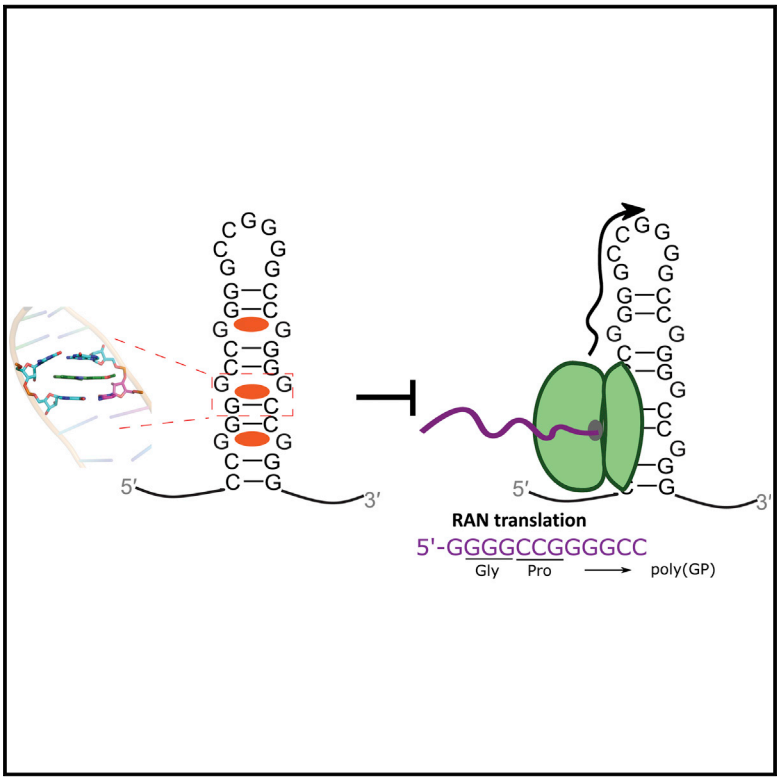
Article

Cell Chemical Biology

The Hairpin Form of r(G4C2)exp in c9ALS/FTD Is Repeat-Associated Non-ATG Translated and a Target for Bioactive Small Molecules

Graphical Abstract



Authors

Zi-Fu Wang, Andrei Ursu, Jessica L. Childs-Disney, ..., Joseph E. Rice, Leonard Petrucelli, Matthew D. Disney

Correspondence

disney@scripps.edu

In Brief

The most common cause of ALS is an expanded RNA repeat [r(G4C2)exp] that folds into two forms *in vitro*, a G-quadruplex and a hairpin. Wang et al. show that the hairpin form is present in cells, undergoes aberrant translation that causes toxicity, and thus is a target for therapeutic development.

Highlights

- An expanded RNA repeat that causes ALS folds into two forms in cells
- The hairpin, not quadruplex, form of the ALS RNA undergoes RAN translation
- Studies were enabled by development of a chemical probe selective for the hairpin

Wang et al., 2019, Cell Chemical Biology *26*, 1–12
February 21, 2019
https://doi.org/10.1016/j.chembiol.2018.10.018

CellPress

Please cite this article in press as: Wang et al., 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 Chemical Biology (2018), https://doi.org/10.1016/j.chembiol.2018.10.018

Cell Chemical Biology
Article

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

Zi-Fu Wang,[1] Andrei Ursu,[1] Jessica L. Childs-Disney,[1] Rea Guertler,[1] Wang-Yong Yang,[1] Viachaslau Bernat,[1] Suzanne G. Rzuczek,[1] Rita Fuerst,[1] Yong-Jie Zhang,[2] Tania F. Gendron,[2] Ilyas Yildirim,[3] Brendan G. Dwyer,[1] Joseph E. Rice,[4] Leonard Petrucelli,[2] and Matthew D. Disney[1,5,*]

[1]Departments of Chemistry and Neuroscience, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458, USA
[2]Department of Neuroscience, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224, USA
[3]Department of Chemistry and Biochemistry, Florida Atlantic University, Jupiter, FL 33458, USA
[4]Department of Medicinal Chemistry, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, 160 Frelinghuysen Road, Piscataway, NJ 08854, USA
[5]Lead Contact
*Correspondence: disney@scripps.edu
https://doi.org/10.1016/j.chembiol.2018.10.018

SUMMARY

The most common genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is an expanded G_4C_2 repeat [(G_4C_2)^{exp}] in *C9ORF72*. ALS/FTD-associated toxicity has been traced to the RNA transcribed from the repeat expansion [r(G_4C_2)^{exp}], which sequesters RNA-binding proteins (RBPs) and undergoes repeat-associated non-ATG (RAN) translation to generate toxic dipeptide repeats. Using *in vitro* and cell-based assays, we identified a small molecule (4) that selectively bound r(G_4C_2)^{exp}, prevented sequestration of an RBP, and inhibited RAN translation. Indeed, biophysical characterization showed that 4 selectively bound the hairpin form of r(G_4C_2)^{exp}, and nuclear magnetic resonance spectroscopy studies and molecular dynamics simulations defined this molecular recognition event. Cellular imaging revealed that 4 localized to r(G_4C_2)^{exp} cytoplasmic foci, the putative sites of RAN translation. Collectively, these studies highlight that the hairpin structure of r(G_4C_2)^{exp} is a therapeutically relevant target and small molecules that bind it can ameliorate c9ALS/FTD-associated toxicity.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS), a devastating neurodegenerative disorder caused by progressive loss of motor neurons, often results in paralysis and death within three to five years of onset (Taylor et al., 2016). Current treatment options for ALS are scarce and only palliative (Taylor et al., 2016). With the discovery that an expanded, transcribed G_4C_2 repeat in *C9ORF72* is the most common genetic cause of ALS and frontotemporal dementia (FTD) (hereafter "c9ALS/FTD"), a new therapeutic target has emerged (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The transcribed repeat RNA [r(G_4C_2)^{exp}] is involved in two putative pathological mechanisms: (1) sequestering proteins involved in RNA biogenesis to form toxic r(G_4C_2)^{exp}-protein foci (Prudencio et al., 2015); and (2) undergoing repeat-associated non-ATG (RAN) translation to produce toxic c9RAN dipeptide repeats (DPRs) in cellular (Almeida et al., 2013; Taylor et al., 2016; Zu et al., 2013), mouse (Chew et al., 2015; O'Rourke et al., 2015), and *Drosophila* (Mizielinska et al., 2014) models (Figure 1A).

Current efforts toward developing c9ALS/FTD therapeutics include both antisense oligonucleotides (ASOs) and small molecules that target r(G_4C_2)^{exp}. Indeed, an ASO directed at *C9ORF72* mRNA improved ALS-associated defects in mouse models (Gendron et al., 2017; Jiang et al., 2016). Small molecules, however, could be a viable alternative, owing to their favorable pharmacological properties amenable to traditional drug development (Su et al., 2014; Yang et al., 2015). As the repeat lies in an intron (DeJesus-Hernandez et al., 2011; Renton et al., 2011), small molecule binding could be advantageous in that it may not affect overall protein expression (Yang et al., 2015) of C9ORF72, which currently has an unknown function. Previously, we discovered that the small molecule **1a** binds r(G_4C_2)^{exp} and inhibits RAN translation and foci formation in both cultured cells and induced neurons, demonstrating the power of small molecules to favorably modulate RNA toxicity (Su et al., 2014).

Interestingly, r(G_4C_2)^{exp} exists in an equilibrium between two folded states, a hairpin and G-quadruplex (Figure 1A) (Su et al., 2014; Taylor, 2014). This equilibrium may be perturbed by temperature or by the presence of K^+ required for G-quadruplex folding (Lane et al., 2008). Often the G-quadruplex is presumed to be the therapeutically relevant structure (Conlon et al., 2016; Fay et al., 2017), as supported by a few studies that showed favorable biological effects with known G-quadruplex-targeting ligands (Simone et al., 2018; Zamiri et al., 2014). In contrast, **1a** binds the hairpin structure of r(G_4C_2)^{exp} and improves c9ALS/FTD-associated defects in iNeurons (Su et al., 2014). Thus, targeting the G-quadruplex or hairpin structure could both be important for the development of c9ALS/FTD therapeutics.

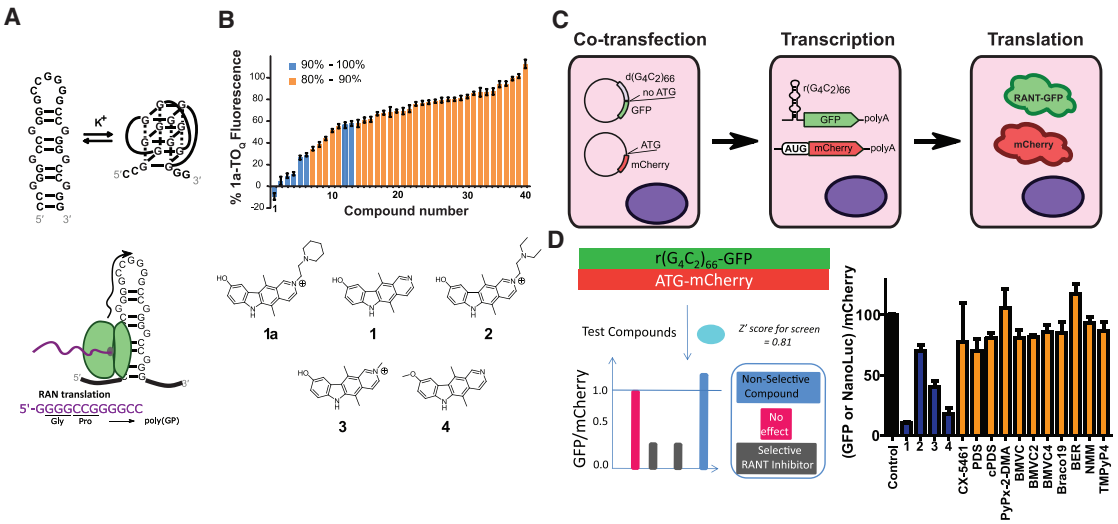


Figure 1. Screening Compounds for Binding to r(G4C2)8 by Displacement of 1a-TOQ and for Inhibiting RAN Translation

(A) Schematics of the equilibrium between the hairpin and the G-quadruplex forms of $r(G_4C_2)^{exp}$ (top) and RAN translation producing poly(GP) (bottom).

(B) Screening results as measured via displacement of **1a-TO$_Q$** (compound structures provided in Data S1) (top). Compounds with >90% similarity to **1a** are indicated with blue bars. Data represent mean values ± SEM of three independent measurements. Chemical structures of **1a** and **1–4** (bottom).

(C) Schematic of a cell-based screen to monitor RAN translation of $r(G_4C_2)^{exp}$.

(D) Left, the cell-based assay can distinguish between selective and non-selective inhibitors of RAN translation by monitoring changes in fluorescence derived from $r(G_4C_2)_{66}$-No ATG-GFP (RAN) and mCherry (canonical translation). Right, inhibitory effect of compounds (10 μM) on RAN translation. Note: (1) concentrations of **2** and **3** were 25 μM and (2) concentration of **CX-5461** was 5 μM due to cytotoxicity. **BMVC** and **BMVC2** were studied with the NanoLuc reporter due to interference with the GFP signal. Error bars are SD.

Thus, we sought to identify more potent inhibitors of $r(G_4C_2)^{exp}$ toxicity than **1a**, compare the activity of hairpin-binding small molecules to known G-quadruplex ligands, and study which fold(s) is therapeutically relevant. Indeed, we found several compounds that inhibited RAN translation better than known G-quadruplex ligands and the ASO Vivo-Morpholino, $mo(G_2C_4)_4$. The most potent compound, **4**, inhibited two pathomechanisms of c9ALS/FTD, sequestration of RNA-binding proteins (RBPs), and generation of toxic DPRs via RAN translation. Nuclear magnetic resonance (NMR) spectroscopy studies and molecular dynamics simulations showed that **4** bound the 1 × 1 nucleotide GG internal loop in the hairpin structure of $r(G_4C_2)^{exp}$. An *in vitro* translation assay and cellular imaging further support that the hairpin structure of $r(G_4C_2)^{exp}$ is RAN translated and a biologically relevant structure for the development of c9ALS/FTD therapeutics.

RESULTS

A Dye Displacement Assay Identifies Small Molecules that Bind More Tightly to $r(G_4C_2)^{exp}$ Than 1a

To identify more potent inhibitors of $r(G_4C_2)^{exp}$ than **1a** (Su et al., 2014), we employed a chemical similarity search based on our lead compound (Data S1). The search yielded a cohort of small molecules (n = 40) exhibiting at least 80% chemical similarity to **1a** with varying charge, structure, and side chains (Data S1). The ability of these **1a**-like small molecules to bind $r(G_4C_2)^{exp}$ *in vitro* was measured via a dye displacement assay. Combining **1a** with TO-PRO-1, a well-known fluorescent RNA intercalator (**1a-TO$_Q$** [STAR Methods and Data S2]), yielded an excellent fluorescent reporter of $r(G_4C_2)^{exp}$ binding, as demonstrated by the 3-fold enhancement of emission signal upon titration of $r(G_4C_2)_8$. Fitting the change in emission of **1a-TO$_Q$** as a function of $r(G_4C_2)_8$ concentration afforded a K_d of 110 ± 14 nM (Data S1). To validate the assay, we competed off the increase in **1a-TO$_Q$**'s fluorescence observed upon binding $r(G_4C_2)_8$ with **1a**, which afforded a dose response as expected. Using **1a**, we calculated the Z factor of this assay, which is equal to 0.88 and thus considered suitable for high throughput screening (Zhang et al., 1999).

Using the **1a-TO$_Q$** displacement assay, we identified ten **1a**-like compounds that displaced >50% of **1a-TO$_Q$** from $r(G_4C_2)_8$ (Figure 1B). Generally, compounds with >90% similarity to **1a** were the most potent (blue bars; Figure 1B). Several trends were observed to elucidate chemical features that drive binding to $r(G_4C_2)_8$. Compounds **2**, **3**, **5**, and **12** have positive charges via *N*-alkylation of the pyridine ring and exhibited potent binding, yet **1** and **4** also demonstrated similar binding affinities despite being uncharged. This suggests that positive charges contribute favorably but are not essential to molecular recognition between the small molecule probe and $r(G_4C_2)^{exp}$. Dose-dependent binding assays demonstrated that the best performing compounds (**1–4**, Figure 1B) are more potent than **1a** with half maximal inhibitory concentration (IC_{50}) values ranging from 0.9 to 8.5 μM (Table 1).

1a-like Small Molecules Inhibit RAN Translation in a Cellular Reporter Assay

To assess the bioactivity of $r(G_4C_2)^{exp}$-binding compounds, we developed a cell-based screen to monitor selective inhibition of RAN translation of $r(G_4C_2)^{exp}$ (Figures 1C and 1D). In brief $(G_4C_2)^{exp}$ was inserted into the 5′ UTR of GFP lacking a canonical

Table 1. IC_{50} Values for the Displacement of 1a-TO_Q and for the Inhibition of r(G_4C_2)$_{66}$-Mediated RAN Translation

Compound	Displacement of **1a-TO_Q** IC_{50} (μM)	Inhibition of RAN Translation IC_{50} (μM)
1a	2.6 ± 1.0	11 ± 1.0
1	1.3 ± 0.5	0.27 ± 0.05
2	1.4 ± 0.4	>25
3	0.9 ± 0.2	12 ± 2.0
4	8.5 ± 2.0	1.6 ± 0.20
mo(G_2C_4)$_4$	–	7.6 ± 1.0

Error is reported as SD.

ATG start codon, akin to a construct previously reported to study RAN translation of r(CGG)^{exp} (Yang et al., 2015). Western blot analysis confirmed that the product generated from this "(G_4C_2)$_{66}$-*No ATG-GFP*" construct is a RAN DPR-GFP fusion, with no detectable amounts of the canonically translated GFP product (Figure S1A).

To assess whether a small molecule selectively inhibits RAN translation, cells were co-transfected with (G_4C_2)$_{66}$-*No ATG-GFP* and a plasmid encoding mCherry (*ATG-mCherry*). Thus, two fluorescent signals could be produced, where GFP fluorescence indicates RAN translation and mCherry fluorescence is a metric of canonical translation. By measuring the relative fluorescence of both proteins, the assay can distinguish whether a small molecule globally affects protein translation non-selectively (both GFP and mCherry are affected), selectively affects canonical translation (only mCherry is affected), or selectively inhibits RAN translation (only GFP is affected) (Figures 1C and 1D). The r(G_4C_2)^{exp} RAN translation assay has an excellent Z factor of 0.81, as measured using compound **1a**, and is therefore suitable for cell-based high throughput screening.

The most promising compounds (**1–4**) from the **1a-TO_Q** displacement screen inhibited RAN translation, with IC_{50} values ranging from 0.27 to 12 μM (blue bars in Figure 1D; Table 1). In fact, **1** and **4** are more potent than **1a**, with IC_{50} values of 0.27, 1.6, and 11 μM, respectively (Table 1). Both **1** and **4** were better inhibitors of r(G_4C_2)^{exp} RAN translation than the oligonucleotide Vivo-Morpholino, mo(G_2C_4)$_4$ (Data S1), which had an IC_{50} of 7.6 μM ($p < 0.001$) (Table 1). Thus, small molecules that target structured RNAs can more potently modulate biological activity than oligonucleotides. This phenomenon is likely due to the kinetic issues of oligonucleotides binding to structured RNA targets, which can be further confounded, in this case, by the oligonucleotide's self-structure (Tran et al., 2014).

To ensure that the relative decrease in the GFP as compared to the mCherry signal occurred as a result of inhibition of RAN translation, a counterscreen for **1–4** was completed with *ATG-GFP* (canonical GFP translation) and *ATG-mCherry* (also canonical translation). As shown in Figures S1B and S1C, **2–4** did not significantly affect the GFP signal from the *ATG-GFP* construct relative to the mCherry signal, while **1** induced a roughly 20% decrease in the GFP signal at 2.5 μM. Therefore, **2–4** are selective RAN translation inhibitors, while **1** has some associated non-specific effects. As **4** was the best selective inhibitor of RAN translation in this cell-based assay, it was studied further.

Compound 4 Exhibits Excellent Affinity and Selectivity for Binding GG Internal Loops in the Hairpin Form of r(G_4C_2)^{exp}

We next investigated the affinity and selectivity of **4** binding to various DNA and RNA targets by using biolayer interferometry (BLI). Compound **4** bound the desired target, r(G_4C_2)$_8$, with a K_d of 0.26 μM, while weakly binding the antisense RNA [r(G_2C_4)$_8$] and a fully base paired RNA [r(G_2C_2)$_8$] (Table 2). In particular, **4** bound r(G_4C_2)$_8$ ~300-fold more tightly than to the antisense r(G_2C_4)$_8$ (K_d = 80 μM) and ~540-fold more tightly than to the r(G_2C_2)$_8$ (K_d = 140 μM). Repeat length [r(G_4C_2)$_n$ where n = 2, 4, and 6] did not affect **4**'s binding affinity (Table S2). The selectivity of **4** binding was studied further with a well-known RNA G-quadruplex called TERRA, r(UAG$_3$(U_2AG_3)$_3$). Although **4** did bind TERRA with a K_d of 8.6 μM, **4** bound 33-fold more tightly to r(G_4C_2)$_8$ (Table 2). With respect to interactions with DNA, **4** binds to d(G_4C_2)$_8$ with a K_d of 0.31 μM; however, **4** does not affect (G_4C_2)$_{66}$-*No ATG-GFP* RNA transcript levels as assessed by RT-qPCR (Figure S1D), suggesting that binding interactions between **4** and G_4C_2 DNA did not affect transcription and are therefore not biologically relevant.

Compound 4 Is Selective for 1 × 1 Nucleotide GG Loops over Other Nucleic Acids *In Vitro* and in Cells

We previously found that r(UG_3C_2)^{exp} produces toxic DPRs via RAN translation, the genetic cause of spinocerebellar ataxia disease type 36 (SCA36) (Liu et al., 2014; Matsuzono et al., 2017; Obayashi et al., 2015). Similar to r(G_4C_2)^{exp}, r(UG_3C_2)^{exp} forms a hairpin structure with a GU internal loop stabilized by hydrogen bonding (Figure S2A). Therefore, r(UG_3C_2)^{exp} serves as an excellent comparison to determine whether **4** binds the 1 × 1 nucleotide GG internal loop in r(G_4C_2)^{exp} selectively. Binding studies revealed that **4** had an ≈1,000-fold lower affinity for binding r(UG_3C_2)$_8$ than r(G_4C_2)$_8$ (Table 2); binding of **4** to r(UG_3C_2)$_8$ was also not detectable by NMR spectroscopy (Figure S2C). Not surprisingly, **4** is unable to inhibit RAN translation of r(UG_3C_2)$_{62}$ in transfected cells (Figure S2B).

NMR Spectroscopy and Molecular Dynamics Simulations Reveal Features of the Molecular Recognition of r(G_4C_2)^{exp} by 4

We next characterized the binding of **4** to r(G_4C_2) repeats using imino proton NMR and circular dichroism (CD) spectroscopies. NMR characterization of **4** binding to r(G_4C_2)$_4$ and to r(G_4C_2)$_8$ showed a new peak appearing at ~10.5 ppm in the absence of K^+ ions, resulting from **4** stabilizing the interactions between the 1 × 1 nucleotide GG internal loop (Figures 2A and S3A). As the imino proton signal for G-quadruplex is nearby that of hydrogen-bonded GG loops, we studied the binding of **4** to r(G_4C_2)$_2$, which cannot form a G-quadruplex. Indeed, the same imino proton signal corresponding to the GG loops in r(G_4C_2)$_4$ and r(G_4C_2)$_8$ was observed (Figure S4). Further, **4** prevents conversion of the hairpin to the G-quadruplex (Figures 2A and S3A). As the concentration of **4** increased, pronounced peak shifts of the GC base pair imino protons at 13.3 and 12.5 ppm were observed. By plotting the change in the NMR signals as a function of [**4**]:[RNA], binding stoichiometry can be determined. As shown in Figure S3A, **4** binds to r(G_4C_2)$_4$ with a stoichiometric

Table 2. Binding Parameters of 4 and r(G_4C_2)$_8$, d(G_4C_2)$_8$, r(G_2C_4)$_8$ (Antisense Repeat), r((G_2C_2)$_4$AAAA(G_2C_2)$_4$) (GC Paired), r(UGGGCC)$_8$, and r(UAG$_3$(U_2AG$_3$)$_3$) (TERRA) by Biolayer Interferometry

RNA	$K_{d,1}$ (M)	$k_{on,1}$ (1/M × s)	$k_{off,1}$ (1/s)	Selectivity for r(G_4C_2)$_8$ (Fold)	$K_{d,2}$ (M)	$k_{on,2}$ (1/M × s)	$k_{off,2}$ (1/s)
r(G_4C_2)$_8$	2.6 (±0.3) × 10^{-7}	2.3 (±0.3) × 10^{4}	6.0 (±0.1) × 10^{-3}	–	1.4 (±0.1) × 10^{-6}	4.0 (±0.2) × 10^{4}	5.6 (±0.3) × 10^{-2}
d(G_4C_2)$_8$	3.1 (±0.1) × 10^{-7}	2.6 (±0.1) × 10^{4}	8.0 (±0.2) × 10^{-3}	1.2	4.0 (±1.2) × 10^{-6}	2.0 (±0.6) × 10^{4}	8.0 (±0.2) × 10^{-2}
r(G_2C_4)$_8$ (antisense)	7.9 (±5.9) × 10^{-5}	1.2 (±0.9) × 10^{3}	9.5 (±0.3) × 10^{-2}	304	NA	NA	NA
r((G_2C_2)$_4$AAAA(G_2C_2)$_4$) (GC paired)	1.4 (±0.9) × 10^{-4}	1.5 (±0.9) × 10^{3}	2.1 (±0.8) × 10^{-1}	538	NA	NA	NA
r(UG_3C_2)$_8$	2.2 (±0.9) × 10^{-4}	1.9 (±0.8) × 10^{2}	4.0 (±0.2) × 10^{-2}	847	NA	NA	NA
r(UAG$_3$(U_2AG$_3$)$_3$)	8.6 (±0.2) × 10^{-6}	3.2 (±0.1) × 10^{3}	2.6 (±0.1) × 10^{-2}	33	NA	NA	NA

k_{on} and k_{off} values were generated by fitting the obtained data using ForteBio's Data Analysis 7.1 software to 1:1 or 2:1 (heterogeneous) binding equation using the entire time range. Values shown for k_{on} and k_{off} are representative of two independent measurements and the errors are the SD. NA, not applicable.

ratio of 4:1 compound:RNA and to r(G_4C_2)$_8$ with a stoichiometric ratio of 6:1.

Nuclear Overhauser effect spectroscopy (NOESY) analysis revealed nuclear Overhauser effects (NOEs) between **4** and the 1 × 1 nucleotide GG internal loops and adjacent GC pairs (Figure S3B). CD spectroscopy confirmed that **4** stabilizes the hairpin structure, as demonstrated by the increase in T_M from 71°C to 78°C after the addition of **4** (Figure 2B). These studies also revealed that **4** has two binding modes, as demonstrated by the two signals at 300 and 330 nm (Figure S3C). Fitting the resulting data from CD studies to a Hill equation indicates that binding of **4** is cooperative (Figure S3C). Collectively, these studies suggest that **4** binds between 1 × 1 nucleotide GG internal loops and GC pairs, stabilizing hydrogen bonding within the loop.

To confirm these results, a duplex model that contains three unique 1 × 1 nucleotide GG internal loops distinguishable by NMR spectroscopy (Figure S4A) was employed. By site-specifically methylating the 1 position of each G in a 1 × 1 nucleotide GG internal loop, we were able to assign which G residues were acting as hydrogen bond donors and acceptors in its interaction with **4** (Figures S4B–S4E). This analysis revealed that: (1) **4** stabilizes hydrogen bonding between the 1 × 1 nucleotide GG internal loop; (2) the 1 × 1 nucleotide GG internal loop hydrogen bonds in a *syn-anti* orientation; and (3) several NOEs between **4** and the imino protons of the guanine in the 1 × 1 nucleotide GG internal loop were observed.

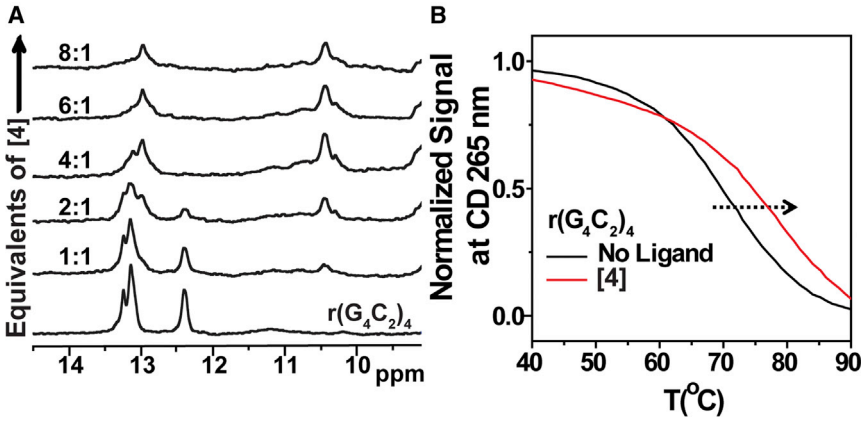
To provide additional insight into the molecular recognition of **4** and r(G_4C_2) repeats, we used molecular dynamics calculations (see Supplemental Information). The lowest binding free energy state from these studies showed **4** stacked between the 1 × 1 nucleotide GG internal loop and the closing GC base pairs, stabilized by π-π interaction between **4** and loop residues (Figures 2C–2E). Hydrogen bonding between 1 × 1 nucleotide GG internal loops in the absence of compound were observed in the imino proton NMR spectrum at low temperature (283K) (Figure S3D) but not at room temperature. The observation that hydrogen bonding within the GG loop in the presence of **4** was observable at room temperature provides additional support that **4** stabilizes the RNA's structure.

G-Quadruplex Ligands Bind r(G_4C_2)^{exp} but Do Not Inhibit RAN Translation

Collectively, the studies described above suggest that **4**'s cellular mode of action is binding to r(G_4C_2)^{exp} in its hairpin form and blocking RAN translation by stabilization of the RNA's structure. However, various reports have suggested that the therapeutically relevant fold of r(G_4C_2)^{exp} is a G-quadruplex (Figure 1A) (Biffi et al., 2014; Burger et al., 2005; Franceschin et al., 2006; Hershman et al., 2008; Huang et al., 2008; Rodriguez et al., 2012; Rzuczek et al., 2010; Siddiqui-Jain et al., 2002; Wang et al., 2014; Xu et al., 2017). We therefore compared the biological activity of 11 well-known G-quadruplex ligands (Data S1) to **4** using our cell-based RAN translation assay (Figures 1C and 1D). (Note: **BMVC** and **BMVC2** were tested using a (G_4C_2)$_{66}$-*No ATG-NanoLuc* construct because they interfere with the fluorescent signal of GFP; **4** was also tested using this system, affording the same IC_{50} as it did for (G_4C_2)$_{66}$-*No ATG-GFP* [Figure S1E].) Interestingly, the G-quadruplex ligands were poor inhibitors of RAN translation, significantly worse than **4** (Figure 1D). The poor inhibitory effect was not due to lack of affinity, as the G-quadruplex ligands demonstrated excellent binding to the quadruplex form of r(G_4C_2)$_4$, induced by addition of K^+ (Table S1). **TMPyP4** (K_d = 50 nM) and **PDS** (K_d = 70 nM) bound more tightly to the quadruplex form than **4** bound to the hairpin (K_d = 260 nM) (Tables 2 and S1). (Note: binding of G-quadruplex ligands to the G-quadruplex form was confirmed by NMR spectroscopy [Figure S5].) Thus, inhibiting RAN translation is not simply a function of binding affinity, rather the affinity for a particular structural form and its presence within a cell.

The Hairpin Is a Biologically Relevant Target for Small Molecule Modulation of r(G_4C_2)^{exp} Biology

The studies above suggest that the hairpin fold is therapeutically relevant for inhibiting r(G_4C_2)^{exp} pathobiology, particularly RAN translation. To gain further insight into the structural form adopted by r(G_4C_2)^{exp} in cells, we conducted both *in vitro* and cell-based experiments. Although both hairpin and G-quadruplex forms are thermodynamically stable, we were interested in studying the kinetically favored form. Imino proton spectra



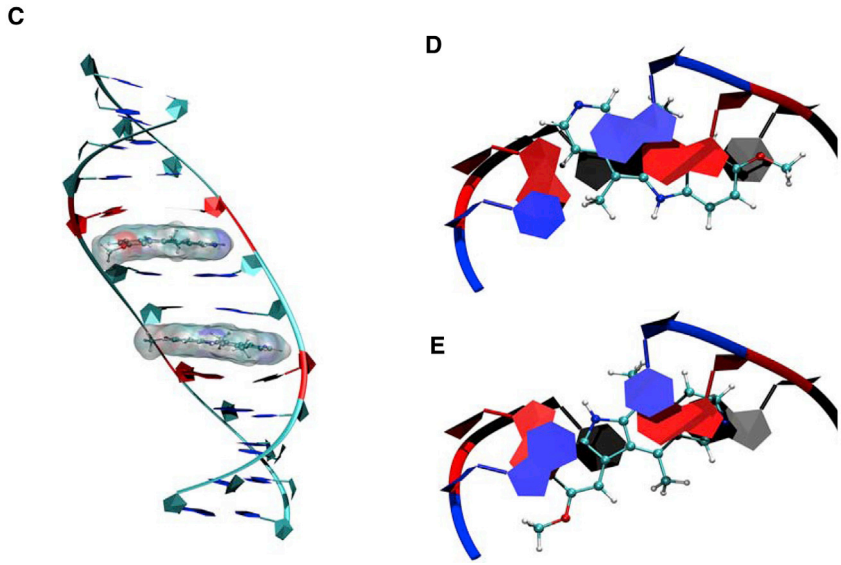


Figure 2. CD and NMR Analysis of 4 Binding to the 1 × 1 Nucleotide GG Internal Loops

(A) Imino proton NMR spectra of r(G_4C_2)$_4$ with increasing amounts of **4**. A new peak appears at ~10.5 ppm upon addition of **4**, consistent with **4** binding and stabilizing the GG pair's hydrogen bonds.

(B) CD melting of r(G_4C_2)$_4$ in the presence and absence of **4**.

(C) Two molecules of **4** interacting with a model r(G_4C_2)^{exp} having two unique 1 × 1 GG internal loop motifs (CGG and GGC) with *syn-anti* orientations.

(D) Top views of **4** interacting with 5′-CGG/5′-CGG.

(E) 5′-GGC/5′-GGC as displayed on the left. Blue, red, and black colored residues represent the first, second, and third base pairs from top, respectively.

show that, upon addition of K^+, hairpin signals decrease and the G-quadruplex signals increase (Figure 3A). A three-state kinetic model was employed to extrapolate the unfolding rate of the hairpin into the G-quadruplex structure via an unfolded intermediate (Figure 3A and Equations 3.1– 3.4 in STAR Methods).

The unfolding rate of hairpin structure (k_1) increased with temperature (Kuo et al., 2015), while the folding rate of G-quadruplex (k_3) was temperature independent (Figures S6A and S6B), consistent with previous G-quadruplex folding studies (Gray and Chaires, 2008; Zhang and Balasubramanian, 2012). These data indicated that a transition state energy is required for hairpin structure unfolding, while there is no substantial barrier for G-quadruplex folding. Using these temperature-dependent measurements (Figure S6A), a sizable transition state energy, $\Delta G^{\ddagger}$ = 20 kcal/mol, was calculated, suggesting that a substantial kinetic barrier exists between the hairpin and unfolded states for r(G_4C_2)^{exp}; the formation of the G-quadruplex is evidently limited by this kinetic barrier (Figure 3B) (Kuo et al., 2015).

Computational analysis predicted that the stability of the hairpin increases as repeat length increases (Figure S6C) (Zuker, 2003; Ash et al., 2013). A series of CD measurements confirmed this prediction, as increasing the number of repeats from 4 to 8 to 12 lowered the observed free energy (Figure 3B). The unfolding time (272–407 min) and the population (0.4–0.6) (Figures S6D and S6E) in the hairpin form increases from 4 to 12 repeats, indicating that the hairpin structure has a higher folding population than the G-quadruplex and suggesting that hairpin form is the kinetically favored.

We next conducted an *in vitro* translation assay to determine whether the hairpin or G-quadruplex form undergoes RAN translation by using (G_4C_2)$_{66}$-*No ATG-NanoLuc* and luciferase signal as a readout. The construct was used to transcribe r(G_4C_2)$_{66}$-No ATG-NanoLuc, which was subsequently folded into a G-quadruplex or hairpin form, as informed by NMR spectroscopy studies. Significant luciferase signal was only observed from *r*(G_4C_2)$_{66}$-*No ATG-NanoLuc* folded into the hairpin form and not the G-quadruplex form (~3.5-fold difference compared with both the G-quadruplex template and a reaction lacking an RNA template; $p < 0.001$) (Figure S2D).

Our *in vitro* studies suggest that the hairpin fold of r(G_4C_2)^{exp} predominates from a kinetic standpoint and is a main contributor of pathobiology, at least in terms of RAN translation. To confirm these findings in cells, we used **4** to develop an r(G_4C_2)^{exp} hairpin imaging agent. In brief, **4** was conjugated with the well-known RNA-binding dye TO-PRO-1, and this hybrid molecule is dubbed **4-TO-PRO** (Figure 3E). (G_4C_2)$_{66}$-transfected cells were incubated with **4-TO-PRO**, revealing strong fluorescent spots in the cytoplasm as determined by fluorescence microscopy (Figure 3E). In contrast, these punctate spots were not visible using TO-PRO-1 alone (Figure S7A). Pre-treating the cells with **4** or RNase significantly reduced these fluorescent spots (Figures 3 and 3F), which suggests that **4** localized to these spots enriched with RNA. According to a previous study by Donnelly et al. (2013), these cytoplasmic spots are the putative sites of RAN translation. Therefore, the observed fluorescent signal from **4-TO-PRO** was most likely due to binding of cytoplasmic r(G_4C_2)^{exp} in the hairpin form.

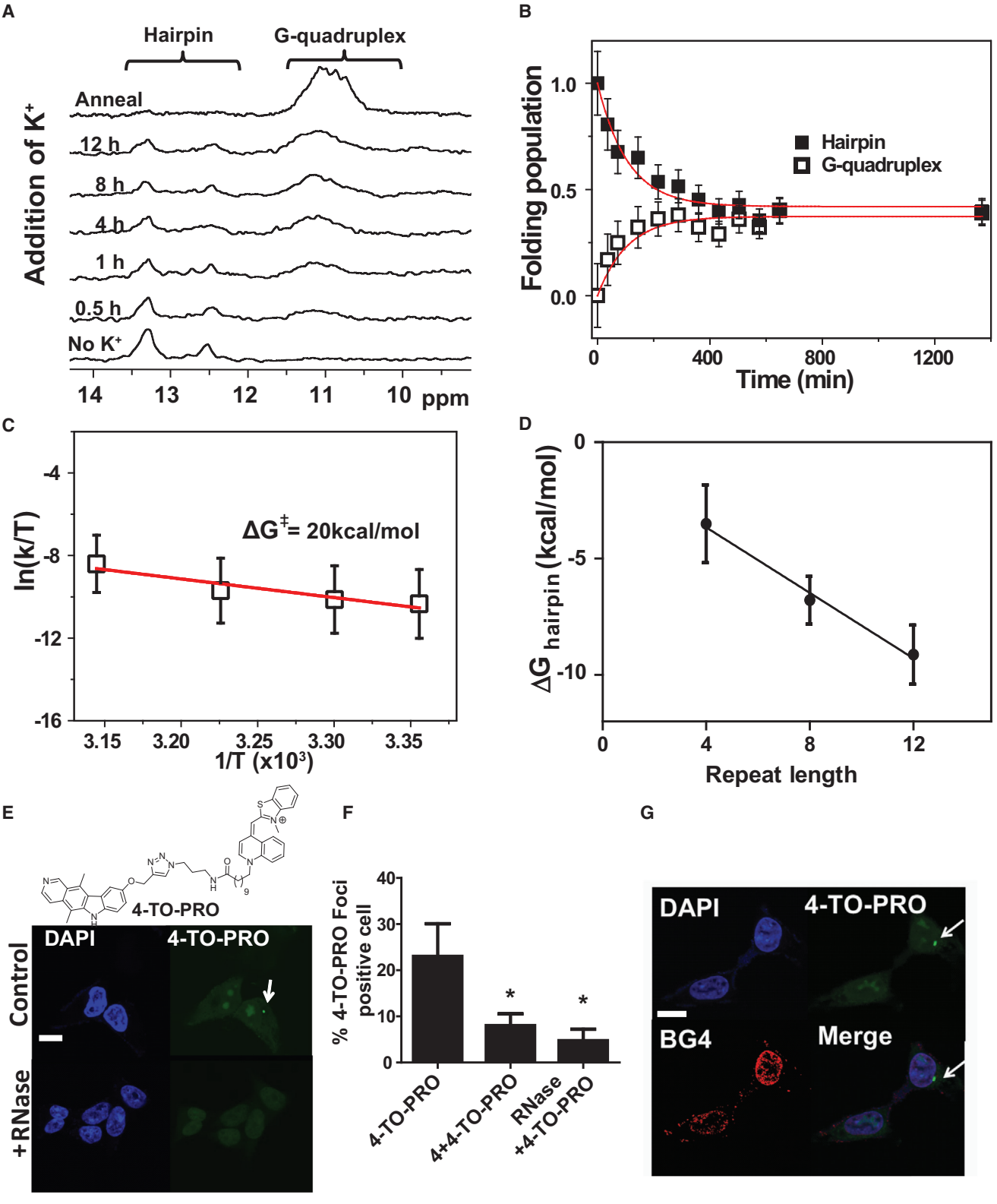


Figure 3. Kinetic and Thermodynamic Study Reveals the Hairpin Structure Is the Major Population in rG_4C_2^{exp} and Also Forms in Cells

(A) Time-resolved imino proton spectra of r(G_4C_2)$_4$ were recorded in Li^+-only buffer first, then at 0.5, 1, 4, 8, and 12 hr post addition of K^+ (37°C), and finally after annealing at 95°C in K^+ ("Anneal").

(legend continued on next page)

To ascertain whether **4-TO-PRO** bound the hairpin or G-quadruplex form of cytoplasmic r(G_4C_2)^{exp}, the G-quadruplex-specific antibody BG4 was employed (Biffi et al., 2014) (STAR Methods). If **4** targets the hairpin form of r(G_4C_2)^{exp} in cells as expected, then BG4 should not co-localize with **4-TO-PRO**; co-localization of BG4 and **4-TO-PRO** would indicate that **4** binds to the G-quadruplex form in cells. BG4 was added to the cells first followed by addition of **4-TO-PRO** to avoid disrupting BG4 binding. Indeed, the BG4 signal did not co-localize with **4-TO-PRO** (Figure 3G), providing evidence that **4** binds the hairpin form of cellular r(G_4C_2)^{exp}, and that the hairpin fold of r(G_4C_2)^{exp} occurs in cells.

Investigating Compound Mode of Action: 4 Blocks Polysome Assembly on r(G_4C_2)^{exp}

We next studied **4**'s mode of action, which could occur by: (1) transcriptional silencing, (2) compound-induced cleavage of the mRNA, (3) blocking ribosomal binding, and/or (4) preventing ribosomal readthrough (stalling). No change in the steady-state levels of r(G_4C_2)^{exp}-containing transcript was observed, as determined by qRT-PCR (Figure S1D); thus **4** does not appear to silence transcription of the mRNA or induce its cleavage.

Polysome profiling studies were then employed to determine whether **4** inhibits ribosome binding or induces stalling. Treatment of HEK293T cells expressing (G_4C_2)$_{66}$-*No ATG-GFP* with **4** reduced the amount of transcript loaded into high-molecular-weight polysomes by 28% ($p < 0.001$), low-molecular-weight polysomes by 18% ($p < 0.01$), and monosome-containing fractions by 12% ($p < 0.05$) (Figures 4A and 4B). These results indicate that **4** acts as a steric block that hinders polysome assembly on r(G_4C_2)^{exp}, thus decreasing levels of toxic DPRs. Previous studies have found that small molecules targeting r(CGG)^{exp} similarly block loading of the corresponding mRNA onto polysomes and inhibit RAN translation (Yang et al., 2015).

Investigating Compound Mode of Action: 4 Blocks Formation of Toxic r(G_4C_2)^{exp}-RBP Complexes

Another pathomechanism of r(G_4C_2)^{exp} is sequestration of RBPs. As **4** effectively blocks ribosomal binding, we hypothesized that **4** could also block RBPs from binding r(G_4C_2)^{exp}, which accumulate in nuclear foci (Mizielinska and Isaacs, 2014). A key protein that binds r(G_4C_2)^{exp} and colocalizes with r(G_4C_2)^{exp} foci is heterogeneous nuclear ribonucleoprotein H (hnRNP H) (Prudencio et al., 2015, Lee et al., 2013). We measured the affinity of hnRNP H for r(G_4C_2)$_8$ using a gel shift assay, yielding a K_d of 65 ± 26 nM (Figure S7B). Therefore, we developed an *in vitro* assay around this complex to assess whether compound binding to r(G_4C_2)^{exp} inhibits RNA-protein complex formation.

The hnRNP H displacement assay was based on a previously developed time-resolved fluorescence resonance energy transfer (TR-FRET) assay used to study whether small molecules inhibited the binding of muscleblind-like 1 protein to r(CUG)^{exp} (Figure 5A) (Chen et al., 2012). In our assay, biotinylated r(G_4C_2)$_8$ forms a complex with hnRNP H-His$_6$. A FRET signal is produced when streptavidin-XL665 binds the biotinylated RNA and Tb-Anti-His$_6$ binds to hnRNP H. Small molecules that bind r(G_4C_2)$_8$ and inhibit the formation of the r(G_4C_2)$_8$-hnRNP H complex reduce the FRET signal. In this assay, **4** inhibited hnRNP H binding with an IC_{50} of 19 μM (Figure 5B).

We next measured reduction of foci upon compound treatment in cells using RNA fluorescence *in situ* hybridization and immunostaining of hnRNP H. As shown in Figure 5C, **4** inhibits foci formation in HEK293T cells expressing (G_4C_2)$_{66}$-*No ATG-GFP*. Quantification confirmed that **4** treatment reduced the number of foci-positive cells and decreased the number of foci in foci-positive cells by 2-fold ($p = 0.0029$; Figure 5B). Thus, **4** inhibits two putative pathomechanisms of c9ALS/FTD—RAN translation and formation of r(G_4C_2)^{exp}-protein complexes—while binding to the 1 × 1 nucleotide GG internal loop.

DISCUSSION

Advancing the development of c9ALS/FTD therapeutics depends on establishing whether the G-quadruplex and hairpin forms are therapeutically relevant structures of r(G_4C_2)^{exp}. Driven by improving upon our previously reported small molecule **1a**, which ameliorated r(G_4C_2)^{exp} toxicity in iNeurons by binding to the hairpin structure (Su et al., 2014), we sought to identify better small molecule modulators of r(G_4C_2)^{exp} toxicity and to use the probes to study r(G_4C_2)^{exp} structures adopted in cells. Indeed, we identified many compounds that bound to r(G_4C_2)^{exp} and inhibited RAN translation better than our previously reported compound **1a**. The most promising compound, **4**, exhibits excellent selectivity and potency toward r(G_4C_2)^{exp}.

Our biophysical studies indicated that **4** binds to the 1 × 1 nucleotide GG internal loop in the hairpin structure in r(G_4C_2)^{exp}, inducing 1 × 1 nucleotide GG hydrogen bonding by reducing the dynamics of the 1 × 1 nucleotide GG internal loop. The 1 × 1 nucleotide GG internal loop in the hairpin is critical for molecular recognition and binding of **4** to the r(G_4C_2)^{exp}. The π-π stacking interaction of **4** with the individual guanine bases in the 1 × 1 nucleotide GG loop and adjacent GC closing pair stabilizes the complex's structure.

The binding of **4** to r(G_4C_2)^{exp} prevents key RNA-protein interactions, thereby inhibiting two putative c9ALS/FTD pathomechanisms of r(G_4C_2)^{exp}: RAN translation of r(G_4C_2)^{exp} and

(B) Kinetic trace analysis of r(G_4C_2)$_4$ as monitored by the imino proton NMR signal of hairpin (solid squares) or G-quadruplex region (open squares) after addition of K^+ at 37°C. The red line shows the fit to a three-state kinetic model.

(C) Activation free energy of r(G_4C_2)$_4$ transition from hairpin structure to unfolded state by Eyring plot of k_1. Data are represented as the mean ± SEM.

(D) Folding free energy of hairpin structure ($\Delta G^\circ_{Hairpin}$) as a function of r(G_4C_2)$_n$ repeat length (n = 4–12), as measured by CD melting temperature. Data are represented as the mean ± SD.

(E) Chemical structure of **4-TO-PRO** and representative images of **4-TO-PRO**-treated (G_4C_2)$_{66}$-transfected cells.

(F) Quantification of foci-positive cells treated with **4-TO-PRO** only (n = 112 cells; 3 biological replicates), pre-treated with **4** and then stained with **4-TO-PRO** (n = 100 cells; 3 biological replicates), or treated with RNase and then stained with **4-TO-PRO** (n = 142 cells; 3 biological replicates). Data are represented as the mean ± SD.

(G) Representative images of cells treated first with BG4 and then **4-TO-PRO**.

Scale bar, 5 μm (in both images). *$p < 0.05$ as determined by a two-tailed Student t test.

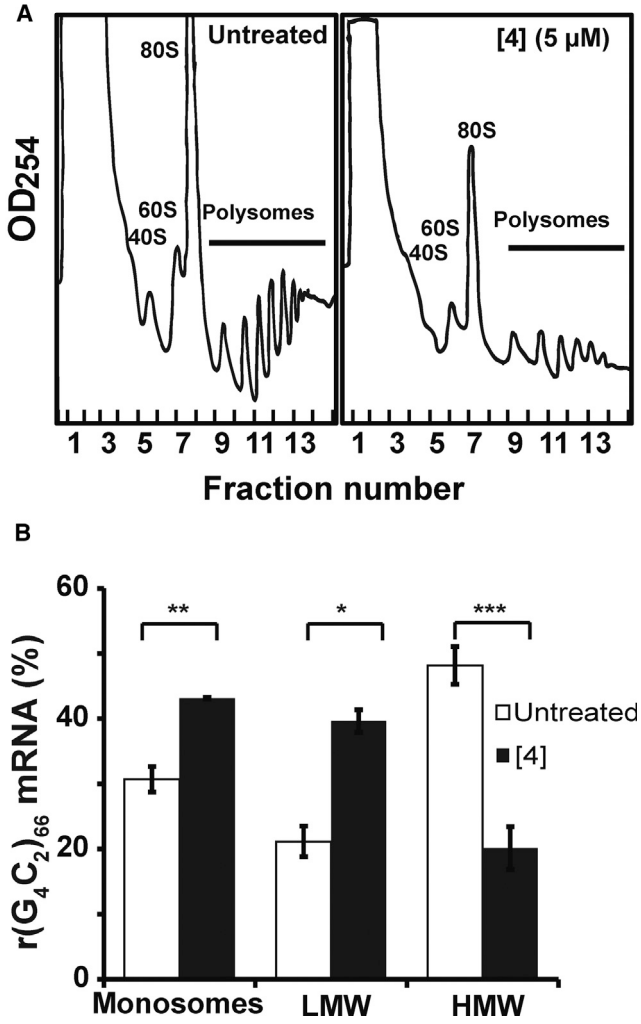


Figure 4. Polysome Profiling Experiments Elucidate 4's Mode of Inhibition of RAN Translation

(A) Polysome fractionation profiles of cell lysates obtained from HEK293T cells transfected with $(G_4C_2)_{66}$-*No ATG-GFP* plasmid upon treatment with **4** or vehicle (DMSO). Polysome fractionation profiles are representative of two independent experiments.

(B) Percentage of $(G_4C_2)_{66}$-*No ATG-GFP* mRNA transcript present within monosome- and polysome-containing fractions with (black) and without (white) treatment of **4**. Fractions labeled as "Monosomes" contain 40S, 60S, and 80S ribosomal subunits; "LMW" indicates low molecular weight polysomes; and "HMW" indicates high molecular weight polysomes. Data are represented as the mean ± SD. *p < 0.05; **p < 0.01; ***p < 0.001, as determined by a two-tailed Student t test.

formation of toxic RBP complexes. **4** inhibits RAN translation by acting as a steric block to prevent polysome assembly. In fact, **4** inhibited RAN translation more potently than the oligonucleotide mo$(G_2C_4)_4$, demonstrating that small molecule probes are a viable alternative to oligonucleotides. While an ASO complementary to the RAN translation initiation site inhibited ribosomal scanning and, hence, RAN translation (Tabet et al., 2018), ASOs are most efficacious when targeting unstructured regions; thus RNA repeat expansions are not ideal targets. Various factors compound the challenges associated with oligonucleotide-based modalities that target r(G_4C_2)^{exp} including: (1) the self-structure of oligonucleotides complementary to r(G_4C_2) repeats (Kovanda et al., 2015)]; (2) off-targets due to sequence, rather than structure-based recognition; and (3) effects on both RAN and canonical ATG translation (Yang et al., 2015).

A previous report found that hnRNP H colocalized with the G-quadruplex antibody BG4 in c9ALS astrocytes and that the BG4 signal was higher in c9ALS cells than control cell lines (Conlon et al., 2016). A structural study, however, showed that the conserved qRRM domain of hnRNP F/H binds to RNA G-rich sequences and prevents formation of G-quadruplexes (Dominguez et al., 2010). Rather than preclude the formation of G-quadruplexes in r(G_4C_2)^{exp}, our results indicate that the hairpin is a relevant target for developing c9ALS/FTD chemical probes, as demonstrated by the ability of **4** to inhibit the two most prominent mechanisms of r(G_4C_2)^{exp} pathology.

We compared the bioactivity of **4** to many G-quadruplex ligands, as G-quadruplexes were previously reported to be the target of small molecules (Fay et al., 2017; Haeusler et al., 2014; Simone et al., 2018; Taylor, 2014; Zamiri et al., 2014). These ligands did not effectively inhibit RAN translation despite binding with low nM affinity to the G-quadruplex form of r(G_4C_2)^{exp}. An *in vitro* translation assay showed that the hairpin form of r$(G_4C_2)_{66}$ undergoes RAN translation, while the G-quadruplex does not (Figure S2D). Other studies have supported this finding: Rode et al. (2016) demonstrated that tRNA shifts the equilibrium between hairpin and G-quadruplex in mRNAs toward the hairpin and increases translational efficacy, while Guo and Bartel (2016) showed that RNA G-quadruplexes are globally unfolded in mammalian cells.

Perhaps the most well-studied G-quadruplex, in terms of both structure (Lim et al., 2009; Luu et al., 2006; Phan et al., 2007) and folding pathway (Ambrus et al., 2006; Gray and Chaires, 2008; Gray et al., 2014; Mashimo et al., 2008; Mashimo and Sugiyama, 2007), is the human telomere repeat sequence d$(TTAGGG)_n$. Although the final fold is dependent upon cation concentration and the loop sequence, three major species predominate. Two small molecules have been discovered that bind a particular telomerase fold; *N*-methylmesoporphyrin IX, which stabilizes a parallel topology (Sabharwal et al., 2014), and epiberberine, which binds the hybrid-2 quadurplex (Lin et al., 2018). Interestingly, the epiberberine-hybrid-2 complex is different than most other small molecule-G-quadruplex complexes (Collie et al., 2015) and perturbs the equilibria between quadruplex forms, converting other structures into the hybrid-2.

Collectively, our studies demonstrate that the hairpin is a therapeutically relevant structure involved in the pathological mechanisms of c9ALS/FTD (Cammas and Millevoi, 2017; Simone et al., 2015). Although a recent study reported that a small molecule reduces r(G_4C_2)^{exp} foci and inhibits RAN translation in iNeurons by targeting r(G_4C_2)^{exp} G-quadruplexes, the study did not investigate whether the hairpin structure could be a target for the compound (Simone et al., 2018). G-quadruplex structures are still a potential target to inhibit translation (Bugaut et al., 2012; Bugaut and Balasubramanian, 2012; Endoh et al., 2013; Murat et al., 2014) based on their high thermostability and structural hierarchy.

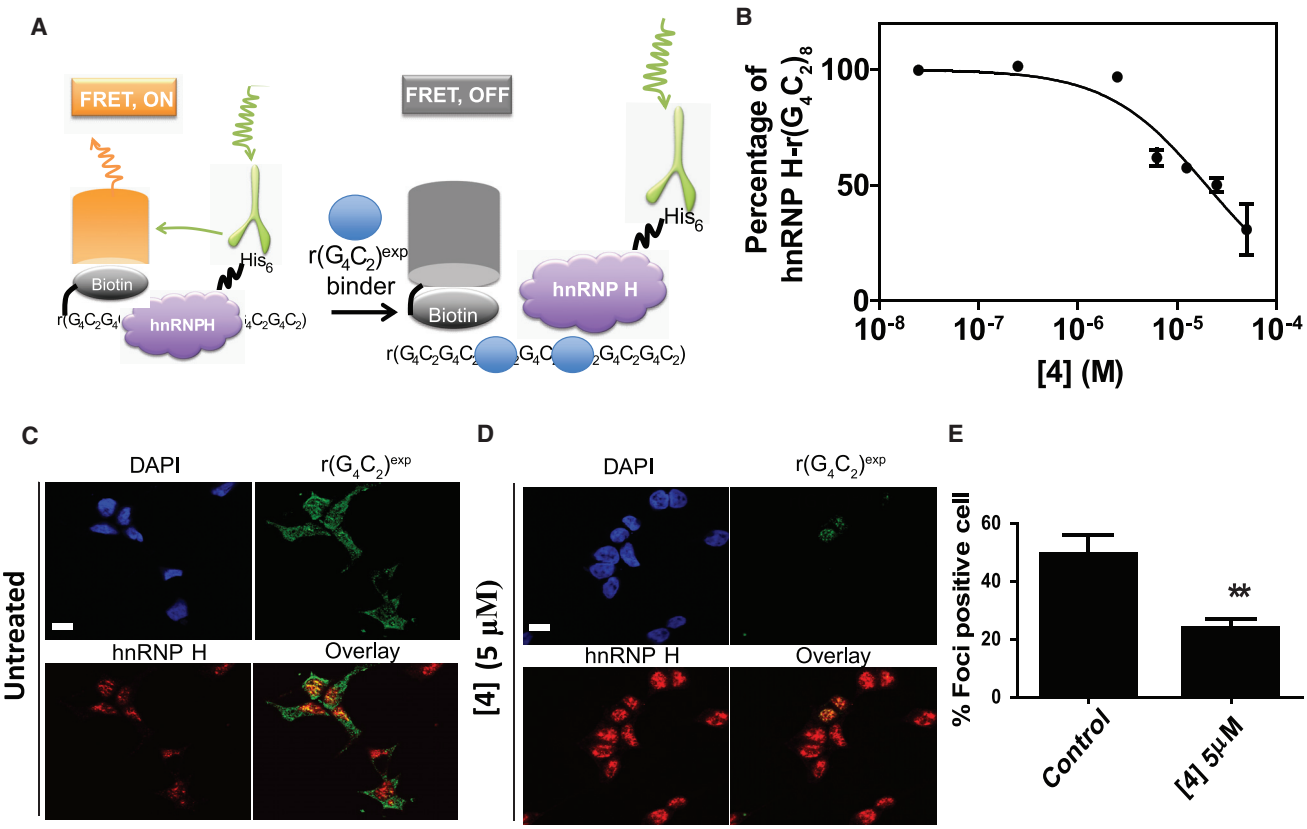


Figure 5. 4 Inhibits Binding of hnRNP H *In Vitro* and in HEK293T Cells Expressing (G_4C_2)$_{66}$-*No ATG-GFP*

(A) Schematic of the *in vitro* TR-FRET assay used to monitor small-molecule inhibition of the r(G_4C_2)$_8$-hnRNP H complex.

(B) Dose-response curve for **4** in the TR-FRET assay.

(C) Representative images of RNA fluorescence *in situ* hybridization (FISH) of r(G_4C_2)^{exp} and immunostaining of hnRNP H in untreated cells.

(D) Representative images of RNA FISH of r(G_4C_2)^{exp} and immunostaining of hnRNP H in **4**-treated cells.

(E) Quantification of foci-positive cells in panels C and D (n = 152; 3 biological replicates for untreated cells and n = 124; 3 biological replicates for **4**-treated cells). **p < 0.01 as determined by a two-tailed Student t test.

SIGNIFICANCE

The most common genetic cause of ALS and FTD is the expanded G_4C_2 [(G_4C_2)^{exp}] repeat in *C9ORF72*. The RNA transcribed from this repeat expansion directly causes "c9ALS/FTD" by sequestering RNA-binding proteins and producing toxic dipeptide repeats via RAN translation. Thus, r(G_4C_2)^{exp} is an important target for chemical probe and lead medicine development. The c9ALS/FTD RNA exists in equilibrium between two forms, a hairpin displaying 1 × 1 nucleotide GG internal loops and a G-quadruplex. We demonstrated that our small molecule, 4, binds the 1 × 1 nucleotide GG internal loop in the hairpin structure of r(G_4C_2)^{exp} and thereby inhibits two putative pathomechanisms of c9ALS/FTD. Surprisingly, known G-quadruplex ligands were poor inhibitors of RAN translation. This finding prompted an investigation into the energetics of r(G_4C_2)^{exp} folding, which revealed that the hairpin form is more kinetically favorable than the G-quadruplex. Further, *in vitro* translation assays showed that the hairpin, not the G-quadruplex, form of r(G_4C_2)^{exp}, undergoes RAN translation. Direct imaging using a fluorogenic hybrid molecule combining 4 with TO-PRO-1 demonstrated that: (1) 4 binds cellular r(G_4C_2)^{exp}, and (2) r(G_4C_2)^{exp} folds into the hairpin structure in cells. Collectively, our studies demonstrate that a small molecule targeting the hairpin structure can inhibit r(G_4C_2)^{exp}-associated toxicity. In summary, this work not only highlights the power of rationally designed chemical probes to identify and validate therapeutically relevant RNA structures in cells but also provides new insight into the role of r(G_4C_2)^{exp} structure in c9ALS/FTD biology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture
  - Authentication of Cell Lines Used
- METHOD DETAILS
  - Compounds
  - Oligonucleotides

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, three tables, and two data files and can be found with this article online at https://doi.org/10.1016/j.chembiol.2018.10.018.

ACKNOWLEDGMENTS

This work was funding by the NIH (DP1 NS096898-02 to M.D.D.; P01 NS099114-01 to M.D.D. and L.P.; R35NS097273 to L.P.), Target ALS (to M.D.D.), and the Nelson Family Fund (to M.D.D). Z.-F.W. is supported in part by the Academia Sinica-TSRI Postdoctoral Fellowship. A.U. acknowledges the Deutsche Forschungsgemeinschaft (DFG) for the DFG Postdoctoral Fellowship.

AUTHOR CONTRIBUTIONS

M.D.D. directed the study, conceived of the ideas, designed the experiments, and designed and generated the (G_4C_2)$_{66}$-No ATG-GFP and (G_4C_2)$_{66}$-No ATG-NanoLuc plasmids. Z.-F.W. designed and performed the G-quadruplex ligands study, NMR kinetic experiments, *in vitro* translation assay, and image study. A.U. performed BLI and polysome profiling experiments. A.U., V.B., W.-Y.Y., and R.F. synthesized the compounds. R.G. performed Western blot and TR-FRET experiments. I.Y. performed molecular modeling. S.G.R. and J.E.R. synthesized the G-quadruplex ligand, PyPx-2-DMA. Y.-J.Z., T.F.G., and L.P. generated (G_4C_2)$_{66}$ and (TG_3C_2)$_{62}$ plasmids and performed PolyGP activity assay. M.D.D., J.L.C.-D., and B.G.D. wrote the manuscript, edited by others.

DECLARATION OF INTERESTS

M.D.D. is a consultant for Expansion Therapeutics.

Received: March 21, 2018
Revised: July 16, 2018
Accepted: October 20, 2018
Published: November 29, 2018

REFERENCES

Almeida, S., Gascon, E., Tran, H., Chou, H.J., Gendron, T.F., Degroot, S., Tapper, A.R., Sellier, C., Charlet-Berguerand, N., Karydas, A., et al. (2013). Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. *126*, 385–399.

Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A., and Yang, D. (2006). Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res. *34*, 2723–2735.

Ash, P.E., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.L., Dejesus-Hernandez, M., van Blitterswijk, M.M., Jansen-West, K., Paul, J.W., 3rd, Rademakers, R., et al. (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron *77*, 639–646.

Bayly, C.I., Cieplak, P., Cornell, W.D., 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.

Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. *6*, 75–80.

Bugaut, A., and Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. *40*, 4727–4741.

Bugaut, A., Murat, P., and Balasubramanian, S. (2012). An RNA hairpin to G-quadruplex conformational transition. J. Am. Chem. Soc. *134*, 19953–19956.

Burger, A.M., Dai, F., Schultes, C.M., Reszka, A.P., Moore, M.J., Double, J.A., and Neidle, S. (2005). The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. Cancer Res. *65*, 1489–1496.

Cammas, A., and Millevoi, S. (2017). RNA G-quadruplexes: emerging mechanisms in disease. Nucleic Acids Res. *45*, 1584–1595.

Carreon, J.R., Mahon, K.P., Jr., and Kelley, S.O. (2004). Thiazole orange-peptide conjugates: sensitivity of DNA binding to chemical structure. Org. Lett. *6*, 517–519.

Case, D.A., Betz, R.M., Cerutti, D.S., Cheatham, T.E., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., et al. (2016). AMBER 16 (University of California).

Chen, C.Z., Sobczak, K., Hoskins, J., Southall, N., Marugan, J.J., Zheng, W., Thornton, C.A., and Austin, C.P. (2012). Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. Anal. Bioanal. Chem. *402*, 1889–1898.

Chew, J., Gendron, T.F., Prudencio, M., Sasaguri, H., Zhang, Y.J., Castanedes-Casey, M., Lee, C.W., Jansen-West, K., Kurti, A., Murray, M.E., et al. (2015). Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science *348*, 1151–1154.

Childs-Disney, J.L., Stepniak-Konieczna, E., Tuan, T., Yildirim, I., Park, H., Chen, C.Z., Hoskins, J., Southall, N., Marugan, J.J., Patnaik, S., et al. (2013). Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. Nat. Commun. *4*, 2044.

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.

Chou, M.Y., Rooke, N., Turck, C.W., and Black, D.L. (1999). hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. Mol. Cell. Biol. *19*, 69–77.

Collie, G.W., Campbell, N.H., and Neidle, S. (2015). Loop flexibility in human telomeric quadruplex small-molecule complexes. Nucleic Acids Res. *43*, 4785–4799.

Conlon, E.G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N.A., and Manley, J.L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife *5*, https://doi.org/10.7554/eLife.17820.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. *117*, 5179–5197.

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.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245–256.

Disney, M.D., Liu, B., Yang, W., Sellier, C., Tran, T., Charlet-Berguerand, N., and Childs-Disney, J.L. (2012). A small molecule that targets r(CGG)^{exp} and improves defects in fragile X-associated tremor ataxia syndrome. ACS Chem. Biol. *7*, 1711–1718.

Dominguez, C., Fisette, J.F., Chabot, B., and Allain, F.H. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. Nat. Struct. Mol. Biol. *17*, 853–861.

Donnelly, C.J., Zhang, P.W., Pham, J.T., Haeusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M., et al. (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron *80*, 415–428.

Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013). Translational halt during elongation caused by G-quadruplex formed by mRNA. Methods *64*, 73–78.

Fay, M.M., Anderson, P.J., and Ivanov, P. (2017). ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. Cell Rep. *21*, 3573–3584.

Franceschin, M., Rossetti, L., D'Ambrosio, A., Schirripa, S., Bianco, A., Ortaggi, G., Savino, M., Schultes, C., and Neidle, S. (2006). Natural and synthetic G-quadruplex interactive berberine derivatives. Bioorg. Med. Chem. Lett. *16*, 1707–1711.

Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.A., Vreven, T., Kudin, K.N., et al. (2004). Gaussian 03 (Gaussian).

Gendron, T.F., Bieniek, K.F., Zhang, Y.J., Jansen-West, K., Ash, P.E., Caulfield, T., Daughrity, L., Dunmore, J.H., Castanedes-Casey, M., Chew, J., et al. (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. *126*, 829–844.

Gendron, T.F., Chew, J., Stankowski, J.N., Hayes, L.R., Zhang, Y.J., Prudencio, M., Carlomagno, Y., Daughrity, L.M., Jansen-West, K., Perkerson, E.A., et al. (2017). Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. Sci. Transl Med. *9*, https://doi.org/10.1126/scitranslmed.aai7866.

Gray, R.D., and Chaires, J.B. (2008). Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. Nucleic Acids Res. *36*, 4191–4203.

Gray, R.D., Trent, J.O., and Chaires, J.B. (2014). Folding and unfolding pathways of the human telomeric G-quadruplex. J. Mol. Biol. *426*, 1629–1650.

Guo, J.U., and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science *353*, https://doi.org/10.1126/science.aaf5371.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195–200.

Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S., and Johnson, F.B. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res. *36*, 144–156.

Huang, F.C., Chang, C.C., Lou, P.J., Kuo, I.C., Chien, C.W., Chen, C.T., Shieh, F.Y., Chang, T.C., and Lin, J.J. (2008). G-quadruplex stabilizer 3,6-bis(1-methyl-4-vinylpyridinium)carbazole diiodide induces accelerated senescence and inhibits tumorigenic properties in cancer cells. Mol. Cancer Res. *6*, 955–964.

Jiang, J., Zhu, Q., Gendron, T.F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J.E., Jafar-Nejad, P., Drenner, K., Schulte, D., et al. (2016). Gain of toxicity from ALS/FTD-linked repeat expansions in C9ORF72 is alleviated by antisense oligonucleotides targeting GGGGCC-containing RNAs. Neuron *90*, 535–550.

Korobov, V., and Ochkov, V. (2011). Chemical Kinetics with Mathcad and Maple (Springer-Verlag), pp. 35–72.

Kovanda, A., Zalar, M., Sket, P., Plavec, J., and Rogelj, B. (2015). Anti-sense DNA d(GGCCCC)$_n$ expansions in C9ORF72 form i-motifs and protonated hairpins. Sci. Rep. *5*, 17944.

Kuo, M.H., Wang, Z.F., Tseng, T.Y., Li, M.H., Hsu, S.T., Lin, J.J., and Chang, T.C. (2015). Conformational transition of a hairpin structure to G-quadruplex within the WNT1 gene promoter. J. Am. Chem. Soc. *137*, 210–218.

Lane, A.N., Chaires, J.B., Gray, R.D., and Trent, J.O. (2008). Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. *36*, 5482–5515.

Lee, Y.B., Chen, H.J., Peres, J.N., Gomez-Deza, J., Attig, J., Stalekar, M., Troakes, C., Nishimura, A.L., Scotter, E.L., Vance, C., et al. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. *5*, 1178–1186.

Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N., and Phan, A.T. (2009). Structure of the human telomere in K$^+$ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. J. Am. Chem. Soc. *131*, 4301–4309.

Lin, C., Wu, G., Wang, K., Onel, B., Sakai, S., Shao, Y., and Yang, D. (2018). Molecular recognition of the hybrid-2 human telomeric G-quadruplex by epiberberine: insights into conversion of telomeric G-quadruplex structures. Angew. Chem. Int. Ed. *57*, 10888–10893.

Liu, W., Ikeda, Y., Hishikawa, N., Yamashita, T., Deguchi, K., and Abe, K. (2014). Characteristic RNA foci of the abnormal hexanucleotide GGCCUG repeat expansion in spinocerebellar ataxia type 36 (Asidan). Eur. J. Neurol. *21*, 1377–1386.

Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L., and Patel, D.J. (2006). Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. J. Am. Chem. Soc. *128*, 9963–9970.

Marky, L.A., and Breslauer, K.J. (1987). Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. Biopolymers *26*, 1601–1620.

Mashimo, T., Sannohe, Y., Yagi, H., and Sugiyama, H. (2008). Folding pathways of hybrid-1 and hybrid-2 G-quadruplex structures. Nucleic Acids Symp. Ser. (Oxf.) *52*, 409–410.

Mashimo, T., and Sugiyama, H. (2007). Folding pathways of human telomeric hybrid G-quadruplex structure. Nucleic Acids Symp. Ser. (Oxf.) *41*, 239–240.

Matsuzono, K., Imamura, K., Murakami, N., Tsukita, K., Yamamoto, T., Izumi, Y., Kaji, R., Ohta, Y., Yamashita, T., Abe, K., et al. (2017). Antisense oligonucleotides reduce RNA foci in spinocerebellar ataxia 36 patient iPSCs. Mol. Ther. Nucleic Acids *8*, 211–219.

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C.E., Clayton, E.L., Devoy, A., Moens, T., Norona, F.E., Woollacott, I.O., Pietrzyk, J., et al. (2014). C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. Science *345*, 1192–1194.

Mizielinska, S., and Isaacs, A.M. (2014). C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia: gain or loss of function? Cur. Opin. Neurol. *27*, 515–523.

Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L., and Bryan, T.M. (2015). Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. Nat. Commun. *6*, 7643.

Murat, P., Zhong, J., Lekieffre, L., Cowieson, N.P., Clancy, J.L., Preiss, T., Balasubramanian, S., Khanna, R., and Tellam, J. (2014). G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. Nat. Chem. Biol. *10*, 358–364.

O'Rourke, J.G., Bogdanik, L., Muhammad, A.K., Gendron, T.F., Kim, K.J., Austin, A., Cady, J., Liu, E.Y., Zarrow, J., Grant, S., et al. (2015). C9orf72 BAC transgenic mice display typical pathologic features of ALS/FTD. Neuron *88*, 892–901.

Obayashi, M., Stevanin, G., Synofzik, M., Monin, M.L., Duyckaerts, C., Sato, N., Streichenberger, N., Vighetto, A., Desestret, V., Tesson, C., et al. (2015). Spinocerebellar ataxia type 36 exists in diverse populations and can be caused by a short hexanucleotide GGCCTG repeat expansion. J. Neurol. Neurosurg. Psychiatry *86*, 986–995.

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.

Phan, A.T., Kuryavyi, V., Luu, K.N., and Patel, D.J. (2007). Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K^+ solution. Nucleic Acids Res. *35*, 6517–6525.

Prudencio, M., Belzil, V.V., Batra, R., Ross, C.A., Gendron, T.F., Pregent, L.J., Murray, M.E., Overstreet, K.K., Piazza-Johnston, A.E., Desaro, P., et al. (2015). Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat. Neurosci. *18*, 1175–1182.

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257–268.

Rode, A.B., Endoh, T., and Sugimoto, N. (2016). tRNA shifts the G-quadruplex-hairpin conformational equilibrium in RNA towards the hairpin conformer. Angew. Chem. Int. Ed. *55*, 14315–14319.

Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310.

Rzuczek, S.G., Pilch, D.S., Liu, A., Liu, L., LaVoie, E.J., and Rice, J.E. (2010). Macrocyclic pyridyl polyoxazoles: selective RNA and DNA G-quadruplex ligands as antitumor agents. J. Med. Chem. *53*, 3632–3644.

Sabharwal, N.C., Savikhin, V., Turek-Herman, J.R., Nicoludis, J.M., Szalai, V.A., and Yatsunyk, L.A. (2014). N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. FEBS J. *281*, 1726–1737.

Siddiqui-Jain, A., Grand, C.L., Bearss, D.J., and Hurley, L.H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc. Natl. Acad. Sci. U S A *99*, 11593–11598.

Simone, R., Balendra, R., Moens, T.G., Preza, E., Wilson, K.M., Heslegrave, A., Woodling, N.S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S., et al. (2018). G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. EMBO Mol. Med. *10*, 22–31.

Simone, R., Fratta, P., Neidle, S., Parkinson, G.N., and Isaacs, A.M. (2015). G-quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome. FEBS Lett. *589*, 1653–1668.

Su, Z., Zhang, Y., Gendron, T.F., Bauer, P.O., Chew, J., Yang, W.Y., Fostvedt, E., Jansen-West, K., Belzil, V.V., Desaro, P., et al. (2014). Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. Neuron *83*, 1043–1050.

Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C.Z., Lin, C.C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., et al. (2018). CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. Nat. Commun. *9*, 152.

Taylor, J.P. (2014). Neurodegenerative diseases: G-quadruplex poses quadruple threat. Nature *507*, 175–177.

Taylor, J.P., Brown, R.H., Jr., and Cleveland, D.W. (2016). Decoding ALS: from genes to mechanism. Nature *539*, 197–206.

Tran, T., Childs-Disney, J.L., Liu, B., Guan, L., Rzuczek, S., and Disney, M.D. (2014). Targeting the r(CGG) repeats that cause FXTAS with modularly assembled small molecules and oligonucleotides. ACS Chem. Biol. *9*, 904–912.

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

Wang, J.M., Huang, F.C., Kuo, M.H., Wang, Z.F., Tseng, T.Y., Chang, L.C., Yen, S.J., Chang, T.C., and Lin, J.J. (2014). Inhibition of cancer cell migration and invasion through suppressing the Wnt1-mediating signal pathway by G-quadruplex structure stabilizers. J. Biol. Chem. *289*, 14612–14623.

Wang, J.M., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2005). Development and testing of a general amber force field, (vol 25, pg 1157, 2004). J. Comput. Chem. *26*, 114.

Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Santos, N.D., Silvester, J., Wei, V., Garcia, J., et al. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nat. Commun. *8*, 14432.

Yang, W.Y., He, F., Strack, R.L., Oh, S.Y., Frazer, M., Jaffrey, S.R., Todd, P.K., and Disney, M.D. (2016). Small molecule recognition and tools to study modulation of r(CGG)^{exp} in fragile X-associated tremor ataxia syndrome. ACS Chem. Biol. *11*, 2456–2465.

Yang, W.Y., Wilson, H.D., Velagapudi, S.P., and Disney, M.D. (2015). Inhibition of non-ATG translational events in cells via covalent small molecules targeting RNA. J. Am. Chem. Soc. *137*, 5336–5345.

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.

Zamiri, B., Reddy, K., Macgregor, R.B., Jr., and Pearson, C.E. (2014). TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. J. Biol. Chem. *289*, 4653–4659.

Zhang, A.Y., and Balasubramanian, S. (2012). The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. J. Am. Chem. Soc. *134*, 19297–19308.

Zhang, J.H., Chung, T.D., and Oldenburg, K.R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J. Biomol. Screen. *4*, 67–73.

Zu, T., Liu, Y., Banez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T.M., Harms, M.B., Falchook, A.E., Subramony, S.H., et al. (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc. Natl. Acad. Sci. U S A *110*, E4968–E4977.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-hnRNP H	Novus Biologicals	Cat#: NBP131648; RRID: 10003534
anti-G-quadruplex, BG4	Sigma-Aldrich	Cat#: MABE917; RRID: 2750936
Chemicals, Peptides, and Recombinant Proteins		
1-40	NCI	N/A
TOQ	Carreon et al., 2004	N/A
1a-amine	Su et al., 2014	N/A
1a-TOQ	This study	N/A
i1-i8	This study	N/A
4-TO-PRO	This study	N/A
CHX (cycloheximide)	Sigma	Cat#: 104450
TO-PRO-1	ThermoFisher	Cat#: T3602
TMPyP4	FisherScientific	Cat#42-535-0
PDS	Sigma	Cat#SML0678
cPDS	Sigma	Cat#SML1176
BER	Sigma	Cat#B3251
Braco-19	Sigma	Cat#SML0560
CX-5461	AdooQ Bioscience	Cat#A11065
NMM	Frontier Scientific	Cat#NMM580
PyPx-2-DMA	Rzuczek et al., 2010	N/A
BMVC	Gift	Dr. Ta-Chau Chang (IAMS, Academia Sinica)
Critical Commercial Assays		
In vitro transcription assay	Promega	Cat# P1320
In vitro translation assay	Promega	Cat# L4540
TR-FRET assay(StreptavidinXL665)	Cisbio	Cat# 610SAXLB
TR-FRET assay (MAb Anti6His-Tb-cryptate)	Cisbio	Cat# 61HISTLB
Experimental Models: Cell Lines		
HEK293T	ATCC	https://www.atcc.org/Products/All/CRL-3216.aspx
Oligonucleotides		
DNA	IDT	N/A
RNA	Dharmoncon	N/A
Recombinant DNA		
$(G_4C_2)_{66}$	Su et al., 2014	N/A
$(TG_4C2)_{66}$	This study	N/A
$(G_4C_2)_{66}$-NoATG-GFP	This study	N/A
$(G_4C_2)_{66}$-NoATG-Nanoluc	This study	N/A
mCherry	Addgene	Cat#54563
pet15b expression vector encoding His-tagged hnRNP H1	Addgene	Cat#2302
Software and Algorithms		
GraphPad 7.03	Prism	https://www.graphpad.com/
Origin 7.5	Originlab	https://www.originlab.com/
Topspin 2.1	Bruker	https://www.bruker.com/products.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to the corresponding author, Prof. Matthew Disney (disney@scripps.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture

HEK293T cells (female) were cultured in 1× DMEM containing 4.5 g/L glucose, 10% FBS, 2 mM L-glutamine, and 1× penicillin/streptomycin (growth medium) at 37°C in 5% CO_2.

Authentication of Cell Lines Used

HEK293T cells were purchased from ATCC and used without further authentication.

METHOD DETAILS

Compounds

All **1a**-like compounds were acquired from the National Cancer Institute (NCI). Compound identity was confirmed by mass spectroscopy and compound purity was assessed by analytical HPLC. All compounds were >95% pure. Methods describing the synthesis of **1a-TO$_Q$** and **4-TO-PRO** are provided in "Synthetic Methods", and their characterization is provided in Data S2. The G-quadruplex ligands, **PDS**, **cPDS** (*carboxyl*PDS), **BER** (Berberine) and **Braco-19** were obtained from Sigma. **CX-5461** was obtained from Adooq Bioscience. The **NMM** (*N*-Methyl mesoporphyrin IX) and **TMPyP4** compounds were obtained from Frontier Scientific and ThermoFisher, respectively. **PyPx-2-DMA** was synthesized as previously described (Rzuczek et al., 2010). The **BMVC** derivatives were provided by Dr. Ta-Chau Chang (IAMS, Academia Sinica).

Oligonucleotides

All oligonucleotides used in this study have been purchased from GE Dharmacon and de-protected according to the manufacturer's recommended procedure. Vivo-Morpholino oligonucleotide was acquired from Gene Tools.

Plasmids

The $(G_4C_2)_{66}$, $(G_4C_2)_{66}$-NoATG-GFP and $(G_4C_2)_{66}$-NoATG-Nanoluc plasmids were constructed using previously described methods (Gendron et al., 2013; Yang et al., 2015). To generate the $(TG_3C_2)_{62}$ plamsid, gDNA from fibroblasts of a Sca36+ patients was used as a templates in a nested PCR strategy using ThermalAce DNA Polymerase (Invitrogen) or AmpliTaq Gold 360 Polymerase (Thermo Fisher). The sequence includes 69 bp 5' of the repeat expansion and 40 bp of 3' flanking sequence. The PCR products were cloned into the pAG3 expression vector (gift of T. Golde, University of Florida), then sequentially ligated using TypeIIS restriction enzymes to generate a $(TG_3C_2)_{62}$ fragment. All $(TGGGCC)_n$ fragments with 5' and 3' flanking sequences were subcloned into the pAG_3 expression vector containing two upstream stop codons in each reading frame, as well as three different C-terminal tags in alternate frames [i.e., (GP)n-HA, (GL)n-Myc and (WA)n-FLAG].

Affinity of 1a-TO$_Q$ for r(G_4C_2)$_8$

The affinity of **1a-TO$_Q$** for r(G_4C_2)$_8$ was completed by measuring the fluorescence intensity of **1a-TO$_Q$** as a function of r(G_4C_2)$_8$ concentration. Briefly, the RNA was folded in 1× Assay Buffer (8 mM K_2HPO_4, pH 7.0, 185 mM KCl, and 1 mM EDTA) by heating at 95°C for 3−5 min and slowly cooling to room temperature for 15 min. BSA and **1a-TO$_Q$** were added to final concentration of 40 μg/mL and 1.2 μM, respectively. The RNA was then serially diluted into 1× Assay Buffer supplemented with 40 μg/mL BSA and 1.2 μM **1a-TO$_Q$**. The resulting curve was fit to a one-site binding model to afford the K_d (Data S1).

Dye Displacement Assay to Study Binding of Compounds to r(G_4C_2)$_8$

The RNA, r(G_4C_2)$_8$, (600 nM final concentration) was folded as described above. Then BSA and **1a-TO$_Q$** were added to final concentration of 40 μg/mL and 1.2 μM, respectively, followed by the compound of interest. The samples were incubated for 15 min at room temperature and fluorescence was measured (excitation/emission wavelengths of 480/530 nm) using a BioTek FLx800 plate reader. Background (emission of **1a-TO$_Q$** in absence of RNA) was subtracted, and the signal was normalized to **1a-TO$_Q$** complex with RNA in the absence of compound.

IC_{50}s were measured analogously by completing serial dilutions of the compound of interest in 1× Assay Buffer containing 600 nM r(G_4C_2)$_8$, 1.2 μM **1a-TO$_Q$**, and 40 μg/mL BSA. IC_{50}s were calculated from normalized data using a four-parameter logistic curve fit (SigmaPlot).

Biolayer Interferometry (BLI)

BLI studies were performed similary a previously reported procedure (Su et al., 2014) on an Octet RED96 system. Briefly, 5′-biotinylated RNAs (60 nM) were folded by heating at 95°C for 5 min followed by slowly cooling to room temperature in 1× BLI Buffer

(80 mM KH_2PO_4, pH 7.0, 185 mM KCl, 1 mM EDTA). Separately, serial dilutions of the compound of interest (ranging from 0.5 μM to 25 μM) were prepared in 1× BLI Buffer. The following incubation times were used during data acquisition (30°C with shaking at 1000 rpm): baseline step (180 s), loading of RNA (900 s), washing (180 s), association of compound (500 s), and dissociation of compound (350 s). The resulting curves were analyzed and processed using the Octet Data Analysis version 7.1 software by subtracting the response of the sensors recorded upon incubation with solutions containing compound and no RNA (parallel reference). The data were then globally fitted assuming reversible binding using the entire time interval for association and dissociation using the 2:1 heterogeneous binding or 1:1 binding curve fitting model.

Recombinant hnRNP H1-His_6 Production, Isolation, and Purification

The pet15b expression vector encoding His-tagged hnRNP H1 was obtained from Addgene (ID: 23020) (Chou et al., 1999) and transformed into Rosetta 2 (DE3) *Escherichia coli* cells. The transformed cells were grown in LB medium at 37°C to OD_{600} = 0.6 before induction with 0.5 mM IPTG and then cultured at 25°C for 4 h. Cells were pelleted and lysed in 20 mM Tris-HCl, pH 7.5, and 500 mM NaCl (Buffer A) via sonication. The lysate was centrifuged and the pellet was washed with lysis buffer before it was solubilized in 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 6 M Urea (Buffer B). After centrifugation, the supernatant was filtered and loaded onto His60 Ni Superflow resin (Clontech). The column was washed with 5 column volumes (CVs) of Buffer B. hnRNP H1 was refolded on the column with a gradient from 100% Buffer B to 100% Buffer A with 20 CVs at a flow rate of 0.1 mL/min. Then the column was washed with 10 CVs of 90% Buffer B and 10% Buffer C (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 500 mM imidazole). The refolded protein was eluted with a gradient to 100% Buffer C over 10 CVs.

Affinity of hnRNP H1 for r(G_4C_2) Repeats

The affinity of hnRNP H1 for r(G_4C_2) repeats was measured by a gel shift assay. The RNA was 5′ end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and gel purified as previously described (Su et al., 2014). Briefly, r(G_4C_2)$_8$ was folded in 1× Folding Buffer (20 mM HEPES, pH 7.5, 110 mM KCl, 10 mM NaCl) by heating to 95°C for 5 min followed by slowly cooling to room temperature. The buffer was adjusted to 1×TR-FRET Assay Buffer (20 mM HEPES, pH 7.5, 110 mM KCl, 10 mM NaCl, 0.1% (w/v) BSA, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 0.05% Tween-20, 5 mM DTT). The RNA was aliquoted and hnRNP H1-His_6 (360 nM final concentration) was added to the first aliquot followed by 1:2 serial dilutions. The samples were incubated at room temperature for 1 h. The RNA and RNA-hnRNP H1-His_6 complex were separated on a native 5% polyacrylamide gel prepared with 1×TBE. The percentage of complex formation as a function of hnRNP H1-His_6 concentration was fit to a one-site binding model to afford the K_d.

TR-FRET hnRNP H1 Binding Assay

The TR-FRET assays were performed as previously described (Disney et al., 2012) except the final concentrations of biotinylated r(G_4C_2)$_8$ RNA and hnRNP H1-His_6 were 80 nM and 75 nM, respectively. The RNA was folded by heating to 95°C for 5 min in 1× Folding Buffer followed by slowly cooling to room temperature. The buffer was adjusted to 1×TR-FRET Assay Buffer followed by addition of compound. The mixture was incubated for 15 min at room temperature and then hnRNP H1-His_6 was added. After an additional 15 min incubation, streptavidin-XL665 (HTRF, Cisbio Bioassays) and anti-His_6-Tb (HTRF, Cisbio Bioassays) were added to a final concentration of 40 nM and 0.44 ng/μL, respectively. After incubating for 1 h, TR-FRET was measured as previously described. IC_{50}s were calculated by fitting the resulting curve Equation 1 with GraphPad Prism.

$$y = B + \frac{A - B}{1 + \left(\frac{IC_{50}}{x}\right)^{hillslope}} \quad \text{(Equation 1)}$$

Verification of RAN Translation Product in (G_4C_2)$_{66}$–NoATG–GFP Plasmid by Western Blot

HEK293T cells were plated in 60 mm cell culture dishes and transfected with 4 μg (G_4C_2)$_{66}$–*NoATG–GFP or ATG–GFP* plasmid using JetPrime transfection reagent (Polyplus). After 4 h, cells were trypsinized and plated into 12-well plates in growth medium. After an additional 2 h, the compound of interest or vehicle was added in growth medium. Cells were lysed in the plate using 300 μL/well of mammalian protein extraction reagent (MPER, Pierce Biotechnology) containing 1 μL of halt protease inhibitor cocktail (Thermo Scientific). Cellular proteins were separated by SDS-PAGE and then transferred to a PVDF membrane. Western blotting was completed using anti-GFP (Santa Cruz) as a primary antibody and anti-IgG-horseradish peroxidase conjugate as the secondary antibody. Chemiluminescent signal was generated by SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific), and the blot was imaged using X-ray film.

Cell-Based Assay to Assess RAN Translation and Poly(GP) Response

HEK293T cells (80% confluent) were batch-transfected in growth medium for 4 h with (G_4C_2)$_{66}$–*No ATG–GFP* and a plasmid encoding mCherry (for normalization) using JetPrime transfection reagent (Polyplus) according to the manufacturer's instructions. The cells were then seeded into a 384-well plate and incubated for 2 h before addition of compound using Biomek NXP Laboratory Automation

Workstation that was equipped with a 100 nL 384-pin head. After treatment for 24 h, the medium was removed, and the cells were lysed in 100 mM potassium phosphate lysis buffer, pH 7.8, and 0.2% 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. Background was determined by measuring the corresponding fluorescence intensities in untransfected cells. The background-corrected ratio of GFP to mCherry was used to determine the effect of compound on RAN translation. Poly(GP) measurements were performed as previously described (Su et al., 2014). HEK293T cells were batch-transfected with $(G_4C_2)_{66}$ *or* $(TG_4C_2)_{62}$ plasmid and treated with different concentrations of **4**.

Effect of Compound Treatment on $(G_4C_2)_{66}$–No ATG–GFP Transcript Levels

Approximately 2×10^6 HEK293T cells were plated in 60 mm cell culture dishes and transfected with 3.5 μg $(G_4C_2)_{66}$–*No ATG–GFP* plasmid using JetPrime transfect agent per manufacturer's protocol. After 4 h, cells were trypsinized and plated into 48-well plates in growth medium. After an additional 2 h, the compound of interest or vehicle was added in growth medium. After overnight incubation, the compound-containing medium was removed, and cells were washed with 1× Dulbecco's phosphate-buffered saline (DPBS). Total RNA was harvested with a Quick RNA^{TM} Mini-Prep Kit (Zymo Research) per the manufacturer's recommended protocol.

The cDNA was generated from 100 ng RNA using a QScript cDNA Synthesis Kit (Quanta Biosciences) according to manufacturer's recommended procedure; qPCR was completed with Power SYBR Green PCR Master Mix (Life Technologies) and 1 μL of 10-fold diluted cDNA on an Applied Biosystems 7900HT Fast Real Time PCR System. The levels of $(G_4C_2)_{66}$–*No ATG–GFP* were measured with primers specific for GFP and normalized to 18S or β-actin values. Table S3 lists the sequences of all PCR primers.

CD Spectroscopy

CD spectroscopic experiments were conducted using a J-815 Jasco spectropolarimeter with a bandwidth of 2 nm at a scan speed of 50 nm/min and a step resolution of 0.2 nm over the spectral range of 210–450 nm. The RNA sample concentration was 10 to 20 μM in 10 mM Na_2HPO_4 buffer with 50 mM LiCl. Thermal melting curves were recorded by a peltier thermal coupler chamber (Jasco), monitored at 265 nm between 20 and 95°C with a temperature ramping rate of 1°C/min. The observed signals were baseline subtracted, and the first derivative zero points were defined as the melting temperature. Shape analysis of the melting curves yielded Van't Hoff enthalpy (ΔH_{vH}) using standard procedures (Marky and Breslauer, 1987). The cooperativity of **4** binding to r(G_4C_2) repeats was measured by plotting the normalized signal intensity at 300 nm or 330 nm in CD spectra as a function of the molar ratio of ([**4**]/[RNA]) and fitted to the following equation:

$$y = x_0 + (x^{n_{Hill}})/(x^{n_{Hill}} + K^{n_{Hill}}) \qquad \text{(Equation 2)}$$

where y is the normalized signal intensity, x_o is initial molar ratio of ([**4**]/ [RNA]); x is the molar ratio of ([**4**]/ [RNA]), n_{Hill} is Hill coefficient, and K is the Hill binding constant.

NMR Spectroscopy

All NMR spectroscopic experiments were performed on a Bruker AVIII 700 MHz NMR spectrometer equipped a cryoprobe (Bruker, USA). 1D imino proton NMR spectra were recorded by a WATERGATE pulse sequence for water suppression. RNA concentrations were typically 0.1 mM for 1D experiments and 0.4 mM for the 2D experiments, and studies were completed in 10 mM sodium phosphate buffer with 100 mM LiCl. Ligand were prepared 10 mM in D_2O or d_6-DMSO, and then added into a solution of RNA. For kinetic experiments, RNA samples were prepared in 10 mM sodium phosphate buffer with 50 mM LiCl as the hairpin folded state, followed by addition of 150 mM KCl. The time for recording an individual NMR spectrum was 30 min. Kinetic curves were fit to a three-state kinetic model and were analyzed by Origin 7.0 and Topspin 2.0 (see below). An internal reference of 0.1 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid was used. The resonances of exchangeable protons were assigned using 2D NOESY with a mixing time of 400 ms.

Analytical Solution of the Three-State Transition Model

A three-state transition model for hairpin state as H, unfolding state as U, and G-quadruplex state as Q was previously described (Korobov and Ochkov, 2011) and is as follows:

$$H(t) \underset{k_2}{\overset{k_1}{\rightleftarrows}} U(t) \underset{k_4}{\overset{k_3}{\rightleftarrows}} Q(t)$$

$$\frac{d[H]}{dt} = -k_1[H] + k_2[U]$$

$$\frac{d[U]}{dt} = k_1[H] + k_4[Q] - (k_2 + k_3)[U] \qquad \text{(Equation 3.1)}$$

$$\frac{d[Q]}{dt} = k_3[U] - k_4[Q]$$

Where k_1 and k_2 are the forward and reverse transition rates associated with the first step between the initial and intermediate states, respectively, while k_3 and k_4 are the forward and reverse transition rates associated with the second step between the intermediate and final states, respectively. The analytical solution for this three-state transition model is given as:

$$H(t) = H_0\left[\frac{k_2k_4}{\gamma_1\gamma_2} - \frac{k_1(\gamma_1 - k_3 - k_4)}{\gamma_1(\gamma_2 - \gamma_1)}e^{-\gamma_1 t} - \frac{k_1(k_3 + k_4 - \gamma_2)}{\gamma_1(\gamma_2 - \gamma_1)}e^{-\gamma_2 t}\right] \quad \text{(Equation 3.2)}$$

$$U(t) = H_0k_1\left[\frac{k_4}{\gamma_1\gamma_2} - \frac{k_4 - \gamma_1}{\gamma_1(\gamma_2 - \gamma_1)}e^{-\gamma_1 t} - \frac{k_4 - \gamma_2}{\gamma_2(\gamma_2 - \gamma_1)}e^{-\gamma_2 t}\right] \quad \text{(Equation 3.3)}$$

$$Q(t) = H_0k_1k_3\left[\frac{1}{\gamma_1\gamma_2} - \frac{1}{\gamma_1(\gamma_2 - \gamma_1)}e^{-\gamma_1 t} - \frac{1}{\gamma_2(\gamma_2 - \gamma_1)}e^{-\gamma_2 t}\right] \quad \text{(Equation 3.4)}$$

Where $\gamma_1\gamma_2 = k_1k_3+k_2k_4+k_1k_4$ and $\gamma_1 + \gamma_2 = k_1 + k_2 + k_3+ k_4$. The initial conditions were $H(0) = H_0$, $U(0) = Q(0) = 0$. The experimental quantities associated with the initial and final states were extracted from the time-resolved spectra, which were used as input for nonlinear regression to extract k_1, k_2, k_3, and k_4 in the Origin 7.5 software (OriginLab Corp., Northampton, MA, USA).

Kinetic data were obtained by analytical solution of the three-state transition model, where normalized kinetic curves showed the imino proton signals of initial state of hairpin (H) and final state of G-quadruplex (Q) as a function of time, which were fitted by Equations 3.2 and 3.4.

RNA FISH and Immunostaining

Nuclear Foci

The RNA FISH and hnRNP H immunostaining were completed as previously described with modifications (Lee et al., 2013). Briefly, HEK293T cells (80% confluent) were batch-transfected as described above. The cells were then seeded into 12-well plates with coverslips and incubated for 2 h before addition of compound. After incubation with compound for 24 h, the cells were fixed in 4% paraformaldehyde for 15 min and washed 5 times with 1× DPBS. After washing with 1× DPBS for 15 min at room temperature, the cells were washed with 0.1% Triton X-100 in 1× DPBS for 5 min at room temperature followed by washing with 40% formamide in 2× SSC buffer for 15 min at room temperature. The RNA FISH probe was then added (5 ng/μL of 5'-Cy3-d(G_2C_4)$_8$-3') in 40% formamide in 2× SSC containing 8 μg/mL BSA, 66 μg/mL yeast tRNA, and 2 mM vanadyl complex. The samples were incubated at 48°C and 5% CO_2 overnight. Following overnight hybridization, the cells were washed three times with 2× SSC and then three times with 1× DPBS at room temperature for 15 min each. To immunostain hnRNP H, cells were fixed with 2% (v/v) formaldehyde in 1× DPBS, permeabilized with 0.1% (v/v) Triton X-100 in 1× DPBS, blocked with 5% goat serum for 40 min, and then incubated with a 1:250 dilution of anti-hnRNP H (Novus Bio) in 1× DPBS for 1 h at room temperature. After washing with 1× DPBS twice, the cells were incubated with a 1:250 dilution anti-rabbit IgG DyLight 650 in 1× DPBS at room temperature for 1 h. After washing (1× DPBS three times at room temperature for 15 min each), cells were stained with 1 μg/mL DAPI in 1× DPBS for 10 min at room temperature and then washed again with 1× DPBS immediately before imaging. Images were collected using an Olympus Fluoview 1000 confocal microscope. A cell was considered foci-positive if green (RNA) and red (hnRNP H) fluorescence signal overlapped in punctate spots. "% Foci-positive Cells" was therefore calculated by dividing the number of foci-positive cells by the total number of cells analyzed and multiplying by 100.

Foci Imaging by 4-TO-PRO and BG4, a G-Quadruplex Antibody

Briefly, HEK293T cells (80% confluent) were batch-transfected as described above. The cells were then seeded into 12-well plates with coverslips and incubated for 2 h before addition of compound. After incubation with compound for 24 h, the cells were fixed in 2% paraformaldehyde for 15 min and washed five times with 1× DPBS. The cells were then washed with 0.1% Triton X-100 in 1× DPBS for 5 min and three times with 1× DPBS. **4-TO-PRO** (0.5 μM) was added to the cells for 2 h at 37°C.

For experiments employing **4** or RNase, the cells were first incubated with **4** (5 μM for 24 h) or RNase (25 mg/mL in 1× DPBS) at 37°C for 2 h and washed two times with 1× DPBS before treatment with **4-TO-PRO** (5 μM for 24 h). For BG4 studies (Moye et al., 2015), the cells were fixed with 2% (v/v) formaldehyde in 1× DPBS, permeabilized with 0.1% (v/v) Triton X-100 in PBS, and blocked with 2% (w/v) milk in 1× DPBS. After washing with 1× DPBS twice, the slides were incubated with BG4 primary antibody (600 nM) for 1 h at 37°C, and washed three times for 5 min with 0.01% Tween-20 in 1× DPBS, and then overlaid with secondary antibody (Rabbit anti-flag Tag antibody, Cell Signaling; 1:800 dilution) for 1 h at 37°C. After washing three times with 0.01% Tween-20 in 1× DPBS, the cells were incubated with anti-rabbit IgG DyLight 650. The cells were stained with 1 μg/mL DAPI in 1× DPBS for 10 min at room temperature and then washed with 1× DPBS. Images were collected using an Olympus Fluoview 1000 confocal microscope. A cell was considered **4-TO-PRO** foci-positive if the cell contained punctate spots with green fluorescence. "% **4-TO-PRO** Foci-positive Cells" was therefore calculated by dividing the number of **4-TO-PRO** foci-positive cells by the total number of cells analyzed and multiplying by 100.

Polysome Profiling

Polysome profiling studies were completed similarly to previously described (Yang et al., 2015, 2016). Briefly, 100 mm dishes were plated with 4.5×10^6 HEK293T cells and batch-transfected as described above. Approximately 2 h after plating, cells were treated with vehicle or **4** (5 μM) and incubated for 16-18 h. Cells were then lysed with 250 μL of ice-cold Cell Lysis Buffer [10 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris, pH 7.5, 1% Triton X-100, 1% sodium deoxycholate, 1 mM DTT supplemented with 0.1 mg/L cycloheximide and 0.2 U/μL RNAsin (Promega)]. The lysate was transferred to an Eppendorf tube, gently vortexed and centrifuged for 5 min at 13,200 rpm and 4°C. The supernatant was transferred to a new tube and stored at −80°C until further use.

Cellular lysates were separated using a 10-50% sucrose gradient in 500 μL fractions. A 100 μL aliquot from each fraction was removed and the RNA isolated by using a Quick RNA™ Mini-Prep (Zymo Research) according to manufacturer's protocol. cDNA was generated from 250 ng of RNA using a QScript cDNA Synthesis Kit according to manufacturer's recommended procedure. The levels of $(G_4C_2)_{66}$–*No ATG–GFP* mRNA in each fraction were measured by qPCR as described above with primers specific for GFP (Table S3). Data from vehicle- and compound-treated samples were normalized as follows: triplicate C_t values were averaged and ΔC_t values were calculated by comparison to β-actin; $\Delta\Delta C_t$ values were afforded by comparing treated to untreated ΔC_t values. Data were then normalized to the fraction with the lowest abundance of GFP.

Verification of RAN Translation Product in $(G_4C_2)_{66}$–NoATG–NanoLuc Plasmid by SDS-PAGE and LC-MS

HEK293T cells were prepared as described above and transfected $(G_4C_2)_{66}$–*NoATG–NanoLuc or ATG-NanoLuc (pNL1.1)* plasmid using JetPrime transfection reagent. Cells were lysed in the plate 300 μL/well of MPER containing 1 μL of halt protease inhibitor cocktail. Cellular proteins were separated by SDS-PAGE. After coomassie staining, the bands correspoding to the molecular weight of *NanoLuc* (~21 kDa) *and Poly(GP)-NanoLuc* (~35 kDa) were excised. The protein-containing gel was treated with 10 mM DTT followed by 50 mM iodoacetamide, and subjected to trypsin digestion overnight. Prior to mass spectrometry analysis, the peptide pools were acidified, desalted with Zip-Tip C18 tip columns and lyophilized to dryness.

Each sample was then reconstructed in 100 μL of 0.1% formic acid, and a 13 μL aliquot was analyzed by an Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific) coupled to an EASY-nLC 1000 system. Peptides were in-line eluted on an analytical RP column (0.075 x 250 mm Acclaim PepMap RLSC nano Viper, Thermo Fisher Scientific), operating at 300 nL/min using the following gradient: : 5–25% B for 40 min, 25–44% B for 20 min, 44-80% B in 10 s, 80% B for 5 min, 80-5% in 10 s, and 5% B for 20 min (solvent A: 0.1% formic acid (v/v); solvent B: 0.1% formic acid (v/v), 80% CH_3CN (v/v)). The Orbitrap Fusion was operated in a data-dependent MS/MS mode using the 10 most intense precursors detected in a survey scan from 380 to 1,400 m/z performed at 120K resolution. Tandem MS was performed by HCD fragmentation with normalized collision energy (NCE) of 30.0%. Protein identification was carried out using Mascot and Sequest algorithms, allowing Oxidation (Met) and Deamination (Q) as variable modifications. Other settings included Carbamidomethylation of Cys as a fixed modification, three missed cleavages, and mass tolerance of 10 and 20 ppm for precursor and fragment ions, respectively. MS/MS raw files were searched against a human database along with porcine trypsin and sequence of protein of interest (*NanoLuc*).

In Vitro Translation Assay

The RNA template for translational assays was transcribed from the $(G_4C_2)_{66}$–*No ATG–NanoLuc* plasmid using an *in vitro* transcription system (RiboMax, Promega). *In vitro* translation reactions (50 μL) using the Flexi Rabbit Reticulocyte Lysate System (Promega) were programmed with 4 μg $(G_4C_2)_{66}$–*No ATG–NanoLuc* RNA and 1 μg Firefly Luciferase RNA (Promega) as templates. The RNA was folded into the G-quadruplex form in buffer containing KCl (384 mM) by heating at 95°C for 10 min followed by slow cooling to 25°C in a total volume of 10 μL. The RNA was folded into the hairpin form in buffer by placing on ice. (NMR spectroscopy shows that only the hairpin is formed at low temperatures.) After folding, the RNA was added to the reticulocyte lysate, amino acid mixtures, and RNasin ribonuclease inhibitor (Promega) in a total volume of 50 μL. Translation was allowed to proceed for 90 min at 30°C, and then luciferase activity was measured by diluting the reactions in GloLysis Buffer followed by addition of Nano-Glo substrate (Promega).

Molecular Dynamics & Computational Methods

Parameterization of 4

A previously described protocol (Childs-Disney et al., 2013, 2014) was applied to study the binding modes of **4** to r(G_4C_2)^{exp}, including force field parameters for **4**. The AMBER GAFF force field (Wang et al., 2005) was used to define the atom types while RESP charges were derived following the RESP protocol (Bayly et al., 1993; Cornell et al., 1993). 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, where Gaussian03 was used to perform these calculations (Frisch et al., 2004).

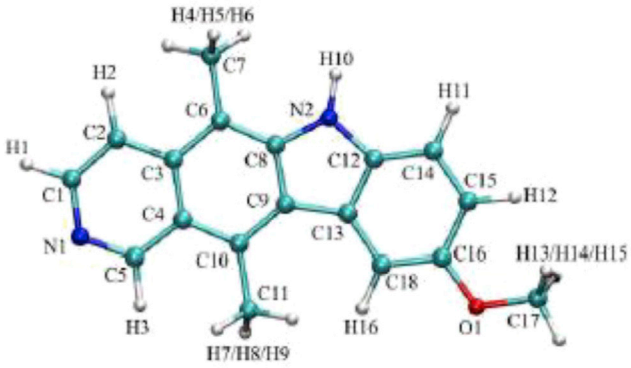
Binding Studies

The folded hairpin structure of G_4C_2 has two unique 1×1 GG motifs; (i) 5′-CGG/5′-CGG, and (ii) 5′-GGC/5′-GGC. Thus, two model systems were designed to study the binding modes of **4** to RNA G_4C_2: (i) r(CCGCGGCGG)$_2$, and (ii) r(CCGGGCCGG)$_2$. For each case, the lowest free energy bound states of **4** were calculated using a dynamic docking methodology we previously designed and successfully applied to study binding modes of two compounds targeting RNA CUG and CCUG repeat expansions (Childs-Disney et al., 2013, 2014). For the RNA, the amber99 force field (Cornell et al., 1995) with revised χ (Yildirim et al., 2010) and α/γ (Wales and Yildirim, 2017) torsional parameters was used. A modified implicit solvent model (GB^{OBC})(Onufriev et al., 2004) with 0.3 M salt concentration was used in all binding

simulations. The stem regions of RNA were restrained to sample around the A-form RNA by inclusion of Watson-Crick (WC) base pairing, and torsional restraints, as well as chirality restraints to keep the chiral centers in proper orientations. The dynamic docking methodology can be summarized as follows: The initial structures of the model RNAs with 1×1 GG motifs have been designed to be in A-form conformations where the 1×1 GG loop residues are in anti-anti state. Compound **4** molecules were placed 40 Å away from RNA loops to create the initial structure for the RNA-**4** complexes. First, we created initial bound states by slowly moving **4** to the 1×1 GG loops of the RNA by increments of 1 Å. The distance between the center of mass (COM) of the heavy atoms of the closing GC bases and the heavy atoms of ring atoms of **4** was used as the reaction coordinate. During this step, WC base pairing and torsional restraints representing A-form RNA was imposed on the RNA except the loop regions to allow **4** to interact with the RNA freely. Once **4** was 0 Å away from the RNA loops as described above, we slowly moved the compound away from the RNA by increments of 1 Å. In this second step, WC base pairing and torsional restraints representing A-form RNA was imposed including the loop regions. This process was repeated 50 times continuously, which provided 50 initial bound states for our next step. We then ran 50 independent MD simulations in implicit solvent using the 50 initial structures. Again, WC base pairing and torsional restraints representing A-form RNA were imposed on the RNA except the loop regions so that **4** was free to sample around the initial conformations. The simulation time of each independent MD was 120 ns creating a total of 6 μs combined trajectory, which was used in cluster analyses. During the whole process, Langevin dynamics with collision frequency of 1 was used with a long-range cutoff of 999 Å. We utilized the pmemd.cuda to perform the MD simulations (Case et al., 2016).

Analyses

An in-house code was written to perform the cluster analyses. First, the 50 MD trajectories were combined for the cluster analyses where root-mean-square deviation (rmsd) was used as the parameter to cluster the snapshots in the combined trajectories. Starting from the first structure in the combined trajectory, rmsd analyses were performed and snapshots with 1 Å were clustered. Note that the RNA duplexes are symmetric, which were considered during the cluster analyses. Once the clusters were created, MMPBSA (Case et al., 2016) analyses were performed to calculate the relative binding free energies (see table below).



Atom names for **4** (see for atom types and charges).

Atom numbers, names, types, and charges for **4**.

Atom Number	Atom Name	Atom Type	Atom Charge	Atom Number	Atom Name	Atom Type	Atom Charge
1	N1	nb	-0.697242	20	H8	hc	0.119629
2	C1	ca	0.379613	21	H9	hc	0.119629
3	H1	h4	0.056724	22	N2	na	-0.449642
4	C2	ca	-0.471363	23	H10	hn	0.382690
5	H2	ha	0.159024	24	C12	ca	0.124979
6	C3	ca	0.281386	25	C13	cp	0.194609
7	C4	ca	-0.328324	26	C14	ca	-0.300241
8	C5	ca	0.490060	27	H11	ha	0.189262
9	H3	h4	0.030793	28	C15	ca	-0.260461
10	C6	ca	-0.081640	29	H12	ha	0.169908
11	C7	c3	-0.180363	30	C16	ca	0.399860
12	H4	hc	0.067851	31	O1	os	-0.368024
13	H5	hc	0.067851	32	C17	c3	0.025384
14	H6	hc	0.067851	33	H13	h1	0.058026
15	C8	ca	0.084953	34	H14	h1	0.058026
16	C9	cp	-0.109642	35	H15	h1	0.058026
17	C10	ca	0.133682	36	C18	ca	-0.419302
18	C11	c3	-0.355354	37	H16	ha	0.182153
19	H7	hc	0.119629				

A sample input file used to perform MMPBSA analysis.
========== mmpbsa.in ==========
MMPBSA.py input file for running PB and GB
&general
startframe=1, endframe=100000, interval=1,
keep_files=1, verbose =2,
/
&gb
igb=5, saltcon=0.300,surften=0.0072,surfoff=0.00, /
&pb
istrng=0.300, fillratio=4.0,inp=1,radiopt=0, indi=1.0,exdi=80,scale=2,linit=1000,prbrad=1.4, cavity_surften=0.0072,cavity_offset=0.00,
/
==========

Synthetic Methods

General Synthetic Methods

All reagents and solvents used for chemical synthesis were purchased from commercial sources and used without further purification unless specified. NMR solvents were acquired from Cambridge Isotope Labs and used as received. Compound characterization is provided in Data S2.

Instrumentation

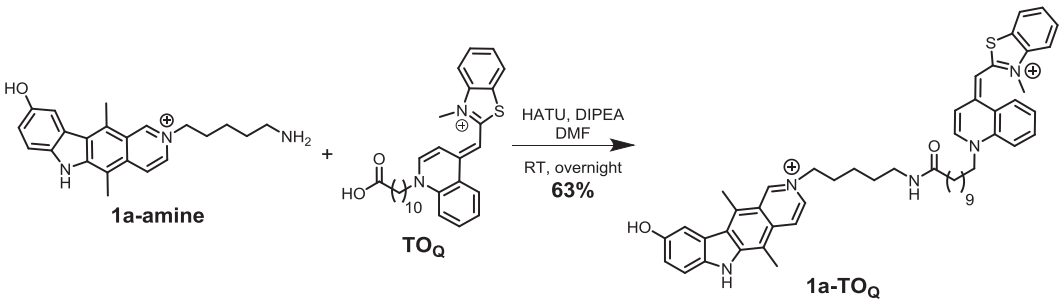
^{1}H- and ^{13}C-NMR spectra were acquired on 400 MHz or 700 MHz Bruker Avance spectrometers. Chemical shifts (δ) are reported in ppm relative to tetramethylsilane or the respective NMR solvent; coupling constants (*J*) are in Hertz (Hz). Abbreviations used are s, singlet; bs, broad singlet; d, doublet; 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.

Reverse-phase HPLC was completed using a Waters 1525 binary HPLC pump equipped with a Waters 2487 dual absorbance detector system. Preparative HPLC separations were completed using Atlantis PrepT3 OBD^{TM} 5 μM 19×150mm column on a gradient of 20% to 100% methanol (MeOH) + 0.1% trifluoroacetic acid (TFA) in H_2O + 0.1% TFA over 60 min and a flow rate of 5 mL/min.

Analytical HPLC separations were completed using SunFire C18 3.5μM 4.6×150mm column on a gradient of 20% to 100% methanol (MeOH) + 0.1% TFA in H_2O + 0.1% TFA over 60 min and a flow rate of 1 mL/min.

Flash chromatography was performed on a Biotage Isolera instrument using pre-packed silica columns purchased from Agela Technologies.

Mass spectra were recorded on a Varian 500-MS IT mass spectrometer. High resolution mass spectra were obtained at the Scripps Florida Mass Spectrometry and Proteomics Laboratory.



Synthetic Scheme for 1a-TO_Q

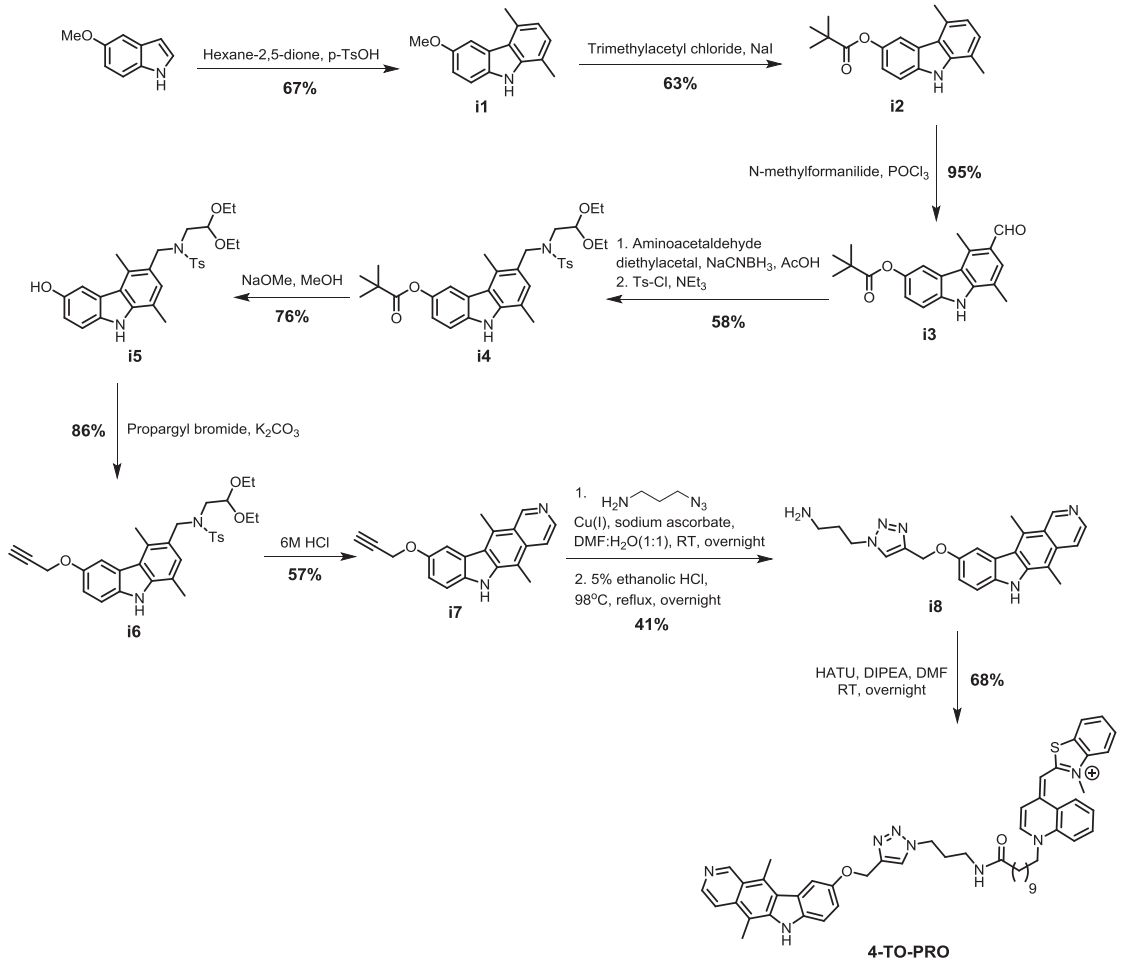
Synthesis of 1a-TO_Q

1a-amine (Su et al., 2014) and **TO_Q** (Carreon et al., 2004) were synthesized as previously described. **TO_Q** (12.5 mg, 22.5 μmol, 1.0 eq) and HATU (8.3 mg, 22.5 μmol, 1.0 eq) were dissolved in anhydrous *N*,*N*-dimethylformamide (DMF; 50 μL), followed by the addition of DIPEA (16 μL, 92 μmol, 4.1 eq). The mixture was vortexed and incubated for 1 min at room temperature. Then an aliquot from a 50 mM **1a-amine** (22.5 μmol, 1.0 eq) stock solution in DMF was added. The mixture was stirred for 16 h at room temperature. The reaction mixture was diluted with MeOH and water and purified by preparative HPLC. The product was isolated as a dark red solid in 63% yield (11.4 mg).

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.97 (s, 1H), 10.02 (s, 1H), 9.42 (s, 1H), 8.75 (d, *J* = 8.0 Hz, 1H), 8.58 (d, *J* = 7.3 Hz, 1H), 8.48 (d, *J* = 7.1 Hz, 1H), 8.42 (d, *J* = 7.2 Hz, 1H), 8.11 – 8.00 (m, 2H), 8.01 – 7.91 (m, 1H), 7.83 – 7.77 (m, 2H), 7.73 (dd, *J* = 10.0, 4.6 Hz, 2H), 7.65 – 7.58 (m, 1H), 7.49 (d, *J* = 8.6 Hz, 1H), 7.42 (t, *J* = 7.6 Hz, 1H), 7.30 (d, *J* = 7.2 Hz, 1H), 7.13 (dd, *J* = 8.6, 2.3 Hz, 1H), 6.87 (s, 1H), 4.67 (t, *J* = 7.3 Hz, 2H), 4.51 (t, *J* = 7.3 Hz, 2H), 4.00 (s, 3H), 3.27 (s, 3H), 3.03 (q, *J* = 6.6 Hz, 2H), 2.80 (s, 3H), 2.01 (dt, *J* = 15.0, 7.4 Hz, 2H), 1.92 (t, *J* = 7.4 Hz, 2H), 1.80 (dt, *J* = 15.0, 7.5 Hz, 2H), 1.48 – 1.42 (m, 2H), 1.35 – 1.09 (m, 16H).

^{13}C NMR (176 MHz, DMSO-d_6) δ 172.01, 160.03, 158.12 (q, TFA), 157.94 (q, TFA), 157.76 (q, TFA), 157.58 (q, TFA), 152.12, 148.48, 146,45, 146.35, 144.89, 144.25, 140.49, 136.96, 136.16, 133.54, 133.28, 132.10, 130.62, 128.25, 126.80, 126.15, 125.82, 124.57, 124.21, 123.90, 123.03, 122.90, 120.29, 119.75, 118.13 (CN, CH_3CN), 118.03, 117.92 (q, TFA), 117.48 116.28 (q, TFA), 113.05, 112.14, 110.07, 109.79, 107.82, 88.08, 59.30, 54.94 (CH_2, CH_2Cl_2), 54.14, 37.95, 35.39, 33.81, 30.44, 28.83, 28.78, 28.75, 28.65, 28.59, 28.53, 25.91, 25.25, 23.00, 14.92, 11.96, 1.18 (CH_3, CH_3CN).

HRMS: calculated for [$C_{51}H_{59}N_5O_2S^{2+}$]: 402.7189; found: 402.7191



Synthetic Scheme for 4-TO-PRO

Synthesis of i1

A solution of 5-methoxyindole (5 g, 33.9 mmol, 1.0 eq), hexane-2,5-dione (6.4 mL, 54.2 mmol, 1.6 eq) and *p*-toluenesulfonic acid (34 mg, 1.7 mmol, 0.05 eq) was sealed in a 20 mL microwave tube and irradiated for 20 min at 160°C. The reaction mixture was transferred to a 250 mL round-bottom flask and recrystallized from ethanol. The crystallized product was washed with ice-cold ethanol and dried under vacuum to provide **i1** as a slightly red solid (5.11 g, yield = 67%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.93 (bs, 1H), 7.70 (d, J=2.4 Hz, 1H), 7.38 (d, J=8.7 Hz, 1H), 7.12 (d, J=7.3 Hz, 1H), 7.08 (dd, J=8.7, 2.4 Hz, 1H), 6.91 (d, J=7.3 Hz, 1H), 3.95 (s, 3H), 2.85 (s, 3H), 2.52 (s, 3H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 207.26, 153.67, 139.70, 134.44, 130.78, 126.14, 125.00, 121.38, 120.49, 117.17, 113.48, 110.94, 106.38, 56.21, 36.94, 29.96, 20.49, 16.60.

HRMS (ESI) calc. for $C_{15}H_{15}NO$ [M+H]$^+$, 226.1154; found 226.1228.

Synthesis of i2

Trimethylacetyl chloride (1.56 mL, 12.72 mmol, 1.5 eq) and sodium iodide (2.6 g, 17.3 mmol, 2.04 eq) were added to a solution of **i1** (1.91 g, 8.48 mmol, 1.0 eq) in 21.4 mL acetonitrile at room temperature. The reaction mixture was heated to reflux (112°C) for 3 h, the solvent was removed under reduced pressure and the crude product was dissolved in ethyl acetate (100 mL). The solution was

washed with water (50 mL), a saturated solution of sodium thiosulfate (50 mL) and brine (20 mL). The organic layer was dried over sodium sulfate, filtered and the solvent was removed under reduced pressure. The crude product was loaded on a 100 g pre-filled Biotage cartridge and was eluted by applying a linear gradient from 4-34% EtOAc in hexanes providing **i2** as a slightly brown solid (1.56 g, yield = 63%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.10 (s, 1H), 7.73 – 7.69 (m, 1H), 7.34 – 7.28 (m, 1H), 7.11 (d, J=7.3 Hz, 1H), 7.06 – 7.01 (m, 1H), 6.88 (d, J=7.3, 1H), 2.72 (s, 3H), 2.50 (s, 3H), 1.48 – 1.43 (m, 9H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 178.49, 144.17, 139.70, 137.17, 130.96, 126.51, 124.73, 121.28, 120.83, 118.63, 117.24, 114.73, 111.04, 39.24, 27.45, 20.43, 16.68.

HRMS (ESI) calcd. for $C_{19}H_{21}NO_2$ $[M+H]^+$, 296.1572; found 296.1651.

Synthesis of i3

Phosphorous oxychloride (679 μL, 7.47 mmol, 1.5 eq) was added dropwise to *N*-methylformanilide (922 μL, 7.47 mmol, 1.5 eq) at room temperature. After stirring for 30 min, a solution of **i2** (1.47 g, 4.98 mmol, 1.0 eq) was added and the reaction mixture was heated to reflux (140°C) for 3 h. The solvent was removed under reduced pressure and the crude product was dried under vacuum for 90 min. An aqueous solution of potassium acetate (10%) was added and the mixture was stirred for 16 h at room temperature. The product was filtered and washed with 1% aqueous HCl (10 mL), water (10 mL) and a mixture of hexanes and toluene (1:1, 10 mL). After drying under vacuum, **i3** was obtained as a brown solid (1.53 g, yield = 95%).

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.82 (s, 1H), 10.36 (s, 1H), 7.89 (d, J=2.3 Hz, 1H), 7.68 (s, 1H), 7.59 (d, J=8.7 Hz, 1H), 7.18 (dd, J=8.7, 2.3 Hz, 1H), 3.07 (s, 3H), 2.55 (s, 3H), 1.36 (s, 9H).

^{13}C NMR (101 MHz, DMSO-d_6) δ 191.47, 177.05, 144.12, 143.02, 138.05, 135.90, 128.25, 125.70, 123.59, 121.03, 119.75, 118.35, 115.17, 111.83, 38.54, 26.91, 16.54, 14.61.

HRMS (ESI) calcd. for $C_{20}H_{21}NO$ $[M+H]^+$, 324.1521; found 324.1602.

Synthesis of i4

A mixture of **i3** (1.43 g, 4.42 mmol, 1.0 eq), aminoacetaldehyde diethylacetal (804 μL, 5.53 mmol, 1.25 eq), acetic acid (443 μL, 7.74 mmol, 1.75 eq) and sodium cyanoborohydride (486 mg, 7.74 mmol, 1.75 eq) in ethanol (34 mL) was heated to 83°C for 5 h. The reaction was quenched by the addition of a saturated aqueous solution of $NaHCO_3$ (30 mL) and the product was extracted with chloroform (3 x 30 mL). The combined organic extracts were washed with brine (20 mL), dried over sodium sulfate and the solvent was removed under reduced pressure giving intermediate product as a brown oil (1.91 g, yield = 98%), which was used without any further purification for the next step.

The intermediate product described above was dissolved in 11 mL THF, triethylamine (1.22 mL, 8.84 mmol, 2.0 eq) and p-toluenesulfonyl chloride (927 mg, 4.86 mmol, 1.1 eq) were added at 0°C. The cooling bath was removed, and the reaction mixture was stirred at room temperature for 16 h. After the addition of ethyl acetate and water (each 20 mL), the phases were separated and the aqueous layer was extracted with ethyl acetate (3 x 20 mL). The combined organic extracts were washed with 0.1 M HCl (1 x 50 mL), saturated aqueous $NaHCO_3$ solution (1 x 50 mL) and brine (1 x 50 mL), dried over sodium sulfate, filtered and the solvent was removed under reduced pressure. The crude product was co-evaporated with silica, loaded on a 100 g pre-filled Biotage cartridge and was eluted by applying a linear gradient from 8-66% EtOAc in hexanes giving **i4** (1.56 g, 59%) as a slightly brown solid. (58% yield over two steps)

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.14 – 8.05 (m, 1H), 7.79 – 7.70 (m, 3H), 7.39 – 7.34 (m, 1H), 7.30 – 7.27 (m, 2H), 7.08 – 7.04 (m, 1H), 6.96 (s, 1H), 4.65 (s, 2H), 4.42 (t, J=5.4 Hz, 1H), 3.57 – 3.46 (m, 2H), 3.34 – 3.25 (m, 2H), 3.19 (d, J=5.4 Hz, 2H), 2.71 (s, 3H), 2.39 (s, 3H), 2.37 (s, 3H), 1.43 (s, 9H), 1.07 (t, J=7.0 Hz, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 178.29, 144.29, 143.23, 139.40, 137.54, 137.48, 130.68, 129.62, 129.28, 127.41, 124.76, 124.27, 121.91, 118.83, 117.01, 115.19, 110.90, 102.07, 63.21, 50.74, 49.52, 39.23, 27.43, 21.58, 16.55, 15.92, 15.37.

HRMS (ESI) calcd. for $C_{33}H_{42}N_2O_6S$ $[M]^+$, 594.2764; found 594.2720.

Synthesis of i5

Sodium methoxide (78 mg, 1.45 mmol, 2.0 eq) was added to a solution of **i4** (430 mg, 0.72 mmol, 1.0 eq) in 3.6 mL methanol and the reaction mixture was stirred for 3 h at room temperature. The reaction was quenched by the addition of water (10 mL) and the product was extracted with ethyl acetate (3 x 20 mL). The combined organic extracts were dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude product was co-evaporated with silica and loaded on a 50 g pre-filled Biotage cartridge and was eluted by applying a linear gradient from 10-80% EtOAc in hexanes providing **i5** as a brown solid (274 mg, yield = 74%).

^{1}H NMR (400 MHz, DMSO-d_6) δ = 10.82 (s, 1H), 8.87 (s, 1H), 7.77 (d, J=8.4 Hz, 2H), 7.53 (d, J=2.3 Hz, 1H), 7.44 (d, J=8.4 Hz, 2H), 7.32 (d, J=8.6 Hz, 1H), 6.91 – 6.87 (m, 2H), 4.47 (s, 2H), 4.04 (t, J=5.4 Hz, 1H), 3.37 – 3.26 (m, 5H), 3.09 (m, 2H), 3.02 (d, J=5.4 Hz, 2H), 2.69 (s, 3H), 2.41 (s, 3H), 2.38 (s, 3H), 0.90 (t, J=7.0 Hz, 6H).

^{13}C NMR (101 MHz, DMSO-d_6) δ 150.29, 143.15, 139.47, 136.35, 134.23, 129.75, 129.53, 128.20, 127.15, 123.91, 122.26, 120.99, 116.82, 114.10, 111.22, 107.31, 100.53, 61.98, 50.77, 49.17, 20.98, 16.64, 15.33, 14.99.

HRMS (ESI) calcd. for $C_{28}H_{34}N_2O_5S$ $[M+H]^+$, 511.2188; found 511.2153.

Synthesis of i6

i5 (100 mg, 0.196 mmol, 1.0 eq) was dissolved in 2 mL acetone. After the addition of K_2CO_3 (81 mg, 0.392 mmol, 3.0 eq) and propargyl bromide (80% in toluene, 37 μL, 0.392 mmol, 2.0 eq) the reaction mixture was heated to 80°C and stirred for 16 hours. The reaction was quenched by the addition of water (5 mL) and the product was extracted with EtOAc (3 x 10 mL). The combined organic extracts were dried over Na_2SO_4, filtered and the solvent was removed under reduced pressure. The resulting yellow oil was purified by flash chromatography applying a linear gradient from 8-66% EtOAc in hexanes, which provided **i6** as a slightly yellow solid (93 mg, yield = 86% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.00 (s, 1H), 7.82 (d, J=2.4, 1H), 7.74 (d, J=8.3 Hz, 2H), 7.37 (d, J=8.8 Hz, 1H), 7.27 (d, J=8.0 Hz, 2H), 7.11 (dd, J=8.8, 2.4 Hz, 1H), 6.94 (s, 1H), 4.78 (d, J=2.4 Hz, 2H), 4.65 (s, 2H), 4.42 (t, J=5.4 Hz, 1H), 3.52 (dq, J=9.4 Hz, 7.0, 2H), 3.30 (dq, J=9.4 Hz, 7.0, 2H), 3.20 (d, J=5.4Hz, 2H), 2.77 (s, 3H), 2.55 (t, J=2.4 Hz, 1H), 2.39 (s, 3H), 2.36 (s, 3H), 1.07 (t, J=7.0 Hz, 6H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 151.68, 143.24, 139.53, 137.42, 135.48, 130.68, 129.59, 129.14, 127.39, 124.92, 123.91, 121.98, 116.97, 114.59, 111.06, 109.01, 102.07, 79.34, 75.46, 63.24, 57.49, 50.90, 49.52, 21.56, 16.53, 15.90, 15.35.

HRMS (ESI) calcd. for $C_{31}H_{36}N_2O_5S$ $[M+H]^+$, 549.2345; found 549.2286.

Synthesis of i7

A solution of **i6** (111 mg, 0.202 mmol) in 1.5 mL dioxane and 220 μL of 6 M HCl was heated to reflux (120°C) for 45 min. The solvent was evaporated, 6 mL 1% aqueous NaCl solution were added and the suspension was stirred for 1 h. The product was filtered, washed with water and rinsed with acetone. The product was removed from the frit and transferred to an Eppendorf vial, suspended in acetone, centrifuged and the acetone was decanted. The acetone washing was repeated three times, and the product was dried *in vacuo* to yield **i7** as a yellow orange solid (39 mg, yield = 57%).

^{1}H NMR (400 MHz, DMSO-d_6) δ 12.13 (s, 1H), 9.94 (s, 1H), 8.51 – 8.25 (m, 2H), 8.01 (d, J=2.4, 1H), 7.61 (d, J=8.8 Hz, 1H), 7.35 (dd, J=8.8, 2.4 Hz, 1H), 4.98 (d, J=2.4 Hz, 2H), 3.63 (t, J=2.4 Hz, 1H), 3.31 (s, 3H), 2.84 (s, 3H).

^{13}C NMR (176 MHz, DMSO-d_6) δ 151.95, 144.77, 144.00, 137.66, 133.77, 133.65, 127.57, 125.65, 122.59, 119.56, 119.44, 117.73, 112.12, 110.12, 109.80, 79.60, 78.35, 56.54, 14.93, 12.06.

HRMS (ESI) calcd. for $C_{20}H_{16}N_2O$ $[M+H]^+$, 301.1263; found 301.1340.

Synthesis of i8

i7 (10 mg, 33.3 μmol, 1 eq) and 3-azido-1-propanamine (10 mg, 100 μmol, 3 eq) were dissolved in a mixture of DMF:H_2O=1:1 (4 mL in total). Then ascorbic acid (2.64 mg, 15 μmol, 0.45 eq) and Cu(I) complex (3 mg, 5 μmol, 0.15 eq) were added under inert atmosphere and the mixture was left to react overnight at room temperature. The solvent was removed and to the solid residue was added 5% ethanolic HCl (5 mL) and the mixture was refluxed overnight at 98°C. Afterwards, the solvent was removed *in vacuo* and the resulting product was re-dissolved in MeOH:H_2O=1:4 and purified using preparative HPLC yielding the desired product as a yellow solid (5.5 mg, yield = 41%).

^{1}H NMR (700 MHz, DMSO-d_6) δ 9.97 (s, 1H), 8.42 (q, J = 6.9 Hz, 2H), 8.34 – 8.31 (s, 1H), 8.08 (s, 1H), 7.89 (s, 3H), 7.60 (d, J = 8.7 Hz, 1H), 7.37 (dd, J = 8.7, 2.3 Hz, 1H), 5.34 (s, 2H), 4.51 (t, J = 6.9 Hz, 2H), 3.32 (s, 3H), 2.83 (s, 5H), 2.18 – 2.07 (m, 2H).

^{13}C NMR (176 MHz, DMSO-d_6) δ 158.70 (q, TFA), 158.50 (q, TFA), 158.29 (q, TFA), 158.09 (q, TFA), 152.91, 144.83, 144.15, 143.13, 137.42, 133.87, 133.66, 127.59, 125.78, 124.73, 122.71, 119.57, 119.48, 118.27 (q, TFA), 117.61, 116.61 (q, TFA), 114.95 (q, TFA), 113.30 (q, TFA), 112.21, 110.12, 109.33, 62.09, 46.65, 36.40, 27.89, 14.96, 11.99.

HRMS (ESI) calcd. for $C_{23}H_{25}N_6O$ $[M+H]^+$, 401.4940; found 401.2058.

Synthesis of 4-TO-PRO

TO_Q (6.43 mg, 13.5 μmol, 1 eq) and HATU (5.14 mg, 13.5 μmol, 1 eq) were dissolved in *N*,*N*-dimethylformamide (DMF; 200 μL), followed by addition of *N*-ethyldiisopropylamine (DIPEA; 5.2 μL, 3.85 mg, 29.76 μmol, 2.2 eq). The mixture was left to react for 15 min at room temperature. Then, **i8** (5.95 mg, 14.88 μmol, 1.1 eq) dissolved in DMF (400 μL) previously neutralized with DIPEA (2.2 eq) was added and the mixture was stirred overnight at room temperature. The reaction mixture was concentrated under reduced pressure and re-dissolved in a mixture of MeOH:H_2O. The desired product was obtained using preparative HPLC as a dark red solid (7.9 mg, yield = 68%).

^{1}H NMR (700 MHz, DMSO-d_6) δ 12.04 (s, 1H), 9.91 (s, 1H), 8.68 (d, J = 8.7 Hz, 1H), 8.56 (d, J = 7.2 Hz, 1H), 8.36 (dd, J = 20.4, 6.9 Hz, 2H), 8.32 (s, 1H), 8.07 (d, J = 8.7 Hz, 1H), 8.03 (d, J = 2.1 Hz, 1H), 7.98 – 7.92 (m, 3H), 7.70 (t, J = 7.4 Hz, 2H), 7.57 – 7.55 (m, 1H), 7.51 (d, J = 8.7 Hz, 1H), 7.37 (t, J = 7.5 Hz, 1H), 7.28 (dd, J = 8.7, 2.3 Hz, 1H), 7.19 (d, J = 7.1 Hz, 1H), 7.04 (s, 1H), 6.76 (s, 1H), 5.30 (s, 2H), 4.52 (t, J = 7.5 Hz, 2H), 4.40 (t, J = 7.1 Hz, 2H), 3.93 (s, 3H), 3.29 (s, 3H), 3.07 (dd, J = 12.6, 6.6 Hz, 2H), 2.81 (s, 3H), 2.05 (t, J = 7.3 Hz, 2H), 1.97 (p, J = 6.9 Hz, 2H), 1.83 – 1.78 (m, 2H), 1.49 – 1.45 (m, 2H), 1.35 – 1.31 (m, 2H), 1.22 (s, 10H).

^{13}C NMR (176 MHz, DMSO-d_6) δ 172.33, 159.72, 158.27 (q, TFA), 158.09 (q, TFA), 157.90 (q, TFA), 157.72 (q, TFA), 152.84, 148.27, 144.75, 144.13, 142.91, 140.29, 137.36, 136.89, 133.78, 133.58, 133.18, 128.09, 126.70, 125.74, 125.70, 124.57, 124.41, 124.13, 123.80, 122.77, 122.66, 119.52, 119.37, 117.99, 117.57 (q, TFA), 117.47, 115.88 (q, TFA), 112.87, 112.12, 110.02, 109.25, 107.76, 87.88, 62.08, 54.12, 47.35, 40.02, 35.63, 35.34, 33.68, 30.09, 28.77, 28.75, 28.70, 28.63, 28.50, 28.44, 25.82, 25.18, 19.70, 14.96, 11.99.

HRMS (ESI) calcd. for $C_{52}H_{57}N_8O_2S^+$, 857.4320; found 857.4278.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample size was determined by studying variation in untreated samples. Each sample was randomly assigned as either treated or untreated and were unblinded. Statistical significance was calculated by using a two-tailed Student *t* test (Prism's GraphPad v. 7.03). Differences between treatment groups were considered statistically significant when $p < 0.05$ ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). Total RNA harvested from cells was excluded from further analysis if the ratio of Abs_{260}/Abs_{280} was <1.8, which indicates poor quality. The number of replicates for each experiment is reported in the figure caption ($n \geq 3$).

DATA AND SOFTWARE AVAILABILITY

N/A.