Supporting information for:

Structural features of small molecules targeting the RNA repeat expansion that causes genetically defined ALS/FTD

Andrei Ursu[1], Kye Won Wang[2], Jessica A. Bush[1], Shruti Choudhary[1], Jonathan L. Chen[1], Jared T. Baisden[1], Yong-Jie Zhang[3], Tania F. Gendron[3], Leonard Petrucelli[3], Ilyas Yildirim[2] and Matthew D. Disney[1]*
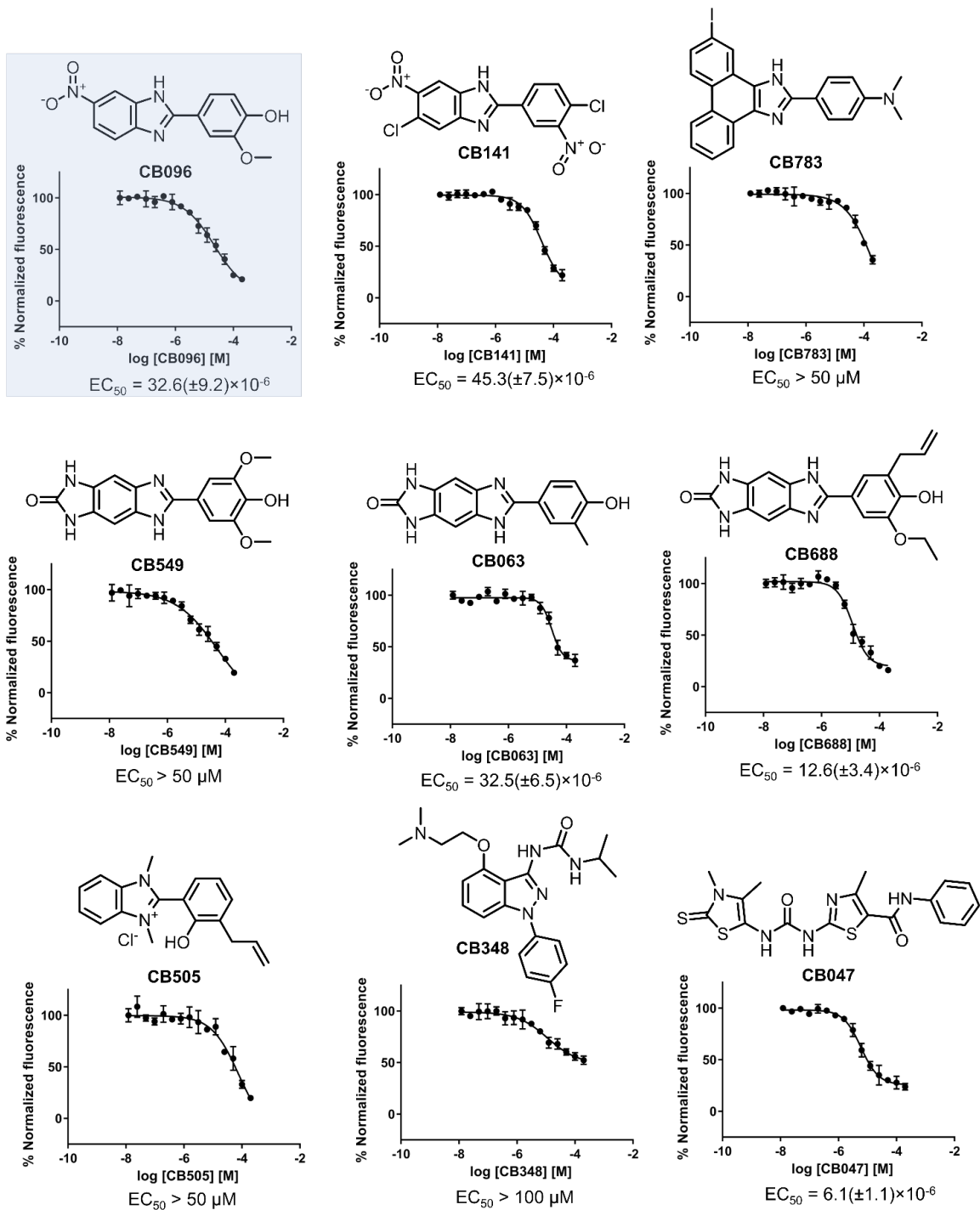
[1] Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458 USA

[2] Department of Chemistry and Biochemistry, Florida Atlantic University, Jupiter, FL 33458 USA

[3] Department of Neuroscience, Mayo Clinic, 4500 San Pablo Rd., Jacksonville, FL 32224 USA

* Author to whom correspondence is addressed; Email: disney@scripps.edu
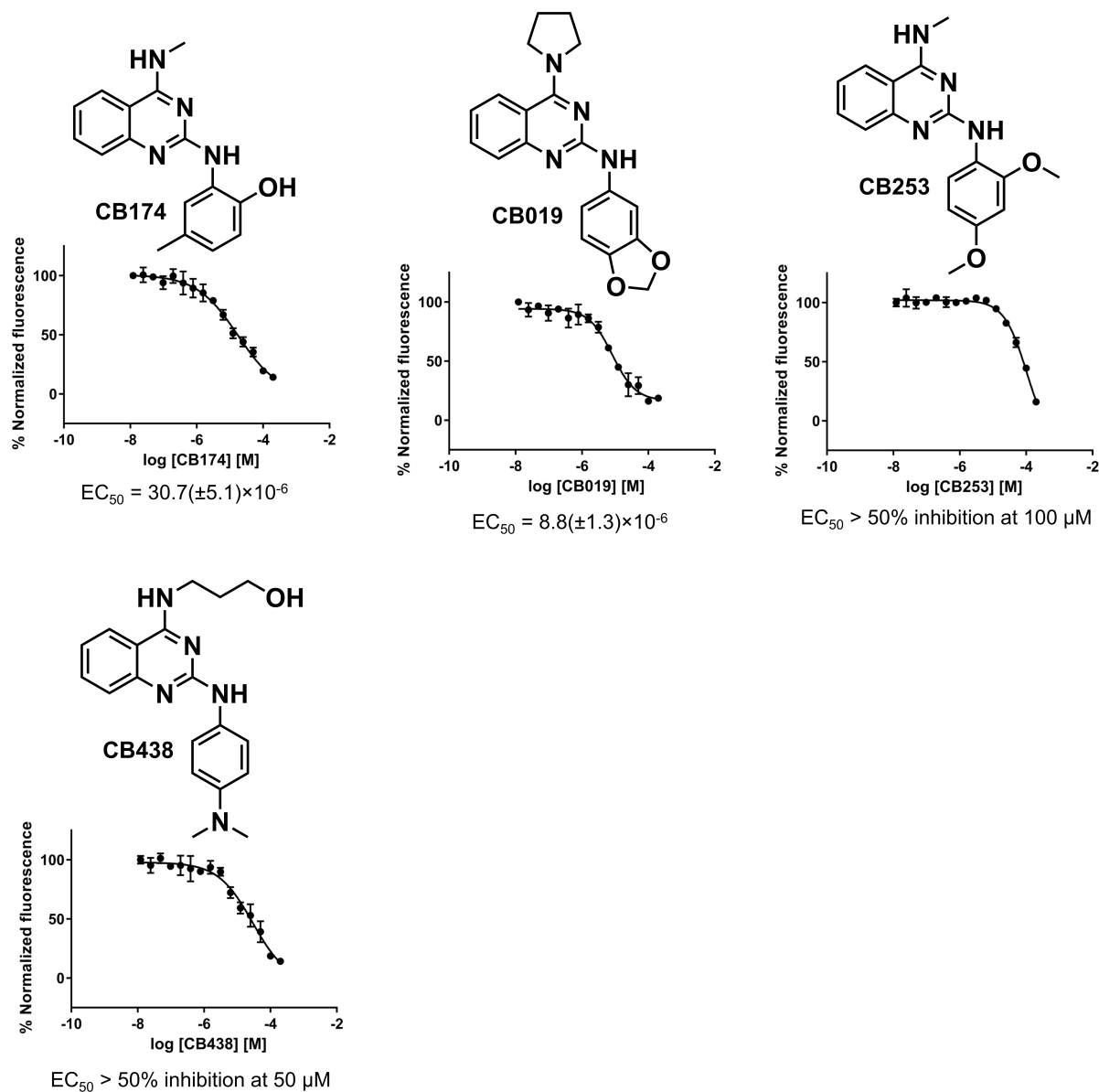
Supplementary Figures

CB096
EC$_{50}$ = 32.6($\pm$9.2)$\times10^{-6}$

CB141
EC$_{50}$ = 45.3($\pm$7.5)$\times10^{-6}$

CB783
EC$_{50}$ > 50 μM

CB549
EC$_{50}$ > 50 μM

CB063
EC$_{50}$ = 32.5($\pm$6.5)$\times10^{-6}$

CB688
EC$_{50}$ = 12.6($\pm$3.4)$\times10^{-6}$

CB505
EC$_{50}$ > 50 μM

CB348
EC$_{50}$ > 100 μM

CB047
EC$_{50}$ = 6.1($\pm$1.1)$\times10^{-6}$

CB174

EC$_{50}$ = 30.7($\pm$5.1)$\times$10^{-6}

CB019

EC$_{50}$ = 8.8($\pm$1.3)$\times$10^{-6}

CB253

EC$_{50}$ > 50% inhibition at 100 µM

CB438

EC$_{50}$ > 50% inhibition at 50 µM

Figure S1. Hit validation of from r(G$_4$C$_2$)$_8$ binders via TO1 displacement *in vitro*. EC$_{50}$ values were determined by plotting the change in normalized fluorescence as a function of compound concentration. Data points are the mean $\pm$ SD of two independent experiments in which each concentration was measured in triplicate. The binding curve was fitted using a four-parameter nonlinear regression equation of the GraphPad Prism 8.1.2 suite.
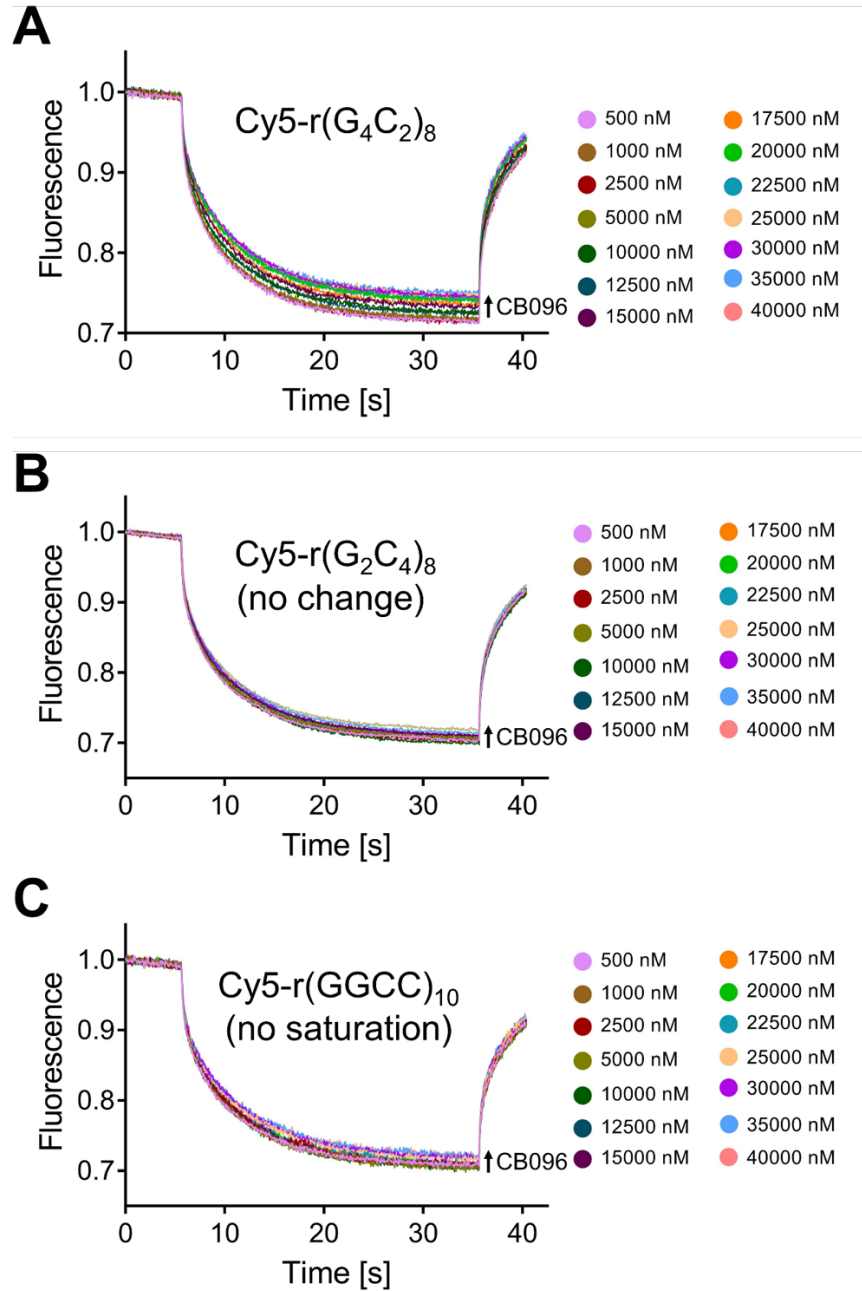
Figure S2. Representative thermophoresis curves of CB096 binding to Cy5-labeled RNAs obtained. Dose dependent addition of CB096: i) increases the thermophoresis of the hairpin form of Cy5-r(G_4C_2)$_8$ (a); ii) leads to minor change of the Cy5-r(G_2C_4)$_8$ (antisense) thermophoresis (b); iii) triggers no saturable change of the Cy5-r(GGCC)$_{10}$ (base pair control) thermophoresis (c). The normalized fluorescence upon dose-dependent addition of CB096 was recorded in the Thermophoresis and T-jump channels. Data points are representative of two independent experiments each performed in duplicate.
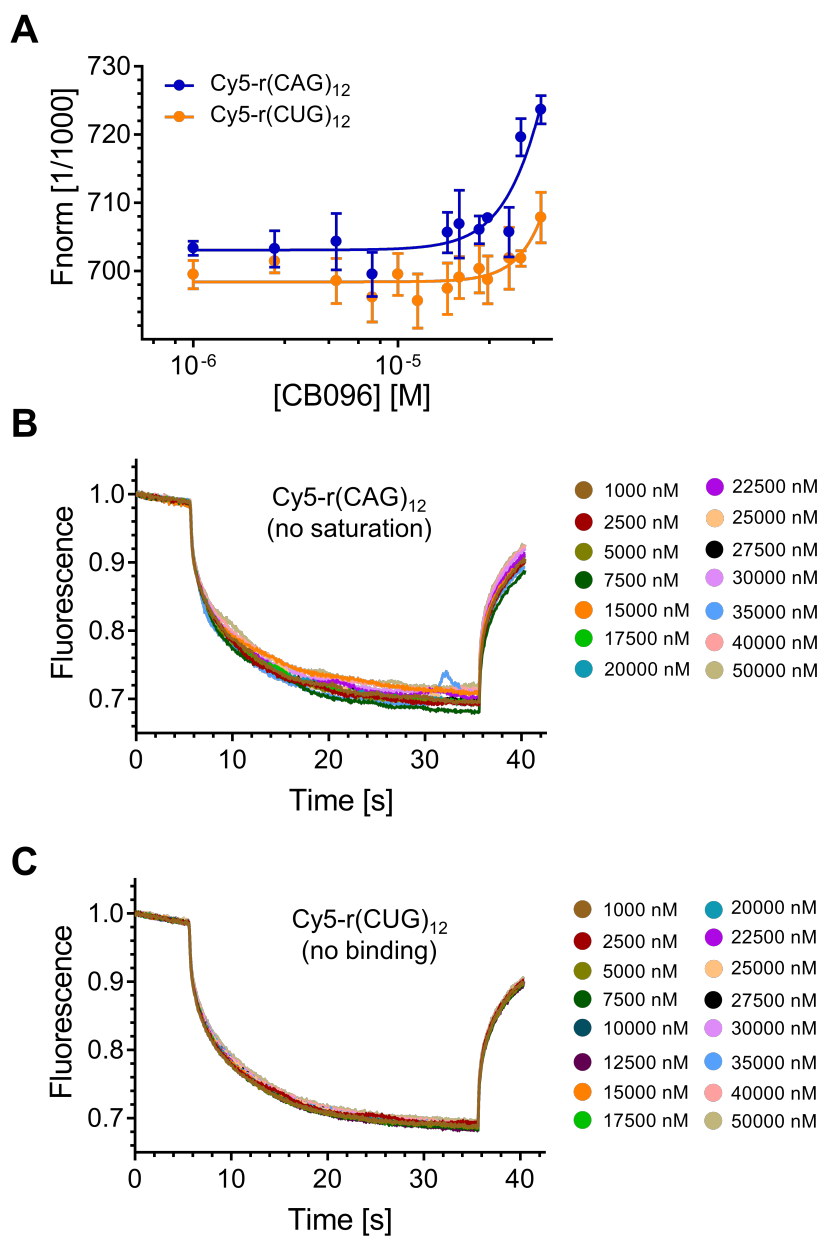
Figure S3. CB096 specifically interacts with the 1×1 GG internal loops within the r(G$_4$C$_2$)^{exp} hairpin RNA. a) No saturable binding was observed in thermophoresis measurement for CB096 and RNAs featuring 1×1 AA and 1×1 UU internal loops within Cy5-r(C<u>A</u>G)$_{12}$ and Cy5-r(C<u>U</u>G)$_{12}$ RNAs, respectively, as assessed via MST. b-c) Representative thermophoresis curves of CB096 interacting with Cy5-r(C<u>A</u>G)$_{12}$ (b) and Cy5-r(C<u>U</u>G)$_{12}$ (c) RNAs. Dose dependent addition of CB096: i) triggers no saturable change of the Cy5-r(C<u>A</u>G)$_{12}$ thermophoresis (b); and ii) leads to minor change of the Cy5-r(C<u>U</u>G)$_{12}$ thermophoresis (c). The normalized fluorescence at each CB096 concentration was recorded in the Thermophoresis and T-jump channels. Data points in panel a were generated from two independent experiments each performed in duplicate.
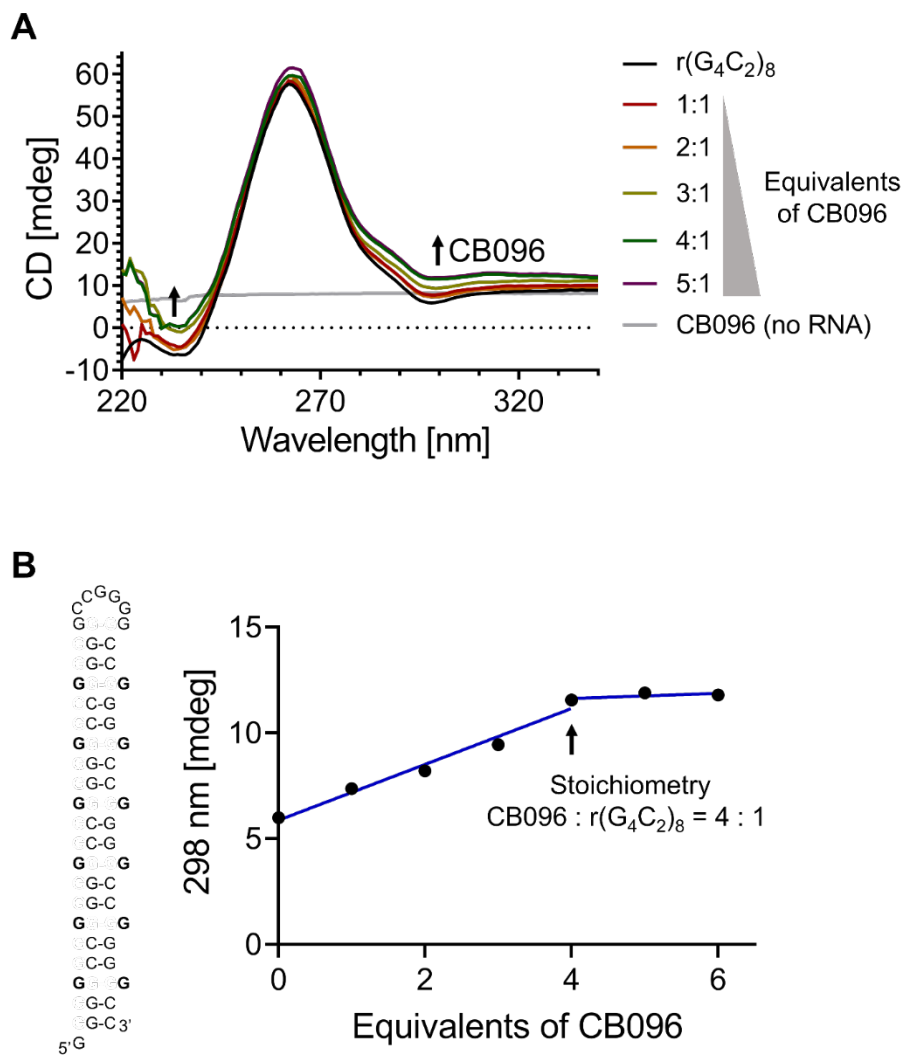
A

B

Figure S4. Representative CD spectrum recorded upon dose dependent addition of CB096 to r(G$_4$C$_2$)$_8$ hairpin. Compound addition triggers an increase of CD signals at 235 and 298 nm (a) that reached saturation at 4 equivalents CB096, *i.e.* a binding stoichiometry of CB096 to r(G$_4$C$_2$)$_8$ of 4:1 (b). The spectra are representative of two independent experiments.
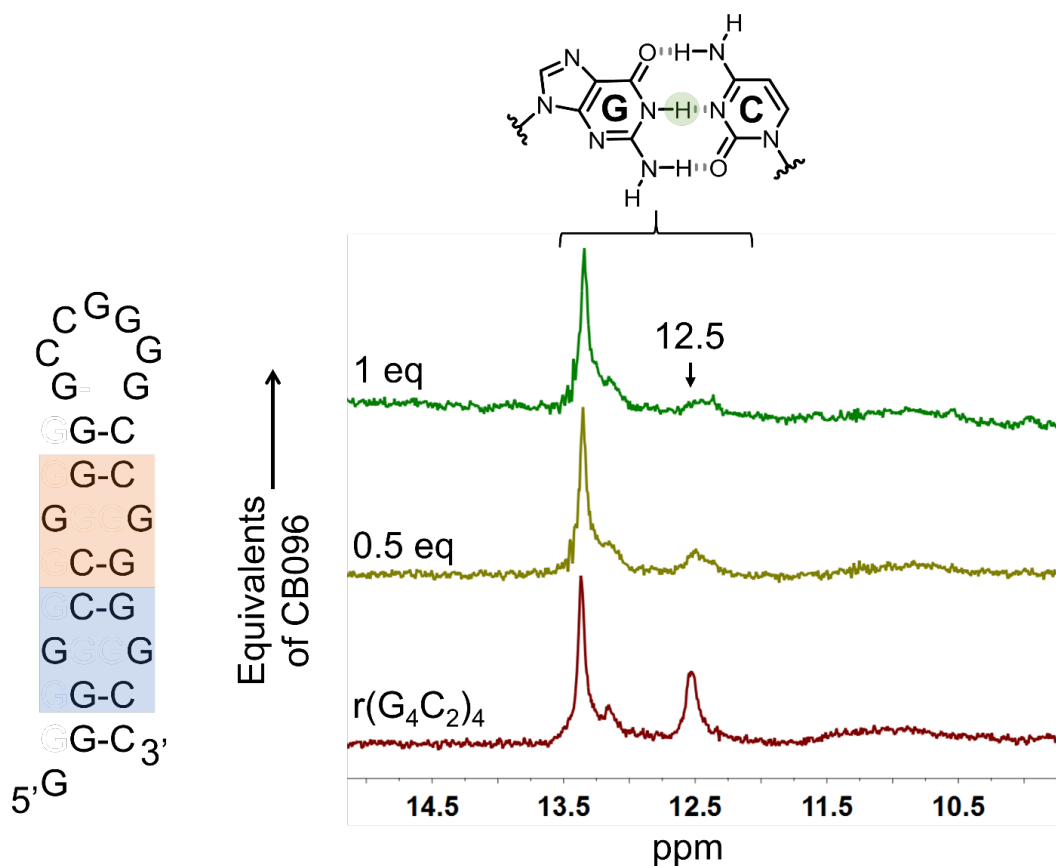
Figure S5. CB096 mode of binding to the r(G$_4$C$_2$)$_4$ hairpin. CB096 broadens the imino proton peak of GC/CG base pairs at 12.5 ppm. Blue labels indicate 5'G<u>G</u>C/3'C<u>G</u>G (<u>G</u> represents the G residues within the 1×1 GG internal loops) internal loops, whereas the orange labels indicate 5'C<u>G</u>G/3'G<u>G</u>C loops. Further addition of CB096 triggered precipitation. The 1D NMR spectra are representative of two independent experiments.
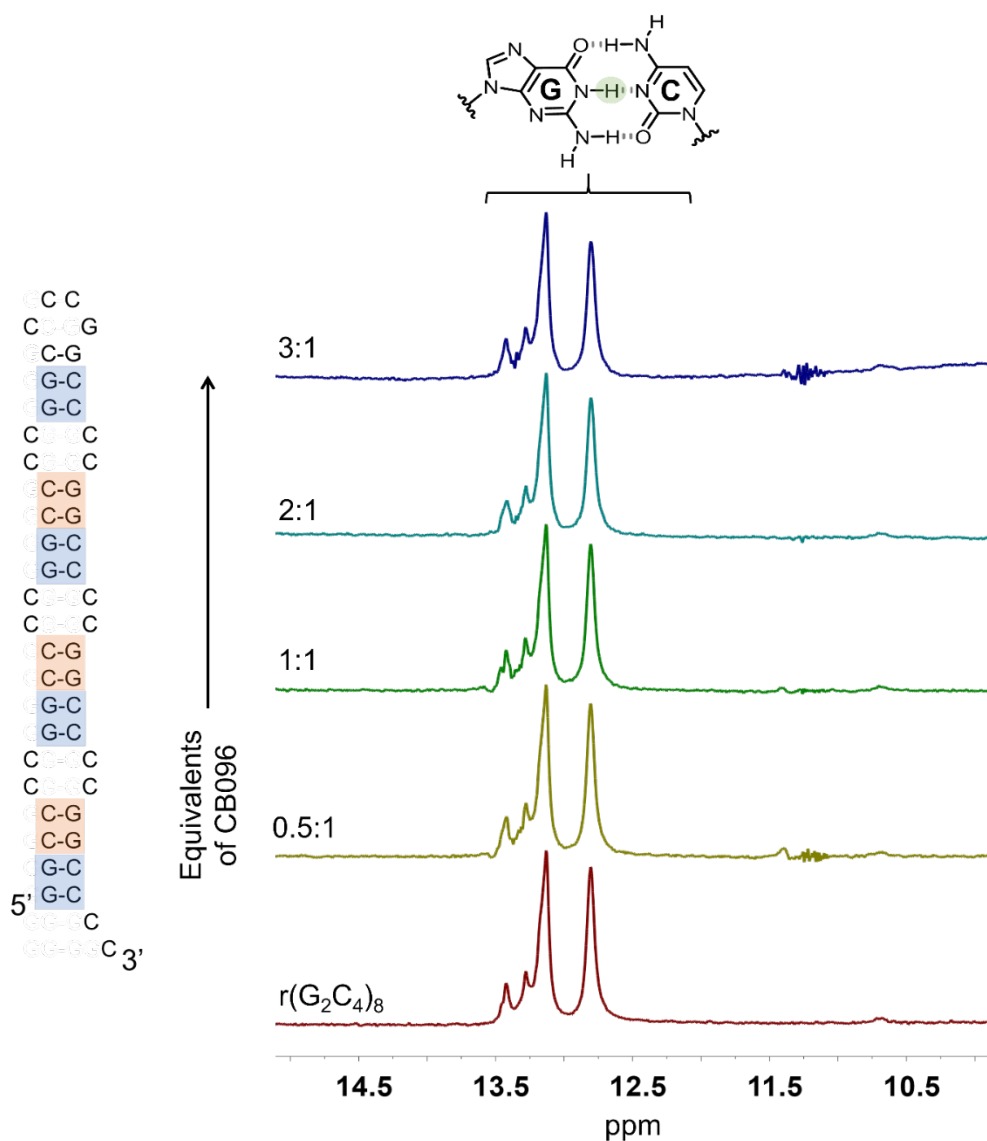
Figure S6. Dose dependent addition of CB096 does not change the imino proton signals of the 5'GG/3'CC (blue squares) and 5'CC/3'GG (orange squares) base pairs at 13.1 ppm and 12.8 ppm within r(G_2C_4)$_8$ (antisense). The 1D NMR spectra are representative of two independent experiments.
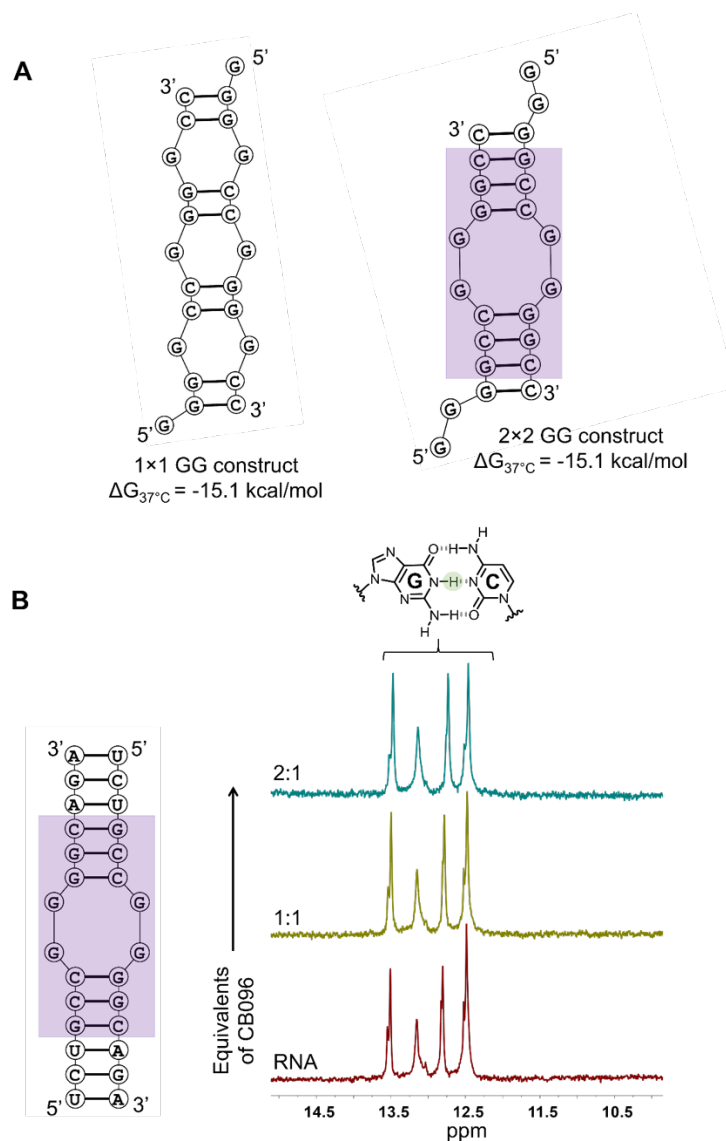
Figure S7. CB096 does not interact with a 2×2 GG internal loop-containing duplex. a) Free energy minimization (RNAstructure software) of the $r(G_4C_2)_2$ duplex predicts secondary structures with 1×1 GG and 2×2 GG loops have the same free energy of folding at 37 °C (ΔG^o_{37}). b) Dose dependent addition of CB096 does not affect the imino proton signal of GC/CG base pairs at 12.5 ppm within the 2×2 GG internal loop containing model construct. The 1D NMR spectra are representative of two independent experiments.
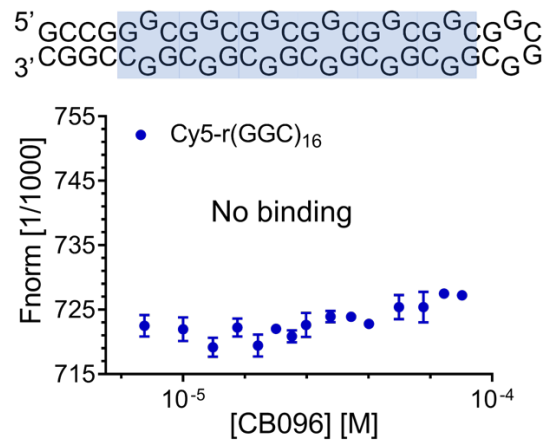
Figure S8. Binding of CB096 to 5'G$\underline{G}$C/3'C$\underline{G}$G sites (highlighted in blue) within Cy5-r(G$\underline{G}$C)$_{16}$ via MST. No change in thermophoresis, *i.e.* no binding, was observed upon dose dependent addition of CB096 to Cy5-r(G$\underline{G}$C)$_{16}$. The secondary structure was predicted by RNAstructure software. The normalized fluorescence was recorded in the Thermophoresis and T jump channels. Data points indicate mean values ± SD from two independent experiments each performed in duplicate.
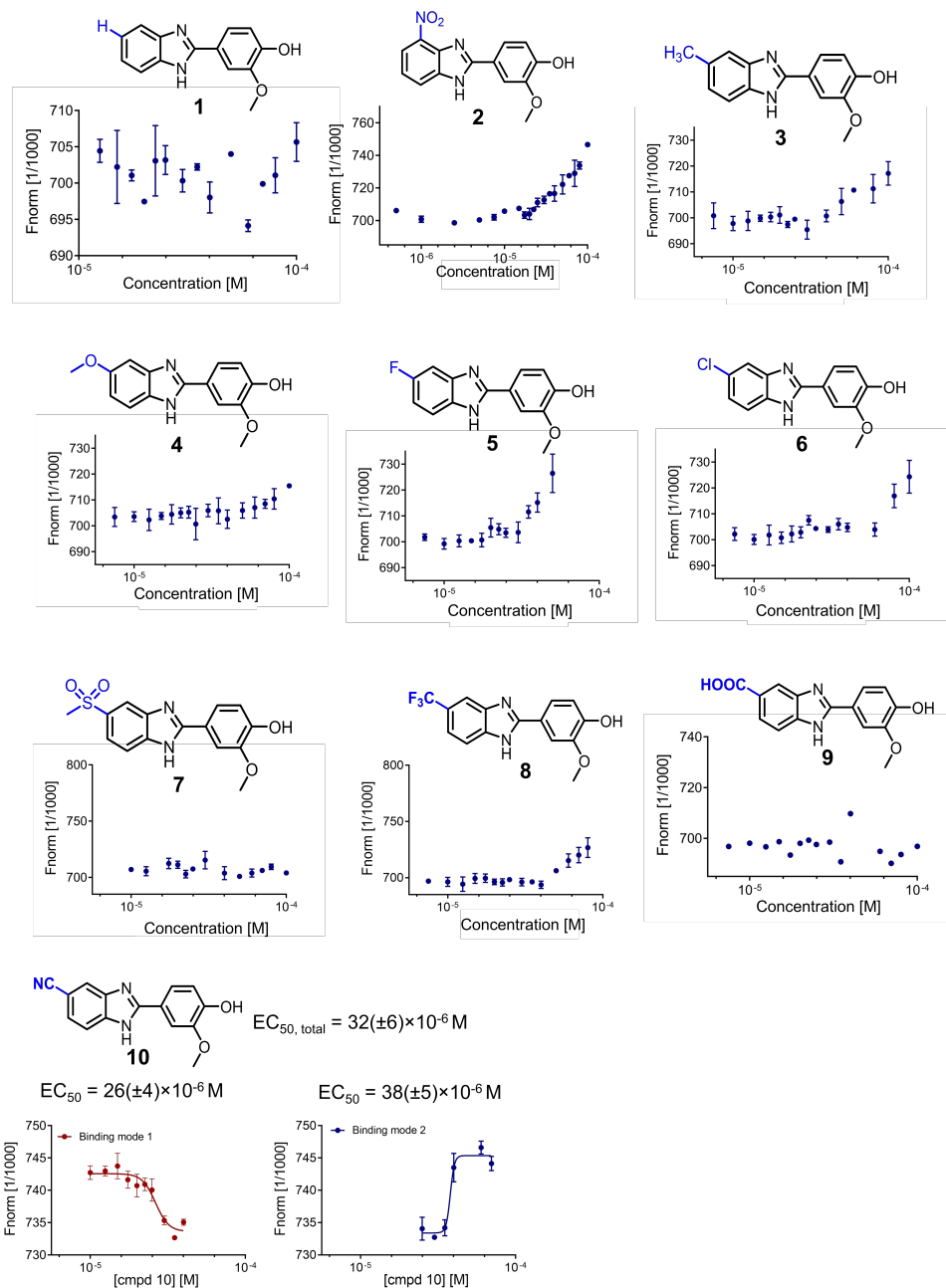
Figure S9. *In vitro* binding of compounds 1-10 to Cy5-r(G$_4$C$_2$)$_8$ hairpin, as measured by MST. No change in thermophoresis or no saturation, *i.e.* no binding, was observed upon dose dependent addition of derivatives 1-10 to Cy5-r(G$_4$C$_2$)$_8$ as shown by the normalized fluorescence signal recorded in the Thermophoresis and T jump channels. Saturable binding was observed for derivative 10 only, with EC$_{50,total}$ value of 32(±6) μM, calculated by plotting the changes in normalized fluorescence as function of compound concentration. EC$_{50,total}$ considered the EC$_{50}$ values for the two observed binding modes and was generated by calculating the square root of the first binding mode (EC$_{50,1}$) multiplied by the value of the second binding mode (EC$_{50,2}$). Data points are reported as the mean ± SD from two independent experiments each performed in duplicate.
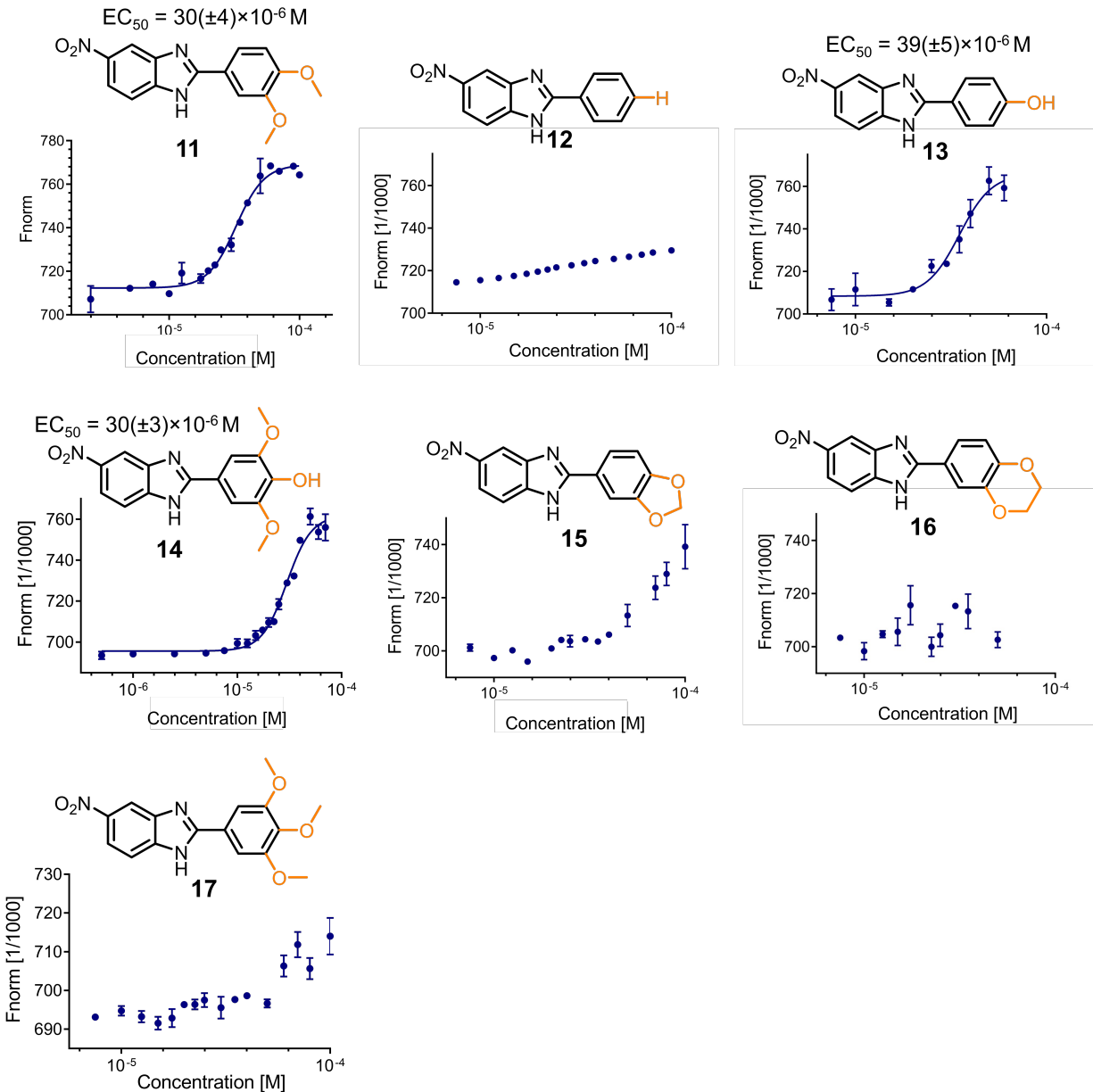
Figure S10. *In vitro* binding of compounds 11-17 to Cy5-r(G_4C_2)$_8$ hairpin, as measured by MST. Saturable binding was observed for derivatives 11, 13, and 14 only, with EC$_{50}$ values of 30(±4) μM, 39(±5) μM, and 30(±3) μM, respectively. EC$_{50}$ values were determined by plotting the changes in normalized fluorescence as function of compound concentration. No change in thermophoresis or no saturation, *i.e.* no binding, was observed addition of derivatives 12,15-17 to Cy5-r(G_4C_2)$_8$. Data points are reported as mean ± SD from two independent experiments each performed in duplicate. Binding curves were fitted using a four-parameter nonlinear regression equation of the GraphPad Prism 8.1.2 suite.
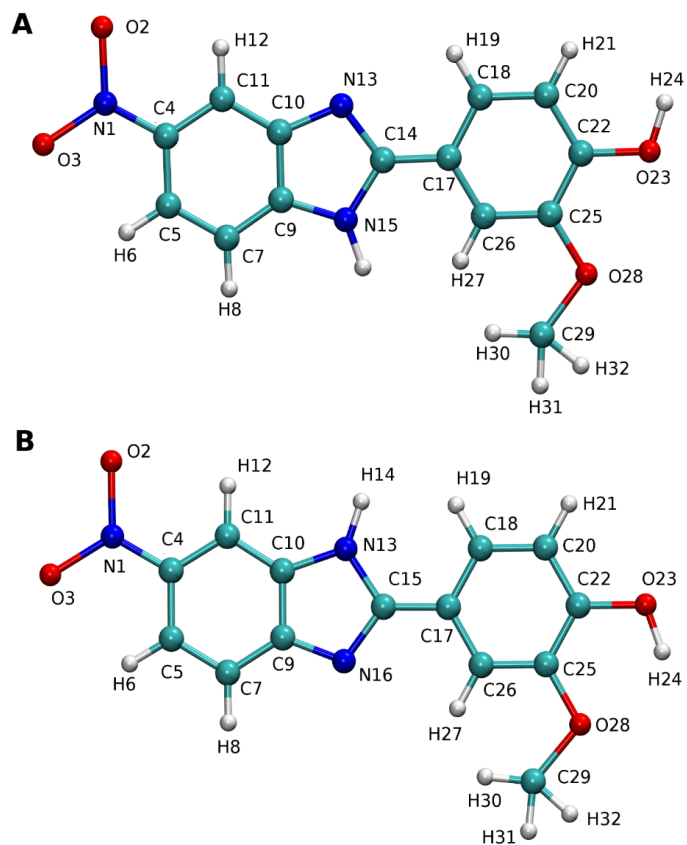
Figure S11. Atom numbering of CB096 (a) and its tautomer CB096$_T$ (b).
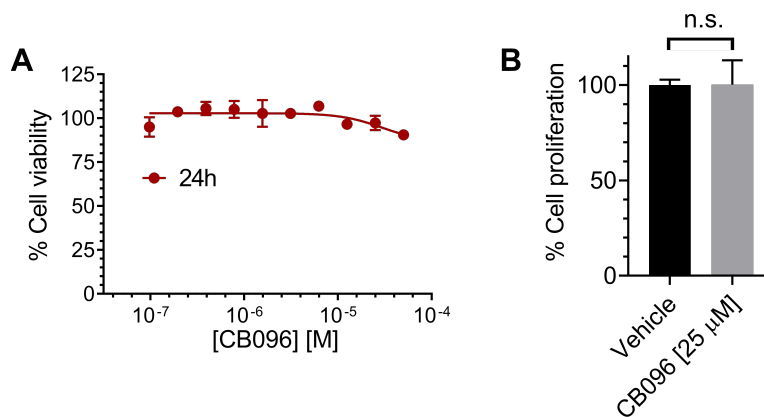
Figure S12. CB096 is not cytotoxic in $r(G_4C_2)_{66}$-transfected HEK293T cells nor affects cell proliferation. a) Cellular viability of $r(G_4C_2)_{66}$-transfected HEK293T cells upon incubation with CB096 for 24 h assessed via CellTiter-Glo. Data points are reported as the mean ± SD from two independent experiments each performed in triplicate. b) CB096 (25 µM) does not alter the cell proliferation of $r(G_4C_2)_{66}$-transfected HEK293T cells. Data points are reported as the mean ± SD relative to vehicle (0.5% DMSO) from two independent experiments each performed in triplicate. n.s. = not significant, as determined by a two-tailed Student t test.
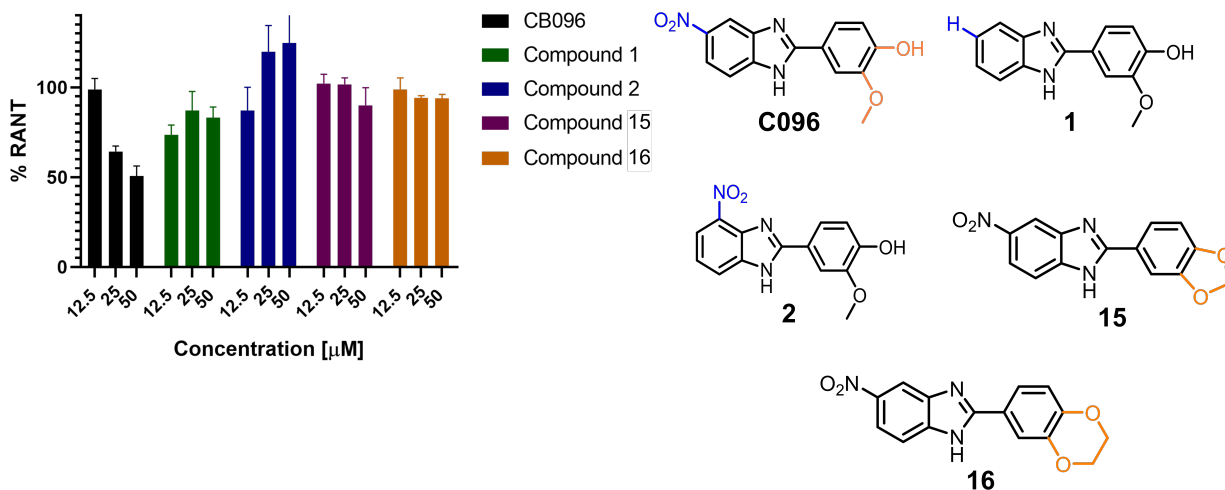
Figure S13. Inhibition of RAN translation by CB096 and compounds 1, 2, 15 and 16. CB096 binds $r(G_4C_2)_8$ hairpin *in vitro* and inhibits RAN translation in cells in a dose dependent manner. Conversely, derivatives 1, 2, 15 and 16 do not bind $r(G_4C_2)_8$ and do not block RAN translation. Data are reported as the mean ± SD from three independent experiments each performed in duplicate.
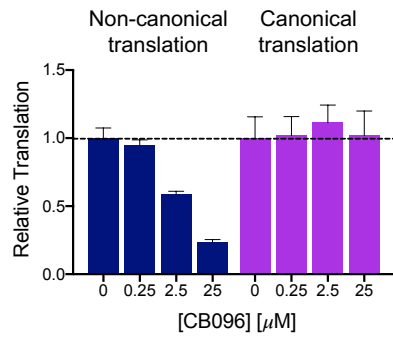
Figure S14. CB096 selectively inhibits RAN translation of $r(G_4C_2)^{exp}$ in HEK293T cells that stably express Lentiviral-S-tdTomato co-transfected with plasmind encoding a $(G_4C_2)_{66}$-*No ATG*-firefly luciferase (RAN translation), Renilla luciferase (canonical translation), and *ATG*-GFP. Relative translation was measured by normalizing the signal from firefly luciferase to Renilla luciferase (non-canonical translation) or the signal from Renilla luciferase to GFP (canonical translation). Data are reported as the mean ± SD (n = 4).

METHODS

General methods

RNA oligonucleotides were purchased from Dharmacon (GE Healthcare) and deprotected according to the vendor's recommended procedure. After deprotection, the RNAs were desalted using PD-10 columns (GE Healthcare) and the concentrations were determined using UV spectrometry to measure the absorbance at 260 nm (extinction coefficients for RNAs were provided by the vendor) and 90 °C using a Beckman Coulter DU 800 spectrophotometer.

The sequences of the RNA hairpins used herein were $r(G_4C_2)_4$, $r(G_2C_4)_8$, $r(GGCC)_{10}$, $r(CGG)_{16}$, $r(CAG)_{12}$, $r(CUG)_{12}$ and $r(GGC)_{16}$: 5'-GCCGGGCGGCGGCGGCGGCGGCGGCGGCG GCGGCGGCGGCGGCCGGC. The sequences of RNA duplexes include: $r(G_4C_2)_2$, ^{5F}C5-$r(G_4C_2)_2$: 5'-GGGG^{5F}CCGGGGCC-3'; ^{5F}C6-$r(G_4C_2)_2$: 5'-GGGGC^{5F}CGGGGCC-3'; ^{5F}C11-$r(G_4C_2)_2$: 5'-GGGGCCGGGG^{5F}CC-3'. RNA secondary structure predictions shown throughout the study were generated using RNA structure software.[1,2] All hairpin RNAs were folded in 10 mM Na_2HPO_4, pH 7, for 4 min at 95 °C and cooled to room temperature. For duplex sequences, 100 mM LiCl solution was added to the RNA solution folded in 10 mM Na_2HPO_4, pH 7 and equilibrated at room temperature.

The RNA-focused library as well as the hit compounds CB096 and derivatives 1, 11 and 12 were purchased from ChemBridge. The characterization of the physicochemical properties of the RNA-focused library was described elsewhere.[3]

IN VITRO STUDIES

High-throughput TO1 displacement screening in 384-well format

The RNA $r(G_4C_2)_8$ was folded in 8 mM Na_2HPO_4, 185 mM NaCl, 2 mM EDTA, pH = 7.4 for 4 min at 95 °C and cooled to room temperature. BSA (40 μg/mL, final concentration) and TO-PRO-1 (TO1, 400 nM) were added to the solution and were incubated for 15 min at room temperature

protected from light (final concentrations of r(G$_4$C$_2$)$_8$ and TO1 were 250 and 200 nM, respectively).

10 μL aliquots of the r(G$_4$C$_2$)$_8$-TO1 solution were dispensed into low volume 384 well black plate

(Greiner Bio-One, 784076) using an Aurora liquid handling system. Compounds (0.1 μL from 10

mM compound stock concentration in 100% DMSO) were added via the Biomek NXP pin transfer

tool onto the plate (final compound concentration of 100 μM in 1% DMSO final concentration).

The reaction mixture was incubated for 15 minutes before reading the fluorescence via a

SpectraMax M5 plate reader (Molecular Device) using 485 nm (excitation) and 515 nm (emission).

Reported Z-factor represent the average of individual Z-factors calculated for each individual plate

considering the vehicle control (DMSO) and Methotrexate (10 μM) spotted in pentuplicate. The

percentage of decrease in TO1 fluorescence was calculated by using the averaged fluorescence

of five vehicle controls on each individual plate according to the following equation:

$$\% \; Decrease \; in \; TO1 \; Fluorescence = \frac{(Average \; Control \; Fluorescence - Well \; Fluoresence)}{Average \; Control \; Fluorescence}$$

where Average Control Fluorescence is the average fluorescence intensity measured in wells that

contain RNA, TO1, and vehicle (n = 5) and Well Fluorescence is the fluorescence intensity

measured from a compound-treated well of interest.

A cut-off of ≥75% decrease in TO1 fluorescence was selected as it represents the most potent

binders of r(G$_4$C$_2$)$_8$ hairpin (n = 26).

Hit validation

Compounds exhibiting ≥75% decrease in TO1 fluorescence were ordered procured from

commercial sources, and false positives were removed by applying stringent selection criteria: (i)

lack of interference with the signal readout due to intrinsic fluorescence; (ii) lack of aggregation at

the tested concentration (visual inspection); and (iii) and dose-dependent displacement of TO1

as described above with a 1% final concentration of DMSO. Dose-dependent studies were

measured in duplicate, and the EC_{50} values were determined by plotting the changes in normalized fluorescence as a function of compound concentration. Data points are reported as the mean ± SD from two independent experiments, each performed in triplicate. Binding curves were fitted using a four-parameter nonlinear regression equation of the GraphPad Prism 8.1.2 suite.

Affinity measurements via microscale thermophoresis (MST)

Fluorescently labeled (5'-Cy5) RNAs were folded in 10 mM Na_2HPO_4, pH 7, and 100 mM LiCl for 4 min at 95 °C and cooled to room temperature. Compound dilutions (with 0.5% final DMSO concentration) were added to the RNA solution (final concentration of 10 nM) and incubated at room temperature for 5 min. MST measurements were performed on a Monolith NT.115 (NanoTemper) instrument using standard capillaries and the following settings: LED power 5%; MST power 80%; fluorescence before 5 s; MST on 30 s; fluorescence after 5 s, and delay 25 s. Individual experiments were measured in duplicate, and EC_{50} values were determined by plotting the changes in normalized fluorescence in the Thermophoresis and T jump channels as a function of small molecule concentration. Data points are reported as the mean ± SD from two independent experiments. Binding curves were fitted using a four-parameter nonlinear regression equation of the GraphPad Prism 8.1.2 suite.

Binding stoichiometry of CB096 to r(G_4C_2)$_8$ via CD spectroscopy

CD experiments were performed on a J-815 Jasco spectropolarimeter over the spectral range of 220–450 nm using a bandwidth of 2 nm and a scan speed of 50 nm/min and a step resolution of 0.2 nm. The RNA sample was folded in 10 mM Na_2HPO_4, pH 7, for 4 min at 95 °C and cooled to room temperature at a final concentration of 10 μM. CB096 was then added from a 10 mM stock solution prepared in DMSO. CD spectra were recorded using a 1 mm Quartz Suprasil cuvette

(Hellma Analytics, 110-128-1-40) at 25 °C and are representative of two independent experiments. Data points were exported and processed using the GraphPad Prism 8.1.2 software suite.

NMR spectroscopy studies

All NMR measurements were performed on a Bruker 600 MHz NMR spectrometer equipped with a cryoprobe (Bruker, USA). 1D imino proton spectra were recorded in 90 % H_2O – 10 % D_2O using a WATERGATE pulse sequence for water suppression. Typical RNA concentration was 0.1 mM, and the recording time for individual NMR spectra was 16 min (512 scans) using a 600 MHz NMR spectrometer. Small molecules were typically added to the folded RNA solution from a 10 mM stock solution in 100% DMSO-d6 (DMSO-d6 content did not exceed 3% of the final volume). NMR spectra were processed using MestreNOVA software using 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS, 0.1 mM) as an internal reference. All NMR spectra are representative of two independent experiments.

COMPUTATIONAL STUDIES

Preparation of 5'C$\underline{GG}$/3'G$\underline{GC}$ and 5'G$\underline{GC}$/3'G$\underline{GC}$ Model Systems. Model RNA systems, r(CCC$\underline{G}$GGG)$_2$ and r(GGG$\underline{G}$CCC)$_2$, representing 5'C$\underline{GG}$/3'G$\underline{GC}$ and 5'G$\underline{GC}$/3'C$\underline{GG}$ sites of r(G$_4$C$_2$)^{exp} were prepared for MD simulations using the nucgen module of AMBER16.[4] The 1×1 GG internal loops in the initial structures were designed to have *syn-anti* orientations. The initial structures were minimized using a mix of positional and geometric restraints to maintain the global structure in A-form and GG base pairs in cis Watson-Crick/Hoogsteen orientations. The systems were then neutralized with 12 Na$^+$ ions[5] and solvated with 4305 TIP3P[6] water molecules in a truncated octahedral box. An extra 5 Na$^+$ and Cl$^-$ ions were included in each system. The Na$^+$ concentrations in both systems were 0.21 M once they had reached equilibration. The AMBER forcefield with revised χ[7] and α/γ[8] torsional parameters was used in the MD simulations.

Molecular Dynamics (MD) Simulations of 5'C<u>G</u>G/3'G<u>G</u>C and 5'G<u>G</u>C/3'C<u>G</u>G. The structures were first minimized and then equilibrated as previously described.[9,10] For each system, 1 µs MD simulation with a 2 ps time step were run after minimization and equilibration. NPT dynamics was utilized in each MD simulation where isotropic positional scaling was turned on with a reference pressure of 1 atm and a pressure relaxation time of 2 ps. SHAKE was used to constrain bonds involving hydrogen atoms.[11] An atom-based long-range hard cutoff of 8.0 Å was used in the production runs. The reference temperature was 300K in all the MD simulations. The pmemd.cuda code of AMBER16 was used to perform the MD simulations.[4]

Analysis. The ptraj module of AMBER16[4] was used to analyze the MD trajectories. X3DNA[12] was also used to extract global structural properties from the MD trajectories such as the overlap areas displayed in Figure 4A.

CB096's Binding Mode to 5'C<u>G</u>G/3'G<u>G</u>C via MD simulations

Parameterization of compounds. The force field parameters of CB096 (**Supplemental Figure S13A**) and CB096$_T$ (the tautomeric structure of CB096; **Supplemental Figure S13B**) were prepared as previously described.[13-16] Each compound was parameterized using the Generalized Amber Force Field.[17] Gaussian09[18] was used to calculate the RESP charges.[19, 20] Structures were first optimized, and then the electrostatic potentials as a set of grid points were calculated using the HF/6-31G* basis set (**Supplemental Tables S1 & S2**).

The CB096-5'C<u>G</u>G/3'G<u>G</u>C complex. Previously, we applied the dynamic binding approach to study small molecule-RNA interactions.[13, 14, 16, 21] The same approach was utilized to investigate the binding properties of CB096 and CB096$_T$ to the 5'C<u>G</u>G3' motif within model r(G_4C_2) RNA duplex. The dynamic binding methodology is a series of simulation technique to investigate small molecule-RNA interactions, where Amber 16 MD package[22] is utilized under implicit solvent model[23] (CG^{OBC}, 0.3 M salt concentrations). First, we determined initial bound conformations for

the CB096 and CB096$_T$ interacting with the 5'C$\underline{G}$G3' binding site by slowly moving the compounds to the RNA using a "reaction coordinate" defined as the distance between the center-of-mass (COM) of the heavy atoms of the closing base pairs and the COM of the heavy atoms of the CB096 (or CB096$_T$). Initially, the reaction coordinate between the compound and the closing base pairs was set to be 40 Å. The compound was then gradually moved towards the closing base pairs by 1 Å intervals using restraints imposed on the reaction coordinate until it became 0 Å. During this process, Watson-Crick (WC) base pairing, torsional and chirality restraints were imposed on the RNA except the loop region so that it maintained the native A-form conformation. Then, the compound was slowly moved away from the closing base pairs by 1 Å intervals until the distance reached to 40 Å. During this process, WC base pairing, torsional and chirality restraints were applied to all the RNA residues to transform the RNA back to its apo structure.

We repeated this forced binding process 50 times sequentially and obtained 50 different initial structures for the CB096/CB096$_T$ - 5'C$\underline{G}$G/3'G$\underline{G}$C binding site complex. We then conducted 50 independent implicit-solvent MD simulations starting from the conformations discovered in the previous step. No restraints were imposed on the loop residues so that the loop freely sampled the conformational space. WC base pairing, torsional, and chirality restraints, however, were imposed on the rest of the RNA residues to maintain the global A-form. One restraint was applied to the reaction coordinate in the MD simulations, so that the distance between the COM of the closing base pairs and COM of the compound did not exceed 10 Å and, thus, made the compound stay around the RNA loop while reorienting itself with respect to the force field to find the optimum bound state.

Each MD simulation was run for 120 ns yielding a total of 6 μs combined MD trajectory (60K snapshots). Cluster analyses were conducted by using an in-house code. Root-mean-square deviation (RMSD) was calculated throughout the trajectory and the snapshots with RMSD ≤ 1.0 Å were clustered into the same group. The binding free energies of the clusters having more than 100 snapshots were calculated using MM-PBSA method.[24] Supplemental Tables S3

and S4 display the results for CB096-5'C$\underline{G}$G/3'G$\underline{G}$C and CB096$_T$-5'C$\underline{G}$G/3'G$\underline{G}$C, respectively. The results indicate that CB096 has a better binding mode to the 5'C$\underline{G}$G/3'G$\underline{G}$C binding site compared to its tautomer CB096$_T$ with ~5 kcal/mol lower binding free energy (**Supplemental Tables S3 & S4**).

Supplementary Table S1. Atom names, types, and charges used to define CB096.

Atom Name	Atom Type	Charge	Atom Name	Atom Type	Charge
N1	no	0.844281	C17	ca	-0.205337
O2	o	-0.486904	C18	ca	-0.136146
O3	o	-0.486904	H19	ha	0.151091
C4	ca	-0.026529	C20	ca	-0.340963
C5	ca	-0.216291	H21	ha	0.177511
H6	ha	0.206435	C22	ca	0.316526
C7	ca	-0.291092	O23	oh	-0.646484
H8	ha	0.200664	H24	ho	0.473607
C9	ca	0.241924	C25	ca	0.284489
C10	ca	0.345253	C26	ca	-0.186112
C11	ca	-0.294583	H27	ha	0.134388
H12	ha	0.230853	O28	os	-0.279257
N13	n2	-0.640793	C29	c3	-0.078466
C14	c2	0.674209	H30	h1	0.091352
N15	na	-0.645754	H31	h1	0.091352
H16	hn	0.406328	H32	h1	0.091352

Supplementary Table S2. Atom names, types, and charges used to define CB096$_T$.

Atom Name	Atom Type	Charge	Atom Name	Atom Type	Charge
N1	no	0.862957	C17	ca	-0.199055
O2	o	-0.493264	C18	ca	-0.188256
O3	o	-0.493264	H19	ha	0.148808
C4	ca	-0.036346	C20	ca	-0.245241
C5	ca	-0.213894	H21	ha	0.195974
H6	ha	0.200056	C22	ca	0.314614
C7	ca	-0.267065	O23	oh	-0.580521
H8	ha	0.195480	H24	ho	0.420537
C9	ca	0.367402	C25	ca	0.137979
C10	ca	0.224604	C26	ca	-0.095830
C11	ca	-0.318666	H27	ha	0.116994
H12	ha	0.234547	O28	os	-0.281539
N13	na	-0.654439	C29	c3	-0.128854
H14	hn	0.423669	H30	h1	0.110359
C15	c2	0.684071	H31	h1	0.110359
N16	n2	-0.662535	H32	h1	0.110359

Supplementary Table S3. Results of the MM-PBSA calculations for CB096-5'C<u>G</u>G/3'G<u>G</u>C complex.

Cluster Name	$\Delta G_{MM\text{-}PBSA}$ (kcal/mol)			Total # of snapshots observed in combined trajectory
cluster_2087	-38.3098	±	3.6351	136
cluster_2083	-38.2729	±	3.3501	201
cluster_2086	-35.6766	±	3.5327	146
cluster_2082	-35.5177	±	3.3708	174
cluster_2085	-35.1116	±	3.7389	156
cluster_2088	-34.7147	±	3.1510	172
cluster_1309	-32.3591	±	2.9876	151
cluster_1306	-31.7013	±	3.1161	220
cluster_1323	-31.2877	±	2.8212	189
cluster_1319	-31.0363	±	2.9796	135
cluster_1314	-30.7889	±	3.3544	186
cluster_1331	-29.7572	±	3.0228	146
cluster_358	-28.5350	±	2.5924	171
cluster_375	-28.3780	±	2.6119	111
cluster_362	-27.8432	±	2.7535	289
cluster_369	-26.6504	±	2.3511	100
cluster_442	-26.5997	±	2.6541	129
cluster_359	-26.0381	±	3.1311	745
cluster_40	-25.8937	±	2.9074	144
cluster_977	-25.6091	±	3.1931	642
cluster_973	-25.6060	±	2.8887	523
cluster_368	-24.7888	±	3.1254	113
cluster_45	-24.6061	±	2.7683	155
cluster_990	-24.2917	±	3.2381	105
cluster_57	-24.1080	±	3.0285	119
cluster_62	-23.7731	±	3.1252	112
cluster_31	-23.0697	±	3.1359	121
cluster_24	-22.4502	±	2.6247	145
cluster_22	-22.2837	±	2.8815	118
cluster_158	-21.4862	±	2.6857	419
cluster_159	-21.1214	±	2.8140	273
cluster_164	-20.6947	±	2.8311	141
cluster_165	-20.5476	±	2.7262	349
cluster_386	-18.7385	±	3.3414	476
cluster_384	-18.2930	±	3.4650	325
cluster_383	-17.7997	±	3.3337	608
cluster_388	-17.6021	±	3.7131	119
cluster_389	-17.3088	±	3.5051	103
cluster_99	-16.9507	±	3.3292	190
cluster_101	-16.7598	±	3.2184	306
cluster_124	-16.5387	±	3.3764	261
cluster_459	-16.3805	±	3.4146	204
cluster_380	-15.9509	±	3.1245	127
cluster_385	-15.8978	±	3.2898	309
cluster_241	-15.1147	±	3.8866	144
cluster_114	-14.4825	±	4.1515	144
cluster_327	-14.4774	±	3.7118	146
cluster_108	-14.1982	±	3.3056	304
cluster_267	-14.1211	±	4.2036	164
cluster_95	-13.9526	±	3.1216	190
cluster_96	-13.9159	±	3.7154	226
cluster_519	-13.8698	±	4.9501	117
cluster_87	-13.8459	±	3.5874	278
cluster_102	-13.7604	±	3.1105	616
cluster_382	-13.4985	±	3.9416	158
cluster_266	-13.4506	±	5.6052	191
cluster_245	-13.2236	±	4.0775	205
cluster_236	-13.1579	±	4.7586	248
cluster_91	-12.9026	±	3.2068	194
cluster_94	-12.5219	±	3.2317	175
cluster_331	-12.1887	±	3.0228	118
cluster_97	-11.9813	±	3.7607	117
cluster_1363	-11.6895	±	3.2304	121
cluster_487	-10.1032	±	3.3089	425

Supplementary Table S4. Results of the MM-PBSA calculations for CB096$_T$-5'C$\underline{G}$G/3'G$\underline{G}$C complex.

Cluster Name	$\Delta G_{\text{MM-PBSA}}$ (kcal/mol)			Total # of snapshots observed in combined trajectory
cluster_2132	**-32.9371**	**±**	**3.3677**	**238**
cluster_2136	-32.5730	±	4.2201	242
cluster_934	-32.2999	±	3.3458	107
cluster_1538	-29.5342	±	2.8160	189
cluster_920	-29.2003	±	3.1827	121
cluster_825	-29.0788	±	3.3713	114
cluster_1175	-29.0423	±	2.4422	105
cluster_1537	-28.5862	±	2.6508	108
cluster_913	-28.4990	±	3.4795	114
cluster_134	-28.4708	±	2.6072	182
cluster_1176	-28.3592	±	2.4346	115
cluster_904	-28.3440	±	3.1245	110
cluster_136	-28.1889	±	2.6470	939
cluster_427	-27.8128	±	2.7866	140
cluster_1117	-27.7955	±	3.3753	101
cluster_341	-27.5654	±	2.5727	111
cluster_909	-27.2871	±	3.8306	231
cluster_167	-27.2593	±	3.0054	158
cluster_163	-27.0794	±	3.0109	105
cluster_138	-26.9226	±	2.5081	165
cluster_1115	-26.8378	±	4.2568	121
cluster_137	-26.7023	±	2.7429	368
cluster_145	-26.5099	±	2.4066	262
cluster_144	-26.4283	±	2.8507	594
cluster_146	-26.3119	±	2.9124	609
cluster_911	-26.2083	±	4.1131	131
cluster_416	-25.8326	±	3.1630	118
cluster_147	-25.6621	±	3.3245	141
cluster_142	-25.5717	±	3.0370	1145
cluster_158	-25.4473	±	2.5918	119
cluster_156	-25.4412	±	2.8547	102
cluster_135	-25.3932	±	2.5309	306
cluster_43	-25.0490	±	2.8737	103
cluster_509	-24.1601	±	3.0768	222
cluster_396	-24.0872	±	2.8728	149
cluster_169	-24.0144	±	2.8901	223
cluster_328	-23.8380	±	2.9049	144
cluster_165	-23.7224	±	2.8580	288
cluster_543	-23.5349	±	3.7203	112
cluster_830	-23.2613	±	2.6464	222
cluster_32	-23.2071	±	2.9203	116
cluster_833	-22.6317	±	2.5474	198
cluster_828	-22.5417	±	2.8628	106
cluster_327	-22.2657	±	2.6976	356
cluster_404	-22.0311	±	3.1319	120
cluster_860	-20.7424	±	2.8592	137
cluster_129	-16.2530	±	3.7666	145
cluster_205	-15.6768	±	3.0266	107
cluster_208	-15.5632	±	3.4594	139
cluster_110	-15.5286	±	2.9056	580
cluster_348	-15.0474	±	3.2987	114
cluster_139	-14.9313	±	2.9333	317
cluster_121	-14.8348	±	3.0479	653
cluster_197	-14.7984	±	3.3079	155
cluster_400	-14.6873	±	3.4565	140
cluster_209	-14.6722	±	3.6763	170
cluster_122	-14.5168	±	3.1401	698
cluster_100	-14.3993	±	3.0601	844
cluster_330	-14.2558	±	3.1549	287
cluster_149	-14.2484	±	3.3336	343
cluster_109	-14.1837	±	3.4238	101
cluster_563	-14.1219	±	3.3276	115
cluster_115	-13.8233	±	3.5485	493
cluster_107	-13.7813	±	3.2851	102

CELLULAR STUDIES

Cell Culture

HEK293T cells (CRL-3216) were purchased from ATCC and maintained in growth medium [Dulbecco's Modified Eagle Medium (DMEM; Corning) supplemented with 10% fetal bovine serum (FBS; Sigma Aldrich), 1% penicillin-streptomycin (P/S; Corning) and 1% glutaGRO supplement (Corning)] at 37 °C and 5% CO_2.

Cytotoxicity of CB096 in $(G_4C_2)_{66}$-transfected HEK293T cells

HEK293T cells (4×10^6 cells) were plated in a 100 mm diameter dish and grown overnight at 37 °C/5% CO_2. The cells were then transfected for at least 6 h with 9 µg of a plasmid encoding *r(G_4C_2)_{66}*-noATG-*GFP*[15] using JetPrime reagent (Polyplus) according to the vendor's procedure. Afterwards, the cells were washed twice with 1× DPBS (10 mL each), collected using trypsin-EDTA (0.25%) (Gibco), plated (8000 cells/20 µL) in a 384-well clear bottom white plate (Corning) in growth medium, and incubated overnight. In addition, reference wells without cells were also generated. The next day, 5 µL of growth medium containing vehicle (0.5% DMSO final concentration/well) or CB096 at varying concentrations (0.5% DMSO final concentration/well) was added, and the cells were incubated for 24 h. After the incubation period, 25 µL CellTiter-Glo reagent (Promega) was added to each well, and the samples were mixed for 5 min on an orbital shaker. After an additional incubation at room temperature for 10 min (protected from light), the luminescent signal was recorded using the Molecular Devices SpectraMax M5 plate reader. The reported data were generated by subtracting the luminescence signal of the reference wells (without cells) and normalized to signal generated from the vehicle-treated samples. The reported results from two independent experiments each performed in triplicate.

Cell proliferation assay in (G$_4$C$_2$)$_{66}$-transfected HEK293T cells

HEK293T cells (~60-70% confluency) in 60 mm diameter dishes were transfected for at least 6 h with 4.5 µg of a plasmid encoding r(G$_4$C$_2$)$_{66}$[15] using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's procedure. Afterwards, cells were washed twice with 1× DPBS (5 mL each), collected using Trypsin-EDTA (0.25%) solution, and plated in a clear 96-well plates. Each well contained 10,000 HEK293T cells in a final volume of 90 µL of growth medium. After overnight incubation, the cells were treated with 10 µL growth medium containing vehicle (0.5% DMSO final concentration/well) or CB096 (25 µM final concentration in 0.5% DMSO/well) and incubated for an addition 24 h at 37 °C. Background absorption at 450 nm and 600 nm (reference wavelength) was measured with a Molecular Devices SpectraMax M5 plate reader. Then, 10 µL of the WST-1 reagent (Sigma Aldrich) was added, and the absorption of the formazan product formed was measured at 450 nm as well as at the reference wavelength of 600 nm (reference wavelength). Background measured prior to addition of WST-1 was then subtracted from all samples. An assessment of cell proliferation was generated by subtracting the reference signal at 600 nm from the absorbance readout at 450 nm and were reported relative to the vehicle-treated samples. The reported results are from two independent experiments, each performed in triplicate.

Inhibition of RAN translation in r(G$_4$C$_2$)$_{66}$-transfected HEK293T cells

Approximately 4×10^6 HEK293T cells were plated in 100 mm diameter cell culture dishes overnight and co-transfected the following day with 9 µg of a plasmid encoding (G$_4$C$_2$)$_{66}$–No ATG–GFP[15] and 1 µg of a plasmid encoding mCherry (for normalization) using JetPrime transfection reagent (Polyplus) according to the manufacturer's instructions. After at least 4 h, cells were washed with 1× DPBS, trypsinized, and plated (8000 cells/well) in 20 µL of growth medium into 384-well clear bottom black plates. After at least 2 h incubation, the compound of interest at the indicated

concentration (final concentration of 0.5% DMSO) or vehicle (0.5 % DMSO) was added in 5 µL growth medium. After overnight incubation, the growth medium was removed, and the cells were lysed in 100 mM potassium phosphate lysis buffer, pH 7.8, and 0.2% (v/v) Triton X-100. Fluorescence was measured using a BioTek FLx800 with 530/25 nm (excitation) and 590/35 nm (emission) filters for mCherry and 485/20 nm (excitation) and 528/20 nm (emission) filters for GFP as previously reported.[15] Background was determined by measuring the corresponding fluorescence intensity of untransfected cells. The background-corrected ratio of GFP to mCherry was used to determine the biological effect of the compounds on RAN translation relative to the vehicle treated samples. The reported results are from two independent experiments performed in triplicate. The effect of compound on canonical translation was completed analogously except a plasmid encoding GFP was used instead of $(G_4C_2)_{66}$–*No ATG*–GFP.

Alternatively HEK293T cells stably expressing Lentiviral-S-tdTomato (Addgene Plasmid #112579)[25] were seeded into 10 cm diameter dishes and grown to ~80% confluency, and then co-transfected with plasmids encoding $(G_4C_2)_{66}$-*No ATG*-firefly luciferase, Renilla luciferase, and *ATG*-GFP using Invitrogen Lipofectamine 3000 following manufacturer's protocol. Following transfection, cells were seeded into 384-well plates at a density of 0.4×10^6 cells/mL and allowed to adhere overnight. Compound was then added for 24 h with a final concentration of <1% DMSO. Luciferase activity was measured using Nano-Glo® Dual-Luciferase® Reporter Assay System (Promega) following the manufacturer's recommendations. Luminescence was measured by a Molecular Devices SpectraMax M5 plates reader and integration time of 500 ms.

Determination of r(G_4C_2)$_{66}$–No ATG–GFP levels via RT-qPCR

Transfection of HEK293T cells was performed as described above. Approximately 5 h post-transfection, the cells were collected and seeded into 24-well plates (125,000 cells per well in 0.8 mL of growth medium). After at least 4 h, vehicle (0.5% DMSO final concentration/well) or CB096 (0.5% DMSO final concentration/well) were added in a final volume of 200 µL, and the cells were

incubated for 24 h. After the incubation period, the medium was removed, and total RNA was extracted using a Quick-RNA Miniprep Kit (Zymo Research) according to the manufacturer's protocol. Approximately 0.2 µg of total RNA, as determined by its absorbance at 260 nm using a Nanodrop UV spectrophotometer, was used for reverse transcription using qScript Kit (Quantbio) according to the manufacturer's protocol. RT-qPCR was performed on 4 ng cDNA on a 7900HT Fast Real Time PCR System (Applied Biosystem) using Power SYBR Green Master Mix (Applied Biosystems) as previously reported.[15]   Expression levels of mRNAs were normalized to 18S.

Rescue of tdTomato Mislocalization

HEK293T cells that stably express Lentiviral-S-tdTomato[25] were seeded into 10 cm diameter dishes and grown to ~80% confluency. They were then co-transfected with plasmids encoding $(G_4C_2)_{66}$-*No ATG*-firefly luciferase, Renilla luciferase, and *ATG*-GFP using Lipofectamine 3000 following manufacturer's protocol. Following transfection, cells were seeded into 24-well glass bottom plates at a density of 0.4×10^6 cells/well and incubated overnight at 37 °C. The following day fresh growth medium containing compound or vehicle was added to each well. As a positive control, cells were treated with 10 µM KPT-350 (KPT). After a 24 h incubation with compound, the cells were fixed with 4% paraformaldehyde prepared in $1\times$ DPBS, and the nuclei were stained with DAPI. Cells were imaged by confocal microscopy at 554/581 nm (tdTomato), 488/510 nm (GFP), and 358/461 nm (DAPI). The ratio of tdTomato in the nucleus and cytoplasm was quantified from three biological replicates; five images were taken from each replicate and 200 cells were evaluated per image. Statistics were measured using a one-way ANOVA test.

Rescue of Stress Granule Formation

HEK293T cells that stably express Lentiviral-S-tdTomato were maintained in growth medium and grown to ~80% confluency in 10 cm diameter dishes. Two types of stress granules were studied:

those induced by $r(G_4C_2)^{exp}$ expression and those chemically induced with sodium arsenite. To study the former, the aforementioned HEK293T cells were were transfected with 4.3 μg of a plasmid encoding $(G_4C_2)_{66}$-*No ATG*-GFP using Lipofectamine 3000 transfection agent per the manufacturer's protocol. To induce the latter, the aforementioned HEK293T cells were transfected with 4.3 μg of a plasmid encoding *ATG*-GFP (akin to $(G_4C_2)_{66}$-*No ATG*-GFP) followed by treatment with 0.5 mM sodium arsenite treatment for 1 h at 37 °C, as previously described.[26]

To determine if compounds could rescue nuclear pore dysfunction, following transfection the cells were seeded into 24-well glass bottom plates at a density of 0.4×10^6 cells/well and incubated overnight at 37 °C. The next day fresh growth medium containing compound or vehicle was added to each well, and the samples were incubated for 24 h at 37 °C. After the treatment period, the cells were fixed in 4% paraformaldehyde prepared in 1$\times$ DPBS for 20 min and then permeabilized by treatment with 0.1% (v/v) Triton X-100 in 1$\times$ DPBS for 10 min, both steps at room temperature. The cells were washed three times with 1$\times$ DPBS, followed by blocking with 1$\times$ TBST supplemented with 5% goat serum for 1 h at room temperature. After, the cells were incubated with Rabbit anti-ATXN2/SCA2 (1:200 dilution; Bethyl Laboratories; Catalog #: A301-118A) in 1$\times$ TBST supplemented with 5% goat serum for 16 h at 4 °C (with shaking). The cells were washed three times with 1$\times$ TBST for 10 min each at room temperature, and the secondary antibody [Goat anti-Rabbit IgG (H+L) Secondary Antibody, DyLight 650; 1:200 dilution; ThermoFisher Scientific; Catalog #: 84546] was added in 1$\times$ TBST supplemented with 5% goat serum for 1 h at room temperature (with shaking). The cells were then washed three times with 1× TBST, stained with 1 μg/ mL DAPI in 1$\times$ PBS for 10 min, and then imaged by confocal microscopy at 554/581 nm (tdTomato), 488/510 nm (GFP), 358/461 nm (DAPI), and 633/649 nm (DyLight 650).

SYNTHETIC METHODS

General Synthetic Methods

All reagents and solvents used for chemical synthesis were purchased from commercial sources (Sigma Aldrich and Alfa Aesar) and used without further purification unless mentioned otherwise. DMSO-d6, MeOD and D_2O solvents for NMR experiments were acquired from Cambridge Isotope Labs and used as received.

Instrumentation

^{1}H- and ^{13}C-NMR spectra were acquired on a 400 MHz Bruker Avance spectrometer and a Bruker 600 MHz spectrometer equipped with a cryoprobe. Chemical shifts (δ) are reported in ppm relative to tetramethylsilane (TMS) or the respective NMR solvent; coupling constants (J) are in Hertz (Hz). Abbreviations used are s, singlet; bs, broad singlet; d, doublet; dd, doublet of doublets; t, triplet; dt, doublet of triplets; td, triplet of doublets; tt, triplet of triplets; bt, broad triplet; q, quartet; m, multiplet; and bm, broad multiplet.

Mass spectra were recorded on a Varian 500-MS IT mass spectrometer.

General procedure for the synthesis of CB096-derivatives (compounds 2-10, 13-17)

Experimental procedure for the synthesis of derivatives 2-10 and 13-17 was slightly modified from previous reports, $^{27\text{-}30}$ *e.g.* the reaction temperature was set to ~100 °C instead of 65 °C.
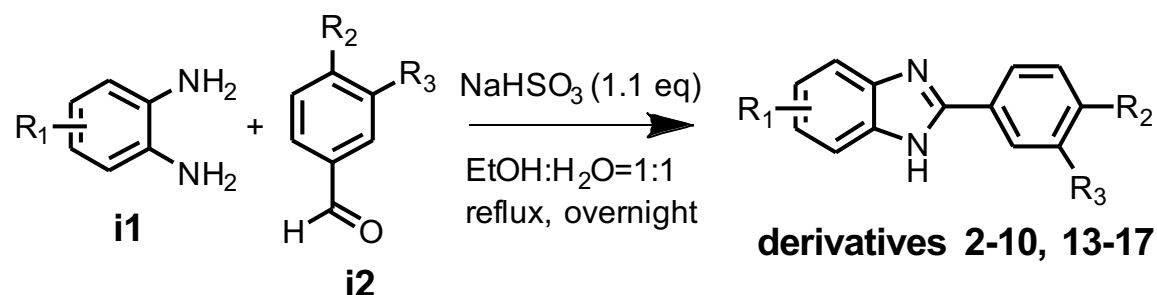
Scheme 1. General procedure for the synthesis of CB096-like derivatives 2-10, 13-17.

To a 0 °C cooled solution of **i2** (4-hydroxy-3-methoxybenzaldehyde; R_2 = OH, R_3 = OMe; 154.14 mg, 1 mmol, 1 eq) in EtOH (2.5 mL) was added dropwise a solution of sodium bisulfite (115 mg, 1.1 mmol, 1.1 eq) in H_2O (2.5 mL) and the resulting mixture was stirred at 0 °C for 30 min. **I1** (o-phenylenediamine; R_1 = -H; 108 mg, 1 mmol, 1 eq) in 1:1 EtOH : H_2O (2.5 mL) was added dropwise and the mixture was stirred at reflux overnight. The volatiles were removed under reduced pressure and the residue was adsorbed onto silica and purified via silica gel chromatography (0 – 100 % EtOAc in Hexanes) to give 4-(1H-benzo[d]imidazol-2-yl)-2-methoxyphenol (**1**, 129.6 mg, 0.54 mmol, 54%).

Compounds 2-10 and 13-17 were obtained in yields varying from 35 to 77% and were also characterized in previous reports.[29-31,32,33,34,28,35] Compounds 2-10 and 13-17 obtained by the procedure herein were fully characterized by means of ^{1}H and ^{13}C-NMR as well as high-resolution mass spectrometry (HRMS).

2-methoxy-4-(4-nitro-1H-benzo[d]imidazol-2-yl)phenol (**compound 2**)
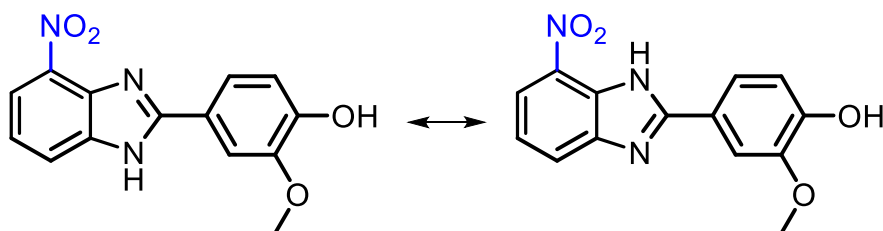
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.41 (s, 1H), 12.93 (s, 1H), 9.73 (d, J = 34.7 Hz, 2H), 7.90 (m, 8H), 7.39 (t, J = 8.0 Hz, 6H), 6.93 (d, J = 8.3 Hz, 6H), 3.92 (s, 6H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 155.60, 155.30, 149.59, 149.37, 147.94, 147.69, 146.80, 138.09, 137.68, 137.54, 132.89, 129.23, 126.03, 121.68, 121.43, 121.22, 120.91, 120.07, 119.99, 118.27, 118.14, 117.61, 115.91, 115.58, 111.66, 110.80, 55.90, 55.79.

HRMS: Calcd. for $C_{14}H_{12}N_3O_4$ 286.0828 [M+H$^+$]; Found: 286.0830.

2-methoxy-4-(5-methyl-1H-benzo[d]imidazol-2-yl)phenol (**compound 3**)
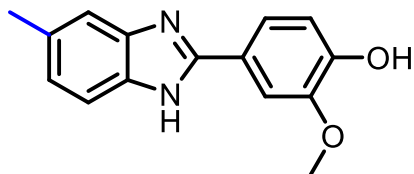
^{1}H NMR (400 MHz, DMSO-d6) δ: 12.65 (s, 1H), 9.53 (s, 1H), 7.75 (t, J = 3.6 Hz, 1H), 7.61 (dd, J = 8.2, 1.9 Hz, 2H), 7.48 (s, 1H), 7.15 (d, J = 3.9 Hz, 2H), 6.91 (d, J = 8.2 Hz, 1H), 3.89 (s, 3H), 3.35 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 151.82, 148.45, 147.85, 121.44, 119.64, 115.68, 110.32, 55.67.

HRMS: Calcd. for $C_{15}H_{15}N_2O_2$ 255.1134 [M+H$^+$]; Found: 255.1138.

2-methoxy-4-(5-methoxy-1H-benzo[d]imidazol-2-yl)phenol **(compound 4)**
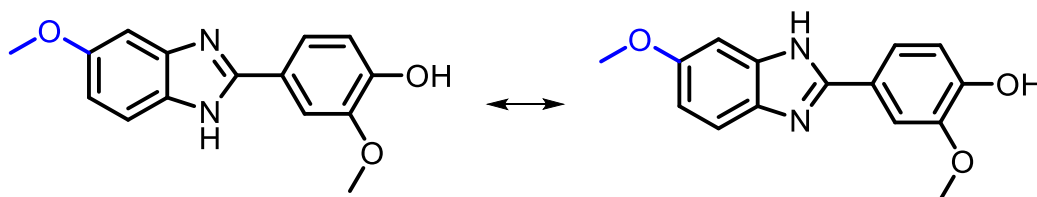
^{1}H NMR (400 MHz, DMSO-d6) δ: 12.49 (s, 2H), 9.48 (d, J = 12.2 Hz, 2H), 7.70 (d, J = 6.8 Hz, 2H), 7.56 (t, J = 9.2 Hz, 2H), 7.48 (d, J = 8.6 Hz, 1H), 7.35 (d, J = 7.8 Hz, 1H), 7.15 (s, 1H), 6.95 (s, 1H), 6.90 (d, J = 8.2 Hz, 2H), 6.79 (t, J = 6.8 Hz, 2H), 3.88 (s, 6H), 3.79 (s, 6H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 155.70, 155.25, 152.13, 151.01, 148.28, 148.11, 144.74, 138.38, 135.58, 129.40, 121.72, 121.61, 119.43, 119.20, 118.80, 115.66, 111.31, 111.10, 110.50, 110.17, 110.02, 101.20, 94.40, 55.84, 55.66, 55.47, 55.40.

HRMS: Calcd. for $C_{15}H_{15}N_2O_3$ 271.1083 [M+H$^+$]; Found: 271.1086.

4-(5-fluoro-1H-benzo[d]imidazol-2-yl)-2-methoxyphenol (**compound 5**)
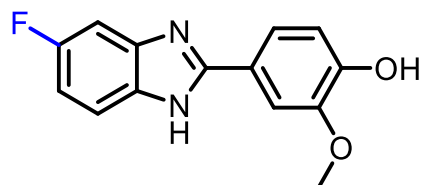
^{1}H NMR (400 MHz, DMSO-d6) δ: 12.75 (s, 1H), 9.58 (s, 1H), 7.73 (d, J = 2.0 Hz, 1H), 7.60 (dd, J = 8.2, 2.0 Hz, 1H), 7.52 (dd, J = 8.4, 4.9 Hz, 1H), 7.34 (d, J = 8.2 Hz, 1H), 7.06 – 6.96 (m, 1H), 6.92 (d, J = 8.2 Hz, 1H), 3.89 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 159.38, 159.19, 157.72 (d, J = 23.5 Hz), 156.40, 153.77, 153.26 (d, J = 121.2 Hz), 148.69, 131.66, 121.15, 119.77, 119.12, 115.74, 114.43 (d, J = 9.6 Hz), 111.48, 110.33, 109.84, 109.67, 109.43, 109.27, 103.99, 103.84, 97.53, 97.35, 55.69.

HRMS: Calcd. for $C_{14}H_{12}FN_2O_2$ 259.0883 [M+H$^+$]; Found: 259.0885.

4-(5-chloro-1H-benzo[d]imidazol-2-yl)-2-methoxyphenol (**compound 6**)
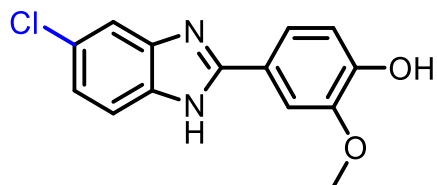
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.10 (s, 1H), 9.67 (s, 1H), 7.95 (s, 1H), 7.77 (s, 2H), 7.66 (dd, *J* = 8.3, 1.7 Hz, 1H), 7.48 (d, *J* = 8.4 Hz, 1H), 6.95 (dd, *J* = 8.2, 0.7 Hz, 1H), 3.90 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 153.22, 148.91, 147.91, 126.05, 121.96, 120.71, 119.98, 115.76, 110.45, 55.72.

HRMS: Calcd. for $C_{14}H_{12}{}^{35}ClN_2O_2$ 275.0587 [M+H$^+$]; Found: 275.0589.

2-methoxy-4-(5-(methylsulfonyl)-1H-benzo[d]imidazol-2-yl)phenol (**compound 7**)
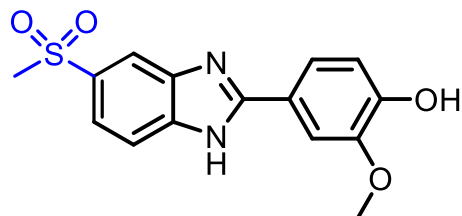
^{1}H NMR (400 MHz, MeOD) δ: 8.14 (s, 1H), 7.80 (dd, J = 8.5, 1.7 Hz, 1H), 7.73 (t, J = 4.1 Hz, 2H), 7.59 (dd, J = 8.3, 2.1 Hz, 1H), 6.94 (dd, J = 7.8, 4.1 Hz, 1H), 3.98 (s, 3H), 3.15 (s, 3H).

^{13}C NMR (151 MHz, MeOD) δ: 157.40, 150.99, 149.54, 135.68, 122.47, 121.76, 121.43, 116.76, 111.43, 56.49, 45.04.

HRMS: Calcd. for $C_{15}H_{15}N_2O_4S$ 319.0753 [M+H$^+$]; Found: 319.0756.

2-methoxy-4-(5-(trifluoromethyl)-1H-benzo[d]imidazol-2-yl)phenol **(compound 8)**
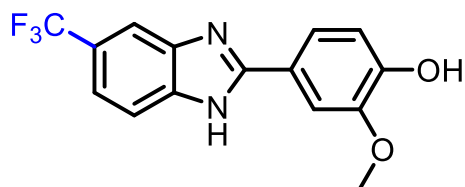
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.11 (s, 1H), 9.69 (d, J = 6.7 Hz, 1H), 7.96 (s, 1H), 7.79 (d, J = 12.8 Hz, 2H), 7.72 – 7.63 (m, 2H), 7.53 – 7.45 (m, 1H), 6.95 (d, J = 8.2 Hz, 1H), 3.91 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 154.49, 149.16, 147.96, 126.07, 124.27, 120.55, 120.23, 118.49, 115.81, 110.59, 55.73.

36

HRMS: Calcd. for $C_{15}H_{12}F_3N_2O_2$ 309.0851 [M+H$^+$]; Found: 309.0854.

2-(4-hydroxy-3-methoxyphenyl)-1H-benzo[d]imidazole-5-carboxylic acid (**compound 9**)
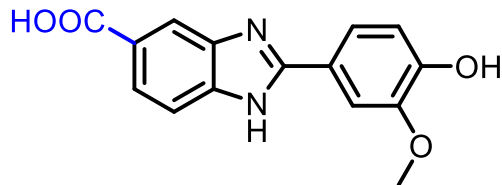
^{1}H NMR (400 MHz, DMSO-d6) δ: 12.97 (s, 1H), 12.66 (s, 1H), 9.64 (s, 1H), 8.12 (s, 1H), 7.80 (d, J = 8.3 Hz, 1H), 7.76 (d, J = 2.0 Hz, 1H), 7.64 (dd, J = 8.2, 2.0 Hz, 2H), 6.94 (d, J = 8.3 Hz, 1H), 3.90 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 167.92, 149.02, 147.93, 124.15, 120.74, 120.11, 115.77, 110.53, 55.73.

HRMS: Calcd. for $C_{15}H_{13}N_2O_4$ 285.0875 [M+H$^+$]; Found: 285.0879.

2-(4-hydroxy-3-methoxyphenyl)-1H-benzo[d]imidazole-5-carbonitrile **(compound 10)**
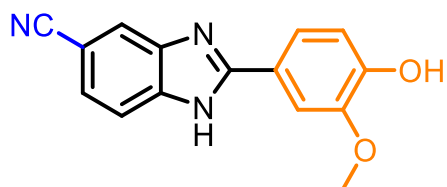
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.21 (s, 1H), 9.74 (s, 1H), 8.06 (s, 1H), 7.76 (d, *J* = 1.8 Hz, 1H), 7.66 (dd, *J* = 8.3, 1.7 Hz, 2H), 7.55 (dd, *J* = 8.3, 1.4 Hz, 1H), 6.94 (d, *J* = 8.3 Hz, 1H), 3.89 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 149.38, 148.01, 125.48, 120.43, 120.31, 120.25, 115.87, 110.67, 103.54, 55.77.

HRMS: Calcd. for $C_{15}H_{12}N_3O_2$ 266.0930 [M+H$^+$]; Found: 266.0933.

4-(5-nitro-1H-benzo[d]imidazol-2-yl)phenol (**compound 13**)
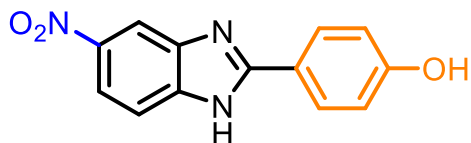
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.36 (s, 1H), 10.20 (s, 1H), 8.46 (s, 1H), 8.07 (dd, J = 18.4, 8.2 Hz, 3H), 7.67 (s, 1H), 6.95 (d, J = 8.7 Hz, 2H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 160.15, 157.10, 155.95, 148.89, 143.32, 142.62, 142.18, 139.93, 134.33, 128.89, 119.91, 118.30, 117.90, 117.63, 116.01, 114.26, 111.33, 107.41.

HRMS: Calcd. for $C_{13}H_{10}N_3O_3$ 256.0722 [M+H$^+$]; Found: 256.0726.

2,6-dimethoxy-4-(5-nitro-1H-benzo[d]imidazol-2-yl)phenol (**compound 14**)
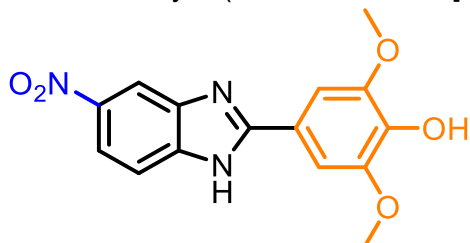
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.38 (s, 1H), 9.14 (d, J = 22.0 Hz, 1H), 8.41 (d, 1H), 8.11 (t, J = 8.1 Hz, 1H), 7.73 (dd, J = 36.2, 8.8 Hz, 1H), 7.52 (s, 2H), 3.90 (s, 6H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 157.18, 155.98, 148.76, 148.24, 143.21, 142.64, 142.19, 139.89, 138.75, 138.47, 134.33, 118.90, 118.81, 118.28, 117.93, 117.57, 114.23, 111.27, 107.32, 104.75, 104.53, 56.17.

HRMS: Calcd. for $C_{15}H_{14}N_3O_5$ 316.0933 [M+H$^+$]; Found: 316.0937.

2-(benzo[d][1,3]dioxol-5-yl)-5-nitro-1H-benzo[d]imidazole **(compound 15)**
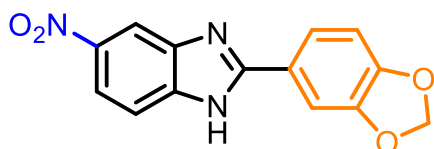
^{1}H NMR (400 MHz, DMSO-d6) δ: 8.41 (d, J = 1.8 Hz, 1H), 8.10 (dd, J = 8.9, 2.2 Hz, 1H), 7.76 (dd, J = 8.2, 1.7 Hz, 1H), 7.73 – 7.67 (m, 2H), 7.13 (d, J = 8.2 Hz, 1H), 6.15 (s, 2H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 155.82, 149.74, 148.09, 142.60, 123.07, 121.95, 117.93, 108.98, 106.84, 101.94.

HRMS: Calcd. for $C_{14}H_{10}N_3O_4$ 284.0671 [M+H$^+$]; Found: 284.0675.

2-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-5-nitro-1H-benzo[d]imidazole **(compound 16)**
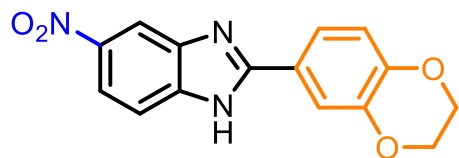
^{1}H NMR (400 MHz, DMSO-d6) δ: 13.42 (s, 1H), 8.47 (s, 1H), 8.09 (d, *J* = 8.0 Hz, 1H), 7.70 (d, *J* = 6.4 Hz, 3H), 7.06 (d, *J* = 9.0 Hz, 1H), 4.33 (s, 4H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 143.79, 122.18, 120.50, 117.92, 115.73, 64.48, 64.19.

HRMS: Calcd. for $C_{15}H_{12}N_3O_4$ 298.0828 [M+H$^+$]; Found: 298.0830.

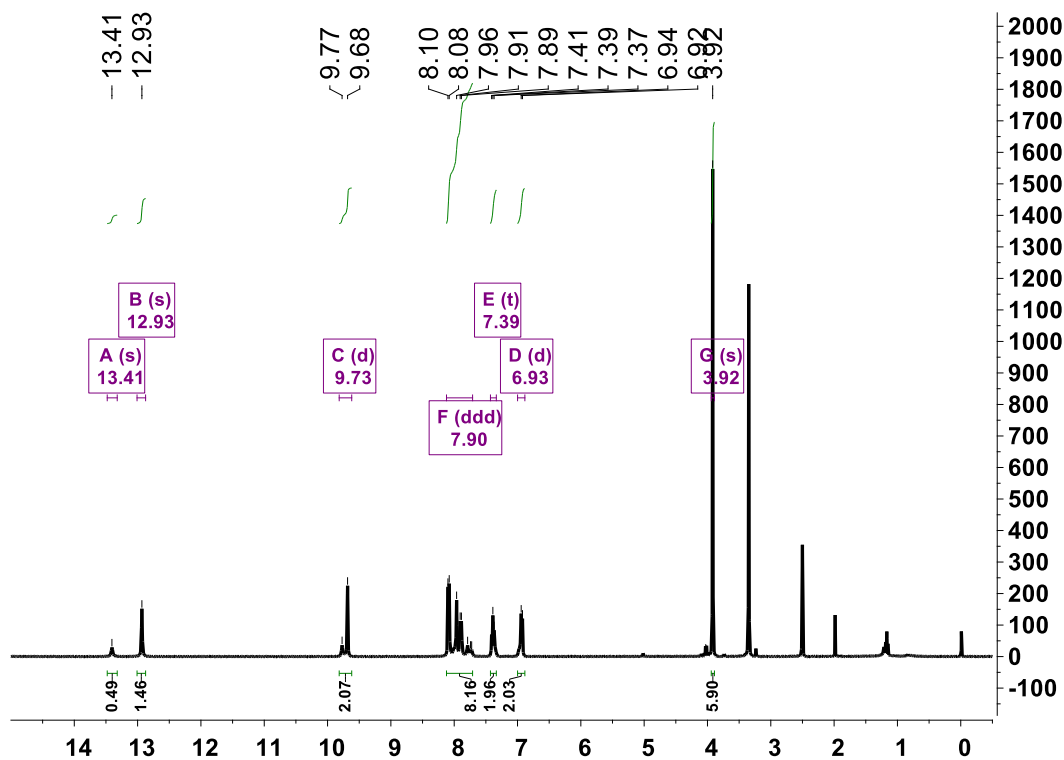5-nitro-2-(3,4,5-trimethoxyphenyl)-1H-benzo[d]imidazole **(compound 17)**
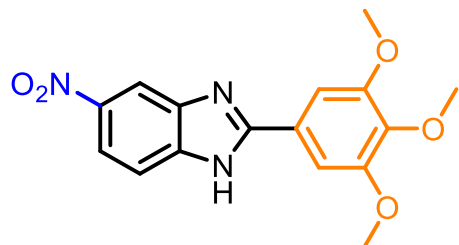
^{1}H NMR (400 MHz, DMSO-d6) δ: 8.47 (d, J = 2.1 Hz, 1H), 8.13 (dd, J = 8.9, 2.2 Hz, 1H), 7.78 (d, J = 8.9 Hz, 1H), 7.57 (s, 2H), 3.92 (s, 6H), 3.76 (s, 3H).

^{13}C NMR (151 MHz, DMSO-d6) δ: 155.66, 153.34, 142.78, 139.91, 123.95, 118.13, 104.49, 60.23, 56.15.

HRMS: Calcd. for $C_{16}H_{16}N_3O_5$ 330.1090 [M+H$^+$]; Found: 330.1095.
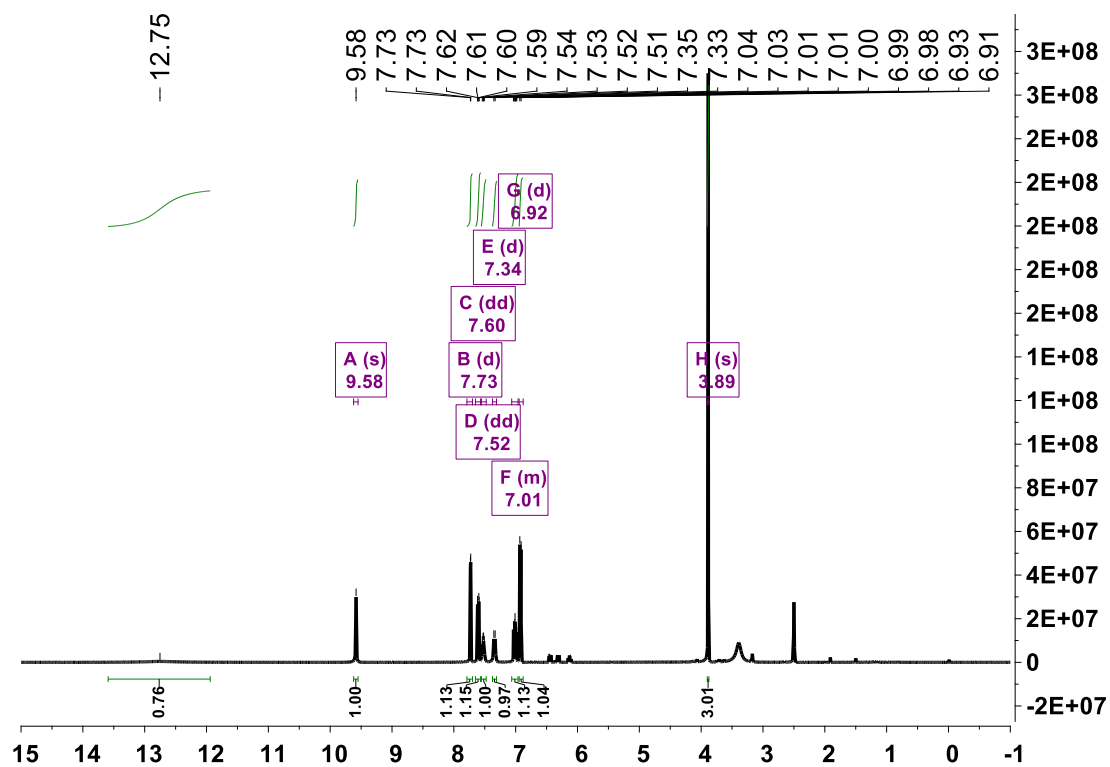
COMPOUND CHARACTERIZATION

Compound 2

X = EtOAc

Compound 3

X = EtOAc

A (s)
12.65

B (s)
9.53

C (t)
7.75

D (dd)
7.61

E (s)
7.48

F (d)
7.15

G (d)
6.91

H (s)
3.89

I (s)
3.35

12.65

9.53

7.75
7.74
7.62
7.60
7.60
7.16
7.15
6.92
6.90
3.35

X

X

X

1.00

1.04

0.97
1.86
0.98
1.95
1.00

2.83
3.13

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2

X = EtOAc

151.82
148.45
147.85

121.44
119.64
115.68
110.32

55.67

X

X

X X

9E+07
8E+07
8E+07
8E+07
7E+07
6E+07
6E+07
5E+07
4E+07
4E+07
4E+07
3E+07
2E+07
2E+07
2E+07
1E+07
5E+06
0
-5E+06

200 170 140 110 80 60 40 20 0

Compound 4

X = EtOAc

X = EtOAc

Compound 5

^{1}H NMR peaks: 12.75, 9.58, 7.73, 7.73, 7.62, 7.61, 7.60, 7.59, 7.54, 7.53, 7.52, 7.51, 7.35, 7.33, 7.04, 7.03, 7.01, 7.01, 7.00, 6.99, 6.98, 6.93, 6.91

A (s) 9.58
B (d) 7.73
C (dd) 7.60
D (dd) 7.52
E (d) 7.34
F (m) 7.01
G (d) 6.92
H (s) 3.89

Integrations: 0.76, 1.00, 1.13, 1.15, 1.00, 0.97, 1.13, 1.04, 3.01

^{13}C NMR peaks: 159.38, 159.19, 157.80, 157.65, 153.66, 152.86, 148.69, 144.39, 144.32, 140.55, 136.83, 136.76, 131.66, 130.93, 121.15, 119.77, 119.12, 115.74, 114.46, 114.40, 111.48, 110.33, 109.84, 109.67, 109.43, 109.27, 103.99, 103.84, 100.77, 97.53, 97.35, 55.69

Compound 6

Compound 7

¹H NMR (400 MHz, MeOD) δ 8.14 (s, 1H), 7.80 (dd, J = 8.5, 1.7 Hz, 1H), 7.73 (t, J = 4.1 Hz, 2H), 7.59 (dd, J = 8.3, 2.1 Hz, 1H), 6.94 (dd, J = 7.8, 4.1 Hz, 1H), 3.98 (s, 3H), 3.15 (s, 3H).

X = EtOAc

Compound 8

¹H NMR (400 MHz, DMSO) 13.11 (s, 4H), 9.69 (d, J = 6.7 Hz, 5H), 7.96 (s, 3H), 7.79 (d, 12.8 Hz, 9H), 7.72 – 7.63 (m, 8H), 7.53 – 7.45 (m, 5H), 6.95 (d, 8.2 Hz, 5H), 3.91 (s, 15H).

X = EtOAc

X = EtOAc

Compound 9

^{1}H NMR peaks: 12.97, 12.66, 9.64, 7.81, 7.79, 7.76, 7.75, 7.65, 7.65, 7.63, 7.63, 6.95, 6.93, 3.90

- F (s) 12.97
- G (s) 12.66
- B (s) 9.64
- H (s) 8.12
- C (d) 7.80
- D (d) 7.76
- I (dd) 7.64
- A (d) 6.94
- E (s) 3.90

Integrations: 0.89, 0.94, 0.92, 0.91, 0.92, 1.01, 1.88, 1.00, 2.98

^{13}C NMR peaks: 167.92, 149.02, 147.93, 124.15, 120.74, 120.11, 115.77, 110.53, 55.73

Compound 10

Compound 13

¹H NMR peaks:
- 13.36
- 10.20
- 8.46
- 8.10
- 8.08
- 8.06
- 8.04
- 7.67
- 6.96
- 6.94

Assignments:
- C (s) 13.36
- B (s) 10.20
- E (dd) 8.07
- D (s) 8.46
- A (d) 6.95
- F (s) 7.67

Integrations:
- 0.84
- 0.94
- 0.89
- 2.99
- 0.99
- 2.00

¹³C NMR peaks:
- 160.15
- 157.10
- 155.95
- 148.89
- 143.32
- 142.62
- 142.18
- 139.93
- 134.33
- 128.89
- 119.91
- 118.30
- 117.90
- 117.63
- 116.01
- 114.26
- 111.33
- 107.41
- 48.67
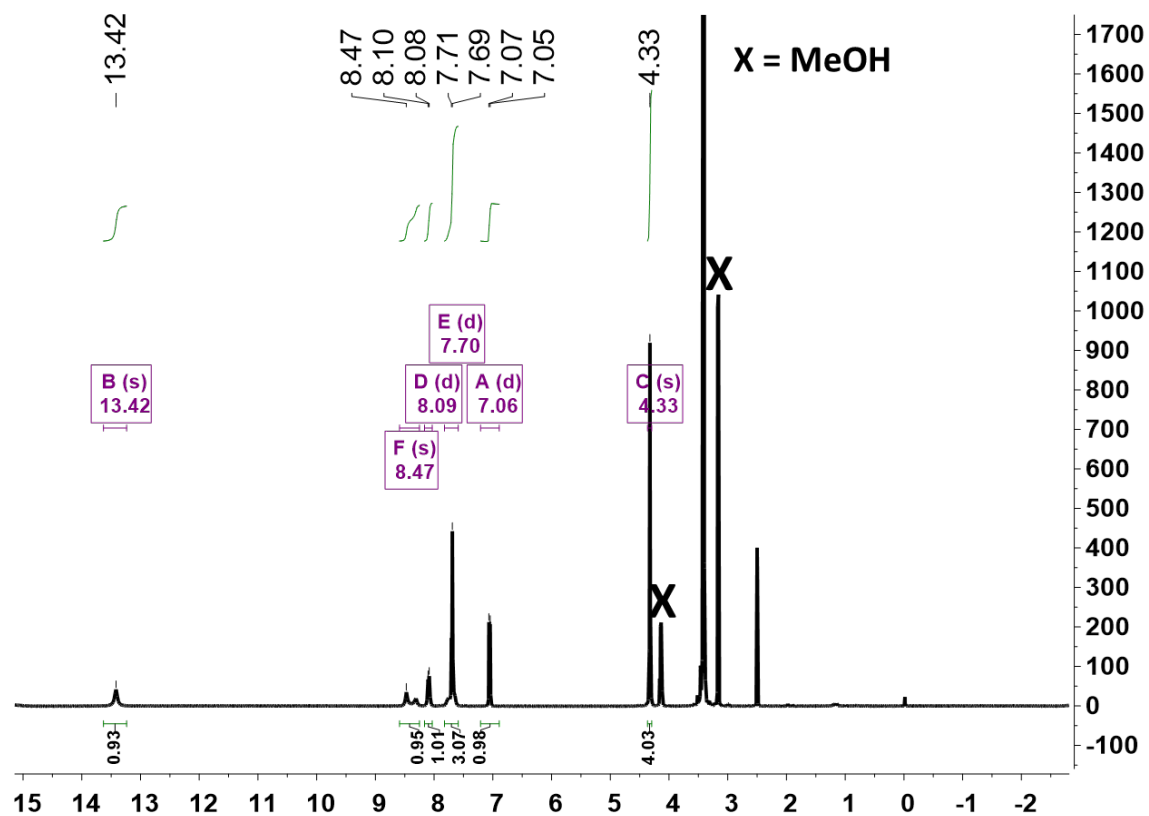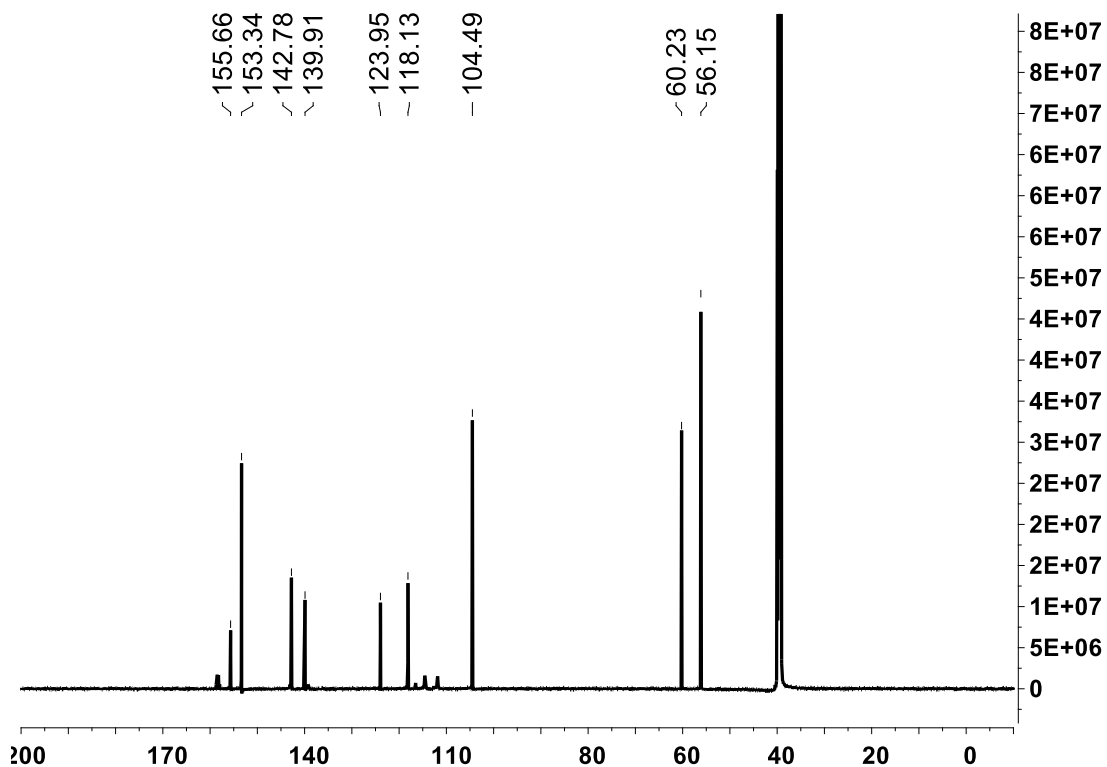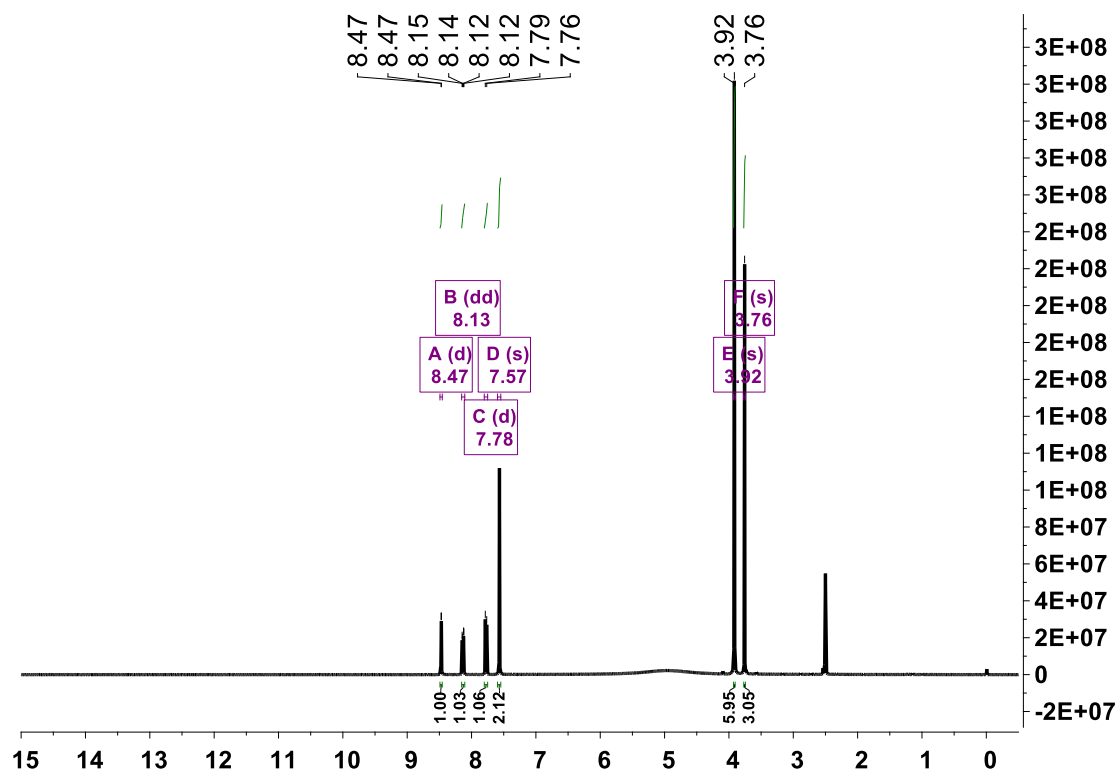
X = MeOH

49

Compound 14

Compound 15

Compound 16
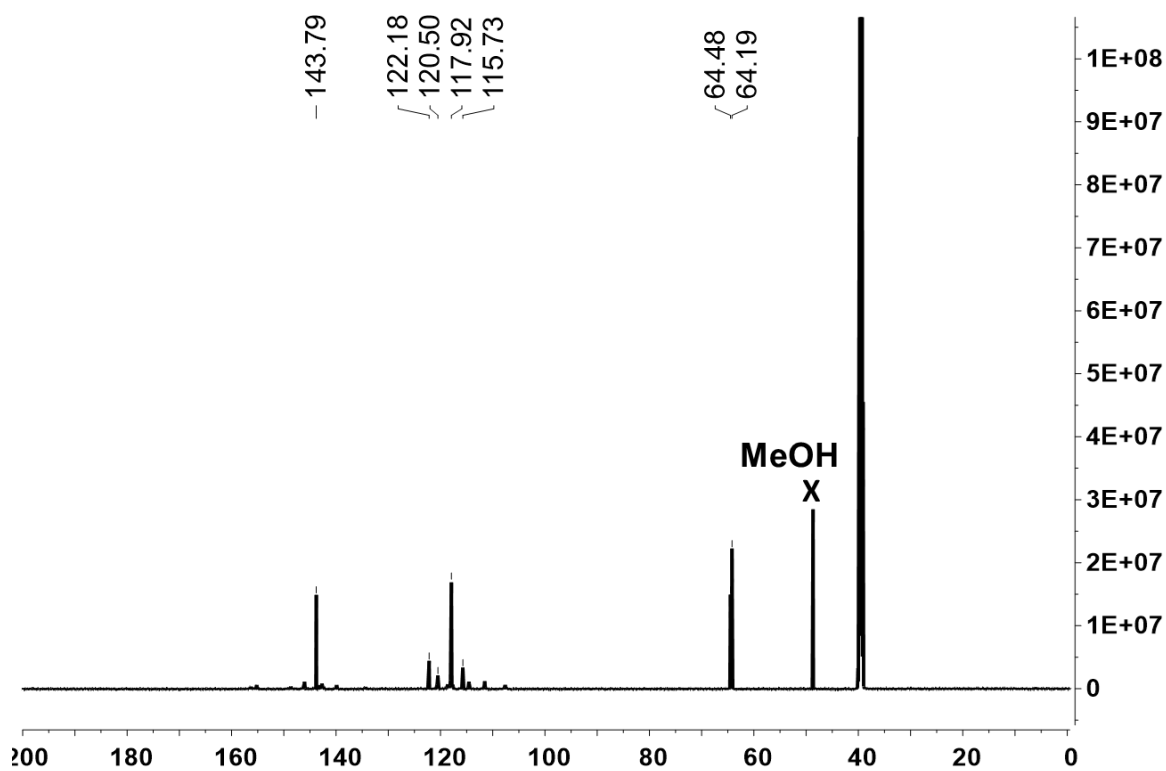
Compound 17

REFERENCES

1. Bellaousov, S., Reuter, J. S., Seetin, M. G., and Mathews, D. H. (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis, *Nucleic Acids Res. 41*, W471-W474.

2. Mathews, D. H. (2014) RNA secondary structure analysis using RNAstructure, *Curr. Protoc. Bioinformatics 46*, 1-25.

3. Haniff, H. S., Graves, A., and Disney, M. D. (2018) Selective small molecule recognition of RNA base pairs, *ACS Comb. Sci. 20*, 482-491.

4. Case, D. A., Betz, R. M., Botello-Smith, W., Cerutti, D. S.,Cheatham, T. E. I., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H.,Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko,A., Lee, T. S., LeGrand, S., Li, C., Lin, T., Luchko, T., Luo, R., Madej, B.,Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Nguyen, H. T.,Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Sagui, C., Simmerling,C. L., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L.,York, D. M., and Kollman, P. A. . (2016) AMBER 16, University of California, San Francisco, San Francisco.

5. Joung, I. S., and Cheatham, T. E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated bomolecular simulations, *J. Phys. Chem. B 112*, 9020-9041.

6. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926-935.

7. Yildirim, I., Stern, H. A., Kennedy, S. D., Tubbs, J. D., and Turner, D. H. (2010) 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.

8. Wales, D. J., and Yildirim, I. (2017) Improving computational pedictions of single-stranded RNA tetramers with revised α/γ torsional parameters for the AMBER frce field, *J. Phys. Chem. B 121*, 2989-2999.

9. Yildirim, I., Park, H., Disney, M. D., and Schatz, G. C. (2013) A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations, *J. Am. Chem. Soc. 135*, 3528-3538.

10. Yildirim, I., Stern, H. A., Sponer, J., Spackova, N., and Turner, D. H. (2009) Effects of restrained sampling space and nonplanar amino groups on free-energy predictions for RNA with imino and sheared tandem GA base pairs flanked by GC, CG, iGiC or iCiG Base pairs, *J. Chem. Theory Comput. 5*, 2088-2100.

11. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys. 23*, 327-341.

12. Lu, X. J., and Olson, W. K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures, *Nat. Protoc. 3*, 1213-1227.

13. Childs-Disney, J. L., Stepniak-Konieczna, E., Tran, T., Yildirim, I., Park, H., Chen, C. Z., Hoskins, J., Southall, N., Marugan, J. J., Patnaik, S., Zheng, W., Austin, C. P., Schatz, G. C., Sobczak, K., Thornton, C. A., and Disney, M. D. (2013) Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules, *Nat. Commun. 4*, 2044.

14. Childs-Disney, J. L., Yildirim, I., Park, H., Lohman, J. R., Guan, L., Tran, T., Sarkar, P., Schatz, G. C., and Disney, M. D. (2014) Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity, *ACS Chem. Biol. 9*, 538-550.

15. Wang, Z.-F., Ursu, A., Childs-Disney, J. L., Guertler, R., Yang, W.-Y., Bernat, V., Rzuczek, S. G., Fuerst, R., Zhang, Y.-J., Gendron, T. F., Yildirim, I., Dwyer, B. G., Rice, J. E., Petrucelli, L., and Disney, M. D. (2019) The hairpin form of r(G4C2)exp in c9ALS/FTD is repeat-associated non-ATG translated and a target for bioactive small molecules, *Cell Chem. Biol. 26*, 179-190.

16. Costales, M. G., Aikawa, H., Li, Y., Childs-Disney, J. L., Abegg, D., Hoch, D. G., Pradeep Velagapudi, S., Nakai, Y., Khan, T., Wang, K. W., Yildirim, I., Adibekian, A., Wang, E. T., and Disney, M. D. (2020) Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer, *Proc. Natl. Acad. Sci. U. S. A. 117*, 2406-2411.

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

18. Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., Cheeseman, J., Scalmani, G., Barone, V., Mennucci, B., and Petersson, G. (2005) Gaussian 09; Gaussian, Inc: Wallingford, CT.

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

20. Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation, *J. Am. Chem. Soc. 115*, 9620-9631.

21. Wang, Z.-F., Ursu, A., Childs-Disney, J. L., Guertler, R., Yang, W.-Y., Bernat, V., Rzuczek, S. G., Fuerst, R., Zhang, Y.-J., Gendron, T. F., Yildirim, I., Dwyer, B. G., Rice, J. E., Petrucelli, L., and Disney, M. D. (2019) The hairpin form of r(G_4C_2)^{exp} in c9ALS/FTD Is repeat-associated non-ATG translated and a target for bioactive small molecules, *Cell Chem. Biol. 26*, 179-190.e112.

22. D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C., Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I., Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails,, and R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016) AMBER 2016, *University of California, San Francisco*.

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

24. Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theory Comput. 8*, 3314-3321.

25. Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S., Gupta, S., Thomas, M. A., Hong, I., Chiu, S. L., Huganir, R. L., Ostrow, L. W., Matunis, M. J., Wang, J., Sattler, R., Lloyd, T. E., and Rothstein, J. D. (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport, *Nature 525*, 56-61.

26. Bernstam, L., and Nriagu, J. (2000) Molecular aspects of arsenic stress, *J. Toxicol. Environ. Health B Crit. Rev. 3*, 293-322.

27. Dikusar, E. A., and Potkin, V. I. (2011) Synthesis of 2-[3-alkoxy-4-(hydroxy, alkoxy, acyloxy)phenyl]-5-methyl-1H-benzimidazoles, *Russ. J.Org. Chem. 47*, 694-701.

28. Alpan, A. S., Sarıkaya, G., Çoban, G., Parlar, S., Armagan, G., and Alptüzün, V. (2017) Mannich-benzimidazole derivatives as antioxidant and anticholinesterase inhibitors: synthesis, biological evaluations, and molecular docking study, *Arch. Pharm. (Weinheim) 350*, e1600351.

29. Penieres-Carrillo, J.-G., Ríos-Guerra, H., Pérez-Flores, J., Rodríguez-Molina, B., Torres-Reyes, Á., Barrera-Téllez, F., González-Carrillo, J., Moreno-González, L., Martínez-Zaldívar, A., Nolasco-Fidencio, J.-J., Matus-Meza, A.-S., and Luna-Mora, R.-A. (2020) Reevaluating the synthesis of 2,5-disubstituted-1H-benzimidazole derivatives by different green activation techniques and their biological activity as antifungal and antimicrobial inhibitor, *J Heterocycl. Chem. 57*, 436-455.

30. Shaukat, A., Mirza, H. M., Ansari, A. H., Yasinzai, M., Zaidi, S. Z., Dilshad, S., and Ansari, F. L. (2013) Benzimidazole derivatives: synthesis, leishmanicidal effectiveness, and molecular docking studies, *Med. Chem. Res. 22*, 3606-3620.

31. Navarrete-Vázquez, G., Moreno-Diaz, H., Estrada-Soto, S., Torres-Piedra, M., León-Rivera, I., Tlahuext, H., Muñoz-Muñiz, O., and Torres-Gómez, H. (2007) Microwave-assisted one-pot synthesis of 2-(substituted phenyl)-1H-benzimidazole derivatives, *Synth. Commun. 37*, 2815-2825.

32. Navarrete-Vázquez, G., Hidalgo-Figueroa, S., Torres-Piedra, M., Vergara-Galicia, J., Rivera-Leyva, J. C., Estrada-Soto, S., León-Rivera, I., Aguilar-Guardarrama, B., Rios-Gómez, Y., Villalobos-Molina, R., and Ibarra-Barajas, M. (2010) Synthesis, vasorelaxant activity and antihypertensive effect of benzo[d]imidazole derivatives, *Bioorg. Med. Chem. 18*, 3985-3991.

33. Savall, B. M., Fontimayor, J. R., and Edwards, J. P. (2009) Selective phenol alkylation for an improved synthesis of 2-arylbenzimidazole H4 receptor ligands, *Tetrahedron Lett. 50*, 2490-2492.

34. Chaturvedi, A. K., Verma, A. K., Thakur, J. P., Roy, S., Bhushan Tripathi, S., Kumar, B. S., Khwaja, S., Sachan, N. K., Sharma, A., Chanda, D., Shanker, K., Saikia, D., and Negi, A. S. (2018) A novel synthesis of 2-arylbenzimidazoles in molecular sieves-MeOH system and their antitubercular activity, *Bioorg. Med. Chem. 26*, 4551-4559.

35. Brătulescu, G. (2017) Mild, one-pot preparation of 2-substituted benzimidazoles from organic

halides, *Synth. Commun. 47*, 811-817.