Supplementary Information For:

Induction and Reversal of Myotonic Dystrophy Type 1 Pre-mRNA Splicing Defects by Small Molecules

Jessica L. Childs-Disney[1], Ewa Stepniak-Konieczna[2], Tuan Tran[1,3], Ilyas Yildirim[4], HaJeung Park[1], Catherine Z. Chen[5], Jason Hoskins[6], Noel Southall[5], Juan J. Marugan[5], Samarjit Patnaik[5], Wei Zheng[5], Chris P. Austin[5], George C. Schatz[4], Krzysztof Sobczak[2,6], Charles A. Thornton[6], and Matthew D Disney[1]*

[1]Department of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way #3A1, Jupiter, FL 33458

[2]Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, 61 251 Poznan, Poland

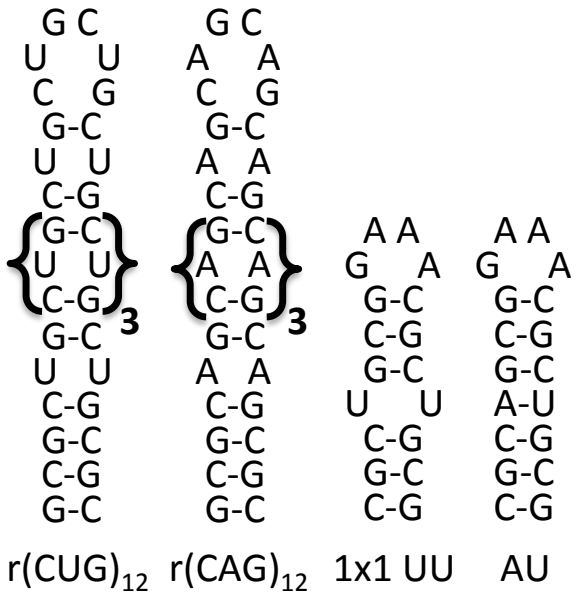
[3]Department of Chemistry, The University at Buffalo, SUNY, Buffalo, NY 14260

[4]Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3113

[5]NIH Chemical Genomics Center, Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD 20892

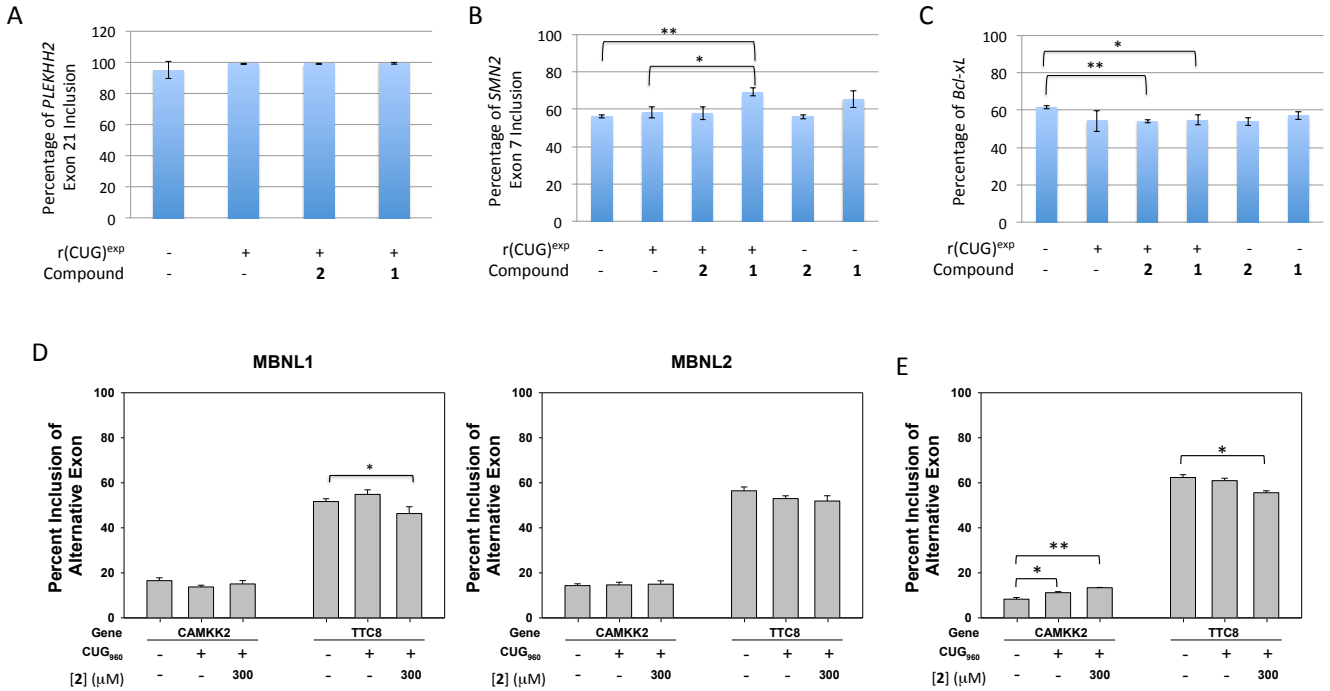
[6]Department of Neurology, School of Medicine and Dentistry, University of Rochester, Rochester, NY 14642

Supplementary Fig. S1



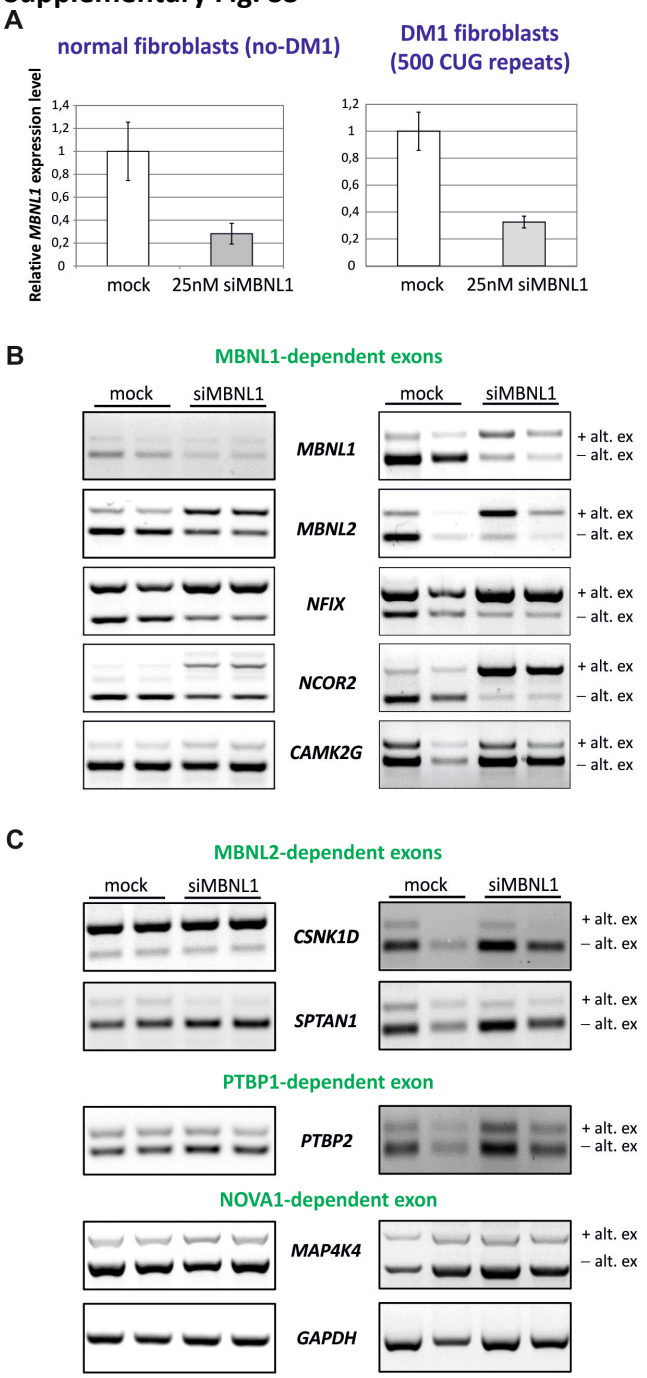
Supplementary Fig. S1: Secondary structures of the RNAs used in affinity measurements. $r(CUG)_{12}$, a mimic of $r(CUG)^{exp}$, contains five 5'CUG/3'GUC repeats. Likewise, $r(CAG)_{12}$, a mimic of $r(CAG)^{exp}$, contains five 5'CAG/3'GAC repeats. The RNA designated as 1x1 UU contains only one copy of the 5'CUG/3'GUC motif while the RNA designated as AU contains a fully paired stem.

Supplementary Fig. S2



Supplementary Fig. S2: Compounds **1** and **2** do not affect pre-mRNA alternative splicing of genes not regulated by MBNL1. Top, compounds **1** and **2** (each tested at 500 μM) do not affect the pre-mRNA splicing of mini-genes whose splicing is not regulated by MBNL1, including *PLEKHH2* (panel A)*, SMN2* (panel B)*, and Bcl-x* (panel C)*.* The percentage of each isoform was determined by RT-PCR (n ≥ 2; error bars are the standard deviations in the measurements). All p-values, as determined by a student t-test, are >0.05 unless indicated. A radioactively labeled forward PCR primer was used for the *PLEKHH2* mini-gene. The *SMN2* and *Bcl-x* isoforms were separated by agarose gel electrophoresis and stained with SYBR gold. D, compound **2** does not significantly affect the alternative splicing of endogenous pre-mRNAs *CAMKK2* and *TTC8* in a HEK 293 model systems of DM1, whether they express MBNL1 or MBNL2. CAMKK2 and TTC8 are not regulated by MBNL1 (n = 2; error bars are the standard deviations in the measurements). All p-values, as determined by a student t-test are >0.05 unless idnicated. E, compound **2** does not significantly affect the alternative splicing of endogenous pre-mRNAs *CAMKK2* and *TTC8* in a HeLa cell model system of DM1 (n =2; error bars are the standard deviations in the measurements). All p-values, as determined by a student t-test are >0.05 unless indicated. "*" indicates p<0.05 while "**" indicates p<0.01 as determined by a student t-test.

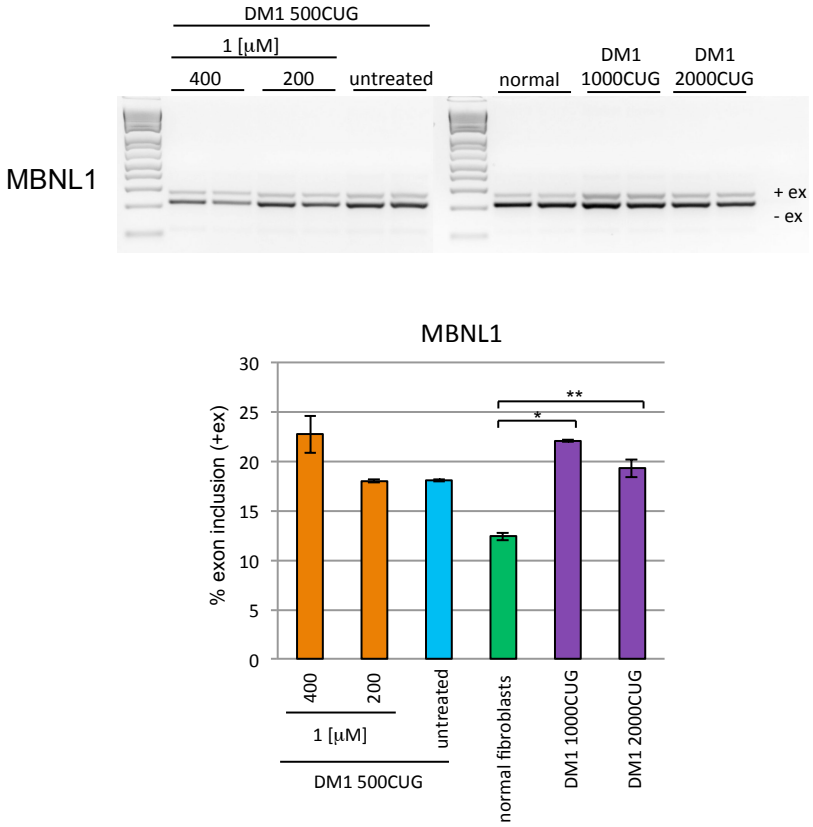
Supplementary Fig. S3



Supplementary Fig. S3: Splicing alterations upon siRNA mediated knock-down of MBNL1 in normal (left panel) and DM1-affected (right panel) fibroblasts. (A) MBNL1 expression level plotted relative to *GAPDH* expression in mock- and siMBNL1-treated normal (left) and DM1 (right) fibroblasts. (B) MBNL1 knock-down induces DM1-like splicing shift in MBNL1-dependent exons (*MBNL1*, *MBNL2*, *NFIX*, *NCOR2*, *CAMK2G*) in normal fibroblast (left panel), and exacerbates splicing alterations in DM1-fibroblasts (right panel). (C) MBNL1

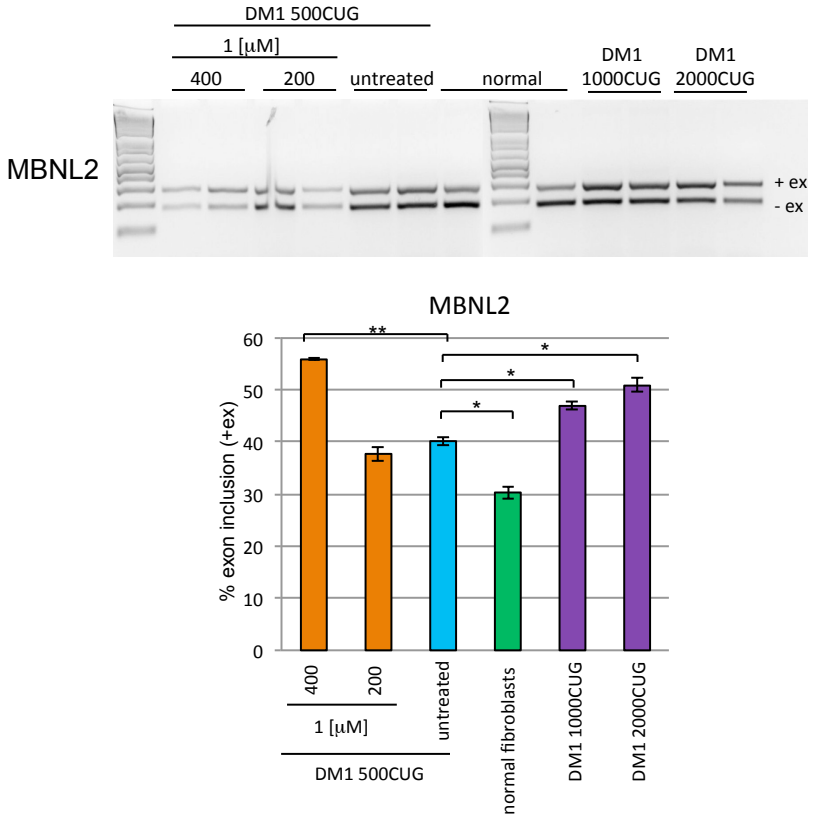
knock-down does not alter the alternative splicing pattern of MBNL2-dependent (*CSNK1D*, *SPTAN1*), PTBP1-dependent (*PTBP2*) or NOVA1-dependent (*MAP4K4*) exons in normal and DM1-fibroblasts. *GAPDH* was used as a loading control. The DM1-like splicing shift for MBNL1-dependent exons is depicted as alternative exon inclusion (+alt. ex), while the normal splicing isoform is depicted as alternative exon exclusion (-alt. ex).

Supplementary Fig. S4



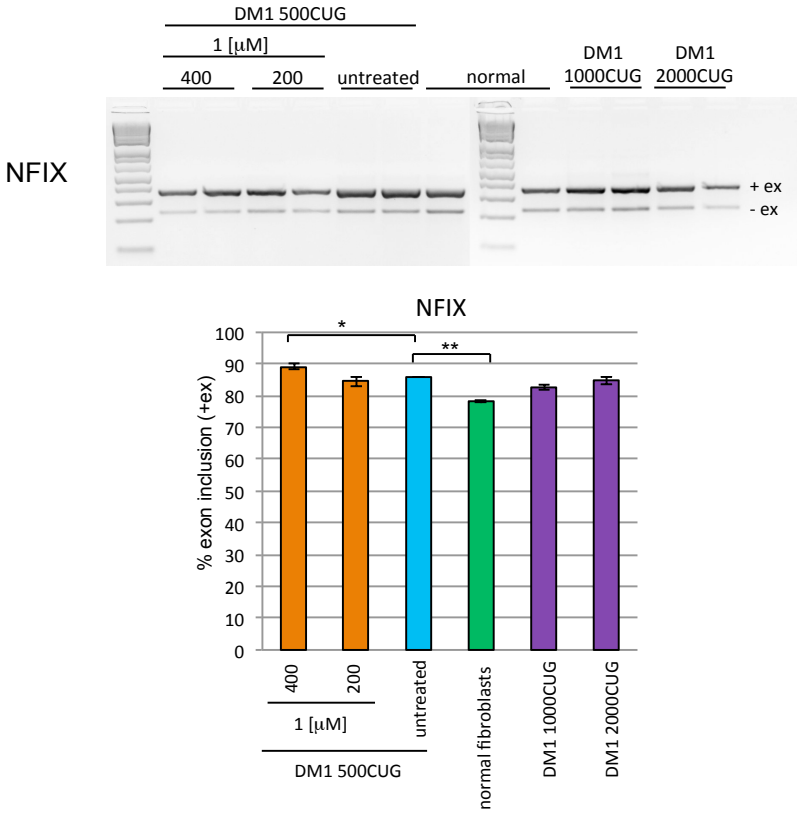
Supplementary Fig. S4: Compound **1** exacerbates the alternative splicing of the MBNL1-dependent exon of *MBNL1* in DM1 fibroblasts. The extent of dysregulation is similar to cells expressing 1000 or 2000 r(CUG) repeats. Top, representative gel images of the alternatively spliced isoforms of *MBNL1* as determined by RT-PCR (n = 2 for all treated fibroblasts; n = 5 for untreated and DMSO-treated DM1 500CUG and normal fibroblasts; n = 4 for untreated DM1 1000CUG and DM1 2000CUG fibroblasts; error bars are the standard deviations in the measurements). Bottom, plot of the data shown in the top panel. The blue bar represents untreated DM1 fibroblasts with 500r(CUG) (DM1 500CUG), the green bar represents normal human fibroblasts, and the violet bar represent DM1 affected control fibroblasts expressing 1000 and 2000 r(CUG) repeats (DM1 1000CUG, DM1 2000CUG). "*" indicates $p<0.05$ while "**" indicates $p<0.01$ as determined by a two-tailed t-test.

Supplementary Fig. S5



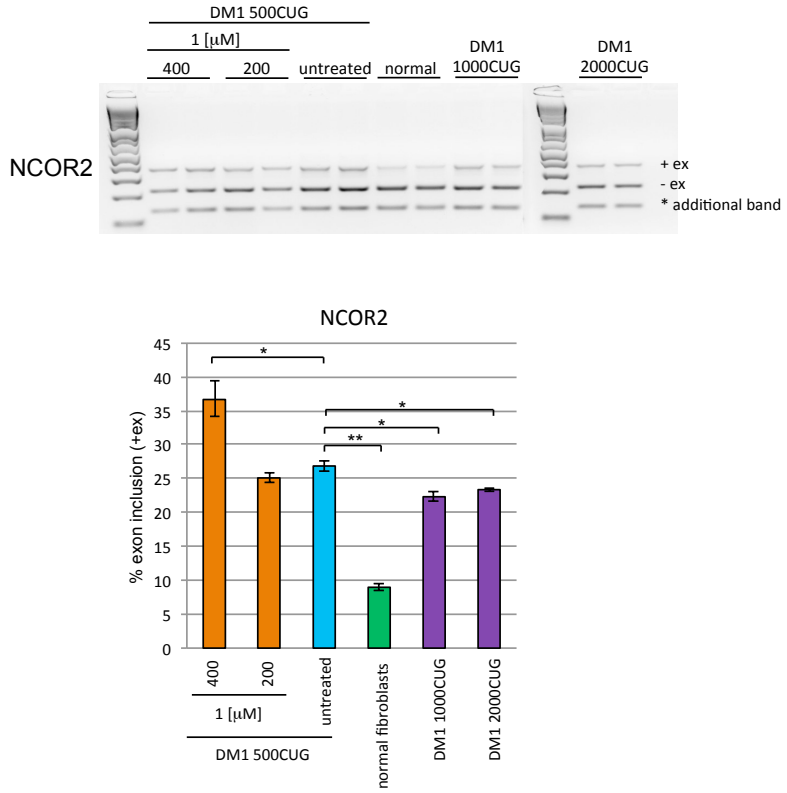
Supplementary Fig. S5: Compound **1** exacerbates the alternative splicing of the MBNL1-dependent exon of *MBNL2* in DM1 fibroblasts. The extent of dysregulation is similar to cells expressing 1000 or 2000 r(CUG) repeats. Top, representative gel images of the alternatively spliced isoforms of *MBNL2* as determined by RT-PCR (n = 2 for all treated fibroblasts; n = 5 for untreated and DMSO-treated DM1 500CUG and normal fibroblasts; n = 4 for untreated DM1 1000CUG and DM1 2000CUG fibroblasts; error bars are the standard deviations in the measurements). Bottom, plot of the data shown in the top panel. The blue bar represents untreated DM1 fibroblasts with 500r(CUG) (DM1 500CUG), the green bar represents normal human fibroblasts, and the violet bar represent DM1 affected control fibroblasts expressing 1000 and 2000 r(CUG) repeats (DM1 1000CUG, DM1 2000CUG). "*" indicates $p<0.05$ while "**" indicates $p<0.01$ as determined by a two-tailed t-test.

Supplementary Fig. S6



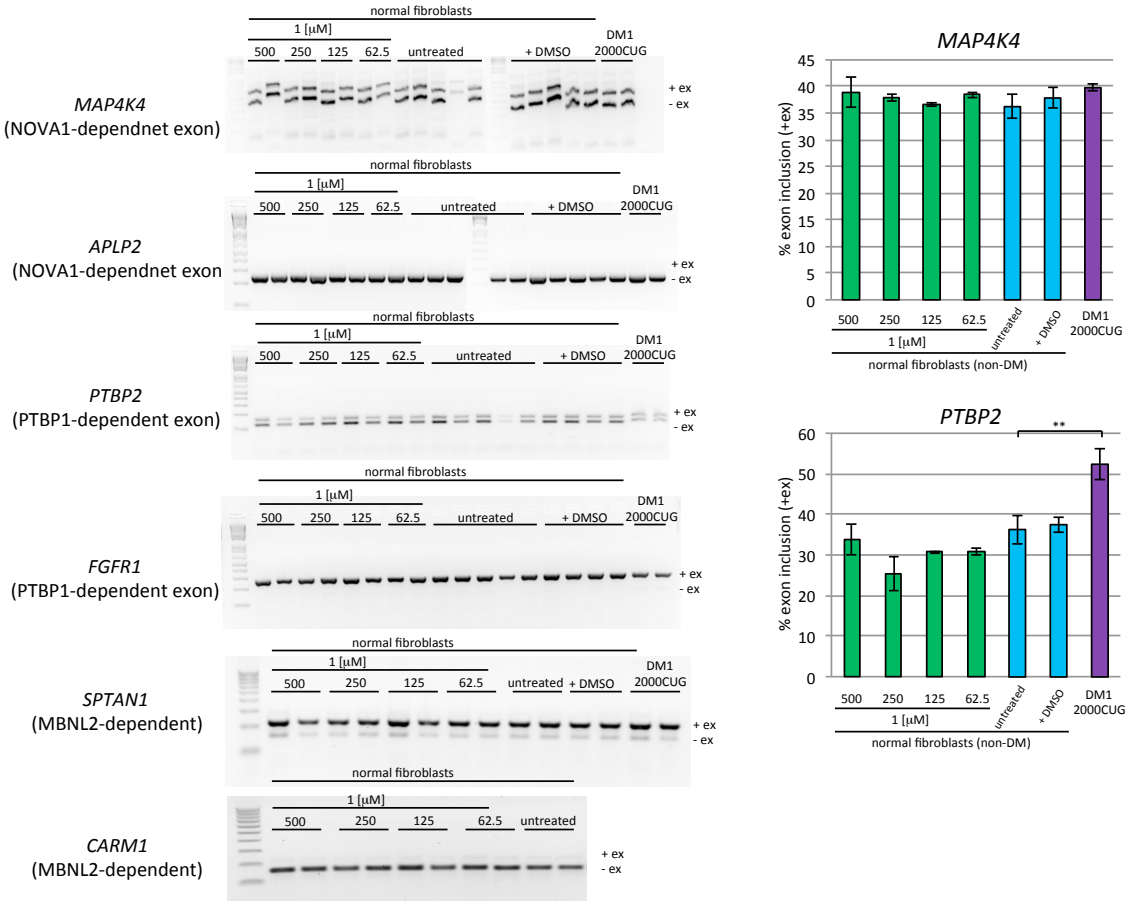
Supplementary Fig. S6: Compound **1** exacerbates the alternative splicing of the MBNL1-dependent exon of *Nfix* in DM1 fibroblasts. The extent of dysregulation is similar to cells expressing 1000 or 2000 r(CUG) repeats. Top, representative gel images of the alternatively spliced isoforms of *Nfix* as determined by RT-PCR (n = 2 for all treated fibroblasts; n = 5 for untreated and DMSO-treated DM1 500CUG and normal fibroblasts; n = 4 for untreated DM1 1000CUG and DM1 2000CUG fibroblasts; error bars are the standard deviations in the measurements). Bottom, plot of the data shown in the top panel. The blue bar represents untreated DM1 fibroblasts with 500r(CUG) (DM1 500CUG), the green bar represents normal human fibroblasts, and the violet bar represent DM1 affected control fibroblasts expressing 1000 and 2000 r(CUG) repeats (DM1 1000CUG, DM1 2000CUG). "*" indicates $p<0.05$ while "**" indicates $p<0.01$ as determined by a two-tailed t-test.

Supplementary Fig. S7



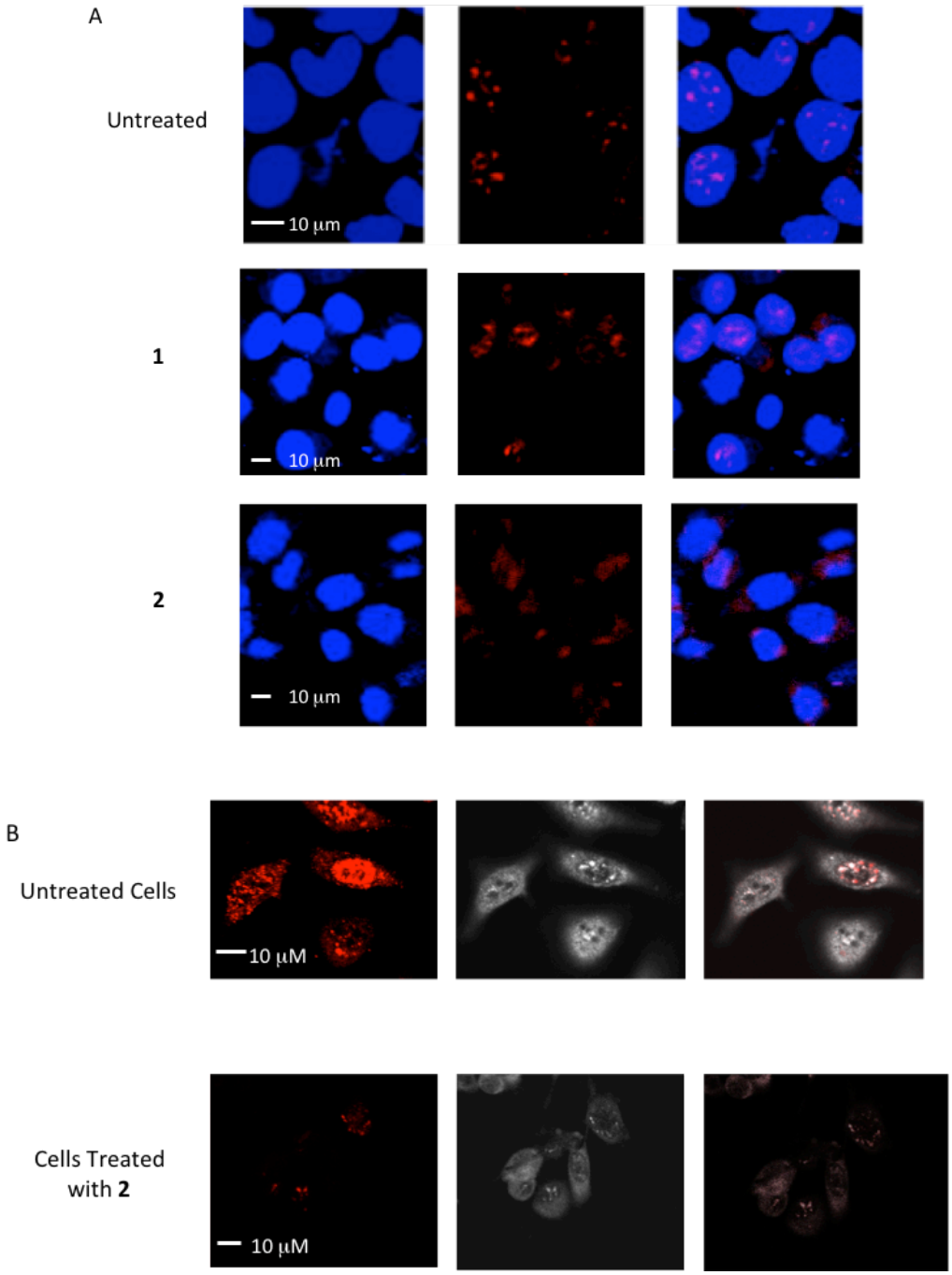
Supplementary Fig. S7: Compound **1** exacerbates the alternative splicing of the MBNL1-dependent exon of *Ncor* in DM1 fibroblasts. The extent of dysregulation is similar to cells expressing 1000 or 2000 r(CUG) repeats. Top, representative gel images of the alternatively spliced isoforms of *Ncor* as determined by RT-PCR (n = 2 for all treated fibroblasts; n = 5 for untreated and DMSO-treated DM1 500CUG and normal fibroblasts; n = 4 for untreated DM1 1000CUG and DM1 2000CUG fibroblasts; error bars are the standard deviations in the measurements). Bottom, plot of the data shown in the top panel. The blue bar represents untreated DM1 fibroblasts with 500r(CUG) (DM1 500CUG), the green bar represents normal human fibroblasts, and the violet bar represent DM1 affected control fibroblasts expressing 1000 and 2000 r(CUG) repeats (DM1 1000CUG, DM1 2000CUG). "*" indicates $p<0.05$ while "**" indicates $p<0.01$ as determined by a two-tailed t-test.

Supplementary Fig. S8



Supplementary Fig. S8: Compound **1** does not affect the alternative splicing of MBNL1-independent exons in normal human fibroblasts. Left, representative gel images of the alternatively spliced isoforms of *Map4K4*, *PTBP2*, *Aplp2*, *Fgfr1, SPTAN1*, and *CARM1* as determined by RT-PCR (n = 2 for treated fibroblasts; n= 5 for untreated and DMSO-treated fibroblasts; error bars are the standard deviations in the measurements). Right, plots of the data shown in the top panel. "**" indicates p<0.01 as determined by a two-tailed t-test.

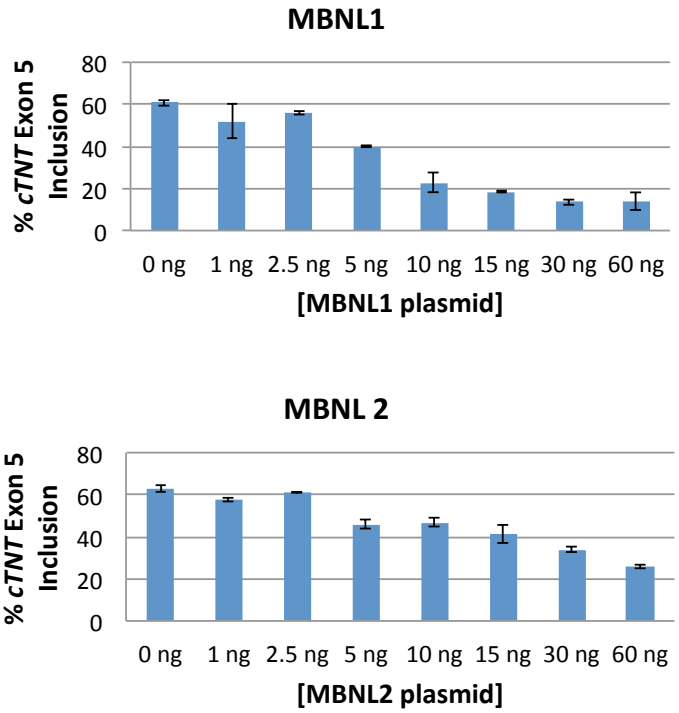
Supplementary Fig. S9



Supplementary Fig. S9: Compounds **1** and **2** reduce the size and number of nuclear foci as determined by fluorescence *in situ* hybridization. A, Addition of 100 µM of **1** and **2** disrupt the formation of nuclear foci to varying extents. All panels: left, DAPI fluorescence, which indicates nuclei; middle, Cy3 fluorescence, which indicates

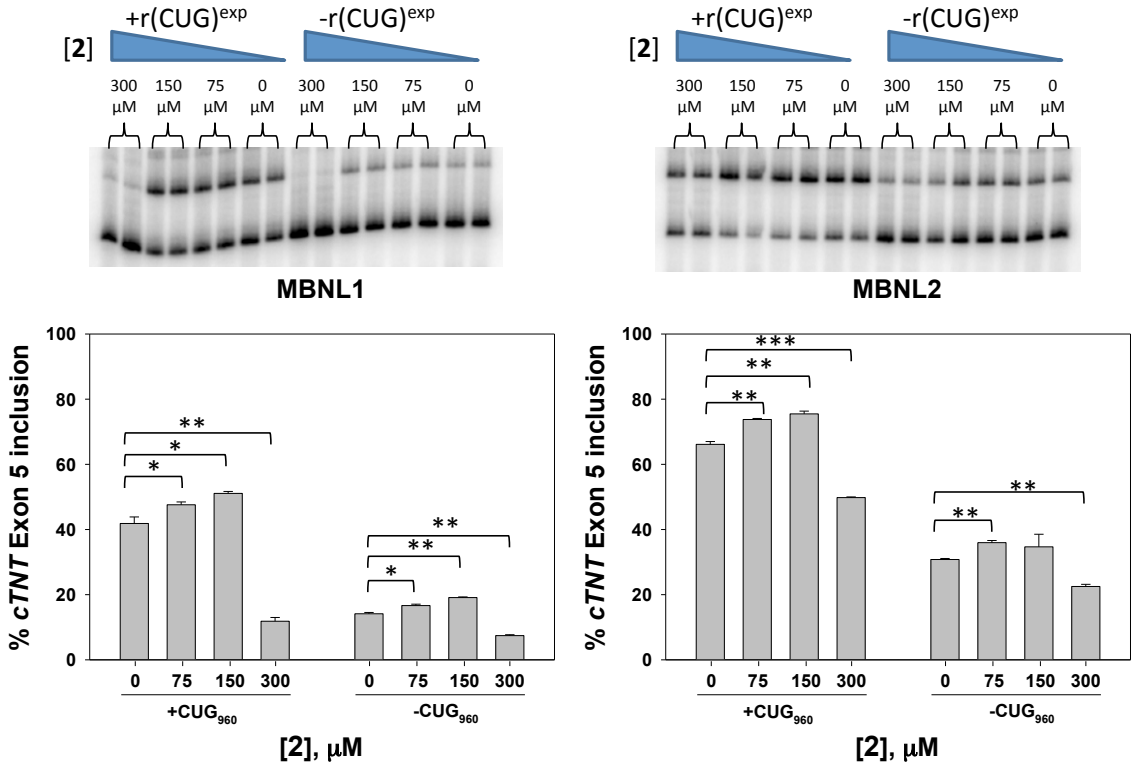
r(CUG)^{exp}; right, overlay of DAPI and Cy3 images. Top, cells were transfected with the DM1 mini-gene and then subjected to FISH. Middle, cells were transfected with the DM1 mini-gene, treated with compound **1**, and subjected to FISH. Bottom, cells were transfected with the DM1 mini-gene, treated with compound **2**, and subjected to FISH. B, Immunohistochemistry of MBNL1 after FISH reveals that the foci present in cells treated with **2** have less MBNL1. All panels: left, Cy3 fluorescence, which indicates r(CUG)^{exp}; middle, Cy5 fluorescence, which indicates MBNL1; right, overlay of Cy3 and Cy5 images.

Supplementary Fig. S10



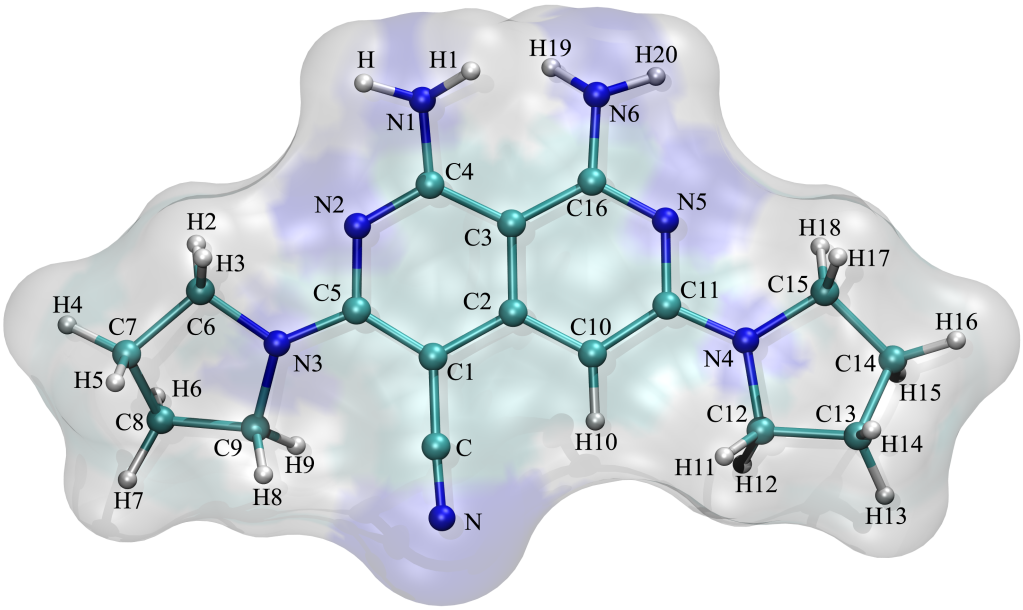
Supplementary Fig. S10: Compound **2** improves MBNL1- and MBNL2-dependent alternative splicing in a DM1 model system. The alternative splicing of cTNT exon 5 is sensitive to the expression level of MBNL1 and MBNL2 in HEK 293T cells. Briefly, HEK cells were co-transfected with plasmids encoding r(CUG)^{exp}, the cTNT mini-gene, and different concentrations of plasmids encoding MBNL1 or MBNL2. Splicing patterns were assayed by RT-PCR as described in the Methods section (n = 2; error bars are the standard deviations in the measurements).

Supplementary Fig. S11



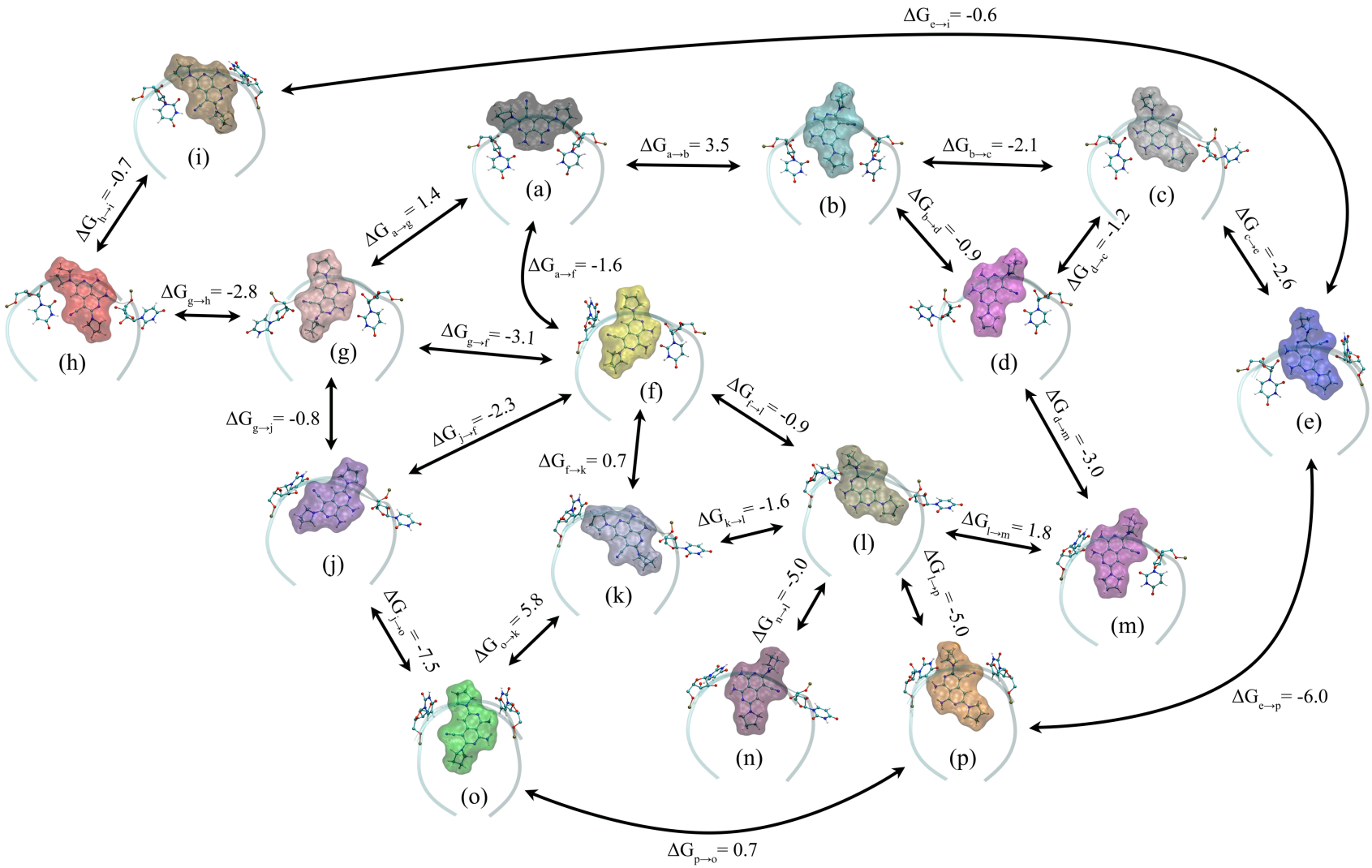
Supplementary Fig. S11: Compound **2** improves the MBNL1- and MBNL2-dependent alternative splicing of cTNT exon 5 in DM1 cellular models. Left, **2** improves MBNL1-dependent alternative splicing of cTNT exon 5. HEK 293T cells were co-transfected with (+CUG_{960}) or without (-CUG_{960}) plasmids encoding the DM1 mini-gene (100 ng), the cTNT mini-gene (100 ng), and MBNL1 (5 ng). Alternative splicing patterns were assayed by RT-PCR (n = 2; error bars are the standard deviations in the measurements). Right, **2** improves MBNL2-dependent alternative splicing of cTNT exon 5. HEK 293T cells were co-transfected with (+CUG_{960}) or without (-CUG_{960}) plasmids encoding the DM1 mini-gene (100 ng), the cTNT mini-gene (100 ng), and MBNL2 (5 ng). Alternative splicing patterns were assayed by RT-PCR (n = 2; error bars are the standard deviations in the measurements). "*" indicates $p<0.05$; "**" indicates $p<0.01$ and "***" indicates $p<0.001$ as determined by a two-tailed t-test.

Supplementary Fig. S12



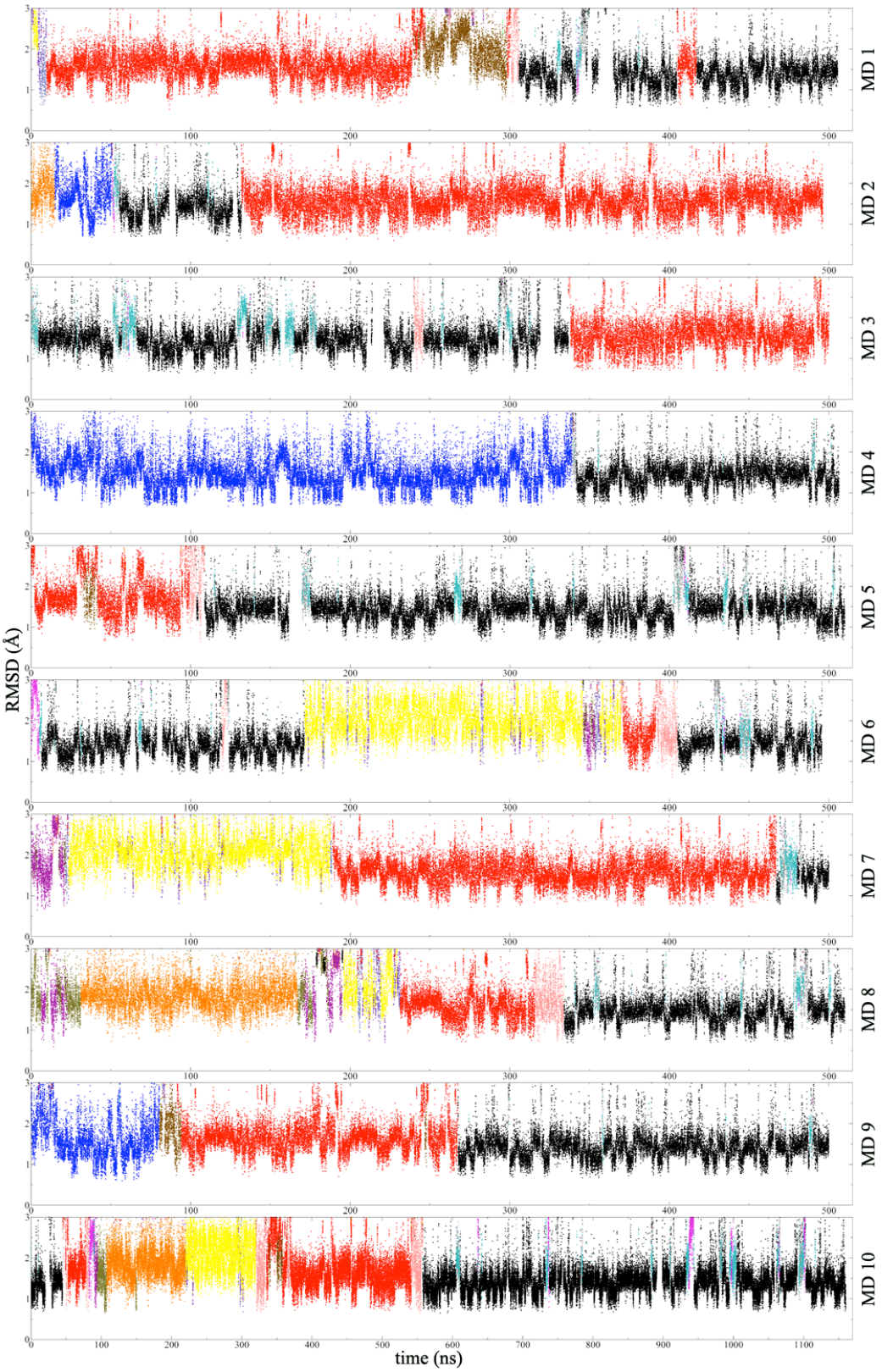
Supplementary Fig. S12: Atom names for **2** (see Table S-2 for atom types and charges).

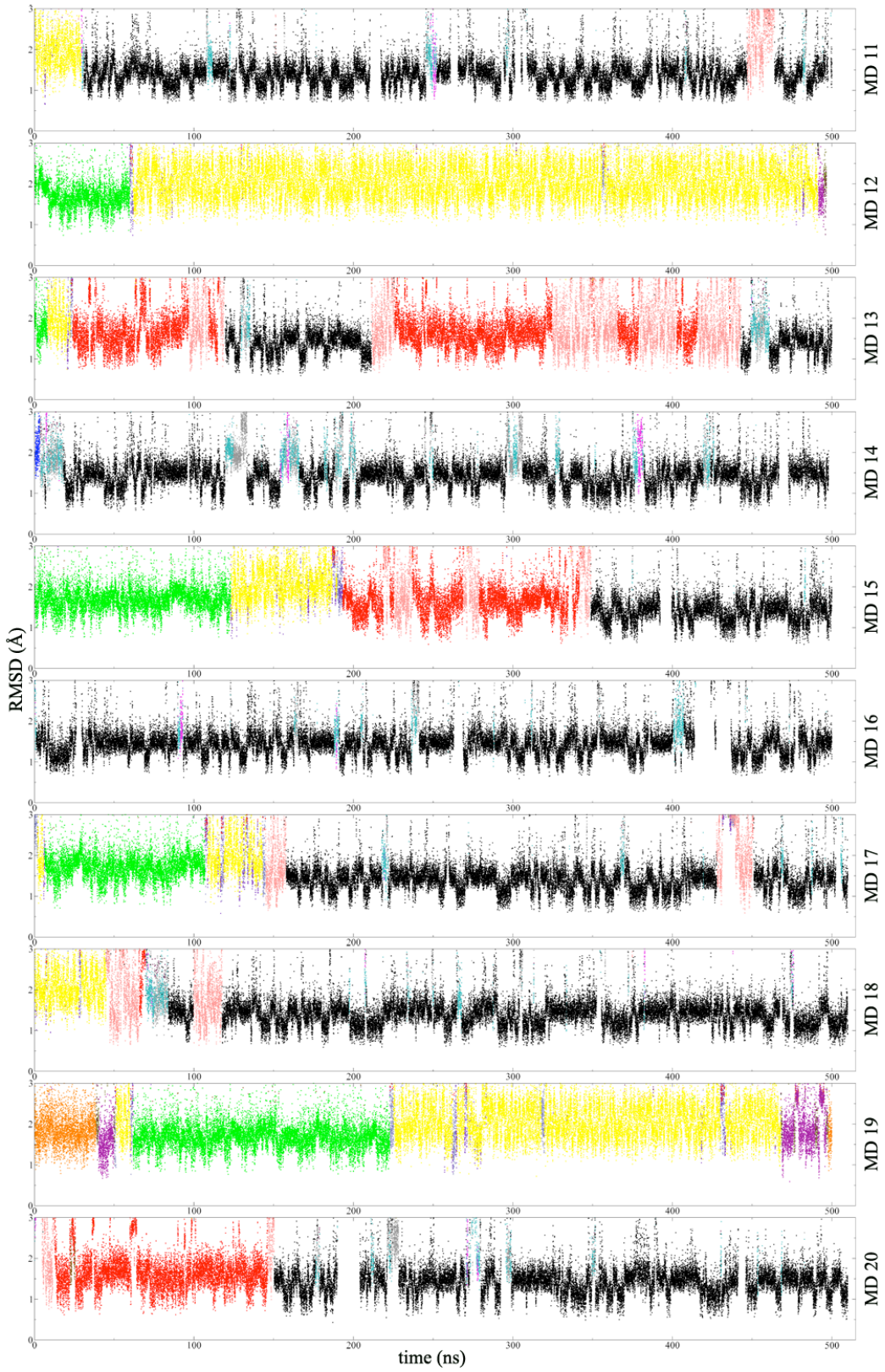
Supplementary Fig. S13

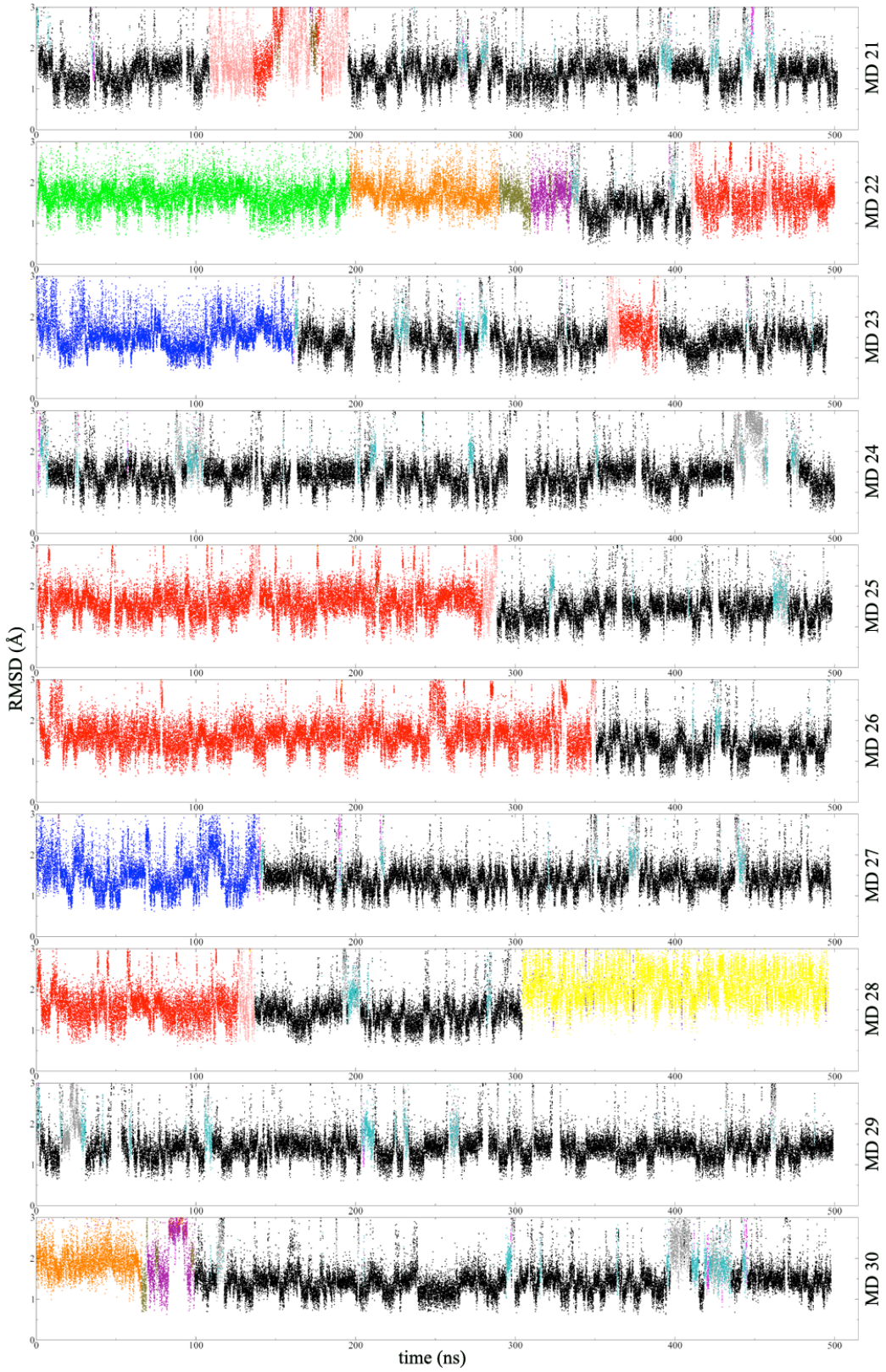


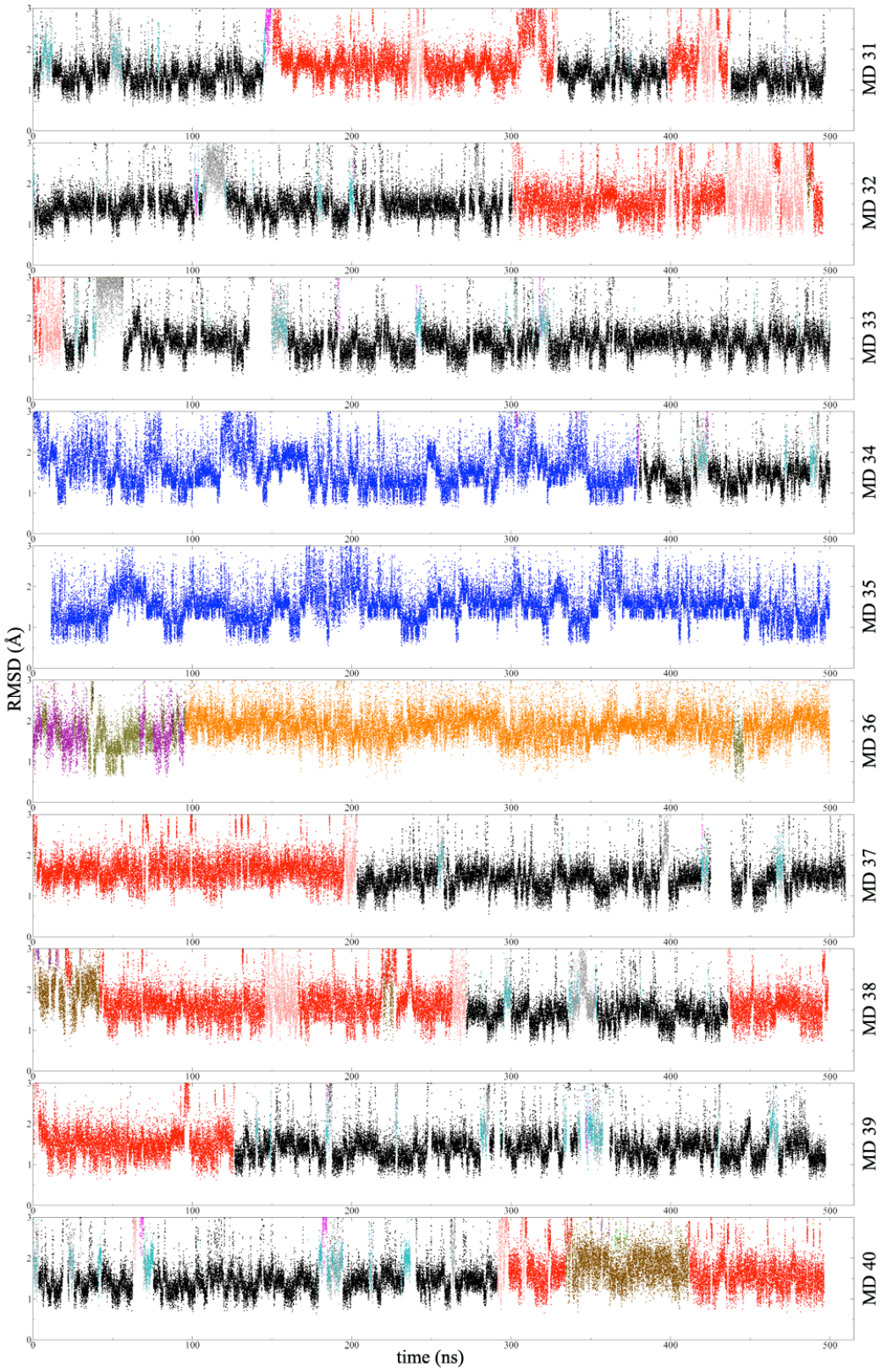
Supplementary Fig. S13: A schematic representation of potential conformational transitions of **2** in the r(CUG) repeat binding process. Table S5 shows the frequency of each binding mode observed in the trajectories and the relative binding free energies computed with MM-GBSA approach. Red and blue colored base pairs are Watson-Crick GC base pairs that flank the UU pair; they are drawn in order to illustrate stacking of **2**. The RNA backbone is highlighted with transparent tubes. Note that upper and lower parts of each binding mode represent the minor and major grooves, respectively. No binding through the major groove side was observed in the simulation.

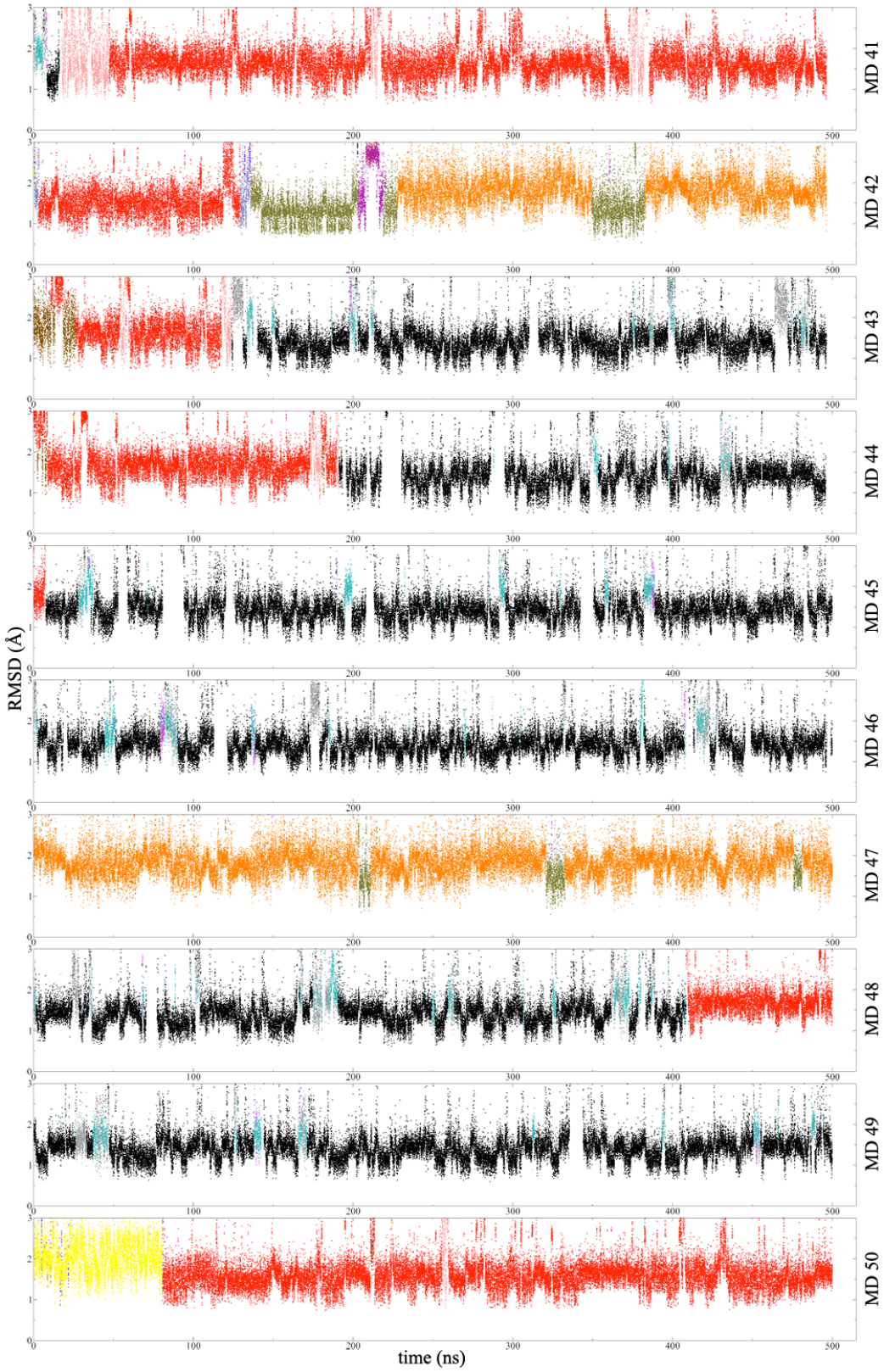
Supplementary Fig. S14




RMSD (Å)
time (ns)
MD 11
MD 12
MD 13
MD 14
MD 15
MD 16
MD 17
MD 18
MD 19
MD 20


RMSD (Å)
time (ns)
MD 21
MD 22
MD 23
MD 24
MD 25
MD 26
MD 27
MD 28
MD 29
MD 30


RMSD (Å)
time (ns)
MD 31
MD 32
MD 33
MD 34
MD 35
MD 36
MD 37
MD 38
MD 39
MD 40


RMSD (Å)
time (ns)
MD 41
MD 42
MD 43
MD 44
MD 45
MD 46
MD 47
MD 48
MD 49
MD 50

Supplementary Fig. S14: RMSD analysis of 50 MD simulations. Different colors represent different binding modes of **2** described in Table S5 and Supplementary Figure S13. MD simulation #10 is over 1.1 µs long while the rest are ~0.5 µs long.

Supplementary Table S1

Supplementary Table S1: Primer sequences used for RT-PCR splicing analyses.

Gene	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
MBNL1	GCTGCCCAATACCAGGTCAAC	TGGTGGGAGAAATGCTGTATGC
MBNL2	TCCTTTACCAAAGAGACAAGCAC	CTCAATGCAGATTCTTGGCATTCC
NCOR2	ACACCCACAACCGGAATGAGCCTG	GGACTTGGCTTTTCGGCTGCTG
NFIX	GAGCCCTGTTGATGACGTGTTCTA	CTGCACAAACTCCTTCAGTGAGTC
CAMK2G	GAGTGTTTGCGCAAGTTCAA	ACAGTGGTTTGTGGCTCCAT
MAP4K4	CCTCATCCAGTGAGGAGTCG	ATCACAGGAAAATCCCACCA
PTBP2	CGACCACCGTTATGTCAGGA	GGTTTCCATCAGCCATCTGT
APLP2	GGAGGAACCAAAGCCTCTCT	GCTCTCCCACTCCAGATCCT
FGFR1	ATGCTAGCAGGGGTCTCTGA	GCTTCCCGATCATCTTCATC
CSNK1D	CATGGAGAGAGAGCGGAAAG	GCACGACAGACTGAAGACCA
CARM1	GCCACAACAACCTGATTCCT	CGTAGTGCATGGTGTTGGTC
SPTAN	GTGAACGATCGTCAGGGTTT	TACGCTTCTCACCCAGTTCC
GAPDH	CATCAATGGAAATCCCATCAC	GGTTTTTCTAGACGGCAGGTC
cTNT	GTTCACAACCATCTAAAGCAAGATG	GTTGCATGGCTGGTGCAGG
INSR	GTACAAGCTTGAATGCTGCTCCTGTCCAAGACAG	GCCCTCGAGCGTGGGCACGCTGGTC
PLEKHH2	CGGGGTACCAAATGCTGCAGTTGACTCTCC	CCGCTCGAGCCATTCATGAAGTGCACAGG
SMN2	GGTGTCCACTCCCAGTTCAA	GCCTCACCACCGTGCTGG
Bcl-x	GGAGCTGGTGGTTGACTTTC	TAGAAGGCACAGTCGAGG
CAMKK2	CCTGGTGAAGACCATGATAC	GGCCCAGCAACTTTCCAC
TTC8	AGCTATTTTAGGCGCAGGAAGT	TTTTCATCCAGCATCATTTCTG

Supplementary Table S2

Table S2: GAMESS variables used in optimization and molecular electrostatic potential (MEP) calculation of **2**.

```
$CONTRL ICHARG=0 MULT=1
        MPLEVL=0 COORD=UNIQUE
        RUNTYP=OPTIMIZE SCFTYP=RHF
        EXETYP=RUN UNITS=ANGS
        MAXIT=200 $END
$BASIS  GBASIS=N31 NGAUSS=6
        DIFFSP=.F. NDFUNC=1 NPFUNC=1
        $END
$DFT    DFTTYP=NONE METHOD=GRID
        $END
$SCF    DIRSCF=.T. CONV=1.0E-08
        FDIFF=.F. $END
$SYSTEM TIMLIM=50000 MWORDS=64
        MEMDDI=0 $END
$STATPT NSTEP=200 OPTTOL=1.0E-06
        purify=.t. HESS=GUESS IHREP=0
        HSSEND=.T. $END
$GUESS  GUESS=HUCKEL $END
$DATA
compound_3_NIH1
C1
.
.
.
$END
```

```
$CONTRL ICHARG=0 MULT=1
        MPLEVL=0 COORD=UNIQUE
        RUNTYP=ENERGY MOLPLT=.T.
        SCFTYP=RHF EXETYP=RUN
        UNITS=ANGS MAXIT=200 $END
$BASIS  GBASIS=N31 NGAUSS=6
        DIFFSP=.F. NDFUNC=1 NPFUNC=1
        $END
$SCF    DIRSCF=.T. CONV=1.0E-08 $END
$SYSTEM TIMLIM=50000 MWORDS=64
        MEMDDI=0 $END
$GUESS  GUESS=HUCKEL $END
$ELPOT  IEPOT=1 WHERE=PDC
        OUTPUT=BOTH $END
$PDC    PTSEL=CONNOLLY
        CONSTR=NONE $END
$DATA
MEP calculation of compound # 3 - NIH1
C1
.
.
.
$END
```

Supplementary Table S3

Supplementary Table S3: Atom name, type, and RESP charges for **2** (Supplementary Fig. S12).

Name	Type	Charge	Name	Type	Charge	Name	Type	Charge
N	n1	-0.512323	C7	c3	-0.035047	H12	h1	0.044467
C	cg	0.311971	H4	hc	0.028198	C13	c3	-0.047273
C1	ca	-0.129484	H5	hc	0.028198	H13	hc	0.032962
C2	ca	-0.019222	C8	c3	-0.035047	H14	hc	0.032962
C3	ca	-0.021933	H6	hc	0.028198	C14	c3	-0.047273
C4	ca	0.496477	H7	hc	0.028198	H15	hc	0.032962
N1	nh	-0.838790	C9	c3	-0.005849	H16	hc	0.032962
H	hn	0.362499	H8	h1	0.054135	C15	c3	-0.003899
H1	hn	0.362499	H9	h1	0.054135	H17	h1	0.044467
N2	nb	-0.492619	C10	ca	-0.305391	H18	h1	0.044467
C5	ca	0.184688	H10	ha	0.164334	N5	nb	-0.451000
N3	nh	-0.044978	C11	ca	0.183393	C16	ca	0.464100
C6	c3	-0.005849	N4	nh	-0.044662	N6	nh	-0.833927
H2	h1	0.054135	C12	c3	-0.003899	H19	hn	0.354728
H3	h1	0.054135	H11	h1	0.044467	H20	hn	0.354728

Supplementary Table S4

Supplementary Table S4: Amber frcmod file used to define missing force field parameters for **2**.

```
frcmod file for NIH
MASS

BOND

ANGLE

DIHE
n1-cg-ca-ca   1    0.000         0.0             1.

IMPROPER
ca-ca-ca-cg         1.1          180.0         2.0      General improper torsional angle
(2 general atom types)

NONBON
```

Supplementary Table S5

Supplementary Table S5: Binding modes for **2**-CUG, number of structures observed in MD trajectories, and relative binding free energies of each mode computed with MM-GBSA (kcal/mol). Colors were used for each binding mode to emphasize conformational transitions seen in Figs. S12 and S13.

Conformation	# of structures	$\Delta G_{binding}$	Color	Color Name
a	46096	-35.00 ± 4.86		Black
b	2195	-31.75 ± 3.58		Cyan
c	2175	-33.82 ± 3.60		Silver
d	267	-32.65 ± 4.33		Magenta
e	6306	-36.43 ± 3.05		Blue
f	6065	-36.58 ± 3.09		Yellow
g	2390	-33.47 ± 3.19		Pink
h	21064	-36.31 ± 3.88		Red
i	815	-37.02 ± 3.31		Ochre
j	249	-34.29 ± 3.57		Indigo
k	219	-35.92 ± 3.56		Ice blue
l	1120	-37.49 ± 3.60		Tan
m	672	-35.66 ± 3.47		Purple
n	83	-32.48 ± 3.08		Maroon
o	2512	-41.76 ± 5.22		Green
p	6151	-42.45 ± 4.55		Orange

SUPPLEMENTARY METHODS

Fluorescence in situ Hybridization (FISH) to Detect Nuclear Foci. FISH assays were completed as previously described using a 2'OMe(CAGCAGCAGCAGCAGCAGC) FISH probe that is labeled with DY547 at the 5' end [27,38]. Immunostaining of MBNL1 was completed as previously described using the MB1a antibody (diluted 1:4), which was generously supplied by Prof. Glenn E. Morris (Wolfson Centre for Inherited Neuromuscular Disease) [33].

Computational Methods:

*Parameterization of **2**.* The system in Supplementary Figure S3 shows the structure of **2**. Atoms types and RESP charges were defined as previously described [47]. The AMBER GAFF force field was used to define atom types [48] while RESP charges were derived following the RESP protocol [49-52]. (See Table S2 for atom names, types, and RESP charges for **2**). The molecule was optimized and the electrostatic potential as a set of grid points was calculated at the HF level using the 6-31G* basis set. These calculations were performed with GAMESS [53]. (See Table S2 for variables used in optimization and MEP calculation).

Molecular Dynamics Simulations. The initial RNA structure, r(CCGCUGCGG), was taken from Protein Data Bank (PDB ID # 2l8C) [36]. A molecule of **2** was added to the structure far away from the RNA. For the RNA, amber99 force field [50,54] with revised χ [55], and α/γ [56] torsional parameter sets were used. For compound **2**, the frcmod shown in Table S3 and residue libraries defined in Table S2 were used. Modified Generalized

Born implicit solvent model (GB^{OBC}), model II [57], with 0.3 M salt concentration was used in all MD simulations. The sander.MPI module of AMBER ver. 11 [58] was used in order to run the MD simulations.

*Creating random initial structures of **2** bound to the 5'CUG/3'GUC region*. Compound **2** binds r(CUG) repeats but the structural properties of the binding mode are unknown. Therefore, a protocol was created where **2** was moved back and forth to the center of 5'CUG/3'GUC region while positional restraints with a restraint force of 1.0 kcal/mol-Å^2 were imposed to the heavy atoms of cytidine and guanosine residues in order to keep the global 3D RNA structure in A-form conformation. This way, the uridine residues were left free to transform to random conformations while **2** was pulled toward the center of 5'CUG/3'GUC region. During the whole process, chirality restraints were imposed to the system in order to keep the residues in their correct orientation. A total of 50 initial conformations were created.

Langevin dynamics with collision frequency of 1 was used during the whole process with a long-range cutoff of 20 Å. The temperature was kept at 300 K. In order to move **2** close to the center of 5'CUG/3'GUC region, a harmonic restraint with a force constant 50 kcal/mol-Å^2 was used between the centers of 5'CUG/3'GUC and **2**; 10000 steps with a 1 fs time step were completed. The final conformation was taken as one of the randomly created initial conformations that were used in production runs. In order to move away **2** from the RNA structure, a harmonic restraint with a force constant 10 kcal/mol-Å^2 was used between the centers of 5'CUG/3'GUC and **2**; 20000 steps with a 1

fs time step were completed. This whole process was iterated 50 times yielding a total of 50 initial structures for the production runs.

Restrained implicit solvent MD simulations. The 50 random conformations created above were used as initial structures in the implicit solvent MD simulations. The same simulation process described above was used in the production runs except that no positional restraints were used; however, dihedral and chirality restraints were used in RNA residues in order to keep the global 3D structure in A-form. Hydrogen bonding restraints were imposed to Watson-Crick GC base pairs. No dihedral or hydrogen bonding restraints were imposed to the UU base pair in order to guarantee sampling of all conformational space. A harmonic distance restraint between the centers of **2** and 5'CUG/3'GUC region was imposed whenever it was greater than 10 Å in order to keep **2** close to the binding region. Individual MD simulations each ~500 ns long were run (50 total) except for one, which was simulated beyond 1.1 μs in order to see potential conformational transformations. Trajectories were written at each 10 ps step. Over 175K CPU time was used for the MD.

Analysis

1. *Cluster Analysis:* First 5 ns of each MD simulation were not included in the analysis. From each trajectory, structures were extracted at intervals of 250 ps making more than 98000 structures for the clustering process. Cluster analysis showed 16 unique binding modes (Supplementary Figure S13 and Supplementary Table S5).

2. *Relative binding free energy calculations using MM-GBSA:* In order to quantify the binding of **2** to the r(CUG) repeats, relative binding free energies were calculated for

each binding mode/cluster with MM-GBSA approach (Supplementary Table S5). The MMPBSA.py module of AMBER v. 12 was used for this purpose [58]. A schematic representation of potential conformational transitions of **2** in the binding process is shown in Supplementary Figure S13. Structures shown in Supplementary Figure S9 are averaged structures using all conformations clustered for each binding mode. The lowest free energy structure is shown in Figure 6B (binding mode (p) – see also Supplementary Table S5 and Supplementary Figure S13).

SUPPLEMENTARY NOTE 1

One of the hallmarks of DM1 is the presence of nuclear foci, which consist r(CUG)^{exp} complexed with various proteins, including MBNL1 [15]. Since **1** and **2** both disrupt the r(CUG)^{exp}-MBNL1 complex, albeit by two distinct mechanisms, both compounds should diminish the number, size and/or intensity of nuclear foci. The luciferase reporter system described above that mimics the translational defect associated with DM1 is an indirect readout of the disruption of nuclear foci.

HeLa cells were transfected with the DM1 mini-gene and then treated with compound **1** or **2**. Nuclear foci were detected using fluorescence *in situ* hybridization and a DY547-labeled oligonucleotide that is complementary to r(CUG)^{exp}. Confocal microscopy was employed to image the cells. As shown in Supplementary Figure S2, addition of 100 μM of **1** or **2** causes a decrease in the size and number of foci per cell relative to untreated cells that express r(CUG)^{exp}. (Similar results were also observed when cells were treated with 500 μM of **1**.) Although **1** and **2** disperse foci to varying degrees, the disruption of nuclear foci is not a prerequisite for compounds to improve pre-mRNA splicing or translation defects associated with DM1, which only requires freeing sufficient amounts of the *DMPK* RNA and MBNL1. Previous studies have shown that the presence of foci and splicing defects can be decoupled [46]; a small molecule that targets r(CAG)^{exp} improves pre-mRNA splicing defects associated with sequestration of MBNL1 even though it does not affect the formation of r(CAG)^{exp}-MBNL1 foci [40].

SUPPLEMENTARY NOTE 2

Analyses show that **2** does not bind to the RNA's major groove. Compound **2** first contacts the UU base pairs occur through minor groove (Supplementary Figure S13, conformation a). After this binding mode (a), there are multiple different transformation pathways for **2** that leads to the lowest free energy conformations of (o) and (p) (Supplementary Figure S13). Binding modes (b), (c), (d), (j), (k), (m), and (n) described in Supplementary Table S5 and Supplementary Figure S13 are transition states. Trajectory analysis show that **2** does not stay in these binding modes for a long time while it spends over 100 ns of MD time in binding modes (a), (e), (f), (h), (o), and (p) (Supplementary Fig. S14).

REFERENCES

15. Cardani, R., Mancinelli, E., Rotondo, G., Sansone, V. & Meola, G. Muscleblind-like protein 1 nuclear sequestration is a molecular pathology marker of DM1 and DM2. *Eur. J. Histochem.* **50**, 177-182 (2006).

27. Warf, M.B., Nakamori, M., Matthys, C.M., Thornton, C.A. & Berglund, J.A. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18551-18556 (2009).

33. Holt, I. et al. Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells* **12**, 1035-1048 (2007).

36. Parkesh, R., Fountain, M. & Disney, M.D. NMR spectroscopy and molecular dynamics simulation of r(CCGCUGCGG)$_2$ reveal a dynamic UU internal loop found in myotonic dystrophy type 1. *Biochemistry* **50**, 599-601 (2011).

38. Parkesh, R. et al. 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.* **134**, 4731-4742 (2012).

40. 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).

46. Ho, T.H., Savkur, R.S., Poulos, M.G., Mancini, M.A., Swanson, M.S. & Cooper, T.A. Colocalization of muscleblind with RNA foci is separable from mis-regulation of alternative splicing in myotonic dystrophy. *J. Cell. Sci.* **118**, 2923-2933 (2005).

47. Eryazici, I., Yildirim, I., Schatz, G.C. & Nguyen, S.T. Enhancing the melting properties of small molecule-DNA hybrids through designed hydrophobic interactions: an experimental-computational study. *J. Am. Chem. Soc.* **134**, 7450-7458 (2012).

48. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157-1174 (2004).

49. Bayly, C.I., Cieplak, P., Cornell, W.D., Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

50. Cornell, W.D., et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc*. **117**, 5179-5197 (1995).

51. Cornell, W.D., Cieplak, P., Bayly, C.I. & Kollmann, P.A. Application of RESP charges to

calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **115**, 9620–9631 (1993).

52. Cieplak, P., Cornell, W.D., Bayly, C. & Kollman, P.A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comp. Chem.* **16**, 1357–1377 (1995).

53. Schmidt, M.W., et al. General atomic and molecular electronic structure system. *J. Comp. Chem.* **14**, 1347–1363 (1993).

54. Wang, J., Cieplak, P. & Kollman, P.A. How well does a restrained electrostaticpotential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2**, 1049–1074 (2000).

55. Yildirim, I., Stern, H.A., Kennedy, S.D., Tubbs, J.D. & Turner, D.H. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J. Chem. Theory Comput.* **6**, 1520-1531 (2010).

56. Perez, A., et al. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **92**, 3817-3829 (2007).

57. Onufriev, A., Bashford, D. & Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383-394 (2004).

58. Case, D.A., et al. AMBER 11. University of California, San Francisco 2010.