Supplementary Information for:

Design of a small molecule that stimulates VEGFA enabled by screening RNA fold-small molecule interactions

Hafeez S. Haniff[a], Laurent Knerr[d], Xiaohui Liu[a], Gogce Crynen[b], Jonas Boström[d], Daniel Abegg[a], Alexander Adibekian[a], Elizabeth Lekah[a], Kye Won Wang[e], Michael Cameron[c], Ilyas Yildirim[e], Malin Lemurell[d], and Matthew D. Disney[a,*]
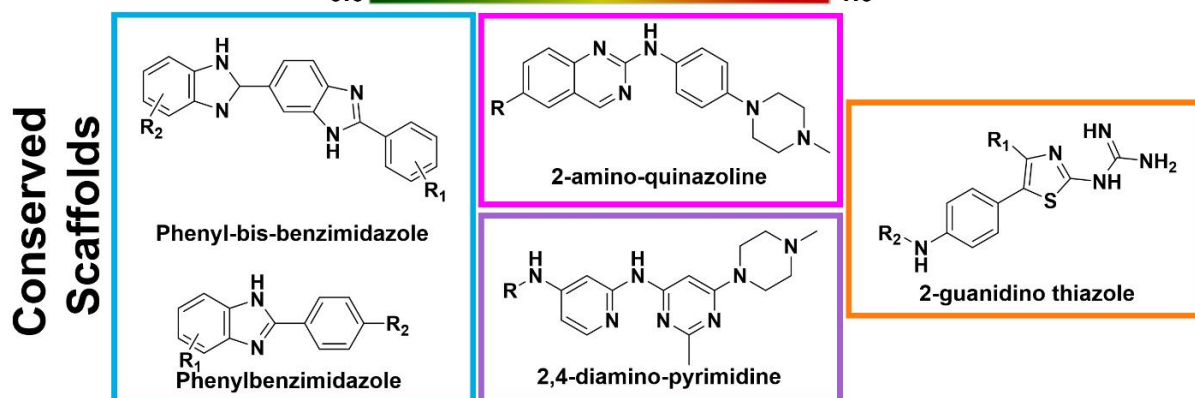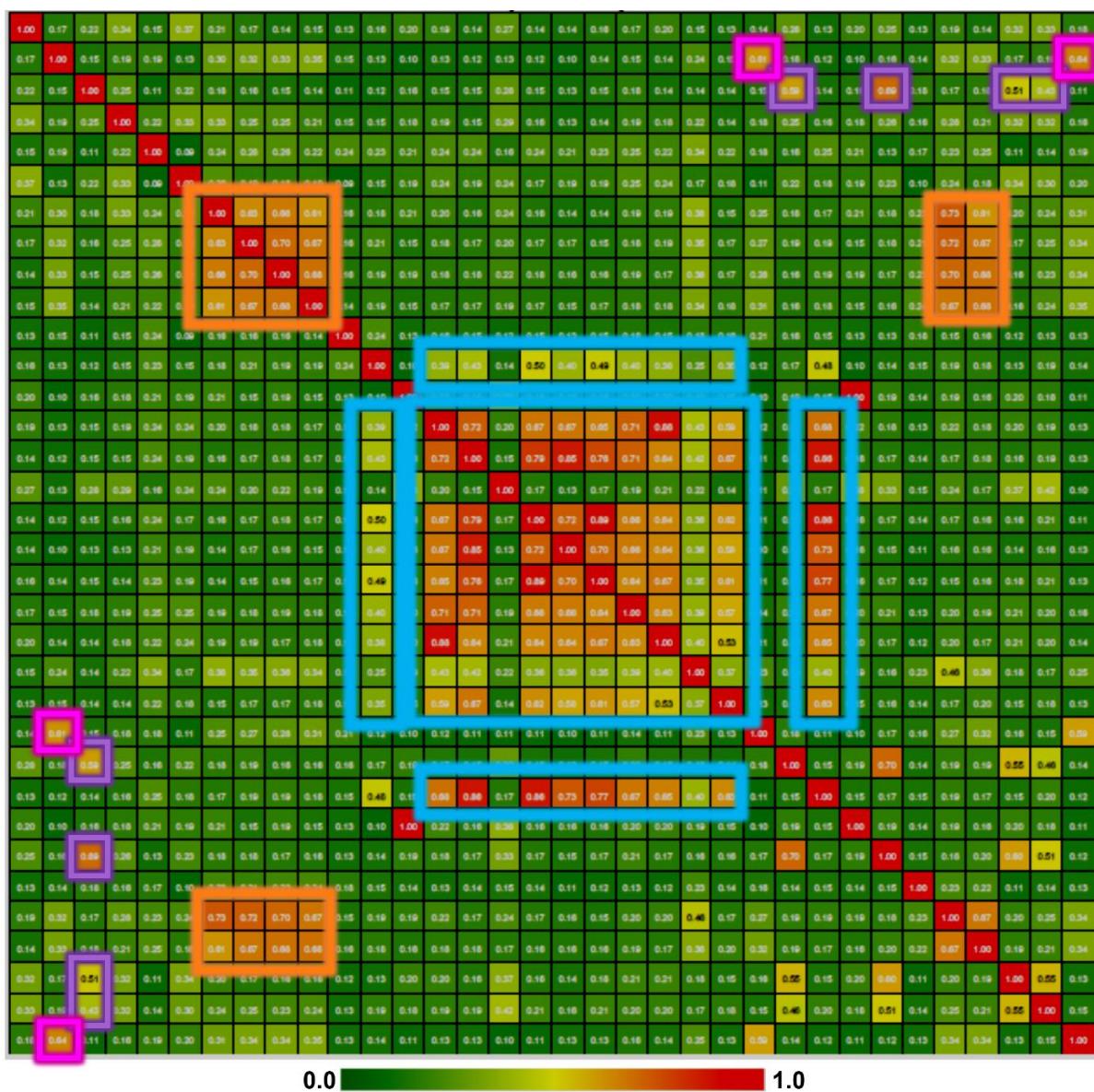
The Scripps Research Institute, [a] Department of Chemistry, [b] Informatics Core, [c] Department of Molecular Medicine 130 Scripps Way, Jupiter, FL 3x3458, USA; [d] Cardiovascular, Renal and Metabolism IMED Biotech Unit, AstraZeneca, Gothenburg, Pepparedsleden, 1, SE-431 83 Mölndal, Sweden; [e] Florida Atlantic University Department of Chemistry and Biochemistry 5353 Parkside Drive, Jupiter FL 33458

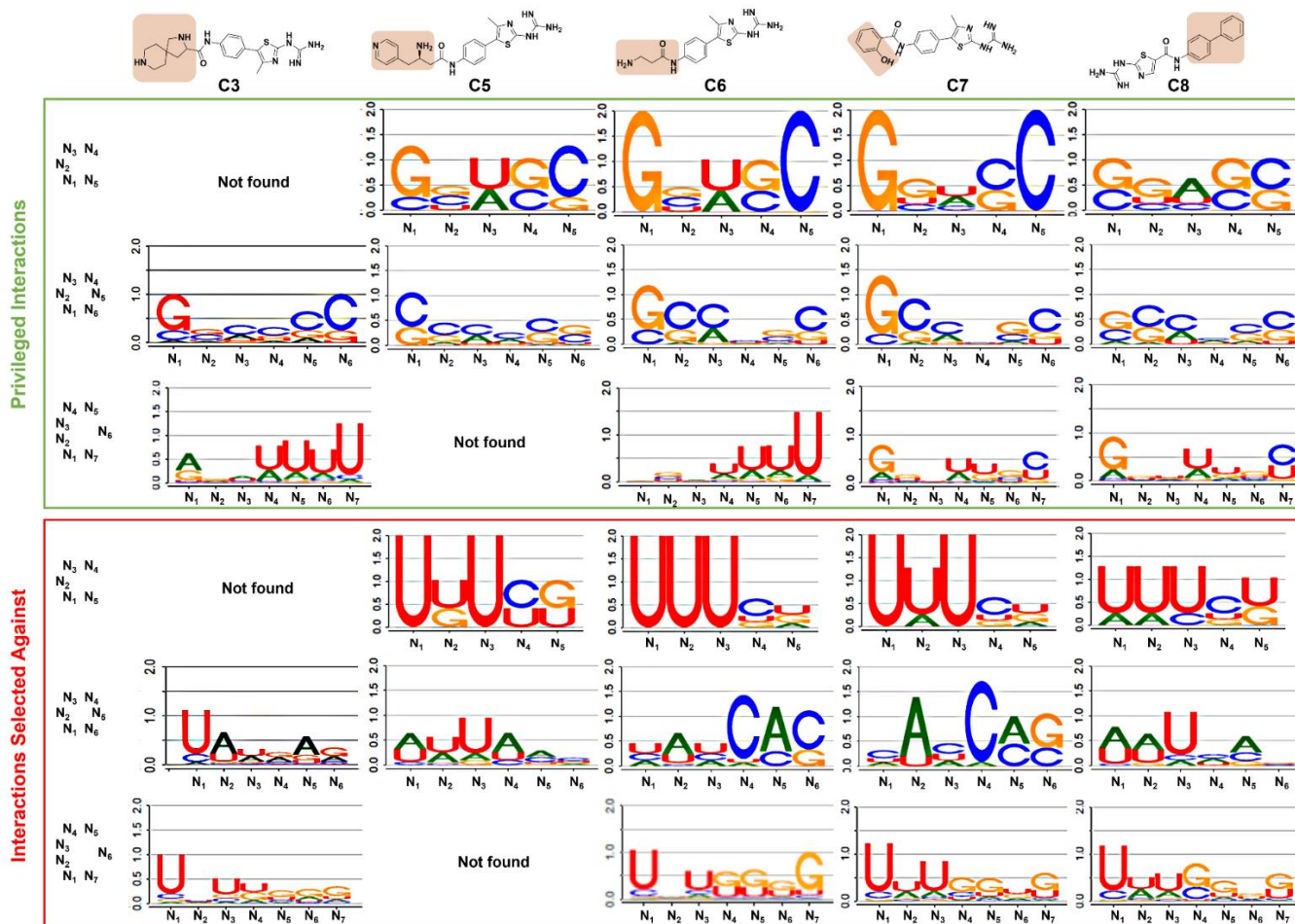*Author to whom correspondence is addressed: disney@scripps.edu

Table of Contents

Supplementary Figure 1: Tanimoto coefficient analysis of hit compounds identified from 2DCS. This analysis revealed a high degree of diversity among the hits, but also discrete clusters that form around conserved scaffolds.

Supplementary Figure 2: LOGOS Analysis of the sequence preference for the 2-guanidino-3-methylthiazole chemotype. This analysis revealed that compounds binding to the **3×2 ILL** and **3×3 ILL** preferred G and C in the positions N_1 and N_5 for **3×2 ILL** and N_1 and N_6 for **3×3 ILL**, suggesting a propensity to form GC pairs. Binders to the **4×3 ILL**, however, showed less homogeneity in their nucleotide preferences, with compounds **4** and **6** having a high prevalence of U at N_7 and A at N_1, suggesting formation of an AU closing pair Compounds **7** and **8** continued to show a preference for a GC closing pairs, as they bind similar RNAs from **3×2 ILL** and **3×3 ILL**. For RNAs from **3×2 ILL** and **4×3 ILL,** U rich sequences generally did not bind compounds, while those in **3×3 ILL** did not exhibit a clear pattern.
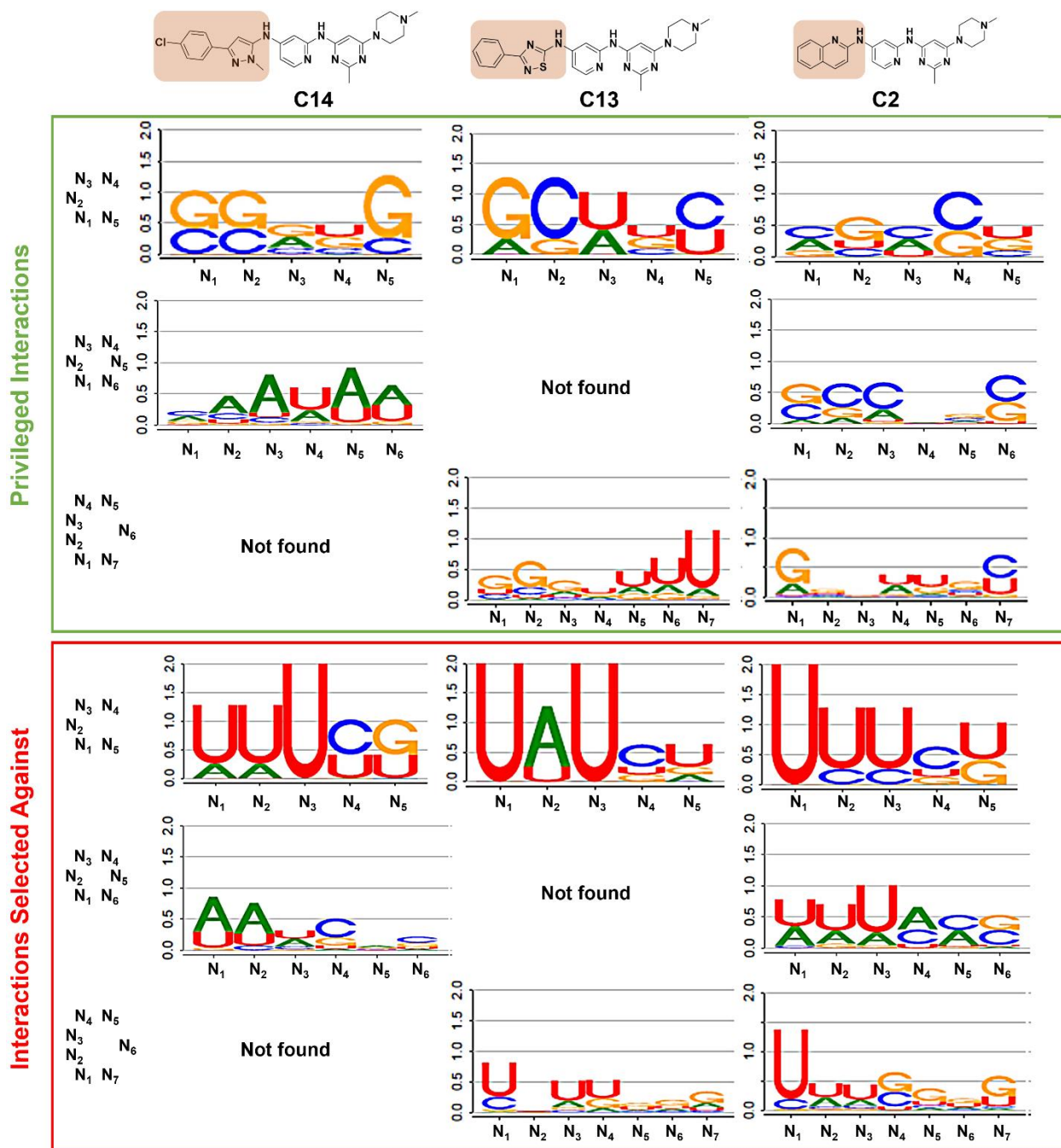
Supplementary Figure 3: LOGOS Analysis of the sequence preference for the 2-aminoquinazoline chemotype. Sequences that bound compounds in this class show a prevalence for G and C in their first and last randomized position for **3×2 ILL** and **4×3 ILL**. Interestingly, compound **14** shows a preference for A-rich sequences in **3×3 ILL** while **13** prefers G- and U-rich sequences in **4×3 ILL**. Globally, U-rich sequence did not bind any of the compounds, with **14** showing equal lack of binding for sequences with A or U in positions N_1 and N_2.
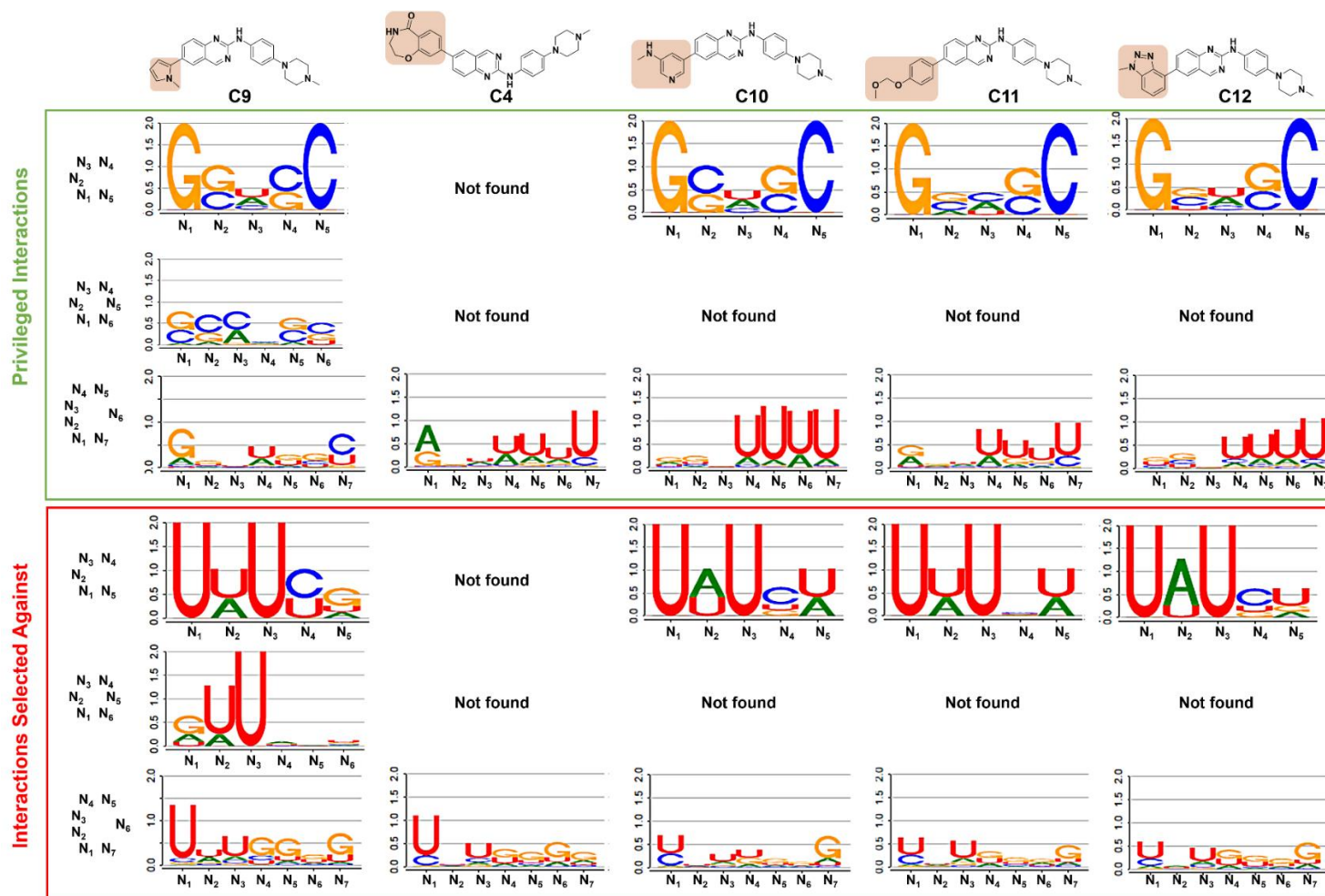
Supplementary Figure 4: LOGOS Analysis of the sequence preference for the 2,4-diaminopyrimidine chemotype. Similar to binders in the 2-guanidino-3-methylthiazole class binders to RNAs from **3×3 ILL** show a prevalence for G and C at positions N_1 and N_5, respectively, indicating a preference for GC closing pairs. Interestingly, compounds from other classes, except **9,** did not bind **3×3 ILL**, and **9** maintains this preference for a GC closing pair in both **3×3 ILL** and **4×3 ILL**. Unlike **9**, compounds **4**, **10**, **11**, and **12** prefer A or G at position N_1 and U at N_7, indicating the potential to form an AU or GU pair. Similar to the 2-guanidino-3-methylthiazole class, sequences in **3×2 ILL** and **3×3 ILL** that do not bind are U-rich while those that do not bind from **4×3 ILL** are rich in Gs and Us. Interestingly, U is selected against at N_1 and G at N_7, suggesting motifs with GU wobbles are selected against in all cases.

Preferred motifs binding many compounds

5'AU^ACA 5'AU^GCA 5'UG^CCUA 5'GC^AUA 5'UG^UCUA
3'UA GU 3'UA GU 3'AU_U AU 3'CG AU 3'AU _U AU

5'GG^AUA 5'UG^CUUA 5'GC^CUA 5'GU^UAC 5'UG^ACU
3'CC AU 3'AU _U AU 3'CG AU 3'CA UG 3'AC GA

5'CU^UAC 5'AU^UAC 5'GG^CUA 5'UA^CUUA 5'UG^UCU
3'GA UG 3'UA UG 3'CC AU 3'AU _U AU 3'AC GA

5'UU^UAC 5'UC^GCU 5'CG^AUA
3'GA UG 3'AG GA 3'GC AU

Unique motifs to Phenyl Thiazoles

5'AU^UAAUA 5'AU^AUA
3'UA_C _GAU 3'UA AU

5'AU^AUA
3'UA AU

Unique motifs to Quinazolines

5'GC^AUA 5'UA^UUUA 5'AU^ACAUA
3'CG AU 3'AU _U AU 3'UA_C _GAU

5'AU^GGGUA 5'CU^UUC
3'UA_G _AAU 3'GA AG

Unique motifs to Unclassified Compounds

5'UG^GUU 5'UU^UAC 5'UU^UAC
3'AC AA 3'AA UG 3'AA UG

5'AA^GGGUA 5'CU^AAC
3'UU_G _GAU 3'GA UG

Unique motifs to Phenyl Benzimidazoles

5'CC^ACA 5'AU^AGU 5'AU^CGG
3'GG AU 3'UA CG 3'UA CU

5'AU^AGAUA 5'GG^CUA
3'UA_C _GAU 3'CC AU

Unique motifs to Aminopyrimidines

5'CC^GUA 5'UG^GAU 5'AU^GGGU
3'GG AU 3'AC UA 3'UA _G CA

5'AG^CUA 5'UC^UCU
3'UC AU 3'AG GA

Supplementary Figure 5: RNA folds that prefer binding to various small molecules.

Motifs not bound by many compounds

5'AU^{UU}UU   5'UU^{UU}UA   5'AU^{UUU}UA   5'UU^{UU}UA   5'UU^{CUG}UA
3'UA $_U$ GA   3'AG $_C$ AU   3'UA$_U$ $_C$AU   3'AA $_C$ AU   3'AA$_U$ $_G$AU

5'AU^{U}AU   5'AU^{U}AC   5'UU^{A}UU   5'UU^{UU}UA   5'AU^{UU}UU
3'UA UG   3'UA UG   3'AA GA   3'AA $_U$ AU   3'UA $_U$ AA

5'UA^{A}UU   5'AU^{U}AU   5'AU^{UAU}UA   5'AU^{UUA}UA   5'UU^{UU}UA
3'AU GA   3'UA UA   3'UA$_U$ $_C$AU   3'UA$_U$ $_C$AU   3'AG $_U$AU

5'AU CU   5'AU^{UUC}UA   5'UU^{AU}UA   5'UU^{UA}UA   5'GU^{U}AC
3'UA$_U$GA   3'UA$_U$ $_C$AU   3'AA $_C$ AU   3'AG $_C$ AU   3'CA UG

Motifs not preferred by Phenyl Thiazoles

5'AU^{CCA}UA   5'AU^{AU}UA   5'AU^{UUU}AA
3'UA$_U$ $_G$AU   3'UA $_C$ AU   3'UA$_U$ $_C$UU

5'GU^{U}AC   5'UU^{UU}UA   5'AU^{UUU}UA
3'UA UG   3'GA $_U$ AU   3'UA$_U$ $_C$AU

5'AU^{AAU}UA
3'UA$_C$ $_G$AU

Motifs not preferred by Quinazolines

5'UU^{U}AU   5'UG^{C}UA   5'AU^{U}GU
3'AA UA   3'AC AU   3'UA UA

Motifs not preferred by Unclassified Compounds

5'AU^{GU}UU 5'AU^{UUA}UA 5'AU^{UUC}UA
3'UA $_U$ GA 3'UA$_U$ $_C$AU 3'UA$_C$ $_U$AU

Motifs not preferred by Phenyl Benzimidazoles

5'AU^{UAA}UA 5'AU^{UGA}AU 5'AU^{UCA}UA
3'UA$_U$ $_G$AU 3'UA$_U$ $_G$AU 3'UA$_U$ $_G$AU

5'AU^{U}AU   5'AU^{U}AC
3'UA UG   3'UA UG

Motifs not preferred by Aminopyrimidines

5'UU^{UU}UA   5'UA^{UU}UA   5'AU^{UC}UU
3'AG $_C$ AU   3'AU $_C$ AU   3'UA $_U$ AA

5'UU^{AU}UA
3'AG $_C$ AU

Supplementary Figure 6: RNA 3D folds not bound by small molecules.
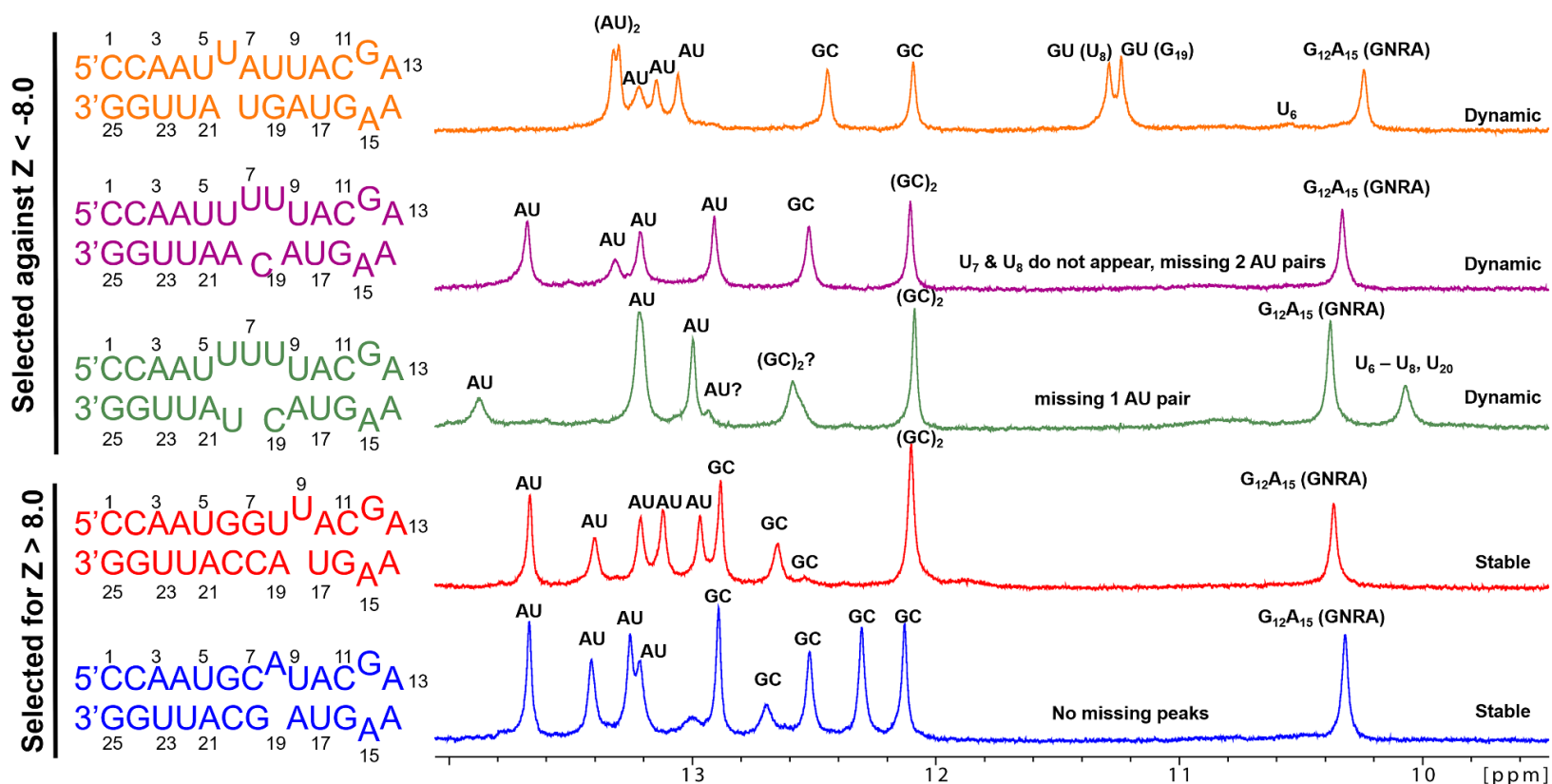
Supplementary Figure 7: 1D imino H^1 NMR of RNAs spanning a range of Z$_{obs}$. To gain insight into how structural dynamics might affect small molecule recognition, nuclear magnetic resonance (NMR) spectroscopy studies of RNAs that are found to generally bind and not to bind small molecules were completed. The RNAs **1** (blue) and **2** (red) were chosen because they appear in the RNAs within the top 0.5% of Z$_{obs}$ scores, i.e., bind avidly, for >50% of the ligands while RNAs **3** (green), **4** (purple), and **5** (orange) were chosen because they appear in the lowest 0.5% of Z$_{obs}$ scores, i.e., do not bind, >50% of the ligands. For these studies, 1D imino proton spectra were analyzed as it gives insights into the stability of structure and dynamics by whether protons are readily assigned. Each of the imino proton resonance was observed in the spectra of RNAs **1** and **2**, which bind small molecules. In contrast, imino proton resonances are missing because of fast exchange with solvent for RNAs **3**, **4**, and **5** that do not bind small molecules. For example, RNAs **3** and **5** are missing one resonance from an AU pair and RNA **4** is missing two resonances from AU pairs. Single stranded uridine residues are also unaccounted for in RNAs **2**, **4**, and **5** because of exchange with solvent. Thus, and in general, RNA motifs that bind ligands may tend to have less dynamics than RNA motifs that do not bind ligands. Spectra were measured once for each RNA.

a

RNAs Tested for binding

hsa-miR-342 BPC 5'Cy5-GGUG**ACG**GGU^{G}A 3'CCAC**UGC**CCA$_A$A

hsa-miR-342 Drosha Site 5'Cy5-GGUGA^{G}GGU^{G}A 3'CCACU$_G$CCCA$_A$A

hsa-miR-100 BPC 5'Cy5-GGCAC**AUA**CC^{G}A 3'CCGUG**UAU**GG$_A$A

hsa-miR-100 Drosha Site 5'Cy5-GGCACA^{A}ACC^{G}A 3'CCGUGU$_A$UGG$_A$A

hsa-miR-214 BPC 5'Cy5-CAUG**UGAC**ug^{G}A 3'GUACA**CUG**AC$_A$A

hsa-miR-214 Drosha Site 5'Cy5-CAUGUG^{U}CUG^{G}A 3'GUACAC$_U$GAC$_A$A

hsa-miR-377 BPC 5'Cy5-GUU**GUG**UG**AAU**UC^{G}A 3'CAA**CAC**ACUU**A**AG$_A$A

hsa-miR-377 Dicer Site 5'Cy5-GUU**G** GUGA^{A}UUC^{G}A 3'CAA**C**$_A$CAC**U** **A**AG$_A$A

Fitness (C1): 0%, 0%, 0%, 43%
Fitness (C10): 0%, 8%, 25%, 61%

b

C1 K$_d$ = 29,100 ± 3,100 nM

C1 K$_d$ >100,000 nM

c

C10 K$_d$ = 29,200 ± 540 nM

C10 K$_d$ > 100,000 nM

d

miRNA Isoforms (* = Functional Site)

e

miRNA Isoforms (* = Functional Site)

S-10

Supplementary Figure 8: Microscale Thermophoresis binding isotherms of C1 and C10 to miR-377, -214, -100, and -342. **a)** Summary of binding constants for **C1** and **C10** to RNA constructs with a range of Fitness Scores. These data show that **C1** only binds to miR-377 (Fitness Score = 43%), while no binding is observed to miR-214, -100, and -342, all of which **C1** has a Fitness Score of 0. Compound **C10** shows decreasing affinity based on the Fitness for each miRNA. Data are reported as mean ± SD (n = 3). **b)** Representative binding isotherm of **C1** binding to the miR-377 Dicer site and its fully base paired control RNA. Datapoints are the mean ± SD (n = 3 experimental replicates). **c)** Representative binding isotherm of **C10** to miR-214's Drosha site and its fully base paired control RNA. Datapoints are mean ± SD (n = 3 experimental replicates). **d)** Fitness Score of **C1** is correlated with reduction of mature miRNA levels, as determined by RT-qPCR, for miR-377, miR-214, - 100, and -342, and their isoforms in HUVECs. Only miRNAs with highly fit interactions between **C1** and a functional site are affected by compound treatment in HUVECs (for miR-377, -421,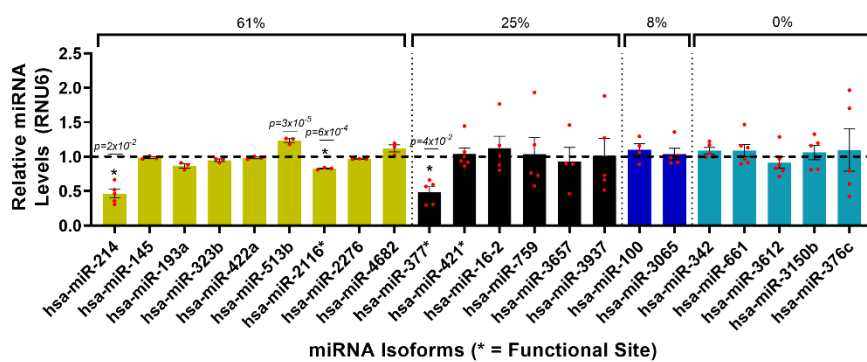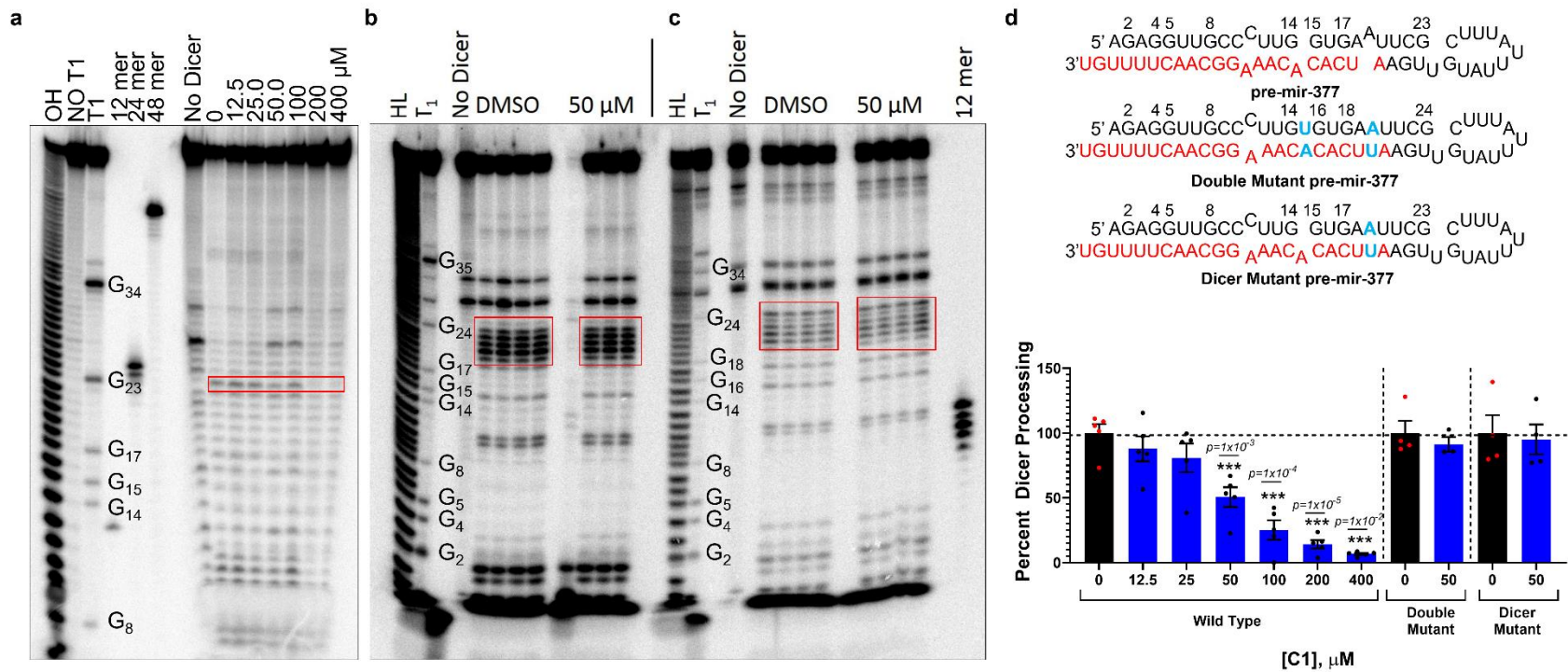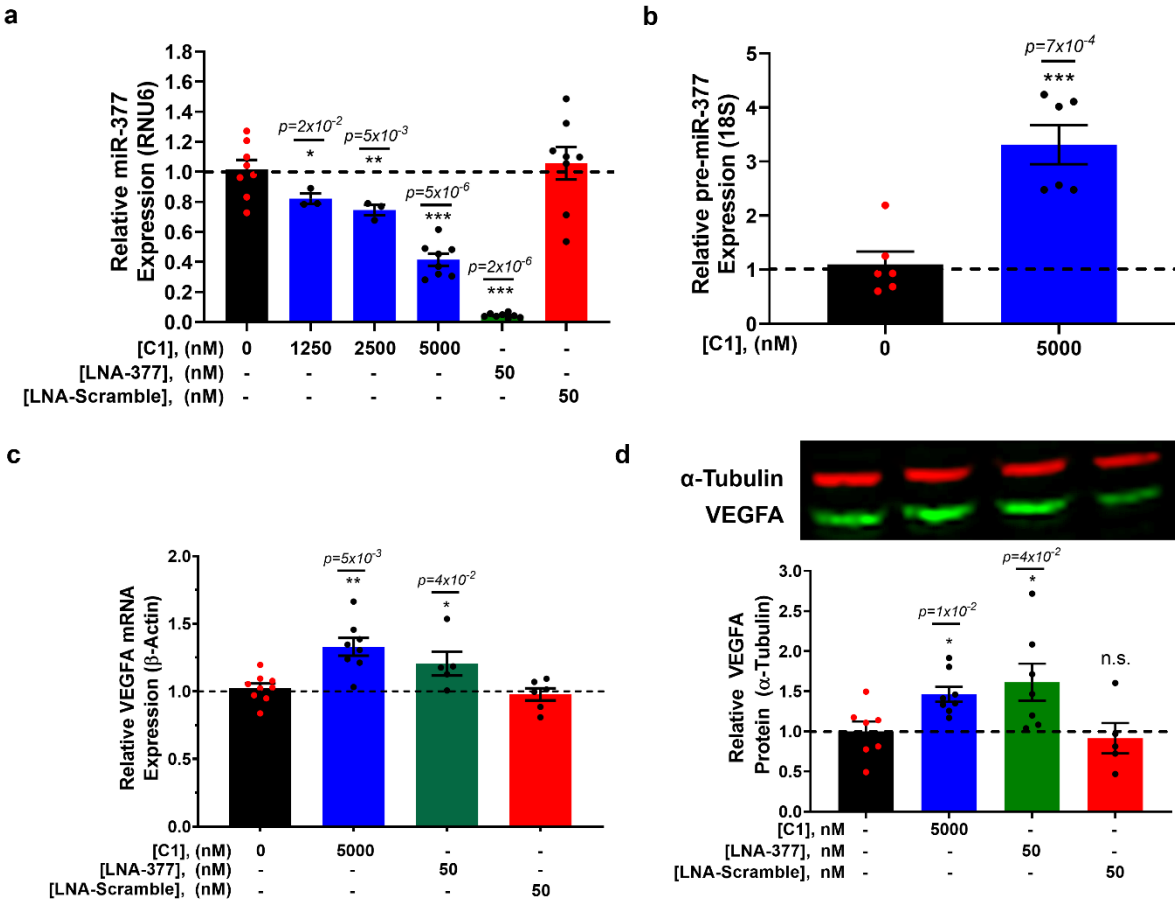 -16-2, -759, -3657, -214, -100, and -342, n = 9 biologically independent replicates; for all other miRNAs, n = 3 biologically independent replicates). Data are reported as mean ± SEM. Statistical significance was determined by a two-sided Student's t-test. **e)** Fitness Score for **C10** is correlated with bioactivity, as determined by reduction of the levels of mature miR-377, miR-214, - 100, and -342, and their isoforms in HUVECs by RT-qPCR. Only miRNAs with fit interactions with **C10** (61% and 25%) in a functional site are affected by compound treatment in HUVECs. Data are reported as a mean ± SEM (n = 4 biologically independent replicates). All p values were determined by a two-sided Student's t-test; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

Supplementary Figure 9: Validation of the SMIRNA methodology by RT-qPCR and RANT assay. **a)** Schematic of r(G$_4$C$_2$) repeat driven ALS, where under disease conditions the repeat expansion can initiate repeat associated non-ATG mediated translation (RANT) of toxic poly(GP) dipeptides. If a small molecule binds to the repeat, then RANT is inhibited. Thus, we designed a reporter with (G$_4$C$_2$)$_{66}$ fused to Nano-Luc to measure RANT, and an SV40 ATG FLuc clone to measure canonical translation.[1] Inforna identified **C20** as a lead binder to r(G$_4$C$_2$) repeats. **b)** MST binding isotherm of **C20** binding to r(G$_4$C$_2$)$_8$ with a K$_d$ of 622 ± 110 nM, with no saturable binding to r(GGCC)$_5$ base paired RNA. Data are reported as mean ± SEM (n = 4). **c)** **C20** selectively inhibits RANT over canonical translation in HEK293T cells expressing reporters of both RANT and canonical translation. Data are reported as mean ± SEM (n = 3 biologically independent replicates. All p values were determined by a two-sided Student's t-test; *, p < 0.05; **, p < 0.01.

Supplementary Figure 10: Inhibition of *in vitro* Dicer processing of pre-miR-377 by C1. **a)** *In vitro* Dicer processing assay of WT pre-miR-377 with **C1** treated in a dose response. **b)** *In vitro* Dicer processing assay for the Dicer Mutant transcript with **C1** at 50 μM. **c)** In vitro Dicer processing assay for Double Mutant transcript with **C1** at 50 μM. **d)** Quantification of Dicer processing for all RNAs shows that **C1** inhibited pre-miR-377 with an IC_{50} of 50 ± 10 μM for the WT RNA while mutation of the A bulge in the Dicer site to an AU pair and to both A bulges ablate protection, indicating the necessity of the binding site for activity. For all panels, n = 4 independent experimental replicates, and data are reported as mean ± SEM. All p values were determined using a two-sided Student's t-test; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.
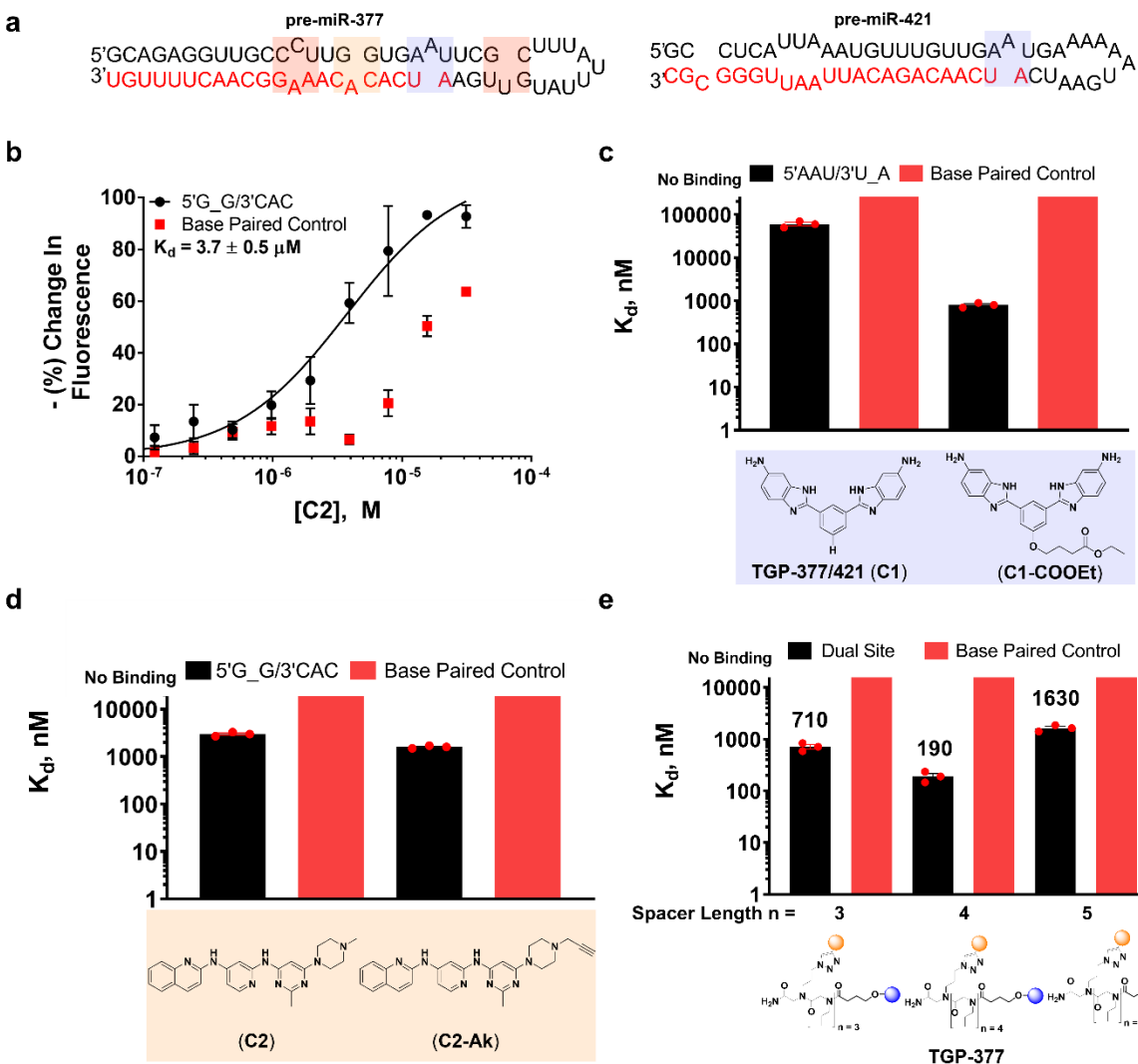
Supplementary Figure 11: Summary of bioactivity for C1. **a)** Compound **C1** reduced levels of miR-377 in HUVECs in a dose-dependent fashion, with an IC_{50} of ~5 μM. Data are reported as mean ± SEM (n = 8 biologically independent replicates). **b)** Levels of pre-miR-377 were increased with **C1** treatment, as expected by its mode of action – inhibition of miR-377 biogenesis. Data are reported as mean ± SEM (n = 6 biologically independent replicates). **c)** Compound **C1** (5 μM) also increased expression of *Vegfa* mRNA by ~30%, similar to an LNA antagomir targeting miR-377 (LNA-377). A scrambled LNA oligonucleotide had no effect on mRNA levels. Data are reported as mean ± SEM (n = 8 biologically independent replicates). **d)** Western blot analysis showed that **C1** increased VEGFA protein levels by ~30%, similar to the change in mRNA levels. A similar change was observed upon treatment with LNA-377, while the scrambled LNA control oligonucleotide had no effect. Data are reported as mean ± SEM (n = 7 biologically independent replicates). All p values were determined by a two-sided Student's t-test; *, p<0.05; **, p<0.01; ***, p<0.001.
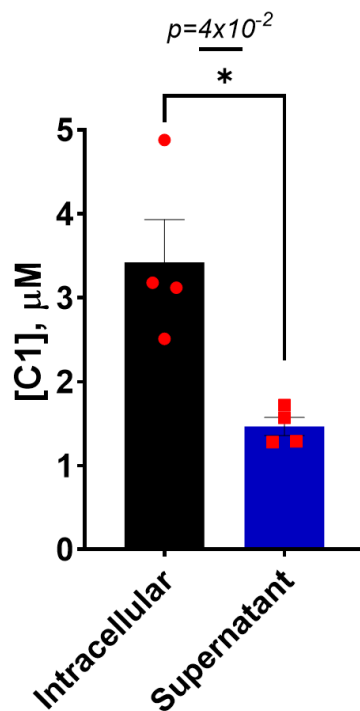
Supplementary Figure 12: Cellular uptake of C1 in HUVECs. HUVECs were treated with **C1** at 5 µM for 48 h, and then medium supernatants and total cell lysates were harvested. Using LC-MS/MS analysis, **C1** was identified and quantified spectroscopically to assess the concentration of **C1** both intracellularly and extracellularly. Compound **C1** either accumulates or is actively up taken into HUVECs by 2.7-fold compared to the extracellular concentration. This supports that higher intracellular concentration are possible to achieve activity. Data are reported as mean ± SEM (n = 4 biologically independent replicates). All p values were determined by a two-sided Student's t-test; * = p<0.05.
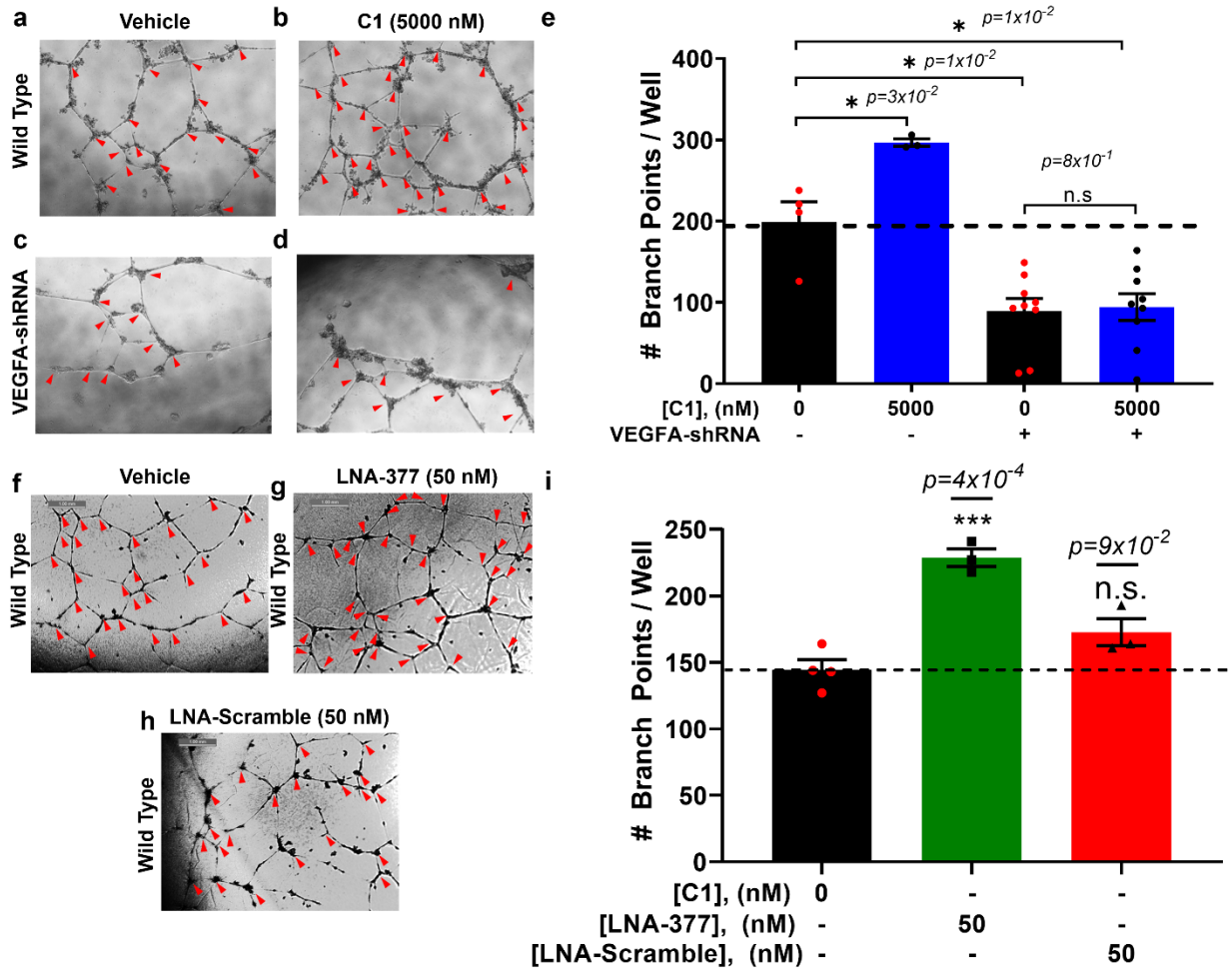
Supplementary Figure 13: Angiogenesis triggered by C1 is VEGFA-dependent. a – b) Representative images for Wild Type HUVECs treated with **a)** vehicle or **b) C1** (5 μM). **a – b)** Representative images for HUVECs transduced to express an anti-*Vegfa* shRNA treated with **c)** vehicle or **d) C1** (5 μM). **e)** Quantification of the number of branching points (red arrows) per well from 9-fields of view. Data are reported as mean ± SEM (n = 4 biologically independent replicates for wild type cells from two independent experiments; n = 9 biologically independent replicates for shRNA cells from two independent experiments). **f – h)** Treatment of wild type HUVECs with **f)** vehicle; **g)** LNA-377 (50 nM); and **h)** LNA-Scramble (50 nM). **i)** Quantification of the number of branching points (red arrows) for 9 fields of view per replicate. Data are reported as mean ± SEM (n = 4 biologically independent replicates from two independent experiments). All p values were calculated using a two-sided Student's t-test; *, p<0.05; ***, p<0.001).

Supplementary Figure 14: Summary of binding data for monomer fragments in lead optimization. a) Binding data for **TGP-377/421** to miR-377's Dicer site and Base Paired control. **b)** Representative Fluorescein labeled RNA binding isotherm for binding of **C2** to the 5'G_G/3'C$\underline{A}$C A bulge in pre-miR-377. **c)** Binding affinity of **TGP-377/421** (65,000 ± 10,000 nM) to **C1-COOEt** shows a >10-fold gain (800 ± 10 nM) in affinity compared to **TGP-377/421**. **d)** Binding affinity of **C2** (K_d = 3,000 ± 100 nM) compared to **C2-Ak** (K_d = 1,600 ± 100 nM) shows no significant gain or loss in affinity. **e)** Binding affinity of dimer library with spacer length of 3, 4, and 5 *N*-n-propyl-glycine spacer units. These studies show that 4 spacer units are optimal (**TGP-377**) by microscale thermophoresis. All binding constants are reported as mean ± SD of the curve fit by fitting a curve generated from the average of n = 3 points per concentration.
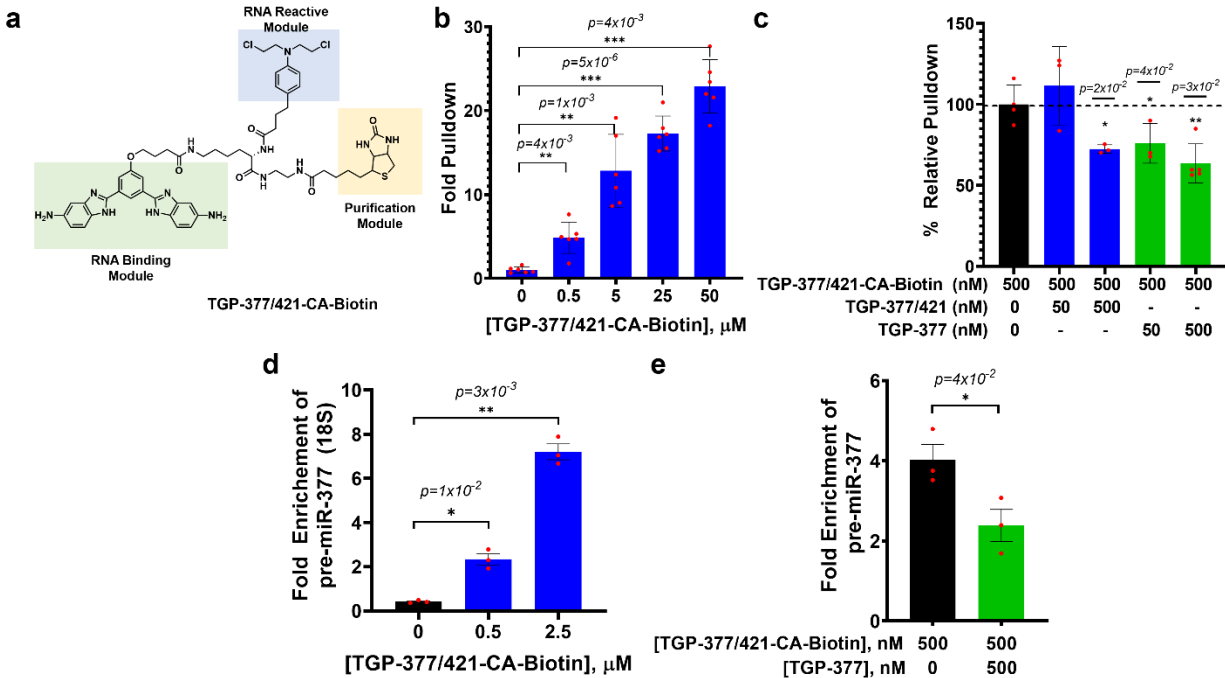
Supplementary Figure 15: Chemical Crosslinking and Isolation by Pull-down (Chem-CLIP).
a) Chemical structure of the **TGP-377/421-CA-Biotin** Chem-CLIP probe. **b) TGP-377/421** pulled down pre-miR-377 dose dependently *in vitro*. Data are reported as mean ± SEM (n = 4 independent experimental replicates). **c)** Competitive Chem-CLIP experiments show that the parent compound, **TGP-377/421**, bound the same site as the Chem-CLIP probe, as an equimolar amount (500 nM) reduced pull-down by ~25%. Interestingly, 10-fold less **TGP-377** (50 nM) is required to reduce pull-down by the same amount. Data are reported as mean ± SEM (n = 6 independent experimental replicates). **d)** Treatment of HUVECs with **TGP-377/421-CA-Biotin** resulted in ~3-fold enrichment of pre-miR-377 when cells were treated with 500 nM of the Chem-CLIP probe and ~8-fold enrichment when treated with 2.5 µM of the Chem-CLIP probe. Data shown are reported as mean ± SEM (n = 3 biologically independent replicates). **e)** pre-treatment of HUVECs with **TGP-377** at 500 nM followed by addition of 500 nM **TGP-377/421-CA-Biotin** resulted in an ~50% reduction in pull-down of pre-miR-377. Data are reported as mean ± SEM (n = 3 biologically independent replicates). All p values were calculated using a two-sided Student's t-test; *, $p < 0.05$; **, $p < 0.01$; ***, $p<0.001$.

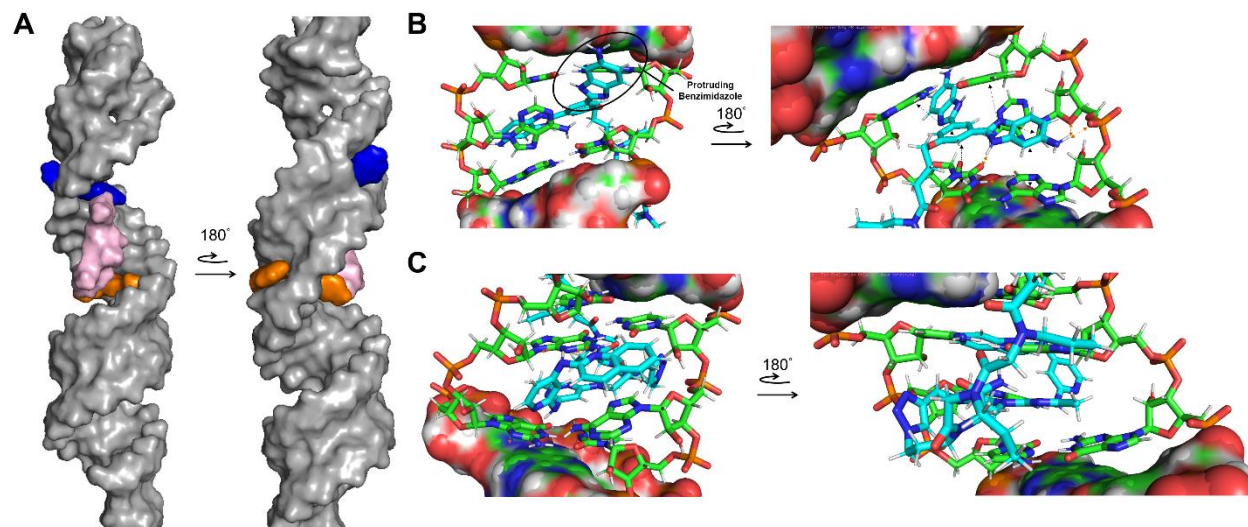**Supplementary Figure 16: Representative poses of MD simulations of TGP-377 bound to pre-miR-377. a)** Energy minimized MD simulation of **TGP-377** bound to pre-miR-377. MD simulations show that both **C1** and **C2** modules intercalate into the helix with potential H-bonding and $\pi-\pi$ interactions with the closing pair bases and the non-canonically paired adenine. **b)** Zoomed in structures of **C1** bound to the 5'AAU/3'U_A bulge (Dicer site) shows clear potential $\pi-\pi$ interactions with the central benzene of C1 and one of the benzimidazole moieties and potential hydrogen bonding interactions to the nucleobases with the benzimidazole. Specifically, **C1**'s central benzene exhibits $\pi-\pi$ interactions with the uridine (U) and adenine (A) of the Dicer site bulges 5' and 3' closing pairs respectively. One of the benzimidazoles stacked benzimidazoles shows similar stacking interactions with the A and U of the 5' and 3' closing pairs respectively. The other benzimidazole protrudes from the minor groove of the helix. **c)** Zoomed in structures of **C2** bound to the 5'G_G/3'CAC bulge adjacent to the Dicer site. Intercalation is also observed with potential $\pi-\pi$ stacking interactions driving binding.
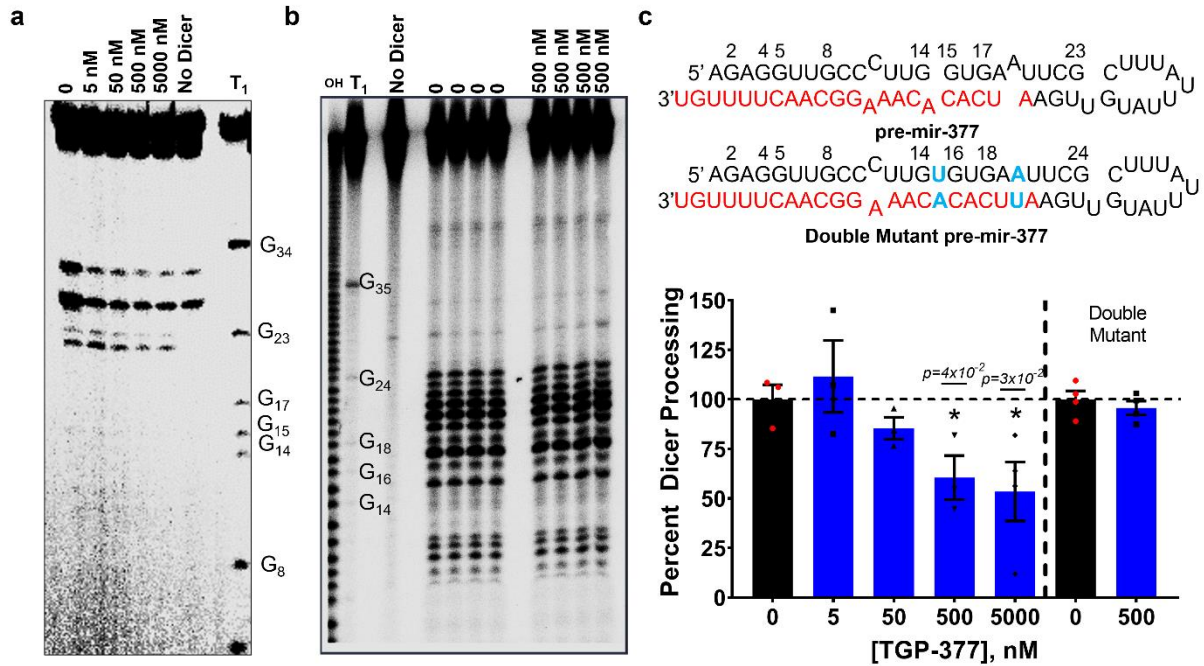
Supplementary Figure 17: Inhibition of *in vitro* Dicer processing of pre-miR-377 by TGP-377. a) *In vitro* Dicer inhibition of wild type pre-miR-377 by **TGP-377** is dose dependent . (n = 3). **b) TGP-377** (500 nM) did not inhibit *in vitro* Dicer inhibition of Double Mutated pre-miR-377 (n = 4. **c)** RNA secondary structures for tested constructs and quantification of Dicer cleavage bands. Data are reported as mean ± SEM. All p values were calculated by a two-sided Student's t-test; *, p<0.05).
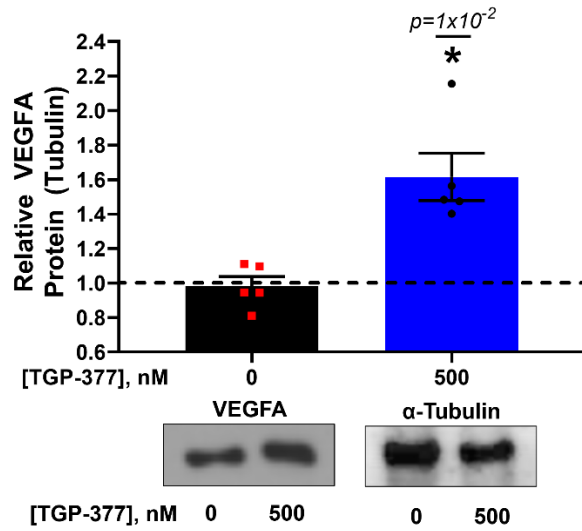
Supplementary Figure 18: Western blot of HUVECs treated with TGP-377. VEGFA protein levels are increased by treatment with **TGP-377**. Data are reported as mean ± SEM (n = 5 biologically independent replicates). All p values were determined by a two-sided Student's t-test; *, p <0.05.
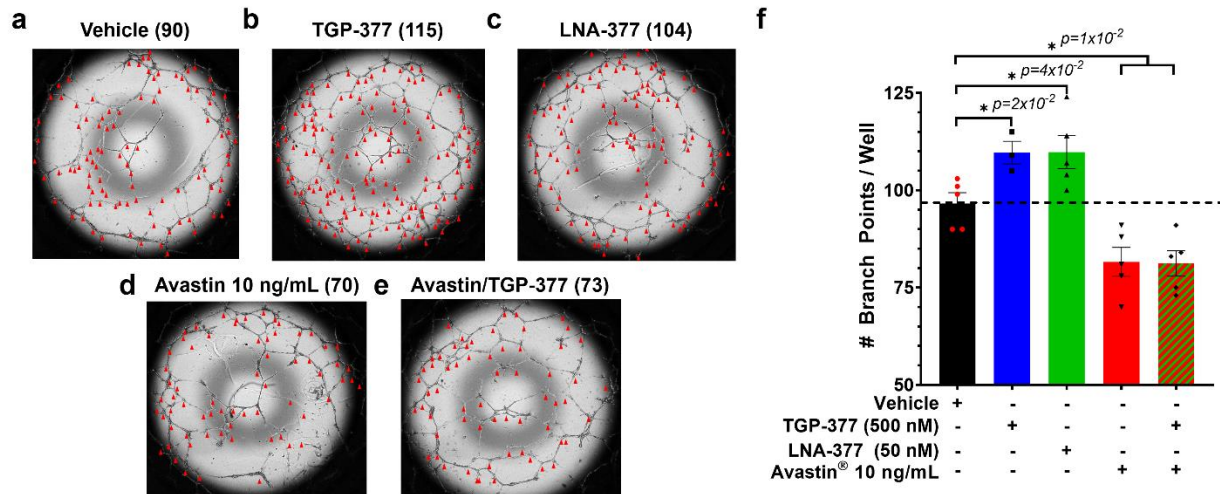
Supplementary Figure 19: Inhibition of phenotype by co-treatment with Avastin. a) Representative image of tubule networks formed by vehicle treated HUVECs. **b)** Representative image of HUVECs treated with **TGP-377** at 500 nM. **c)** Representative image of HUVECs treated with LNA-377 at 50 nM. Both LNA and **TGP-377** show similar increases in branch points. **d)** Representative image of HUVECs treated with Avastin shows a 20-30% decrease in tubule branching. **e)** Co-treatment of HUVECs with **TGP-377** and Avastin shows no enhancement in tubule branching suggesting inhibited compound function. **f)** Plot of mean tubule branch points per well. For all panels, data are reported as mean ± SEM (n = 5 biologically independent replicates from two independent experiments). All p values were determined by a two-sided Student's t-test; * = p<0.05.

A

B

C

D

RNA1-C2, Cluster_2347

E

RNA2-C1, Cluster_474
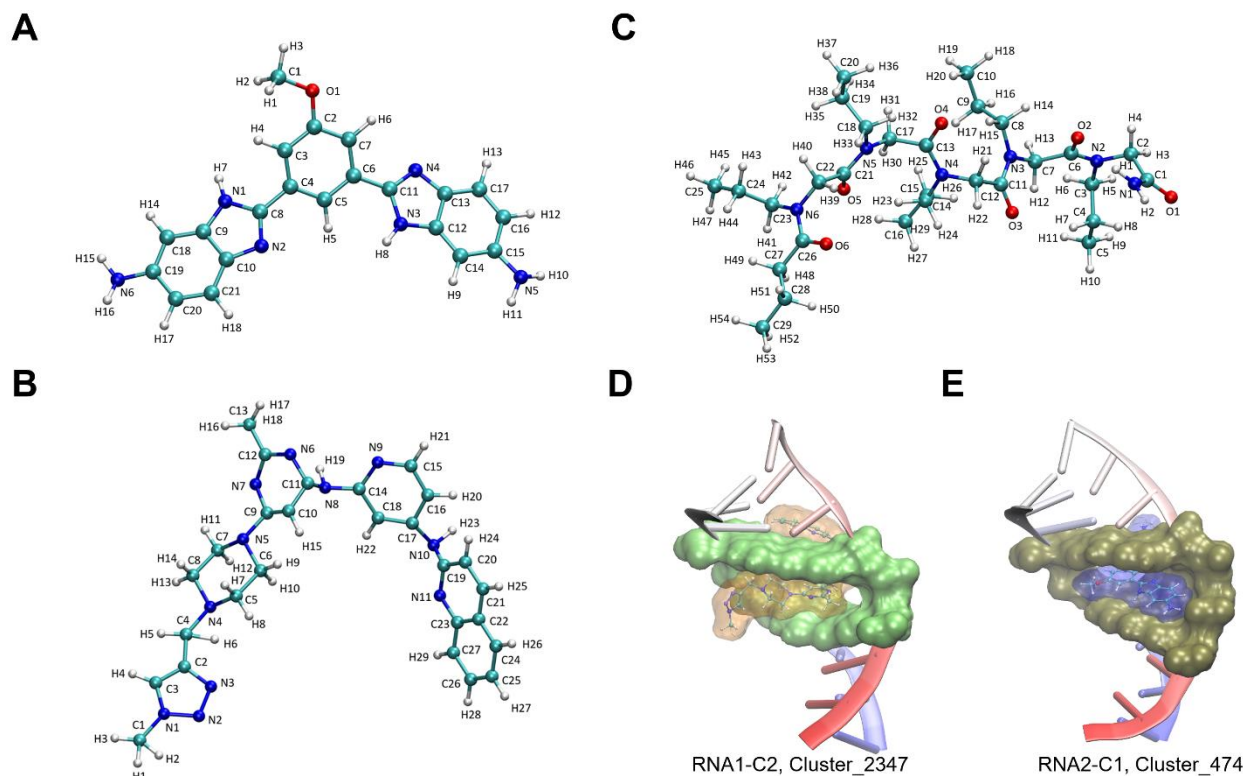
Supplementary Figure 20: Atomic diagrams and free energy binding models for TGP-377-pre-miR-377 modeling. a) Atom names used for Compound 1 (**C1**). **b)** Atom names used for Compound 2 (**C2**). **c)** Atom names used for the peptoid linker. **d – e)** The lowest binding free energy structures of (**d**) RNA1-C2 and (**e**) RNA2-C1 complexes. The molecular surface of A-bulge site (5'-GG-3'/5'-CAC-3') in RNA1 is displayed in green color and that of A-bulge site (5'-AAU-3'/5'-AU-3') in RNA2 is in tan color. The rest of the RNAs are displayed in NewCartoon representation. CPK model was used to represent **C2** and **C1** with transparent molecular surfaces highlighted in orange and blue colors, respectively. The corresponding computational parameters for each panel can be found in **Supplementary Tables 7 – Supplementary Figure 11**.

Supplementary Discussion

A statistical analysis of the RNA folds that bind small molecules was completed to define patterns in RNA folds that bind small molecules avidly and also to define the RNA folds that do not. LOGOS were generated for the RNAs with the highest and lowest 0.5% Z_{obs} (**Supplementary Figures 2, 3, and 4**).[2] LOGOS assign each nucleotide in a sequence alignment to bits of information with the height of the letter corresponding to the amount of information bits that each nucleotide in a sequence contributes to a consensus motif. The larger a nucleotide height, the greater number of data bits that are contributing to the consensus sequence. By comparing the LOGOS for related compounds, or DiffLOGOs,[3] one can define features that govern binding for the RNA folds.

Interestingly, inspection of the motifs that bound these compounds with $Z_{obs} > 8.0$ showed that single nucleotide bulges comprised 57% of all the motifs bound, followed by 1×1 internal loop, and 3×2 internal loops each comprising 14% of the total bound motifs. As motif size increases, such as going from a single nucleotide bulge to a two or three nucleotide bulge, the number of motifs bound diminishes starkly to <1%, with analogous changes observed for symmetric internal loops.

In general, motifs that are rich in GC base pairs have smaller unpaired regions, such as single nucleotide bulges derived from the **3×2** and **3×3 ILL**s. In contrast, U-rich internal loops of all sizes generally do not bind to small molecules (**Supplementary Figures 2, 3, and 4**). Compounds binding the **4×3 ILL** show less homogeneity in nucleotide preference and this is likely due to the large diversity of structures possible in the library. Nevertheless, for **4×3 ILL**, bulkier compounds like **3** and **6** exhibited a preference for AU rich folds compared to less bulkier compounds such as **7** and **8**, which did not show as large of a prevalence.

Translating these nucleotide preferences into 3D RNA folds shows that single nucleotide bulges comprise 70% (p < 0.0001) of the motifs bound, while 1×2 and 3×2 internal loops comprise 30% (p < 0.0001; **Supplementary Figure 5**). Interestingly, for 3D folds that do not bind compounds, the inverse is true, with bulges comprising 30% (p < 0.0001) and 1×2 and 3×2 internal loops comprising 70% (p < 0.0001) of the motifs that do not bind (**Supplementary Figure 6**). Further analysis also revealed that the incidence of 5'GC closing pairs in bound motifs was 36% (p < 0.0001). Motifs that did not bind also showed a higher prevalence of AU (81%, p < 0.0001) and GU (13%, p < 0.0001) loop closing pairs (**Figure 1C, and Supplementary Figure 5,** and **6**)

It is known that RNA structure can be dynamic, and both local and global dynamics can be observed.[4] Dynamics have been broadly shown to affect the ligand binding capacity of SMIRNAs and thus, based on observed patterns in the RNA folds that form, we studied the effect of dynamics on the unbound RNA target. For example, large uridine-rich internal loops do not, in general, bind SMIRNAs, as they are dynamic. In contrast, RNA motifs derived from **3×2 ILL** that display bulges, i.e., part of the randomized region forms base pairs, have lower potential for adopting multiple conformations than larger motifs and are thus likely to bind SMIRNAs. Indeed, dynamics has been shown to be a key determinant in a variety of RNA functions and local dynamics could affect ligand binding capacity broadly.[5,6] If the structure of a target RNA is too dynamic, ligand binding potential could be diminished. That is, a defined small molecule binding pocket may not be present, or the affinity of an ill-defined pock could be low due to energetic and entropic penalties associated with locking out multiple conformations. Previous studies have provided some insight into RNA structural dynamics such as smaller loops are not typically as conformationally dynamic as their larger counterparts and U-rich motifs can have dynamic character.[5,7]

Nuclear magnetic resonance (NMR) spectroscopy studies of RNAs that generally bind and do not bind small molecules were completed (**Supplementary Figure 7**). The RNAs 1 (blue) and 2 (red) were chosen because they appear in the RNAs within the top 0.5% of Z_{obs} scores, i.e., bind avidly, for >50% of the ligands while RNAs 3 (green), 4 (purple), and 5 (orange) were chosen because they appear in the lowest 0.5% of Z_{obs} scores, i.e., do not bind, >50% of the ligands. As shown in Figure S7, RNAs that bind small molecules have defined structure, as evidenced by the presence of all possible imino proton peaks. In contrast, various peaks from imino protons were absent in the spectra of RNAs that generally do not bind small molecules (**Supplementary Figure 7**). Our data suggest that a defined RNA structure is a contributing factor for SMIRNA binding, at least for the scaffolds studied herein. That is, binding is not observed to RNA motifs that are overtly dynamic and SMIRNA binding to them with sufficient affinity and selectivity could be a challenge.

Synthetic Methods and Characterization
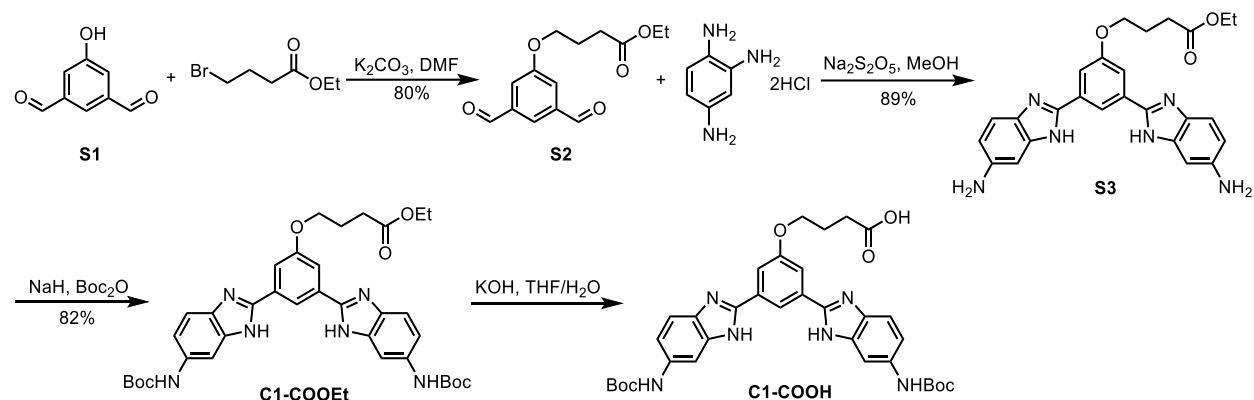
Supplementary Scheme 1. Synthesis of **C1-COOH**.

Compound S2: A solution of compound **S1**[8] (292 mg, 2 mmol), ethyl 4-bromobutyrate (585 mg, 3 mmol) and K_2CO_3 (414 mg, 3 mmol) in 10 mL DMF was stirred at 70 ^{o}C overnight. Then the mixture was filtered, and the filtrate was concentrated *in vacuo*. The residue was purified by column chromatography to give compound **S2** as a colorless oil (420 mg, 1.6 mmol, 80%).

^{1}H NMR (400 MHz, CDCl$_3$) δ (ppm) 10.05 (s, 2H), 7.96 (s, 1H), 7.64 (s, 2H), 4.15 (m, 4H), 2.54 (t, *J*=7.2 Hz, 1H), 2.18 (m, 2H), 1.27 (t, *J*=7.1 Hz, 3H);

^{13}C NMR (100 MHz, CDCl$_3$) δ (ppm) 190.9, 172.9, 160.0, 138.3, 124.2, 119.8, 67.6, 60.6, 30.6, 24.3, 14.2;

HRMS (m/z): calculated for $C_{14}H_{17}O_5$ [M+H]$^+$ 265.1071, found: 265.1078.

Compound S3: To a solution of Compound **S2** (250 mg, 0.95 mmol) in 10 mL MeOH was added a solution of $Na_2S_2O_5$ (376 mg, 1.9 mmol) in 2 mL H_2O at 0 ^{o}C. The mixture was stirred at 0 ^{o}C for 1 h. Then the 1,2,4-triaminobenzene hydrochloride salt (360 mg, 1.9 mmol) was added to the mixture. The reaction was stirred at reflux for 5 h. To the reaction was added H_2O after it cooled down to room temperature. The precipitate was collected by filtration and washed thoroughly by H_2O and diethyl ether to give compound **S3** as a brown solid (400 mg, 0.85 mmol, 89%).

^{1}H NMR (400 MHz, DMSO-d^6) δ (ppm) 8.59 (s, 1H), 7.83 (d, *J*=1.0 Hz, 2H), 7.57 (d, *J*=8.6 Hz, 2H), 7.19 (s, 2H), 6.97 (dd, *J*=8.7, 1.9 Hz, 2H), 4.22 (t, *J*=6.2 Hz, 2H), 4.10 (q, *J*=7.1 Hz, 2H), 2.55 (t, *J*=7.4 Hz, 2H), 2.10 (t, *J*=7.12 Hz, 2H);

^{13}C NMR (100 MHz, DMSO-d^6) δ (ppm) 172.6, 159.4, 148.9, 137.5, 133.0, 130.2, 117.4, 116.0, 115.9, 115.8, 114.1, 102.0, 87.3, 80.0, 30.1, 24.2, 14.2;

HRMS (m/z): calculated for $C_{26}H_{27}N_6O_3$ [M+H]$^+$ 471.2139, found: 471.2188.

Compound C1-COOEt(S4): A solution of compound **S3** (170 mg, 0.362 mmol), NaH (60%, 57.9 mg, 1.447 mmol) and Boc$_2$O (315 mg, 1.447 mmol) in 4 mL DMF was stirred at r.t. overnight. The mixture was extracted by ethyl acetate from H_2O for 3 times. The combined organic layer was washed by brine, dried over anhydrous. Na_2SO_4 and concentrated *in vacuo*. The residue was

purified by column chromatography to give compound **C1-COOEt(S4)** as a brown solid (200 mg, 0.30 mmol, 82%).

¹H NMR (400 MHz, CD₃OD) δ (ppm) 8.26 (s, 1H), 8.10 (s, 2H), 7.76 (d, *J*=1.4 Hz, 2H), 7.60 (d, *J*=8.8 Hz, 2H), 7.34 (dd, *J*=8.9, 1.8 Hz, 2H), 4.22-4.14 (m, 4H), 2.57 (t, *J*=7.4 Hz, 2H), 2.16 (m, 2H), 1.56 (s, 18H), 1.27 (t, *J*=7.1 Hz, 3H);

¹³C NMR (100 MHz, CD₃OD) δ (ppm) 174.7, 161.9, 155.0, 148.9, 139.9, 134.9, 130.1, 128.3, 119.5, 119.3, 117.8, 115.6, 103.4, 81.4, 69.2, 61.7, 31.4, 28.7, 25.5, 14.5;

HRMS (m/z): calculated for $C_{36}H_{43}N_6O_7$ [M+H]⁺ 671.3188, found: 671.3259.

Compound C1-COOH(S5): To a solution of Compound **S4** (200 mg, 0.30 mmol) in THF was added a solution of KOH (33.6 mg, 0.6 mmol) in H₂O. The mixture was stirred at 50 °C overnight. Then the mixture was neutralized by 3 M HCl and concentrated *in vacuo* to give **C1-COOH(S5),** which was directly used in the next step without further purification.
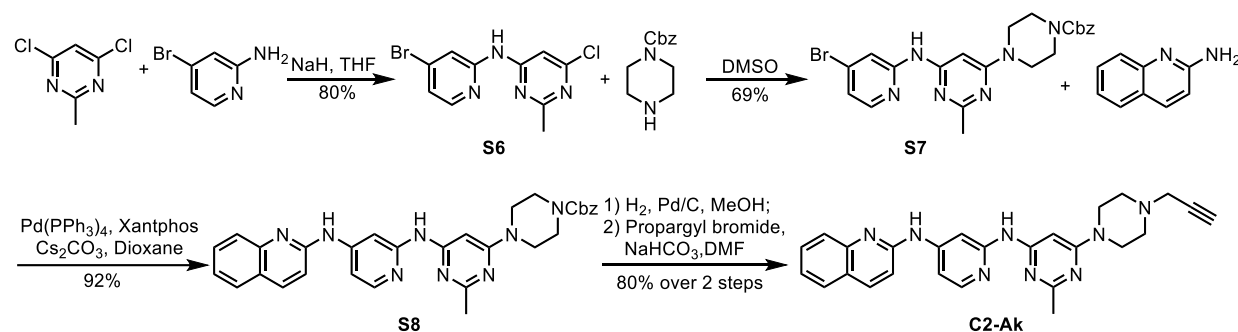
Supplementary Scheme 2. Synthesis of AZR B4-Alkyne **C2-Ak**

Compound S6: To a solution of 4,6-Dichloro-2-methylpyrimidine (3.26 g, 20 mmol), 2-amino-4-bromopyridine (3.46 g, 20 mmol) in 50 mL THF was added NaH (60%, 1.2 g, 30 mmol). The reaction mixture was stirred at 60 °C overnight. Then the mixture was acidified by 3 M HCl and extracted by ethyl acetate for 3 times. The combined organic layer was washed by brine, dried over *anhydrous.* Na₂SO₄ and concentrated *in vacuo*. The residue was washed with MeOH:H₂O (1:1) to give **S6** as a brown solid (4.8 g, 80%).

¹H NMR (400 MHz, DMSO-*d⁶*) δ (ppm) 10.64 (s, 1H), 8.22 (d, *J*=5.4 Hz, 1H), 7.80 (s, 1H), 7.27 (dd, J=5.4, 1.72 Hz, 1H), 2.48 (s, 3H);

¹³C NMR (100 MHz, DMSO-*d⁶*) δ (ppm) 167.5, 159.9, 159.3, 154.0, 148.9, 132.8, 120.7, 115.6, 103.7, 25.3;

HRMS (m/z): calculated for $C_{10}H_9BrClN_4$ [M+H]⁺ 298.9694, found: 298.9723.

Compound S7: To a solution of Compound **S6** (2 g, 6.69 mmol) in 50 mL DMSO was added benzyl piperazine-1-carboxylate (7.4 g, 33.44 mmol). The mixture was stirred at 130 °C for 3 h. The reaction mixture was diluted with EA and washed with H₂O for 3 times. The organic phase

S-27

was washed with brine, dried over *anhydrous.* Na_2SO_4 and concentrated *in vacuo*. The residue was purified by column chromatography. The product was further washed with diethyl ether to give **S7** as a white solid (2.53 g, 69%).

¹H NMR (400 MHz, DMSO-*d⁶*) δ (ppm) 9.85 (s, 1H), 8.13 (d, *J*=5.4 Hz, 1H), 7.87 (d, *J*=1.6 Hz, 1H), 7.39-7.31 (m, 5H), 7.27 (dd, *J*=5.4, 1.76 Hz, 1H), 3.60-3.45 (m, 8H), 2.33 (s, 3H);

¹³C NMR (100 MHz, DMSO-*d⁶*) δ (ppm) 165.4, 162.7, 159.3, 155.1, 154.5, 148.7, 136.8, 132.3, 128.4 127.9, 127.6, 119.1, 114.9, 84.3, 88.3, 43.1, 25.7;

HRMS (m/z): calculated for $C_{23}H_{26}BrN_6O_2$ [M+H]⁺ 483.1139, found: 483.1180.

Compound S8: A solution of compound **S7** (1.93 g, 4 mmol), 2-aminoquinoline (864 mg, 6 mmol), $Pd(PPh_3)_4$ (462 mg, 0.4 mmol), Xantphos (464 mg, 0.8 mmol) and Cs_2CO_3 (2.608 g, 8 mmol) in dioxane was stirred at reflux under the atmosphere of Argon. Then the mixture was filtered and the filtrated *in vacuo*. The residue was purified by C18 column and the product was washed with ether to give compound **S8** as a white solid (2 g, 3.66 mmol, 92%).

¹H NMR (400 MHz, DMSO-*d⁶*) δ (ppm) 11.49 (s, 1H), 10.83 (s, 1H), 8.65 (s,1H), 8.29 (d, *J*=8.8 Hz, 1H), 8.19 (d, *J*=7.0 Hz, 1H), 8.00 (d, *J*=8.3 Hz, 1H), 7.87 (d, *J*=7.3 Hz, 1H) 7.72 (m,1H), 7.47 (m,1H), 7.39-7.31 (m, 5H), 7.30-7.22 (m, 2H), 6.20 (s, 1H), 5.12 (s, 2H), 3.61-3.52 (m, 8H), 2.49 (s, 3H);

¹³C NMR (100 MHz, DMSO-*d⁶*) δ (ppm) 164.7, 161.8, 158.7, 158.5, 158.4, 154.5, 152.5, 152.3, 150.8, 146.0, 138.2, 136.8, 129.8, 128.5, 127.9, 127.7, 127.3, 124.8, 124.5, 114.4, 107.8, 97.9, 84.0, 66.4, 43.1, 42.8, 25.2;

HRMS (m/z): calculated for $C_{31}H_{31}N_8O_2$ [M+H]⁺ 547.2564, found: 547.2609.

Compound C2-Ak: To a solution of compound **S8** (1 g, 1.83 mmol) in 20 mL MeOH was added 10% Pd/C (100 mg). The mixture was stirred at r.t. under the atmosphere of H_2 overnight. Then the reaction was filtered, and the filtrate was concentrated *in vacuo*. The residue was dissolved in 10 mL of DMF and to this solution was added propargyl bromide (261 mg, 2.20 mmol) and $NaHCO_3$(185 mg, 2.2 mmol). The reaction mixture was stirred at r.t. overnight. Then the reaction mixture was filtered and concentrated *in vacuo*. The residue was purified by column chromatography to give compound **C2-Ak** as a white solid (652 mg, 1.45 mmol, 80% over 2 steps).

¹H NMR (400 MHz, CD₃OD) δ (ppm) 8.72 (s, 1H), 8.19 (d, *J*=8.8 Hz, 1H), 8.07 (d, *J*=7.0, 1H), 8.01 (d, *J*=8.4 Hz, 1H), 7.80 (d, *J*=8.1 Hz, 1H), 7.7 (m, 1H), 7.46 (m, 1H), 7.17 (d, *J*=6.9 Hz, 1H), 7.12 (s, 1H), 5.99 (s, 1H), 3.71 (s, 4H), 3.40 (d, *J*=2.4 Hz, 2H), 2.73 (t, *J*=2.3 Hz, 1H), 2.52 (s, 3H);

¹³C NMR (100 MHz, CD₃OD) δ (ppm) 166.4, 163.9, 160.2, 153.9, 153.1, 139.7, 138.2, 131.2, 128.8, 128.6, 126.5, 126.2, 115.1, 109.2, 99.1, 84.6, 78.8, 75.4, 52.3, 47.3, 44.8, 25.3;

HRMS (m/z): calculated for $C_{26}H_{27}N_8$ [M+H]⁺ 451.2353, found: 451.2376.
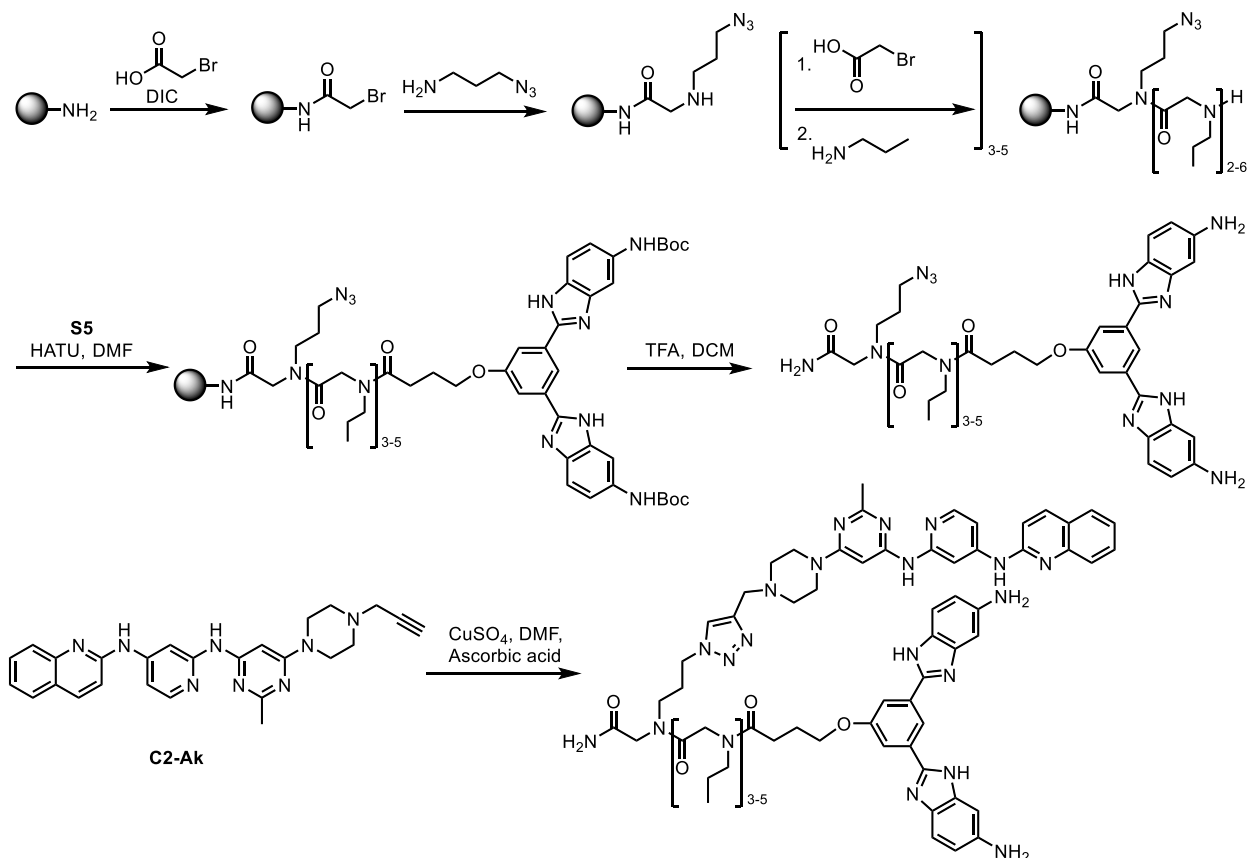
Supplementary Scheme 3. General protocol for dimer synthesis.

General Protocol for Peptoid Synthesis: Peptoids were synthesized via standard resin-supported oligomerization protocol. Rink Amide Resin (555 mg, 0.6 mmol) was activated with 1 M HCl/dioxane in DCM(4 M HCl dioxane was diluted with DCM) for 30 min. After that, solvent was removed and washed with DMF and DCM for 3 times respectively.

Coupling Step: To the resin was added 3 mL of 1 M bromoacetic acid in DCM (3 mmol, 5 eq) and DIC (3.0 mmol, 519 uL). The resin was shaken at r.t. for 2h. Then the solvent was removed, and the resin was washed with DMF for 3 times.

Displacement step: To the resin was added 5 mL DMF and 3-azidopropylamine. The resin was shaken at r.t. for 2 h. Then the solvent was removed, and the resin was washed with DMF for 3 times.

Peptoid Chain Extension: a) To the resin was added 5 mL DMF, bromoacetic acid and DIC. The resin was shaken at r.t. for 2 h. Then the solvent was removed, and the resin was washed with DMF for 3 times') To the resin was added 5 mL DMF and propyl amine. The resin was shaken at r.t. for 2 h. Then the solvent was removed, and the resin was washed with DMF for 3 times. Steps a) and b) were repeated for another 1-5 times

Coupling of C1-COOH to the Peptoid: To the resin was added 5 mL DMF, compound **C1-COOH** (3 eq) and HATU (3 eq). The resin was shaken at r.t. overnight. Then the solvent was removed, and the resin was washed with DMF for 3 times and then DCM for 3 times. Then the resin was treated with 30% TFA in DCM and shaken at r.t. for 30 min. The crude was collected, concentrated in *vacuo* and purified by HPLC.

Click C2-Ak to the Peptoid: A solution of the above purified peptoid product, **C2-Ak** (1 eq), CuSO$_4$•5H$_2$O (1 eq) and ascorbic acid (1 eq) in DMF was stirred at r.t. overnight. Then the mixture was purified by HPLC to give dimer as a yellow solid. The dimers were purified by preparative HPLC (Waters 1525 Binary HPLC Pump equipped with a Waters 2487 Dual Absorbance Detector system) with either a reverse phase Atlantis® Prep T3 C18 5 µM column or a Sunfire Prep C18 5 µM 19 × 150 mm column. The HPLC separations employed a linear gradient from 0% to 100% B in A over 60 min and a flow rate of 5 mL/min (A: water + 0.1% (v/v) trifluoroacetic acid (TFA); B: methanol + 0.1% (v/v) TFA.). The purity of the final dimers was tested on a reverse phase Waters Symmetry C18 5 µm 4.6 × 150 mm column at room temperature. A flow rate of 1 mL/min and a linear gradient of 0–100% B in A. Absorbance was monitored at 254 nm. All dimers were ≥95% pure.
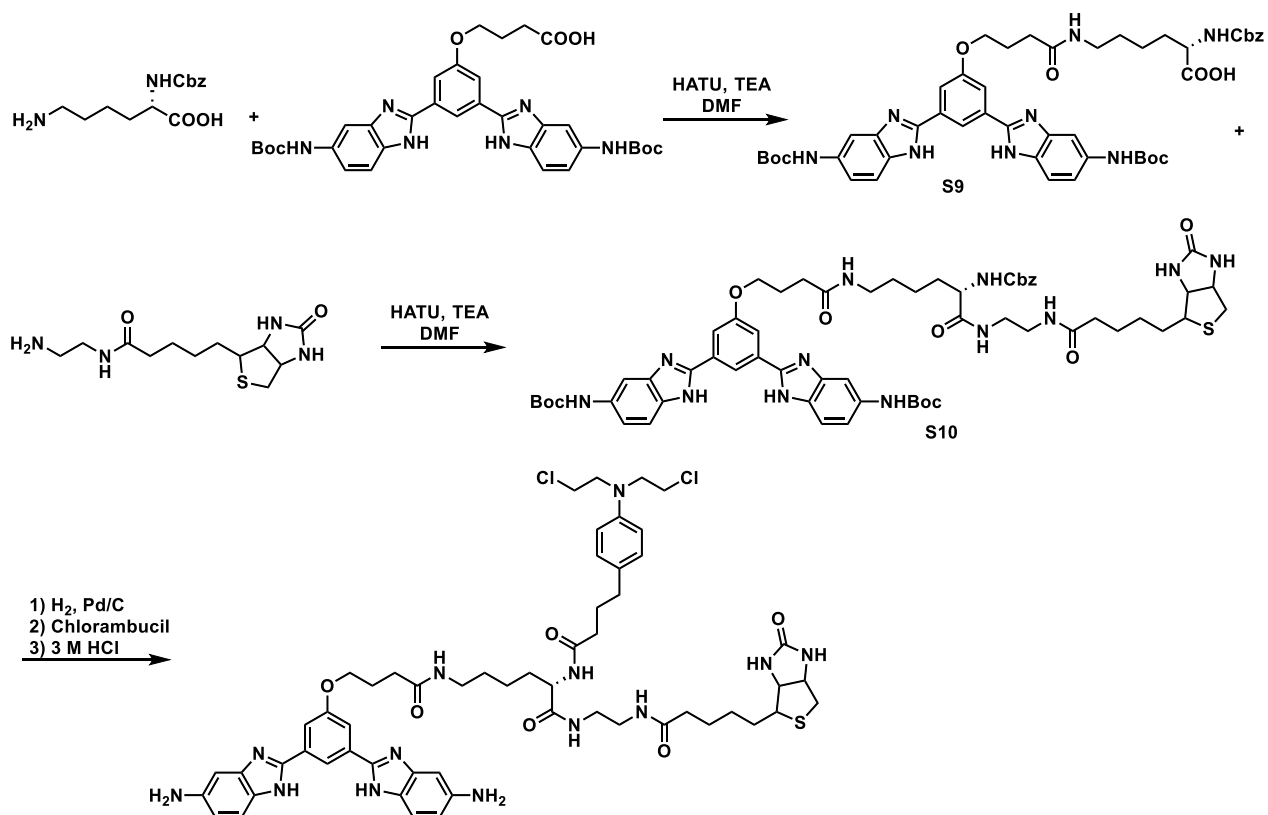

**Supplementary Scheme 4.** Synthesis of Chem-CLIP probe.

A solution of **C1-COOH** (1 equiv.), HATU (1.2 equiv.) and TEA (1.5 equiv.) in DMF was stirred at r.t. for 20 min, followed by the addition of Z-Lys-OH. The mixture was stirred at r.t. for another 1 h. The product was purified by HPLC to give **S9**. **S9** (1 equiv.), Biotin-amine (1.2 equiv.), HATU (1.1 equiv.), and TEA (3 equiv.) in DMF was stirred at r.t. for 30 min. The product was purified by HPLC to give **S10**. The **S10** and Pd/C in MeOH was stirred under the atmosphere of H_2 overnight at r.t. and then the Pd/C was filtered off. The filtrate was concentrated *in vacuo*. Chlorambucil (1.2 eq), HATU (1.2 eq) and TEA (2 eq) in DMF was stirred at r.t. for 20 min and then the above amine in DMF was added to this solution and the reaction mixture was stirred at r.t. for 30 min. The product was purified by HPLC and then dissolved in 3 M HCl in MeOH/H_2O(3:1) at 0 ^{o}C. The reaction was slowly warmed to room temperature and stirred at r.t. for 3 h. The Chem-CLIP product was purified by HPLC.
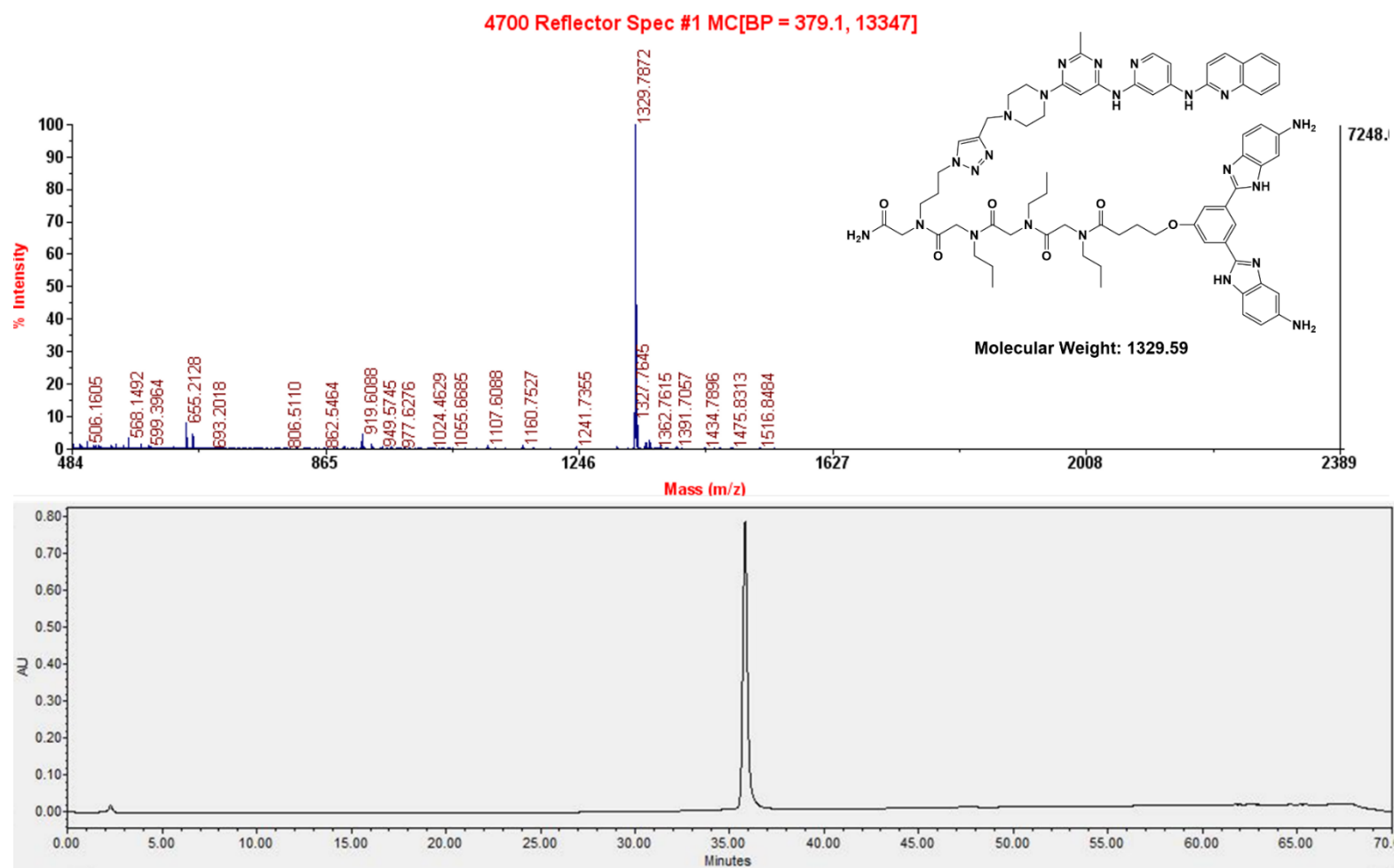
Supplementary Figure 21: Characterization of dimer n = 3. (Top) MALDI-MS, (Bottom) Analytical HPLC 254 nm.
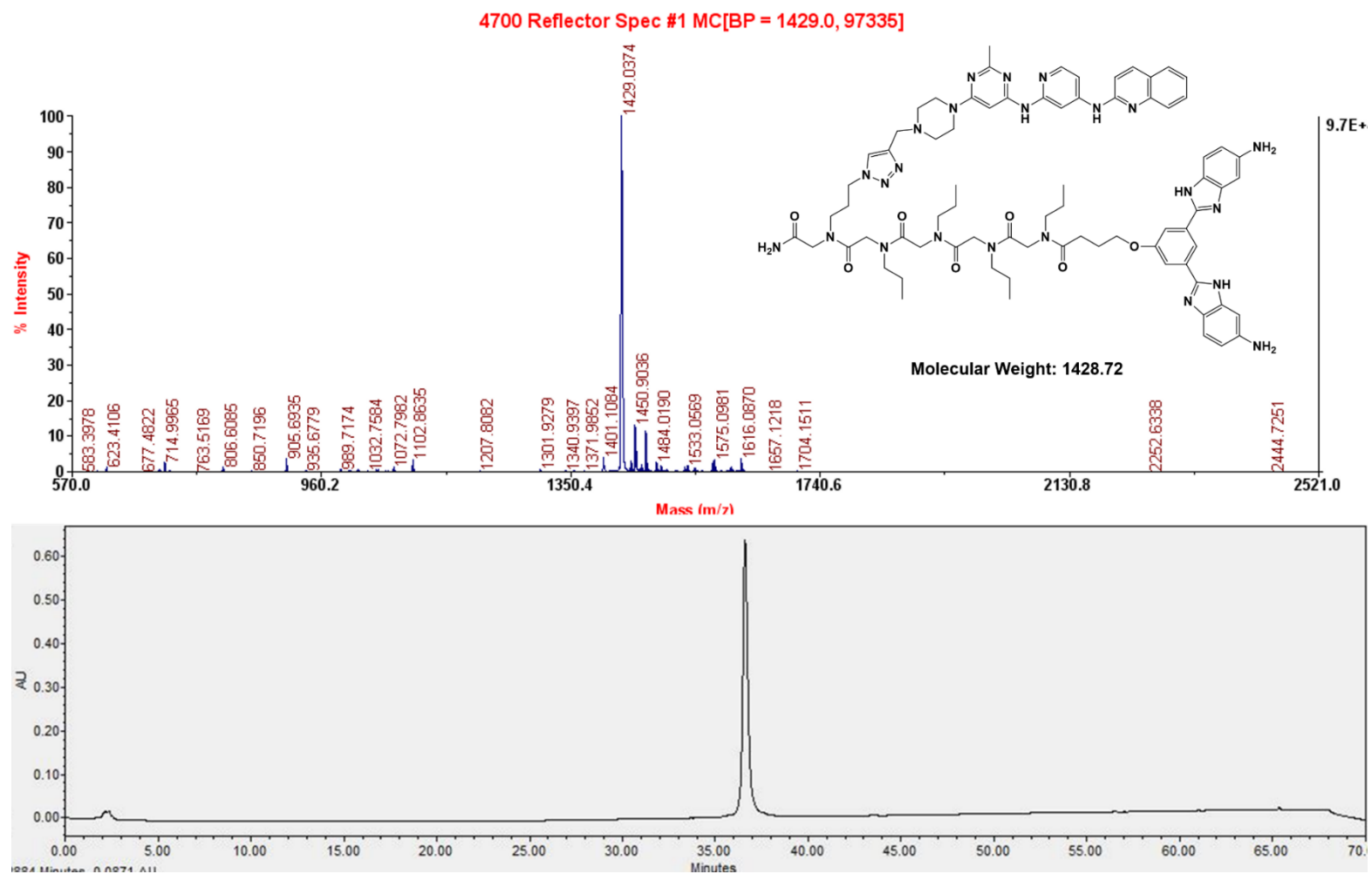
Supplementary Figure 22: Characterization of dimer n = 4. (Top) MALDI-MS, (Bottom) Analytical HPLC 254 nm.

Supplementary Figure 23: Characterization of dimer n = 5. (Top) MALDI-MS, (Bottom) Analytical HPLC 254 nm.
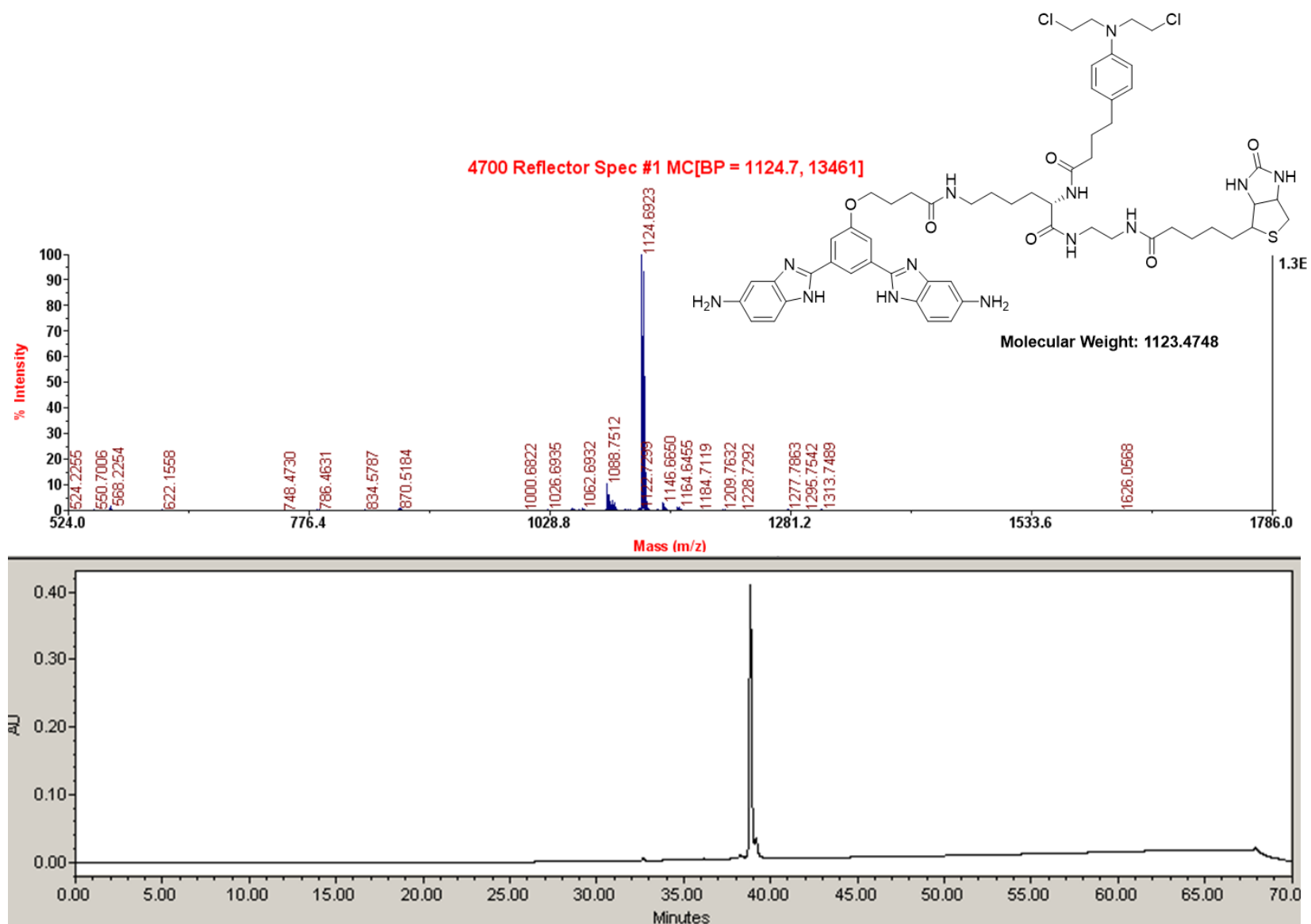
Supplementary Figure 24: Characterization of Chem-CLIP probe (TGP-377/421-CA-Biotin). (Top) MALDI-MS, (Bottom) Analytical HPLC 254 nm.

Supplementary Methods

Dicer Inhibition Assay. Transcripts of pre-miR-377, Dicer Mutant, or the Double mutant were 5'-end labeled with γ-^{32}P and purified on a denaturing 15% polyacrylamide gel. Purified RNAs were incubated with either **TGP-377/421** or **TGP-377** at concentrations of 50,000 nM and 500 nM, respectively, for 15 min, followed by addition of recombinant Dicer (0.0017 U/µL final), and incubation at 37°C for 1 h. RNA fragments were resolved on a denaturing 15% polyacrylamide gel, imaged, and quantified.

Cell Viability. HUVECs were seeded into a 96-well tissue culture plate (Corning) at 5,000 cells/well and allowed to adhere for 12 h. Cells were then treated with DMSO (0.8%); **TGP-377/421** at concentration of 5, 10, 20, 40 and 80 µM; or **TGP-377** at concentrations of 5, 50, 500 and 5000 nM for 48 h. **C2** viability was tested by dosing HUVECs with **C2** at 1.25, 2.5, 5, and 10 µM for 48 h. Cell viability was measured using Cell Titer Fluor viability reagent (Promega) according to the manufacturer's protocol.

RNA isolation and RT-qPCR. Total RNA was extracted from cells using Quick-RNA Miniprep (Zymo Research) per the manufacturer's protocol. For measurement of mature miRNA levels, 200 ng of total RNA was used in reverse transcription reactions using mi Script II RT kit (Qiagen) per the manufacturer's protocol. For measurement of precursor and mRNA levels, 400 ng of total RNA was used in reverse transcriptase reactions using qScript reverse transcriptase (Quantabio) per the manufacturer's protocol. RT-qPCR primers can be found in **Supplementary Table 13**. RT-qPCR samples were prepared using Power SYBR Green PCR Master Mix (Applied Biosystems) and amplification was completed on a 7900HT Fast Real Time PCR System in 384-well format (Applied Biosystems). RNA expression was normalized to RNU6 small nuclear RNA, 18S or β-Actin mRNA for mature, precursor, and mRNA genes, respectively.

Lentiviral transduction. HUVEC cells were transduced to express an anti-Vegfa shRNA as follows: To generate viral particles the following plasmids were transfected into HEK293T cells: (i) anti-VEGFA shRNA plasmid clone set (NM_1204384.1-Genecopoeia); (ii) packaging plasmid (psPAX2-Addgene); and (iii) envelop plasmid (pMD2.G-Addgene), using Jetprime (Polyplus) according to the manufacturers protocol in the following molar ratio (1.0 : 0.55 : 1.3 pmol). After changing the transfection medium, supernatants were harvested at 12, 24 and 48 h. Supernatants were filtered through a 0.4 µm syringe filter and dosed into 100 mM dishes of HUVECs in equal volumes to growth medium. To sort cells, they were grown to 80 – 90% confluency and then detached using Acutase (Innovative Cell Technologies Inc.), washed with 1× PBS and resuspended in sort buffer (1× PBS supplemented with 5 mM EDTA, 25 mM HEPES, pH 7.0, 5% FBS (heat denatured), and10 units/mL DNase II) at a density of 5 x 10^6 cells/mL. GFP positive cells (shRNA expressing) were sorted on a BD FACS Aria Fusion™ cell sorter. Lentivirus titers were measured using Lenti-Go-Stix (Takara) according to the manufacturer's protocol. Titers were greater than 5x10^6 IFU/mL at time of transduction.

RAN Translation Assay. HEK2937 cells were grown to 80% confluency in a 100 mm dish in 1× DMEM supplemented with 10% FBS and 1× Pen/Strep. The media was changed to 10 mL of Opti-MEM (Life Technologies) and the cells were co-transfected with 1 µg of r(G$_4$C$_2$)$_{66}$-Nano-Luc and 1 µg of SV40-Firefly-Luc for 6 hr. After transfection, the cells were washed 3× with 1× DPBS and then fresh media was added, and the cells allowed to recover for 12 h. Once recovered, the cells were then seeded into Greiner 384-well white clear bottom plates (781098) at 50% confluency and allowed to adhere for 8 h. Once adhered cells were treated with **C20** in a dose response from 6.25 µM to 0.781 µM by 1:2 dilution. Cell viability was measured using 5× Cell Titer Fluor (Promega) according to the manufacturers protocol. Luciferase signal was measured

using the Dual Glow-Nano-Luciferase kit (Promega) according to the manufacturers protocol. RANT data was normalized to viability and then DMSO to obtain percent Luciferase activity.

Global proteomics profiling using LC-MS/MS. Cells were resolubilized in 1× PBS, lysed via sonication and protein concentration was determined using a Bradford assay (Bio-Rad). Samples (20 μg) were denatured with 6 M urea in 50 mM NH_4HCO_3, reduced with 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) for 30 min, and finally alkylated with 25 mM iodoacetamide for 30 min in the dark. Samples were diluted to 2 M urea with 50 mM NH_4HCO_3, and digested with trypsin (1 μL of 0.5 μg/μL) in the presence of 1 mM $CaCl_2$ for 12 h at 37 °C. Samples were acidified with acetic acid to a final concentration of 5%, desalted over a self-packed C18 spin column, and dried. Samples were analyzed by LC-MS/MS (see below) and the MS data was processed with MaxQuant (see below).

LC-MS/MS analysis. Peptides were resuspended in water with 0.1% formic acid (FA) and analyzed using EASY-nLC 1200 nano-UHPLC coupled to Q Exactive HF-X Quadrupole-Orbitrap mass spectrometer (Thermo Scientific). The chromatography column consisted of a 30 cm long, 75 μm i.d. microcapillary capped by a 5 μm tip and packed with ReproSil-Pur 120 C18-AQ 2.4 μm beads (Dr. Maisch GmbH). LC solvents were 0.1% FA in H_2O (Buffer A) and 0.1% FA in 90% MeCN: 10% H_2O (Buffer B). Peptides were eluted into the mass spectrometer at a flow rate of 300 nL/min over a 240 min linear gradient (5-35% Buffer B) at 65 °C. Data were acquired in data-dependent mode (top-20, NCE 28, R = 7'500) after full MS scan (R = 60'000, m/z 400-1'300). Dynamic exclusion was set to 10 s, peptide match to prefer, and isotope exclusion was enabled.

MaxQuant analysis. MS data were analyzed with MaxQuant[9] (V1.6.1.0) and searched against the human proteome (Uniprot) and a common list of contaminants (included in MaxQuant). The first peptide search tolerance was set at 20 ppm; 10 ppm was used for the main peptide search; and fragment mass tolerance was set to 0.02 Da. The false discovery rate for peptides, proteins and sites identification was set to 1%. The minimum peptide length was set to 6 amino acids and peptide re-quantification, label-free quantification (MaxLFQ), and "match between runs were enabled. The minimal number of peptides per protein was set to two. Methionine oxidation was searched as a variable modification and carbamidomethylation of cysteines was searched as a fixed modification.

LC-MS/MS Analysis of Cellular Concentration. C1 was quantitated in medium and cells using LC-MS/MS analysis. Cell medium or pelleted cells were treated with 4x v:v acetonitrile containing 1 μM carbamazepine which served as an internal standard. Precipitated protein was removed by filtration through a 0.2-micron filter membrane prior to LC-MS/MS analysis. Instrument setting are provided in the table below in **Supplementary Method Tables 1 to 3**.

Supplementary Method Table 1. Instrument conditions	
Instrument	LC (Shimadzu UFLC XR)
Compound	C1
Column	Phenomenex Synergi 4µ Polar-RP 80A
Mobile phase	A: Water with 0.1% Formic Acid B: Acetonitrile with 0.1% Formic Acid
Flow rate (ml/min)	0.35
Temperature (°C)	35
Injection volume(µl)	10

Supplementary Method Table 2. Gradient conditions

Time (min)	Mobile phase A (%)	Mobile phase B (%)
0.2	90	10
0.5	90	10
2.0	5	95
3.0	5	95
4.0	90	10
5.9	90	10

Supplementary Method Table 3. MS (API5500) conditions

Compound	C1
MRM(+)	341.2/107.2
Collision Gas	7
Curtain GAS	20
Ion Source Gas1	40
Ion Source Gas2	40
Ion Spray Voltage	4500
Temperature	500
Collision Energy	30
De-clustering Potential	70
Entrance Potential	10

Chemoinformatics Analysis. All computation of physiochemical properties was completed using Instant J Chem (Chem Axon) or RD Kit on an academic license. Scaffold analysis was completed using Scaffold Hopper (Tripod Development). Molecular fingerprint similarities were obtained using the RDKit Morgan fingerprint analysis with a, radius of 2 and 2048 bits - Landrum, G. RDKit: Open-source cheminformatics. http://www.rdkit.org.

Computational Methods

Parameterization of RNA and RNA binder. Two RNA molecules, r(5'-CUUGGUGA-3'/5'-UCACACAAA-3') (RNA1), and r(5'-GUGAAUUCG-3'/5'-UGAAUCAC-3') (RNA2-Dicer site), were generated to represent the two binding sites in pre-mir-377 (**Figure 2B**) by nucgen module of AMBER 16[10]. Amber99[11] force field with revised χ[12] and α/γ[13] torsional parameters were used to describe the RNAs. Watson-Crick (WC) base pairing, torsional, and chirality restraints were used to maintain the A-form geometry during the targeted binding process.

Three different compounds were prepared to build the RNA binder, TGP-377: compound 1 (C1), compound 2 (C2), and the peptoid based linker (**Figure 2C and Supplementary Figure 20A-C**). We attached -$C_3H_4N_3$ and -OCH_3 to C1 and C2, respectively, to join C1 and C2 to the peptoid linker. The number of repeat units for the linker is four as shown in **Figure 2C**. The Generalized Amber Force Field (GAFF)[14] was used to parameterize the compounds and the linker. Gaussian09[15] was used to calculate the RESP[16] charges for C1 and C2 as described before[1,17,18]. Initially, we optimized the structures, and then calculated the electrostatic potentials at a set of grid points using the HF/6-31G* basis set.[16,19] The atomic charges for the linker were calculated using AM1-BCC[20] method built in Antechamber[21]. All the force field parameters and charge values are displayed in **Supplementary Tables 8-10**.

Binding studies. We applied the dynamic binding method described before[1,17,18] to find the global minimum structures of RNA1-C2 and RNA2-C1 complexes. All the simulations for the binding study were performed under the conditions of modified implicit solvent model (GB^{OBC})[22] with 0.3M salt concentrations. The detailed processes of the dynamic binding are as follows: Initially, the compound (C1 or C2) was placed 40 Å away from the RNA with A-bulge (RNA1 or RNA2). We utilized a reaction coordinate, which is defined as the distance between the center-of-mass (COM) of the heavy atoms of the flanking bases (with respect to A-bulge) and the COM of the heavy atoms of the compound, to study the dynamic binding process. Next, we forced the compound to interact with the binding site so that we can determine initial binding modes for the MD simulations. In the initial binding process, the compound was slowly moved towards the bulge site, which we defined as 'move-close' process, by decrements of 1 Å until the reaction coordinate became 0 Å. During this 'move-close' process, we maintained the RNA structure in an A-form geometry by imposing Watson-Crick (WC) base pairing restraints on the WC base-pairs, and torsional and chirality restraints on all the RNA residues except the A-bulge region. We called this restraint set as 'flex-restraints'. This made the bulge site to act naturally while interacting with the compound. Once the reaction coordinate hit 0 Å, we then moved the compound away from the bulge-site, which we defined as 'move-away' process, by increments of 1 Å until the distance reached 40 Å. During this so-called 'move-away' process, the RNA was forced to return back to an A-form geometry by imposing WC base pairing, torsional, and chirality restraints on all the RNA residues so that the apo-structure of A-bulge was maintained when the compound was far away from the RNA. This process was repeated fifty times sequentially to create fifty different initial bound states for the RNA-compound complex. These initial states were then used in MD studies to perform fifty independent implicit-solvent MD simulations. In these MD simulations, a modified 'flex-restraints' were imposed on the system where the restraints on the reaction

coordinate was set to kick in when the COM distances was greater than 10 Å. This allowed the compound to reorient itself according to the force field to find the preferred binding state. Each MD simulation was run for 120 ns producing a total of 6 µs combined MD trajectory, which was used in the cluster analysis.
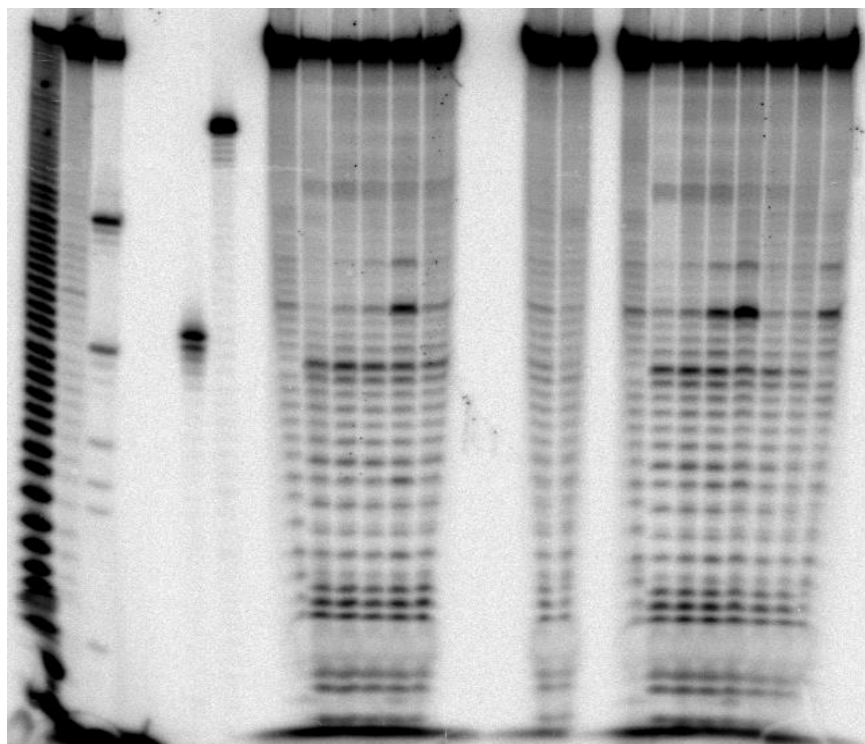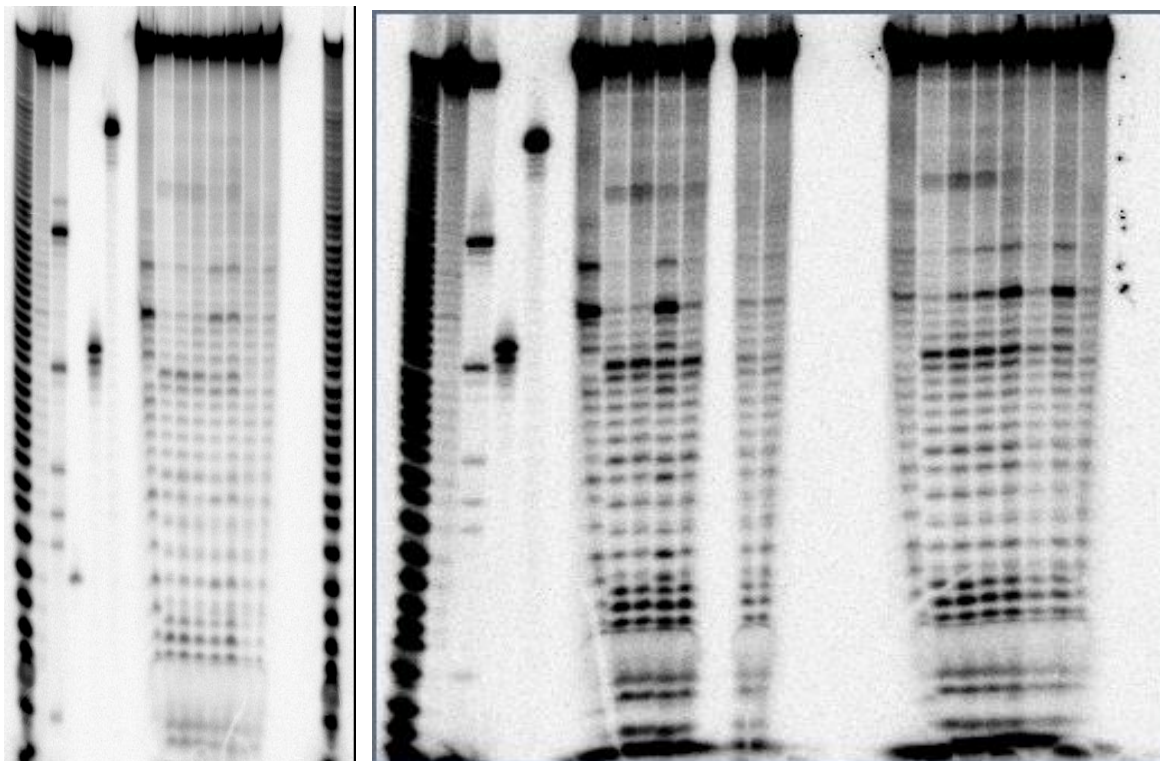
Cluster Analysis. Cluster analyses was carried out using an in-house code. The combined MD trajectories included 60K snapshots stored every 0.1 ns of the MD simulations. Root-mean-square deviation (RMSD) was calculated through the whole snapshots, and the snapshots with RMSD <= 1.5 Å for RNA1-C2 and with RMSD <=1.0 Å for RNA2-C1 were clustered into the same group. The different cutoff criterion was set in order to obtain sufficient number of clusters. The cpptraj[23] module of AMBER16 was used for the RMSD calculation. Symmetry states of C1 and C2 were considered while calculating the RMSD.

Relative binding free energy calculations using MM-PBSA. MM-PBSA analyses were conducted on each cluster to determine the lowest binding free energy states for the RNA1-**C2** and RNA2-**C1** complexes. The MMPBSA.py module of AMBER16 was used and applied on each cluster having more than 100 snapshots. The results of relative binding free energies for RNA1-**C2** and RNA2-**C1** are displayed in **Supplementary Tables 11** and **12**.
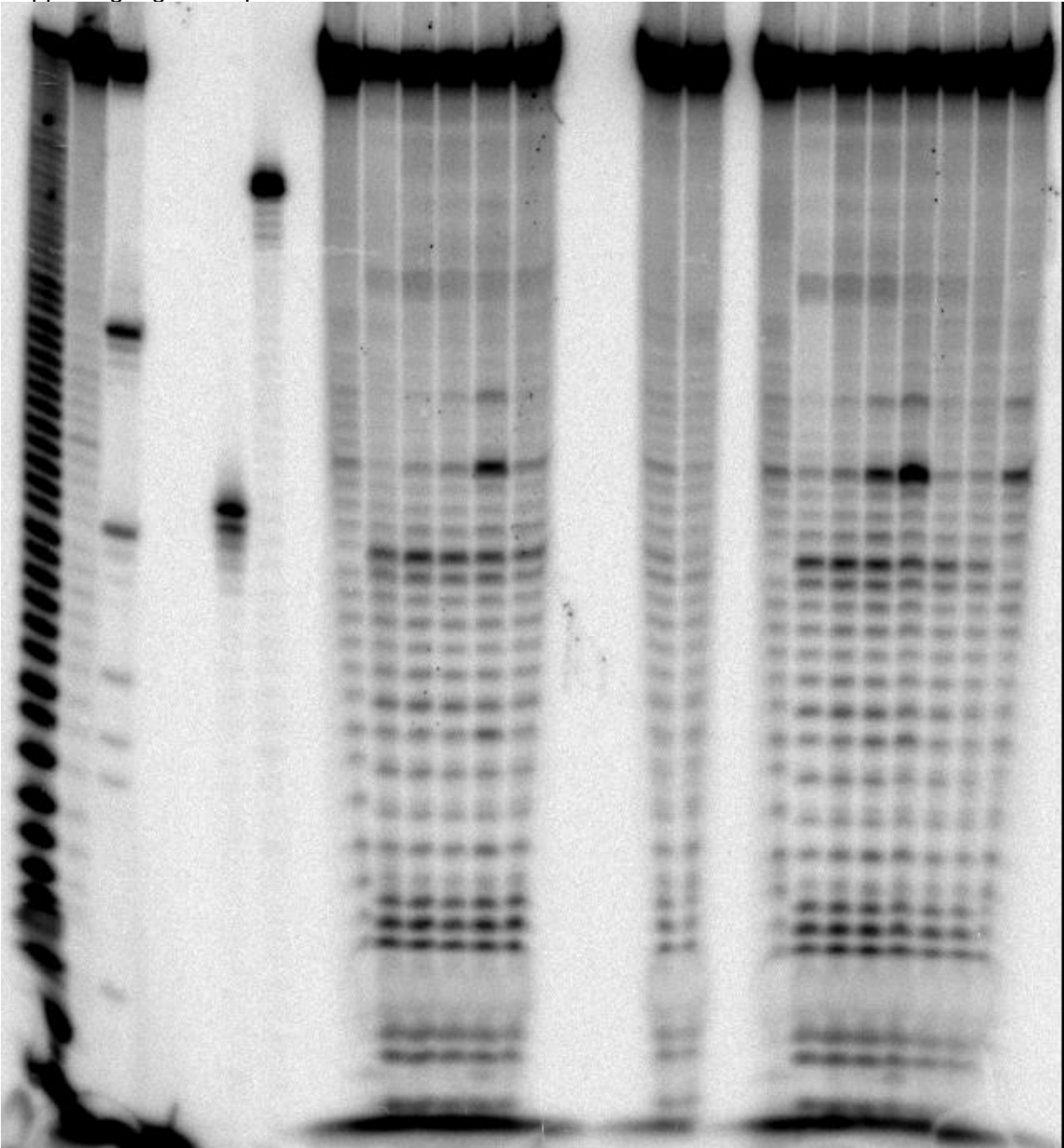
Building the 3D structure of TGP-377/pre-miR-377. The full sequence of miR-377 (**Figure 2B**) was used to model the bound state of **TGP-377** to pre-miR-377. The RNA strand folded and formed base pairs with its complementary parts of the same strand, and the hairpin structure was expressed based on the position shown in **Figure 2B**. The lowest binding free energy states of RNA1-C2 and RNA2-C1 (**Supplementary Figure 20D and 20E**) were used to model the bound state of **TGP-377** to pre-miR-377. The two A-bulge sites seen in pre-miR-377 were replaced with the structures shown in **Supplementary Figure 15**. VMD (Visual Molecular Dynamics)[24] was used to homology model the bound states on pre-miR-377. A peptoid linker was then attached at the ends of the two compounds and minimized to create **TGP-377**, which interacts with the pre-miR-377 at the A-bulge sites as shown in **Figure 2D**.
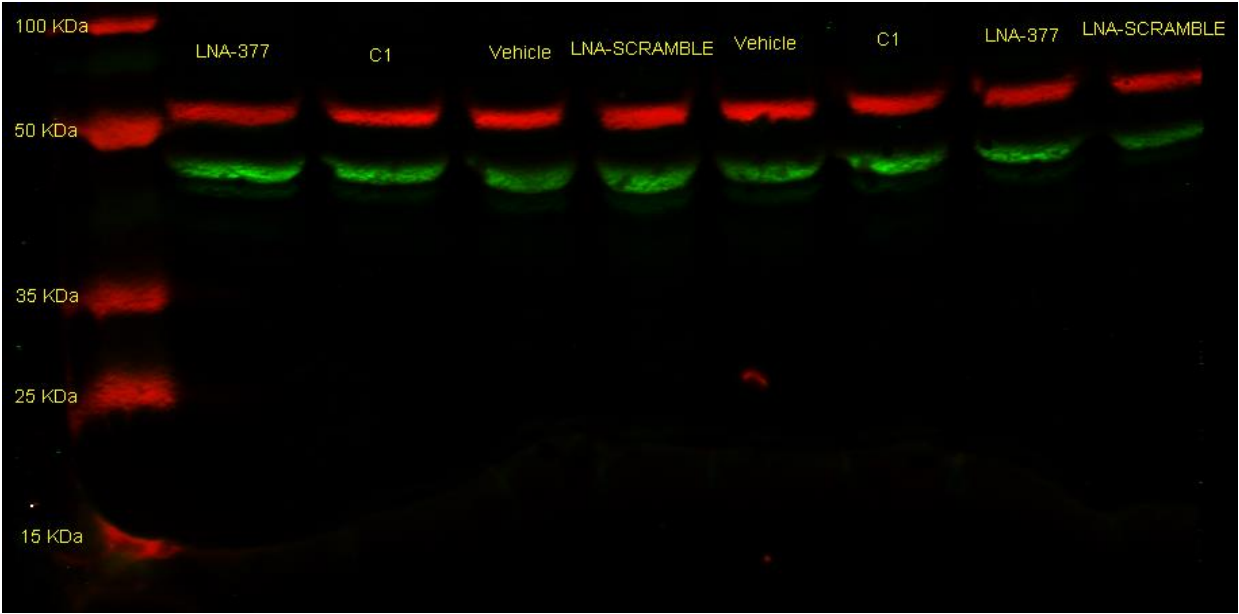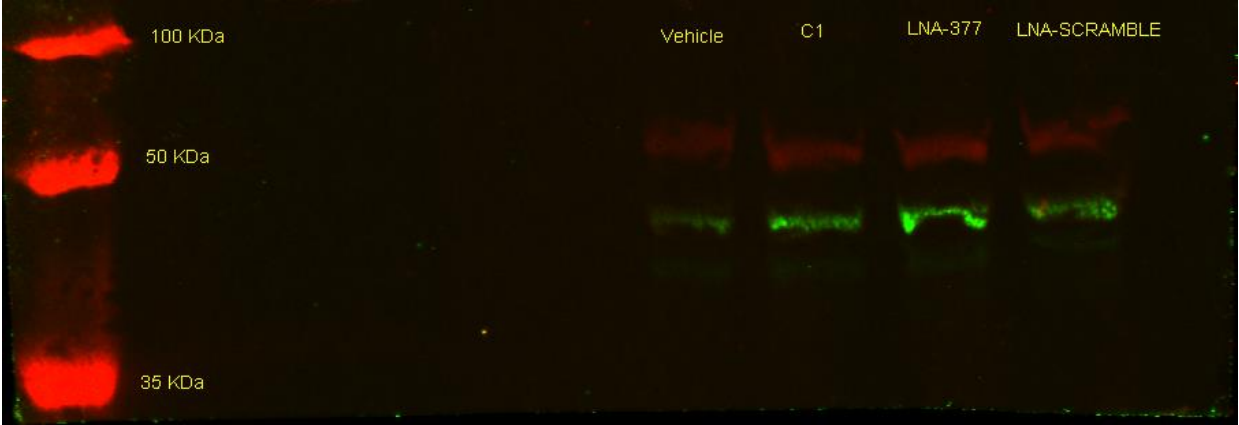
Full Length Gels

Supporting Figure 10 panel A (Gels 1 - 3)
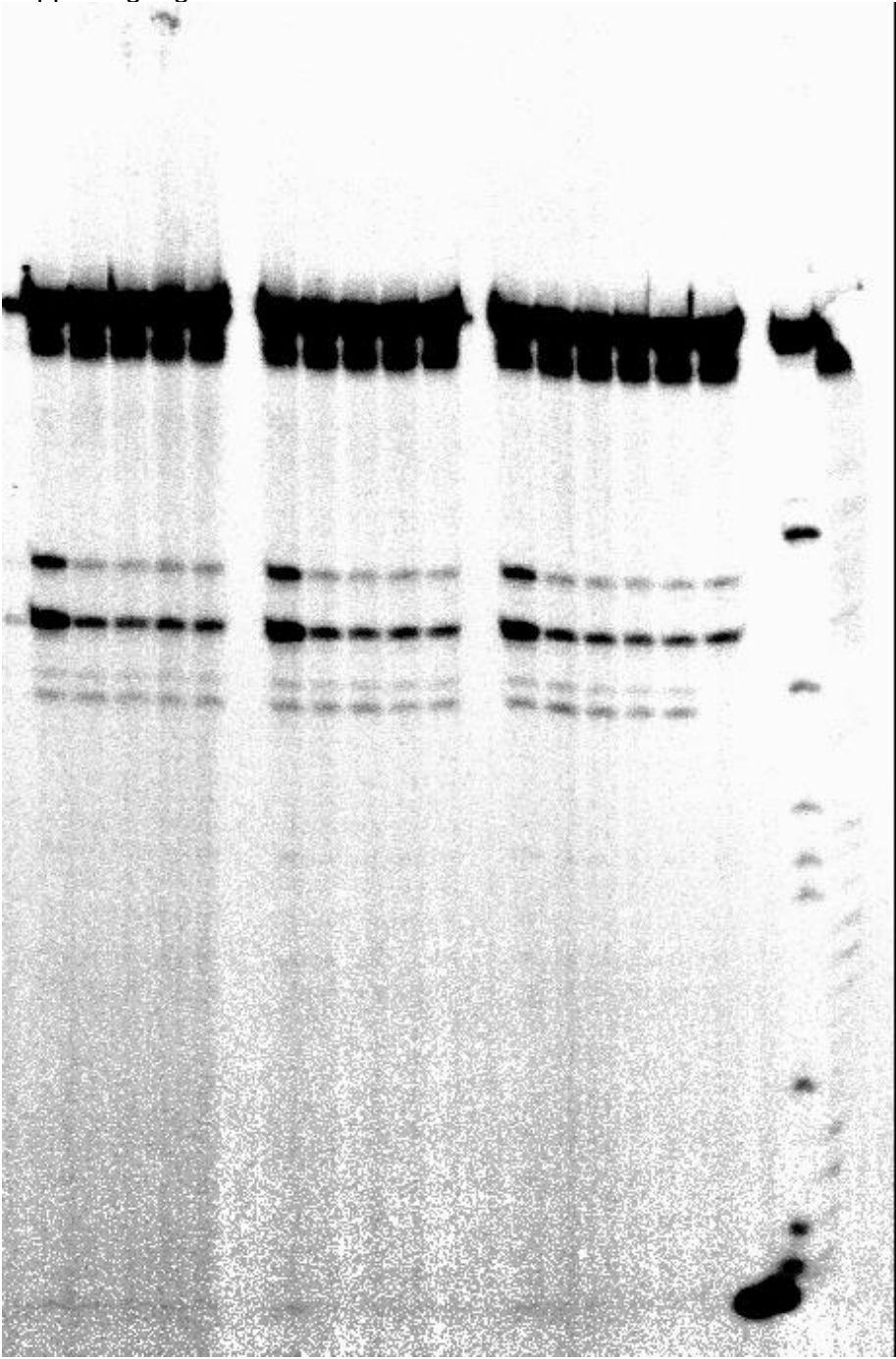
Supporting Figure 10 panel B and C

Supporting Figure 11 Panel D

100 KDa

Vehicle   Vehicle   LNA-377   LNA-Scramble   LNA-377   LNA-Scramble
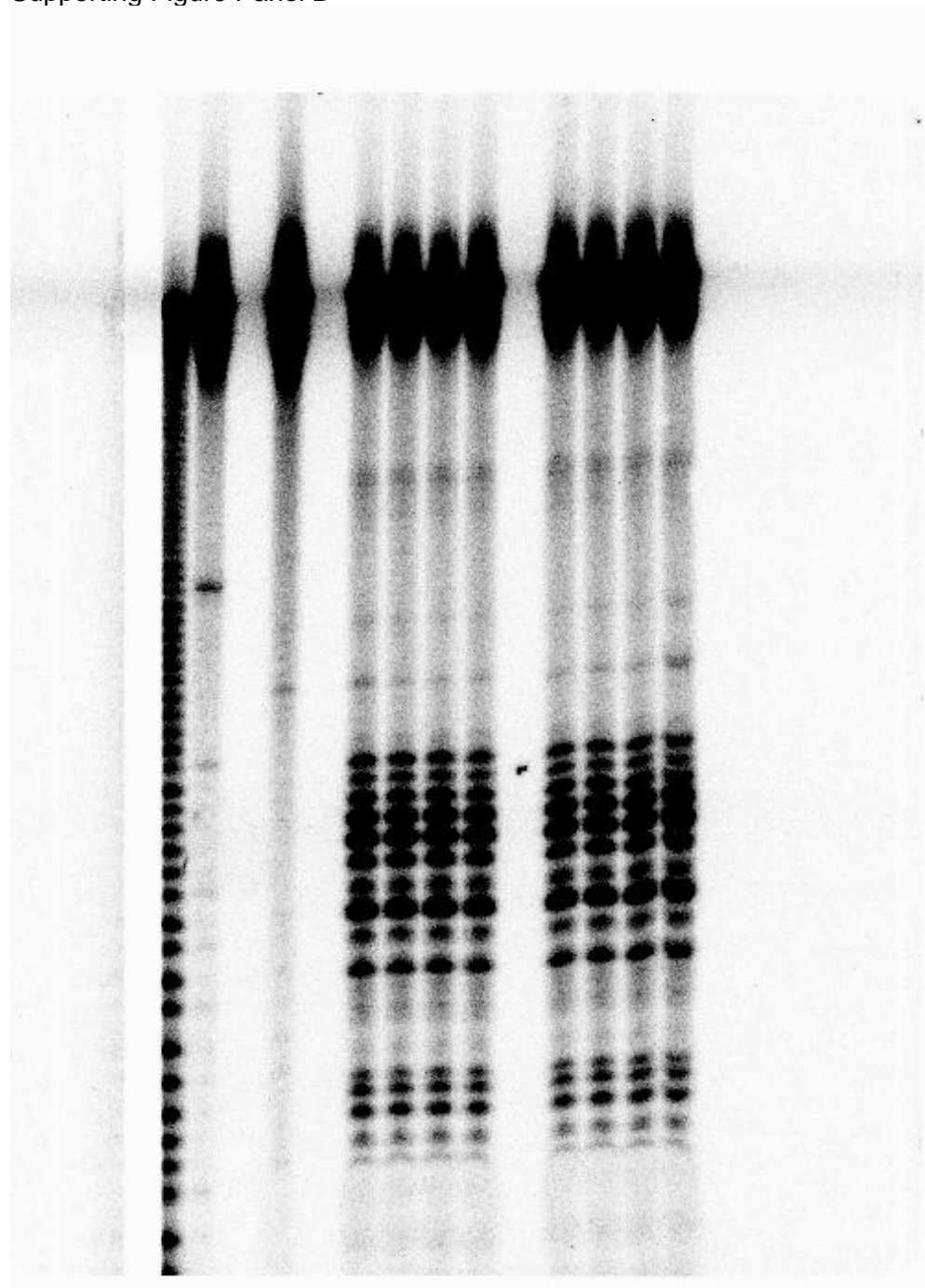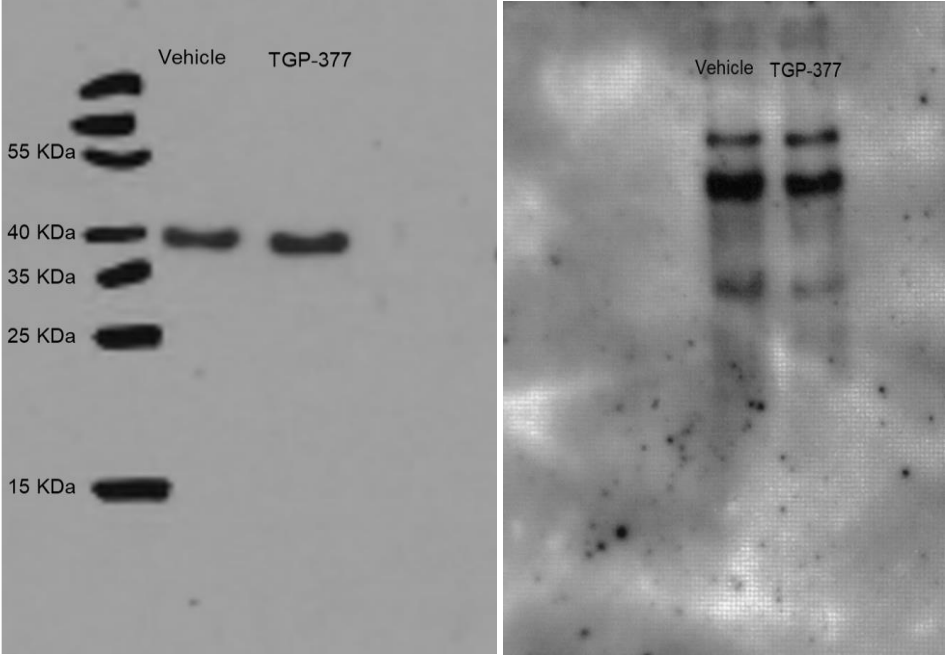
50 KDa

40 kDa

35 KDa

Supporting Figure 17 Panel A

Supporting Figure Panel B

Supporting Figure 18

REFERENCES

1   Wang, Z.-F. *et al.* The hairpin form of r(G4C2)exp in c9ALS/FTD is repeat-associated non-ATG translated and a target for bioactive small molecules. *Cell Chem. Biol.* **26**, 179-190 (2019).

2   Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097-6100 (1990).

3   Nettling, M. *et al.* DiffLogo: a comparative visualization of sequence motifs. *BMC Bioinformatics* **16**, 387-395 (2015).

4   Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789-796 (2009).

5   Stelzer, A. C., Kratz, J. D., Zhang, Q. & Al-Hashimi, H. M. RNA dynamics by design: biasing ensembles towards the ligand-bound state. *Angew. Chem. Int. Ed. Engl.* **49**, 5731-5733 (2010).

6   Stelzer, A. C. *et al.* Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **7**, 553-559 (2011).

7   Chen, J. L., VanEtten, D. M., Fountain, M. A., Yildirim, I. & Disney, M. D. Structure and dynamics of RNA repeat expansions that cause Huntington's disease and myotonic dystrophy type 1. *Biochemistry* **56**, 3463-3474 (2017).

8   Jain, A. K., Reddy, V. V., Paul, A., K, M. & Bhattacharya, S. Synthesis and evaluation of a novel class of G-quadruplex-stabilizing small molecules based on the 1,3-phenylene-*bis*(piperazinyl benzimidazole) system. *Biochemistry* **48**, 10693-10704 (2009).

9   Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

10   D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S.

LeGrand, P. Li, C., Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I., Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, & R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman AMBER 2016. *University of California, San Francisco* (2016).

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

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

13   Wales, D. J. & Yildirim, I. 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 (2017).

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

15   Frisch, M. *et al.* Gaussian 09; Gaussian, Inc: Wallingford, CT, 2009. (2005).

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

17   Childs-Disney, J. L. *et al.* Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat. Commun.* **4**, 2044-2055 (2013).

18   Childs-Disney, J. L. *et al.* Structure of the myotonic dystrophy type 2 RNA and designed small molecules that reduce toxicity. *ACS Chem. Biol.* **9**, 538-550 (2014).

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

20   Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623-1641 (2002).

21   Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247-260 (2006).

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

23   Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084-3095 (2013).

24   Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38 (1996).