Supporting Information for:

Structure-specific cleavage of an RNA repeat expansion with a dimeric small molecule is advantageous over sequence-specific recognition by an oligonucleotide

Raphael I. Benhamou[1], Alicia J. Angelbello[1], Ryan J. Andrews[2], Eric T. Wang[3], Walter N. Moss[2], and Matthew D. Disney[1,*]

[1]Department of Chemistry, The Scripps Research Institute, 110 Scripps Way, Jupiter, FL 33458

Email: disney@scripps.edu

[2]Roy J. Carver Department of Biophysics, Biochemistry and Molecular Biology, Iowa State University, Ames, IA 50011, United States of America

[3]Department of Molecular Genetics & Microbiology, Center for NeuroGenetics, UF Genetics Institute, University of Florida, 2033 Mowry Road, Gainesville, FL 32610

Table of Contents:

I. Supplementary Tables & Figures

Table S1. Genes containing r(CCUG)	
Gene	Number of r(CCUG)
AGO1	8
WT CNBP intron 1	7
Mutant CNBP intron1	expansion
MBNL1	8x2

Table S2. Sequences of primers used for PCR			
Gene	Forward Primer (5'-3')	Reverse Prime (5'-3')	Purpose
MAP4K4	CCTCATCCAGTGAGGAGTCG	TGGTGGGAGAAATGCTGTATGC	RT-PCR
IR	CCAAAGACAGACTCTCAGAT	AACATCGCCAAGGGACCTGC	RT-PCR
GAPDH	AAGGTGAAGGTCGGAGTCAA	AATGAAGGGGTCATTGATGG	qPCR
CNBP Intron1	ATTCCAAGGTTGGTTGAAGC	AACCCAAACCAATGAAGCTG	qPCR
CNBP mature mRNA	AAACTGGTCATGTAGCCATCAAC	AATTGTGCATTCCCGTGCAAG	qPCR
AGO1	CCCTAAGATCGACGTGTACCACTA	ACCACTTCCCGGTTGACTCTA	qPCR
MBNL1	TTCATCCACCCCCACATTTA	TTGGCTAGTTGCATTTGCTG	qPCR

Table S3. Summary of RNA secondary structure prediction.	
Gene	ΔG° of r(CCUG) repeat Formation (kcal mol^{-1})
CNBP WT	-22.8
CNBP $CCUG_{2000}$	-1995.6
AGO1	-21.5
MBNL1	-15.5

Table S4. Energy interaction between ASO and gene sequences	
Gene	ΔG° (kcal mol^{-1})
CNBP WT	-37.6
CNBP $CCUG_{2000}$	-36.7
AGO1	-36.9
MBNL1	-20.2

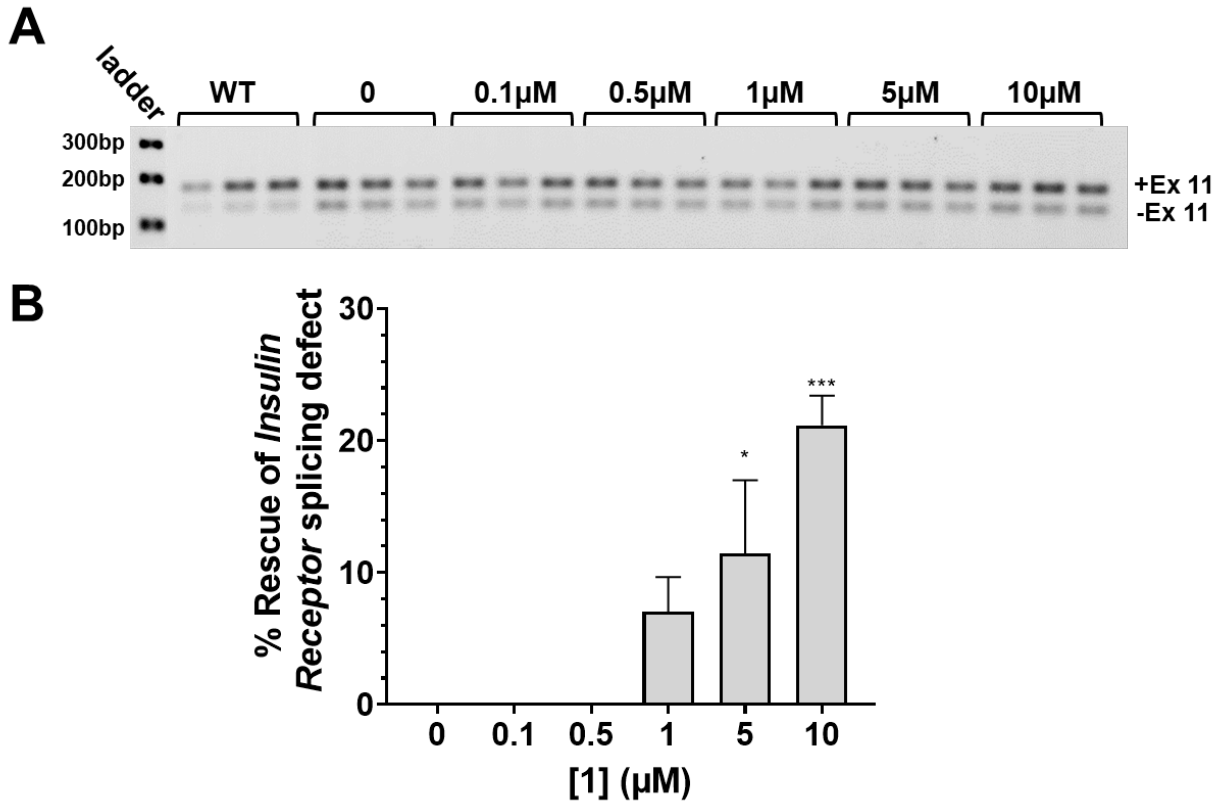


Figure S1: Ability of **1** to rescue splicing events in DM2 fibroblasts. (*A*) Representative gel image of *IR* exon 11 alternative splicing in **1**-treated and untreated DM2 fibroblasts, as assayed by RT-PCR. (*B*) Quantification of RT-PCR analysis of the *IR* exon 11 alternative splicing. Error bars represent SD. **P* < 0.5; ****P* < 0.001; as determined by a one-way ANOVA relative to 0 (n = 3)

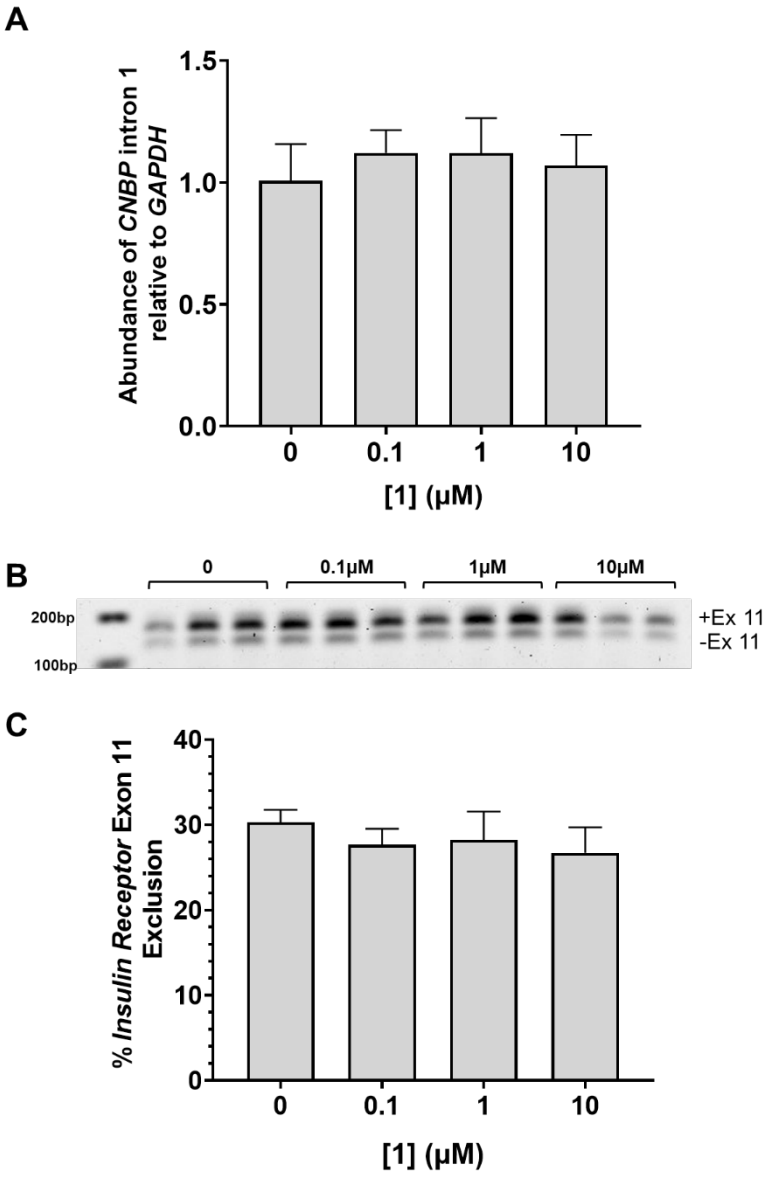


Figure S2: Evaluation of **1** in healthy fibroblasts. (*A*) Effect of **1** on *CNBP* intron 1 levels as measured by RT-qPCR. (*B*) Representative gel image of *IR* exon 11 splicing in WT fibroblasts treated with **2**, as determined by RT-PCR. (*C*) Quantification of gels to assess *IR* exon 11 exclusion. Error bars represent SD (n=3).

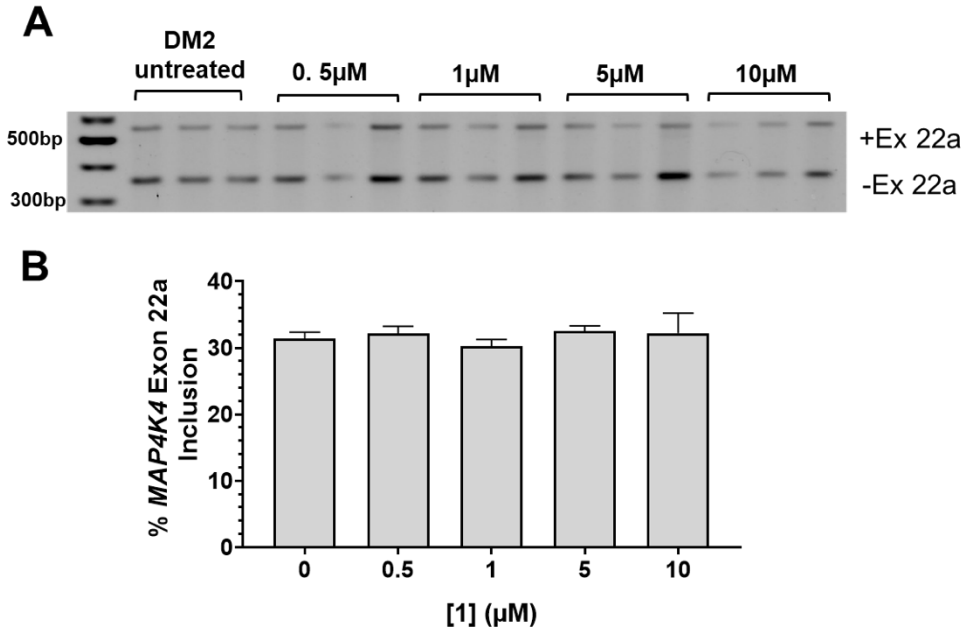


Figure S3: Evaluation of **1** in DM2 fibroblasts. (*A*) Representative gel image of *MAP4K4* exon 22a alternative splicing (non-MBNL1 regulated) in DM2 fibroblasts treated with **1**. (*B*) Quantification of *MAP4K4* exon 22a splicing. Error bars represent SD (n=3).

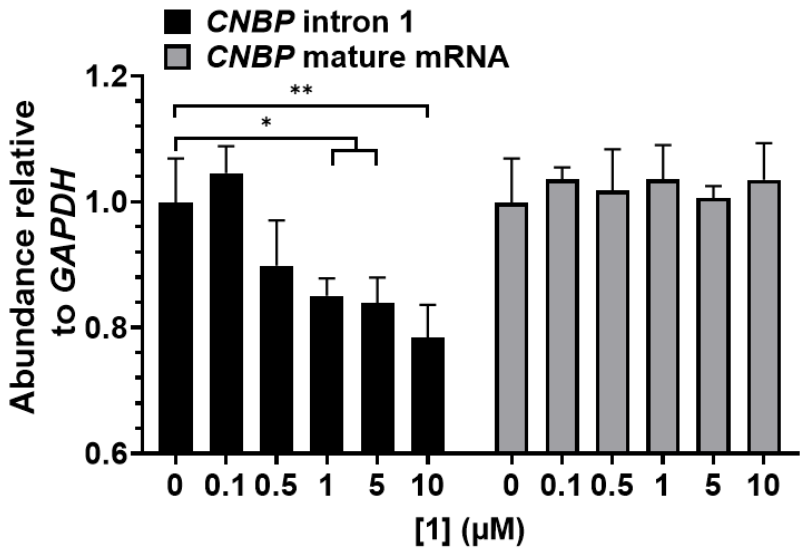


Figure S4: Evaluation of **1** in DM2 fibroblasts. Effect of **1** on *CNBP* intron 1 and *CNBP* mature mRNA levels as measured by RT-qPCR. Error bars represent SD (n=3). **P* < 0.5; ***P* < 0.01; as determined by a one-way ANOVA relative to 0 (n = 3)

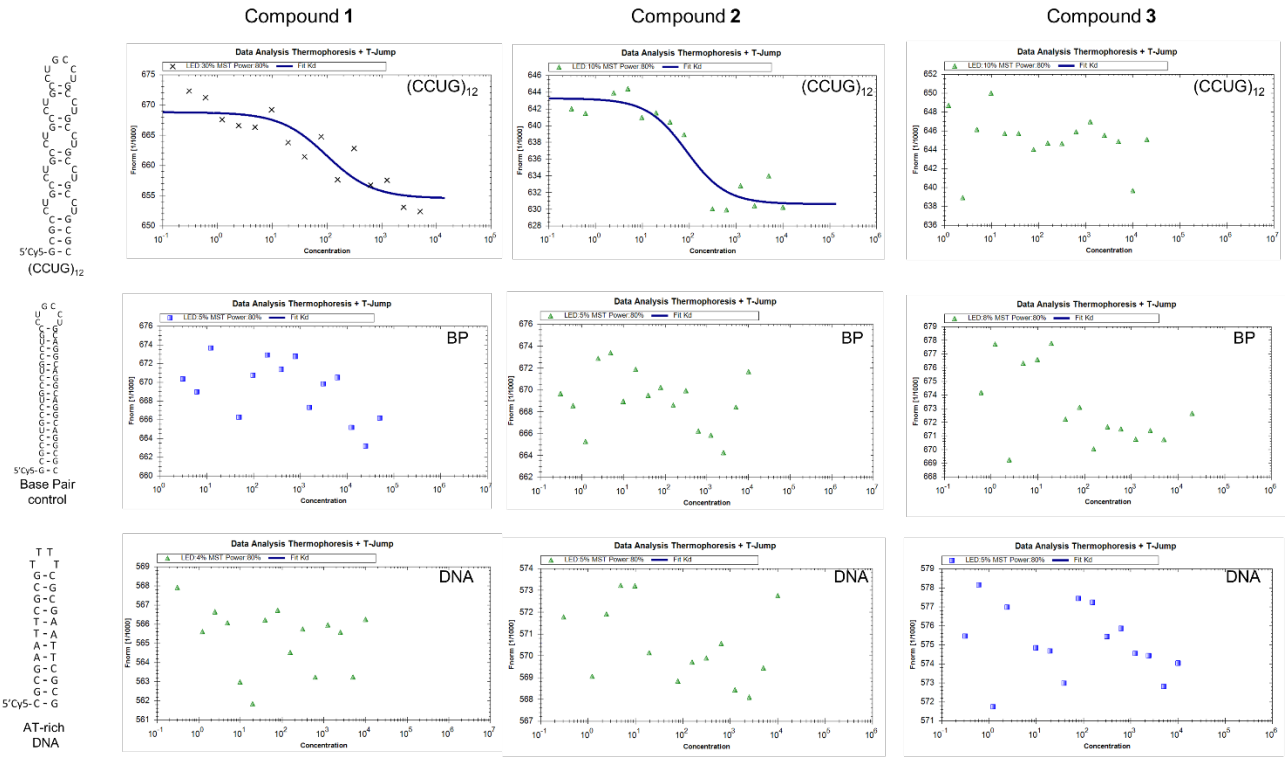


Figure S5: Representative binding affinity data from three replicates of an MST study for compounds **1**, **2**, and **3** binding to r(CCUG)$_{12}$, a fully base paired RNA, and an AT-rich DNA. Binding curves were calculated using a single site model to fit the curve.

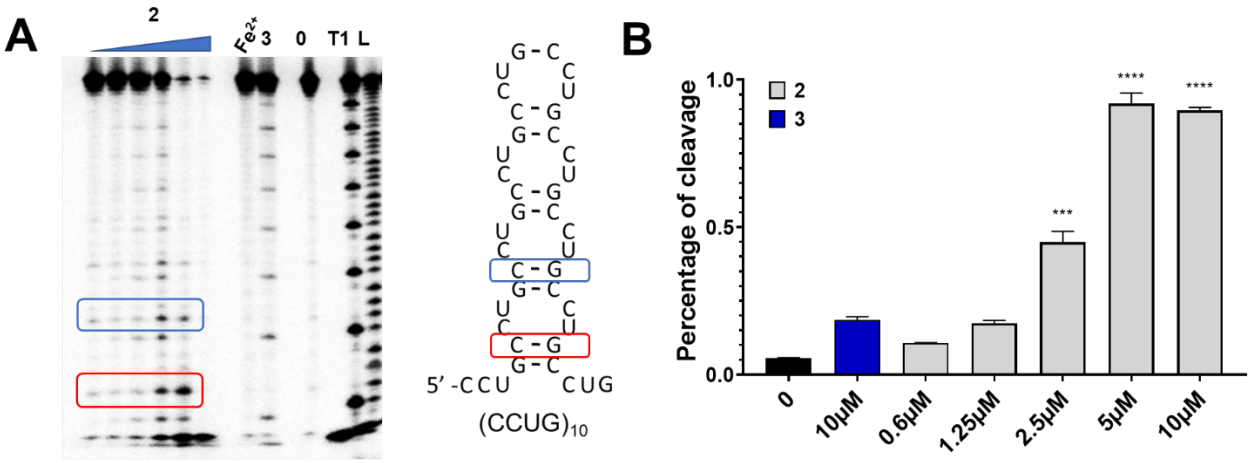


Figure S6: Cleavage of r(CCUG) repeats by **2** *in vitro*. (*A*) *In vitro* cleavage of 5'-^{32}P-labeled r(CCUG)$_{10}$ by **2** and **3**. Blue and red boxes indicate sites cleaved by **2**. (*B*) Average quantification of cleavage from three independent experiments. Error bars represent SD. ***P < 0.001, ****P < 0.0001 as determined by a one-way ANOVA relative to 0 (n = 3).

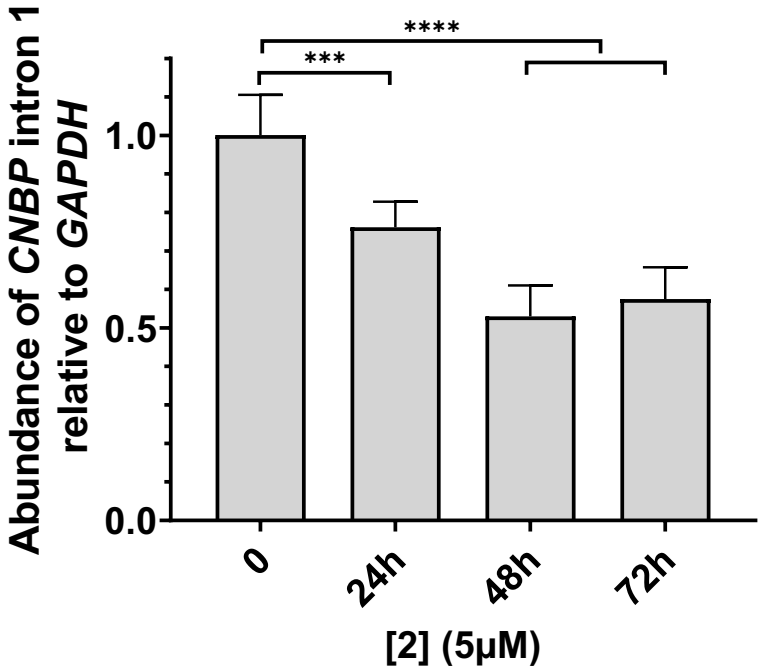


Figure S7: Time course evaluation of **2** in DM2 fibroblasts. RT-qPCR of *CNBP* intron 1 levels in DM2 fibroblasts treated with **2** at 24, 48, and 72 h. Error bars represent SD. ***$P < 0.001$, ****$P < 0.0001$ as determined by a one-way ANOVA relative to 0 (n = 3).

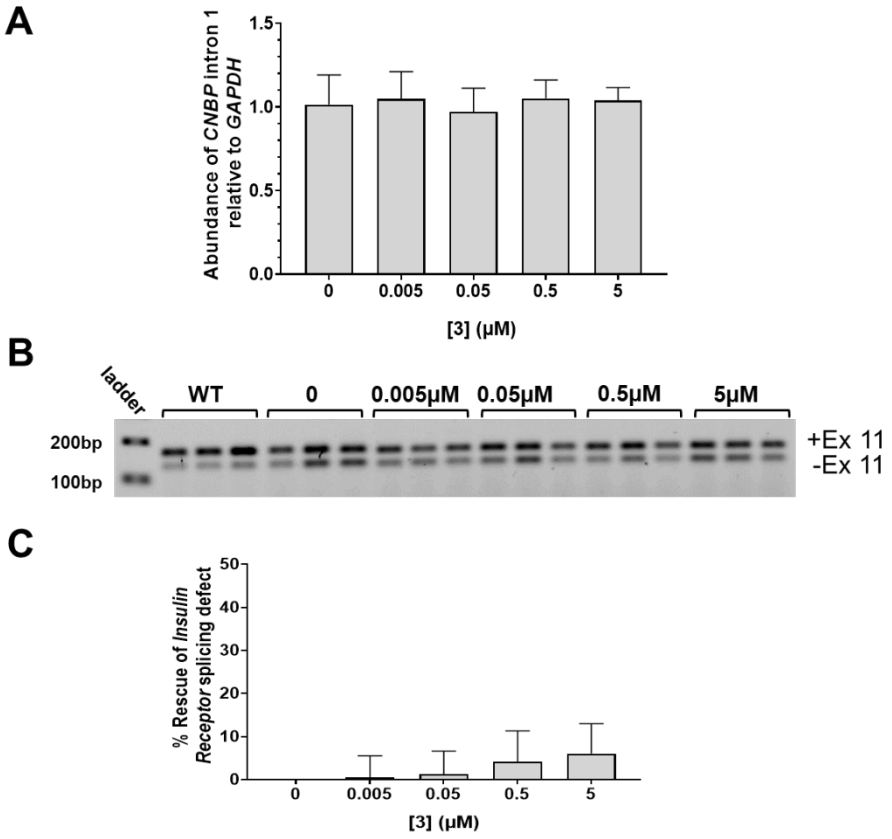


Figure S8: Evaluation of **3** in DM2 fibroblasts. (*A*) RT-qPCR of *CNBP* intron 1 levels in DM2 fibroblasts treated with **3**. (*B*) Representative gel image of the RT-PCR analysis of *IR* exon 11 splicing in DM2 fibroblasts treated with **3**. (*C*) Quantification of *IR* exon 11 splicing in DM2 fibroblasts treated with **3**. Error bars represent SD (n = 3).

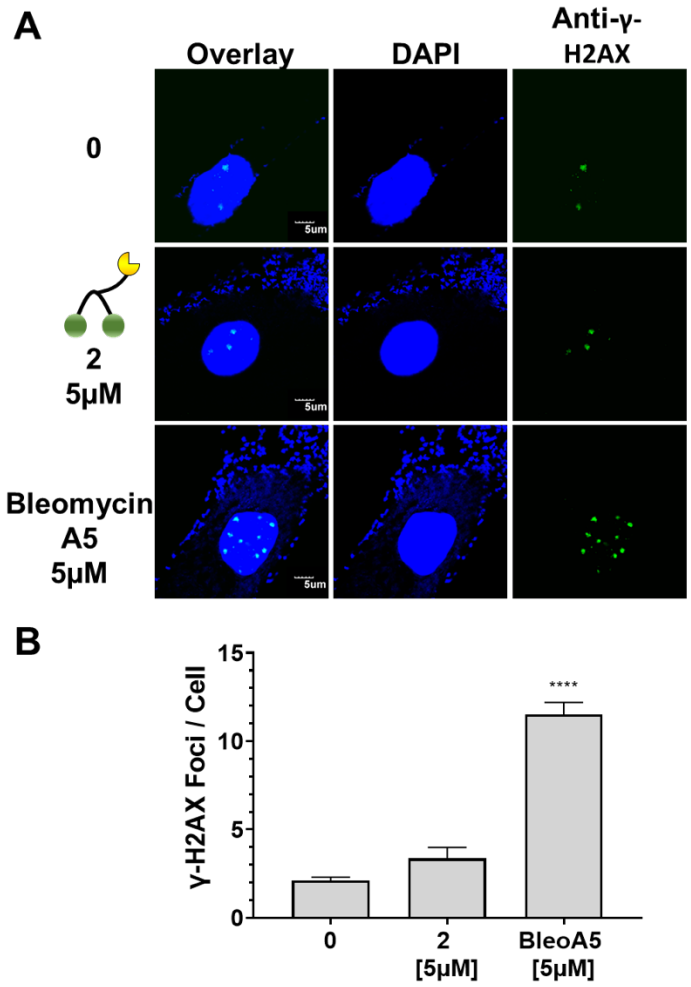


Figure S9: Effects of small molecules on the DNA damage response pathway. (*A*) Representative images from γ-H2AX immunofluorescence to assess DNA damage in DM2 fibroblasts upon treatment with **2** or bleomycin A5. (*B*) Quantification of the number of γ-H2AX foci/cell. Error bars represent SEM. ****$P < 0.0001$, as determined by a one-way ANOVA relative to 0 (n = 3, 40 nuclei counted per replicate).

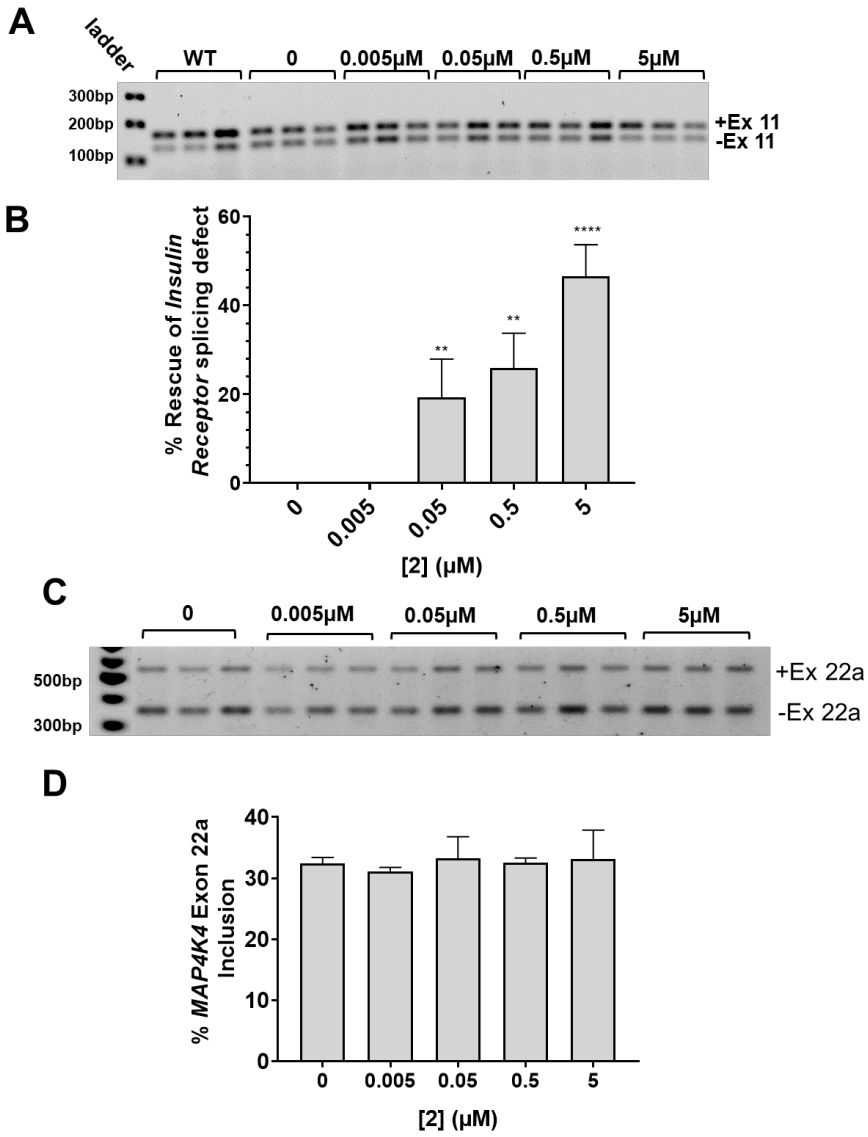


Figure S10: Evaluation of **2** in DM2 fibroblasts. (*A*) Representative gel image of *IR* exon 11 in DM2 fibroblasts treated with **2**. (*B*) Quantification of rescue of the *IR* exon 11 splicing defect by **2**. (*C*) Representative gel image of *MAP4K4* exon 22a splicing (non-MBNL1 regulated) in DM2 fibroblasts treated with **2**. (*D*) Quantification of *MAP4K4* exon 22a splicing. Error bars represent SD. $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by a one-way ANOVA relative to 0 (n = 3).

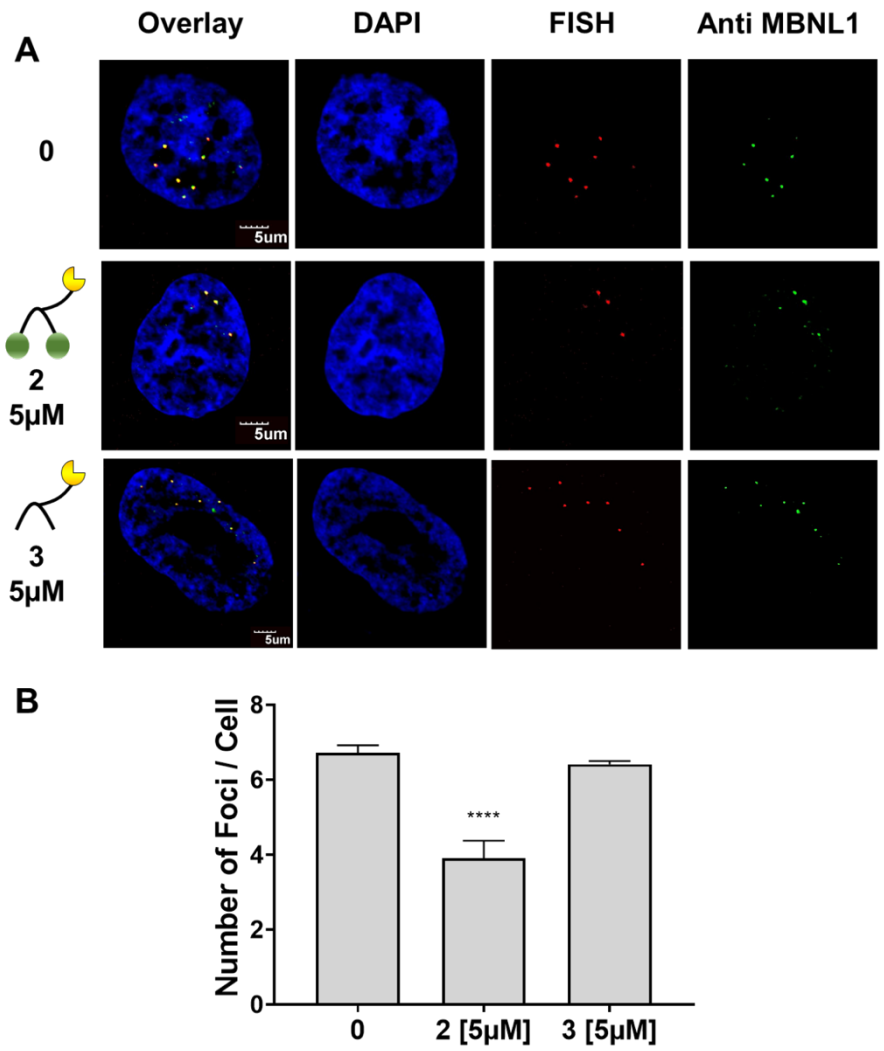


Figure S11: RNA-FISH experiments to assess nuclear foci in DM2 fibroblasts. (A) Representative images of r(CCUG)^{exp}-MBNL1 foci in DM2 fibroblasts treated with **2** and **3**. (B) Quantification of r(CCUG)^{exp}-MBNL1 foci/cell; n = 3 biological replicates, 40 nuclei counted per replicate. Error bars represent SD. ****P < 0.0001, as determined by a one-way ANOVA relative to 0.

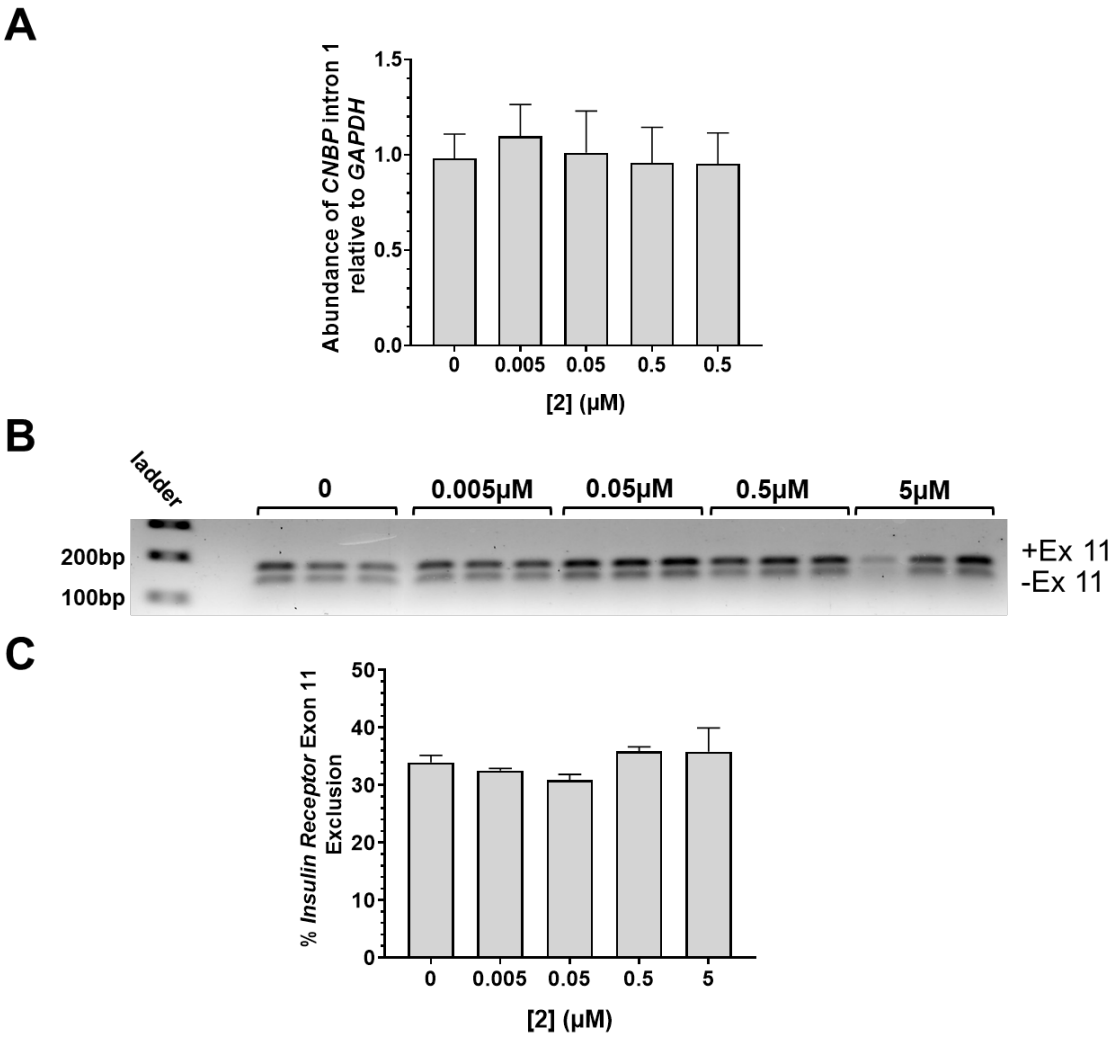


Figure S12: Evaluation of **2** in WT fibroblasts. (*A*) RT-qPCR analysis of *CNBP* intron 1 abundance in WT fibroblast treated with **2**. (*B*) Representative gel image of *IR* exon 11 splicing in WT fibroblast treated with **2**, as determined by RT-PCR. (*C*) Quantification of *IR* exon 11 exclusion. Error bars represent SD (n = 3)

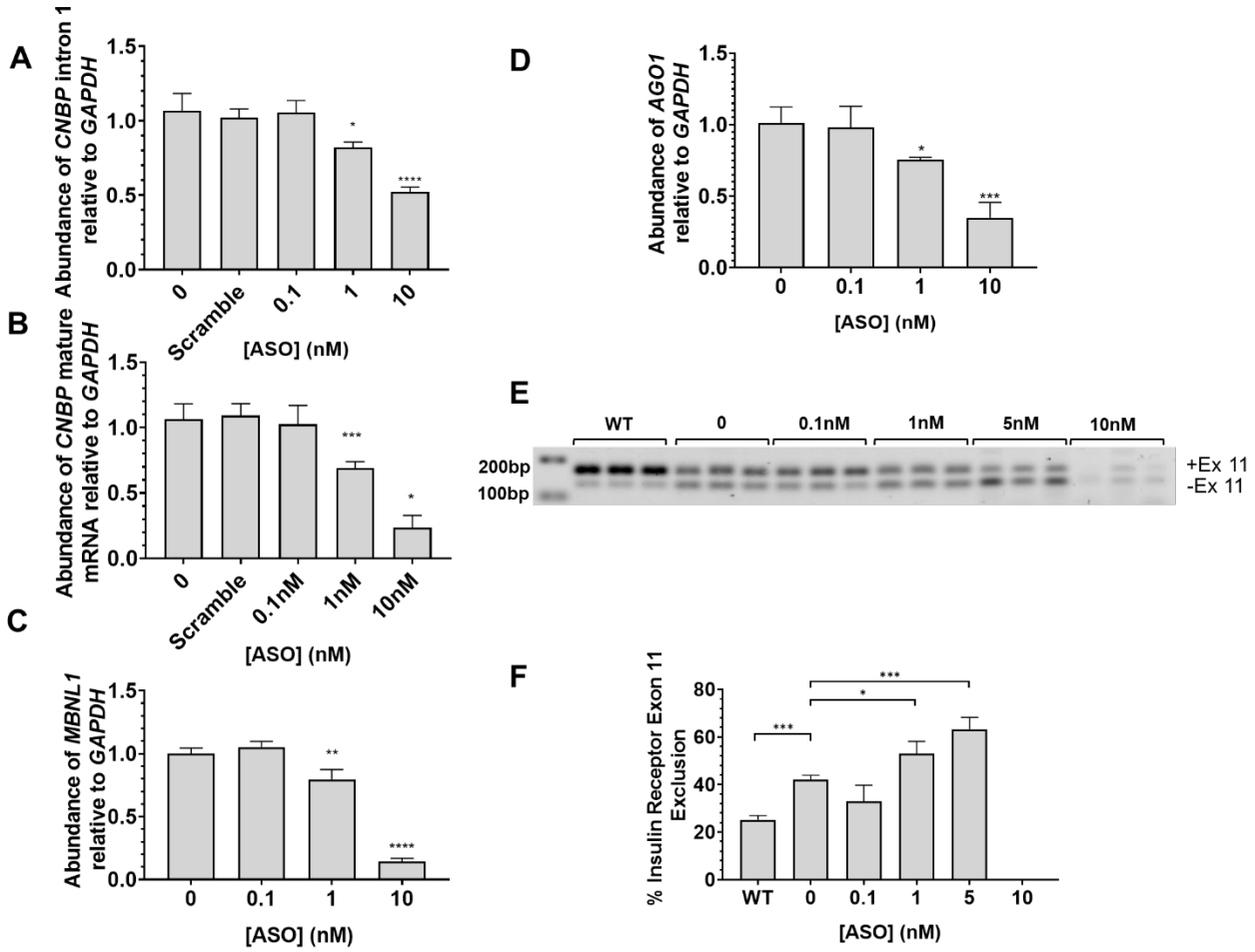


Figure S13: Evaluation of an ASO targeting r(CCUG) sequence in DM2 fibroblasts. (*A*) RT-qPCR of *CNBP* intron 1 abundance in DM2 fibroblasts treated with ASO. The scramble ASO was evaluated at 10 nM. (*B*) RT-qPCR of *CNBP* mature mRNA abundance in DM2 fibroblasts treated with ASO. The scramble ASO was evaluated at 10 nM. (*C*) RT-qPCR of *MBNL1* abundance in DM2 fibroblasts treated with ASO. (*D*) RT-qPCR of *MBNL1* abundance in DM2 fibroblasts treated with ASO. (*E*) Representative gel image of *IR* exon 11 splicing in DM2 fibroblasts treated with **3**. (*F*) Quantification of *IR* exon 11 exclusion. Error bars represent SD. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001 as determined by a one-way ANOVA relative to 0 (n = 3).

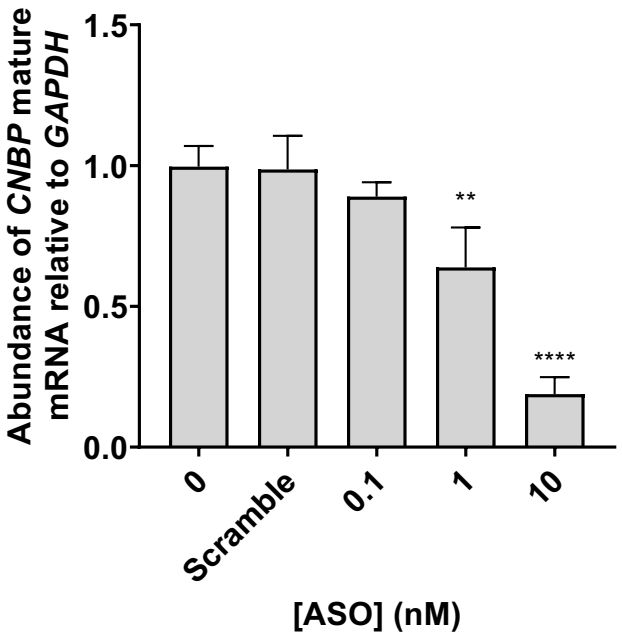


Figure S14: RT-qPCR analysis of *CNBP* mature mRNA abundance in WT fibroblasts. The scramble ASO was evaluated at 10 nM. Error bars represent SD. $^{**}P < 0.01$, $^{****}P < 0.0001$ as determined by a one-way ANOVA relative to 0 (n = 3).

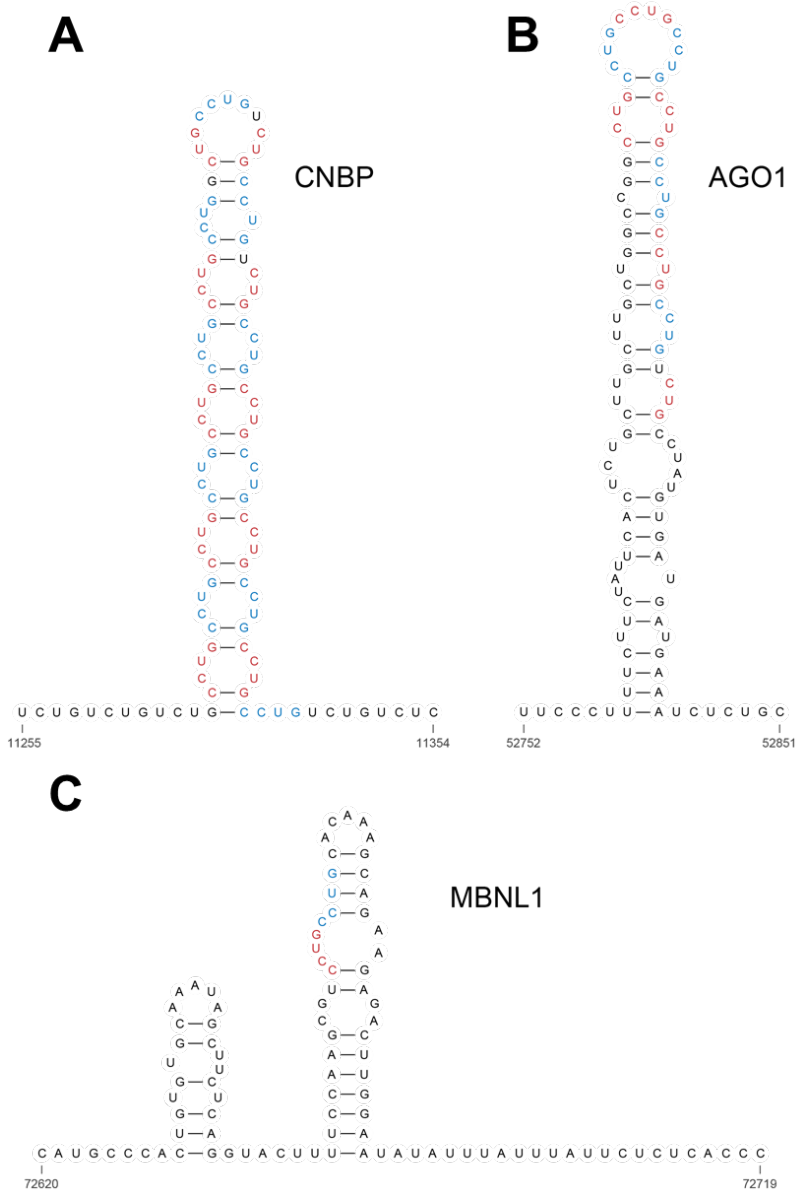
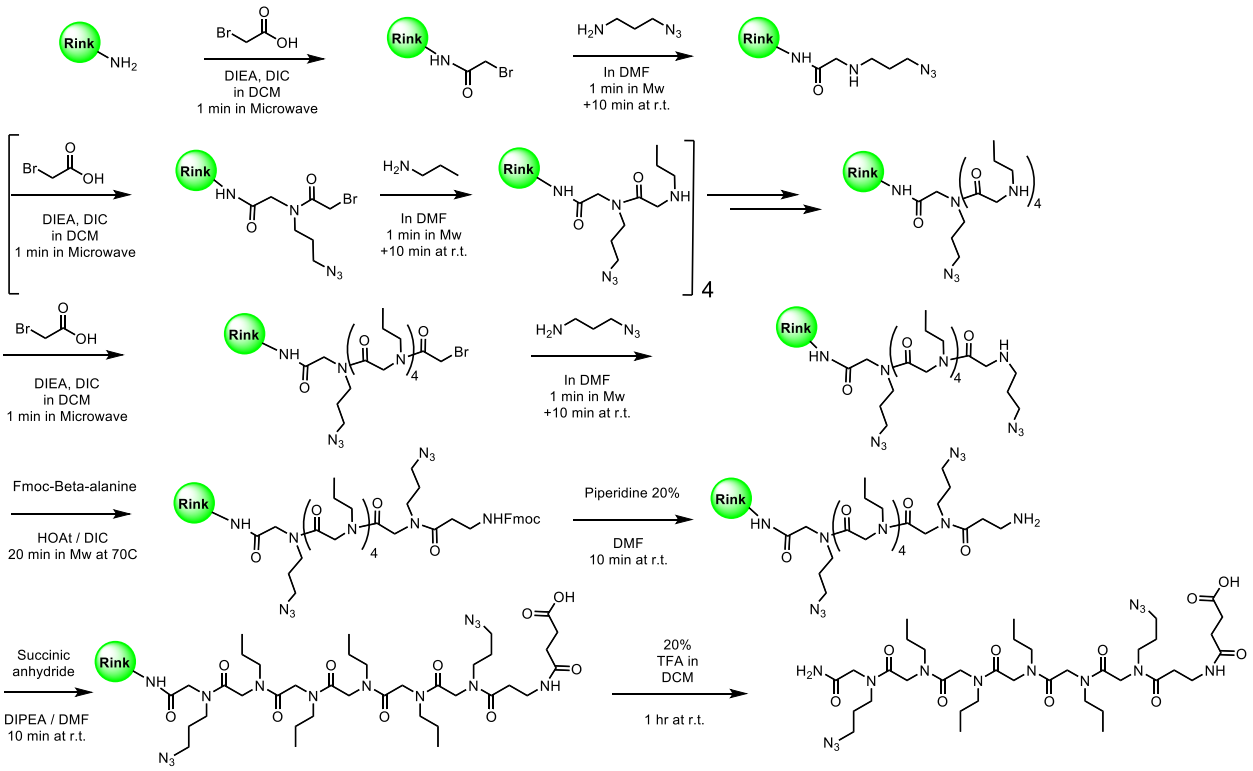


Figure S15: RNA secondary structures of transcripts that have r(CCUG) repeats. (*A*) Structure of *CNBP* WT. (*B*) Structure of *AGO1*. (*C*) Structure of *MBNL1*. r(CCUG) repeats are colored in alternating red and blue colors to easily identify them. Structure predictions were made using the program ScanFold-Scan implementing RNAfold to generate models for windows scanning across repeat regions in each gene plus/minus 100 nt up and down-stream respectively.

II. General protocol for compound synthesis:

Peptoid synthesis general protocol:



Scheme 1: Synthetic route for peptoid synthesis.

The peptoid backbone was synthesized as previously described.[1] Briefly, the peptoid oligomers were synthesized at room temperature. Fmoc-protected Rink amide resin (100-200 mesh) with a substitution level 0.40 mmol/g (500 mg, 200 µmol) was swollen in DCM (5 mL) for 20 min, drained and deprotected with 5 mL of 20% piperidine in DMF for 40 min shaking, followed by draining and then washing with DCM (3x5 mL) and DMF (3x5 mL).

Coupling step: To the resin-bound amine, bromoacetic acid (500mg, 3.65mmol, 5eq), DIC (0.57mL, 3.65mmol, 5eq), and *N*,*N*-Diisopropylethylamine (DIEA) (0.6 mL, 3.65 mmol, 5eq) were added and the resin was microwaved for 1 min using a 700 W microwave set to 10% power and shaken for 10 min. The resin was then drained and washed with DMF (3x5 mL).

Displacement step: a) Introduction of a click counterpart: After washing, DMF (5 mL) and 3-azidopropylamine (210 µL, 2.1mmol, 3eq) were added and the resin was

microwaved for 1 min using a 700 W microwave set to 10% power and shaken for 15 min at room temperature. The resin was then drained and washed with DCM (3x5mL) and DMF (3x5mL). b) Coupling step: After washing, the coupling step was repeated as described above. c) Chain extension with a spacer: After washing with DMF, DMF (5 mL) and propylamine (300 µL, 3.65mmol, 5eq) were added sequentially to the resin and the resin was microwaved for 1 min using a 700 W microwave set to 10% power and shaken for 15 min at room temperature. The resin was then drained and then washed with DCM (3x5mL) and DMF (3x5mL).

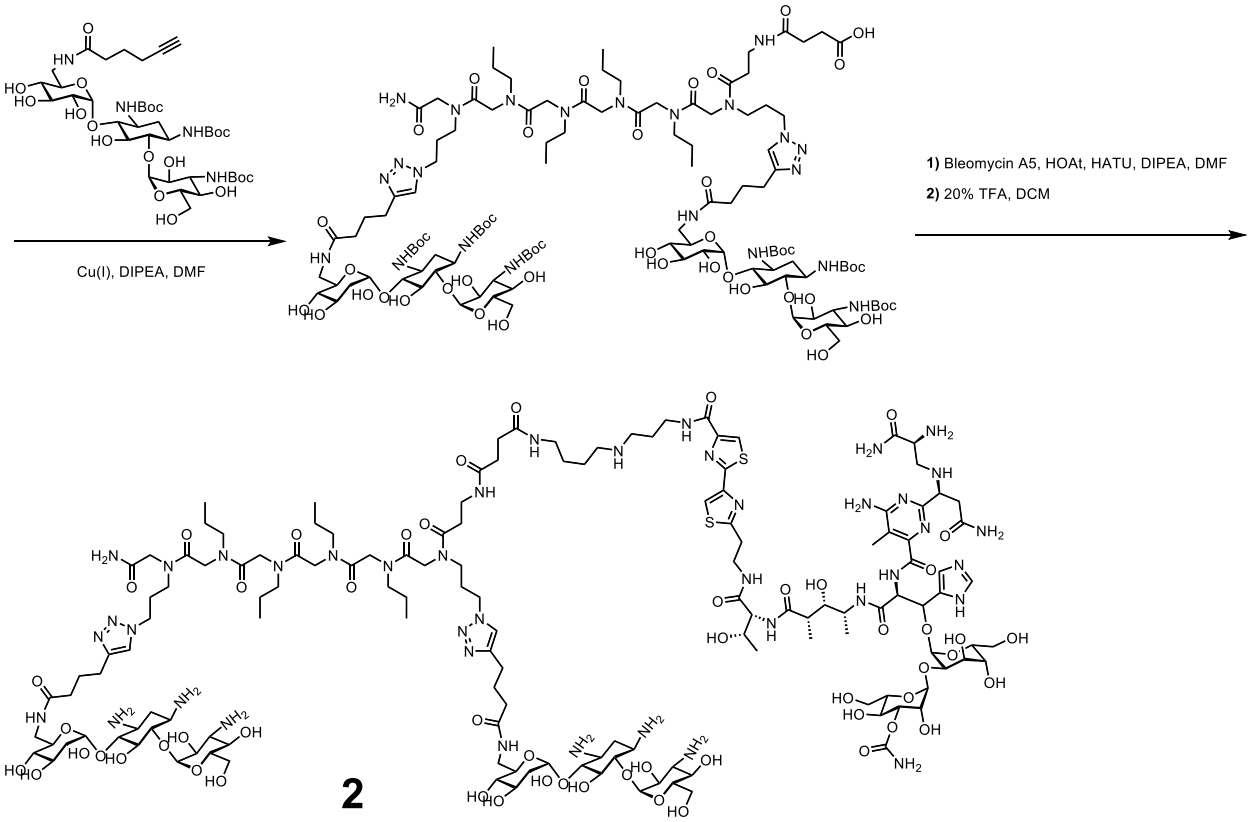
The coupling step and chain extension step were repeated three additional times to obtain a peptoid spacer with 4 propylamine units. Following the final propyl amine addition, bromoacetic acid coupling and introduction of click counterpart steps were repeated.

Final spacer: To attach the final spacer, a solution of Fmoc-beta-alanine (1.1gr, 3.65mmol, 5eq), 1-Hydroxy-7-azabenzotriazole (HOAt) (824mg, 7.3mmol, 10eq), DIC (0.57mL, 3.65mmol, 5eq) and DIEA (0.6 mL, 3.65mmol, 5eq) in 5 mL of DMF was added to the resin. The solution was heated to 70°C for 20 min using the microwave, drained and washed with DCM (3x5mL) and DMF (3x5mL). Then Fmoc was deprotected with 5 mL of 20% piperidine in DMF for 10 min, followed by draining and then washing with DCM (3x5 mL) and DMF (3x5 mL). Next, a solution of succinic anhydride (365mg, 3.65mmol, 5eq) and DIEA (0.6 mL, 3.65mmol, 5eq) in 5 mL of DMF was added to the resin and shaken at room temperature for 10 min and then washed with DCM (3x5mL) and DMF (3x5mL) to afford the peptoid with a free carboxylic acid.

Resin cleavage: The resin-bound peptoid was cleaved by adding 5mL of 20% TFA in DCM and shaking for 10 min at room temperature. Then solvent was removed by evaporation to afford a yellow oil. The oil was purified via HPLC performed with a linear gradient from 20% to 100% B (methanol + 0.1%TFA) in A (water + 0.1%TFA) over 60 min and a flow rate of 5 mL/min. tR = 43 min.

Synthesis of compound 1. Compound **1** was synthesized as previously described.[1]

Synthesis of compound 2:



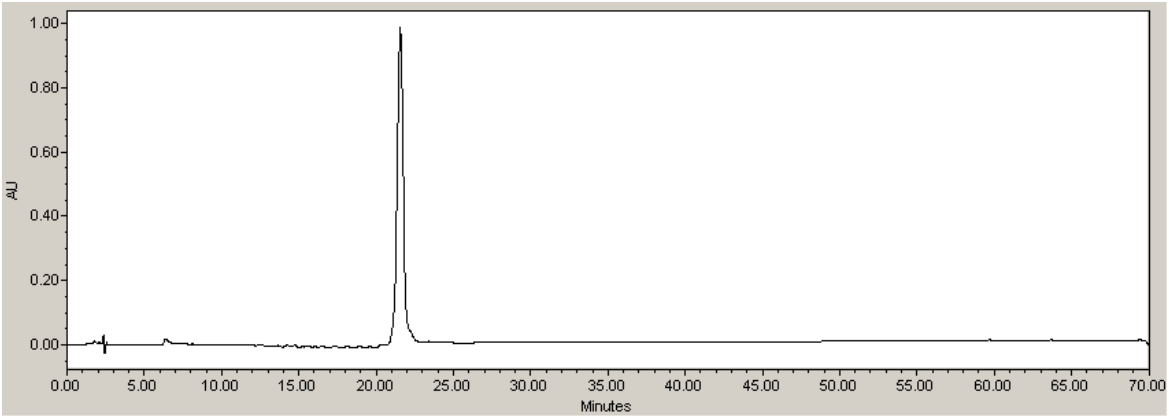
Scheme 2: Synthetic route for compound **2**.

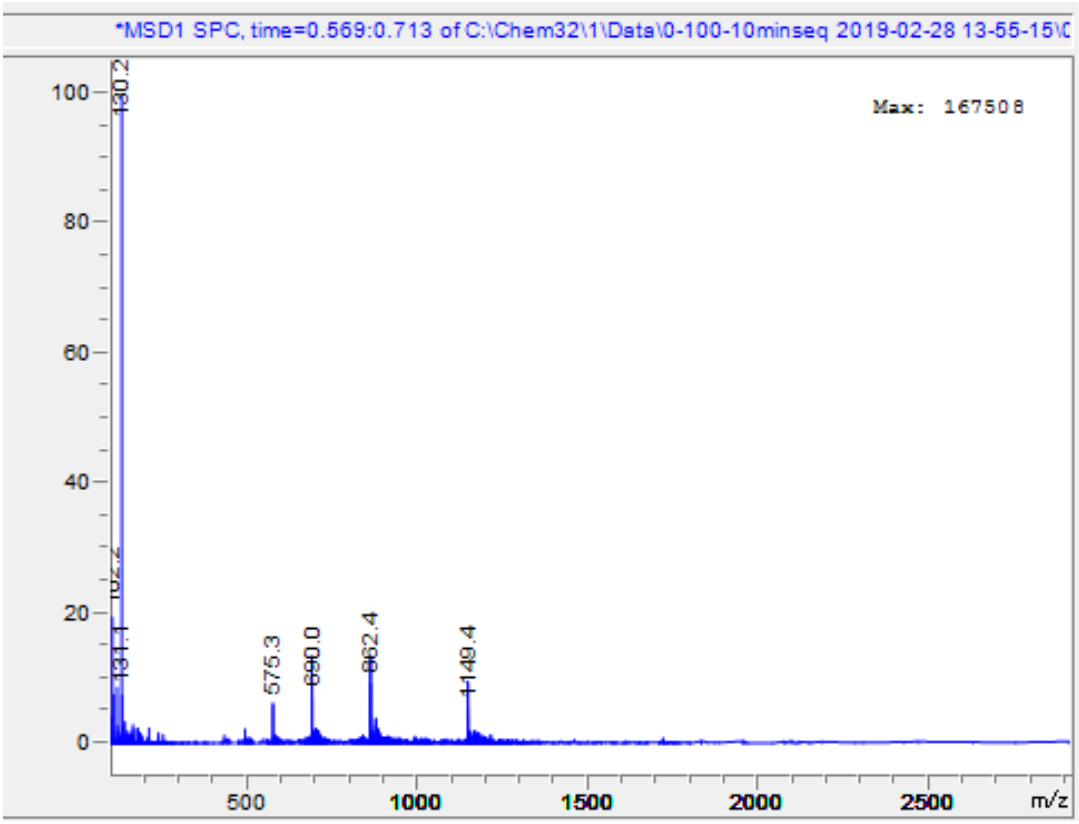
Click Reaction. Peptoid (1 mg, 1.35 µmol, 1 eq). was treated with a solution of Boc-Kanamycin-hexynoate (1.6 mg, 2.7 µmol, 2 eq) (synthesized as previously reported[4]), Cu(I) catalyst (0.15mg, 0.27µmol, 0.2 eq), and DIEA (0.5 µL, 2.7 µmol, 2 eq) in 2 mL DMF. The reaction mixture was heated to 80°C overnight, and the reaction progress was monitored by MALDI MS. After the peptoid starting material was no longer detectable, the product was purified via HPLC with a linear gradient from 20% to 100% B (methanol + 0.1%TFA) in A (water + 0.1%TFA) over 60 min and a flow rate of 5 mL/min, compound fractions were collected at 55-56min. The pure fractions were collected, and the solvent was concentrated under vacuum.

Bleomycin coupling. The Bleomycin A5 coupling reaction was performed by adding HOAt (0.5 mg, 3.6 µmol, 4 eq), HATU (1.3 mg, 3.6 µmol, 4 eq), DIEA (0.7 µL, 3.6 µmol, 4 eq) and Copper coordinated-Bleomycin A5 (2.9 mg, 1.8 µmol, 2 eq) (Bleocin, EMD Millipore) in DMF. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then subjected to HPLC purification by first using 0.1 mM EDTA in water (pH 6.3) for 15 min followed by 100% water for 15 min, and then the target product

was separated with a linear gradient from 0% to 100% B (methanol + 0.1%TFA) in A (water + 0.1%TFA) over 60 min and a flow rate of 5 mL/min, compound fraction was collected at 49min. The pure fraction was collected, and the solvent was concentrated under vacuum.

Boc deprotection. The Boc-protected compound **2** was dissolved in 20% TFA in DCM and stirred for 20 min. Then solvent was removed under vacuum, to afford transparent oil. The oil was dissolved in water and subjected to HPLC purification using a linear gradient from 0% to 100% B (methanol + 0.1%TFA) in A (water + 0.1%TFA) over 60 min and a flow rate of 5 mL/min, compound fraction was collected at 29min. The pure fraction was concentrated under vacuum and **2** was obtained as a transparent oil. Purity was evaluated on a reverse phase Waters Symmetry C18 5 µm 4.6 × 150 mm column at room temperature with a flow rate of 1 mL/min and a linear gradient of 0–100% B (methanol + 0.1%TFA) in A (water + 0.1%TFA) over 60min. Absorbance was monitored at 220 nm and 254 nm. The final compound was analyzed by LC-MS using an Agilent 1260 Series LC-MS. A gradient of 0-100% methanol in water plus 0.1% formic acid over 10 min was used for analysis. LC-MS: [M/3+H]+ calculated: 1149.4; [M/3+H]+ observed: 1149.4. Mass spectra were recorded on an Applied Biosystems MALDI ToF/ToF Analyzer 4800 Plus using an α-cyano-4-hydroxycinnamic acid matrix. MALDI: [M+Na]+ calculated: 3465.6242; [M+Na]+ observed: 3465.7007.





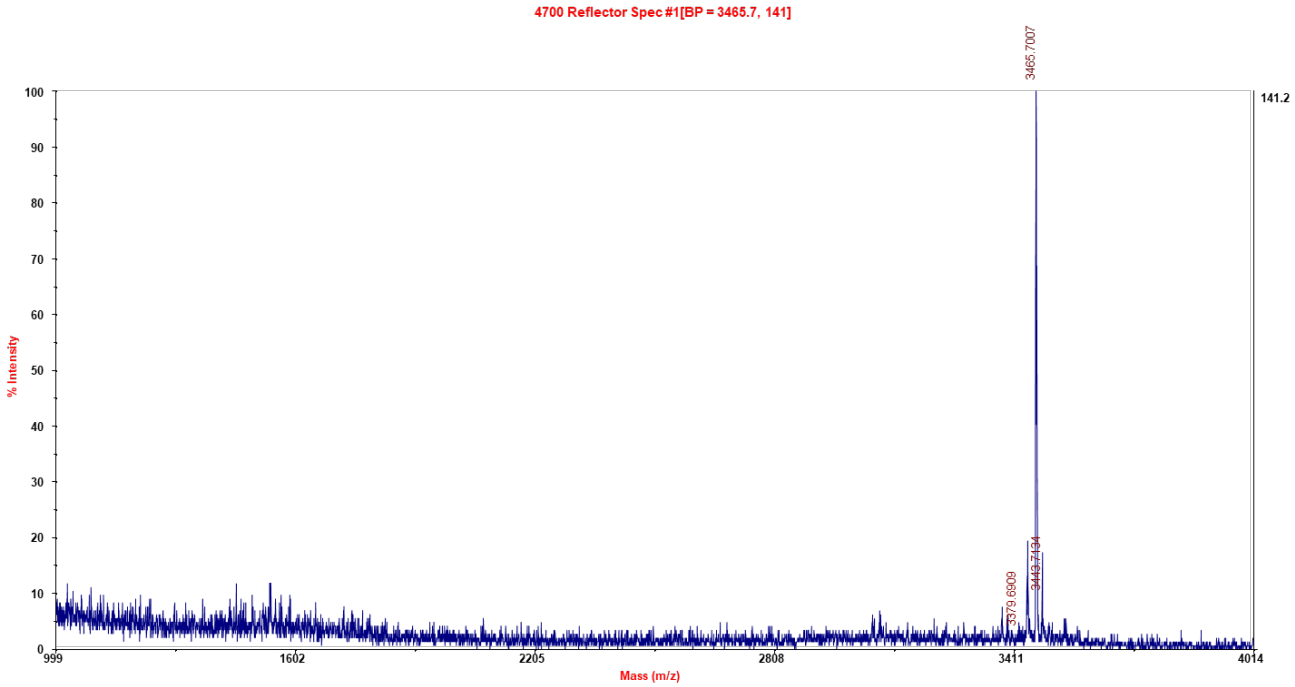
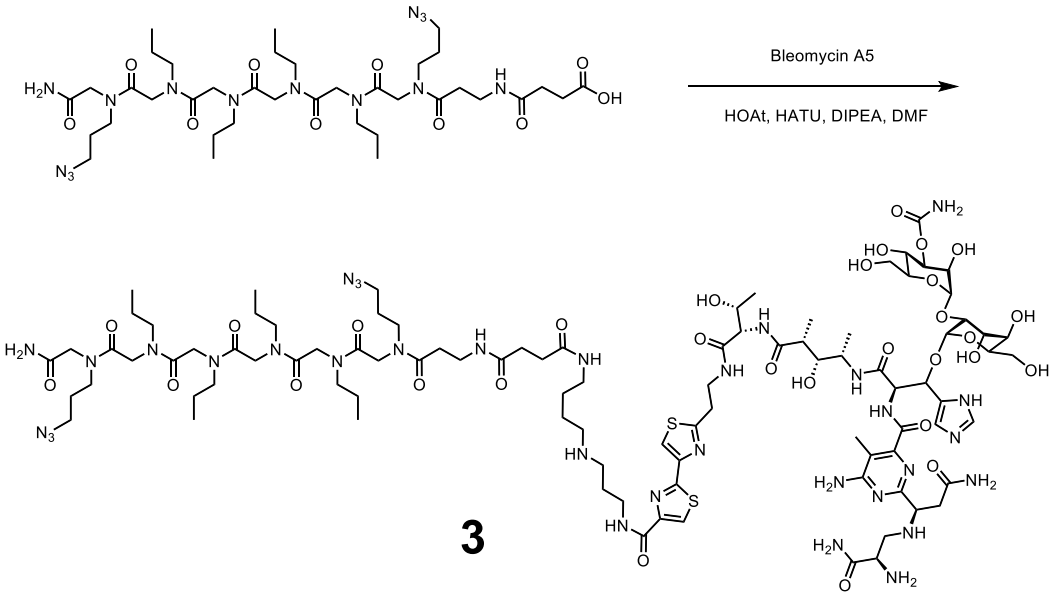


Figure S16. Characterization of compound **2** by HPLC, LC-MS spectrum and MALDI mass spectrum.

Synthesis of compound 3:



Scheme 3: Synthetic route for compound **3**.

Bleomycin coupling. The Bleomycin A5 coupling reaction was performed by adding HOAt (1 mg, 7.2 µmol, 4 eq), HATU (2.6 mg, 7.2 µmol, 4 eq), DIEA (1.4 µL, 7.2 µmol, 4 eq) and Copper coordinated-Bleomycin A5 (4.8 mg, 3.6 µmol, 2 eq) (Bleocin, EMD Millipore) in DMF to the peptoid. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then subjected to HPLC purification by first using 0.1 mM EDTA in water (pH 6.3) for 15 min followed by 100% water for 15 min and then the target product was separated with a linear gradient from 0% to 100% B (methanol +0.1%TFA) in A (water +0.1%TFA) over 60 min and a flow rate of 5 mL/min, compound fraction was collected at 41min. The pure fraction was collected, and the solvent was concentrated under vacuum to afford **3** as a transparent oil. Purity was evaluated on a reverse phase Waters Symmetry C18 5 µm 4.6 × 150 mm column at room temperature with a flow rate of 1 mL/min and a linear gradient of 0–100% B (methanol +0.1%TFA) in A (water +0.1%TFA). Absorbance was monitored at 220 nm and 254 nm. The final compound was analyzed by LC-MS. A gradient of 0-100% acetonitrile in water plus 0.1% formic acid over 10 min was used for analysis. LC-MS: [M/2+Na]+ calculated: 1166.5; [M/2+Na]+ observed: 1166.5.

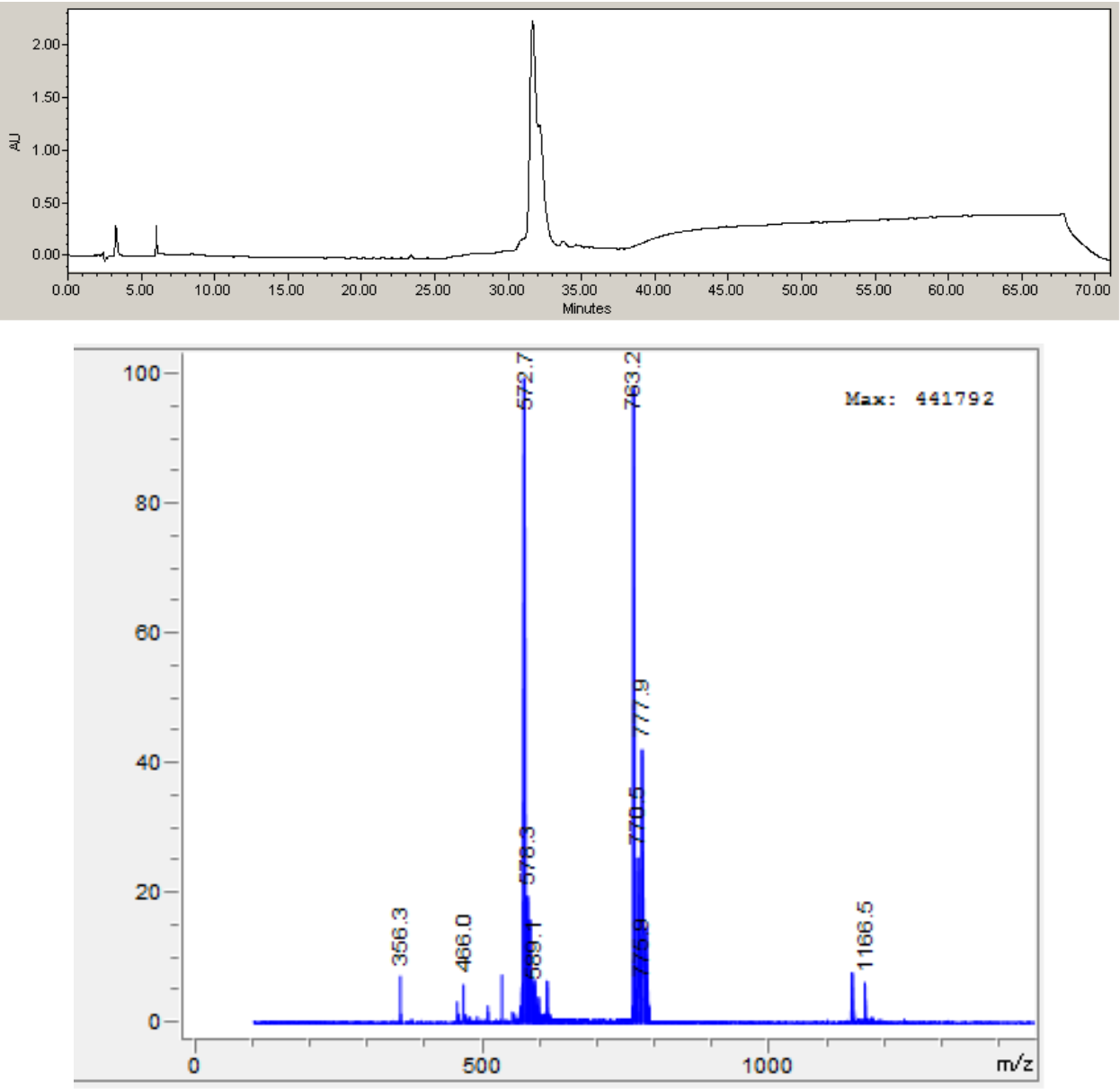


Figure S17. Characterization of compound **3** by HPLC, and LC-MS spectrum.

III. References

1. Lee, M. M., Pushechnikov, A., and Disney, M. D. (2009) Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2, *ACS Chem. Biol. 4*, 345-355.