



# A pothole-filling strategy for selective targeting of rCUG-repeats associated with myotonic dystrophy type 1

J. Dinithi R. Perera<sup>a,1</sup>, Shivaji A. Thadke<sup>a,1</sup>, Savani W. Thrikawala<sup>a,1</sup>, Isha Dharmi<sup>a</sup>, V. M. Hridya<sup>b</sup>, Arnab Mukherjee<sup>b</sup>, Ananya Paul<sup>c</sup>, W. David Wilson<sup>c</sup>, Keith W. R. Tan<sup>d</sup>, Nicholas Z. W. Chan<sup>d</sup>, Anh Tuân Phan<sup>d</sup>, and Danith H. Ly<sup>a,2</sup>

Affiliations are included on p. 9.

Edited by Gregory Petsko, Brigham and Women's Hospital, Boston, MA; received March 26, 2025; accepted December 5, 2025

We present an alternative approach to conventional small-molecule and antisense strategies for selectively targeting expanded CUG-RNA repeats associated with Myotonic Dystrophy type 1. Our alternatively designed nucleic acid ligands uniquely integrate advantageous features from both existing methods: They are compact (only three units in length), structurally resembling small molecules, yet recognize RNA targets through directional hydrogen-bonding similar to antisense oligonucleotides. Notably, these ligands exhibit greater specificity and selectivity than either approach alone. This enhanced specificity results from their bifacial recognition mechanism, wherein mismatches on one binding interface are reciprocally mirrored on the complementary face. Additionally, their short length significantly amplifies specificity, as even a single mismatch substantially reduces the overall binding free energy, effectively minimizing off-target interactions. Unlike conventional oligonucleotides, these ligands avoid binding single-stranded RNA and only recognize defined hairpin motifs via a “pothole-filling” mechanism. This method amplifies recognition specificity and selectivity, circumventing the thermodynamic penalties associated with RNA unfolding. This proof-of-concept study thus lays a foundation for developing versatile nucleic acid ligands capable of selectively targeting not only pathogenic CUG-RNA repeats in Myotonic Dystrophy type 1 but also other disease-associated triplet-repeat expansions prevalent in various neuromuscular disorders.

DM1 | Janus bases | bifacial nucleic acid recognition | cooperativity | ligands

Myotonic dystrophy type 1 (DM1) is the most common adult-onset muscular dystrophy, affecting 1 in 8,000 individuals worldwide, with no effective treatment (1). This autosomal dominant disorder is characterized by progressive muscle wasting and cognitive impairment, caused by a CTG-repeat expansion in the 3'-untranslated region (3'-UTR) of the *DMPK* gene (2, 3), where the normal range of 5 to 35 repeats expands to 80→5,000 in affected individuals. The primary driver of DM1 is RNA toxic gain-of-function, where transcription of the expanded allele produces long CUG-RNA repeats (rCUG<sup>exp</sup>) that form imperfect hairpin structures (4–6), sequestering muscleblind-like protein 1 (MBNL1) and other RNA splicing regulators (2). This entrapment disrupts RNA processing and leads to widespread splicing defects (7), which are central to disease pathology. Although reduced *DMPK* expression contributes to DM1, MBNL1 sequestration plays a greater role in symptom manifestation (3). Other factors, such as protein kinase C activation (8) and repeat-associated non-ATG (RAN) translation (9), may also contribute. Fundamentally, DM1 is a disorder of dysregulated RNA splicing, driven by the aberrant binding of expanded rCUG<sup>exp</sup> to MBNL1 and related proteins, resulting in their functional loss (10).

Several strategies have been explored to develop therapeutic interventions for DM1, including gene editing (11), transcriptional (12) and translational regulation (13), RNA degradation (14), and modulation of RNA–protein interactions (15). Notably, the latter approach has shown promise, with research groups such as Disney (16, 17), Zimmerman (18), and Berglund (19, 20) demonstrating that small-molecule ligands can be designed to bind rCUG<sup>exp</sup> and disrupt their interaction with MBNL1. Several lead compounds have successfully restored MBNL1 function and corrected misspliced RNAs in both cell cultures and animal models and are in clinical trials (21). Similarly, targeting rCUG<sup>exp</sup> with antisense oligonucleotides, specifically morpholino and gapmer variants with 2'-O-methoxyethyl modifications, Thornton (13, 22) demonstrated that DM1 phenotypes could be reversed in a mouse model. Together, these studies establish rCUG<sup>exp</sup> as a molecular driver of and a bona fide therapeutic target for DM1.

Despite recent progress (15, 23), RNA targeting via small molecules and antisense oligonucleotides continues to face significant hurdles. Small molecules, while therapeutically promising, are challenging to rationally design due to RNA's lack of defined binding

## Significance

This work presents a class of triplet nucleic acid ligands that combine the targeting precision of antisense oligonucleotides with the compactness of small molecules to selectively engage expanded CUG-RNA repeats implicated in Myotonic Dystrophy type 1. Featuring bifacial (Janus) recognition elements and a conformationally preorganized gamma peptide nucleic acid backbone, these ligands exhibit unusually high binding cooperativity, specificity, and selectivity for pathogenic over normal-length repeats. While cellular uptake remains an area for improvement, the design shows broad potential for targeting both RNA triplet-repeat expansions and other native RNA structures with essential physiological functions.

Author contributions: J.D.R.P., S.A.T., A.M., W.D.W., A.T.P., and D.H.L. designed research; J.D.R.P., S.A.T., S.W.T., I.D., V.M.H., A.P., K.W.R.T., and N.Z.W.C. performed research; J.D.R.P., S.A.T., S.W.T., I.D., V.M.H., A.M., A.P., W.D.W., K.W.R.T., A.T.P., and D.H.L. analyzed data; and J.D.R.P., S.A.T., S.W.T., A.M., W.D.W., A.T.P., and D.H.L. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2026 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

<sup>1</sup>J.D.R.P., S.A.T., and S.W.T. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: dly@andrew.cmu.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2507065123/-/DCSupplemental>.

Published January 9, 2026.

pockets, resulting in reliance on screening methods, extensive optimization, and difficulties achieving specificity and selectivity (24). Antisense oligonucleotides offer programmable target engagement but suffer from poor pharmacokinetics due to their large molecular size (25). Additionally, targeting structured RNA involves overcoming substantial thermodynamic barriers, complicating effective binding. Prioritizing affinity alone may enhance binding efficiency but often compromises specificity and selectivity, leading to off-target effects and increased cytotoxicity. These limitations highlight the need for alternative strategies that simultaneously enhance RNA-targeting precision, maintain specificity and selectivity, and provide synthetic versatility, advancing the development of effective RNA-based therapeutics.

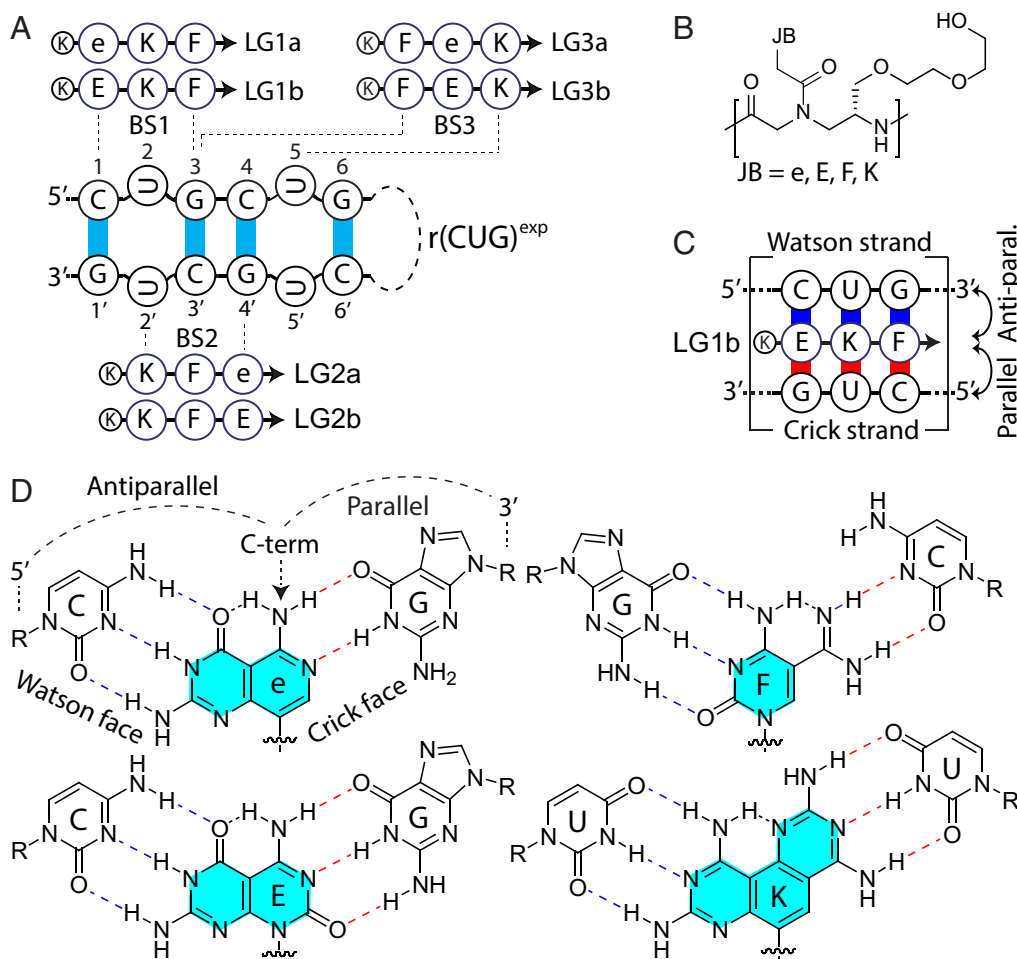
We introduce a ligand design strategy that integrates advantageous aspects of small-molecule and antisense approaches—compact size combined with high specificity and selectivity—to effectively target rCUG<sup>exp</sup>. Utilizing short, three-unit bifacial nucleic acid ligands as a proof-of-concept, these ligands structurally resemble small molecules yet achieve precise RNA recognition through directional hydrogen-bonding. Central to this design are bifacial (Janus) bases and conformationally preorganized gamma peptide nucleic acid (γPNA) backbones, which collectively enhance hydrogen-bonding, base-stacking interactions, cooperativity, and binding efficiency while minimizing entropic penalties. The bifacial architecture further improves specificity by ensuring reciprocal mirroring of mismatches, significantly reducing off-target binding, and enhancing selectivity by preferentially recognizing structured double-stranded RNA motifs over single-stranded regions. Unlike antisense oligonucleotides, these ligands employ a “pothole-filling” mechanism, recognizing RNA in its native secondary and tertiary

conformations, thus bypassing the substantial thermodynamic barriers and kinetic inhibition associated with RNA unfolding. This approach provides a powerful and versatile solution for precise RNA targeting, overcoming critical limitations inherent to traditional antisense and small-molecule strategies.

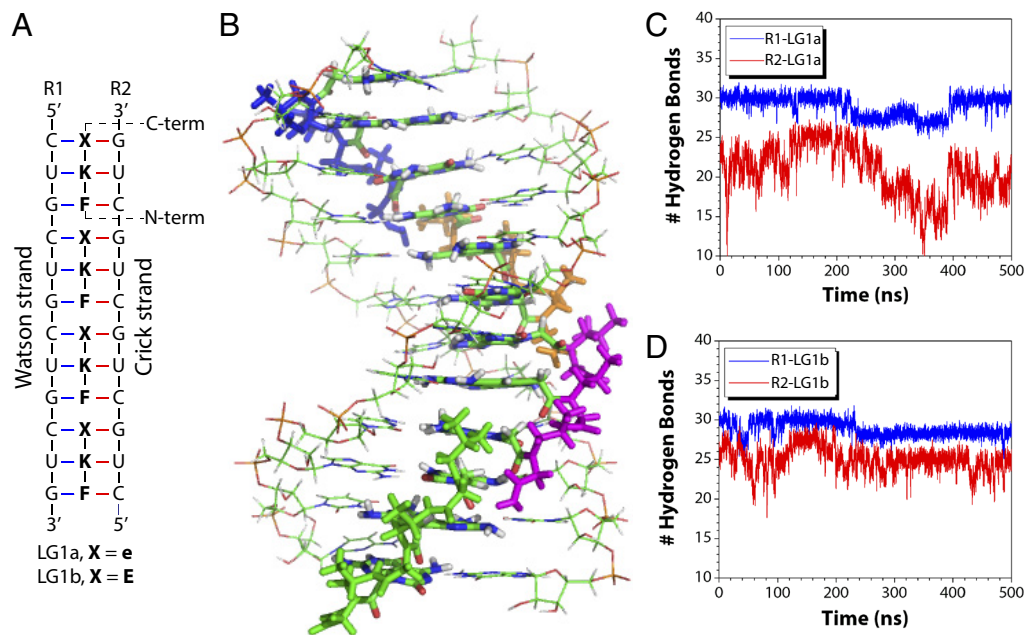
## Results and Discussion

**Rationale.** This study is part of a broader initiative aimed at developing a comprehensive set of 16 Janus bases capable of recognizing all possible RNA base-pair combinations, including canonical and noncanonical pairs. Each Janus base is strategically optimized for uniformity in size, shape, chemical functionality, and tautomeric stability, ensuring modularity and precise hydrogen-bonding with complementary RNA bases. rCUG-repeats were selected as a model system due to their well-characterized hairpin structure formed by repeating (rCUG/rGUC)-units. These structures exhibit reduced strand-invasion barriers facilitated by internal U<>U mismatches, making them particularly suited to demonstrate the effectiveness of this bifacial recognition strategy.

This base-paired triplet-repeat RNA target offers an ideal opportunity to evaluate short, three-unit nucleic acid ligands, leveraging enhanced hydrogen-bonding, improved intra- and intermolecular base-stacking, and an entropically favorable, conformationally preorganized γPNA backbone (26, 27). As illustrated in Fig. 1A, these triplet ligands are strategically designed to target rCUG-repeats by shifting recognition sites in increments of one repeat unit. The first-generation Janus base (e) in “Series a” forms five hydrogen bonds with RNA C–G pairs (28, 29), while the second-generation



**Fig. 1.** Ligand design. (A) Illustration of the rCUG-repeat hairpin structure, highlighting three potential ligand binding sites (BS1, BS2, BS3) and their corresponding ligands. Ligand series a and b feature three JB recognition elements (e or E, F, and K). e (first-generation) forms five hydrogen bonds with the C–G base-pair, while E (second-generation) forms six. The arrow indicates the ligand’s N terminus, and K represents *L*-lysine. (B) Chemical structure of the conformationally preorganized, right-handed, MiniPEG-containing gamma peptide nucleic acid (RH-MPyPNA), with JB’s covalently attached. (C) Representative binding mode of LG1b with the cognate RNA hairpin. The Watson face (blue) is antiparallel, with the RNA 5′-end facing the ligand’s C-terminus, while the Crick face (red) is parallel, with the RNA 3′-end facing the ligand’s C-terminus. (D) Depiction of hydrogen-bonding interactions between JB’s and designated RNA base-pairs.



**Fig. 2.** Molecular dynamics simulations. (A) Illustration of ligand–RNA complexes, showing four consecutively bound LG1a (X = e) and LG1b (X = E) ligands. For clarity, the C-terminal lysine residue is omitted. (B) Optimized structure of the LG1b–RNA complex, with four consecutively bound ligands represented in different colors. (C and D) Comparison of hydrogen-bonds formed between LG1a and LG1b with the Watson strand (R1, blue) and Crick strand (R2, red) of the RNA duplex over 500 ns simulations.

base (E) in “Series b” establishes six hydrogen bonds, further increasing binding affinity and specificity. These Janus bases (e, E, F, and K) are incorporated into the  $\gamma$ PNA backbone (Fig. 1B), and their binding orientation, binding mode, and hydrogen-bonding patterns are detailed in Fig. 1C and D. Altogether, this study establishes a rational design framework for precise RNA targeting, significantly advancing fundamental understanding and offering potential insights for future applications.

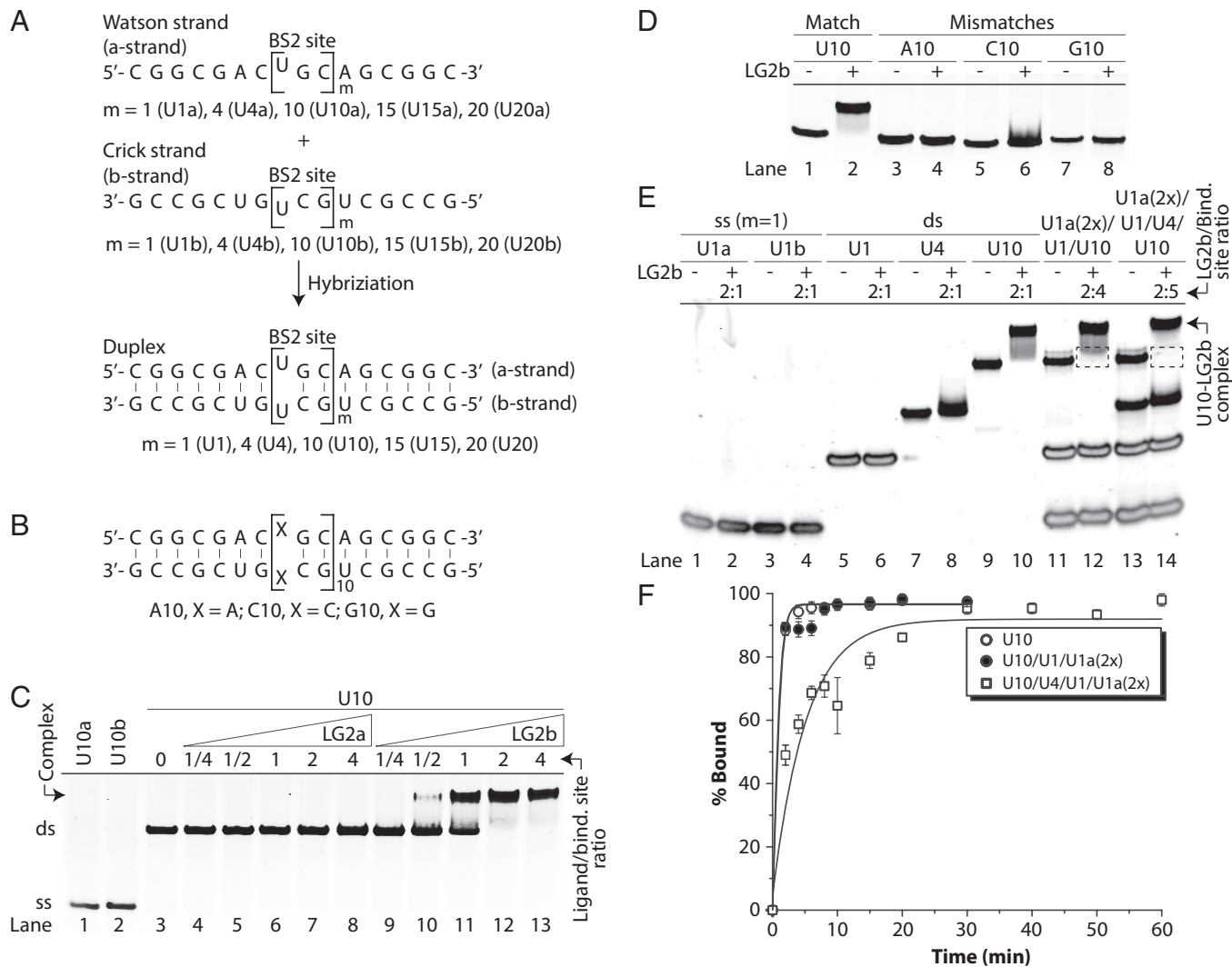
**Molecular Modeling.** We performed molecular dynamics simulations to evaluate ligand design and hydrogen-bonding interactions between Janus bases and RNA base-pairs, using ligands LG1a and LG1b bound to an RNA duplex containing four consecutive binding sites (Fig. 2A). Systems were constructed, solvated, energy-minimized, and simulated for 500 ns in four sets to assess complexes with single, double, triple, and quadruple ligand occupancy. The results revealed progressively increased complex stability as additional ligands bound consecutively, with the quadruple-ligand complex exhibiting the highest stability, demonstrating enhanced binding cooperativity (Fig. 2B).

Through molecular dynamics simulations, the LG1b–RNA complex maintained structural stability, whereas LG1a exhibited distortions, reflecting the superior hydrogen-bonding capability of base E in LG1b compared to base e in LG1a. Ligand binding was more favorable with the Watson strand (R1) than the Crick strand (R2) (Fig. 2C and D), consistent with the established antiparallel binding preference of  $\gamma$ PNA to DNA or RNA (30). Distance analyses confirmed stable hydrogen-bonding interactions ( $\sim 2$  Å) between bases E, F, and K, and their corresponding RNA base-pairs (SI Appendix, Fig. S1). These results demonstrate that triplet nucleic acid ligands could be rationally designed to selectively recognize and bind pathogenic rCUG<sup>exp</sup>, potentially distinguishing expanded repeats from wild-type sequences.

**Chemical Synthesis.** The synthesis of carboxymethyl nucleobases (e, E, F, and K) and their corresponding  $\gamma$ PNA building blocks has been previously reported (SI Appendix, Fig. S2) (31). Ligands were assembled via solid-phase synthesis, then cleaved, deprotected,

purified by HPLC, and characterized by MALDI-TOF (SI Appendix, Figs. S3 and S4 and Table S1). For this study, ligands LG2a and LG2b (Fig. 1A) were selected due to prior evidence showing that ligands targeting terminal mismatched RNA base-pairs exhibit superior binding compared to fully complementary ligands (LG1a and LG1b) (28), likely driven by kinetic factors. Additionally, we synthesized LG2c by adding (*L*-Arg)<sub>6</sub> to a lysine sidechain containing a disulfide bond to enhance cellular uptake and intracellular (cytoplasmic and nuclear) cargo release, addressing the limited permeability of LG2b (SI Appendix, Figs. S5 and S6). UV-vis spectroscopy of the deprotected building blocks (SI Appendix, Fig. S7), individually measured in water and a 1:1 mixture of H<sub>2</sub>O/MeOH at ambient temperature, and collectively within ligands in sodium phosphate buffer at 95 °C (SI Appendix, Fig. S8), showed consistent agreement. The extinction coefficients measured at 260 nm were K = 20,370 ± 1,475; F = 5,787 ± 391; e = 4,383 ± 200; E = 14,683 ± 895 M<sup>-1</sup> cm<sup>-1</sup>.

**Assessment of Ligand Binding.** Electrophoretic mobility shift assays (EMSAs) were performed to evaluate ligand–RNA binding. RNA substrate U10 (Fig. 3A), containing 10 perfectly matched ligand-binding sites, was incubated with varying concentrations of ligands LG2a or LG2b under physiologically relevant conditions (10 mM NaPi, 137 mM NaCl, 150 mM KCl, 2 mM MgCl<sub>2</sub>, pH 7.4, 37 °C) (32), unless stated otherwise. Samples were analyzed using nondenaturing PAGE and visualized with SYBR-Gold staining. As shown in Fig. 3C, LG2a displayed negligible binding even at high ligand-to-target ratios (lanes 4 to 8; up to 4:1 ratio, lane 8). In contrast, LG2b exhibited  $\sim 60\%$  binding at a 1:1 ratio (lane 11 vs. lane 6), achieving complete binding at a 2:1 ratio, as indicated by the clear formation of shifted complexes and reciprocal disappearance of free RNA bands. The sharply defined, stable shifted bands persisted throughout the 2-h electrophoresis at room temperature, suggesting minimal complex dissociation. These results confirm that LG2b (bearing base E instead of e) binds RNA targets significantly more effectively than LG2a, in agreement with molecular dynamics simulations. Enhanced binding by LG2b



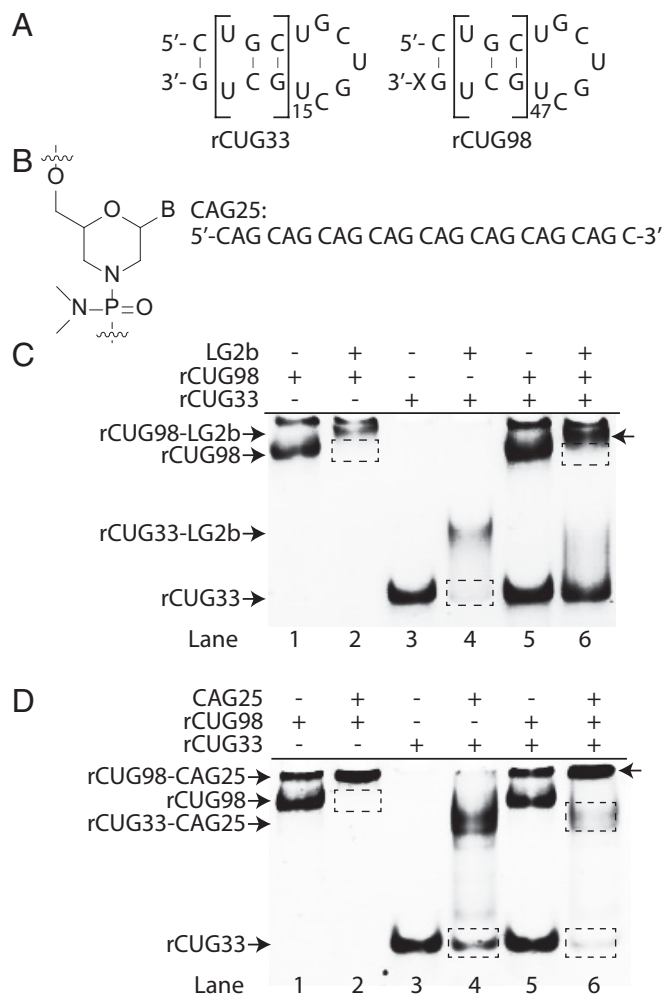
**Fig. 3.** Ligand binding assessment with model RNAs. (A) Sequences of model RNA duplexes containing varying numbers of LG2a/b binding sites, formed by hybridizing equimolar Watson and Crick strands. (B) Sequences of mismatched RNA duplexes. (C) EMSA results comparing LG2a (lanes 4 to 8) and LG2b (lanes 9 to 13) binding to U10. U10 concentration: 500 nM (5  $\mu$ M ligand-binding sites); ligand-to-binding site ratios indicated. (D) LG2b binding comparison between perfectly matched U10 (U $\leftrightarrow$ U, lanes 1 & 2) and mismatched A10 (A $\leftrightarrow$ A, lanes 3 & 4), C10 (C $\leftrightarrow$ C, lanes 5 & 6), and G10 (G $\leftrightarrow$ G, lanes 7 & 8) RNAs. RNA concentration: 500 nM (5  $\mu$ M ligand-binding sites); LG2b: 10  $\mu$ M (2:1 ligand-to-binding site ratio). (E) LG2b binding analysis with single-stranded (U1a, U1b) and double-stranded (U1, U4, U10) RNAs, individually (lanes 1 to 10) and in combinations (lanes 11 to 14) to assess selectivity. LG2b: 10  $\mu$ M; RNA concentrations adjusted to maintain a 5  $\mu$ M binding site concentration (2:1 ligand-to-binding site ratio in lanes 1 to 10). In lanes 11 to 14, U1a and U1b concentrations were doubled to compensate for the absence of a complementary strand. Lanes 11 & 12: U1a (10  $\mu$ M), U1 (5  $\mu$ M), U10 (0.5  $\mu$ M)—20  $\mu$ M ligand-binding sites. Lanes 13 & 14: Same as lanes 11 & 12, with U4 (1.25  $\mu$ M) added—25  $\mu$ M ligand-binding sites. (F) Competitive binding over time under the same conditions as lanes 10, 12, and 14 in (E). Samples were prepared in physiological ionic strength buffer (10 mM sodium phosphate, 137 mM NaCl, 150 mM KCl, 2 mM MgCl<sub>2</sub>, pH 7.2), incubated at 37  $^{\circ}$ C for 30 min, separated by 15% nondenaturing PAGE, and stained with SYBR Gold.

likely arises from stronger hydrogen-bonding and more favorable base-stacking interactions, resulting in improved target affinity.

Although the shifted bands indicated ligand–RNA binding, whether this interaction was sequence-specific or driven by nonspecific electrostatic or hydrophobic interactions remained unclear. To address this question, LG2b was incubated with single-base mismatched RNA targets: A10 (A–A), C10 (C–C), and G10 (G–G) (Fig. 3B), using a 2:1 ligand-to-binding site ratio sufficient for complete binding to the matched U10 target. No shifted bands were observed with any mismatched sequences (Fig. 3D, lanes 3 to 8), confirming the sequence-specific binding of LG2b to U10. This high specificity likely arises from the bifacial nature of the ligand interactions, enabling precise recognition and discrimination of RNA sequences.

**Evaluation of LG2b Binding Selectivity.** A significant challenge in targeting rCUG-repeats is distinguishing pathogenic length from normal-length sequences. To assess the selectivity of LG2b, we

performed competitive binding assays using single-stranded RNAs (U1a and U1b) and their corresponding duplexes containing one (U1), four (U4), or ten (U10) binding sites (Fig. 3A). This approach allowed us to determine whether LG2b recognizes structured RNA duplexes exclusively or also binds single-stranded RNA. LG2b showed no detectable binding to single-stranded RNA (U1a or U1b; Fig. 3E, lanes 1 to 4) or shorter duplexes (U1 and U4; lanes 5 to 8), whereas robust binding was observed with the longer duplex U10 (lanes 9 to 10). Importantly, excess single-stranded U1a and duplexes U1 or U4 did not disrupt LG2b binding to U10 (lanes 11 to 14), further confirming LG2b's selectivity for longer rCUG-repeat duplexes. Kinetic analysis demonstrated rapid and complete binding to U10 within 5 min (Fig. 3F). However, in the presence of excess U4, binding exhibited a lag phase, reaching full occupancy in about 20 min. This indicates weaker and less stable interactions with shorter repeats, allowing rapid dissociation and preferential reassociation with longer, more stable duplexes



**Fig. 4.** Assessment of ligand binding selectivity. (A) RNA targets with normal (rCUG33) and pathogenic (rCUG98) repeat-lengths. (B) Sequence of the morpholino oligonucleotide CAG25. (C) Evaluation of LG2b's binding selectivity for rCUG33 and rCUG98, using LG2b (5  $\mu$ M), rCUG33 (0.33  $\mu$ M,  $\sim$ 5  $\mu$ M ligand-binding sites), and rCUG98 (0.10  $\mu$ M,  $\sim$ 5  $\mu$ M ligand-binding sites). (D) Comparison of LG2b and CAG25 binding selectivity for normal and pathogenic rCUG-repeat lengths. The LG2b, rCUG33, and rCUG98 concentrations matched those in (C), while CAG25 (1.2  $\mu$ M) was used at a stoichiometric ratio with its binding sites in rCUG33 (0.33  $\mu$ M  $\sim$  1.2  $\mu$ M) and rCUG98 (0.1  $\mu$ M  $\sim$  1.2  $\mu$ M). All samples were prepared under physiologically relevant ionic strength and temperature. The top bands in (C and D) likely represent large rCUG98 complexes formed through intermolecular hybridization, which could not be resolved by nondenaturing PAGE under the given experimental conditions.

such as U10. Collectively, these findings highlight LG2b's strong selectivity for longer rCUG-repeat sequences, underscoring its downstream potential for targeting triplet-repeat expansions.

Although LG2b showed promising selectivity, the U10 duplex (10 binding sites) does not represent pathogenic repeat lengths seen in DM1. To determine if LG2b distinguishes pathogenic from normal-length rCUG-repeats, we compared its binding to pathogenic-length rCUG98 (98 repeats, 47 ligand-binding sites) and normal-length rCUG33 (33 repeats, 15 binding sites) duplexes in their native hairpin motif (Fig. 4A). rCUG98 was chosen due to its established pathogenic role in DM1 (22). For reference, we included the antisense morpholino oligonucleotide CAG25 (Fig. 4B), containing eight CAG-repeats, previously demonstrated to effectively disrupt rCUG<sup>exp</sup>-MBNL1 complexes and correct misspliced mRNAs in DM1 cells and animal models (22). LG2b and CAG25 were incubated separately with equimolar binding sites of rCUG98 and rCUG33 duplexes at a 1:2

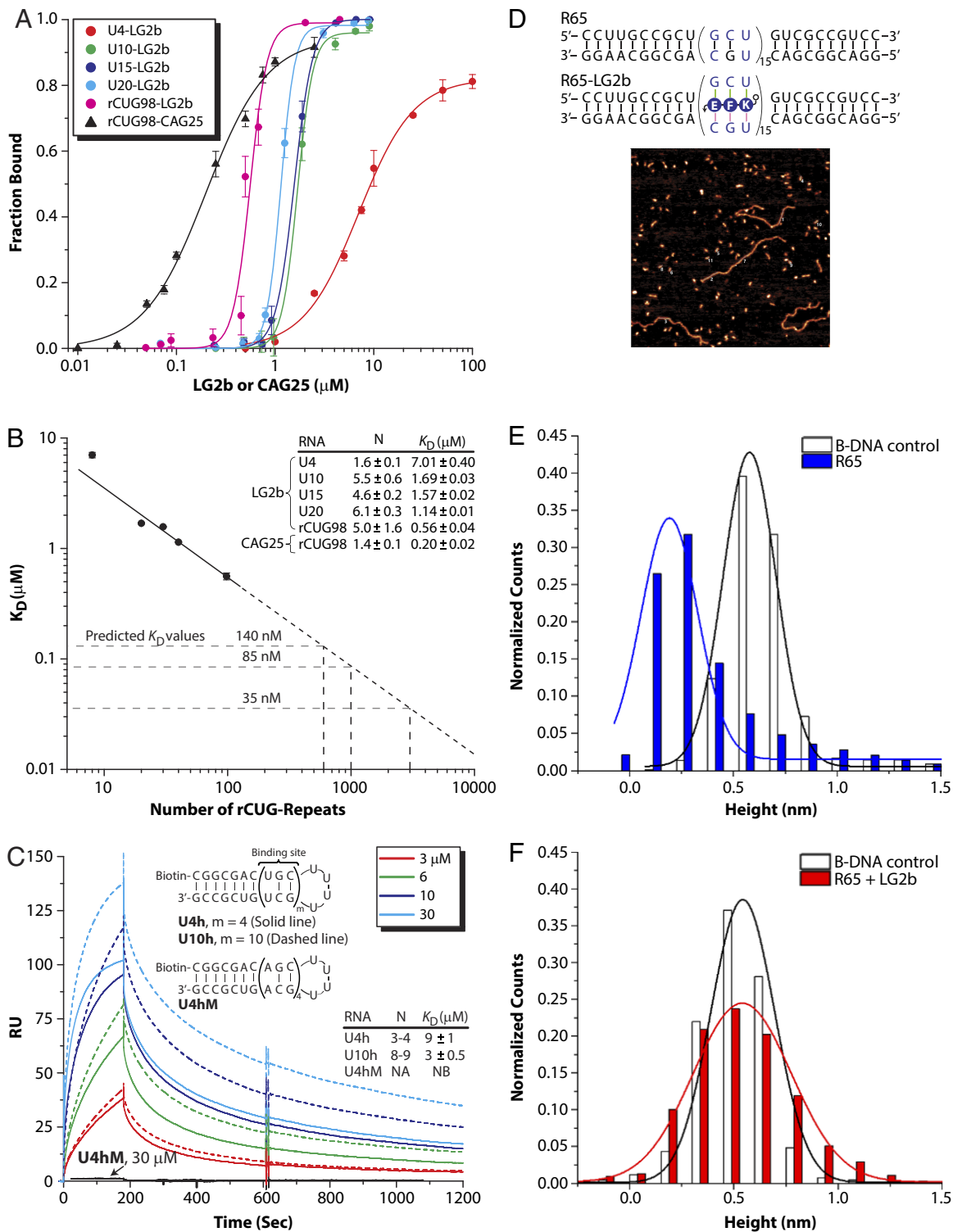
ligand-to-binding site ratio. Consistent with earlier observations, LG2b preferentially bound pathogenic-length rCUG98 with no detectable interaction with rCUG33 (Fig. 4C, lanes 5 & 6). In contrast, CAG25 showed nearly equal binding to both rCUG98 and rCUG33 under the same conditions (Fig. 4D, lanes 5 & 6). The observed preferential binding of LG2b to longer repeats highlights strong cooperativity, likely driven by intermolecular  $\pi$ - $\pi$  stacking between adjacent ligands, enabled by their extended aromatic surfaces.

**Quantification and Validation of Ligand Binding.** We determined the dissociation constants ( $K_d$ s) of LG2b for rCUG-repeated RNAs of varying lengths using EMSAs. Initial experiments provided approximate  $K_d$  values, which were subsequently refined through precise measurements across a 100-fold ligand concentration range centered around these initial estimates (SI Appendix, Fig. S9). As a positive control, we included the CAG25 morpholino oligonucleotide, whose affinity ( $K_d$ ) for rCUG98 is well established (22). All EMSAs were performed in triplicate using 15% PAGE, and RNA-ligand (or RNA-CAG25) complexes were visualized by SYBR-Gold staining. The fraction of RNA bound at each LG2b concentration was quantified using ImageJ, and resulting data were plotted and fitted to a Hill-Langmuir binding model using Origin software to determine  $K_d$  values.

As expected, the  $K_d$  values for LG2b decreased progressively with increasing rCUG-repeat length, reflecting enhanced binding affinity and greater cooperativity (Fig. 5A). This trend was further supported by an increase in the Hill coefficient ( $N$ ), rising from  $\sim$ 1.6 for shorter repeats (U4) to  $\sim$ 5 for pathogenic-length rCUG98. Specifically, LG2b exhibited  $K_d$  values of  $7.01 \pm 0.4$ ,  $1.69 \pm 0.03$ ,  $1.57 \pm 0.02$ ,  $1.14 \pm 0.01$ , and  $0.56 \pm 0.04$   $\mu$ M for U4, U10, U15, U20, and rCUG98, respectively. In comparison, CAG25 bound rCUG98 with a  $K_d$  of  $0.20 \pm 0.02$   $\mu$ M, closely matching its previously reported affinity ( $232 \pm 47$  nM) (22). Although CAG25 displayed slightly higher affinity ( $\sim$ twofold lower  $K_d$ ) for rCUG98, its binding cooperativity was significantly lower ( $N = 1.4$ ) compared to LG2b ( $N \sim 5$ ) (Fig. 5B). This difference is evident from the broader binding profile of CAG25 compared to the sharper and more cooperative binding curves observed for LG2b (Fig. 5A).

In neuromuscular disorders involving triplet-repeat expansions, selectively targeting pathogenic RNA transcripts without affecting normal lengths is crucial, as many implicated genes play essential roles in normal development (33). Nonspecific targeting risks disrupting their normal functions, causing harmful side effects. DM1 is unique among such disorders due to the extreme expansion of rCUG-repeats, often exceeding 1,000 repeats, compared to typical expansions of 50 to 200 repeats in other diseases. LG2b's bifacial design specifically addresses this challenge by preferentially binding longer pathogenic repeats with enhanced selectivity. Additionally, the observed linear correlation between LG2b's  $K_d$  values and rCUG-repeat length predicts a  $K_d$  of approximately 35 nM for transcripts with 5,000 repeats (Fig. 5B)—an affinity comparable to FDA-approved small molecules targeting structured protein sites (34). Overall, these results underscore the potential of this ligand class as molecular tools for studying RNA-repeat expansion pathology and as candidates for therapeutic development.

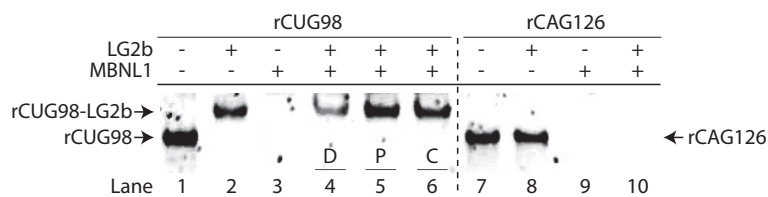
To validate EMSA results, we employed surface plasmon resonance (SPR) to determine  $K_d$  values for representative RNA targets. We selected three RNA hairpins—U4h (four binding sites), U10h (ten binding sites), and a single-base mismatched control (U4hM)—designed to facilitate purification and efficient surface immobilization. Each RNA target was 5'-biotinylated and individually immobilized on streptavidin-coated flow cells. SPR



**Fig. 5.** Quantification of LG2b binding. (A) Hill plots showing LG2b binding to U4, U10, U15, U20, and rCUG98, along with the binding of CAG25 to rCUG98 for comparison, as determined by EMSA. (B) Correlation between  $K_D$  values and rCUG-repeat lengths. Filled circles indicate experimental data, the solid line represents the linear regression fit, and the dashed line shows extrapolation. (C) SPR profiles of LG2b binding to U4h (4 sites, line), U10h (10 sites, dashes), and mismatched U4hM. (D) Sequence of R65 without and with bound LG2b, alongside an AFM image of R65-LG2b complexes (oval shapes) and reference linear 1,434-bp B-DNA (string-like feature). (E) Histogram comparing the height distribution of R65 and reference DNA. (F) Histogram comparing the height distribution of R65-LG2b complexes and reference DNA. Gaussian fits indicate heights of R65 (0.191  $\pm$  0.011 nm), R65 + LG2b (0.540  $\pm$  0.007 nm), and B-DNA (0.560  $\pm$  0.003 nm), with a 0.348 nm increase upon complex formation.

experiments were performed at 25 °C in physiologically relevant buffer, with varying concentrations of LG2b injected over immobilized RNAs at a constant flow rate of 100  $\mu\text{L}/\text{min}$ , followed by a washing step to observe ligand dissociation. Reference signals from a blank flow cell were subtracted to yield response units (RU) directly proportional to the bound ligand.

SPR sensograms for LG2b binding to U4h and U10h are shown in Fig. 5C. Although both targets exhibited similar association rates (on-rates), the dissociation rate (off-rate) was approximately three times slower for U10h compared to U4h, resulting in  $K_d$  values of 3  $\mu\text{M}$  for U10h and 9  $\mu\text{M}$  for U4h, closely matching EMSA results. Under identical conditions, LG2b showed no



**Fig. 6.** Evaluation of LG2b's ability to disrupt the CUG98–MBNL1 complex. Displacement, prevention, and competitive binding assays assessing LG2b's effect on the rCUG98–MBNL1 interaction under different conditions (D: displacement, P: prevention, C: competitive). rCUG98, MBNL1, and LG2b concentrations were 0.01  $\mu$ M (0.5  $\mu$ M ligand binding sites), 0.50  $\mu$ M, and 2.50  $\mu$ M, respectively.

detectable binding to the mismatched control (U4hM), confirming the specificity observed by EMSA. Notably, SPR-derived Hill coefficients (N) were approximately twice as high as those obtained by EMSA, reflecting differences in experimental conditions between the two methods. Nonetheless, both techniques consistently demonstrated increased binding cooperativity with repeat length, reinforcing the robustness of these observed trends.

#### Evaluation of Ligand Binding by Atomic Force Microscopy (AFM).

High-resolution AFM imaging has previously been used to reveal structural details of nucleic acid complexes, including DNA G-quadruplexes (35). To independently verify LG2b binding, we analyzed the structural topology of a model RNA target (R65) using AFM. R65 consists of 15 (rGCU/rUGC)-repeats—near the upper limit of chemical synthesis—and, consistent with prior designs, includes sufficiently long CG-rich flanking regions to ensure duplex stability (Fig. 5D). Binding experiments were conducted under the same conditions as EMSA (SI Appendix, Fig. S10), and RNA–ligand complexes were immobilized on a mica surface for imaging. Contour maps generated by AFM allowed visualization of ligand-induced structural changes, using a double-stranded B-DNA of known length as an internal control. The LG2b-bound R65 complex was expected to show increased contour height compared to unbound R65 due to ligand insertion expanding the RNA helix. Since RNA molecules align along their helical axis on mica, changes in contour height directly reflect alterations in helix width, providing structural confirmation of LG2b binding.

Measured contour heights were  $0.191 \pm 0.011$  nm for unbound R65,  $0.540 \pm 0.007$  nm for R65–LG2b complexes, and  $0.560 \pm 0.003$  nm for the control B-DNA (Fig. 5E and F). The consistent contour height of B-DNA, regardless of LG2b presence, confirmed the ligand's lack of interaction with double-stranded DNA. The significant increase in contour height (0.348 nm) observed exclusively for R65 upon LG2b binding strongly suggests that ligand interaction occurred via a pothole-filling mechanism. Unlike typical groove-binding interactions with small molecules or proteins (36)—which generally produce minimal height changes—this substantial height differential provides robust structural evidence of specific ligand–RNA binding. These AFM findings support the sequence-specificity and selective targeting of rCUG-repeated RNAs by LG2b.

#### Displacement and Prevention of MBNL1 Binding by LG2b.

A primary driver of DM1 pathology is gain-of-function RNA toxicity, wherein expanded CUG-repeat transcripts form hairpin structures that sequester RNA splicing regulators, particularly MBNL1. This sequestration leads to nuclear retention of the MBNL1–rCUG<sup>exp</sup> complexes, reduced DMPK protein production, and disrupted MBNL1 activity, resulting in widespread RNA-splicing defects characteristic of DM1. To determine whether LG2b can compete effectively with MBNL1 for binding to rCUG98, we conducted competitive binding assays in three distinct formats: 1) *displacement*—LG2b added to preformed rCUG98–MBNL1 complexes; 2) *prevention*—LG2b preincubated with rCUG98 before adding MBNL1; and 3) *competitive*—all components

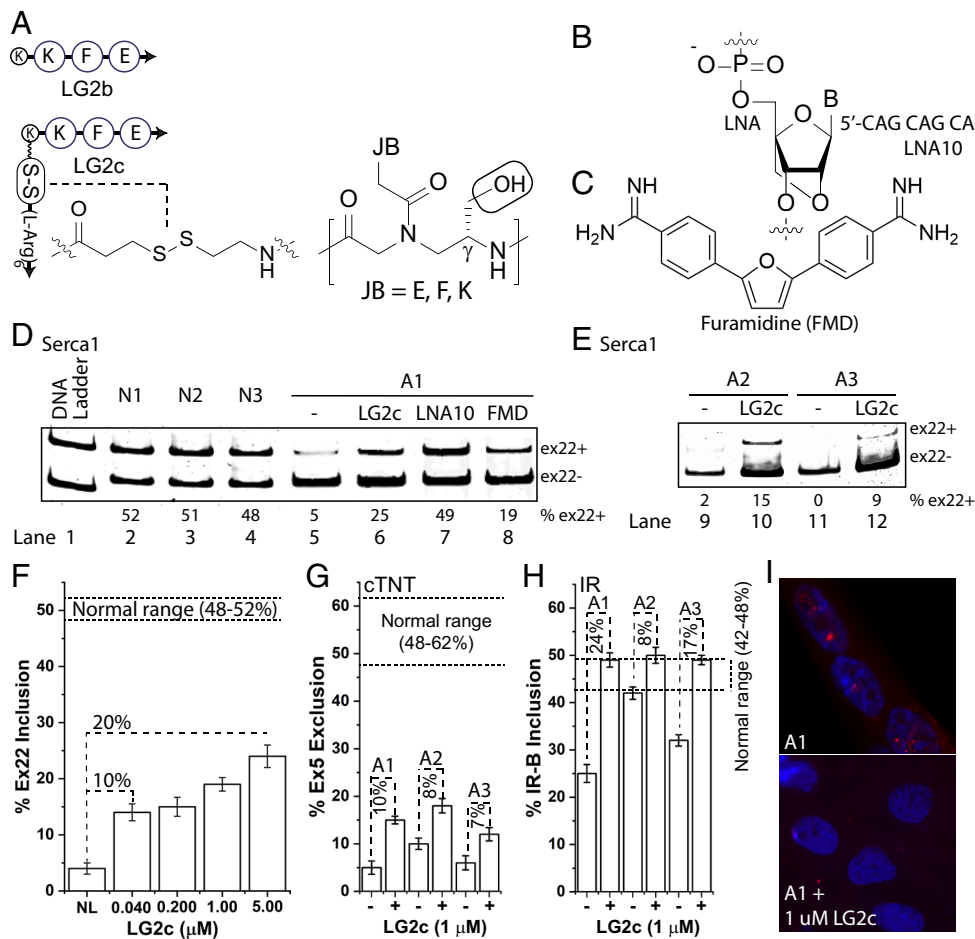
added simultaneously. In the absence of MBNL1, the rCUG98–LG2b complex formed a discrete, strongly shifted band (Fig. 6, lanes 1 & 2). Conversely, the rCUG98–MBNL1 complex appeared as a diffuse smear without a clearly defined band (lane 3), indicative of heterogeneous binding sites—a pattern consistent with previous reports (22). A similar pattern was observed with the negative control rCAG126, indicating that while MBNL1 bound to rCAG126 (lane 9), it was not displaced by LG2b (lane 10), which does not bind rCAG126 (lane 7 vs. lane 8).

In all three competitive binding formats—displacement (lane 4), prevention (lane 5), and competitive (lane 6)—a distinct shifted band consistent with the rCUG98–LG2b complex was observed (lane 2 vs. lane 1). Adding LG2b to the (MBNL1 + rCUG98) sample restored the rCUG98–LG2b band (lane 4 vs. lane 3). This demonstrates LG2b's ability to effectively disrupt MBNL1 binding, whether MBNL1 was prebound (displacement) or introduced subsequently (prevention). Notably, in the displacement assay, the intensity of the rCUG98–LG2b complex band was approximately threefold lower compared to the prevention and competitive assays (lane 4 vs. lanes 5 and 6), indicating that displacement of prebound MBNL1 by LG2b occurs less efficiently. This observation aligns with previous findings by Zimmerman (37) for small-molecule ligands, where the  $IC_{50}$  for displacement (560 nM) was nearly two orders of magnitude higher than for prevention (3 nM), reinforcing that preemptive ligand binding is significantly more effective in inhibiting MBNL1–RNA interactions.

#### Restoration of RNA Missplicing in DM1 Patient-Derived Myotubes.

To determine whether the ligand could rescue DM1-associated RNA missplicing, we developed LG2c, an LG2b analog incorporating a disulfide-linked C-terminal hexa-arginine domain to enhance cellular uptake and facilitate intracellular cargo release (38). Additionally, the  $\gamma$ -miniPEG sidechain was replaced with a hydroxymethyl group to improve RNA binding by reducing steric hindrance (Fig. 7A). DM1 patient-derived myotubes were treated with LG2c alongside positive controls (SI Appendix, Figs. S11 and S12), locked nucleic acid (LNA, Fig. 7B) (39) and small-molecule furamidine (FMD, Fig. 7C) (19), followed by total RNA extraction, cDNA synthesis, and PCR analysis of three genes: Serca1 (sarcolemmal/endoplasmic reticulum calcium ATPase 1), cTNT (cardiac troponin T), and IR (insulin receptor), chosen for their known splicing defects and associations with DM1 symptoms. Serca1 is linked to delayed muscle relaxation (40), cTNT to cardiac conduction defects and arrhythmia (2), and IR to insulin resistance (41). Fig. 7D–H present EMSA results comparing LG2c-treated myotubes with positive controls and unaffected cells, while Fig. 7I shows representative nuclear staining of treated and untreated cells.

Inspection of Fig. 7D reveals that at 1  $\mu$ M, LG2c restored inclusion of Serca1 exon 22 (ex22+) in affected myotubes A1 by approximately 20%, compared to 44% for LNA (125 nM) and 14% for FMD (1  $\mu$ M). While LG2c's effect is comparable to FMD, it is notably weaker than LNA. However, unlike LNA, which required a delivery vehicle (Lipofectamine 2000), both LG2c and FMD were effective without one, potentially offering advantages for downstream applications.



**Fig. 7.** Assessment of LG2c's ability to restore RNA missplicing. (A) LG2c shares LG2b's recognition elements but features a C-terminal hexa-arginine domain linked via a disulfide bridge to enhance cellular uptake and promote intracellular cargo release. It also featured a  $\gamma$ -hydroxymethyl sidechain in place of a miniPEG, designed to improve RNA binding. (B and C) Show the chemical structures of the positive controls: locked nucleic acid (LNA) and small-molecule furamidine (FMD), respectively. (D) ESMA results showing exon 22 (ex22+) inclusion in Serca1 mRNA from DM1 patient-derived myotubes (A1) treated with 1  $\mu$ M LG2c, compared to LNA (0.125  $\mu$ M) and FMD (1  $\mu$ M), and (normal) unaffected controls (N1, N2, N3). (E) Treatments of additional DM1 patient-derived myotubes with LG2c (1  $\mu$ M). (F) Concentration-dependent effects of LG2c on ex22+ in Serca1 mRNA of A1 myotubes. (G and H) Effects of LG2c (1  $\mu$ M) on exon 5 (ex5-) exclusion in cTNT and the inclusion of IR-B isoform of IR mRNA, respectively. (I) Fluorescent detection of nuclear foci of DM1 patient-derived myotubes (A1) with and without treatment with 1  $\mu$ M LG2c. Blue: nuclei, red: nuclear foci.

Similar trends were observed in two additional DM1 myotube lines, A2 and A3 (Fig. 7E), albeit with slightly reduced responses. LG2c's splicing correction was shown to be concentration-dependent, increasing from ~10% at 40 nM to ~20% at 5  $\mu$ M (Fig. 7F and SI Appendix, Fig. S13). Comparable improvements were observed in the exclusion of exon 5 (ex5-) of cTNT (Fig. 7G and SI Appendix, Fig. S14) and the inclusion of IR-B isoform of IR (Fig. 7H and SI Appendix, Fig. S15), with the latter showing a more pronounced effect. Supporting these findings, nuclear staining revealed a marked reduction in nuclear foci following treatment with 1  $\mu$ M LG2c compared to untreated cells, with representative images shown in Fig. 7I. Notably, the reduction in nuclear foci appears more substantial than the extent of splicing restoration, though the reason for this discrepancy remains unclear. Nonetheless, the data consistently demonstrate that LG2c improves splicing defects in DM1 myotubes, with efficiency comparable to that of small-molecule FMD.

## Conclusions

This study marks a significant step forward in the rational design of RNA targeting molecules, showcasing how short nucleic acid ligands can achieve high specificity through bifacial recognition. While challenges remain, particularly in enhancing cellular uptake and achieving full functional rescue, the ability to selectively target pathogenic RNA repeats without affecting normal transcripts highlights the downstream therapeutic and diagnostic potentials of this approach. Such a molecular platform is not only suitable for rCUG-repeats or related RNA triplet-repeat expansions but also holds promise for targeting a diverse array of native RNA structures with essential physiological roles.

## Materials and Methods

**Monomers and Oligomers.** The synthesis of  $\gamma$ PNA monomers with Janus bases E, F, and K have been reported (31); and oligomer synthesis is detailed in the Supporting Information.

**Molecular Dynamics Simulations of the LG1 Series Ligands.** The eKF and EKF triads were constructed using Chimera and optimized with the HF/6-31G basis set in Gaussian. The RNA-ligand complex (RNA-LIG-RNA) was built using the NAB module and incorporated these triads. A modified MeyPNA backbone from crystal structure 3PAO was grafted onto the helical RNA framework, forming a 12-base pair triple helix. This structure underwent energy minimization via the steepest descent method, was solvated with TIP3P water and ionized to maintain physiological conditions. After a second energy minimization, the system was gradually heated to 300 K with a 25 kcal/mol/Å position restraint on the RNA-ligand complex. Six short simulations progressively released these restraints, culminating in an unrestrained 500 ns NPT simulation at 300 K and 1 bar using a Nose-Hoover thermostat and a Parrinello-Rahman barostat. All MD simulations were performed in GROMACS with a 2 fs time step. Electrostatic interactions were calculated using the particle mesh Ewald method with a 10 Å cutoff. The Amber99-parmbcs0 force field with  $\chi$ OL3 modifications was used for RNA, while ligand parameters were derived from RED-generated amber atom types and charges.

**RNA Sample Preparation.** RNA targets with varying rCUG-repeat lengths and CG-rich terminal regions were selected, with the latter to enhance duplex stability under physiological conditions (Fig. 3A). Both perfect-match (U<>U) and mismatch variants (A<>A, C<>C, G<>G) were examined (Fig. 3B). Intermolecular duplexes, formed by hybridizing two complementary RNA strands, were used instead of hairpins to allow more flexible target design and to compare ligand binding to single-stranded RNA vs. r(CUG/GUC)<sub>n</sub>-duplexes. All samples were prepared in physiological relevant buffer (10 mM NaPi, 137 mM NaCl, 150 mM KCl, 2 mM MgCl<sub>2</sub>, and pH7) and incubated with ligands at 37 °C.

**Preparation of rCUG98 and rCAG126 Transcripts.** Following established protocols (22), rCUG98 and rCAG126 transcripts were enzymatically synthesized using T7 RNA polymerase with plasmids pCUGexp-tail and pCAGexp-tail as templates. The plasmids were linearized with *Xba*I (rCUG98) or *Sac*I (rCAG126) digestion at 37 °C for 2 h, separated on a 1% agarose gel, and purified using the QIAquick Gel Extraction Kit. Transcription was performed in a 40  $\mu$ L reaction containing 2  $\mu$ g of digested plasmid DNA, 3.5 mM rNTPs, 4  $\mu$ L enzymatic mix (T7 RNA polymerase with RNase inhibitors, NEB), and 1 $\times$  reaction buffer. The reaction proceeded at 37 °C for 4 h, followed by annealing at 70 °C for 5 min and rapid cooling to 0 °C. Transcripts were purified via 10% nondenaturing PAGE, visualized by UV-shadowing, excised, and extracted into 1 $\times$  PBS buffer.

**Electrophoretic Mobility-Shift Assay (EMSA).** RNA and RNA–ligand complexes were separated on 8 to 15% nondenaturing PAGE (1 $\times$  TBE, 10 V/cm), stained with SYBR-Gold, and imaged. Band intensities were quantified using ImageJ. For  $K_D$  determination, binding assays were performed at constant RNA concentration with varying ligand levels, and fraction bound vs. log[ligand] data were fitted using a Hill binding model.

**Surface Plasmon Resonance (SPR).** SPR measurements were performed with a four-channel Biacore T200 optical biosensor system. Flow cell 1 was left blank, while flow cells 2 to 4 were immobilized with 5′-biotin-labeled RNA sequences. The SPR experiments were performed at 25 °C in degassed and filtered 1 $\times$  PBS, 2 mM MgCl<sub>2</sub>, 150 mM KCl; pH 7.4. Solutions of different known LG2b concentrations were injected over the immobilized RNA surface at a flow rate of 100  $\mu$ L/min. Compound solution flow was then replaced by buffer flow resulting in dissociation of the complex. After each cycle, the sensor chip surface was regenerated with 1M KCl as a regeneration buffer for 60 s followed by three running buffer injections (each 60 s) to yield unbound RNA and a stable baseline for the following cycles. The reference response from the blank cell was subtracted from the response in each flow cell containing RNA to give a signal (RU, response units) that is directly proportional to the amount of bound compound.

#### Atomic Force Microscopy (AFM).

**Sample preparation.** Each sample contained 30 ng of R65/R65-LG2b, 45 ng of double-stranded DNA, and excess water, adjusted to a total volume of 20  $\mu$ L. A freshly cleaved mica surface was rinsed with deionized water, blow-dried with compressed air, and treated with 20  $\mu$ L of 16.7  $\mu$ M spermidine. The spermidine was then blow-dried and rinsed with deionized water. A drop of 20  $\mu$ L of the sample was deposited on the mica surface. After an incubation time of 5 min, the mica surface was blow-dried using compressed air, rinsed with deionized water and dried once again before AFM imaging.

**Imaging.** All AFM images were acquired using a Cypher ES SPM (Asylum Research, Oxford Instruments) in AC Air Topography mode at 25 °C. ScanAsyst-fluid+ probes (Bruker Corp., USA) with a stiffness of 0.7 N/m were used. A pulsed blue laser (BlueDrive) with a  $\sim$ 5  $\mu$ m circular spot size was focused on the cantilever base for photothermal excitation, while a superluminescent diode (SLD) laser (3  $\times$  9  $\mu$ m spot size) was used for detecting cantilever deflections.

**Image analysis.** Image analysis was performed using custom software developed in MATLAB and CellProfiler, with histograms generated in OriginLab. Height images were extracted from raw IGO Pro binary wave files (.ibw) and converted into 16-bit intensity images for processing. Using CellProfiler's "EnhanceOrSuppressFeatures" module, long, thin features (Control B-DNA) and spot-like features (RNA particles) were enhanced. Particles were identified based on intensity thresholds, determined using the Global Otsu method, which minimizes variance within foreground and background pixel classes. Control B-DNA and RNA particles were distinguished via size and shape filtering. Identified particles were then converted into image masks and overlaid onto the original height image in MATLAB. Overlapping pixels were extracted and plotted as height histograms in OriginLab. Finally, CellProfiler's "MeasureObjectShapeSize" module was used to measure the major and minor axis lengths of particles. The aspect ratio (major/minor axis) was calculated and plotted as a histogram in OriginLab.

**Competitive Binding of LG2b and MBNL1 with rCUG98.** Samples were prepared in a physiologically relevant buffer with 0.1  $\mu$ g/ $\mu$ L BSA and incubated at 37 °C. Binding conditions: D (Displacement)–rCUG98 and MBNL1 were

preincubated, followed by LG2b addition; P (Prevention)–CUG98 and LG2b were pre-incubated, followed by MBNL1 addition; Competitive—all three components were incubated together. rCUG98, MBNL1, and LG2b concentrations were 0.01  $\mu$ M (0.5  $\mu$ M ligand binding sites), 0.50  $\mu$ M, and 2.50  $\mu$ M, respectively. Samples were resolved on an 8% nondenaturing PAGE gel (15 V/cm, 3 h) and stained with SYBR-Gold.

**Cell Culture.** Fibroblasts derived from DM1 patients (cell lines A1: GM03132, A2: GM03759 and A3: GM03989 expressing DMPK transcript with  $\sim$ 2,000 CTG-repeats) and control fibroblasts derived from non-DM1 patients (cell lines N1: GM03377, N2: GM03651, N3: GM03652) were purchased from the Coriell Cell Repositories. Cells were grown in Dulbecco's modified Eagle's Medium (DMEM) (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific), 1% antibiotic antimycotic solution (Sigma-Aldrich), and 1% nonessential amino acids solution (Thermo Fisher Scientific), in 5% CO<sub>2</sub> atmosphere at 37 °C. At the differentiation stage, MyoD-transduced fibroblasts were cultured in a differentiation medium containing DMEM/F-12 (HAM) 1:1 (Thermo Fisher Scientific), 2% donor equine serum (DES) (Cytiva), and 1% antibiotic antimycotic solution. The medium was changed every 2 to 3 d.

**MyoD Transduction.** The Adenovirus MyoD was obtained from Vector Biolabs with an infectivity of 1  $\times$  10<sup>7</sup> MOI/ $\mu$ L. The fibroblasts at 30% confluence grown in normal growth medium were transduced with 2% MyoD (20 MOI/cell). Seventy-two hours after incubation in 5% CO<sub>2</sub> atmosphere at 37 °C, the virus-containing transduction medium was replaced with fresh differentiation medium. The MyoD-transduced fibroblasts were cultured for 10 d to differentiate into myotubes.

**Treatments.** LG2c (40 nM to 5  $\mu$ M) and Furamidine (1  $\mu$ M) were incubated with myotubes after 10 d of differentiation. Cell transfection experiments were performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions with PO-LNA-CAG-10 at 125 nM final concentration. The total RNA was isolated from cells after 72 h of treatment, using RNeasy (Qiagen) with in-column gDNA removal according to the manufacturer's protocol and the RNA concentrations were determined using NanoDrop (Thermo Fisher Scientific).

**PCR Reactions.** First-strand cDNA was synthesized using 1  $\mu$ g of total RNA using QuantiTect Reverse Transcription Kit (Qiagen) according to the manufacturer's instructions and subsequent PCR was performed using 2  $\mu$ L of the cDNA. All PCRs were performed at an initial holding temperature of 95 °C for 3 min, 30 cycles of denaturation at 95 °C for 30 s, annealing at 50 to 60 °C for 30 s, and elongation at 72 °C for 30 s, and a final elongation temperature of 72 °C for 10 min and 4 °C holding temperature. Primer sequences and PCR conditions are specified in *SI Appendix, Table S2*. The RT-PCR products were separated in 5% nondenaturing PAGE (15 V/cm, 1.5 h) and stained with SYBR-Gold. The amounts of PCR products were quantified by ImageJ.

**Nuclear Staining.** Myotubes were treated with 1  $\mu$ M LG2c for 3 d, fixed with 4% paraformaldehyde, permeabilized with 70% EtOH, hybridized with Cy3-(CAG)<sub>3</sub> probe, and imaged on an Olympus-IX80 with a 100 $\times$ -oil objective. Blue: nuclei, red: nuclear foci.

**Data, Materials, and Software Availability.** Study data are included in the article and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank Professor Charles Thornton for providing pdwdT98CTG and pdwT7127CAG plasmids and for his guidance, and the following funding agencies for their financial support: The NIH (R21NS098102), the NSF (CHE-1609159), and the DSF Charitable Foundation (DSF-MCS Cancer Moonshot and DSF-MCS GPS2).

Author affiliations: <sup>a</sup>Department of Chemistry and Institute for Biomolecular Design and Discovery, Carnegie Mellon University, Pittsburgh, PA 15213; <sup>b</sup>Department of Chemistry, Indian Institute of Science Education and Research, Pune, Maharashtra 411008, India; <sup>c</sup>Department of Chemistry, Georgia State University, Atlanta, GA 30302; and <sup>d</sup>School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore

- J. Hartman, T. Patki, N. E. Johnson, Diagnosis and management of myotonic dystrophy type 1. *JAMA* **331**, 1227–1228 (2024).
- A. V. Philips, L. T. Timchenko, T. A. Cooper, Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**, 737–741 (1998).
- R. N. Kanadia *et al.*, A muscleblind knockout model for myotonic dystrophy. *Science* **302**, 1978–1980 (2003).
- K. Sobczak, M. de Mezer, G. Michlewski, J. Krol, W. J. Krzyzosiak, RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res.* **31**, 5469–5482 (2003).
- B. M. Mooers, J. S. Logue, J. A. Berglund, The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16626–16631 (2005).
- R. Parkesh, M. Fountain, M. D. Disney, NMR spectroscopy and molecular dynamic simulations of  $\mu$ (CCGCGCGG)<sub>2</sub> reveal a dynamic UU internal loop found in myotonic dystrophy type 1. *Biochemistry* **50**, 599–601 (2011).
- H. Du *et al.*, Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat. Struct. Mol. Biol.* **17**, 187–193 (2010).
- N. M. Kuyumcu-Martinez, G. S. Wang, T. A. Cooper, Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol. Cell* **28**, 68–78 (2007).
- M. Banez-Coronel, L. P. W. Ranum, Repeat-associated non-AUG (RAN) translation: Insights from pathology. *Lab. Invest.* **99**, 929–942 (2019), 10.1038/s41374-019-0241-x.
- I. Malik, C. P. Kelley, E. T. Wang, P. K. Todd, Molecular mechanism underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 (2021).
- M. Scudato *et al.*, Genome editing of expanded CTG repeats within the human *DMPK* gene reduces nuclear RNA foci in the muscle of DM1 mice. *Mol. Ther.* **27**, 1372–1388 (2019).
- B. S. Pinto *et al.*, Impeding transcription of expanded microsatellite repeats by deactivated Cas9. *Mol. Cell* **68**, 479–490 (2017).
- T. M. Wheeler *et al.*, Targeting nuclear RNA for *in vivo* correction of myotonic dystrophy. *Nature* **488**, 111–115 (2012).
- R. Batra *et al.*, Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. *Cell* **170**, 899–912 (2017).
- J. L. Childs-Disney *et al.*, Targeting RNA structures with small molecules. *Nat. Rev. Drug Discov.* **21**, 736–762 (2022).
- S. G. Rzuczek *et al.*, Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **13**, 188–193 (2017).
- A. J. Angelbello *et al.*, Precise small-molecule cleavage of an  $\mu$ (CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7799–7804 (2019).
- A. H. Jahromi *et al.*, A novel CUGexp-MBNL1 inhibitor with therapeutic potential for myotonic dystrophy type 1. *ACS Chem. Biol.* **8**, 1037–1043 (2013).
- J. R. Jenquin *et al.*, Furamidine rescues myotonic dystrophy type 1 associated mis-splicing through multiple mechanisms. *ACS Chem. Biol.* **13**, 2708–2718 (2018).
- K. Reddy *et al.*, A CTG repeat-selective chemical screen identifies microtubule inhibitors as selective modulators of toxic CUG RNA levels. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20991–21000 (2019).
- M. Pascual-Gilabert, R. Artero, A. Lopez-Castel, The myotonic dystrophy type 1 drug development pipeline: 2022 Edition. *Drug Discov. Today* **28**, 1–11 (2022).
- T. M. Wheeler *et al.*, Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science* **325**, 336–339 (2009).
- M. Egli, M. Manoharan, Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acid Res.* **51**, 2529–2573 (2023).
- S. Kovachka *et al.*, Small molecule approaches to targeting RNA. *Nat. Rev. Chem.* **8**, 120–135 (2024).
- A. Khorova, J. K. Watts, The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **35**, 238–248 (2017).
- A. Dragulescu-Andrasi *et al.*, A simple gamma-backbone modification preorganizes peptide nucleic acid into a helical structure. *J. Am. Chem. Soc.* **128**, 10258–10267 (2006).
- J. I. Yeh *et al.*, Crystal structure of chiral gammaPNA with complementary DNA strand: Insights into the stability and specificity of recognition and conformational preorganization. *J. Am. Chem. Soc.* **132**, 10717–10727 (2010).
- S. A. Thadke *et al.*, Design of bivalent nucleic acid ligands for recognition of RNA-repeated expansion associated with Huntington's disease. *Biochemistry* **57**, 2094–2108 (2018).
- S. A. Thadke *et al.*, Shape selective bifacial recognition of double helical DNA. *Commun. Chem.* **1**, 79 (2018).
- P. E. Nielsen, Peptide nucleic acid. A molecule with two identities. *Acc. Chem. Res.* **32**, 624–630 (1999).
- S. A. Thadke, J. D. R. Perera, S. W. Thrikawala, D. H. Ly, Synthesis of Janus bases for targeting C-G and U-U pairs of CUG-RNA repeats associated with myotonic dystrophy type 1. *ACS Omega* **10**, 49252–49259 (2025).
- S. W. Santoro, G. F. Joyce, A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4262–4266 (1997).
- J. Nasir *et al.*, Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell* **81**, 811–823 (1995).
- H. Lu, P. J. Tonge, Drug-target residence time: Critical information for lead optimization. *Curr. Opin. Chem. Biol.* **14**, 467–474 (2010).
- K. Bose, C. J. Lech, B. Heddi, A. T. Phan, High-resolution AFM structure of DNA G-wires in aqueous solution. *Nat. Commun.* **9**, 1959 (2018), 10.1038/s41467-018-04016-y.
- P. Schon, Atomic force microscopy of RNA: State of the art and recent advances. *Semin. Cell Dev. Biol.* **73**, 209–219 (2018).
- A. H. Jahromi, M. Honda, S. C. Zimmerman, M. Spies, Single-molecule study of the CUG repeat-MBNL1 interaction and its inhibition by small molecules. *Nucleic Acid Res.* **41**, 6687–6697 (2013).
- E. G. Stanzl, B. M. Trantow, J. R. Vargas, P. A. Wender, Fifteen years of cell-penetrating, guanidinium-rich molecular transporters: Basic science, research tools, and clinical applications. *Acc. Chem. Res.* **46**, 2944–2954 (2013).
- A. Wojtkowiak-Szlachcic *et al.*, Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy. *Nucleic Acids Res.* **43**, 3318–3331 (2015).
- X. Lin *et al.*, Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15**, 2087–2097 (2006).
- R. S. Savkur, A. V. Philips, T. A. Cooper, Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.* **29**, 40–47 (2001).