor is minimally absorbed following oral administration in human plasma (below the limit of quantification). To avoid potential systemic toxicity caused by higher doses, Tenapanor was designed to be restricted to the lumen of the gastrointestinal tract, where its target, NH3 protein, is highly expressed. Moreover, there is a growing body of evidence for the potential therapeutic application of chimeric chemical probes, such as proteolysis targeting chimeras (PROTACs), a bleomycin-SMIRNA conjugate (Cugamycin),[47] and RIBOTACs.[3] Consequently, charting the bRo5 chemical space is likely to reveal novel therapeutically beneficial modalities.

As we continue to identify novel, functional, conserved and structured RNA motifs, these emerging modalities will greatly expand on the types of RNAs that can be targeted with SMIRNAs. In conclusion, exciting times are ahead with the continued exploration of the potential of small molecule chemical probes targeting both functional and non-functional structured RNA motifs to explore RNA biology and affect a broad spectrum of human disorders.

Conflicts of interest

M. D. D. is a founder of Expansion Therapeutics.

References

1 M. D. Disney, *J. Am. Chem. Soc.*, 2019, **141**, 6776–6790.
2 D. K. Hendrix, S. E. Brenner and S. R. Holbrook, *Q. Rev. Biophys.*, 2005, **38**, 221–243.
3 M. G. Costales, H. Aikawa, Y. Li, J. L. Childs-Disney, D. Abegg, D. G. Hoch, S. P. Velagapudi, Y. Nakai, T. Khan, K. W. Wang, I. Yildirim, A. Adibekian, E. T. Wang and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 2406–2411.
4 S. M. Meyer, C. C. Williams, Y. Akahori, T. Tanaka, H. Aikawa, Y. Tong, J. L. Childs-Disney and M. D. Disney, *Chem. Soc. Rev.*, 2020, DOI: 10.1039/d0cs00560f.
5 T. A. Cooper, L. Wan and G. Dreyfuss, *Cell*, 2009, **136**, 777–793.
6 D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker and D. H. Turner, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7287–7292.
7 R. J. Andrews and W. N. Moss, *Biochim. Biophys. Acta, Gene Regul. Mech.*, 2019, **1862**, 194380.
8 R. J. Andrews, J. Roche and W. N. Moss, *PeerJ*, 2018, **6**, e6136.
9 D. Mitchell, 3rd, S. M. Assmann and P. C. Bevilacqua, *Curr. Opin. Struct. Biol.*, 2019, **59**, 151–158.
10 C. Feng, D. Chan and R. C. Spitale, *Methods Mol. Biol.*, 2017, **1648**, 247–256.
11 A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neubock and I. L. Hofacker, *Nucleic Acids Res.*, 2008, **36**, W70–W74.
12 R. J. Andrews, L. Baber and W. N. Moss, *Sci. Rep.*, 2017, **7**, 17269.
13 J. L. Chen, W. N. Moss, A. Spencer, P. Zhang, J. L. Childs-Disney and M. D. Disney, *PLoS One*, 2019, **14**, e0219210.
14 P. Zhang, H.-J. Park, J. Zhang, E. Junn, R. J. Andrews, S. P. Velagapudi, D. Abegg, K. Vishnu, M. G. Costales, J. L. Childs-Disney, A. Adibekian, W. N. Moss, M. M. Mouradian and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 1457–1467.
15 R. J. Andrews, L. Baber and W. N. Moss, *Methods*, 2019, **S1046-2023**(19), 30172-0, DOI: 10.1016/j.ymeth.2019.11.001.
16 S. P. Velagapudi, M. G. Costales, B. R. Vummidi, Y. Nakai, A. J. Angelbello, T. Tran, H. S. Haniff, Y. Matsumoto, Z. F. Wang, A. K. Chatterjee, J. L. Childs-Disney and M. D. Disney, *Cell Chem. Biol.*, 2018, **25**, 1086–1094.
17 S. P. Velagapudi, Y. Luo, T. Tran, H. S. Haniff, Y. Nakai, M. Fallahi, G. J. Martinez, J. L. Childs-Disney and M. D. Disney, *ACS Cent. Sci.*, 2017, **3**, 205–216.
18 J. L. Childs-Disney, T. Tran, B. R. Vummidi, S. P. Velagapudi, H. S. Haniff, Y. Matsumoto, G. Crynen, M. R. Southern, A. Biswas, Z.-F. Wang, T. L. Tellinghuisen and M. D. Disney, *Chem*, 2018, **4**, 2384–2404.
19 C. M. Connelly, F. A. Abulwerdi and J. S. Schneekloth, Jr., *Methods Mol. Biol.*, 2017, **1518**, 157–175.
20 S. P. Velagapudi, S. M. Gallo and M. D. Disney, *Nat. Chem. Biol.*, 2014, **10**, 291–297.
21 S. S. Chavali, R. Bonn-Breach and J. E. Wedekind, *J. Biol. Chem.*, 2019, **294**, 9326–9341.
22 A. Donlic, B. S. Morgan, J. L. Xu, A. Liu, C. Roble, Jr. and A. E. Hargrove, *Angew. Chem., Int. Ed.*, 2018, **57**, 13242–13247.
23 R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem. Rev.*, 2010, **110**, 2579–2619.
24 C. S. Eubanks, J. E. Forte, G. J. Kapral and A. E. Hargrove, *J. Am. Chem. Soc.*, 2017, **139**, 409–416.
25 F. A. Abulwerdi, W. Xu, A. A. Ageeli, M. J. Yonkunas, G. Arun, H. Nam, J. S. Schneekloth, Jr., T. K. Dayie, D. Spector, N. Baird and S. F. J. Le Grice, *ACS Chem. Biol.*, 2019, **14**, 223–235.

26 A. L. Garner, D. A. Lorenz and E. E. Gallagher, in *Methods Enzymol*, ed. A. E. Hargrove, Academic Press, 2019, vol. 623, pp. 85–99.
27 A. H. Jahromi, Y. Fu, K. A. Miller, L. Nguyen, L. M. Luu, A. M. Baranger and S. C. Zimmerman, *J. Med. Chem.*, 2013, **56**, 9471–9481.
28 R. Simone, R. Balendra, T. G. Moens, E. Preza, K. M. Wilson, A. Heslegrave, N. S. Woodling, T. Niccoli, J. Gilbert-Jaramillo, S. Abdelkarim, E. L. Clayton, M. Clarke, M. T. Konrad, A. J. Nicoll, J. S. Mitchell, A. Calvo, A. Chio, H. Houlden, J. M. Polke, M. A. Ismail, C. E. Stephens, T. Vo, A. A. Farahat, W. D. Wilson, D. W. Boykin, H. Zetterberg, L. Partridge, S. Wray, G. Parkinson, S. Neidle, R. Patani, P. Fratta and A. M. Isaacs, *EMBO Mol. Med.*, 2018, **10**, 22–31.
29 J. Li, M. Nakamori, J. Matsumoto, A. Murata, C. Dohno, A. Kiliszek, K. Taylor, K. Sobczak and K. Nakatani, *Chemistry*, 2018, **24**, 18115–18122.
30 J. Sztuba-Solinska, G. Chavez-Calvillo and S. E. Cline, *Bioorg. Med. Chem.*, 2019, **27**, 2149–2165.
31 C. M. Connelly, N. H. Moon and J. S. Schneekloth, *Cell Chem. Biol.*, 2016, **23**, 1077–1090.
32 A. Di Giorgio and M. Duca, *MedChemComm*, 2019, **10**, 1242–1255.
33 S. G. Rzuczek, L. A. Colgan, Y. Nakai, M. D. Cameron, D. Furling, R. Yasuda and M. D. Disney, *Nat. Chem. Biol.*, 2017, **13**, 188–193.
34 D. L. Boger and H. Cai, *Angew. Chem., Int. Ed.*, 1999, **38**, 448–476.
35 A. M. Ardekani and M. M. Naeini, *Avicenna J. Med. Biotechnol.*, 2010, **2**, 161–179.
36 B. Liu, J. L. Childs-Disney, B. M. Znosko, D. Wang, M. Fallahi, S. M. Gallo and M. D. Disney, *BMC Bioinf.*, 2016, **17**, 112.
37 S. P. Velagapudi, M. D. Cameron, C. L. Haga, L. H. Rosenberg, M. Lafitte, D. R. Duckett, D. G. Phinney and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 5898–5903.
38 T. J. Kelly, A. L. Souza, C. B. Clish and P. Puigserver, *Mol. Cell. Biol.*, 2011, **31**, 2696–2706.
39 M. G. Costales, C. L. Haga, S. P. Velagapudi, J. L. Childs-Disney, D. G. Phinney and M. D. Disney, *J. Am. Chem. Soc.*, 2017, **139**, 3446–3455.
40 M. G. Costales, D. G. Hoch, D. Abegg, J. L. Childs-Disney, S. P. Velagapudi, A. Adibekian and M. D. Disney, *J. Am. Chem. Soc.*, 2019, **141**, 2960–2974.
41 A. Grover, H. Houlden, M. Baker, J. Adamson, J. Lewis, G. Prihar, S. Pickering-Brown, K. Duff and M. Hutton, *J. Biol. Chem.*, 1999, **274**, 15134–15143.
42 Y. Liu, E. Peacey, J. Dickson, C. P. Donahue, S. Zheng, G. Varani and M. S. Wolfe, *J. Med. Chem.*, 2009, **52**, 6523–6526.
43 J. L. Chen, P. Zhang, M. Abe, H. Aikawa, L. Zhang, A. J. Frank, T. Zembryski, C. Hubbs, H. Park, J. Withka, C. Steppan, L. Rogers, S. Cabral, M. Pettersson, T. T. Wager, M. A. Fountain, G. Rumbaugh, J. L. Childs-Disney and M. D. Disney, *J. Am. Chem. Soc.*, 2020, **142**, 8706–8727.
44 T. T. Wager, X. Hou, P. R. Verhoest and A. Villalobos, *ACS Chem. Neurosci.*, 2010, **1**, 435–449.
45 M. G. Costales, B. Suresh, K. Vishnu and M. D. Disney, *Cell Chem. Biol.*, 2019, **26**, 1180–1186.
46 P. P. Graczyk, *J. Med. Chem.*, 2007, **50**, 5773–5779.
47 A. J. Angelbello, S. G. Rzuczek, K. K. McKee, J. L. Chen, H. Olafson, M. D. Cameron, W. N. Moss, E. T. Wang and M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 7799–7804.
48 X. Liu, H. S. Haniff, J. L. Childs-Disney, A. Shuster, H. Aikawa, A. Adibekian and M. D. Disney, *J. Am. Chem. Soc.*, 2020, **142**(15), 6970–6982.
49 P. Mu, Y. C. Han, D. Betel, E. Yao, M. Squatrito, P. Ogrodowski, E. de Stanchina, A. D'Andrea, C. Sander and A. Ventura, *Genes Dev.*, 2009, **23**, 2806–2811.
50 H. Ratni, M. Ebeling, J. Baird, S. Bendels, J. Bylund, K. S. Chen, N. Denk, Z. Feng, L. Green, M. Guerard, P. Jablonski, B. Jacobsen, O. Khwaja, H. Kletzl, C.-P. Ko, S. Kustermann, A. Marquet, F. Metzger, B. Mueller, N. A. Naryshkin, S. V. Paushkin, E. Pinard, A. Poirier, M. Reutlinger, M. Weetall, A. Zeller, X. Zhao and L. Mueller, *J. Med. Chem.*, 2018, **61**, 6501–6517.

CHEMICAL REVIEWS

pubs.acs.org/CR

Review

DNA-Encoded Chemistry: Drug Discovery from a Few Good Reactions

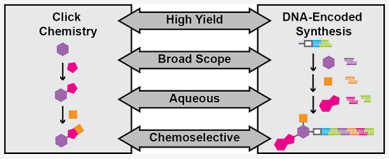
Patrick R. Fitzgerald and Brian M. Paegel*

Cite This: https://dx.doi.org/10.1021/acs.chemrev.0c00789

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Click chemistry, proposed nearly 20 years ago, promised access to novel chemical space by empowering combinatorial library synthesis with a "few good reactions". These click reactions fulfilled key criteria (broad scope, quantitative yield, abundant starting material, mild reaction conditions, and high chemoselectivity), keeping the focus on molecules that would be easy to make, yet structurally diverse. This philosophy bears a striking resemblance to DNA-encoded library (DEL) technology, the now-dominant combinatorial chemistry paradigm. This review highlights the similarities between click and DEL reaction design and deployment in combinatorial library settings, providing a framework for the design of new DEL synthesis technologies to enable next-generation drug discovery.



CONTENTS

1. INTRODUCTION

Early stage drug discovery in both the pharmaceutical industry and academia has driven technology development efforts to revolutionize methods for generating and screening large collections of compounds for biological activity. What began as amassing compounds for robotic high-throughput screening (HTS) has evolved dramatically in recent years with the development of DNA-encoded library (DEL) technology.[1] DELs—combinatorial libraries of drug-like molecules, each barcoded with a DNA sequence that encodes the attached library member's structure—can be prepared at unprecedented scales of diversity and efficiently screened for ligands of a purified protein target. The many published examples of DEL screening hits becoming leads or even clinical candidates have in turn spurred intense interest in developing new chemistry to generate DELs of ever-increasing structural diversity that maximize the probability of successfully discovering additional leads. However, the constraints of DEL-compatible reaction development (e.g., in water, quantitative yield, dilute reactants, DNA-orthogonal) pose a great challenge to modern synthetic organic chemistry.

A surprising majority of the properties that define DNA-compatible chemistry coincide with those of click reactions. As initially described, the click philosophy is a pragmatist's approach to chemistry: achieve diverse chemical function from "a few good reactions".[2] Click reactions proceed with

Special Issue: Click Chemistry

Received: July 28, 2020

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

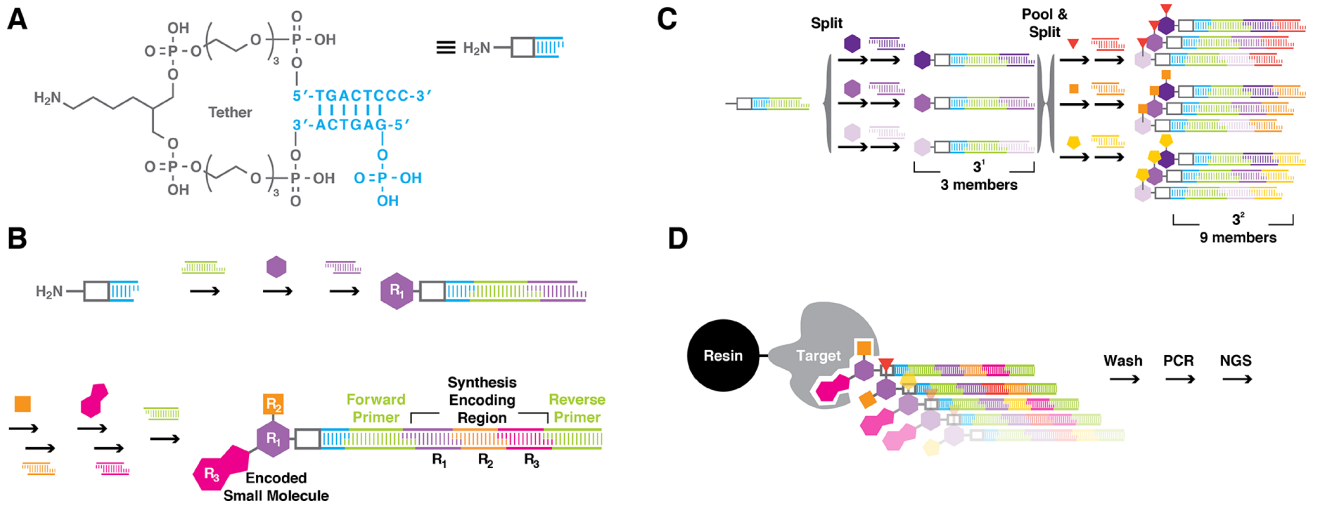


Figure 1. DNA-encoded library synthesis and screening. (A) The linchpin "headpiece" DNA molecule is a covalently tethered (gray) dsDNA heteroduplex (cyan) displaying a primary amine for chemical synthesis and a 5′-CC-3′ dinucleotide overhang and 5′ phosphate for enzymatic cohesive end ligation. (B) Encoded synthesis proceeds in interleaved steps of enzymatic ligation and building block coupling. Synthesis begins with ligation of a 5′-phosphorylated dsDNA (green) to the headpiece. Coupling of the position 1 building block (R_1, purple hexagon) and enzymatic encoding (purple dsDNA) follow. Synthesis continues with analogous coupling and encoding steps for the position 2 building block (R_2, orange square) and position 3 building block (R_3, magenta bicycle) and a final dsDNA ligation (green). Green sequences flanking the synthesis encoding region are constant PCR primer binding sites. The product is a small molecule attached to a DNA whose sequence encodes the synthesis history of the small molecule. (C) Split-and-pool diversification entails parallel position 1 building block coupling and encoding (purple hues) to yield three different example DEL members. A second encoded split-and-pool synthesis step (orange hues) yields $3^2 = 9$ different DEL members. (D) Protein target (gray) immobilized to resin captures all molecules in the library that are ligands. Washing removes unbound species, leaving the bound fraction encoding DNA sequences for PCR amplification, sequencing, and structure decoding.

quantitative yield and with minimal side-product formation in aqueous or inoffensive solvent, are broad in substrate scope, and ideally require no chromatographic purification. These criteria are also highly advantageous for DEL preparation. For example, while 20 or more DEL-compatible reactions have been developed, a single DEL generally employs only two to three robust reactions to maximize library quality. Furthermore, DEL generation often entails parallel synthesis using hundreds (if not thousands) of substrates, thus demanding broad scope, and chromatography is impossible past the first coupling cycle. *The defining elements of click reactions in essence describe the ideal reaction for DEL.*

In this review, we spotlight the remarkable relationship between click chemistry and DEL reaction development. In the first part of this review, we begin with a brief description of DEL technology and establish a framework for evaluating reaction development. We then apply these criteria to analyze six commonly practiced DEL reactions and three emerging reaction formats that potentially expand the scope of DEL synthesis. We restrict coverage to DNA-*encoded* synthesis (i.e., the DNA sequence encodes the synthetic history, it does not template or direct library synthesis) from 2008 to 2020, with some relevant background. DNA-templated[3,4] and dual pharmacophore[5] libraries are alternative encoded library approaches that have evolved significantly over the years and have delivered hits against numerous targets.[6−8] However, reaction development for these types of libraries diverges significantly, since they use purified and validated DNA−small molecule heteroconjugates to yield well-defined combinations of building blocks. Our analysis of prospective strategies for expanding DEL synthesis, such as photocatalytic transformations and solid-phase reversible immobilization, is speculative, as the state of the art is changing rapidly. However, we expect that this broader framework for reaction development will guide future efforts to achieve higher-quality and more structurally novel DELs, thereby delivering ever more successful screening outcomes in drug discovery.

2. CLICK CONSTRAINTS ESTABLISH A FRAMEWORK FOR DEL DESIGN

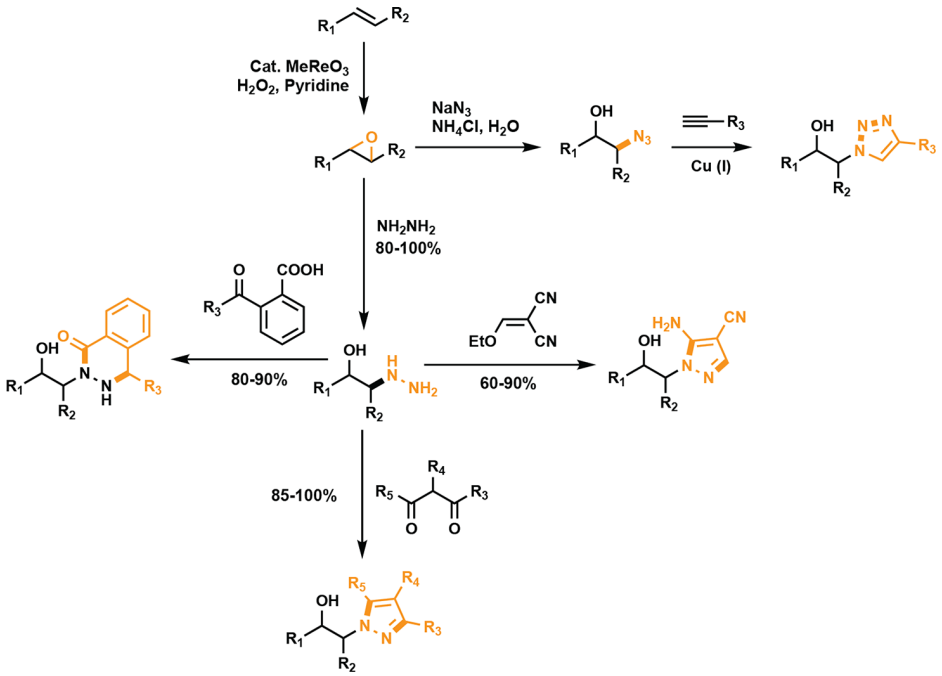
2.1. DEL Fundamentals

Despite major technological advances in genome sequencing, structural biology, and computational drug design, drug discovery remains an empirical science. Screening large (100,000−5,000,000) compound collections by HTS is routine in the pharmaceutical industry and a handful of academic centers. These collections vary in composition depending on the screening center but often comprise compounds synthesized in house, natural product extracts, and compounds purchased through external vendors.[9,10] Advances in laboratory automation and analytical instrumentation have delivered the capacity to screen >100,000 compounds per day, but HTS remains a costly endeavor that is limited by the chemical diversity inherent to collected compounds.

As HTS took root in industry, the rate of screening began to outstrip the rate at which chemical diversity could be generated, driving the development of more efficient chemical library synthesis methods. Combinatorial chemistry emerged as a prospective solution to the synthesis bottleneck. In contrast to HTS compound collections, which grow linearly with acquisition, combinatorial synthesis exponentially diversifies a target scaffold by split-and-pool strategies.[11,12] Combinatorial libraries could be screened against purified proteins or against whole cells,[13] but limited throughput of analytical methods available at the time for determining the hit structures after screening resulted in underpowered experiments.[14,15]

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Scheme 1. Structural Diversification Using a Few Good Reactions[a]



[a]Adapted with permission from ref 2. Copyright 2001 Angewandte Chemie International Edition.

The advent of DEL technology in 2009 resurrected the field of combinatorial chemistry and gave rise to a powerful new mode of designing and searching chemical space. Originating from a seminal thought experiment of Lerner and Brenner in 1992,[16] DNA-encoded synthesis matured as a technology through disclosures from Neri[17] and, critically, Morgan,[1] who described DEL as it is widely practiced today (Figure 1). The power of DEL lies in linking compound identity with DNA-encoding tags. The DEL can be affinity panned as a highly complex mixture against the immobilized protein target of interest, and the specifically bound fraction is amplified and deep sequenced[18] to reveal large collections of hit structures en masse. The statistical power of these experiments eclipses that of combinatorial chemistry from decades past by many orders of magnitude, simultaneously lending high confidence in hit authenticity and revealing detailed structure−activity relationship trends.[19] The analytical throughput advantages of DNA deep sequencing, however, are contingent on the library chemistry, yielding solely the intended product while minimally compromising the fidelity of the encoding DNA.

2.2. Applying the Click Constraints to DEL

The quantitative yield, bioorthogonality, and other constraints associated with the ideal DEL reaction are highly evocative of the broader foundations of click reaction development. To the contemporary chemical biologist, click chemistry usually refers to the Cu(I)-catalyzed azide−alkyne cycloaddition (CuAAC) reaction for bioconjugation,[20−22] but click chemistry was originally a pragmatic, philosophical approach to defining a new mode of ligand discovery.[2] Given that the number of permutations of H, C, N, O, P, S, F, Cl, and Br atoms yielding a drug-like molecule is vast (10^{63})[23] compared to pharmaceutical industry compound collections (10^{6}), the click philosophy suggested looking for leads in the most strategically accessible regions of chemical space to expedite the process.[2,24]

The click ideals for ligand discovery simply restrict all searches to molecules that are easy to make.[2] As a corollary to this rule, the philosophy strongly advocates creating chemical diversity by using only the most efficient "click" reactions for coupling two building blocks rather than performing multiple reactions. Even a few reactions with broad scope can yield libraries with sufficient diversity by capitalizing on readily available and deep building block sets. Aspects of click chemistry that resonate strongly with current practices in the field of DEL and underscore important commonalities between the reactions that define the two approaches include (1) broad scope and high yield and (2) viable in water or benign solvent under mild conditions. Click chemistry has also evolved, becoming intertwined with bioorthogonal chemistry.[25] This final aspect of click chemistry also reflects an increasingly important consideration in DEL reaction development: (3) DNA-orthogonal reactivity. The following sections provide in-depth analysis of recent DEL reaction development efforts in the context of these click chemistry constraints.

2.2.1. Maximize Reaction Scope and Yield. High reaction yield is important to all chemists, but yield requirements and reaction considerations vary with application, such as in process chemistry, total synthesis, medicinal chemistry, or combinatorial chemistry. Likewise, the importance of yield carries different concerns in click and DEL. The primary concern for yield in the context of click chemistry relates to expedient and efficient synthesis of individual molecules to populate large compound libraries. In contrast, DEL practitioners strive for maximum reaction yield during split-and-pool synthesis to ensure that all intended library members are prepared. As the composition of a full DEL cannot be analytically measured, nor each member individually purified, DEL generally adheres to the first principle of click, employing only high-yielding reactions.

Click reactions are by definition very high yielding. While this was not quantitatively defined, originally published click

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

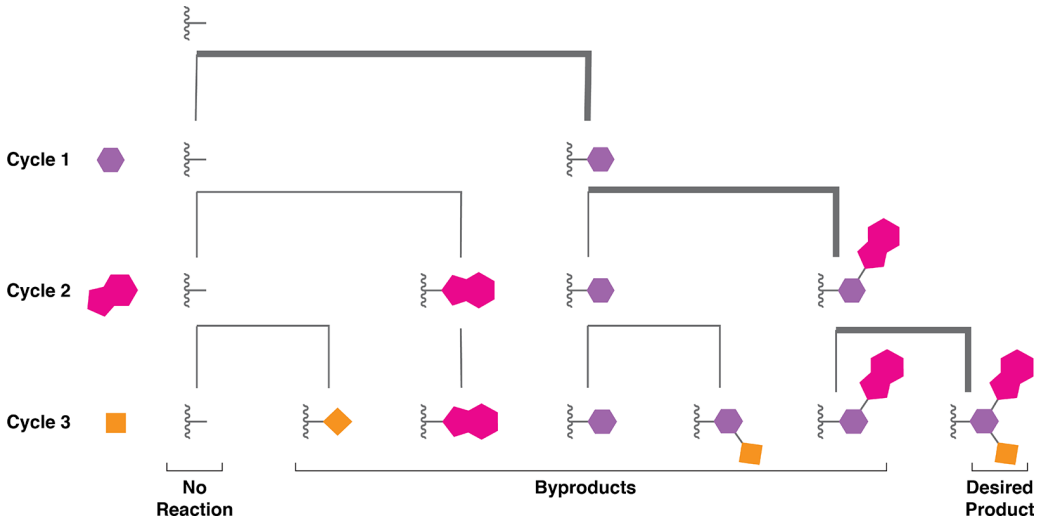


Figure 2. DNA-encoded library synthesis decision tree. Combinatorial synthesis yields an exponentially increasing diversity of products. The decision tree maps all possible molecular outcomes of a three-cycle DEL synthesis (excluding unanticipated side reactions). The attachment point to DNA is shown with a wavy line. The DEL synthesis begins with coupling of a trifunctional hub (purple hexagon) and ends with two capping groups (magenta, orange). Three successful building block couplings lead to formation of the desired product three-cycle DEL member (bold decision branches). Incomplete building block coupling leads to one of five different truncate byproducts, or in the case of no coupling, an unmodified DNA results. The scheme is representative of the triazine DEL, which are the product of two sequential nucleophilic aromatic substitutions of a cyanuric chloride hub using amine building blocks. The desired product, byproducts, and no reaction all share the same encoding sequence. As a consequence, schemes with fewer steps employing only high-yielding chemistry universally lead to higher fidelity library outputs by maximizing the number of encoding tags that display the desired product. Adapted from ref 42. Copyright 2016 American Chemical Society.

reactions proceeded with >60% yields, with many reactions achieving >95% yield.[2] These reactions produced C-heteroatom functionality through four main classes of reactions, including C−C multiple bond addition (epoxidation, aziridination, dihydroxylation, Michael addition), nucleophilic ring opening, cycloaddition (Diels−Alder, Huisgen dipolar addition), and non-aldol carbonyl chemistry (formation of oxime ethers, hydrazones). Notably, the majority of these reactions are fusion reactions that can be conducted stoichiometrically, proceeding to completion without generating undesired side products. These few reactions can be combined in just two to three steps to generate new structures with interesting function (Scheme 1).

A high-yielding reaction with broad scope can be incredibly enabling for discovery science. The CuAAC reaction is perhaps the quintessential example. When it was discovered shortly after the formalization of click principles,[20,21] CuAAC offered significant regioselectivity and rate enhancement advantages over the uncatalyzed Huisgen condensation. CuAAC became a reaction of choice for large-scale library synthesis due to its high yield, orthogonal reactivity, broad scope, and mild conditions.[24] In fact, researchers at Lexicon Pharmaceuticals used CuAAC and the first-generation click reactions to generate a 200k-member library of individually purified compounds (25−50 mg each).[24] Despite comprising few reactions, click chemistry has demonstrated the potential to create diverse libraries by sampling readily available building block sets[24,26−33] and continues to enable access to new chemical spaces by virtue of sustained reaction development, such as the recently disclosed sulfur(VI) fluoride exchange (SuFEx) chemistry.[34−37]

Like click chemistry, DEL seeks to generate massive diversity expediently from large collections of building blocks. For example, the seminal disclosure of DEL technology[1] described two library plans for diversifying a central triazine scaffold with large building block pools (32−384 building blocks/cycle), resulting in impressive library diversity (7×10^6 and 8×10^8 members). However, subsequent computational analysis[38,39] has shown that these higher-cycle-number DELs contain large swathes of chemical space that deviate from typical drug-like molecules. Increasing the number of synthesis cycles generally increases molecular weight and all other relevant physicochemical property distributions (MW, logP, TPSA, H-bond donors/acceptors).[39,40] These considerations collectively prescribe constraining synthesis to two to three cycles using large diversity sets. For example, a two-cycle DEL, which is likely the most densely populated with drug-like matter, will require two parallel synthesis steps of >1000 building blocks per coupling, underscoring the importance and scale of reaction scope in the state of the art.

To synthesize such a drug-like library with high chemical diversity, the requisite building blocks must be readily available and all members of these large building block sets must couple to the growing encoded compound in high yield to ensure interpretable screening data. Given the nature of split-and-pool synthesis, the presence and diversity of reaction byproducts increases exponentially with the number of additional synthetic steps, while the fraction of desired product corresponding to a given sequence decreases exponentially (Figure 2). As reaction yield is inevitably variable across building blocks, this means that, among similarly potent library members, those with higher synthesis yield will be preferentially identified as hits during affinity selection. In fact, in computational screening simulations of three-cycle libraries ($100 \times 100 \times 100$) that parametrized and modeled the outcome of variable synthesis yields (average 60%, standard deviation 20% per step), library member enrichment was more highly correlated with synthesis yield than with target association constant.[41,42] These simulations support the hypothesis that low yielding reactions decrease the signal-to-noise ratio in DEL, potentially leading to false SAR.

Further experimental analysis has substantiated the importance of incorporating high-yielding reactions. Sixteen DELs

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Table 1. Amide Formation Conditions

activating agent	conditions	scope investigated	library	year
PyBOP/DIPEA[a]		AAs L, F, G, K, Y, P	5	1994[103]
HBTU/HOBt/DIPEA	0.1 M AA/HBTU/HOBT; 0.3 M DIPEA	(L) AAs R, Q, K, V, and T; (D) V	8.2×10^5 (5 cycles)	1994[104]
DMT-MM or EDC/Sulfo-NHS (sNHS)	templated (60 nM NH_2-DNA, 120 nM COOH-DNA, 50 mM DMT-MM in 0.1 M MOPS buffer pH 7.0, 1 M NaCl, 16 h, 25 °C	2 NH_2-DNA, 4 COOH-DNA		2002[105]
EDC/HOAt	DEAE immobilization, 50 mM Fmoc-AA/EDC, 5 mM HOAt, MeOH, 30 min, RT	33 AAs	10^6 (6 cycles, DNA-routed)[106]	2004[107]
DMT-MM or HATU/DIPEA	0.7 mM NH_2-DNA, 28 mM DMT-MM/Fmoc-AA	96 AAs (yield not reported)	8.4×10^7 tetrapeptides	2005[108]
EDC/sulfo-NHS[a]	50 μM NH_2-DNA, 4 mM Fmoc-AA/EDC, 10 mM sNHS, 70% DMSO, 30% 80 mM TEA, pH 9	20 AAs and 200 RCOOH	4×10^3 (2 cycles)	2008[18]
DMT-MM[a]	0.7 mM NH_2-DNA, 27 mM Fmoc-AA, 27 mM DMT-MM, 18% DMF, 82% 130 mM borate pH 9.5, 18 h, 4 °C	>192/400 Fmoc-AAs (yield >70%)	DEL A: 7×10^6 (3 cycles), DEL B: 8×10^8 (4 cycles)	2009[1]
	0.44 mM $COOH_2$-DNA, 22 mM amine, 44 mM DMT-MM, 111 mM phosphate pH 5.5, 44 mM HCl, 94% H_2O, 6% MeCN, 72 h, RT, 50 additional equiv of DMTMM added at 48 and 60 h	383 amines		
9 tested, EDC/HOAt/DIPEA recommended[a]	4 μM NH_2-DNA, 9 mM EDC, 2 mM HOAt, 9 mM DIPEA, 56% MOPS pH 8, 44% DMSO, 16 h, RT	543 RCOOH tested for DMT-MM and EDC/HOAt/DIPEA		2016[97]
DIC/HOAt[b]	resin-bound DNA, resin-bound 1° amine: 40 mM Fmoc-AA, 40 mM HOAt, 57 mM DIC, DMF 1 h, 37 °C; resin-bound 2° amine: 80 mM Fmoc AA/Oxyma/TMP, 100 mM DIC, 3 h, 37 °C	1 Fmoc-AA	6×10^4 (2 cycles)[109]	2016[86]
HATU/DIPEA[a,b]	71 mM Boc-AA/DIPEA/HATU, 43% DMA, 57% 250 mM borate pH 9.5, 2 h, RT	Boc-Phe-OH		2019[54]
DMT-MM[a,b]	60 μM NH_2-DNA, 60 mM AA/DMT-MM, 120 mM *N*-methyl morpholine, 58% DMSO 42% 50 mM MOPS pH 8.2, 24 h, RT	126 AAs in library	1.4×10^6 macrocycles	2019[47]
DEPBT/DIPEA[a]	0.24 mM NH_2-DNA, 24 mM COOH/DEPBT/DIPEA, 100 mM borate pH 9.5, 63% H_2O, 37% MeCN, 2 h, RT	57 N-Boc AAs, 23 nitro benzoic acids	7.5×10^7	2019[98]
DMT-MM (reverse acylation)	0.2 mM COOH-DNA, 20 mM amine/DMTMM, 100 mM MES buffer pH 5.8, 65% H_2O, 35% MeCN, ON, RT	183 N-Boc diamines, 23 nitro anilines		
DMT-MM		library included 57 N_3-AAs (Fmoc/Boc) and 1846 RCOOH (used in two positions)	6.6×10^8 (3 cycles)	2020[72]

[a]Analysis of DNA conjugates by HPLC(A260) and LCMS. [b]Quantitative analysis of DNA compatibility by qPCR.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

were synthesized and screened against a phosphodiesterase and a kinase.[43] The library that produced the most hit clusters in these experiments was generated by employing robust acylation/Fmoc deprotection conditions; library productivity did not correlate with numeric size. While this study was limited to Roche's DELs and only two protein targets, it is now routine practice across many groups to screen building block sets and exclude monomers that are problematic for DEL synthesis.[1,44−49] Notably, both of GSK's DEL-derived clinical candidates were the product of libraries prepared using robustly validated acylation or nucleophilic substitution of cyanuric chloride.[19,50−53] There is no agreed upon yield threshold for including a building block in a DEL synthesis, but standard practice seems to suggest that a yield >70% is suitable.[1,44−49]

Yield determination for large building block sets requires high-throughput, automated workflows that are sufficiently sensitive to analyze the rather unusual DNA−small molecule heteroconjugate products of DEL synthesis. DEL reactions are usually performed at the nanomole scale, and the product of building block coupling typically results in a mass shift of ∼100−300 Da on DNA starting material >5000 Da. LCMS accompanied by UV absorbance detection is the gold standard for DEL reaction yield determination. Reaction crude is separated, and relative abundance of starting materials and products is quantitated by the DNA tag UV absorbance (λ = 260 nm), which is the dominant contributor to the molar extinction coefficient. The DNA tag also dominates chromatographic character; thus, the same LC method usually provides adequate separation for all DNA-building block conjugates. LC method optimization (heating, hexafluoroisopropanol mobile phase, ion-pairing reagents, column particle size) has further enhanced separation efficiency and sensitivity while reducing analysis time.[1,44,54] MS analysis enables facile peak assignment and identification of side products; some also favor using MS abundance over UV absorbance for quantitation.[55] LCMS analysis cannot offer insight into reaction performance in a true combinatorial synthesis setting because DELs are too complex, but the workflow is routinely deployed for building block validation.

2.2.2. Employ Water- and Air-Compatible Reactions. Just as analytical characterization of DEL compounds is shaped by the dominant physical properties of DNA, optimal synthesis conditions must also accommodate DNA. DNA's hydrophilic nature and the lability of its glycosidic bonds impose the strictest constraints on DEL synthesis, limiting reaction conditions to aqueous solution with organic cosolvent, moderate temperatures (4−95 °C), and moderate pH (4−10). These coincide with click conditions, which prize a simple reaction setup using a benign solvent (often water) and insensitivity to water/oxygen.[2,24] The similarity between click and DEL reaction development becomes even more apparent in the context of click-enabled bioorthogonal labeling chemistry (discussed later) where robust and selective reaction performance under aqueous conditions is necessary, for example, to preserve native protein structure or cell viability.

Using water as the solvent during click chemistry library generation is profoundly and counterintuitively enabling. Aqueous synthesis epitomizes ease of implementation, and water's physical and chemical properties are favorable for characteristic reactions. While the low solubility of many organic compounds in water may seem like a liabilty, low compound solubility can be offset by the high free energy of organic compounds poorly solvated with water.[2] In fact, some reactions perform better in this fashion and some of the very best click reactions, such as 1,3-dipolar azide/alkyne cycloaddition, proceed best when reactants are "on water".[24] The nucleophilicity of water is also a potential liability, but again, it is offset by water's behavior specifically in the click context. Water's polar nature and tendency to form H-bonds allows it to facilitate rapid changes in H-bonding required for nucleophilic additions (epoxide/aziridine ring opening) that feature prominently in click. Moreover, water's high specific heat capacity (4.18 $J \cdot g^{-1} \cdot K^{-1}$) allows the solvent to double as a heatsink for highly exothermic click reactions. The low volatility and high surface tension of water also make it an ideal solvent for automated liquid handling, which facilitates large-scale parallel synthesis.

Water is similarly ideal and virtually required for solution-phase DEL chemistry because it is the only solvent in which the DNA polyanion is appreciably soluble. However, this solubility is still minimal (<10 mM) compared to the concentrations of reactants found in traditional organic synthesis (0.1−1 M), which introduces kinetic reaction constraints. While click chemistry (as initially envisioned) may be more closely aligned with organic synthesis conditions, click chemistry extended to bioorthogonal labeling is quite reminiscent of DEL. Like DEL, bioorthogonal labeling chemistry also faces the challenge of limiting reactant concentration in the form of minimal biological reactant (micromolar or lower concentration of antibody, native protein, metabolite, etc.). Bioorthogonal chemistry development often solves this challenge by designing reactions with large kinetic rate constants, a feature that is also desirable for DEL reactions.[25,56] Indeed, several routine bioorthogonal labeling reactions have been applied to DEL, including CuAAC,[47,48,57−63] SuFEx,[35] strain-promoted cycloaddition,[64] and inverse electron demand Diels−Alder reactions.[65] Furthermore, using high-throughput experimentation and reaction progress kinetic analysis protocols, it is now possible to design the highest yielding reactions within kinetic constraints by, for example, employing excess small molecule building blocks, different catalysts, and/or ligands, higher temperature, solvent admixtures, or some complex combination of these parameters.[66] Even with these advances, DEL reaction development remains a challenging endeavor, especially considering the difficulties associated with designing aqueous reactions.

Water is an enabling solvent for many click reactions, but it is a challenging solvent for most of the medicinal chemistry transformations that feature prominently in DEL. Water is necessary to solubilize DNA, but organic cosolvent (DMA, ACN, MeOH) is also necessary to improve building block solubility, making reaction development an exercise in compromise. The suggested percentage of organic solvent that is compatible with on-DNA synthesis ranges from <50[67] to <80%.[66] In our analysis of commonly employed DEL reactions, we observe that the majority of reactions are performed with <50% organic cosolvent (Tables 1−5) though certain on-DNA reactions tolerate higher percentages of organic solvent.[64,68−70] Balancing building block solubility with DNA solubility is a major challenge in DEL, so organic solvent percentage optimization is routine. Interestingly, as was observed for click reactions, DEL reactions can sometimes proceed even when building blocks are not fully soluble.[71] Methods for integrating solid-phase synthesis with DEL have emerged for circumventing building block solubility issues and for enabling water-sensitive reactions (discussed in detail later).

2.2.3. Prioritize High Chemoselectivity. In addition to having broad scope and high yield while proceeding in aqueous

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

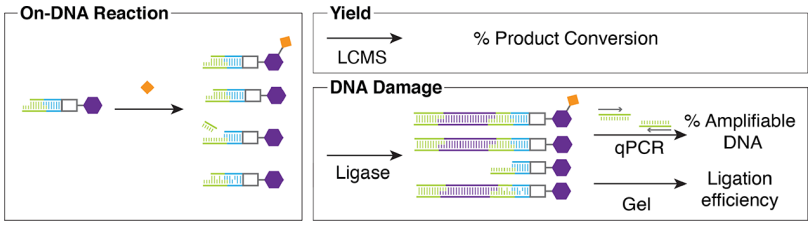


Figure 3. DNA compatibility assay. Comprehensive qPCR and LCMS characterization of DEL reactions is critical for preparing high-quality libraries. A candidate reaction is evaluated by coupling one or several model building blocks (orange diamond) in an on-DNA reaction with an elaborated DNA headpiece. The elaborated headpiece displays the appropriate reaction site (purple hexagon) for the candidate reaction and a partial DNA encoding tag. The resulting reaction mixture is analyzed by LCMS to obtain the on-DNA % conversion. The mixture is also enzymatically ligated (T4 DNA ligase) to dsDNA oligonucleotide modules that install binding sites for a Taqman exonuclease probe (purple) and closing primer (green). The ligase reaction is analyzed for ligation yield by gel electrophoresis and for the molecules of amplifiable DNA remaining by qPCR. The qPCR analysis is also conducted on DNA that has experienced control conditions (e.g., incubation in buffer) and the result used to calculate % amplifiable DNA remaining after exposure to conditions of chemical synthesis. Adapted from ref 54. Copyright 2019 American Chemical Society.

solvent, click reactions exhibit high chemoselectivity, a critical characteristic of DEL reactions. The specific considerations of reaction orthogonality are interesting to compare between these applications and even suggest new reactions for implementation in DEL. Click reactions used for parallel library synthesis leverage orthogonal reactivity to allow for sequential transformations without the need for protecting groups or purification. Click reactions used for bioconjugation demand conditions that are inert to biological nucleophiles, and in many cases, the reaction must occur in the complex intracellular milieu. The most useful transformations for DEL often share these elements; highly chemoselective reactions allow for multiple sequential building block coupling reactions while minimally interfering with the DNA encoding tag.

Avoiding complicated protecting group strategies is a major motivating factor for employing click chemistry in library synthesis. This is achieved through orthogonal, high-yielding chemistry. For example, in the library generation strategy we described earlier (Scheme 1), nucleophilic epoxide-ring opening produces a free hydroxyl group and an azide or hydrazine group depending on the nucleophile. This reaction is followed by non-aldol carbonyl or cycloaddition reaction for the respective nucleophile. When performed in water, these reactions are inert to the −OH group that would otherwise require protection. Similarly, these reactions are also inert to most amine nucleophiles, allowing a broad range of building blocks.

Reactions in DEL share with click a need for chemoselectivity, as both are library synthesis strategies involving highly diverse pools of building blocks that display different functional groups. As such, orthogonal reactivity in both building block sets and protecting group strategies (if used) is critical. For example, the Schreiber lab recently synthesized a stereochemically rich, 100,000-member library using orthogonal amidation, N-Boc deprotection, reductive amination, sulfonylation, and Suzuki coupling reactions.[44] In another recent example, researchers at Pfizer synthesized a library using amidation, reductive amination, sulfonylation, and carbamoylation coupling reactions, while azide, Fmoc, and Boc functionalities served as NH_2 protecting groups.[72]

Highly chemoselective reactions are also important in DEL for the critical reason that side reactions with the DNA tags can compromise library selection analysis. This is evocative of click applied to bioconjugation, such as activity-based protein profiling,[73−75] preparation of antibody−drug conjugates,[76,77] and metabolic labeling in cells or animals.[78−82] Optimizing these reactions to perform under physiological conditions while avoiding reactivity with off targets, such as proteins or intracellular thiols, is a challenging endeavor. These concerns, as well as avoiding cellular toxicity, make the constraints of bioorthogonal reaction development stricter than those for DEL, but DNA still possesses structural features that can be problematic. The main structural liabilities present in DNA are the reactive heteroatoms of the nucleobases (exocyclic amines and purine N3, N7), the nucleophilic 3′-OH necessary for enzymatic ligation, the glycosidic bond, and the phosphodiester backbone.[83,84] Several reaction conditions, such as low pH and high temperature, lead to DNA damage through depurination and concomitant phosphodiester strand scission, resulting in loss of encoded information. Radical species induce strand cleavage by oxidative abstraction of H from the deoxyribose-phosphate backbone and by introducing mutations by oxidation of guanine, the most easily oxidized nucleobase.[85] The potential pathways that introduce DNA damage are diverse and complex, which requires general methods for empirically determining the extent of DNA damage from a set of reaction conditions.

Analytical characterization of DNA integrity during DEL reaction development is typically performed via LCMS, but these measurements cannot directly report whether the DNA remains amplifiable in PCR and intelligible by sequencing. However, analytical methods relying on tag ligation and qPCR now provide a reliable assessment of DNA damage.[54,86] In the first of these approaches, solid-phase DEL reactions are performed in the presence of DNA-functionalized "sensor beads", which are subsequently harvested postsynthesis for qPCR analysis to measure the quantity of amplifiable DNA.[86] Pfizer adapted this approach to the conventional on-DNA DEL synthesis format.[54] Their DNA compatibility assay begins by coupling building blocks to a DNA construct displaying both an appropriate site for coupling and an overhang for enzymatic DNA ligation (Figure 3). LCMS analysis provides reaction yield while enzymatic ligation of a qPCR Taqman probe sequence to the overhang, and then analysis of the product by gel electrophoresis and qPCR provides quantitative assessment of the amplifiable DNA remaining. DNA-compatible reactions maximize the recovery of DNA measured by qPCR.

Using this approach, the Pfizer team demonstrated high yield and DNA compatibility of several common deprotection and coupling reactions. Notably, the conditions demonstrated for Boc-deprotection avoid trifluoroacetic acid in favor of thermal deprotection conditions (250 mM borate buffer, pH 9.5, 18 h, 90 °C).[54,67] Pfizer researchers subsequently implemented the validated amidation and Boc-deprotection reactions in a library

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

setting as discussed above.[72] DNA-damage assays using qPCR are generalizable to other reactions as well, since both Pfizer and GSK have used this assay format to determine the compatibility of photoredox reactions.[69,87] Unfortunately, qPCR cannot directly identify the cause of DNA damage, but it nevertheless provides a rapid platform for hypothesis testing; qPCR assays are readily parallelized (96-well plates) for simultaneous and high-precision analysis of standards, controls, and sample replicates.

Although qPCR readily provides a holistic evaluation of postsynthesis DNA integrity, there are some liabilities. First, mutagenic damage is silent. Some reactions invoke known mutagens (e.g., hydroxylamine[44]) or induce mutagenesis while leaving the DNA intact (e.g., deamination). Incorporation of Sanger sequencing as an additional assay can detect whether mutagenesis has occurred.[44,69] Appropriate encoding language design can also mitigate mutagenesis issues by increasing the genetic distance between sequences in the encoding sets.[62] Second, assigning an exact cutoff point for damage acceptability is difficult, since systematic studies are still lacking. Nonetheless, these quantitative analyses assist planning library synthesis and aid in understanding screening outcomes. Although PCR can amplify single molecules from a library, the notoriously low yield of affinity selection means that 10^4-10^6 amplifiable copies of each library member are needed as input to detect signal after multiple rounds of selection.[88−90] As quantitative assessment of DNA damage by qPCR becomes more widely practiced, correlations between overall DNA fidelity and DEL quality are likely to emerge.

Regardless of the DNA damage threshold for describing useful DEL reactions, developing such chemistry will always be challenging. Reactions must not only be DNA compatible but should also be high yielding for a broad scope of building blocks, enable sequential couplings, and proceed in aqueous conditions, evocative of click chemistry development. In fact, the strong overlap of reaction conditions between click and DEL suggests that click reactions would make excellent starting points for developing DEL-compatible reactions. As mentioned earlier, several biocompatible/click reactions have already been investigated in the DEL setting, including CuAAC,[47,48,57−60,62,63] SuFEx,[35] strain-promoted cycloaddition,[64] and inverse electron demand Diels−Alder reactions.[65] Additionally, several more common click reactions such as Diels−Alder[17] and epoxide ring opening[91] have also been implemented in DEL.

Enzymatic transformations are at the frontier of DEL chemistry development. Enzymatic reactions occur in water with limiting substrate concentrations and, by virtue of enzyme structure, are highly chemoselective (suggesting DNA compatibility). They are also typically high yielding and feature large kinetic rate enhancements. Thus far, however, these advantages have only been explored in a proof-of-concept on-DNA synthesis.[92] The model reaction sequence began with carbamoylation between DNA-NH_2 and nine 2-ethyl-aminoglycosides, followed by modification using one of four enzymes (β-galactosyltransferase, sialyl transferase, trans sialidase, galactose oxidase). Scaling enzyme production for DEL synthesis and expensive cofactors may hinder widespread adoption of the approach. Narrow substrate scope could also render the approach incompatible with the relatively large and structurally diverse building block pools used in DEL. Nonetheless, this initial study sets the stage for future DEL synthesis using biocatalysis and, more broadly, demonstrates that novel linkages and structures are attainable in DEL using reactions that embody virtually all aspects of the click philosophy.

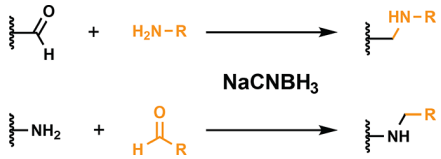
3. REACTION CONSTRAINTS APPLIED TO DEL ANALYSIS

3.1. Practiced Reactions

We have thus far established a connection between click reaction constraints and DEL reaction development to inform and prioritize future DEL design. High-yielding reactions simplify hit deconvolution and reduce false negative rates. The physical properties of the DNA encoding tag strictly constrain reactions to aqueous conditions, which in turn limits the types of reactions that are possible on DNA. Finally, enforcing high chemoselectivity reduces the probability of damaging DNA, an emerging constraint of interest in the field as new DNA analysis methods emerge (qPCR, Sanger sequencing) and integrate with routine measurements of reaction yield (LCMS).

In the following section, we apply the click criteria as a framework to overview the state of the art in DEL reactions and designs. Vipergen recently enumerated the complete set of reactions for DEL.[93] In contrast, we focus our discussion on six widely practiced reactions and their implementation in libraries. Generally, each section describes the reaction's advantages for library diversification, the evolution of reaction conditions to broaden scope and/or utility, and the DNA compatibility of the reaction.

Scheme 2. Amide Bond Formation



3.1.1. Amide Bond Formation. Amide bond forming reactions feature prominently in DEL for their versatility and DNA compatibility. Amidation usually involves the condensation of an amine and a carboxylic acid, two of the largest commercially available building block pools,[94] owing to its popularity in medicinal chemistry.[95] Further, amide chemistry permits ready diversification through heterobifunctional amino acids (another large pool of building blocks), which can be fashioned into both linear and macrocyclic products. Structurally diverse diamines and diacids that are amenable to decoration via amide chemistry provide additional strategies for accessing novel chemical space.[1] Finally, the routinely high yields and DNA compatibility of this reaction meet many of the click criteria.

Amine-terminated headpiece DNA is the most common starting material for DEL synthesis,[39] so amidation with functionalized carboxylic acid building blocks (Fmoc-amine, Boc-amine, aryl halide, etc.) is a nearly ubiquitous first step of library synthesis.[43,93,96] Many libraries have also relied on amide formation as a critical diversification reaction.[53,72] Activating agents are necessary reactants for converting carboxylic acids to active esters to promote condensation with amines. There is conflicting evidence for which activation conditions provide the highest average yields while minimizing DNA damage. Neri performed a thorough analysis of eight different activating conditions for their ability to couple carboxylic acids to on-DNA

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

amines, finding that the combination of EDC/HOAt/DIPEA outperformed all others.[97] However, many DEL groups routinely use DMT-MM for modification of DNA-conjugated amines[49,72] or other coupling reagents.[54,98] Reaction conditions (organic cosolvent, buffer composition, pH, DNA concentration, and building block selection) varied widely between these studies, possibly explaining the diversity in optimal conditions.

While amidation of on-DNA amines tends to employ basic buffer conditions (pH 8−9.5), amidation of on-DNA carboxylic acids tends to employ acidic buffer conditions.[1,98,99] DMT-MM is the dominant coupling reagent for this reaction. Gillingham's recent comparison of reaction conditions for acylation with 126 amino acids determined that DMT-MM outperformed EDC/HOAt/DIPEA for coupling amino acids to on-DNA carboxylic acids.[47]

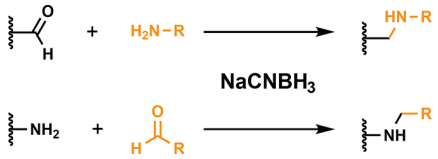
Building block validation is critical for chemistries like amidation that use deep, structurally diverse building block pools. To give a sense of scale, some commercial suppliers offer >28,000 carboxylic acids alone. Many pharmaceutical companies also curate internal, proprietary building block collections.[100] As a consequence, broad and novel chemical functionalities (aromatic, cyclic, heterocyclic, aliphatic, bridged, etc.) can be added to a library using robust acylation reaction conditions. Running contrary to the widespread notion that amides are an intrinsic liability, both DEL compounds in late stage clinical trials feature amides.[19,50−53] The RIP1K candidate consists of a benzoxazipinone linked by an amide bond to a benzyl-isoxazole,[53] and the parent DEL of the sEH candidate was synthesized by nucleophilic aromatic substitution of cyanuric chloride with amines/amino acids followed by acylation.[51]

The high yield and broad reaction scope of amidations have led to the success of these and other library screens by tailoring amide bond forming reactions to aqueous reaction conditions. Outside of DEL, amide bond coupling reagent development has received significant attention within the chemical synthesis community,[101,102] but these chemistries are not usually designed to accommodate aqueous reaction conditions. DEL addresses the challenge of aqueous chemistry by most often employing EDC or DMT-MM as activating agents, which likely succeed in aqueous conditions, since they are water-soluble cationic salts.

Critically, amide formation occurs under mild conditions and does not promote amidation of DNA's numerous nucleophilic sites. Amide formation does not require extreme pH, mutagenic reagents, metal catalysts, or high temperature. Moreover, qPCR analysis of recovered DNA has shown that amidation conditions minimally affect DNA amplifiability. Under aqueous reaction conditions, Stress et al. found that acylation of dsDNA with acids or amino acids activated by DMT-MM leaves 80% amplifiable DNA remaining,[47] and Ratnayake et al. found that acylation of headpiece DNA using HATU/DIPEA leaves the DNA completely unharmed.[54] Similarly, numerous DNA-compatible conditions now exist for removing common protecting groups associated with amidation, such as Fmoc, Boc, and azides.[54,72,86] Thus, between the array of coupling reagents and orthogonal nucleophile protection strategies available, amide formation represents the best-developed reaction class in DEL.

3.1.2. Reductive Amination/Alkylation. Reductive amination is another very popular reaction in DEL. It is one of the 10 most frequent transformations in traditional medicinal chemistry[95] and was one of the first reactions to be demonstrated for DNA-templated synthesis,[105] foretelling its utility in DEL. This reaction furnishes the C−N bonds found in many drugs[110] and is operationally simple, making it attractive for library preparation. Reductive amination reactions are generally high yielding and exhibit a broad scope for both amine and aldehyde reaction partners (Table 2). While aqueous solvent is traditionally avoided for reductive amination,[111] reaction adaptation has allowed facile implementation in water at library scale. Finally, in the context of DEL, this reaction is reasonably tolerant of additional functional groups and reaction conditions are DNA compatible.

Scheme 3. Reductive Amination/Alkylation



Reductive amination/alkylation generally occurs as a sequential one-pot reaction. In the first step, a primary or secondary amine reacts reversibly with a carbonyl, resulting in loss of water and concomitant imine or imminium ion formation. This species is subsequently reduced in the second step—frequently with borohydrides—rendering the reaction irreversible through formation of the desired C−N bond. Reagent selection is important because strong reducing agents reduce aldehydes and ketones, leading to unproductive alcohol synthesis. In more traditional organic synthesis, several reducing agents ($NaCNBH_3$, $NaBH_4$, $NaBH(OAc)_3$, etc.) are used in slight excess (2−3 equiv),[111] but DEL has thus far almost exclusively used $NaCNBH_3$ and in large excess (Table 2). The reaction is quite attractive for DEL due to the sheer quantity of commercially available amines and aldehydes, which also tend to be the most economical.[94]

Reductive amination of on-DNA aldehydes has been a highly fruitful coupling mode for DEL synthesis. This direction of the reductive amination has been implemented in published work four times.[46,63,92,112] The largest amine sets included 2259 and 2341 amines,[63,112] but these publications did not describe building block validation or structural diversity. Of the published reductive amination reactions, GSK disclosed the most comprehensive scope for this transformation with an evaluation of 813 amines coupling to a model benzaldehyde DNA substrate.[46] Under the specified reaction conditions, 216 amines achieved yields >70% and were subsequently included in the library. Generally, primary amines coupled more efficiently than secondary amines and aromatic amines coupled more efficiently than aliphatic amines. However, the majority of amines did not couple efficiently, again highlighting the importance of building block validation.

Reductive alkylation of on-DNA amines has similarly been a productive reaction for DEL synthesis, leveraging abundant commercial collections of amino acids and aldehydes. To avoid bis-alkylation during library synthesis, on-DNA amine reactants are often restricted to secondary amines, though conditions for monoalkylation of primary amines have been described.[67] Conditions for reductive alkylation of the secondary amine of proline-conjugated DNA were disclosed in the initial patent describing DEL filed by Praecis Pharmaceuticals in 2004.[108] Since then, reductive alkylation of on-DNA amines in library synthesis has been published twice in recent reports from the Schreiber lab[44] and X-Chem pharmaceuticals/Arrakis therapeutics.[113] Schreiber's conditions achieved a broad scope,

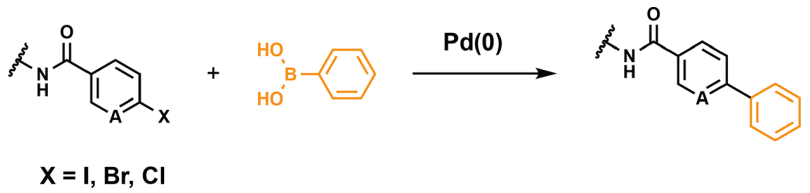
https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Table 2. Reductive Amination Conditions

reducing agent	conditions	scope investigated	library	year
$NaCNBH_3$	60 nM NH_2-DNA, 60 nM CHO-DNA, 3 mM $NaCNBH_3$, 0.1 M MES pH 6.0, 1.5 h, 25 °C	1 NH_2-DNA, 1 CHO-DNA, or 1 glyoxal-DNA		2002[105]
$NaCNBH_3$	0.76 mM proline-DNA, 30 mM RCHO/$NaCNBH_3$, 75% 150 mM phosphate pH 5.5, 25% DMF, 2 h, 80 °C or 1 mM CHO-DNA, 30 mM RNH_2/$NaCNBH_3$, 2 h, 80 °C	proline-DNA and CHO-DNA		2005[108]
$NaCNBH_3$[a]	50 μM CHO-DNA, 500 mM amine/$NaCNBH_3$, 300 mM MOPS, pH 7.4, 16 h, 37 °C	12 amines		2014[116]
$NaCNBH_3$	0.63 mM CHO-DNA, 50 mM amine/$NaCNBH_3$, 62.5% 500 mM phosphate pH 5.5, 12.5% H_2O/MeCN/DMF, 16 h, 60 °C	2259 amines (no validation)	3.3×10^8	2015[63]
$NaCNBH_3$	0.5 mM CHO-DNA, 50 mM amine/$NaCNBH_3$, 50% 250 mM phosphate pH 5.5, 25% DMA/ACN, 16 h, RT	218/831 amines (yield >70%)	3.5×10^6	2015,[117] 2016[46]
$NaCNBH_3$[a]	0.63 mM NH_2-DNA, 50 mM R-CHO/$NaCNBH_3$, 62.5%, 25% DMF, 12.5% DMA 250 mM phosphate pH 5.5, 16 h, 60 °C	636 R-CHO in library, >50% yield in validation with 1° and 2° amine-DNA substrates	3.4×10^7	2015[45]
$NaCNBH_3$[b]	resin-bound 2° amine, (1) 0.5 M R-CHO, 1% AcOH in DMF, 10 min, RT; (2) 0.5 M $NaCNBH_3$, 1% AcOH in MeOH, 1 h, RT	4-iodobenzaldehyde		2016[86]
$NaCNBH_3$[c]	33 μM carbohydrate/CHO-DNA, 333 mM amine/$NaCNBH_3$, 16 h, 37 °C	1 carbohydrate/CHO-DNA with 19 RNH_2, 7 carbohydrate/CHO-DNA with benzylamine		2017[92]
$NaCNBH_3$[a]	0.19 mM 2° amine-DNA, 37.5 mM RCHO/$NaCNBH_3$, 25% $NaHCO_3$, 37.5% 1 M phosphate pH 4.2/DMF, 8 h, 37 °C	4 NR_2H-DNA with 20 aldehydes, 1 NR_2H-DNA with 118 R-CHO (72/118 >70% yield)	1.07×10^6	2019[44]
$NaCNBH_3$[a]	500 mM ketone, 1 M $NaCNBH_3$, 400 mM $B(OH)_3$, 80% NMP, 20% H_2O, 20 h, 60 °C	17 ketones		2019[114]
$NaCNBH_3$	100 equiv of RCHO, 100 equiv of $NaCNBH_3$, O/N, 60 °C	1347 aldehydes	1.9×10^8	2020[72]

[a]Analysis of DNA conjugates by HPLC(A260) and LCMS. [b]Quantitative analysis of DNA compatibility by qPCR. [c]Analysis of DNA conjugates by MALDI-TOF.

Scheme 4. Suzuki−Miyaura Cross-Coupling



coupling 72 out of 117 aldehydes with an azetidine-DNA substrate in >70% yield (previous studies coupling 20 aldehydes per scaffold showed little reactivity difference between azetidine- and pyrolidine-DNA). However, 22 aldehydes were entirely unreactive, suggesting that certain classes of aldehydes are not suitable for this chemistry. In the other reported application, Litovchik et al. reacted 85 aldehydes, a similar sized building block set, with 1024 secondary amine DNA conjugates, but building block validation data were not provided. Taken together, these studies substantiate the reaction of aldehydes with on-DNA secondary amines as useful for DEL synthesis.

Expanding the reductive alkylation reaction scope to include addition of ketones would be highly beneficial for generating Csp^3 amine bonds. Scripps Research/Pfizer discovered aqueous phase conditions for reductive alkylation of on-DNA amines with ketones.[114] Nine of 14 ketones coupled with a primary amine at yields >70%. Boric acid was key for this transformation, but the reaction also employed significantly higher concentrations of ketone and $NaCNBH_3$ (500 mM) than similar transformations with aldehyde building blocks (30−50 mM, see Table 2). Overall, the reaction scope will require further study to confirm suitability for library synthesis.

In addition to the high yields achievable by reductive C−N bond formation in aqueous solution, these reactions are also orthogonal to other reactions, thereby facilitating more complex library design. Multifunctional building blocks can combine amidation reactions, reductive alkylation/amination, and Suzuki couplings. Recently, small collections of trifunctional building blocks have become commercially available for this purpose (Enamine), but so far, library synthesis with trifunctional building blocks has incorporated only custom building blocks. In all cases, the order of reaction is important. Schreiber suggested that reductive alkylation should precede Suzuki coupling to avoid side reactions with the on-DNA amine, while GSK suggested that Suzuki coupling should precede reductive amination to avoid dehalogenation.[46] The relative orthogonality of reductive amination enables incorporation of other chemistries as well. For example, a recent library featured chemoenzymatic installation of aldehyde-labeled sugars, which were subsequently modified by reductive amination.[92]

Like amide formation, reductive C−N bond formation generally invokes reactivity that is orthogonal to DNA and is another feature that makes the chemistry attractive for DEL. Quantitative analysis of the DNA compatibility of this reaction is limited to solid-phase DEL synthesis methods, but those studies indicated that there is no detectable effect on DNA amplifiability.[86] Even for untested reaction conditions, though, they are relatively mild, suggesting compatibility. For example, although the reaction usually proceeds at pH 5.5, this is not sufficiently acidic to promote depurination, which proceeds only sluggishy at pH > 5.0.[115] Similarly, one would not predict redox cross-reactivity of DNA with the commonly used borohydride

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

reductants. Finally, the reaction occurs at mild temperature as well, with most <60 °C. Altogether, reductive amination adheres to the suggested reaction criteria for DEL.

3.1.3. Suzuki−Miyaura Cross-Coupling. Since its discovery in 1981, the Suzuki−Miyaura cross-coupling reaction has become one of the most popular reactions in medicinal chemistry[95] and it received a citation for the 2010 Nobel Prize in chemistry. Approximately 30 drugs contain biaryl C−C linkages,[118] suggesting that this reaction has high potential for generating suitable chemical matter. While designing Suzuki couplings that are compatible with DEL synthesis is particularly challenging, extensive reaction optimization has yielded improved scope and the reaction has been implemented in several libraries.

Among the now many DEL-compatible chemistries, the Suzuki cross-coupling has been unique in its ability to deliver sufficiently robust C−C bond construction as to be useful in a library setting. Initially, Suzuki reactions were implemented in DEL using $Pd(PPh_3)_4$ as the catalyst and the scope was limited to the reaction of on-DNA aryl iodides (primarily) or pyrimidinyl bromides (far less explored) with excess boronates to yield the corresponding biaryl linkage (Table 3, entries 3, 4, and 6).[45,46,63,119] Alternative Pd ligands explored recently have provided higher yields for aryl iodide building blocks,[44] and some ligands catalyze cross-coupling with challenging aryl chloride-DNA substrates,[46,120] and other coupling partners,[121,122] expanding the scope of Suzuki cross-couping in DEL.

The utility of the Suzuki reaction in library synthesis derives from its relatively broad substrate scope and its orthogonality with other common bond construction strategies. The first disclosed DEL[45] that used Suzuki coupling was a three-cycle library (Table 3, entry 4) designed around 44 trifunctional building blocks (Fmoc-protected amine, aryl iodide, carboxylic acid). These trifunctional building blocks were used to amidate NH_2-DNA and were then further elaborated in cycles two and three by Suzuki coupling with 265 boronic acids followed by amine capping reactions with 2976 electrophilic building blocks (carboxylic acids, sulfonyl chlorides, aldehydes, isocyanates, and heteroaryl chlorides). A screen for BCATm inhibitors using this 34.7-million-member Suzuki DEL yielded an optimized hit compound with IC_{50} = 2 μM in an enzyme activity assay. A second DEL from GSK used these Suzuki conditions in conjunction with trifunctional aldehydes and reductive amination[46] to yield a 3.5-million-member DEL of biaryls. X-Chem reported similar reaction conditions for DEL synthesis, coupling 222 bromoaryl carboxylates by Suzuki with 667 boronic acids, but yields for this transformation were not reported.[63]

Further studies of Suzuki reactions in on-DNA synthesis expanded the reaction scope to other aryl halide species for accessing other building block pools. Researchers at GSK discovered that precombining POPd with the Buchwald sSPhos precatalyst (1:2) enabled cross-coupling to challenging pyridinyl and aryl chlorides.[123] To improve operational simplicity, Li and Huang reported that the commercially available precatalyst sSPhosPd G2 also catalyzed the aryl chloride cross-coupling reaction.[120] In this work, eight Buchwald precatalysts were tested, and only sSPhosPd G2 provided appreciable yield, possibly due to its water solubility derived from sulfation; the parent SPhosPd G2 compound lacking sulfation (but otherwise identical) did not catalyze the reaction. The optimized conditions using the sSPhosPd G2 catalyst were evaluated in a substrate scope study of 6 aryl/heteroaryl chloride, 1 heteroaryl bromide, and 1 heteroaryl iodide headpiece coupled with 84

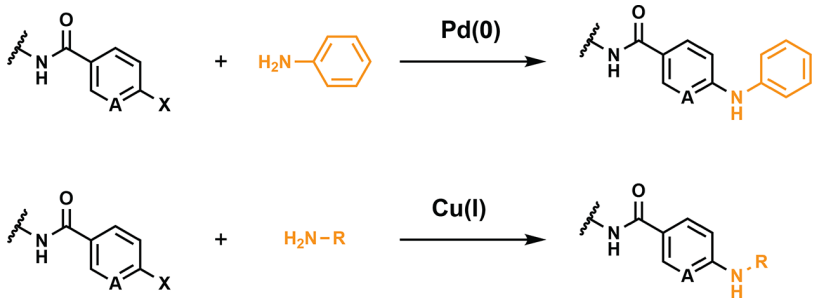
Table 3. Suzuki−Miyaura Cross-Coupling Conditions

catalyst	conditions	scope investigated	library	year
$Pd(PPh_3)_4$[a]	1 mM ArI-DNA, 20 mM Ar-$B(OR)_2$,[b] 1 mM $Pd(PPh_3)_4$, 40 mM Na_2CO_3, 90 min, 80 °C	15 R-$B(OH)_2$, 8 boronic esters, 2 ArI-DNA, 1 ArBr-DNA, 3 PyBr-DNA		2015[119]
$Pd(PPh_3)_4$[a]	0.73 mM ArI-DNA, 30 mM Ar-$B(OR)_2$,[b] 60 mM Na_2CO_3, 0.4 mM $Pd(PPh_3)_4$, 5 h, 80 °C	44 ArI (>70% yield in two-step validation) and 265 Ar-$B(OR)_2$ (>50% yield in validation)	3.4×10^7	2015[45]
$Pd(PPh_3)_4$[a]	0.5 mM ArI-DNA, 50 mM Cs_2CO_3, 40 mM Ar-$B(OR)_2$,[b] 0.6 mM $Pd(PPh_3)_4$, 75% aqueous, 25% DMA, 3 h, 80 °C; 15 mM $NaS_2CN(C_2H_5)_2$ Pd scavenger	222 ArBr, 667 Ar-$B(OR)_2$	3.3×10^8	2015[63]
$Pd(PPh_3)_4$[c]	resin-bound ArI, 344 mM Ar-$B(OH)_2$, 690 mM DIPEA, 0.13 mM $Pd(PPh_3)_4$, NMP, 7 h, 70 °C	4-(iPr)PhB$(OH)_2$		2016[86]
POPd/sSPHOS ($Pd_2(dba)_3$, POPd1, POPd2)[a]	0.83 mM (het)ArCl-DNA, 33 mM KOH, 17 mM Ar-$B(OR)_2$,[b] 1.7 mM sSPHOS, 0.83 mM POPd, 3 h, 80 °C	3 (het)ArCl-DNA with 8 Ar-$B(OR)_2$ each		2016[123]
sSPhos-Pd-G2[a]	0.42 mM ArX-DNA, 170 mM CsOH, 42 mM Ar-$B(OR)_2$,[b] 0.84 mM sSPhos-Pd-G2, 81% H_2O, 11% dioxane, 8% DMA, 15 min, 80 °C	8 (het)ArX-DNA with 84 Ar-$B(OR)_2$ each		2018[120]
$Pd(OAc)_2$[a]	0.19 mM $ArSO_2$F-DNA, 77 mM Ar-$B(OH)_2$, 192 mM TEA, 3.8 mM $Pd(OAc)_2$, 60% H_2O, 40% dioxane, 2 h, 25 °C; $NaS_2CN(C_2H_5)_2$ Pd scavenger	33 Ar-$B(OH)_2$ with $ArSO_2$F-DNA; 8 (het)ArX-DNA with PhB$(OH)_2$		2019[121]
$Pd(dppf)Cl_2$[a]	0.13 mM ArI-DNA, 25 mM Ar-$B(OH)_2$, 0.15 mM $Pd(dppf)Cl_2$, 20/51/13/16 100 mM carbonate pH 8.2/1 M phosphate pH 9.2/EtOH/MeCN, 1 h, 90 °C	241 Ar-$B(OH)_2$ (yield 117/241 >85%)	10^6	2019[44]

[a]Analysis of DNA conjugates by HPLC(A260) and LCMS. [b]$(OR)_2$ = $(OH)_2$ or pinacol. [c]Quantitative analysis of DNA compatibility by qPCR.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Scheme 5. Buchwald and Ullmann Cross-Coupling



boronates. For all headpieces tested, >60% of the building blocks coupled in >50% yield, but some boronates (e.g., sterically hindered, chlorinated, fluorinated) remained low yielding. Finally, a fluorosulfonate electrophile (accessible via phenols) enabled more permissive coupling at lower temperature under ligand-free conditions.[121]

Despite improvements in ligands and other reaction parameters, Suzuki coupling conditions are among the more demanding and can be quite detrimental for DNA. Suzuki reactions are commonly alkaline and require both high temperature and metal catalyst. All of these are depurination hazards, and indeed, X-Chem observed that library recovery following Suzuki coupling was 3-fold lower than the other reactions used to generate this library.[63] Similarly, our laboratory observed that Suzuki coupling catalyzed by $Pd(PPh_3)_4$ yielded only 30% amplifiable DNA after the reaction.[86] The potential for Pd species to induce DNA damage is well appreciated: several unique library purification strategies have emerged from Pd-mediated reaction development, such as spin filtration,[45] centrifugation,[44] and addition of metal scavengers, including sodium sulfide[67] and sodium diethyldiothiocarbamate.[119] Additionally, ligation analysis[45] and Sanger sequencing[120] data have provided further characterization of the Suzuki reaction's DNA compatibility. Collectively, Suzuki−Miyaura cross-coupling is highly advantageous for increasing druglikeness of DELs, but caution is necessary, as common reaction conditions compromise DNA fidelity.

3.1.4. Buchwald−Hartwig and Ullmann Cross-Coupling. The rise of Suzuki coupling in DEL has driven strong parallel interest in C−N cross-coupling reactions by way of Buchwald−Hartwig (Pd-catalyzed) or Ullmann type couplings (Cu-catalyzed). The Buchwald and Ullmann reactions forge aryl C−N bonds, which occupy chemical space distinct from C−N bonds formed by reductive amination. Buchwald coupling is a particularly popular transformation in medicinal chemistry for its simplicity, the prevalence of C−N bonds in natural products, and the pharmacokinetic versatility of the secondary amine linkage. From a combinatorial chemistry perspective, C−N cross-coupling reactions are more attractive than C−C Suzuki cross-coupling reactions because amines are much more commercially abundant than aryl boronates and they are substrates for a wider array of reactions. However, the small reaction scale in DEL offers opportunities for exploring complementary C−N and C−C cross-couplings, modularly increasing library diversity.[98,124]

Despite the appeal of C−N cross-coupling for DEL diversification, discovering suitable reaction conditions has proven difficult. GSK's initial disclosure of C−N cross-coupling for DNA-linked substrates included both Buchwald- and Ullmann-type reactions with large amine sets, but yields were generally low.[124] The Buchwald reaction, employing tBuXPhos precatalyst G1, was evaluated by coupling 6329 primary aromatic amines to a DNA-linked aryl iodide, but for 93% of building blocks, the yield was <50%, highlighting the challenge of achieving broad scope. The Cu(I)-catalyzed Ullmann reaction proved fruitful using primary aliphatic amines and amino acids, but yields were similarly low. Although these initial Buchwald and Ullmann conditions yielded a narrow building block scope, both were used for library construction (screening yet to be disclosed), paving the way for future reaction development.

The first major improvement to Buchwald coupling in DEL stemmed from the use of third-generation Buchwald catalysts. Incorporating these catalysts[125,126] in DEL[49,98,121] expanded the amine and aryl halide scope. In the largest study, Eli Lilly found that 310/867 primary aromatic amines coupled in >70% conversion with a model aryl bromide DNA conjugate. This compares favorably with GSK's previous conditions that required higher temperature, more base, and more reactive aryl iodide substrate to identify only 177/6329 primary aromatic amines coupling in >70% yield. Direct comparison of six reactions under both conditions uniformly proceeded in higher yield using Lilly's method. Lilly's conditions also expanded the aryl halide scope to include several aryl bromide building blocks. In a two-step validation of 471 aryl bromide carboxylates, 225 acylated DNA in >70% yield and 105 of these underwent further Buchwald coupling in >70% yield. The utility of the developed conditions was demonstrated when Lilly synthesized a Buchwald DEL that was used in over 140 selection experiments.

More recent optimization of Buchwald for DEL improved the reaction scope through use of a pyridine-enhanced precatalyst (PEPPSI) activated by ascorbate.[127] The optimized conditions required careful PEPPSI catalyst selection, high temperature (95 °C), and DMF cosolvent but expanded both the amine and aryl halide scope. While a small number of aryl chloride couplings suggested the potential suitability of these building blocks, the major focus of the work was identifying suitable aniline and secondary amine coupling partners. To this end, 197/328 and 123/292 aniline building blocks coupled to aryl bromide headpiece DNA and pyridinyl bromide headpiece DNA, respectively. Secondary amines remained challenging with 23/92 and 15/92 coupling in >70% yield to the same aryl bromide and pyridinyl bromide headpieces, respectively. The aryl halide scope investigation was more limited, but the newly developed reaction conditions were nevertheless applied to the synthesis of a 6×10^7 member library featuring 165 aryl halides in the second cycle followed by 386 anilines and 92 cyclic secondary amines in the third cycle.

The Ullmann coupling remains the most promising reaction for coupling aryl halides with aliphatic amines. Once again,

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

ligand optimization was critical to enabling Ullmann-type couplings for DEL. Whereas commercial catalysts drove Buchwald coupling, Novartis tested 8 potential ligands and then synthesized 13 derivatives to arrive at an optimal ligand.[68] Using this ligand and optimized conditions, 5/8 aliphatic amines coupled to the aryl iodide headpiece in >70% yield. In another study, 8 aryl iodide headpiece DNA substrates were reacted with 12 aliphatic amines. Both aryl iodides with ring-adjacent substituents and sterically hindered amines uniformly coupled poorly, but sterically unhindered aryl iodides and amines generally coupled in >70% yield. While the newly designed catalyst and optimized conditions have not yet been applied to library synthesis, they again highlighted the importance of catalyst selection.

Although Buchwald and Ullmann reaction conditions often resemble those of Suzuki couplings, C−N couplings are in several instances less damaging to DNA. For example, Ruff and Berst found that Ullmann coupling (Table 4, entry 2) leaves 65% amplifiable DNA,[68] while Ratnayake et al. found Buchwald (Table 4, entry 6) coupling left 73% amplifiable DNA by qPCR.[54] In the absence of quantitative data, the DNA compatibility of reactions employing PEPPSI precatalyst remains speculative. The higher temperature and increased basicity under PEPPSI precatalyst conditions (95 °C) are likely to increase DNA damage, warranting further study. As in Suzuki couplings, Pd species in Buchwald couplings may be scavenged post reaction with thiol-containing compounds.[49] Cu may be scavenged with EDTA.[68]

Together with Suzuki, Buchwald and Ullmann reactions comprise a suite of highly valuable cross-couplings that use commercially available and deep building block pools to access chemical matter of high druglikeness. The scope has rapidly and dramatically improved, both increasing the diversification potential and allowing the DEL designer to pull from a collection of aryl halides to construct either C−N or C−C bonds. Finally, while the DNA compatibility appears suitable for all of these metal-mediated couplings, arriving at these conditions has been a multivariate battle to enhance reaction kinetics through proper catalyst selection while minimizing reaction time. Metal scavenging has also played a crucial role and one that has not been fully characterized to date. Further quantitative studies, particularly of Ullmann, will be helpful in establishing a click-like set of cross-coupling conditions for the field.

3.1.5. Triazine Substitution. Nucleophilic displacement on the cyanuric chloride scaffold has yielded diverse published DELs. The seminal disclosure of DEL[1] showcased this scaffold and chemistry sequence, which uses economical and plentiful amines in three cycles of chemistry for readily accessing numerically large libraries (10^6−10^8). Libraries employing this chemistry have furnished several novel inhibitors,[1,128−131] including the lead for GSK's sEH inhibitor.[51] Although computational analysis identified potential liabilities,[38] the cyanuric chloride scaffold is still a logical starting material for creating branched libraries.

The simplicity of cyanuric chloride functionalization is a major aspect of its appeal for library synthesis. Cyanuric chloride is a trifunctional scaffold elaborated through symmetric and dynamic site reactivity[132] instead of protecting groups or orthogonal functionalities in other DEL scaffolds. Synthesis of the Praecis/GSK triazine "DEL A" entailed mild initial nucleophilic chloride substitution (1 h, 4 °C), a longer second substitution on the less-activated ring system (16 h, 4 °C), and

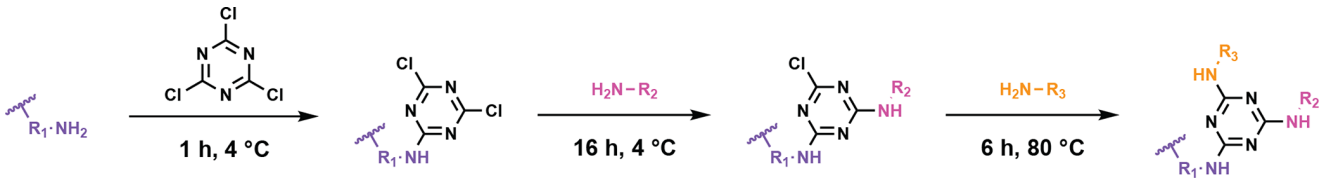
Table 4. Buchwald−Hartwig and Ullmann C−N Cross-Coupling Conditions

catalyst	conditions	scope investigated	library	year
tBuXPhos Pd G1[a]	0.4 mM ArI-DNA, 400 mM $ArNH_2$, 400 mM CsOH, 0.8 mM tBuxPhos Pd G1, 48% aq, 52% DMA, 3 h, 100 °C	1033/6329 1° $ArNH_2$ >50% yield	1.7×10^8	2017[124]
$CuSO_4$, ascorbate, (proline)[a]	0.38 mM ArI-DNA, 280 mM 2° amine, 11 mM $CuSO_4$/proline/KOH, 14 mM ascorbate, 2 h, 100 °C	96/557 AAs (yield >50%); 300/2776 aliphatic 1° RNH_2 (>50% yield)	3×10^7	
$Cu(OAC)_2$, ascorbate, 2-((2,6-dimethoxyphenyl)amino)-2-oxoacetic acid[a]	62.5 μM ArI-DNA, 500 mM amine, 25 mM $Cu(OAC)_2$, 50 mM ascorbate, 200 mM ligand, 500 mM K_3PO_4, 1:3 DMSO:H_2O, 3 h, 40 °C	8 ArI-DNA with 12 2° amines		2018[68]
tBuBrettPhos Pd G3[a]	0.31 mM $ArSO_2F$-DNA, 62.5 mM $ArNH_2$, 312.5 mM TEA, 3.1 mM tBuBrettPhos Pd G3	1 $ArSO_2F$-DNA with 24 $ArNH_2$ and 8 $ArSO_2F$-DNA with aniline		2019[121]
BrettPhos Pd G3[a]	0.25 mM ArX-DNA, 50 mM aniline, 1.25 mM BrettPhos Pd G3, 125 mM CsOH, 50/50 H_2O/1-methoxy-2-propanol, 20 min, 80 °C; cysteine Pd scavenger	222 anilines (yield not provided)	7.5×10^7	2019[98]
tBuXPhos Pd G3[a]	0.74 mM ArX-DNA, 112 mM $ArNH_2$, 112 mM NaOH, 11 mM tBuXPhos Pd G3, 2 h, 60 °C	867 $ArNH_2$ with 1 ArI-DNA and 1 ArBr-DNA, 471 ArBr-DNA with 3-fluoro aniline	4×10^7	2019[49]
tBuXPhos Pd G3[a,b]	0.8 mM ArI-DNA, 63 mM amine, 230 mM NaOH, 1 mM tBuXPhos d G3, 1 h, 80 °C			2019[54]
Pd-PEPPSI-iPent^{Cl}, ascorbate[a]	0.25 mM ArX-DNA, 250 mM CsOH, 5 mM ascorbate, 0.5 mM Pd-PEPPSI-iPent^{Cl}, 15 min, 95 °C; cysteine Pd scavenger	ArBr-DNA with 328 anilines and 92 2° amines; PyBr-DNA with 292 anilines and 92 2° amines	6.2×10^7	2020[127]

[a]Analysis of DNA conjugates by HPLC(A260) and LCMS. [b]Quantitative analysis of DNA compatibility by qPCR.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Scheme 6. Nucleophilic Substitution



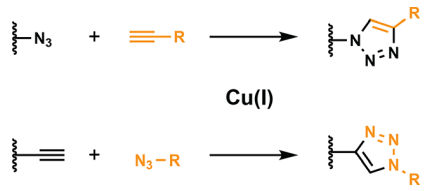
an aggressive third substitution on the most deactivated ring system (6 h, 80 °C).[1]

The relatively large scope of nucleophiles that reacts with the cyanuric chloride scaffold is also a highly attractive feature of the chemistry. Competent nucleophile classes include amines, thiols, and alcohols,[132] but amines predominate for their utility in several different DEL chemistries. GSK identified 340/1000 amines that performed at least one step of substitution in >70% yield. This reaction tolerated most classes of primary and secondary amines including aliphatic, aromatic, or cyclic amines. Substitution with phenolic acids, amino acids, or diamines introduces opportunities for a fourth cycle, increasing library size but reducing library member druglikeness.[1,38] Incorporation of thiol nucleophiles or using the final aryl chloride in a Suzuki coupling poses alternative routes for exploiting the versatility of the cyanuric chloride scaffold while retaining druglike character.[133]

Despite the common synthesis of libraries focused around cyanuric chloride, a quantitative understanding of the DNA compatibility of the associated S_NAr reactions remains lacking. For the first substitution, electrophilic cyanuric chloride reacts with amine-functionalized DNA to introduce the core library structure but has potential for off target reactivity with nucleobase amines. The second and third nucleophilic substitutions appear mild, requiring low amine concentrations and temperatures <80 °C. In comparison, Buchwald coupling with higher amine concentration, strong base, and metal catalyst leaves 73% amplifiable DNA.

The range of lead compounds discovered from triazine libraries is remarkable. Examples include metalloprotease inhibitors exhibiting high selectivity between highly homologous enzymes,[128,129] an LFA1-ICAM1 protein−protein interaction inhibitor,[130] and an OXA-48 β-lactamase inhibitor[131] to highlight a few target class firsts for DEL screening. Finally, the sEH inhibitor in late stage clinical trials exemplifies a few good reactions (amidation, cyanuric chloride substitution) enabling drug discovery through DEL.[51]

Scheme 7. CuAAC



3.1.6. Copper-Catalyzed Azide Alkyne Cycloaddition.

The quintessential click reaction, CuAAC, has served myriad purposes in encoded libraries. It was first shown in a DNA-templated reaction[57] and has since been deployed in DEL for introducing diversity elements,[47,48,58,61] macrocyclization,[59,60] coupling DNA to resin for solid-phase DEL synthesis,[62] and even chemical ligation of encoding tags in place of enzymatic ligation[63] (Table 5). CuACC is altogether convenient for DEL syntheses due to its facile reaction setup, broad scope, aqueous compatibility, and chemoselectivity, but commercial abundance of azides and alkynes is low relative to most other building block classes.

CuAAC in DELs employs conditions similar to those of bioconjugation. Most often, in situ reduction of Cu(II) with ascorbate furnishes the active Cu(I) catalyst, which tris-triazole ligands subsequently stabilize. The reaction benefits from sealing after gentle sparging with inert gas to protect the catalyst from decomposition by reaction with O_2. As is the case for bioconjugation,[22] an excess of coupling reagent and gentle heating (45−60 °C) enhance reaction kinetics. Ligand and solvent optimization generally maximizes reaction rate[134−136] while minimizing oxidative damage to off targets.[137] While the majority of DEL applications have used TBTA as the ligand, Neri recently explored a more water-soluble TBTA derivative[138] and further ligand exploration may prove fruitful.

The chemoselectivity of CuAAC has prompted several innovative applications in DEL beyond just building block coupling. For example, CuAAC has been used for encoding tag ligation in place of enzymatic methods. X-Chem described library synthesis starting from ssDNA functionalized with a 5′-amine for diversification and a 3′-propargyl group for CuAAC ligation of encoding tags. Additional ssDNA oligonucleotides functionalized with a 3′-silane-protected propargyl group and a 5′-azide[63,139] enabled efficient chemical tag ligation. The Klenow fragment could read through DNA triazole linkages, but efficiency was low. This strategy enabled split-and-pool library synthesis on ssDNA with encoding tag ligation by CuAAC to yield a 334-million-member DEL. Selection of this library identified a potent inhibitor of sEH (IC_{50} = 2 nM). As an additional example of CuAAC enabling novel library design through orthogonality, it has been the reaction of choice for attaching the DNA headpiece to resin for solid-phase DEL synthesis (see section 3.2.2 below).[62]

Macrocyclization is another important application showcasing the versatility of the CuAAC in DEL design. GSK created a 6-mer macrocyclic peptide library with theoretical diversity of 2.3×10^{12} peptide backbones cyclized via CuAAC.[59] This application depended on the high yield and orthogonality of CuAAC. Selection patterns indicated that macrocylic library members were more potent ligands than linear counterparts obtained from an aliquot of the DEL that was not subjected to CuAAC. A similar strategy using CuAAC to cyclize scaffolds was also applied to DNA-recorded peptoid macrocycle libraries.[60] While acylation was the chosen reaction for macrocylization in a recent library from Stress et al., CuAAC was critical for attaching 21 non-peptidic macrocycle precursors to encoding tags.[47]

Given its roots as a library diversification strategy, CuAAC has appeared in several DEL syntheses. In one example, aliphatic and benzylic halides were converted in situ to azides by nucleophilic substitution with NaN_3 and then coupled to propargyl glycine, pyrimidine, and benzodiazepine scaffolds.[58]

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Table 5. CuAAC Conditions

catalyst	conditions	scope investigated	library	year
CuCl	2 μM alkyne DNA, 100 mM RN_3, 10 mM CuCl, 9:1 MeCN:H_2O	1 alkyne-DNA, 2 R-N_3		2011[57]
TBTA, $CuSO_4$, ascorbate	0.5 mM alkyne-DNA, 0.5 mM N_3-DNA, 1 mM $CuSO_4$, 2 mM ascorbate, 0.5 mM TBTA, O/N, RT	3′-propargyl-DNA and 5′-N_3-DNA	3.3×10^8	2015[63]
TBTA, $CuSO_4$, ascorbate	solid-phase: 40 mg of alkyne-resin (17.2 μmol of resin, 4 mM sites), 16 μM N_3-DNA, 4.7 mM $CuSO_4$, 23 mM ascorbic acid, 8 μM TBTA, 50% DMSO, 30 mM TEAA, pH 7.5, 0.04% Tween 20, 4 h, 40°	resin-bound alkyne, N_3-DNA		2015[62]
TBTA, CuBr[a]	0.35 mM N_3-DNA, 2.9 mM CuBr, 11.4 mM TBTA, 42.9 mM alkyne (generated in situ from R-CHO), 16 h, RT	N_3-DNA with 4 alkynes (in situ generated and commercially)		2015[67]
TBTA, $CuSO_4$, ascorbate[a]	pseudo-solid-phase: DEAE sepharose, 50 pmol of alkyne-DNA (400 nM on resin), 200 μM RN_3 (generated in situ), 10 μM TBTA, 10 μM ascorbate, 0.5 μM $CuSO_4$, O/N, 45 °C; EDTA Cu scavenger.	104 RN_3	3.4×10^4	2016[58]
$CuSO_4$, ascorbate	intramolecular reaction: 1 mM azide-/alkyne-DNA, 3.8 mM $CuSO_4$, 3.8 mM ascorbate, 30 min, 60 °C	propargylglycine with azido acetic acid linked by variable hexapeptide	2.4×10^{12}	2018[59]
TBTA, $CuSO_4$, ascorbate[a]	pseudo-solid-phase synthesis: DEAE, 0.7 nmol of N_3-macrocycle-DNA, 20 mM alkyne, 10 mM TBTA, 2.5 mM $CuSO_4$, 10 mM ascorbate, 4 h, 25 °C	136 alkynes to constant macrocyclic scaffold	3.5×10^7	2018[120]
THPTA, $CuSO_4$, ascorbate	0.13 mM propyne DNA, 12 mM N_3-natural product, 10 mM THPTA, 10 mM $CuSO_4$, 20 mM ascorbate, 1:1 DMSO:H_2O, O/N, RT	110 natural products (N_3-labeled by photoreactive diazarene linker)	110	2019[121]
TBTA, $CuSO_4$, ascorbate[a,b]	scaffold attachment: 0.21 mM alkyne DNA, 0.42 mM N_3-macrocycle precursor, 1.1 mM ascorbate, 1 mM $CuSO_4$, 1.1 mM TBTA, 425 mM TEA buffer pH 7.2, 9:1 H_2O:DMSO, 20 min, RT	21 N_3-macrocycle precursors, 663 alkyne diversifying elements	1.4×10^6	2019[47]
TBTA derivative, $CuSO_4$, ascorbate[a]	63 μM N_3-iodophenylalanine oligonucleotide, 2.5 mM alkyne, 6.3 μM $Cu(OAc)_2$, 26 μM ligand, 630 μM sodium ascorbate, 530 μM K_2CO_3, 94% aq, 6% DMSO, 3 h, 35 °C	73/116 alkynes (yield >75%)		2019[138]
THPTA, $CuSO_4$, ascorbate	1.6 μM alkyne-DNA, 400 μM RN_3 (generated in situ), 200 μM THPTA, 200 μM $CuSO_4$, 320 μM ascorbate, 96% H_2O, 4% DMF, 1 h, 40 °C	pool of 78 alkyne-DNA with 104 RN_3	8.1×10^3	2020[61]

[a]Analysis of DNA conjugates by HPLC(A260) and LCMS. [b]Quantitative analysis of DNA compatibility by qPCR.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Prior to library synthesis, building blocks were validated by coupling to a propargylamido−DNA conjugate, identifying 82/102 building blocks that coupled quantitatively. In alternative strategies, alkyne building blocks diversified azide-functionalized scaffolds. Stress et al. coupled 663 alkynes to azido homoalanine-functionalized library en route to a 1.4 × 10^6-member macrocycle library,[47] Li et al. coupled 136 alkynes to an azidolysine-functionalized scaffold during synthesis of a 3.5 × 10^7-member macrocycle library,[48] and Favalli found that 72/115 alkynes coupled to an azido iodiphenylalanine ssDNA conjugate in >75% yield.[138]

Although CuAAC has seen limited application as a diversification strategy in DEL, synthetic methodology development is providing new opportunities for exploration. Perhaps the most significant impediment has been the relative paucity of commercially available azides and alkynes compared to, for example, amines or aldehydes. Novel azides are accessible from alkyl/aryl halide starting materials,[58] and drawing on a larger building block set, several approaches generate azides from primary amines.[140−142] However, on-demand building block generation is a fairly recent innovation in the field,[55] and deployment in library synthesis is not yet published.

CuAAC is generally high yielding and prized for its orthogonality, suggesting high compatibility with DNA, but the Cu(I) catalyst introduces known hazards to DNA. For example, Cu mediates oxidative DNA damage and radical chemistry.[97,143] Measurements of DNA damage from CuAAC vary with conditions. Skopic et al. ascribed a loss of 50% of amplifable DNA post CuAAC to oxidative damage.[58] In a separate study, CuAAC preserved 74% of amplifiable dsDNA, but only 10% of amplifiable ssDNA remained.[47] Avoiding oxidative DNA damage is a multivariate problem requiring optimization of Cu, reducing agent, ligand, and even solvent. Ligand optimization, particularly as it relates to solvent choice, is the likeliest starting point for further investigation.

3.2. Prospective DEL Reaction Schemes

Emergent synthesis technologies are both rapidly expanding the chemical space that DEL can access and enabling new screening modalities that have the potential to unlock previously intractable targets. For example, on-DNA photoredox coupling is delivering novel structural diversity via radical-mediated reactivity and solid-phase synthesis-inspired approaches are circumventing the aqueous reaction constraint of on-DNA synthesis while supporting activity-based DEL screening. The framework for evaluating the utility of these new technologies remains unchanged: rigorous analytical characterization of reaction yield and DNA compatibility are determinants of effective synthesis methodology development, and building block availability continues to influence adoption prospects. Altogether, this section seeks to highlight the potential for new synthesis technologies to expand the scope of useful DEL reactions.

3.2.1. Covalent Attachment to Solid Support. Syntheses involving covalent binding of library members to solid supports have enabled both novel DEL reaction development and screening modalities. Drawing inspiration from conventional automated DNA synthesis, protected oligonucleotides attached to controlled porous glass (CPG) solid supports have been modified under reaction conditions likely to be incompatible with unprotected DNA to yield functionalized ssDNAs for initiating library synthesis. Libraries have also been prepared directly on polymeric solid supports using enzymatic encoding tag ligation reminiscent of conventional DELs but circumventing the solubility limitations of DNA during chemical synthesis steps. Solid-phase DELs also introduce the unique ability to conduct activity-based screening by exploiting the polyvalent library member display on each bead.

Recent exploration of DEL synthesis initiated on CPG has suggested that solid-phase synthesis procedures can expand the scope of DNA-compatible chemistries. Building on the earliest work in DNA-encoded combinatorial chemistry,[103] Brunschweiger disclosed methodologies for accessing modified heterocycles. Initiating synthesis on CPG-linked DNA removes the aqueous reaction constraint, and nucleobase protection increases the chemical robustness of DNA tags. Linkers composed of only pyrimidines such as hexathymidine ("hexT") or an alternating TC linker were most robust.[144,145] Exposure to 10% TFA or certain metal ions (Sc(III), Rh(II), Ru(II), Pd(0), and Pd(II)) still resulted in significant DNA damage detected by HPLC,[144,145] but on-CPG synthesis using protected oligonucleotides has nonetheless furnished previously inaccessible DNA conjugates.

On-CPG synthesis has introduced innovative approaches to library coding and access to several new chemistries on DNA. Brunschweiger's procedure begins with nucleobase-protected ssDNA bound to CPG at its 3′ terminus; the 5′ terminus displays an amine for compound synthesis. On-CPG reactions have included acid-catalyzed Pictet Spengler reaction to form β-carbolines and gold-catalyzed pyrazoline and spirocycle formation.[144,146] Several multicomponent reactions, including Zn(II)-catalyzed aza Diels−Alder, Povarov, Biginelli, Castagnoli-Cushman, 1,3-dipolar azomethine ylide−alkene cycloaddition, and isocyanide reactions, have also been developed.[61,144−148] After the initial on-CPG reaction, products are cleaved from solid support and splint-ligated to the encoding tag. These reactions showcased the ability of on-CPG synthesis to enable otherwise difficult transformations that also explore Csp^3-rich chemical space.

An alternative DNA-encoded solid-phase synthesis (DESPS) approach has provided additional unique opportunities for reaction development and library screening. Drawing inspiration from the canonical one-bead-one-compound combinatorial synthesis strategy popularized by Lam,[12,149] solid-phase DELs are synthesized and screened entirely on beads[62,104] (Figure 4). Solid-phase technology development in DEL remained largely dormant until the 2015 publication of DESPS, which married the enzymatic encoding paradigm of DEL with the water-free and automation advantages of solid-phase synthesis.[62] Solid-phase DELs are constructed on a bifunctional linker that supports compound synthesis and substoichiometric DNA-encoding tag ligation. Like traditional DEL, split-and-pool synthesis delivers exponential diversification of chemical structure, while enzymatic ligation of encoding tags records each bead's synthetic history. Unlike conventional on-DNA DEL synthesis, DNA solubility considerations are irrelevant in DESPS; thus, building block couplings are conducted in organic solvent.

High yield and DNA compatibility remain critical for successful solid-phase DEL reaction development. To evaluate chemical reactions for solid-phase DEL synthesis, our laboratory developed DNA-encoded reaction rehearsal. Solid-phase reactions of unknown yield or DNA compatibility are conducted on synthesis resin that has been mixed with magnetic sensor beads displaying a fully constructed DNA tag.[86] After the reaction, sensor beads are separated from synthesis beads.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

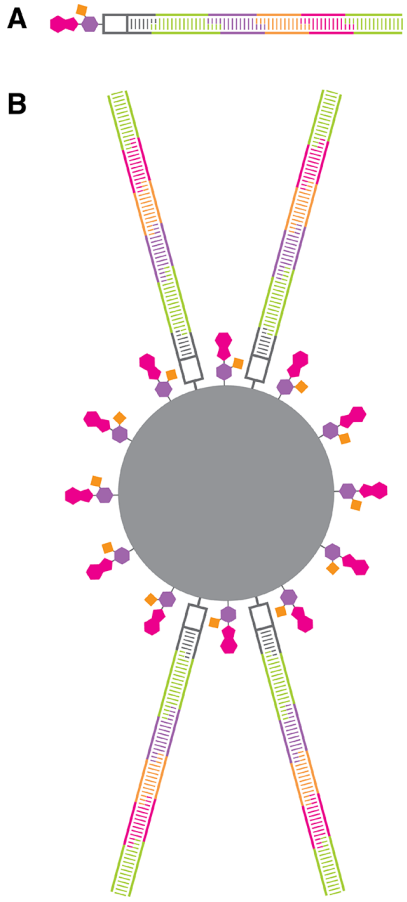


Figure 4. Solid-phase versus conventional on-DNA DELs. (A) A conventional three-cycle DEL member contains a DNA-encoding tag that is stoichiometrically conjugated to the encoded small molecule. (B) A solid-phase three-cycle DEL member comprises a bead that polyvalently displays the DNA-encoding tag and the encoded small molecule. DNA functionalization is substoichiometric to small molecule loading. Solid-phase DELs are prepared using analogous encoded split-and-pool combinatorial synthesis.

LCMS is used to determine the reaction yield from synthesis resin cleavage product, and qPCR is used to measure the amount of amplifiable DNA remaining on the sensor beads. Several common coupling reactions, such as amidation, Suzuki cross-coupling, and reductive amination, were studied.[86] Along with azide reduction, protecting group removal (Fmoc, Mtt, and Alloc) proceeded to completion with acceptable DNA damage.[86] Several more specialized reactions, such as a proline-catalyzed aldol reaction,[150] Knoevenagel condensation,[151] and Mannich reaction,[152] provide additional reactions for library generation. A variety of aldehyde reactions were also found to be both high yielding and highly DNA compatible.[153] While many DNA-compatible solid-phase reactions have been disclosed, only amidation,[109] nucleophilic displacement,[151,154] and Knoevenagel condensation[151] have been used to prepare solid-phase DELs.

The primary driver for developing solid-phase DEL technology has been achieving the ability to screen libraries in functional assays. As each DEL bead polyvalently displays one library member, it is possible to generate high local concentrations of an individual library member in the vicinity of the bead for screening. This is in stark contrast to conventional on-DNA DELs, which are inherently complex, inseparable mixtures. Synthesis of libraries on photocleavable linkers and microfluidic compartmentalization of beads allowed automated and miniaturized off-DNA screening of solid-phase DELs.[155,156] Using this technology, activity-based DEL screens identified inhibitors of the phosphodiesterase autotaxin,[109] while fluorescence polarization competition binding assays identified ligands of the receptor tyrosine kinase DDR1.[157] By separating the DEL member from the encoding tag, solid-phase DELs are poised to remove the barrier to investigating nucleic acid binding proteins[40] and enable direct interrogation of cellular signaling.

3.2.2. Solid-Phase Reversible Immobilization. Sustained interest in performing chemical modification of oligonucleotides in organic solvent has also motivated studies of solid-phase reversible immobilization to eliminate the DNA solubility constraint. These "pseudo-solid-phase" strategies leverage the polyanionic phosphate backbone of DNA to immobilize it on cationic resin for suspension in organic solvent. Conceivably, this facilitates translation of organic synthesis conditions to DNA-encoded synthesis conditions. In addition, pseudo-solid-phase chemistry offers ease of resin washing to allow rapid exchange of reagents and repeat couplings for increasing synthesis yield. Finally, these approaches expand the chemical space accessible to DEL by permitting transformations that do not tolerate water.

In 2004, Halpin and Harbury described amidation and an extensive suite of deprotection strategies for the automated chemical modification of DNA substrates immobilized on DEAE sepharose.[107,158] This approach enabled repeated cycles of synthesis and selection for two 10^6-member libraries.[106] In an investigation of several resin types, DEAE sepharose provided optimal DNA immobilization and proof-of-concept experiments demonstrated Fmoc-peptide synthesis on immobilized DNA.[107] Since then, pseudo-solid-phase reactions on DNA using DEAE sepharose have been used for amide bond formation,[48,58,159,160] peptoid synthesis,[161] reductive amination,[116] and CuAAC.[58,61,138] Despite clear utility, whether DEAE-based synthesis provides yield improvements compared to reactions in solution remains unclear.

Subsequent development of pseudo-solid-phase synthesis on DNA has focused on optimizing resin properties. The DEAE-functionalized sepharose core (Figure 5) displays numerous hydroxyl groups that are potentially cross-reactive, the resin retains sufficient water to render water-sensitive chemistries inaccessible, and the DEAE becomes deprotonated above pH ~8, abrogating DNA binding. Two parallel efforts sought to address this limitation with an essentially identical solution. Reversible adsorption to solid support (RASS)[114] and amphiphilic polymer-facilitated transformations under anhydrous conditions (APTAC)[162] both employed a resin with an organic core (PEG or PS) and quaternary amine functionality. These resins lack potentially cross-reactive nucleophiles, the organic resin core can be readily dehydrated, and quaternary amine DNA binding is pH independent.

These resin improvements offered renewed incentive to develop pseudo-solid-phase reactions that were difficult under aqueous conditions. RASS first enabled Ni-catalyzed Csp^2–Csp^3 cross-coupling of an aryl iodide-modified DNA with soluble redox-active esters (RAE), providing 82% yield in a model reaction.[114] Reactions with an additional 42 RAEs proceeded in >60% yield, while the same reaction on DEAE resin or free in solution was unproductive. A Ni-catalyzed electrochemical amination of on-DNA aryl iodide was then

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

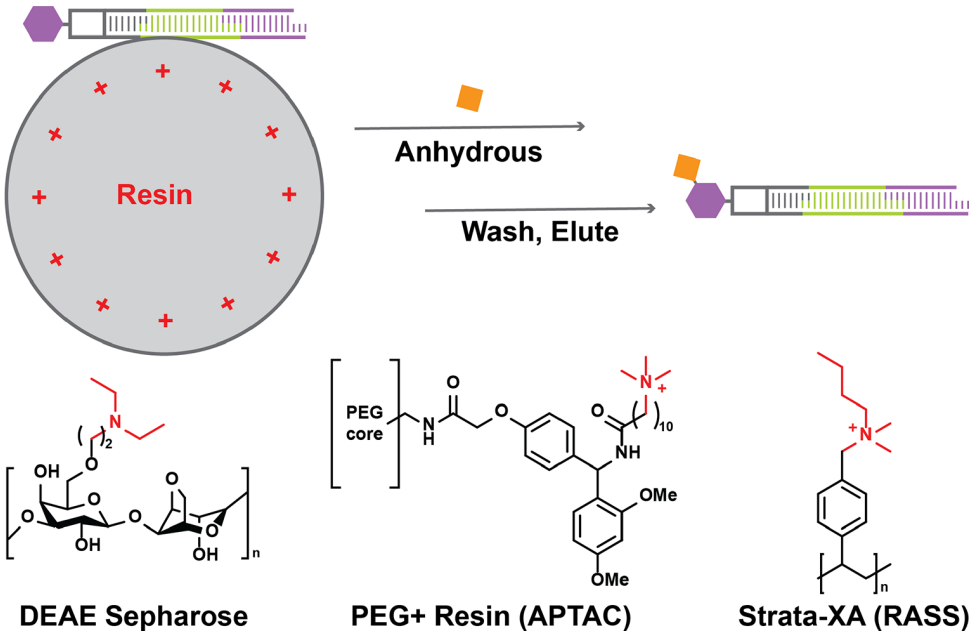
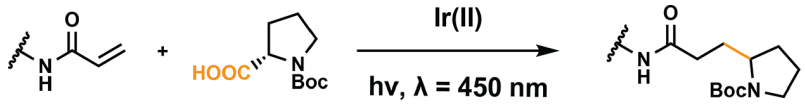


Figure 5. Solid-phase reversible immobilization-enabled DEL synthesis. The cationic bead surface immobilizes DNA conjugates through ionic interactions with the DNA phosphodiester backbone. Resin is combined with DNA in aqueous media, washed with water-miscible solvent, and then washed with dry solvent to provide near-anhydrous conditions for DNA modification. Anion exchange resins used in this process include DEAE sepharose, a modified ChemMatrix resin (PEG+), and Strata-XA.

Scheme 8. Photoredox Giese Reaction



investigated, yielding 74% conversion in a model reaction, while the corresponding reaction performed free in solution phase reaction leads to DNA loss, presumably due to electrode absorption. Yields for this reaction were roundly modest in a 21-substrate scope study (19 reactions <55%), but this was the first electrochemical modification of DNA substrates. Subsequently, RASS afforded novel C−S and S−N bonds.[163] Aryl iodide-DNA and thiols/thiophenols were coupled using a water-sensitive Ni catalyst; 29 thiols reacted with variable C−S bond forming yield (19−83%). In a second demonstration, bromohexanamide-DNA was reacted with thiols or sodium sulfinates; 12 sulfinate salts reacted in variable yields (29−83%), and 6 thiols/thiophenols all reacted in high yield (>70%).

The APTAC method has similarly expanded the scope of DEL-compatible chemistry, but fewer reactions have been disclosed.[162] While Umpolong addition and tin amine protocol (SnAP) reactions[164] were demonstrated as potential APTAC reactions, these use rare building blocks and the reaction scope is still unproven. APTAC was explored to enable photocatalytic decarboxylative cross-coupling, which uses abundant carboxylic acids. Using a custom LED-illuminated 96-well photo reactor (λ = 470 nm), APTAC permitted Ni/Ir dual catalyst cross-coupling between aryl iodide-DNA and 25 aliphatic carboxylic acids. This transformation provided high yields (>70%) for 12 carboxylic acids.

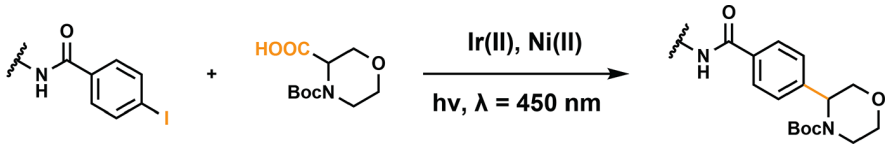
Both RASS and APTAC warrant further investigation for introducing anhydrous reactions to library synthesis. These strategies require bead handling, which increases the complexity of library synthesis but provides access to new chemical space. Reactions such as decarboxylative cross-coupling increase the Csp^3 character and introduce stereochemistry, while C−S cross-couplings form bonds that are not currently found in DEL. Abundant building blocks for these reactions increase their utility for library synthesis, but RASS and APTAC do not yet appear in published DELs. An analysis of resin binding capacity, however, supports scalability, and similar pseudo-solid-phase synthesis with DEAE resin has yielded several DELs.[48,106,160] In summary, solid-phase reversible immobilization seems generally promising, granting access to useful functionality through anhydrous conditions. Further studies are likely to confirm that these approaches exhibit the requisite high yield, broad scope, and DNA compatibility of other canonical DEL chemistry formats.

3.2.3. Photoredox Catalysis. Photoredox reactions have attracted significant attention in medicinal chemistry for their ability to generate novel C−C bonds, and they appear to be equally promising for DEL bond construction. Photoredox couplings between C-centered radicals and alkene groups (Giese coupling or defluorinative alkylation) can produce Csp^3−Csp^3 linkages. Similarly, dual catalytic photoredox cross-couplings between heteroaryl-halide-conjugated DNA and C-centered radicals can generate valuable Csp^2−Csp^3 linkages. The photoredox reactions demonstrated thus far have great appeal for DEL synthesis, as they proceed with fast kinetics despite dilute DNA concentrations, use abundantly available building block sets, and generate highly desirable functionality. Also, although radical-mediated, several studies suggest that these reactions do not significantly affect DNA recovery. s8

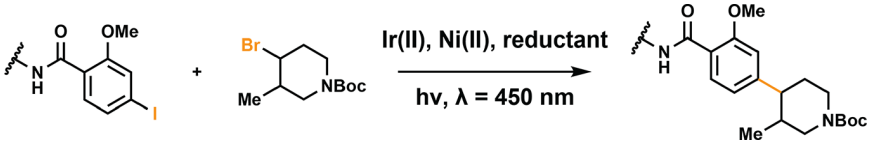
Photochemistry was one of the first reaction modes to receive attention for its inherent compatibility with DNA. Halpin and

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Scheme 9. Decarboxylative Cross-Coupling



Scheme 10. Cross Electrophile Coupling



Harbury highlighted the use of the nitroveratryloxycarbonyl (NVOC) group for mild and selective nucleophile protection in DNA-routed synthesis,[107] and both *o*-nitroveratryl and *o*-nitrobenzyl linkers have seen heavy use in oligonucleotide synthesis for their ability to mediate mild and chemoselective DNA phosphodiester cleavage. Later, Liu's powerful reaction discovery platform uncovered an on-DNA photocatalytic azide reduction,[57] and GSK used NVOC deprotection in a triazine library,[128] setting the stage for implementation of photoredox chemistry in DEL. Drawing inspiration from the corresponding off-DNA reactions,[165,166] Pfizer disclosed a photocatalytic Giese coupling[167] between acrylamide DNA and Boc-protected phenylalanine. After optimization, the desired 1,4 product was obtained in 89% yield. Boc-protected α-amino acids generally coupled in high yield (25/29 conversion >65%), while other radical precursors were only preliminarily investigated. Further, the reaction tolerated several radical acceptors; 6/9 alkene-conjugated DNA species coupled with Boc-phenylalanine in >65% conversion.

Photoredox-catalysis-based reaction development for DEL has continued to be an intense area of investigation. Novel dual catalytic reaction cycles for decarboxylative arylation were high yielding and DEL compatible.[69,162,168] In the dual catalytic approach, the photocatalyst (Ir or suitable organic dye) generates an α-carbon radical through decarboxylation, which inserts in Ni-activated aryl halides to yield the corresponding Csp^3-Csp^2 linkages and regenerate the catalysts. The reaction conditions varied widely, employing 60% aqueous,[168] 20% aqueous with $MgCl_2$ to aid DNA solubility,[69] or nearly anhydrous conditions on solid support.[162] The reaction scope was similarly variable. Pfizer's conditions were best suited for coupling Boc-protected α-amino acids as 26/29 such substrates reacted in >70% yield, while 10/21 heteroaryl halide DNA conjugates coupled to N-Boc morpholine carboxylic acids in >65% conversion.[168] Molander's conditions were simpler, as they did not require exclusion of air, but yields were generally lower. Only 4 of 15 reactions between various Boc-protected α-amino acids and aryl/pyridinyl halides surpassed 60% yield.[69] Novartis's conditions were the most intensive, requiring both solid support and water/air exclusion, but they provided access to aliphatic carboxylic acids (12/25 >70% converison with aryl iodide substrate).[162]

Adaptation of cross electrophile coupling has further expanded the scope of photoredox chemistry.[70,169] Pfizer identified a novel bis(carboxamidine) ligand that provided 92% yield in a model reaction between aryl iodide DNA and *N*-Boc-3-bromopiperidine.[169] In a scope study of 26 aryl halides reacting with a DNA-linked aryl halide, primary and secondary alkyl bromides reacted favorably for many building blocks, but tertiary or sterically hindered alkyl bromides coupled poorly. In contrast, Molander developed cross electrophile coupling conditions that proceeded under a standard atmosphere using 80% DMSO as the reaction solvent.[169] Optimal conditions included an Ir photocatalyst and a Ni catalyst, but triethylamine was used as a mild reductant. The optimized reaction coupled 15 alkyl bromides with 4-bromobenzoic acid-tagged DNA (43−84% yield). Then, 15 heteroaryl bromides/iodides were tested against two alkyl bromides (34−74% yield).

Several additional photoredox reactions have been developed, increasing both reaction and building block scope. Molander demonstrated defluorinative alkylation using a range of radical precursors, including N-Boc/Fmoc-α-amino acids, alkyl 1,4-dihydropyridines (DHP), alkyl bis(catecholato)silicates,[69] and methyl trimethylsilyl amines.[70] Similarly, methyl trimethylsilyl amines[70] and alkyl DHP radical precursors[69] were coupled to aryl halide DNA. Finally, Pfizer recently showed photoredox [2 + 2] cycloaddition as another potential reaction for implementation in DEL.[169]

Altogether, photoredox couplings appear to be one of the most promising additions to the DEL repertoire. Certain photoredox conditions, including decarboxylative cross-coupling,[162] cross electrophile coupling,[87] and [2 + 2] cycloaddition,[169] minimally impacted DNA. DNA recovery was 48, 67, and 60% post reaction, respectively. Of the available photoredox reactions, decarboxylative cross-coupling reactions are particularly attractive given that they use versatile, stable, and abundant building block sets, introduce attractive library functionalities, and complement other cross-couplings found in DEL. All photoredox couplings introduce unique challenges for optimizing catalytic cycles, designing high-throughput LEDs, and devising enclosures to exclude air. Both Pfizer and Novartis have already designed 96-well-plate illuminators for this specific purpose,[162,168] hinting that the future of photoredox coupling in DEL may be bright.

4. CONCLUSIONS

DEL synthesis is a powerful new technology for drug discovery that depends critically on reaction development. In this review, we compared click reaction constraints with desirable features of DEL reactions, providing an outline for evaluating reaction utility. This comparison emphasized both the convenience of an aqueous reaction setup and the importance of building block validation and quantitative postsynthesis DNA recovery assessment. Analysis of six of the most popular reactions in modern DEL as well as emerging synthesis technologies demonstrated the utility of the click-inspired framework.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

The success of DEL technology is ultimately measured by the quality of lead molecules that result from screening. Employing reactions that maximize library quality is critical for providing reliable data to discover molecules that can be readily optimized to druglike compounds. Additionally, exploring new bond construction is advantageous for pioneering new chemical space, particularly for the purposes of investigating increasingly difficult targets. Despite the challenges associated with designing robust reactions for DEL synthesis, several established reactions have realized dramatic performance improvements in DEL and novel reaction development efforts have greatly expanded in recent years. It will be exciting to see how future library design influences the next generation of DEL-derived clinical candidates and how these new synthesis technologies enable other efforts to circumvent the canonical limits of druggability.

AUTHOR INFORMATION

Corresponding Author

Brian M. Paegel − *Departments of Pharmaceutical Sciences, Chemistry, & Biomedical Engineering, University of California, Irvine, Irvine, California 92617, United States; Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States;* orcid.org/0000-0002-6531-6693; Email: bpaegel@uci.edu

Author

Patrick R. Fitzgerald − *Skaggs Doctoral Program in the Chemical and Biological Sciences, The Scripps Research Institute, Jupiter, Florida 33458, United States;* orcid.org/0000-0002-9324-2134

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrev.0c00789

Notes

The authors declare the following competing financial interest(s): B.M.P. declares a significant financial interest in 1859, Inc., which seeks to commercialize some technologies reviewed here.

Biographies

Patrick R. Fitzgerald earned his B.S. degree in chemistry from the University of California, Davis in 2016. After a year developing GCMS and LCMS assays at Quest Diagnostics (San Juan Capistrano, CA), he undertook doctoral studies with Paegel at Scripps Research (Jupiter, FL). In 2019, he moved with Paegel to the University of California, Irvine, where he continues his research developing advanced DNA-encoded library screening technology with an emphasis on antibacterial discovery and characterization.

Brian M. Paegel holds the rank of Professor at the University of California, Irvine, with appointments in the Departments of Pharmaceutical Sciences, Chemistry, and Biomedical Engineering. He earned a B.S. degree in chemistry from Duke University and a Ph.D. in chemistry from UC Berkeley under the direction of Richard Mathies, whose laboratory pioneered high-throughput DNA sequencing technology for the Human Genome Project. He undertook postdoctoral studies with Gerald Joyce (then at Scripps Research and now at Salk Institute, La Jolla, CA) applying laboratory automation and instrumentation engineering to problems in molecular evolution and the chemical origins of life. In 2009, he started his independent faculty career in the Department of Chemistry at Scripps Research (Jupiter, FL) where he developed miniaturized and automated technology for drug discovery, including solid-phase DNA-encoded library synthesis and microfluidic-activity-based DEL screening. He was recruited back to the UC system in 2019 to help build the new School of Pharmacy and Pharmaceutical Sciences at Irvine. His lab is developing cellular measurement strategies for DEL and using activity-based DEL to pursue classically undruggable targets. Paegel has received a Ruth L Kirschstein NRSA, a Pathway to Independence Award, the NIH Director's New Innovator Award, and the NSF CAREER Award for his work in reaction miniaturization and chemical evolution.

ACKNOWLEDGMENTS

This work was supported by the Department of Defense (W81XWH-19-1-0718) and a grant award from the National Institutes of Health (GM120491) to B.M.P.

REFERENCES

(1) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; et al. Design, Synthesis and Selection of DNA-Encoded Small-Molecule Libraries. *Nat. Chem. Biol.* **2009**, *5*, 647−654.

(2) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004−2021.

(3) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. DNA-Templated Organic Synthesis and Selection of a Library of Macrocycles. *Science* **2004**, *305*, 1601−1605.

(4) Hansen, M. H.; Blakskjaer, P.; Petersen, L. K.; Hansen, T. H.; Hoøjfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. V. A Yoctoliter-Scale DNA Reactor for Small-Molecule Evolution. *J. Am. Chem. Soc.* **2009**, *131*, 1322−1327.

(5) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. Encoded Self-Assembling Chemical Libraries. *Nat. Biotechnol.* **2004**, *22*, 568−574.

(6) Petersen, L. K.; Blakskjær, P.; Chaikuad, A.; Christensen, A. B.; Dietvorst, J.; Holmkvist, J.; Knapp, S.; Korínek, M.; Larsen, L. K.; Pedersen, A. E.; et al. Novel p38α MAP Kinase Inhibitors Identified from yoctoReactor DNA-Encoded Small Molecule Library. *MedChemComm* **2016**, *7*, 1332−1339.

(7) Zhou, Y.; Li, C.; Peng, J.; Xie, L.; Meng, L.; Li, Q.; Zhang, J.; Li, X. D.; Li, X.; Huang, X.; et al. DNA-Encoded Dynamic Chemical Library and Its Applications in Ligand Discovery. *J. Am. Chem. Soc.* **2018**, *140*, 15859−15867.

(8) Neri, D.; Lerner, R. A. DNA-Encoded Chemical Libraries: A Selection System Based on Endowing Organic Compounds with Amplifiable Information. *Annu. Rev. Biochem.* **2018**, *87*, 479−502.

(9) Pereira, D. A.; Williams, J. A. Origin and Evolution of High Throughput Screening. *Br. J. Pharmacol.* **2007**, *152*, 53−61.

(10) Volochnyuk, D. M.; Ryabukhin, S. V.; Moroz, Y. S.; Savych, O.; Chuprina, A.; Horvath, D.; Zabolotna, Y.; Varnek, A.; Judd, D. B. Evolution of Commercially Available Compounds for HTS. *Drug Discovery Today* **2019**, *24*, 390−402.

(11) Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−93.

(12) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **1991**, *354*, 82−84.

(13) Silen, J. L.; Lu, A. T.; Solas, D. W.; Gore, M. A.; Maclean, D.; Shah, N. H.; Coffin, J. M.; Bhinderwala, N. S.; Wang, Y.; Tsutsui, K. T.; et al. Screening for Novel Antimicrobials from Encoded Combinatorial Libraries by Using a Two-Dimensional Agar Format. *Antimicrob. Agents Chemother.* **1998**, *42*, 1447−1453.

(14) Kennedy, J. P.; Williams, L.; Bridges, T. M.; Daniels, R. N.; Weaver, D.; Lindsley, C. W. Application of Combinatorial Chemistry Science on Modern Drug Discovery. *J. Comb. Chem.* **2008**, *10*, 345−354.

(15) Kodadek, T. The Rise, Fall and Reinvention of Combinatorial Chemistry. *Chem. Commun.* **2011**, *47*, 9757−9763.
(16) Brenner, S.; Lerner, R. A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 5381−5383.
(17) Buller, F.; Zhang, Y.; Scheuermann, J.; Schäfer, J.; Bühlmann, P.; Neri, D. Discovery of TNF Inhibitors from a DNA-Encoded Chemical Library Based on Diels-Alder Cycloaddition. *Chem. Biol.* **2009**, *16*, 1075−1086.
(18) Mannocci, L.; Zhang, Y.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. High-Throughput Sequencing Allows the Identification of Binding Molecules Isolated from DNA-Encoded Chemical Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17670−17675.
(19) Satz, A. L. What Do You Get from DNA-Encoded Libraries? *ACS Med. Chem. Lett.* **2018**, *9*, 408−410.
(20) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596−2599.
(21) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67*, 3057−3064.
(22) Presolski, S. I.; Hong, V. P.; Finn, M. Copper-Catalyzed AzideAlkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.* **2011**, *3*, 153−162.
(23) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-based Drug Design: a Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16*, 3−50.
(24) Kolb, H. C.; Sharpless, K. B. The Growing Impact of Click Chemistry on Drug Discovery. *Drug Discovery Today* **2003**, *8*, 1128−1137.
(25) Lang, K.; Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **2014**, *9*, 16−20.
(26) Meldal, M.; Tomøe, C. W. Cu-Catalyzed Azide - Alkyne Cycloaddition. *Chem. Rev.* **2008**, *108*, 2952−3015.
(27) Tron, G. C.; Pirali, T.; Billington, R. A.; Canonico, P. L.; Sorba, G.; Genazzani, A. A. Click Chemistry Reactions in Medicinal Chemistry: Applications of the 1,3-Dipolar Cycloaddition Between Azides and Alkynes. *Med. Res. Rev.* **2008**, *28*, 278−308.
(28) Hein, C. D.; Liu, X. M.; Wang, D. Click Chemistry, a Powerful Tool for Pharmaceutical Sciences. *Pharm. Res.* **2008**, *25*, 2216−2230.
(29) Moorhouse, A. D.; Moses, J. E. Click Chemistry and Medicinal Chemistry: A case of "Cyclo- Addiction. *ChemMedChem* **2008**, *3*, 715−723.
(30) Hou, J.; Liu, X.; Shen, J.; Zhao, G.; Wang, P. G. The Impact of Click Chemistry in Medicinal Chemistry. *Expert Opin. Drug Discovery* **2012**, *7*, 489−501.
(31) Jiang, X.; Hao, X.; Jing, L.; Wu, G.; Kang, D.; Liu, X.; Zhan, P. Recent Applications of Click Chemistry in Drug Discovery. *Expert Opin. Drug Discovery* **2019**, *14*, 779−789.
(32) Thirumurugan, P.; Matosiuk, D.; Jozwiak, K. Click Chemistry for Drug Development and Diverse Chemical-Biology Applications. *Chem. Rev.* **2013**, *113*, 4905−4979.
(33) Wang, X.; Huang, B.; Liu, X.; Zhan, P. Discovery of Bioactive Molecules From CuAAC Click-Chemistry-Based Combinatorial Libraries. *Drug Discovery Today* **2016**, *21*, 118−132.
(34) Dong, J.; Krasnova, L.; Finn, M. G.; Barry Sharpless, K. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angew. Chem., Int. Ed.* **2014**, *53*, 9430−9448.
(35) Liu, Z.; Li, J.; Li, S.; Li, G.; Sharpless, K. B.; Wu, P. SuFEx Click Chemistry Enabled Late-Stage Drug Functionalization. *J. Am. Chem. Soc.* **2018**, *140*, 2919−2925.
(36) Zheng, Q.; Woehl, J. L.; Kitamura, S.; Santos-Martins, D.; Smedley, C. J.; Li, G.; Forli, S.; Moses, J. E.; Wolan, D. W.; Barry Sharpless, K. SuFEx-Enabled, Agnostic Discovery of Covalent Inhibitors of Human Neutrophil Elastase. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 18808−18814.
(37) Wan, X.; Yang, T.; Cuesta, A.; Pang, X.; Balius, T. E.; Irwin, J. J.; Shoichet, B. K.; Taunton, J. Discovery of Lysine-Targeted eIF4E Inhibitors through Covalent Docking. *J. Am. Chem. Soc.* **2020**, *142*, 4960−4964.
(38) Zhu, H.; Flanagan, M. E.; Stanton, R. V. Designing DNA Encoded Libraries of Diverse Products in a Focused Property Space. *J. Chem. Inf. Model.* **2019**, *59*, 4645−4653.
(39) Franzini, R. M.; Randolph, C. Chemical Space of DNA-Encoded Libraries: Miniperspective. *J. Med. Chem.* **2016**, *59*, 6629−6644.
(40) Goodnow, R. A.; Davie, C. P. *Annual Reports in Medicinal Chemistry*, 1st ed.; Elsevier Inc.: 2017; Vol. *50*, pp 1−15.
(41) Satz, A. L. DNA Encoded Library Selections and Insights Provided by Computational Simulations. *ACS Chem. Biol.* **2015**, *10*, 2237−2245.
(42) Satz, A. L. Simulated Screens of DNA Encoded Libraries: The Potential Influence of Chemical Synthesis Fidelity on Interpretation of Structure-Activity Relationships. *ACS Comb. Sci.* **2016**, *18*, 415−424.
(43) Eidam, O.; Satz, A. L. Analysis of the Productivity of DNA Encoded Libraries. *MedChemComm* **2016**, *7*, 1323−1331.
(44) Gerry, C. J.; Wawer, M. J.; Clemons, P. A.; Schreiber, S. L. DNA Barcoding a Complete Matrix of Stereoisomeric Small Molecules. *J. Am. Chem. Soc.* **2019**, *141*, 10225−10235.
(45) Deng, H.; Zhou, J.; Sundersingh, F. S.; Summerfield, J.; Somers, D.; Messer, J. A.; Satz, A. L.; Ancellin, N.; Arico-Muendel, C. C.; Bedard, K. L.; et al. Discovery, SAR, and X-ray Binding Mode Study of BCATm Inhibitors from a Novel DNA-Encoded Library. *ACS Med. Chem. Lett.* **2015**, *6*, 919−924.
(46) Ding, Y.; Franklin, G. J.; DeLorey, J. L.; Centrella, P. A.; Mataruse, S.; Clark, M. A.; Skinner, S. R.; Belyanskaya, S. Design and Synthesis of Biaryl DNA-Encoded Libraries. *ACS Comb. Sci.* **2016**, *18*, 625−629.
(47) Stress, C. J.; Sauter, B.; Schneider, L. A.; Sharpe, T.; Gillingham, D. A DNA-Encoded Chemical Library Incorporating Elements of Natural Macrocycles. *Angew. Chem., Int. Ed.* **2019**, *58*, 9570−9574.
(48) Li, Y.; de Luca, R.; Cazzamalli, S.; Pretto, F.; Bajic, D.; Scheuermann, J.; Neri, D. Versatile Protein Recognition by the Encoded Display of Multiple Chemical Elements on a Constant Macrocyclic Scaffold. *Nat. Chem.* **2018**, *10*, 441−448.
(49) De Pedro Beato, E.; Priego, J.; Gironda-Martínez, A.; González, F.; Benavides, J.; Blas, J.; Martín-Ortega, M. D.; Toledo, M. Á.; Ezquerra, J.; Torrado, A. Mild and Efficient Palladium-Mediated C-N Cross-Coupling Reaction between DNA-Conjugated Aryl Bromides and Aromatic Amines. *ACS Comb. Sci.* **2019**, *21*, 69−74.
(50) Lazaar, A. L.; Yang, L.; Boardley, R. L.; Goyal, N. S.; Robertson, J.; Baldwin, S. J.; Newby, D. E.; Wilkinson, I. B.; Tal-Singer, R.; Mayer, R. J.; et al. Pharmacokinetics, Pharmacodynamics and Adverse Event Profile of GSK2256294, a Novel Soluble Epoxide Hydrolase Inhibitor. *Br. J. Clin. Pharmacol.* **2016**, *81*, 971−979.
(51) Belyanskaya, S. L.; Ding, Y.; Callahan, J. F.; Lazaar, A. L.; Israel, D. I. Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *ChemBioChem* **2017**, *18*, 837−842.
(52) Harris, P. A.; King, B. W.; Bandyopadhyay, D.; Berger, S. B.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Finger, J. N.; et al. DNA-Encoded Library Screening Identifies Benzo[b][1,4]-oxazepin-4-ones as Highly Potent and Monoselective Receptor Interacting Protein 1 Kinase Inhibitors. *J. Med. Chem.* **2016**, *59*, 2163−2178.
(53) Harris, P. A.; Berger, S. B.; Jeong, J. U.; Nagilla, R.; Bandyopadhyay, D.; Campobasso, N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; et al. Discovery of a First-in-Class Receptor Interacting Protein 1 (RIP1) Kinase Specific Clinical Candidate (GSK2982772) for the Treatment of Inflammatory Diseases. *J. Med. Chem.* **2017**, *60*, 1247−1261.
(54) Ratnayake, A. S.; Flanagan, M. E.; Foley, T. L.; Smith, J. D.; Johnson, J. G.; Bellenger, J.; Montgomery, J. I.; Paegel, B. M. A Solution Phase Platform to Characterize Chemical Reaction Compatibility with DNA-Encoded Chemical Library Synthesis. *ACS Comb. Sci.* **2019**, *21*, 650−655.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

(55) Zambaldo, C.; Geigle, S. N.; Satz, A. L. High-Throughput Solid-Phase Building Block Synthesis for DNA-Encoded Libraries. *Org. Lett.* **2019**, *21*, 9353−9357.
(56) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. Inverse Electron Demand Diels-Alder Reactions in Chemical Biology. *Chem. Soc. Rev.* **2017**, *46*, 4895−4950.
(57) Chen, Y.; Kamlet, A. S.; Steinman, J. B.; Liu, D. R. A Biomolecule-Compatible Visible-Light-Induced Azide Reduction from a DNA-Encoded Reaction-Discovery System. *Nat. Chem.* **2011**, *3*, 146−153.
(58) Škopić, M. K.; Bugain, O.; Jung, K.; Onstein, S.; Brandherm, S.; Kalliokoski, T.; Brunschweiger, A. Design and Synthesis of DNA-encoded Libraries Based on a Benzodiazepine and a Pyrazolopyrimidine Scaffold. *MedChemComm* **2016**, *7*, 1957−1965.
(59) Zhu, Z.; Shaginian, A.; Grady, L. C.; O'Keeffe, T.; Shi, X. E.; Davie, C. P.; Simpson, G. L.; Messer, J. A.; Evindar, G.; Bream, R. N.; et al. Design and Application of a DNA-Encoded Macrocyclic Peptide Library. *ACS Chem. Biol.* **2018**, *13*, 53−59.
(60) Shin, M. H.; Lee, K. J.; Lim, H. S. DNA-Encoded Combinatorial Library of Macrocyclic Peptoids. *Bioconjugate Chem.* **2019**, *30*, 2931−2938.
(61) Kunig, V. B. K.; Potowski, M.; Akbarzadeh, M.; Klika Škopić, M.; Santos Smith, D.; Arendt, L.; Dormuth, I.; Adihou, H.; Andlovic, B.; Karatas, H.; et al. TEAD-YAP Interaction Inhibitors and MDM2 Binders from DNA-Encoded Indole-Focused Ugi Peptidomimetics. *Angew. Chem., Int. Ed.* **2020**, DOI: 10.1002/anie.202006280.
(62) MacConnell, A. B.; McEnaney, P. J.; Cavett, V. J.; Paegel, B. M. DNA-Encoded Solid-Phase Synthesis: Encoding Language Design and Complex Oligomer Library Synthesis. *ACS Comb. Sci.* **2015**, *17*, 518−534.
(63) Litovchick, A.; Dumelin, C. E.; Habeshian, S.; Gikunju, D.; Guié, M.-a.; Centrella, P.; Zhang, Y.; Sigel, E. A.; Cuozzo, J. W.; Keefe, A. D.; et al. Encoded Library Synthesis Using Chemical Ligation and the Discovery of sEH Inhibitors from a 334-Million Member Library. *Sci. Rep.* **2015**, *5*, 10916.
(64) Westphal, M. V.; Hudson, L.; Mason, J. W.; Pradeilles, J. A.; Zécri, F. J.; Briner, K.; Schreiber, S. L. Water-Compatible Cycloadditions of Oligonucleotide-Conjugated Strained Allenes for DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* **2020**, *142*, 7776−7782.
(65) Li, H.; Sun, Z.; Wu, W.; Wang, X.; Zhang, M.; Lu, X.; Zhong, W.; Dai, D. Inverse-Electron-Demand Diels-Alder Reactions for the Synthesis of Pyridazines on DNA. *Org. Lett.* **2018**, *20*, 7186−7191.
(66) Wang, J.; Lundberg, H.; Asai, S.; Martín-Acosta, P.; Chen, J. S.; Brown, S.; Farrell, W.; Dushin, R. G.; O'Donnell, C. J.; Ratnayake, A. S.; et al. Kinetically Guided Radical-Based Synthesis of $C(sp^3)-C(sp^3)$ linkages on DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E6404−E6410.
(67) Satz, A. L.; Cai, J.; Chen, Y.; Goodnow, R.; Gruber, F.; Kowalczyk, A.; Petersen, A.; Naderi-Oboodi, G.; Orzechowski, L.; Strebel, Q. DNA Compatible Multistep Synthesis and Applications to DNA Encoded Libraries. *Bioconjugate Chem.* **2015**, *26*, 1623−1632.
(68) Ruff, Y.; Berst, F. Efficient Copper-Catalyzed Amination of DNA-Conjugated Aryl Iodides Under Mild Aqueous Conditions. *MedChemComm* **2018**, *9*, 1188−1193.
(69) Phelan, J. P.; Lang, S. B.; Sim, J.; Berritt, S.; Peat, A. J.; Billings, K.; Fan, L.; Molander, G. A. Open-Air Alkylation Reactions in Photoredox-Catalyzed DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* **2019**, *141*, 3723−3732.
(70) Badir, S. O.; Sim, J.; Billings, K.; Csakai, A.; Zhang, X.; Dong, W.; Molander, G. A. Multifunctional Building Blocks Compatible with Photoredox-Mediated Alkylation for DNA-Encoded Library Synthesis. *Org. Lett.* **2020**, *22*, 1046−1051.
(71) Luk, K.-C.; Satz, A. L. *A Handb. DNA-Encoded Chem.*; John Wiley & Sons, Inc.: Hoboken, NJ, 2014; pp 67−98.
(72) Kung, P.-P.; Bingham, P.; Burke, B. J.; Chen, Q.; Cheng, X.; Deng, Y.-L.; Dou, D.; Feng, J.; Gallego, G. M.; Gehring, M. R.; et al. Characterization of Specific N -α-Acetyltransferase 50 (Naa50) Inhibitors Identified Using a DNA Encoded Library. *ACS Med. Chem. Lett.* **2020**, *11*, 1175−1184.
(73) Speers, A. E.; Adam, G. C.; Cravatt, B. F. Activity-Based Protein Profiling in vivo Using a Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 4686−4687.
(74) Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative Reactivity Profiling Predicts Functional Cysteines in Proteomes. *Nature* **2010**, *468*, 790−797.
(75) Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. Global Profiling of Lysine Reactivity and Ligandability in the Human Proteome. *Nat. Chem.* **2017**, *9*, 1181−1190.
(76) Oller-Salvia, B.; Kym, G.; Chin, J. W. Rapid and Efficient Generation of Stable AntibodyDrug Conjugates via an Encoded Cyclopropene and an Inverse-Electron-Demand DielsAlder Reaction. *Angew. Chem., Int. Ed.* **2018**, *57*, 2831−2834.
(77) Peplow, M. Click Chemistry Targets Antibody-drug Conjugates for the Clinic. *Nat. Biotechnol.* **2019**, *37*, 835−837.
(78) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 3192−3193.
(79) Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **2008**, *320*, 664−667.
(80) Rossin, R.; Verkerk, P. R.; Van Den Bosch, S. M.; Vulders, R. C.; Verel, I.; Lub, J.; Robillard, M. S. In vivo Chemistry for Pretargeted Tumor Imaging in Live Mice. *Angew. Chem., Int. Ed.* **2010**, *49*, 3375−3378.
(81) Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling. *Angew. Chem., Int. Ed.* **2012**, *51*, 5852−5856.
(82) Rashidian, M.; Keliher, E. J.; Dougan, M.; Juras, P. K.; Cavallari, M.; Wojtkiewicz, G. R.; Jacobsen, J. T.; Edens, J. G.; Tas, J. M.; Victora, G.; et al. Use of 18f-2-Fluorodeoxyglucose to Label Antibody Fragments for Immuno-positron Emission Tomography of Pancreatic Cancer. *ACS Cent. Sci.* **2015**, *1*, 142−147.
(83) Gates, K. S. *Reviews of Reactive Intermediate Chemistry*; John Wiley & Sons, Inc.: Hoboken, NJ, 2006; pp 333−378.
(84) Götte, K.; Chines, S.; Brunschweiger, A. Reaction Development for DNA-encoded Library Technology: From Evolution to Revolution? *Tetrahedron Lett.* **2020**, *61*, 151889.
(85) Burrows, C. J.; Muller, J. G. Oxidative Nucleobase Modifications Leading to Strand Scission. *Chem. Rev.* **1998**, *98*, 1109−1151.
(86) Malone, M. L.; Paegel, B. M. What is a "DNA-Compatible" Reaction? *ACS Comb. Sci.* **2016**, *18*, 182−187.
(87) Kölmel, D. K.; Ratnayake, A. S.; Flanagan, M. E.; Tsai, M. H.; Duan, C.; Song, C. Photocatalytic [2 + 2] Cycloaddition in DNA-Encoded Chemistry. *Org. Lett.* **2020**, *22*, 2908−2913.
(88) Sannino, A.; Gabriele, E.; Bigatti, M.; Mulatto, S.; Piazzi, J.; Scheuermann, J.; Neri, D.; Donckele, E. J.; Samain, F. Quantitative Assessment of Affinity Selection Performance by Using DNA-Encoded Chemical Libraries. *ChemBioChem* **2019**, *20*, 955−962.
(89) Satz, A. L.; Hochstrasser, R.; Petersen, A. C. Analysis of Current DNA Encoded Library Screening Data Indicates Higher False Negative Rates for Numerically Larger Libraries. *ACS Comb. Sci.* **2017**, *19*, 234−238.
(90) McCarthy, K. A.; Franklin, G. J.; Lancia, D. R.; Olbrot, M.; Pardo, E.; O'Connell, J. C.; Kollmann, C. S. The Impact of Variable Selection Coverage on Detection of Ligands from a DNA-Encoded Library Screen. *SLAS Discovery* **2020**, *25*, 515−522.
(91) Fan, L.; Davie, C. P. Zirconium(IV)-Catalyzed Ring Opening of on-DNA Epoxides in Water. *ChemBioChem* **2017**, *18*, 843−847.
(92) Thomas, B.; Lu, X.; Birmingham, W. R.; Huang, K.; Both, P.; Reyes Martinez, J. E.; Young, R. J.; Davie, C. P.; Flitsch, S. L. Application of Biocatalysis to on-DNA Carbohydrate Library Synthesis. *ChemBioChem* **2017**, *18*, 858−863.
(93) Madsen, D.; Azevedo, C.; Micco, I.; Petersen, L. K.; Hansen, N. J. V. *Progress in Medicinal Chemistry*, 1st ed.; Elsevier B.V.: 2020; Vol. 59, pp 181−249.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

(94) Kalliokoski, T. Price-Focused Analysis of Commercially Available Building Blocks for Combinatorial Library Synthesis. *ACS Comb. Sci.* **2015**, *17*, 600−607.
(95) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59*, 4443−4458.
(96) Richter, H.; Satz, A. L.; Bedoucha, M.; Buettelmann, B.; Petersen, A. C.; Harmeier, A.; Hermosilla, R.; Hochstrasser, R.; Burger, D.; Gsell, B.; et al. DNA-Encoded Library-Derived DDR1 Inhibitor Prevents Fibrosis and Renal Function Loss in a Genetic Mouse Model of Alport Syndrome. *ACS Chem. Biol.* **2019**, *14*, 37−49.
(97) Li, Y.; Gabriele, E.; Samain, F.; Favalli, N.; Sladojevich, F.; Scheuermann, J.; Neri, D. Optimized Reaction Conditions for Amide Bond Formation in DNA-Encoded Combinatorial Libraries. *ACS Comb. Sci.* **2016**, *18*, 438−443.
(98) Du, H.-C.; Simmons, N.; Faver, J. C.; Yu, Z.; Palaniappan, M.; Riehle, K.; Matzuk, M. M. A Mild, DNA-Compatible Nitro Reduction Using B 2 (OH) 4. *Org. Lett.* **2019**, *21*, 2194−2199.
(99) Li, J. Y.; Miklossy, G.; Modukuri, R. K.; Bohren, K. M.; Yu, Z.; Palaniappan, M.; Faver, J. C.; Riehle, K.; Matzuk, M. M.; Simmons, N. Palladium-Catalyzed Hydroxycarbonylation of (Hetero)aryl Halides for DNA-Encoded Chemical Library Synthesis. *Bioconjugate Chem.* **2019**, *30*, 2209−2215.
(100) Helal, C. J.; Bundesmann, M.; Hammond, S.; Holmstrom, M.; Klug-Mcleod, J.; Lefker, B. A.; McLeod, D.; Subramanyam, C.; Zakaryants, O.; Sakata, S. Quick Building Blocks (QBB): An Innovative and Efficient Business Model to Speed Medicinal Chemistry Analog Synthesis. *ACS Med. Chem. Lett.* **2019**, *10*, 1104−1109.
(101) Valeur, E.; Bradley, M. Amide Bond Formation: Beyond the Myth of Coupling Reagents. *Chem. Soc. Rev.* **2009**, *38*, 606−631.
(102) El-Faham, A.; Albericio, F. Peptide Coupling Reagents, More Than a Letter Soup. *Chem. Rev.* **2011**, *111*, 6557−6602.
(103) Nielsen, J.; Brenner, S.; Janda, K. D. Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry. *J. Am. Chem. Soc.* **1993**, *115*, 9812−9813.
(104) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. Generation and Screening of an Oligonucleotide-Encoded Synthetic Peptide Library. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 10700−10704.
(105) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. Expanding the Reaction Scope of DNA-Templated Synthesis. *Angew. Chem., Int. Ed.* **2002**, *41*, 1796−1800.
(106) Halpin, D. R.; Harbury, P. B. DNA Display II. Genetic Manipulation of Combinatorial Chemistry Libraries for Small-Molecule Evolution. *PLoS Biol.* **2004**, *2*, e174.
(107) Halpin, D. R.; Lee, J. A.; Wrenn, S. J.; Harbury, P. B. DNA Display III. Solid-phase Organic Synthesis on Unprotected Dna. *PLoS Biol.* **2004**, *2*, e175.
(108) Morgan, B.; Hale, S.; Arico-Muendel, C. C.; Clark, M.; Wagner, R.; Kavarana, M. J.; Creaser, S. P.; Franklin, G. J.; Centrella, P. A.; Israel, D. I.; et al. Methods for Synthesis of Encoded Libraries. 2005.
(109) Cochrane, W.; Malone, M. L.; Dang, V. Q.; Cavett, V. J.; Satz, A. L.; Paegel, B. M. Activity-Based DNA-Encoded Library Screening. *ACS Comb. Sci.* **2019**, *21*, 425−435.
(110) Afanasyev, O. I.; Kuchuk, E.; Usanov, D. L.; Chusov, D. Reductive Amination in the Synthesis of Pharmaceuticals. *Chem. Rev.* **2019**, *119*, 11857−11911.
(111) Podyacheva, E.; Afanasyev, O. I.; Tsygankov, A. A.; Makarova, M.; Chusov, D. Hitchhiker's Guide to Reductive Amination. *Synthesis* **2019**, *51*, 2667−2677.
(112) Cuozzo, J. W.; Centrella, P. A.; Gikunju, D.; Habeshian, S.; Hupp, C. D.; Keefe, A. D.; Sigel, E. A.; Soutter, H. H.; Thomson, H. A.; Zhang, Y.; et al. Discovery of a Potent BTK Inhibitor with a Novel Binding Mode by Using Parallel Selections with a DNA-Encoded Chemical Library. *ChemBioChem* **2017**, *18*, 864−871.
(113) Litovchick, A.; Tian, X.; Monteiro, M. I.; Kennedy, K. M.; Guié, 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*, 2026.
(114) Flood, D. T.; Asai, S.; Zhang, X.; Wang, J.; Yoon, L.; Adams, Z. C.; Dillingham, B. C.; Sanchez, B. B.; Vantourout, J. C.; Flanagan, M. E.; et al. Expanding Reactivity in DNA-Encoded Library Synthesis via Reversible Binding of DNA to an Inert Quaternary Ammonium Support. *J. Am. Chem. Soc.* **2019**, *141*, 9998−10006.
(115) An, R.; Jia, Y.; Wan, B.; Zhang, Y.; Dong, P.; Li, J.; Liang, X. Non-Enzymatic Depurination of Nucleic Acids: Factors and Mechanisms. *PLoS One* **2014**, *9*, e115950.
(116) Franzini, R. M.; Samain, F.; Abd Elrahman, M.; Mikutis, G.; Nauer, A.; Zimmermann, M.; Scheuermann, J.; Hall, J.; Neri, D. Systematic Evaluation and Optimization of Modification Reactions of Oligonucleotides With Amines and Carboxylic Acids for the Synthesis of DNA-Encoded Chemical Libraries. *Bioconjugate Chem.* **2014**, *25*, 1453−1461.
(117) Yang, H.; Medeiros, P. F.; Raha, K.; Elkins, P.; Lind, K. E.; Lehr, R.; Adams, N. D.; Burgess, J. L.; Schmidt, S. J.; Knight, S. D.; et al. Discovery of a Potent Class of PI3Kα Inhibitors with Unique Binding Mode via Encoded Library Technology (ELT). *ACS Med. Chem. Lett.* **2015**, *6*, 531−536.
(118) Wishart, D. S. DrugBank: a Comprehensive Resource for in silico Drug Discovery and Exploration. *Nucleic Acids Res.* **2006**, *34*, D668−D672.
(119) Ding, Y.; Clark, M. A. Robust Suzuki-Miyaura Cross-Coupling on DNA-Linked Substrates. *ACS Comb. Sci.* **2015**, *17*, 1−4.
(120) Li, J. Y.; Huang, H. Development of DNA-Compatible Suzuki-Miyaura Reaction in Aqueous Media. *Bioconjugate Chem.* **2018**, *29*, 3841−3846.
(121) Xu, H.; Ma, F.; Wang, N.; Hou, W.; Xiong, H.; Lu, F.; Li, J.; Wang, S.; Ma, P.; Yang, G.; et al. DNA-Encoded Libraries: Aryl Fluorosulfonates as Versatile Electrophiles Enabling Facile On-DNA Suzuki, Sonogashira, and Buchwald Reactions. *Adv. Sci.* **2019**, *6*, 1901551.
(122) Qu, Y.; Liu, S.; Wen, H.; Xu, Y.; An, Y.; Li, K.; Ni, M.; Shen, Y.; Shi, X.; Su, W.; et al. Palladium-Mediated Suzuki-Miyaura Cross-Coupling Reaction of Potassium Boc-Protected Aminomethyltrifluoroborate with DNA-Conjugated Aryl Bromides for DNA-Encoded Chemical Library Synthesis. *Biochem. Biophys. Res. Commun.* **2020**, DOI: 10.1016/j.bbrc.2020.04.027.
(123) Ding, Y.; DeLorey, J. L.; Clark, M. A. Novel Catalyst System for Suzuki-Miyaura Coupling of Challenging DNA-Linked Aryl Chlorides. *Bioconjugate Chem.* **2016**, *27*, 2597−2600.
(124) Lu, X.; Roberts, S. E.; Franklin, G. J.; Davie, C. P. On-DNA Pd and Cu Promoted C-N Cross-Coupling Reactions. *MedChemComm* **2017**, *8*, 1614−1617.
(125) Bruno, N. C.; Tudge, M. T.; Buchwald, S. L. Design and Preparation of New Palladium Precatalysts for C−C and C−N Cross-Coupling Reactions. *Chem. Sci.* **2013**, *4*, 916−920.
(126) Bruneau, A.; Roche, M.; Alami, M.; Messaoudi, S. 2-Aminobiphenyl Palladacycles: the "Most Powerful" Precatalysts in C-C and C-Heteroatom Cross-Couplings. *ACS Catal.* **2015**, *5*, 1386−1396.
(127) Chen, Y. C.; Faver, J. C.; Ku, A. F.; Miklossy, G.; Riehle, K.; Bohren, K. M.; Ucisik, M. N.; Matzuk, M. M.; Yu, Z.; Simmons, N. C-N Coupling of DNA-Conjugated (Hetero)aryl Bromides and Chlorides for DNA-Encoded Chemical Library Synthesis. *Bioconjugate Chem.* **2020**, *31*, 770−780.
(128) Deng, H.; O'Keefe, H.; Davie, C. P.; Lind, K. E.; Acharya, R. A.; Franklin, G. J.; Larkin, J.; Matico, R.; Neeb, M.; Thompson, M. M.; et al. Discovery of Highly Potent and Selective Small Molecule ADAMTS-5 Inhibitors that Inhibit Human Cartilage Degradation via Encoded Library Technology (ELT). *J. Med. Chem.* **2012**, *55*, 7061−7079.
(129) Ding, Y.; O'Keefe, H.; DeLorey, J. L.; Israel, D. I.; Messer, J. A.; Chiu, C. H.; Skinner, S. R.; Matico, R. E.; Murray-Thompson, M. F.; Li, F.; et al. Discovery of Potent and Selective Inhibitors for ADAMTS-4 through DNA-Encoded Library Technology (ELT). *ACS Med. Chem. Lett.* **2015**, *6*, 888−893.
(130) Kollmann, C. S.; Bai, X.; Tsai, C. H.; Yang, H.; Lind, K. E.; Skinner, S. R.; Zhu, Z.; Israel, D. I.; Cuozzo, J. W.; Morgan, B. A.; et al. Application of Encoded Library Technology (ELT) to a Protein-

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

Protein Interaction Target: Discovery of a Potent Class of Integrin Lymphocyte Function-Associated Antigen 1 (LFA-1) Antagonists. *Bioorg. Med. Chem.* **2014**, *22*, 2353−2365.
(131) Taylor, D. M.; Anglin, J.; Park, S.; Ucisik, M. N.; Faver, J. C.; Simmons, N.; Jin, Z.; Palaniappan, M.; Nyshadham, P.; Li, F.; et al. Identifying Oxacillinase-48 Carbapenemase Inhibitors Using DNA-Encoded Chemical Libraries. *ACS Infect. Dis.* **2020**, *6*, 1214−1227.
(132) Banerjee, R.; Brown, D. R.; Weerapana, E. Recent Developments in the Synthesis of Bioactive 2,4,6-Trisubstituted 1,3,5-Triazines. *Synlett* **2013**, *24*, 1599−1605.
(133) Saleh, M.; Abbott, S.; Perron, V.; Lauzon, C.; Penney, C.; Zacharie, B. Synthesis and Antimicrobial Activity of 2-Fluorophenyl-4,6-Disubstituted [1,3,5]Triazines. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 945−949.
(134) Rodionov, V. O.; Presolski, S. I.; Díaz, D. D.; Fokin, V. V.; Finn, M. G. Ligand-Accelerated Cu-Catalyzed Azide-Alkyne Cycloaddition: A Mechanistic Report. *J. Am. Chem. Soc.* **2007**, *129*, 12705−12712.
(135) Presolski, S. I.; Hong, V.; Cho, S. H.; Finn, M. G. Tailored Ligand Acceleration of the Cu-Catalyzed Azide-Alkyne Cycloaddition Reaction: Practical and Mechanistic Implications. *J. Am. Chem. Soc.* **2010**, *132*, 14570−14576.
(136) Besanceney-Webler, C.; Jiang, H.; Zheng, T.; Feng, L.; Soriano Del Amo, D.; Wang, W.; Klivansky, L. M.; Marlow, F. L.; Liu, Y.; Wu, P. Increasing the Efficacy of Bioorthogonal Click Reactions for Bioconjugation: A Comparative Study. *Angew. Chem., Int. Ed.* **2011**, *50*, 8051−8056.
(137) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem.* **2009**, *121*, 10063−10067.
(138) Favalli, N.; Bassi, G.; Zanetti, T.; Scheuermann, J.; Neri, D. Screening of Three Transition Metal-Mediated Reactions Compatible with DNA-Encoded Chemical Libraries. *Helv. Chim. Acta* **2019**, *102*, e1900033.
(139) Keefe, A. D.; Clark, M. A.; Hupp, C. D.; Litovchick, A.; Zhang, Y. Chemical Ligation Methods for the Tagging of Dna-encoded Chemical Libraries. *Curr. Opin. Chem. Biol.* **2015**, *26*, 80−88.
(140) Fischer, N.; Goddard-Borger, E. D.; Greiner, R.; Klapötke, T. M.; Skelton, B. W.; Stierstorfer, J. Sensitivities of Some Imidazole-1-Sulfonyl Azide Salts. *J. Org. Chem.* **2012**, *77*, 1760−1764.
(141) Gironda-Martínez, A.; Neri, D.; Samain, F.; Donckele, E. J. DNA-Compatible Diazo-Transfer Reaction in Aqueous Media Suitable for DNA-Encoded Chemical Library Synthesis. *Org. Lett.* **2019**, *21*, 9555−9558.
(142) Meng, G.; Guo, T.; Ma, T.; Zhang, J.; Shen, Y.; Sharpless, K. B.; Dong, J. Modular Click Chemistry Libraries for Functional Screens Using a Diazotizing Reagent. *Nature* **2019**, *574*, 86−89.
(143) Abel, G. R.; Calabrese, Z. A.; Ayco, J.; Hein, J. E.; Ye, T. Measuring and Suppressing the Oxidative Damage to DNA During Cu(I)-Catalyzed Azide-Alkyne Cycloaddition. *Bioconjugate Chem.* **2016**, *27*, 698−704.
(144) Škopić, M. K.; Salamon, H.; Bugain, O.; Jung, K.; Gohla, A.; Doetsch, L. J.; Santos, D. D.; Bhat, A.; Wagner, B.; Brunschweiger, A. Acid- and Au(I)-Mediated Synthesis of Hexathymidine-DNA-Heterocycle Chimeras, an Efficient Entry to DNA-Encoded Libraries Inspired by Drug Structures. *Chem. Sci.* **2017**, *8*, 3356−3361.
(145) Potowski, M.; Losch, F.; Wünnemann, E.; Dahmen, J. K.; Chines, S.; Brunschweiger, A. Screening of Metal Ions and Organocatalysts on Solid Support-Coupled DNA Oligonucleotides Guides Design of DNA-Encoded Reactions. *Chem. Sci.* **2019**, *10*, 10481−10492.
(146) Klika Škopić, M.; Willems, S.; Wagner, B.; Schieven, J.; Krause, N.; Brunschweiger, A. Exploration of a Au(I)-Mediated Three-Component Reaction for the Synthesis of DNA-Tagged Highly Substituted Spiroheterocycles. *Org. Biomol. Chem.* **2017**, *15*, 8648−8654.
(147) Potowski, M.; Kunig, V. B.; Losch, F.; Brunschweiger, A. Synthesis of DNA-coupled Isoquinolones and Pyrrolidines by Solid Phase Ytterbium- and Silver-Mediated Imine Chemistry. *MedChemComm* **2019**, *10*, 1082−1093.
(148) Kunig, V. B. K.; Ehrt, C.; Dömling, A.; Brunschweiger, A. Isocyanide Multicomponent Reactions on Solid-Phase-Coupled DNA Oligonucleotides for Encoded Library Synthesis. *Org. Lett.* **2019**, *21*, 7238−7243.
(149) Lam, K. S.; Lebl, M.; Krchnák, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, *97*, 411−448.
(150) Shu, K.; Kodadek, T. Solid-Phase Synthesis of β-Hydroxy Ketones Via DNA-Compatible Organocatalytic Aldol Reactions. *ACS Comb. Sci.* **2018**, *20*, 277−281.
(151) Pels, K.; Dickson, P.; An, H.; Kodadek, T. DNA-Compatible Solid-Phase Combinatorial Synthesis of β-Cyanoacrylamides and Related Electrophiles. *ACS Comb. Sci.* **2018**, *20*, 61−69.
(152) Tran-Hoang, N.; Kodadek, T. Solid-Phase Synthesis of β-Amino Ketones Via DNA-Compatible Organocatalytic Mannich Reactions. *ACS Comb. Sci.* **2018**, *20*, 55−60.
(153) Paciaroni, N. G.; Ndungu, J. M.; Kodadek, T. Solid-Phase Synthesis of DNA-Encoded Libraries: Via an "Aldehyde Explosion" Strategy. *Chem. Commun.* **2020**, *56*, 4656−4659.
(154) Mendes, K. R.; Malone, M. L.; Ndungu, J. M.; Suponitsky-Kroyter, I.; Cavett, V. J.; McEnaney, P. J.; MacConnell, A. B.; Doran, T. D.; Ronacher, K.; Stanley, K.; et al. High-throughput Identification of DNA-encoded IgG Ligands that Distinguish Active and Latent Mycobacterium Tuberculosis Infections. *ACS Chem. Biol.* **2017**, *12*, 234−243.
(155) Price, A. K.; MacConnell, A. B.; Paegel, B. M. HνSABR: Photochemical Dose-Response Bead Screening in Droplets. *Anal. Chem.* **2016**, *88*, 2904−2911.
(156) MacConnell, A. B.; Price, A. K.; Paegel, B. M. An Integrated Microfluidic Processor for DNA-Encoded Combinatorial Library Functional Screening. *ACS Comb. Sci.* **2017**, *19*, 181−192.
(157) Hackler, A. L.; FitzGerald, F. G.; Dang, V. Q.; Satz, A. L.; Paegel, B. M. Off-DNA DNA-Encoded Library Affinity Screening. *ACS Comb. Sci.* **2020**, *22*, 25−34.
(158) Halpin, D. R.; Harbury, P. B. DNA Display I. Sequence-Encoded Routing of DNA populations. *PLoS Biol.* **2004**, *2*, e173.
(159) Leimbacher, M.; Zhang, Y.; Mannocci, L.; Stravs, M.; Geppert, T.; Scheuermann, J.; Schneider, G.; Neri, D. Discovery of Small-Molecule Interleukin-2 Inhibitors from a DNA-Encoded Chemical Library. *Chem. - Eur. J.* **2012**, *18*, 7729−7737.
(160) Krusemark, C. J.; Tilmans, N. P.; Brown, P. O.; Harbury, P. B. Directed Chemical Evolution With an Outsized Genetic Code. *PLoS One* **2016**, *11*, e0154765.
(161) Wrenn, S. J.; Weisinger, R. M.; Halpin, D. R.; Harbury, P. B. Synthetic Ligands Discovered by in Vitro Selection. *J. Am. Chem. Soc.* **2007**, *129*, 13137−13143.
(162) Ruff, Y.; Martinez, R.; Pellé, X.; Nimsgern, P.; Fille, P.; Ratnikov, M.; Berst, F. An Amphiphilic Polymer-Supported Strategy Enables Chemical Transformations under Anhydrous Conditions for DNA-Encoded Library Synthesis. *ACS Comb. Sci.* **2020**, *22*, 120−128.
(163) Flood, D. T.; Zhang, X.; Fu, X.; Zhao, Z.; Asai, S.; Sanchez, B. B.; Sturgell, E. J.; Vantourout, J. C.; Richardson, P.; Flanagan, M. E.; et al. RASS-Enabled S/P−C and S−N Bond Formation for DEL Synthesis. *Angew. Chem., Int. Ed.* **2020**, *59*, 7377−7383.
(164) Vo, C. V. T.; Luescher, M. U.; Bode, J. W. SnAP Reagents for the One-Step Synthesis of Medium-Ring Saturated N-Heterocycles from Aldehydes. *Nat. Chem.* **2014**, *6*, 310−314.
(165) Chu, L.; Ohta, C.; Zuo, Z.; MacMillan, D. W. Carboxylic Acids as a Traceless Activation Group for Conjugate Additions: A Three-Step Synthesis of (±)-Pregabalin. *J. Am. Chem. Soc.* **2014**, *136*, 10886−10889.
(166) Bloom, S.; Liu, C.; Kölmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; Macmillan, D. W. Decarboxylative Alkylation for Site-selective Bioconjugation of Native Proteins via Oxidation Potentials. *Nat. Chem.* **2018**, *10*, 205−211.
(167) Kölmel, D. K.; Loach, R. P.; Knauber, T.; Flanagan, M. E. Employing Photoredox Catalysis for DNA-Encoded Chemistry: Decarboxylative Alkylation of α-Amino Acids. *ChemMedChem* **2018**, *13*, 2159−2165.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX

(168) Kölmel, D. K.; Meng, J.; Tsai, M. H.; Que, J.; Loach, R. P.; Knauber, T.; Wan, J.; Flanagan, M. E. On-DNA Decarboxylative Arylation: Merging Photoredox with Nickel Catalysis in Water. *ACS Comb. Sci.* **2019**, *21*, 588−597.
(169) Kölmel, D. K.; Ratnayake, A. S.; Flanagan, M. E. Photoredox Cross-Electrophile Coupling in DNA-Encoded Chemistry. *Biochem. Biophys. Res. Commun.* **2020**, DOI: 10.1016/j.bbrc.2020.04.028.

https://dx.doi.org/10.1021/acs.chemrev.0c00789
Chem. Rev. XXXX, XXX, XXX−XXX