

Supporting Information

Small molecule-based pattern recognition to classify RNA structure

Christopher S. Eubanks, Jordan E. Forte, Gary J. Kapral and Amanda E. Hargrove

Department of Chemistry, Duke University, Durham, NC 27708, United
States.

Corresponding Author

* E-mail: amanda.hargrove@duke.edu; Tel.: 919-660-1522

Contents

S1. Materials and Synthetic Methods	S3
S2. RNA Training Set Design	S14
S3. RNA Solid Phase Synthesis and Purification	S22
S4. Plate Reader Assay: Method and Data	S25
S5. Training Set Principal Component Analysis and Loading Plots	S36
S6. TAR Structural Classification	S39
S7. Cheminformatic Analysis of Aminoglycoside Receptors	S43
S8. NMR Spectra and HPLC Chromatograms	S45
S9. References	S68

S1. Materials and Synthetic Methods

General Information

All reactions were performed under a nitrogen atmosphere passed through drierite absorbents (Fisher Scientific), unless otherwise indicated. Reagents and anhydrous solvents (excluding CH_2Cl_2) were purchased from Glen Research, Sigma-Aldrich, Acros, Fisher, ChemImpex, and Oakwood Chemicals and were used as received without further purification. Anhydrous CH_2Cl_2 was obtained using a Pure Solv (Innovative Technology) solvent purification system at Duke University. Deuterated chloroform (CDCl_3), methanol (CD_3OD), dichloromethane (CD_2Cl_2), and dimethyl sulfoxide ($[\text{d}_6]\text{DMSO}$) for all NMR experiments were purchased from Cambridge Isotope Laboratories, and the former was deacidified using potassium carbonate prior to use. Deionized water was obtained from an ELGA PURELAB Flex (Veolia Water Technologies) water purification system. Diethylpyrocarbonate (DEPC) treated water was used for RNase free solutions and was produced with water was incubated with DEPC (1% v/v) overnight, followed by autoclaving at 120°C for 30 minutes. All microwave syntheses were performed using a Biotage Initiator+. Synthesized products were purified using flash column chromatography using 230-400 mesh silica gel (Silicycle). ^1H -NMR, ^{13}C -NMR, and ^{31}P -NMR spectra were obtained using a Varian Unity 500 MHz or a 400 MHz Varian Inova spectrometer at the Department of Chemistry at Duke University. Chemical shifts are reported in ppm in reference to the peaks of deuterated solvents (CDCl_3 , CD_3OD , CD_2Cl_2 , and $[\text{d}_6]\text{DMSO}$). Mass spectra were obtained using ESI-MS on an Agilent LC/MSD Trap. RNA oligomers were synthesized on a MerMade 6/12 Oligonucleotide Synthesizer (BioAutomation) and purified using a 3-5 micron polydivinylbenzene 4,4'-Dimethoxytrityl affinity column. Plate reader assays were run on a SpectraMax I3 (Molecular Devices) and used Corning 4514 384 well plates.

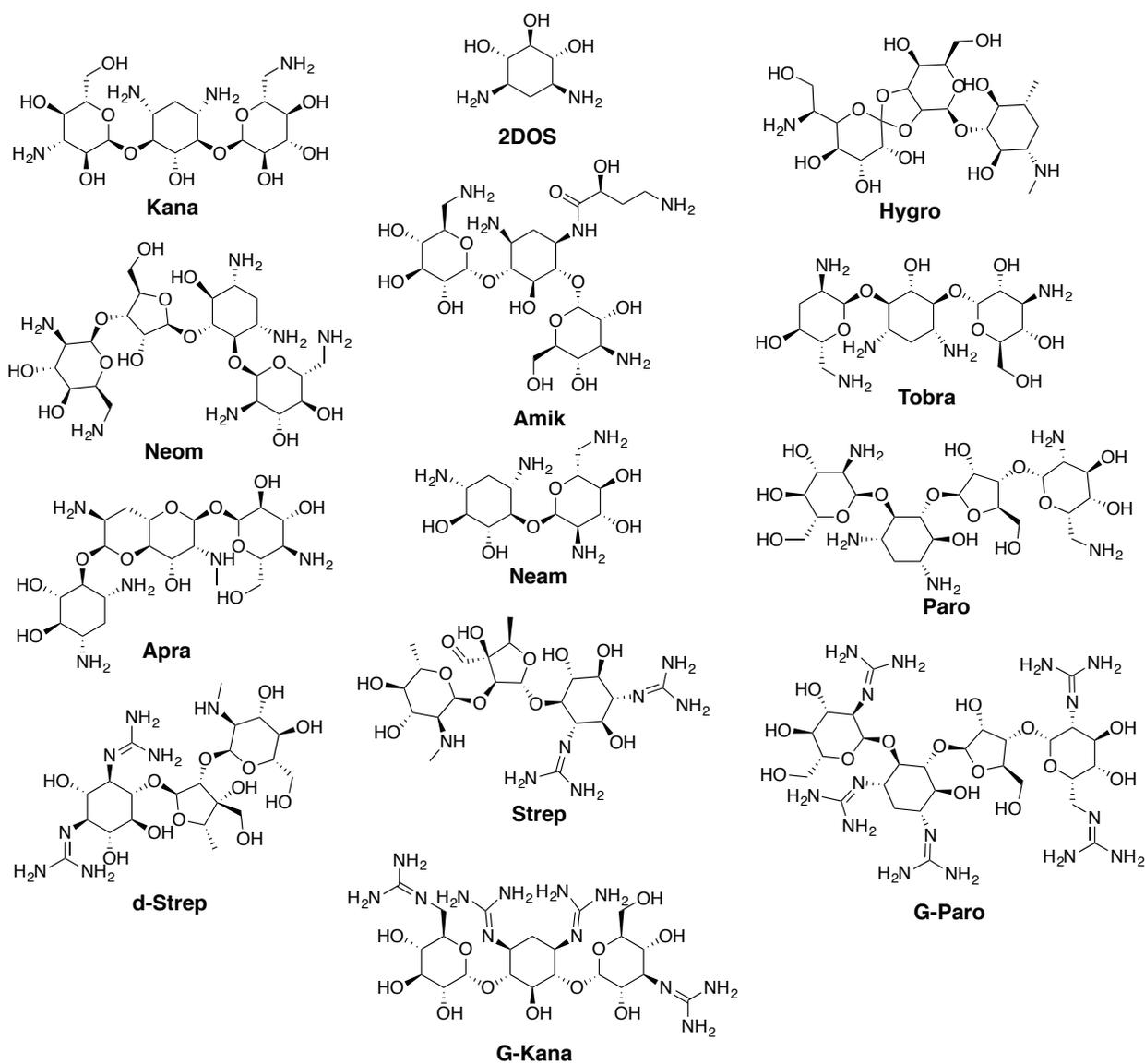
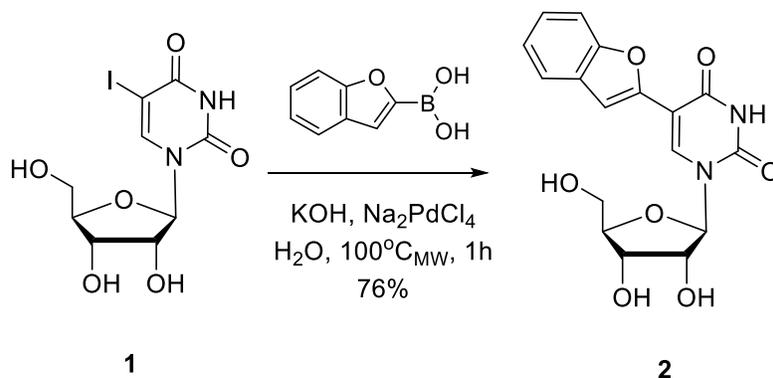


Fig S1-1. The aminoglycoside receptor library. Hygromycin B (**Hygro**), tobramycin (**Tobra**), and paromomycin (**Paro**) were removed from the library to improve the predictive power of the principal component analysis.

Synthesis of protected benzofuranyluridine derivative for solid phase synthesis

Synthesis of 5-(2-benzofuranyl)-uridine (**2**)

Adapted from Gallagher-Duval et al.¹

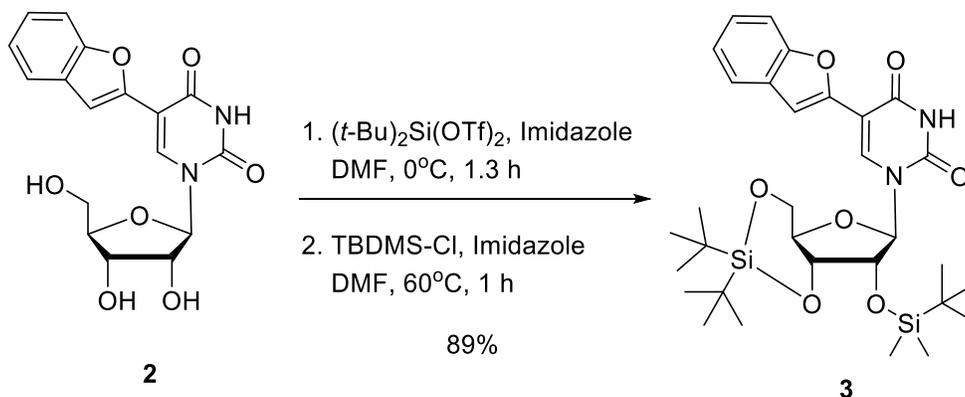


In a microwave vial, 5-iodouridine (**1**, 260.9 mg, 0.70 mmol), 2-benzofuranylboronic acid (150.8 mg, 0.93 mmol, 1.3 equiv.), and potassium hydroxide (81.4 mg, 1.45 mmol, 2.0 equiv.) were dissolved in 4.0 mL degassed water bubbled with argon. A solution of Na₂PdCl₄ in degassed water (1.0 mL, 0.007 mmol, 0.1 mol%) was added (0.14 M total). The solution was stirred under microwave irradiation at 100°C, high absorption, for 1 hour. The solid product was cooled, collected by filtration, and rinsed with water, ethyl ether (1 mL), then hexanes (5 mL x 3). Crude product was dissolved in a 1:1 MeOH/acetone solution and cooled and concentrated under nitrogen to precipitate impurities. Solution was filtered, collected, concentrated *in vacuo*, dissolved in ~1 mL water and lyophilized to yield **2** as a white solid (190.4 mg, 0.53 mmol, 76%).

¹H NMR ([d₆]DMSO, 500 MHz, 30°C) δ 11.76 (s, 1H), 8.84 (s, 1H), 7.62 (dd, 1H, *J* = 8.0, 0.5 Hz), 7.55 (d, 1H, *J* = 8.0 Hz), 7.30-7.26 (m, 1H), 7.23-7.20 (m, 1H), 5.88 (d, 1H, *J* = 4.5 Hz), 5.46 (d, 1H, *J* = 5 Hz), 5.35 (t, 1H, *J* = 4.0 Hz), 5.11 (d, 1H, *J* = 5.5 Hz), 4.16 (q, 1H, *J* = 5.0 Hz), 4.09 (q, 1H, 5.5 Hz), 3.97-3.94 (m, 1H), 3.79 (dt, 1H, *J* = 11.0, 4.0 Hz), 3.67 (dt, 1H, *J* = 12.0, 3.5 Hz). ¹³C NMR ([d₆]DMSO, 126 MHz) δ 161.0, 153.7, 150.3, 149.7, 138.0, 129.5, 124.9, 123.7, 121.7, 111.5, 105.4, 104.6, 89.4, 85.3, 75.1, 70.1, 60.8. HRMS (ESI+) Calculated for C₁₇H₁₇N₂O₇ - 361.1030; Found - 361.1031 (± 1.2 ppm).

Synthesis of 5',3'-O-Bis(*t*-butylsilyl)-2'-O-(*t*-butyldimethylsilyl)-5-(2-benzofuranyl)-uridine (**3**)

Adapted from Ghanty et al.²

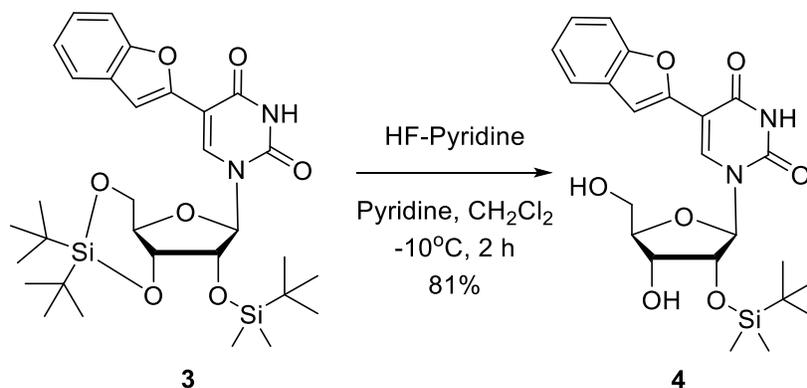


Compound (**2**) (319.0 mg, 0.88 mmol) was dissolved in anhydrous *N,N*-dimethylformamide (DMF, 4.4 mL, 0.2 M) with 3 Å molecular sieves stirring in an ice bath under nitrogen. Di-*tert*-butylsilyl-bis(trifluoromethanesulfonate) ((*t*-Bu)₂Si(OTf)₂) (490 μL, 1.5 mmol, 1.7 equiv.) was added drop-wise over 45 minutes and then allowed to react at room temperature for 15 minutes. The solution was diluted using 0.3 mL DMF, an additional (*t*-Bu)₂Si(OTf)₂ (50 μL, 0.15 mmol) was added, and the reaction proceeded for 20 minutes. The solution was quenched with imidazole (793.0 mg, 11.6 mmol, 13.0 equiv.) for 10 minutes while stirring. *Tert*-butyldimethylsilyl chloride (572.9 mg, 3.82 mmol, 4.3 equiv.) was added and a reflux condenser was attached to the reaction flask. Solution was heated to 60°C and reacted for 1 hour. Reaction was cooled on an ice bath and quenched with 5 mL water. Precipitate was collected by vacuum filtration and washed with a minimal amount of chilled (4°C) MeOH. Product in the MeOH filtrate was isolated via silica column flash chromatography (86:9:4:1 hexanes:EtOAc:MeOH:TEA). The remaining solid product was dissolved in EtOAc and concentrated *in vacuo*. Both pure products were combined to give **3** as a white solid (485.3 mg, 0.75 mmol, 89%).¹H NMR (CDCl₃, 500 MHz) δ 9.12 (s, 1H), 8.05 (s, 1H), 7.61 (d, 1H, *J* = 7.5 Hz), 7.52 (dd, 1H, *J* = 0.5 Hz), 7.43 (dd, 1H, *J* = 8.0, 1.0 Hz), 7.32 (td, 1H, *J* = 8.5, 1.5 Hz), 7.26 (td, 1H, *J* = 7.5, 1.0 Hz), 5.88 (s, 1H), 4.65 (dd, 1H, *J* = 9.5, 5.0 Hz), 4.38 (d, 1H, *J* = 4.5 Hz),

4.30 (td, 1H, $J = 9.5, 5.0$ Hz), 4.21 (dd, 1H, $J = 10.5, 9.0$ Hz), 4.09 (dd, 1H, $J = 10.0, 4.5$ Hz), 1.12 (s, 9H), 1.08 (s, 9H), 0.99 (s, 9H), 0.25 (s, 3H), 0.20 (s, 3H). ^{13}C NMR (CDCl_3 , 126 MHz) δ 160.1, 153.9, 148.9, 147.7, 135.0, 129.4, 125.0, 123.4, 121.7, 110.8, 106.7, 106.3, 93.9, 76.2, 75.8, 75.0, 68.0, 27.8, 27.2, 26.1, 23.2, 20.6, 18.5, -4.1, -4.8. HRMS (ESI+) Calculated for $\text{C}_{31}\text{H}_{47}\text{N}_2\text{O}_7\text{Si}_2$ - 615.2916; Found - 615.2920 (± 0.6 ppm).

Synthesis of 2'-O-(t-butylidimethylsilyl)-5-(2-benzofuranyl)-uridine (**4**)

Adapted from Ghanty et al.²

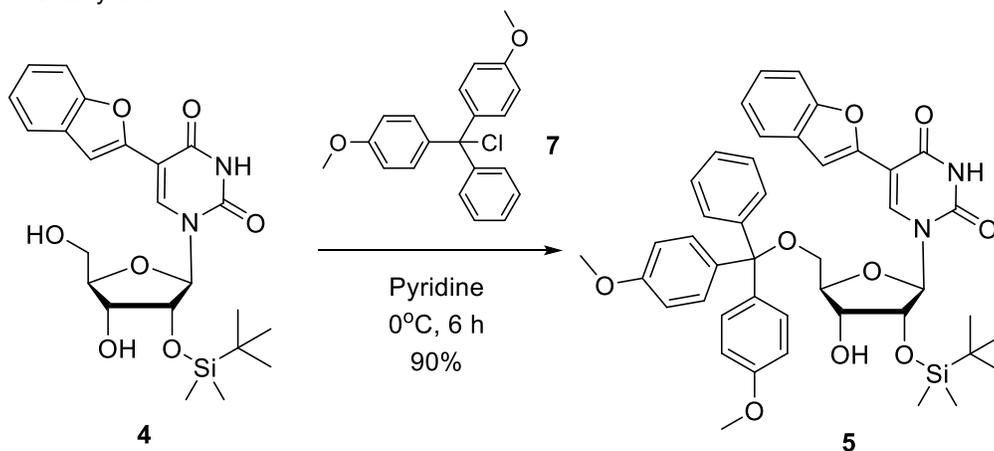


Anhydrous (**3**) (171.0 mg, 0.28 mmol) was dissolved in anhydrous CH_2Cl_2 (1.4 mL, 0.2 M) in a polypropylene test tube while stirring under nitrogen in an ice/salt bath. A separate solution was prepared by diluting HF-pyridine (36.1 μL , 1.39 mmol, 5.0 equiv.) with anhydrous pyridine (224 μL) at 0°C in a separate polypropylene test tube. The HF-pyridine solution was slowly added to the former solution and reacted for 2 hours. Solution was then diluted with 1.4 mL CH_2Cl_2 , quenched with a 2.8 mL of a saturated, aqueous Na_2CO_3 solution, and then allowed to warm to room temperature. The solution was washed with a saturated aqueous Na_2CO_3 solution (5 mL), and then the CH_2Cl_2 layer was washed with NaHCO_3 (5 mL x 2), and again with a brine solution (5 mL x 2). Organic layer was then dried with anhydrous Na_2SO_4 , filtered, and dried *in vacuo*. Crude product was purified with silica column flash chromatography (73:24:3 hexanes:EtOAc:MeOH) to yield **4** as a white solid (106.6 mg, 0.22 mmol, 81%). ^1H NMR

(CD₃OD, 400 MHz) δ 8.88 (s, 1H), 7.52 (d, 1H, J = 7.2 Hz), 7.48 (d, 1H, J = 8.0 Hz), 7.32 (s, 1H), 7.23 (td, 1H, J = 8.0, 1.2 Hz), 7.16 (t, 1H, J = 7.6 Hz), 5.78 (d, 1H, J = 4.0 Hz), 4.39 (t, 1H, J = 4.4 Hz), 4.19 (t, 1H, J = 5.2 Hz), 4.11-4.07 (m, 1H), 3.98 (dd, 1H, J = 12.4, 2.4 Hz), 3.83 (dd, 1H, J = 12.4, 2.4 Hz), 0.90 (s, 9H), 0.11 (s, 6H). ¹³C NMR (CD₂Cl₂, 126 MHz) δ 160.1, 154.0, 149.6, 148.1, 137.8, 129.3, 124.9, 123.3, 121.5, 110.8, 106.8, 105.8, 92.7, 85.7, 75.1, 71.0, 62.2, 25.6, 18.1, -4.8, -5.1. HRMS (ESI+) Calculated for C₂₃H₃₁N₂O₇Si - 475.1895; Found - 475.1896 (\pm 1.6 ppm).

Synthesis of 5'-O-(4,4'-Dimethoxytrityl)-2'-O-(*t*-butyldimethylsilyl)-5-(2-benzofuranyl)-uridine (**5**)

Adapted from Ghanty et al.²

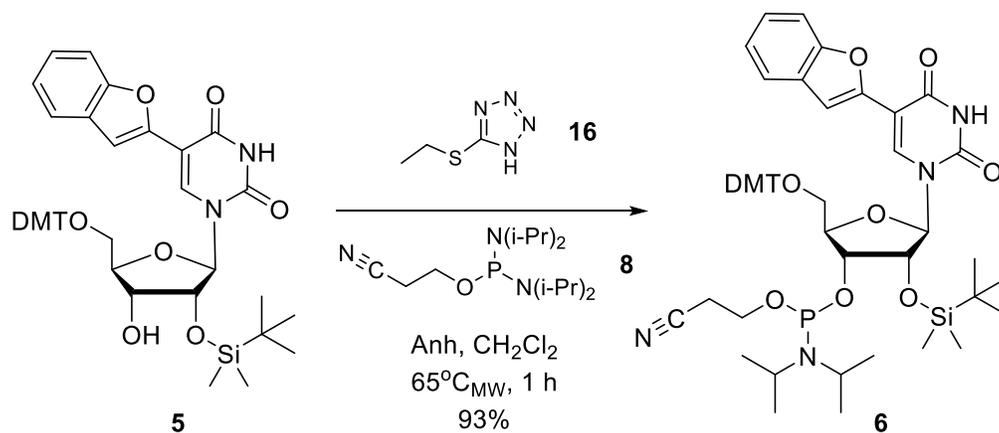


To a solution of dried (**4**) (117.1 mg, 0.25 mmol) in anhydrous pyridine (2.5 mL, 0.1 M) stirring in an ice/salt bath under nitrogen, 4,4'-dimethoxytrityl chloride (**7**) (430.4 mg, 1.27 mmol, 5.0 equiv.) was added. The solution was then allowed to warm to room temperature and stirred for 6 hours. Solution was quenched with 2 mL MeOH for 5 minutes and dried *in vacuo*. The solid was dissolved in EtOAc and washed with saturated aqueous NaHCO₃ (5 mL x 2) and then brine (5 mL x 2). The organic layer was dried using anhydrous Na₂SO₄, filtered, and dried *in vacuo*. Crude product was purified with silica column flash chromatography (50:49:1 hexanes:EtOAc:TEA) and dried *in vacuo* to yield **5** as a yellow foam (180.2 mg, 0.23 mmol, 90%). ¹H NMR (CDCl₃, 500 MHz) δ 9.60 (br s, 1H), 8.56 (s, 1H), 7.57 (d, 2H, J = 7.5 Hz), 7.49

(s, 1H), 7.47-7.41 (m, 5H), 7.21 (t, 2H, $J = 8.0$ Hz), 7.10 (t, 1H, $J = 7.0$ Hz), 7.05 (t, 1H, $J = 7.5$ Hz), 6.84 (t, 1H, $J = 7.0$ Hz), 6.74-6.70 (m, 4H), 6.22 (d, 1H, $J = 5.5$ Hz), 5.84 (d, 1H, 8.0 Hz), 4.60 (t, 1H, $J = 5.5$ Hz), 4.24-4.20 (m, 1H), 4.17-4.14 (m, 1H), 3.73 (dd, 1H, $J = 10.5, 2.0$ Hz), 3.63 (s, 3H), 3.61 (s, 3H), 3.27 (dd, 1H, $J = 11.0, 3.0$ Hz), 2.73 (br s, 1H), 0.92 (s, 9H), 0.15 (s, 3H), 0.13 (s, 3H). ^{13}C NMR (CDCl_3 , 126 MHz) δ 160.1, 158.5, 158.5, 153.4, 149.5, 147.1, 144.5, 135.7, 135.4, 134.7, 130.1, 130.0, 128.7, 128.1, 127.8, 126.9, 124.0, 122.5, 120.7, 113.2, 113.1, 110.8, 107.3, 106.1, 87.7, 86.9, 84.0, 75.9, 71.1, 63.2, 55.1, 25.6, 17.9, -4.7, -5.1. HRMS (ESI+) $\text{C}_{44}\text{H}_{48}\text{N}_2\text{O}_9\text{SiNa}$ - 799.3021; Found - 799.3015 (± 0.8 ppm).

Synthesis of 5'-O-(4,4'-Dimethoxytrityl)-3'-O-[(2-cyanoethoxy)(N,N-diisopropylamino)phosphino]-2'-O-(t-butyldimethylsilyl)-5-(2-benzofuranyl)-uridine (6**)**

Adapted from the Krishnamurthy et al.^{3,4}



To an oven-dried microwave vial purged with argon, anhydrous (**5**) (150.0 mg, 0.193 mmol) and 5-ethylthio-1H-tetrazole (**16**) (35.4 mg, 0.27 mmol, 1.4 equiv.) were added sequentially and maintained under argon. The mixture was dissolved using anhydrous CH_2Cl_2 (1.2 mL, 0.16 M). Reagent (**8**) (80 μL , 0.25 mmol, 1.3 equiv.) was then added dropwise via syringe. The solution was stirred under microwave irradiation at 65°C , low absorption, 30 sec prestir, for 1 hour. Solution was concentrated to dryness, and purified by silica column flash chromatography (gradient from 10% EtOAc in Hexanes + 3% TEA to 80% EtOAc in Hexanes + 3% TEA, loaded using ~ 2 mL CH_2Cl_2) and dried *in vacuo* to yield **6** as a white solid (176.4 mg,

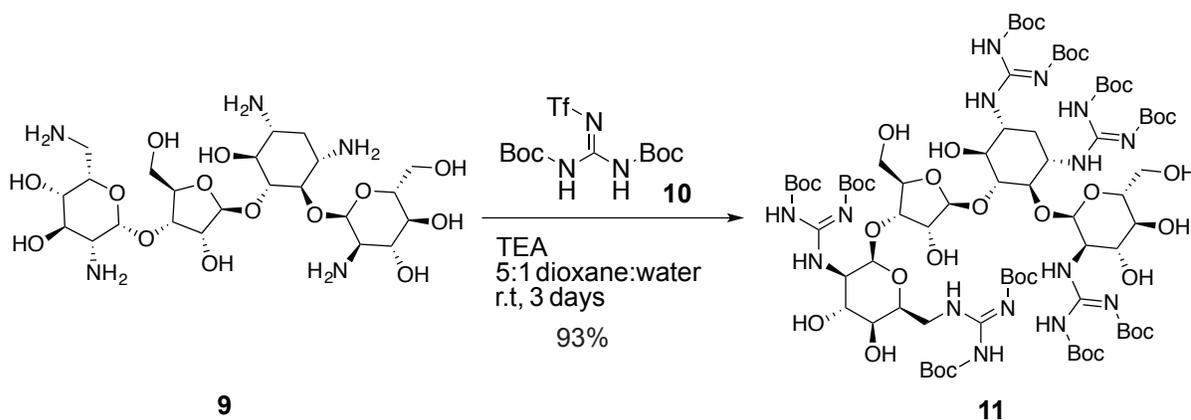
0.181 mmol, 93%). Both major products observed by TLC were isolated and determined to be diastereomers based on identical mass spectral characterization and ^{13}P -NMR chemical shifts characteristic of phosphoramidites (150.95 and 148.84 ppm). An H-phosphonate impurity of 2.6% of the total yield was observed (8.01 and 7.57 ppm) but is not expected to interfere with the solid phase synthesis because it is unable to react with the 5' hydroxyl group of the preceding nucleoside. Diastereomer 1: ^1H NMR (CDCl_3 , 500 MHz): δ 8.87 (br s, 1H), 8.63 (s, 1H), 7.60 (d, 2H, $J = 7.5$ Hz), 7.48-7.43 (m, 6H), 7.23 (t, 2H, $J = 8.0$ Hz), 7.12 (t, 1H, $J = 7.0$ Hz), 7.03 (t, 1H, $J = 7.5$ Hz), 6.78 (t, 1H, $J = 7.5$ Hz), 6.73 (d, 4H, $J = 8.5$ Hz), 6.18 (d, 1H, $J = 6.5$ Hz), 5.64 (d, 1H, $J = 8.5$ Hz), 4.58 (t, 1H, $J = 5.5$ Hz), 4.17-4.12 (m, 1H), 3.80 (d, 1H, $J = 10.5$ Hz), 3.63 (s, 3H), 3.62 (s, 3H), 3.60-3.50 (m, 4H), 3.18 (dd, 1H, $J = 11.0, 2.5$ Hz), 2.33-2.21 (m, 2H), 1.18 (d, 6H, $J = 6.5$ Hz), 1.14 (d, 6H, $J = 6.5$ Hz), 0.88 (s, 9H), 0.08 (s, 6H). ^{13}C NMR (CDCl_3 , 126 MHz) δ 160.0, 158.5, 153.4, 149.3, 147.3, 144.6, 135.8, 135.4, 135.0, 130.19 (peak overlap), 130.17, 128.7, 128.3, 127.9, 126.9, 124.0, 122.5, 120.6, 117.2, 113.2, 113.1, 110.8, 107.1, 105.8, 87.5, 86.9, 84.1, 75.09 (peak overlap), 75.06, 73.0, 72.9, 62.8, 57.5, 57.4, 55.2, 43.4, 43.3, 25.7, 25.6, 24.74, 24.67, 24.61 (peak overlap), 20.11, 20.05, 18.0, -4.7, -4.9. Diastereomer 2: ^1H NMR (CDCl_3 , 400 MHz): δ 8.80 (br s, 1H), 8.56 (s, 1H), 7.55 (d, 2H, $J = 8.4$ Hz), 7.44-7.39 (m, 6H), 7.18 (t, 2H, $J = 7.6$ Hz), 7.10-7.05 (m, 1H), 7.03-6.99 (m, 1H), 6.79-6.74 (m, 1H), 6.71-6.66 (m, 4H), 6.27 (d, 1H, $J = 7.2$ Hz), 5.64 (d, 1H, $J = 8.4$ Hz), 4.62 (dd, 1H, $J = 7.2, 4.4$ Hz), 4.22 (s, 1H), 4.10 (dd, 1H, $J = 12.8, 4.4$ Hz), 4.01-3.93 (m, 1H), 3.92-3.84 (m, 1H), 3.71 (dd, 1H, $J = 11.2, 1.6$ Hz), 3.60 (s, 3H), 3.57 (s, 3H), 3.55-3.48 (m, 2H), 3.15 (dd, 1H, $J = 10.4, 2.4$ Hz), 2.74-2.61 (m, 2H), 1.13 (d, 6H, $J = 6.8$ Hz), 0.90 (d, 6H, $J = 6.8$ Hz), 0.87 (s, 9H), 0.07 (d, 6H, $J = 4.8$ Hz). ^{13}C NMR (CDCl_3 , 126 MHz) δ 160.0, 158.6, 158.5, 153.4, 149.7, 147.3, 144.5, 135.6, 135.2, 134.8, 130.13, 130.08, 128.7, 128.2, 128.0, 127.0, 124.0, 122.6, 120.7, 117.9, 113.31, 113.26, 110.9, 107.4, 105.9, 87.2, 86.6, 84.0, 77.6, 75.8, 72.2, 72.1, 63.2, 59.2, 59.1, 55.1, 42.9, 42.8, 29.7, 25.8, 24.65, 24.58, 24.5 (peak overlap), 20.6, 20.5, 18.1, 1.0, -4.50, -4.53, -4.8. Combined Diastereomers: ^{31}P NMR (CDCl_3 , 162 MHz) δ 150.95, 148.84;

diastereomer 1 corresponds to 150.95 while diastereomer 2 corresponds to 148.84. HRMS (ESI+) Calculated for $C_{53}H_{66}N_4O_{10}PSi$ 977.4280; Found 977.4299 (± 1.9 ppm).

Synthesis of guanidinylated paromomycin and kanamycin

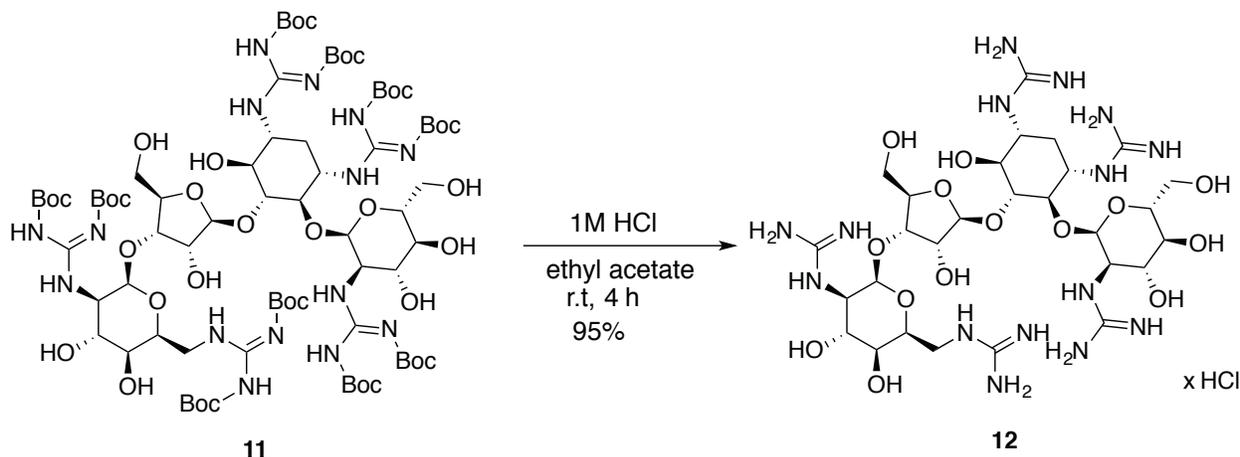
The following reactions were performed based on literature procedures.⁵

Boc-guanidinylated Paromomycin (9)



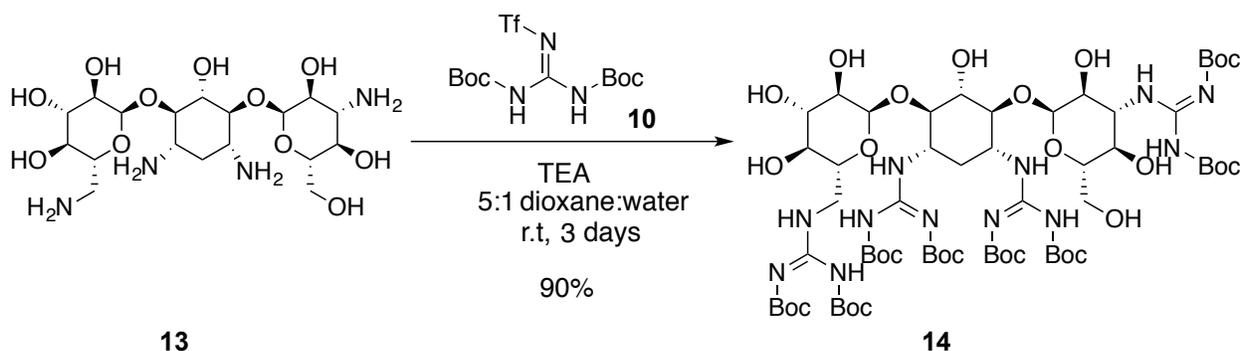
To an oven-dried 5 mL round bottom flask, paromomycin (5 amines, 12 mg, 0.032 mmol, 1 equiv) was dissolved in H_2O (0.27 mL, 0.02 M) and 1,4-dioxane (1.35 mL, 0.02 M). While stirring, N,N'-di-Boc-N''-triflylguanidine (10) (125 mg, 0.32 mmol, 10 equiv.) was added. After five minutes of stirring, triethylamine (0.05 mL, 0.38 mmol, 12 equiv.) was added slowly and the reaction was stirred for three days. After 3 days, the mixture was extracted with CH_2Cl_2 (10 mL x 3) and washed with brine (10 mL x 3). The organic layer was dried using anhydrous Na_2SO_4 , filtered, and dried *in vacuo*. The resulting residue was purified by silica column flash chromatography (90:10 CH_2Cl_2 , MeOH) and dried *in vacuo* to yield 11 as a white solid (54 mg, 0.0295 mmol, 93%). 1H NMR ($CDCl_3$, 400 MHz): δ 1.49 (s, 90H). HRMS (ESI+) Calculated for $C_{78}H_{137}N_{15}O_{34}$ [M+2H] - 913.96, Found - 913.9725 (± 0.5 ppm).

Deprotection of Boc-Guanidinoparomomycin (12)



In an oven-dried 5 mL round bottom flask, **11** (40 mg, 0.021 mmol, 1 equiv.) was dissolved ethyl acetate (0.548 mL, 0.04M) and HCl (0.047 mL, 1 M). The solution was stirred at room temperature. After 4 hours, the solution was diluted with toluene (3 mL), concentrated *in vacuo*, and then diluted with water (3 mL). Subsequent lyophilization of the water provided **12** as a white solid powder (16.5 mg, 0.0199 mmol, 95%). HRMS (ESI+) Calculated for $\text{C}_{28}\text{H}_{55}\text{N}_{15}\text{O}_{14}$ [M+H] – 825.4, Found – 825.4101 (± 1.2 ppm). LCMS showed a single peak at 11.9 min and had a single mass at 864.2 (M+K).

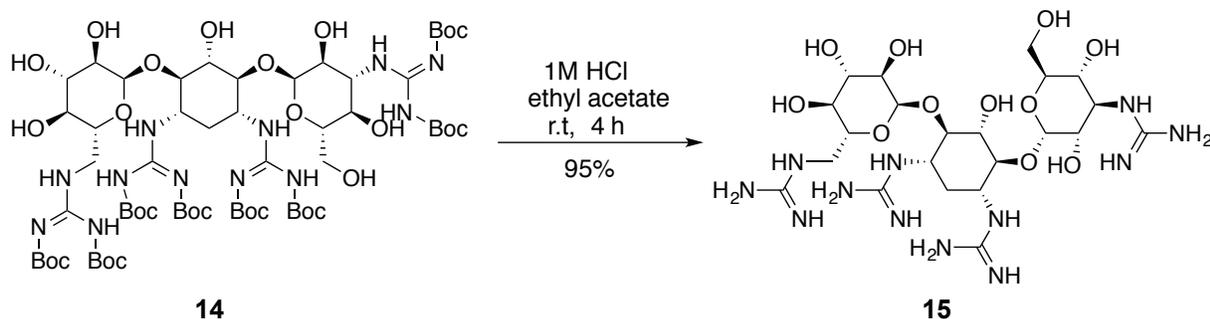
Boc-guanidinylation of Kanamycin (**14**)



To an oven-dried 5 mL round bottom flask, kanamycin (4 amines, 20 mg, 0.041 mmol, 1 equiv.) was dissolved in H_2O (0.35 mL, 0.02M) and 1,4-dioxane (1.71 mL, 0.02 M). While stirring, N,N'-di-Boc-N''-triflylguanidine (**10**) (129 mg, 0.33 mmol, 8 equiv.) was added. After five

minutes of stirring, triethylamine (0.065 mL, 0.41 mmol, 10 equiv.) was added slowly and the reaction was stirred for three days. After 3 days, the mixture was extracted with CHCl_2 (10 mL x 3) and washed with brine (10 mL x 3). The organic layer was dried using anhydrous Na_2SO_4 , filtered, and dried *in vacuo*. Resulting residues was purified by silica column flash chromatography (90:10 CH_2Cl_2 , MeOH) and dried *in vacuo* to yield **14** as a white solid (53 mg, 0.037 mmol, 90%). ^1H NMR (CDCl_3 , 400 MHz): δ 1.51 (s, 72H). HRMS (ESI+) Calculated for $\text{C}_{62}\text{H}_{108}\text{N}_{12}\text{O}_{27}$ [M+2H] – 727.3, Found – 727.3806 (\pm 1.3 ppm).

Deprotection of Boc-Guanidinokanamycin (**15**)



In an oven-dried 5 mL round bottom flask, **14** (50 mg, 0.034 mmol, 1 equiv.) was dissolved in ethyl acetate (0.86 mL, 0.04M) and HCl (0.074 mL, 1 M). The solution was stirred at room temperature. After 4 hours, the solution was diluted with toluene (3.2 mL), concentrated *in vacuo*, and then diluted with water (4 mL). Subsequent lyophilization of the water provided **15** a white solid powder (21.1 mg, 0.032 mmol, 95%). HRMS (ESI+) Calculated for $\text{C}_{22}\text{H}_{43}\text{N}_{12}\text{O}_{11}$ [M+H] – 651.3, Found – 651.3322 (\pm 0.7 ppm). LCMS showed a single peak at 11.7 min and had a major mass of 650.9 (M+H).

S2. RNA Training Set Design

RNAstructure⁶ was employed to determine the most stable structures of RNA sequences with 95% predicted probable structural motif formation. Each sequence contained a constant stem and hairpin sequence, except for the hairpin training set motifs, which had a constant stem sequence. Variable sequences and number of nucleotides were inserted at the secondary structures of interest to increase the diversity of the RNA training set. SI Table 1 lists the 60 sequences that were determined by RNAstructure, as well as the GC content and sequence length. As a secondary analysis, MC-FOLD⁷ was used to check the RNA sequences, with the correct structure having ~ -4 kcal/mol Gibbs free energy compared to any other structures found. The RNA constructs were further assessed using the FARFAR de novo RNA protocol⁸ to determine an ensemble of 20 structures per sequence and identify the most flexible sites for BFU nucleoside insertion. This procedure involved taking each RNA sequence and subjecting it to the FARFAR algorithm available on the Rosie webserver;⁹ this algorithm uses fragment-based assembly to build the 3D RNA structure stepwise, with the goal of finding a diverse set of structures with low energies that could represent the dynamic ensemble of 3D RNA structures found in solution. We ran the base FARFAR algorithm with the following conditions: 1) vary bond lengths and angles; 2) optimize the RNA after fragment assembly; 3) use a bulge-favorable entropic score term; and 4) use the latest (2012) force-field. Monte Carlo simulations were run for 10,000 cycles, producing 1000 structures, which were then clustered based on RMSD to the lowest energy structure. To ensure diversity in 3D structure, we selected the lowest energy structure and the lowest energy exemplars of the next 19 lowest energy clusters. Studying these structures in the KiNG molecular viewer,¹⁰ we identified nucleotides in the secondary structure motif region for each of the sixteen RNA sequences that were chosen as the RNA training set. Nucleotides that were part of the bulge or the loop were compared across each of the 20 RNA conformations to assess their predicted dynamics. Uracil had highest priority if it was present. If the nucleotide was flipped out of the stack in more than 5 of

the 20 RNA structures (**Figure S2-6**), we considered it dynamic enough to be a candidate for replacement with BFU. This process was repeated for each of the 16 training set sequences.

Table S2-1. RNA sequences of RNA training set, N represents a variable nucleotide (nt). A number of nucleotides were variable in each structure motif: Bulge (2-4 nt), Internal loop (3x3 nt), Asymmetrical Internal Loop (3x 1-2 nt), Hairpin (4-6 nt), and Stem (6 nt).

Secondary Motifs	RNA Sequences
Bulge	GGACAC NNN CAGAGUACCUCUGGUGUCC
Internal Loop	GGACAC NNN CAGAGUACCUCUG NNN GUGUCC
Asymmetric Internal Loop	GGACAC NNN CAGAGUACCUCUG NN GUGUCC
Hairpin	GGACACUGGACAC NNNN GUGUCCAGUGUCC
Stem	GGACAC NNNNNN CAGAGUACCUCUG NNNNNN GUGUCC

Table S2-2: RNA Training Set Sequences. The RNA synthesized have their corresponding code in parentheses.

RNA ID	Sequence	Sequence Length	GC Content (%)
IL-001 (IL A)	GUCUGGACAC AUG CAGAGUACCUCUG AGA GUGUCCAGAC	39	56.4
IL-002	GUCUGGACAC UAC CAGAGUACCUCUG CAC GUGUCCAGAC	39	56.4
IL-003	GUCUGGACAC GAA CAGAGUACCUCUG CGA GUGUCCAGAC	39	56.4
IL-004 (IL B)	GUCUGGACAC CCC CAGAGUACCUCUG ACA GUGUCCAGAC	39	69.7
IL-005	GUCUGGACAC CUG CAGAGUACCUCUG ACU GUGUCCAGAC	39	56.4
IL-006	GUCUGGACAC GAU CAGAGUACCUCUG CAA GUGUCCAGAC	39	53.8
IL-007	GUCUGGACAC UCC CAGAGUACCUCUG UAU GUGUCCAGAC	39	53.8
IL-008	GUCUGGACAC AGA CAGAGUACCUCUG GAG GUGUCCAGAC	39	56.4
IL-009 (IL C)	GUCUGGACAC AGU CAGAGUACCUCUG UAA GUGUCCAGAC	39	51.2
IL-010	GUCUGGACAC AUC CAGAGUACCUCUG CCU GUGUCCAGAC	39	56.4
IL-011	GUCUGGACAC GAC CAGAGUACCUCUG CGU GUGUCCAGAC	39	58.9
IL-012	GUCUGGACAC CAA CAGAGUACCUCUG CGA GUGUCCAGAC	39	56.4
AIL-001 (AIL A)	GUCUGGACAC AGA CAGAGUACCUCUG A GUGUCCAGAC	37	54.1
AIL-002 (AIL B)	GUCUGGACAC AUA CAGAGUACCUCUG GC GUGUCCAGAC	38	55.3
AIL-003 (AIL C)	GUCUGGACAC AGU CAGAGUACCUCUG UA GUGUCCAGAC	38	52.6
AIL-004	GUCUGGACAC CCC CAGAGUACCUCUG AC GUGUCCAGAC	38	60.5
AIL-005	GUCUGGACAC CUG CAGAGUACCUCUG CU GUGUCCAGAC	38	57.9
AIL-006	GUCUGGACAC GAU CAGAGUACCUCUG A GUGUCCAGAC	37	54.1
AIL-007	GUCUGGACAC UCC CAGAGUACCUCUG U GUGUCCAGAC	37	56.8
AIL-008	GUCUGGACAC AGA CAGAGUACCUCUG GG GUGUCCAGAC	38	57.9
AIL-009	GUCUGGACAC AGU CAGAGUACCUCUG U GUGUCCAGAC	39	51.3

AIL-010	GUCUGGACAC AUC CAGAGUACCUCUG CU GUGUCCAGAC	38	55.3
AIL-011	GUCUGGACAC GAC CAGAGUACCUCUG CU GUGUCCAGAC	38	57.9
AIL-012	GUCUGGACAC CAA CAGAGUACCUCUG C GUGUCCAGAC	37	56.8
BG-001 (Bulge A)	GUCUGGACAC UC CAGAGUACCUCUG GUGUCCAGAC	36	56.7
BG-002	GUCUGGACAC GA CAGAGUACCUCUG GUGUCCAGAC	35	58.8
BG-003	GUCUGGACAC AG CAGAGUACCUCUG GUGUCCAGAC	35	58.8
BG-004 (Bulge B)	GUCUGGACAC GCU CAGAGUACCUCUG GUGUCCAGAC	36	58.3
BG-005	GUCUGGACAC UCG CAGAGUACCUCUG GUGUCCAGAC	36	58.3
BG-006	GUCUGGACAC ACG CAGAGUACCUCUG GUGUCCAGAC	36	58.3
BG-007	GUCUGGACAC ACU CAGAGUACCUCUG GUGUCCAGAC	36	55.6
BG-008	GUCUGGACAC GUG CAGAGUACCUCUG GUGUCCAGAC	36	58.3
BG-009 (Bulge C)	GUCUGGACAC UGU CAGAGUACCUCUG GUGUCCAGAC	36	55.6
BG-010 (Bulge D)	GUCUGGACAC GAUA CAGAGUACCUCUG GUGUCCAGAC	37	54
BG-011	GUCUGGACAC AGUA CAGAGUACCUCUG GUGUCCAGAC	37	54
BG-012	GUCUGGACAC UGAA CAGAGUACCUCUG GUGUCCAGAC	37	54
HP-001	CAUGUGCUGGACAU GGAA AUGUCCAGCACAUG	32	56.7
HP-002 (HP A)	CAUGUGCUGGACAU GCUA AUGUCCAGCACAUG	32	56.7
HP-003	CAUGUGCUGGACAU GACA AUGUCCAGCACAUG	32	56.7
HP-004	CAUGUGCUGGACAU CAGG AUGUCCAGCACAUG	32	59.5
HP-005	CAUGUGCUGGACAU UACA AUGUCCAGCACAUG	32	54
HP-006	CAUGUGCUGGACAU AGAC AUGUCCAGCACAUG	32	56.7
HP-007	CAUGUGCUGGACAU CUUC AUGUCCAGCACAUG	32	56.7

HP-008	CAUGUGCUGGACAU AUAC AUGUCCAGCACAUG	32	54
HP-009	CAUGUGCUGGACAU CGAUA AUGUCCAGCACAUG	33	55.2
HP-010 (HP B)	CAUGUGCUGGACAU AUAUG AUGUCCAGCACAUG	33	52.6
HP-011	CAUGUGCUGGACAU GCAUA AUGUCCAGCACAUG	33	55.2
HP-012 (HP C)	CAUGUGCUGGACAU ACAGUG AUGUCCAGCACAUG	33	55.2
STM-001 (Stem A)	GGACAU GAUCUG CAGACUACGUCUG CAGAUC AUGUCC	37	51.3
STM-002	GGACAU ACAGAU CAGACUACGUCUG AUCUGU AUGUCC	37	45.9
STM-003	GGACAU ACUUCA CAGACUACGUCUG UGAAGU AUGUCC	37	45.9
STM-004	GGACAU CUGGAC CAGACUACGUCUG GUCCAG AUGUCC	37	56.7
STM-005	GGACAU GCUUAC CAGACUACGUCUG GUAAGC AUGUCC	37	51.3
STM-006 (Stem B)	GGACAU UCACGC CAGACUACGUCUG GCGUGA AUGUCC	37	56.7
STM-007	GGACAU CAGACG CAGACUACGUCUG CGUCUG AUGUCC	37	56.7
STM-008	GGACAU GGAGAG CAGACUACGUCUG CUCUCC AUGUCC	37	56.7
STM-009	GGACAU UCUCUC CAGACUACGUCUG GAGAGA AUGUCC	37	51.3
STM-010 (Stem C)	GGACAU ACGAUC CAGACUACGUCUG GAUCGU AUGUCC	37	51.3
STM-011	GGACAU GAACUA CAGACUACGUCUG UAGUUC AUGUCC	37	45.9
STM-012	GGACAU CUACAG CAGACUACGUCUG CUGUAG AUGUCC	37	51.3

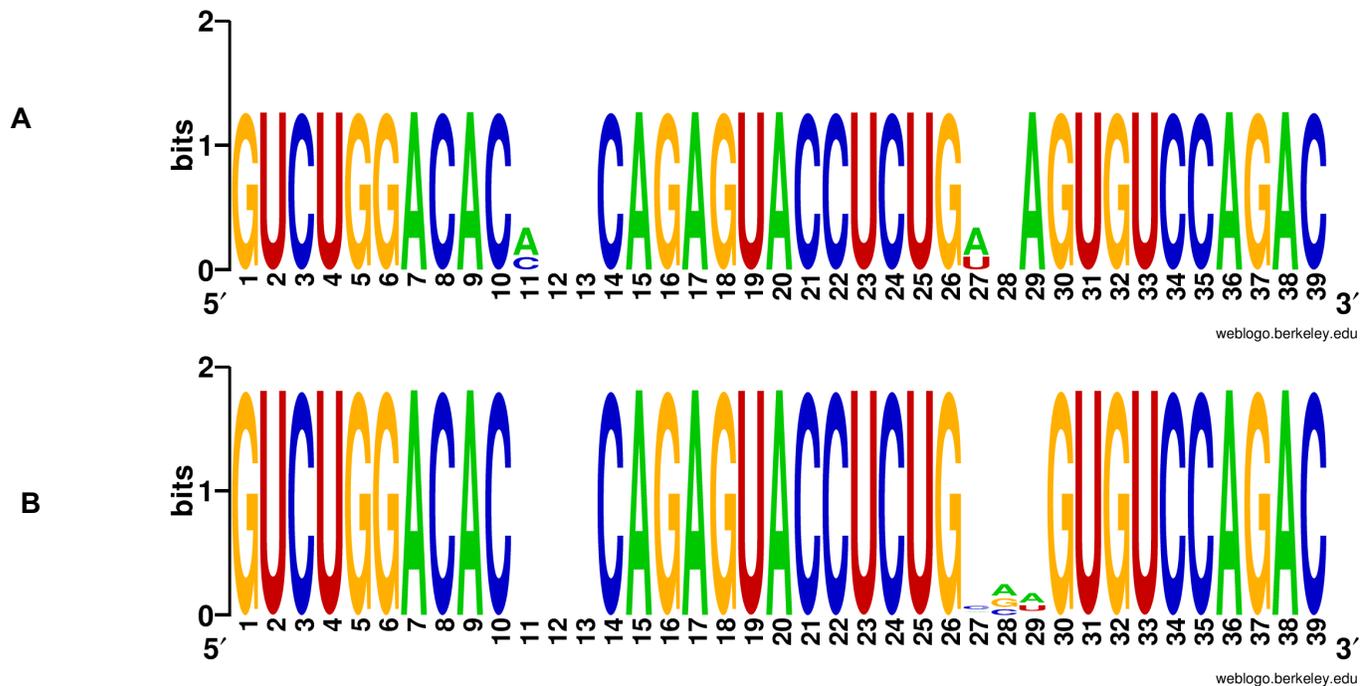


Figure S2-1: Sequence logo of Internal Loop sequences from the training set (A) and the extended dataset (B)

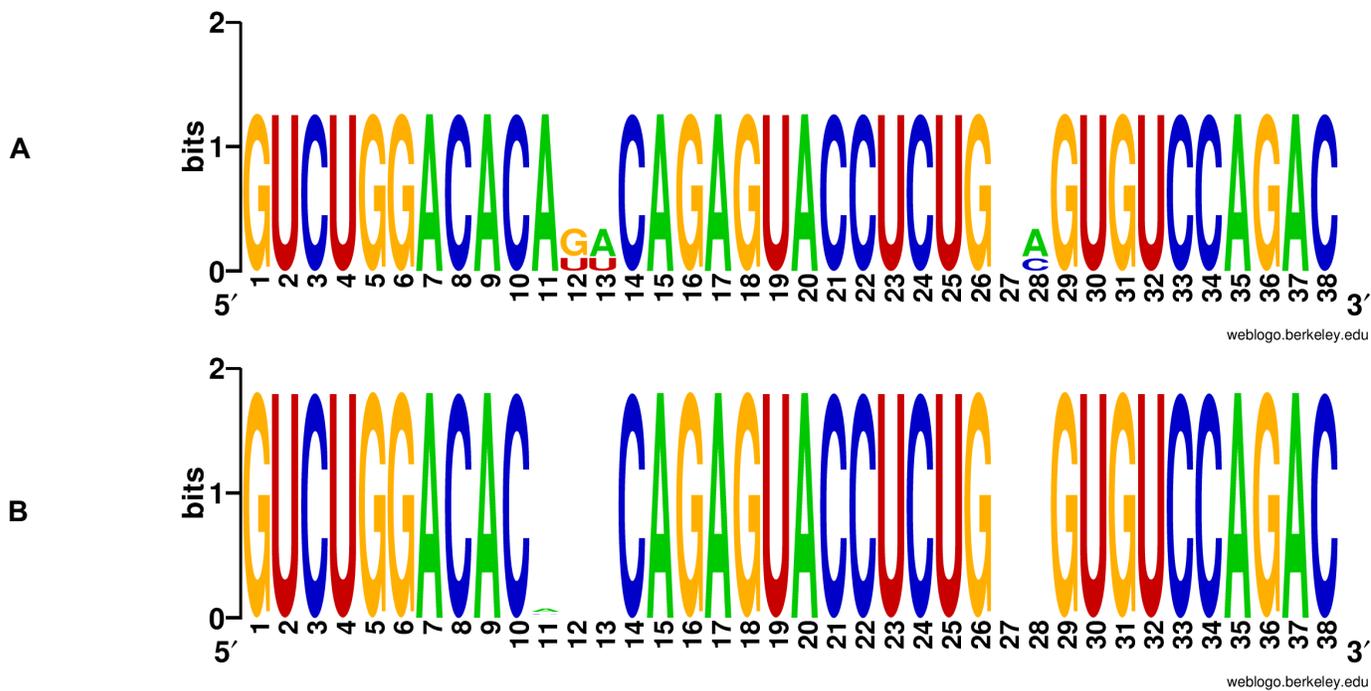


Figure S2-2: Sequence logo of Asymmetric Internal Loop sequences from the training set (A) and the extended dataset (B)

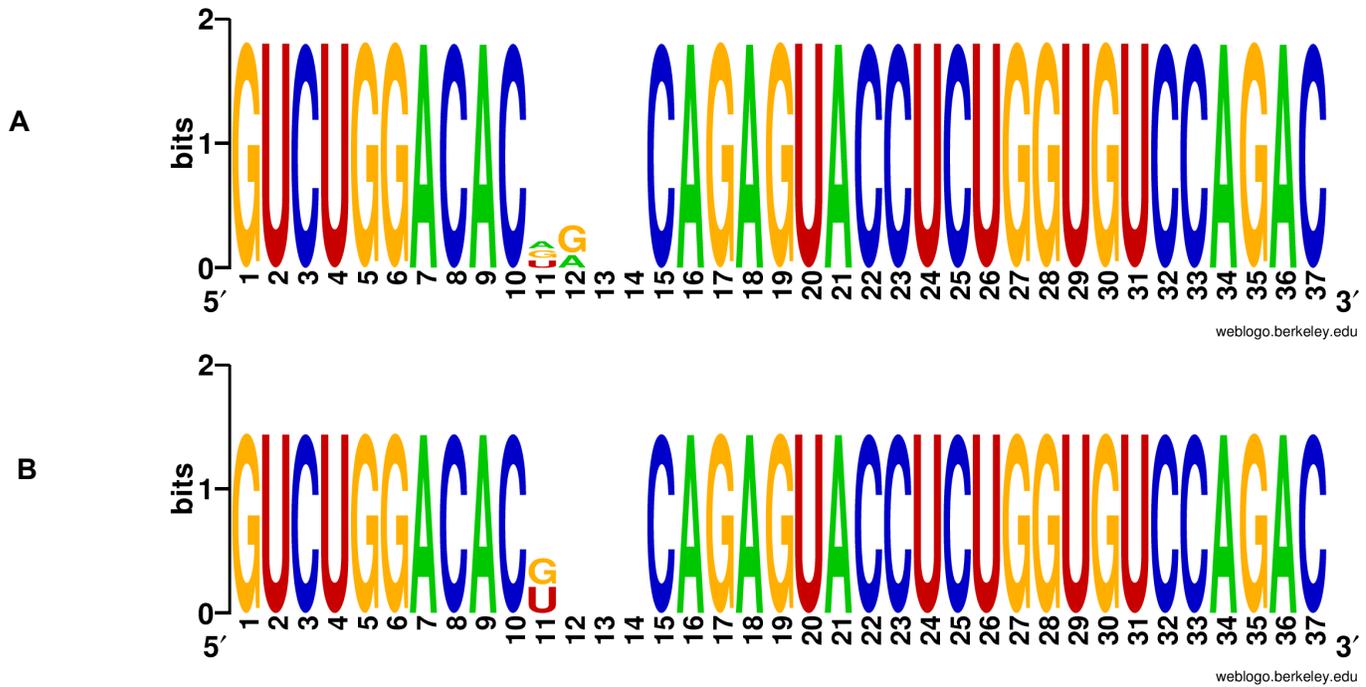


Figure S2-3: Sequence logo of Bulge sequences from the training set (A) and the extended dataset (B)

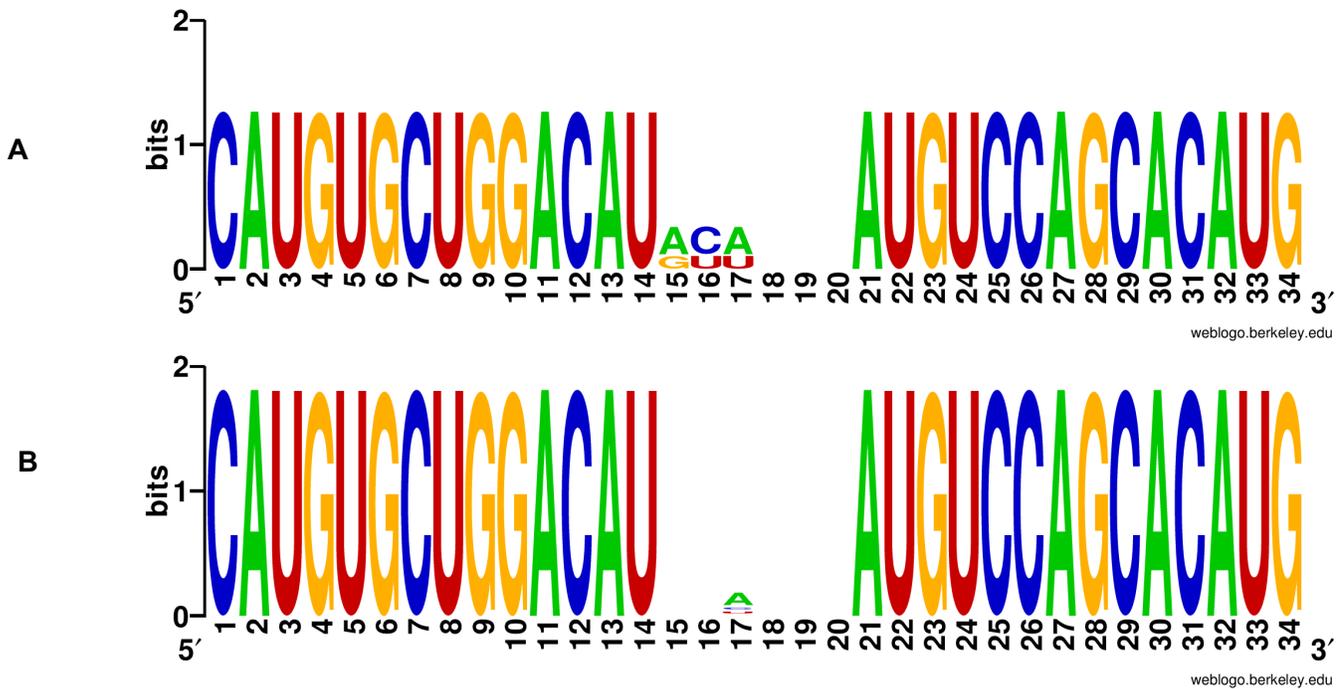


Figure S2-4: Sequence logo of hairpin sequences from the training set (A) and the extended dataset (B)

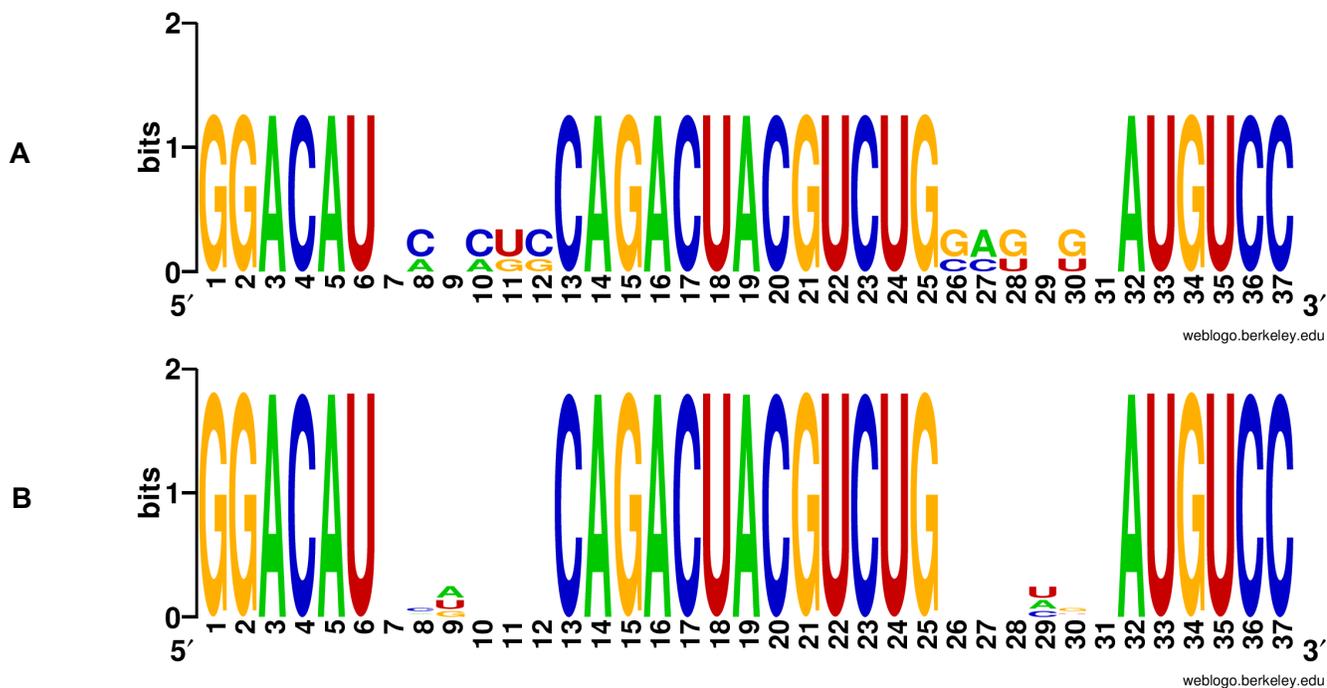


Figure S2-5: Sequence logo of Stem sequences from the training set (A) and the extended dataset (B)

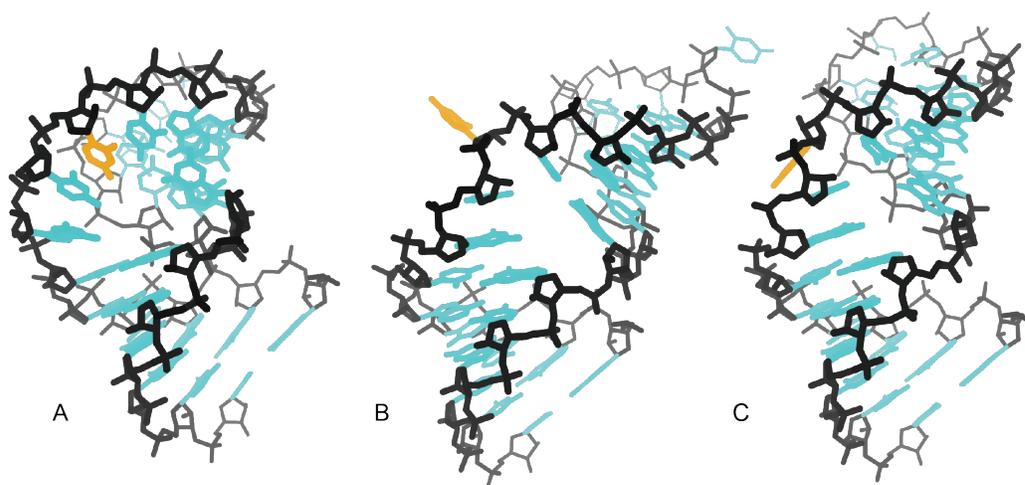
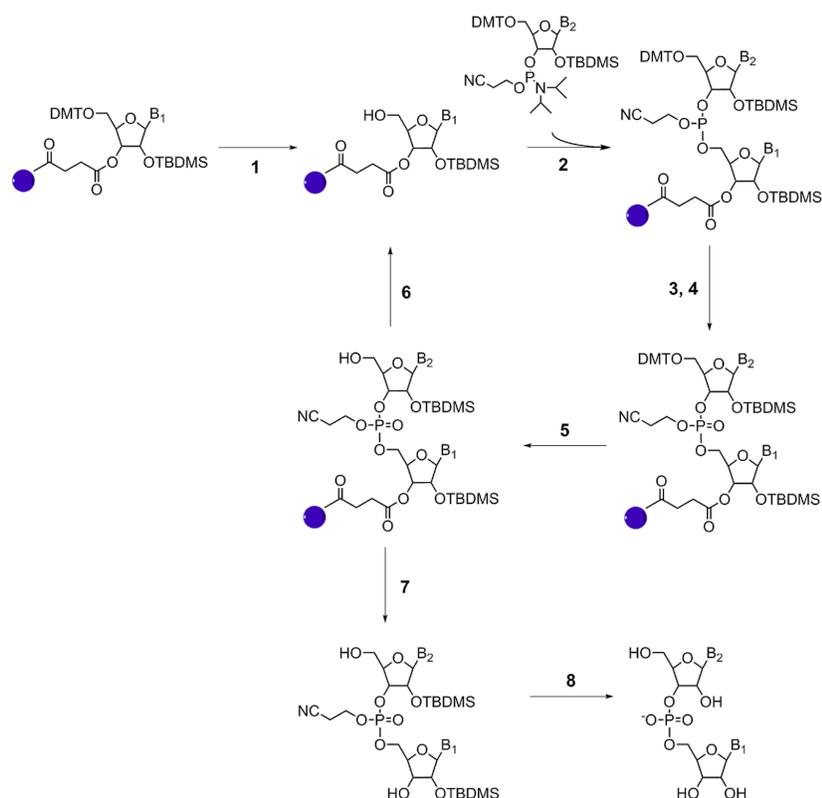


Figure S2-6: Three predicted structures of the BG-B sequence. A, B, and C depict the RNA backbone in black. The uracil marked for replacement with BFU is in gold. A and C show the uracil stacking with the other bulged bases, while B shows it flipped into the solvent.

S3. RNA Solid Phase Synthesis and Purification



Scheme S3-1. Outline of solid phase synthesis cycle. Key: 1. Detritylation; 2. Activation and Coupling; 3. Capping; 4. Oxidation, I₂; 5. Detritylation; 6. Next cycle; 7. Cleavage from solid-phase; 8. Deprotections.

A solid phase synthesizer was used for synthesis of all RNA sequences and allowed for selective modification with the BFU nucleoside. In total, 16 sequences of the RNA training set, a BFU modified bulge (U25) and hairpin (G33) TAR were synthesized. Orthogonally protected adenosine, guanosine, cytosine, uracil, and BFU phosphoramidite were dissolved in anhydrous acetonitrile. Polystyrene columns were used with the first 3' nucleotide attached. For each nucleotide, the 4,4' dimethoxytrityl group was removed with trichloroacetic acid. The next nucleotide was incorporated using a tetrazole catalyst. Unreacted 5' hydroxyl groups are capped with acetic anhydride and N-methylimidazole dissolved in tetrahydrofuran/pyridine. Finally, the phosphodiester bond is oxidized with iodine in pyridine/water mixture, which

increases the stability of the RNA sequence. The synthesis is then repeated until the RNA sequence has been fully synthesized.

Purification of the RNA sequences began with cleavage of the RNA off the solid phase support and deprotection of the bases of the sequences with 333 μL x3 ammonium methylamine for 7.5 minutes and then allowed to incubate for two hours. Afterwards, the solutions were dried down using a vacuum concentrator until crystals were formed. To deprotect the 2' hydroxyl group, crystals were dissolved in 115 μL dimethyl sulfoxide, 60 μL triethylamine, and 75 μL of triethylamine: hydrogen fluoride (30%) and heated for 2.5 hours at 65 °C. After cooling to room temperature, 1.75 mL of quenching buffer (Glen Research) is added to the solutions. Finally, a polydivinylbenzene 4,4' dimethoxytrityl affinity column was used to purify the RNA sequence. The columns are pre-conditioned with 0.5 mL acetonitrile and 1 mL 2M triethylammonium acetate (TEAA). The RNA sequence solutions are added to the column, washed with 1 mL of a 1:9 acetonitrile:TEAA solution, 1 mL water, 2 mL trifluoroacetic acid, and 2 mL water. Subsequently the RNA is eluted from the column, and the 4,4' dimethoxytrityl group is removed, with 1 mL of 1 M ammonium bicarbonate. Ethanol precipitation is used as the final step in RNA purification. The RNA sequences were dissolved in phosphate buffer (10 mM NaH_2PO_4 , 25 mM NaCl, 4 mM MgCl_2 , 0.5 mM EDTA, pH 7.3), and the concentration was analyzed using a Nanodrop spectrophotometer. The purity of the RNA was determined with 20% polyacrylamide gel electrophoresis (PAGE) run with 1X tris, borate, and EDTA (TBE) buffer. Gels were made by mixing 75 mL 20% polyacrylamide solution (19:1 acrylamide:bis-acrylamide), 750 μL ammonium persulfate, and 75 μL tetramethylethylenediamine (TEMED) and poured into a glass gel support. After 1 hour of polymerization, the gel was pre-run without sample for 1 hour at 11 W. The gels were run with sample at 11 W for ~2.5 hours. Afterwards, the gel was stained with Diamond dye[®] for 30 minutes, followed by washing with water 2x15 minutes.

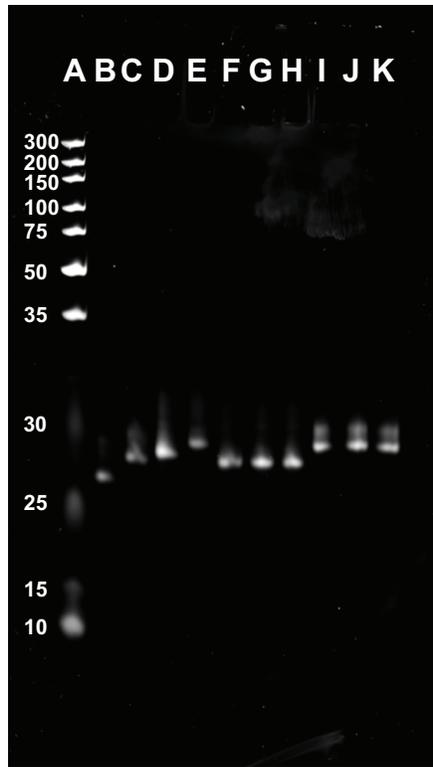


Figure S3-1. 20% PAGE gel run at a constant 11W for 2.5 hours in 1X TBE buffer. The columns from left to right; A) ladder, B) Blg A, C) Blg B, D) Blg C, E) Blg D, F) Hp A, G) Hp B, H) Hp C, I) AL A, J) AL B, K) AL C.

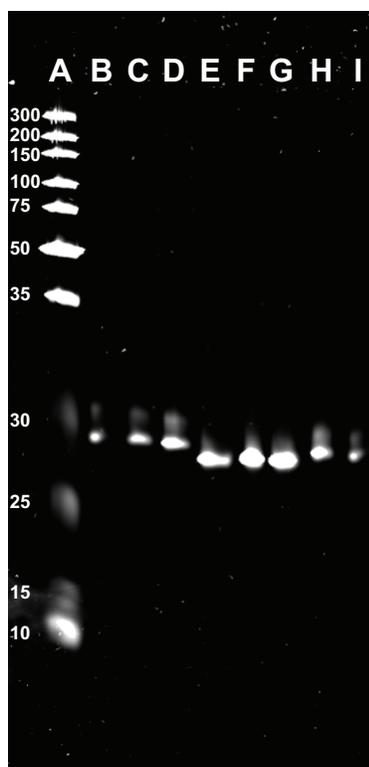


Figure S3-2. 20% PAGE gel run at a constant 11W for 2.5 hours in 1X TBE buffer. The columns from left to right; A) ladder, B) Stm A, C) Stm B, D) Stm C, E) IL A, F) IL B, G) IL C, H) TAR Blg, I) TAR Hp.

S4. Plate Reader Assay: Method and Data

A serial dilution of small molecule aminoglycosides (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 2, 2.8, 3.2, and 4 μM final concentration in 384 well plate) was performed in a 96 well plate. In a 384 well plate, 10 μL of each small molecule solution is added in triplicate, followed by 10 μL of 400 nM RNA (final concentration: 200 nM). The plate is shaken on an orbital shaker for 10 minutes at 100 RPM in order to confirm proper mixture of the small molecule and RNA, followed by centrifugation at 3220 x g for 1 minute to remove any possible air bubbles. After incubation for 15 minutes in the dark, the plate is scanned at excitation of 322 nm and emission of 455 nm, 50 flashes/read (**Fig S4-1**). The following graphs are the fluorescence titration results of each RNA sequence to the small molecule library (**Fig S4-2**).

Each experimental point was taken as the average of three wells in a single experiment. Errors were calculated from the standard deviation of the triplicate experiments. The error for each concentration was averaged to obtain the average error for each aminoglycoside:RNA titration (Table S4-1).

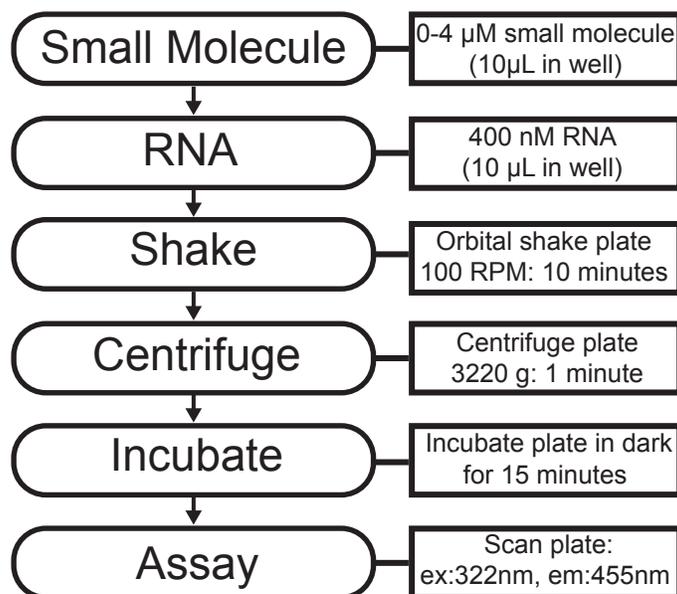
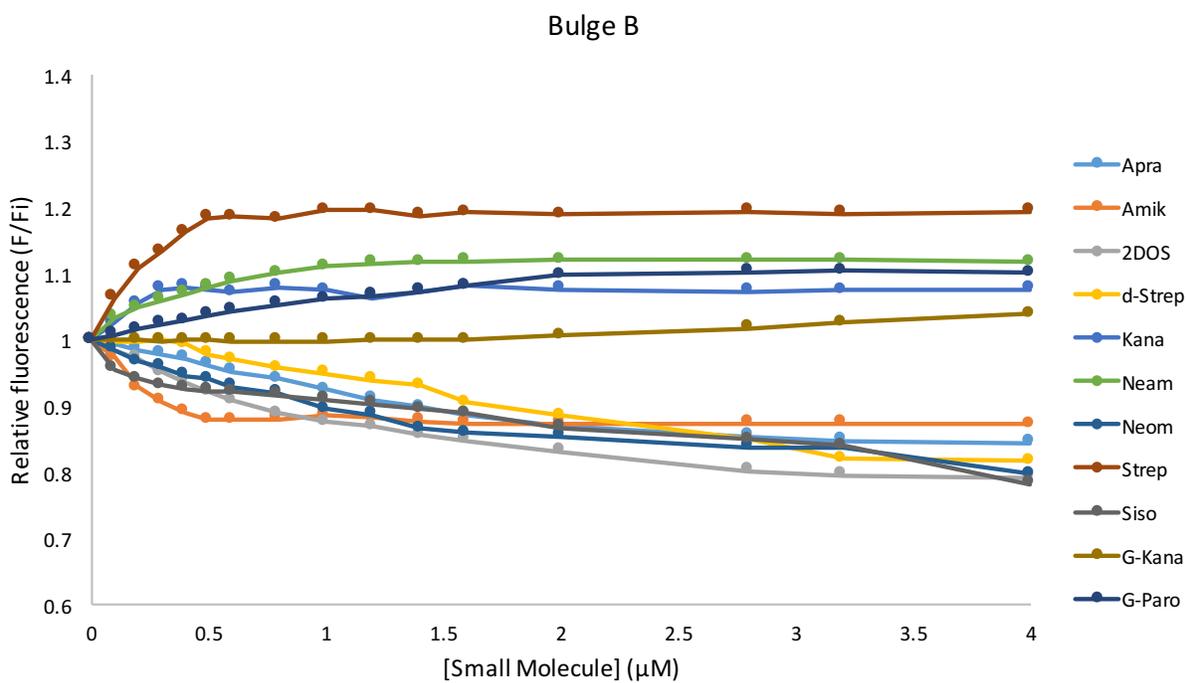
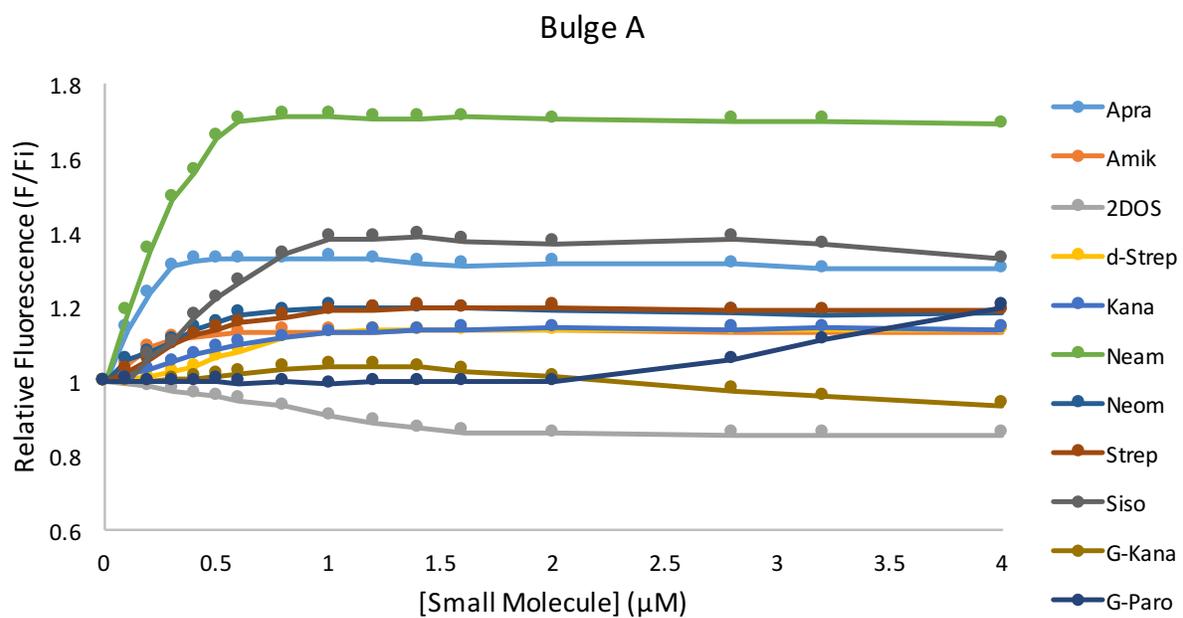
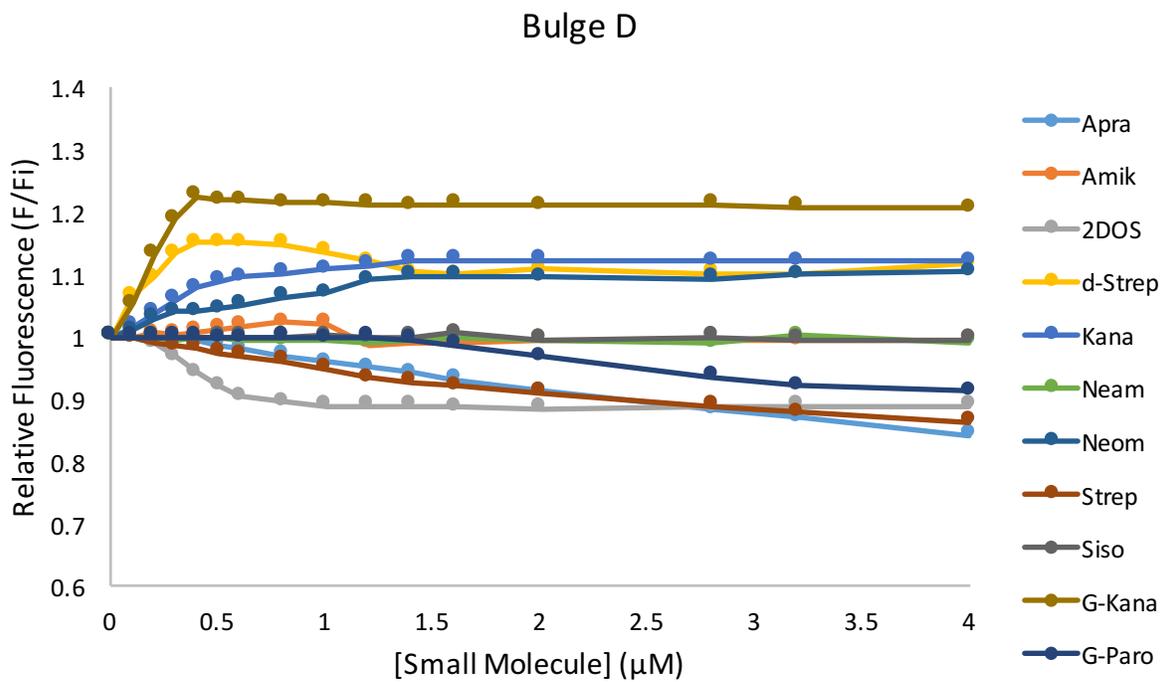
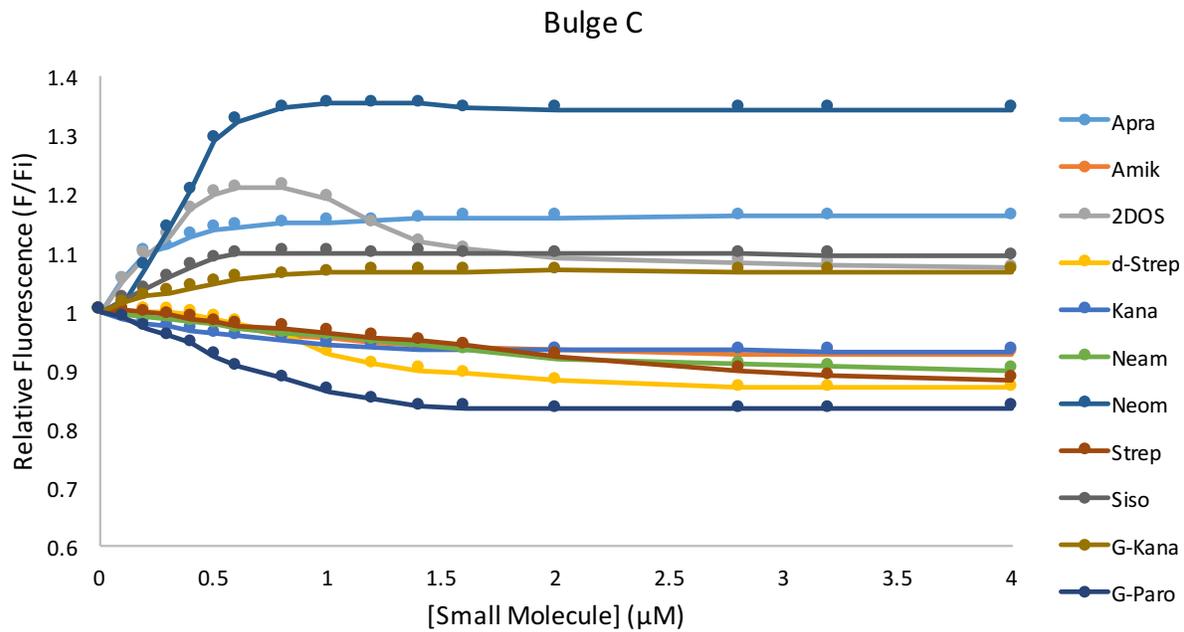
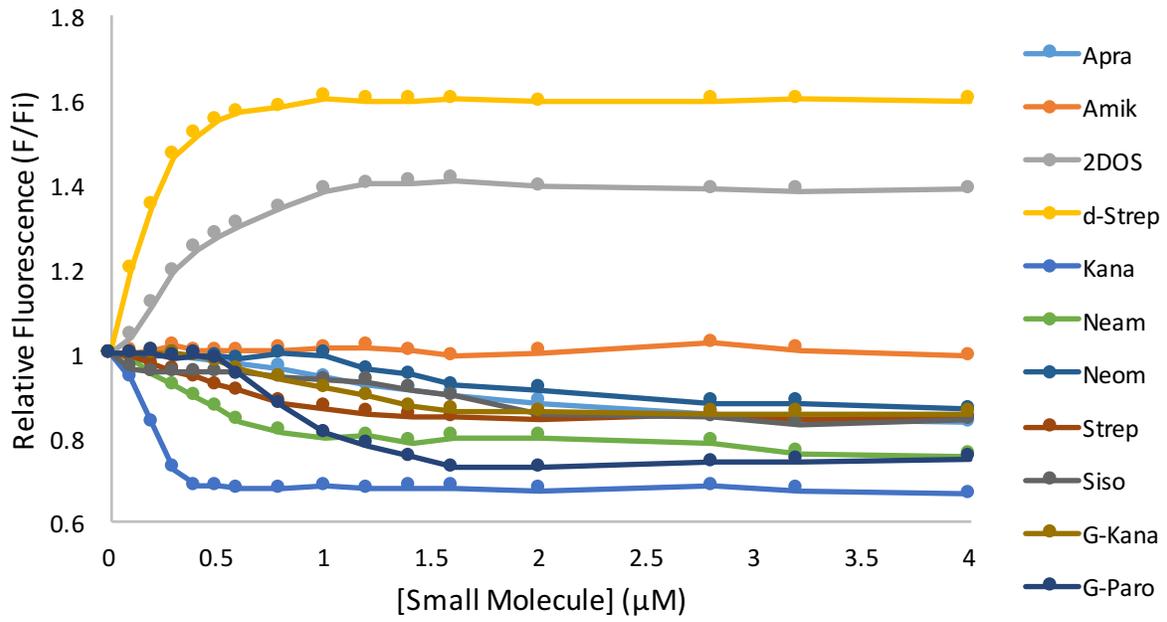


Figure S4-1. Aminoglycosides (0-4 μM) and the labeled RNA (200 nM) were diluted in 10mM NaH₂PO₄, 25mM NaCl, 4mM MgCl₂, and 0.5mM EDTA at pH 7.3. The aminoglycoside and RNA were combined and shaken in a 384 well plate, followed by centrifugation. After incubating the plate, the wells were excited at 322 nm and scanned for emission at 455 nm.

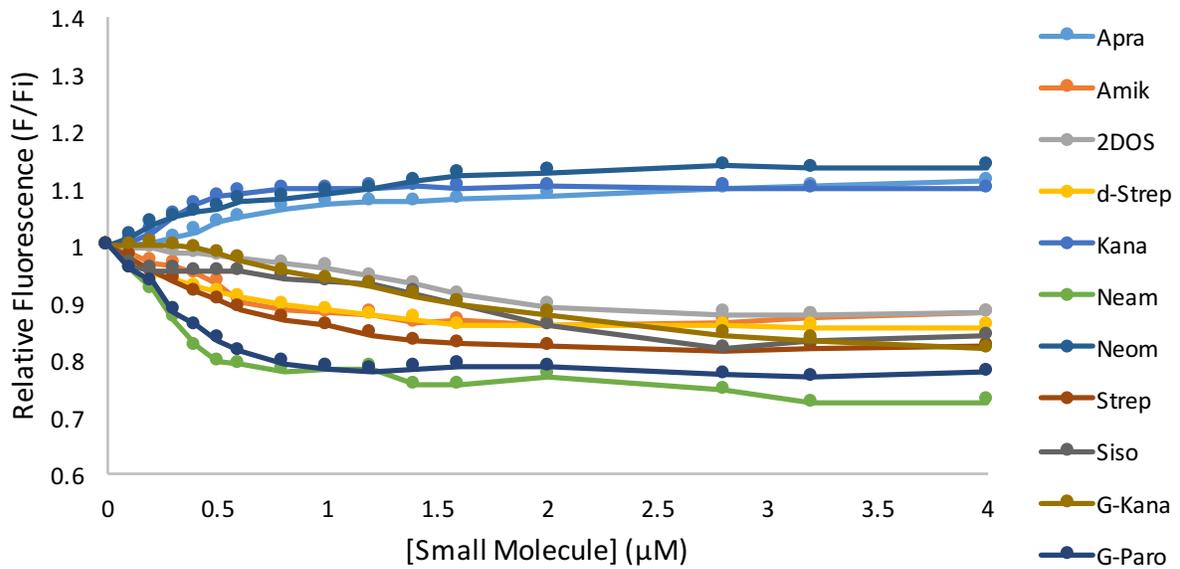




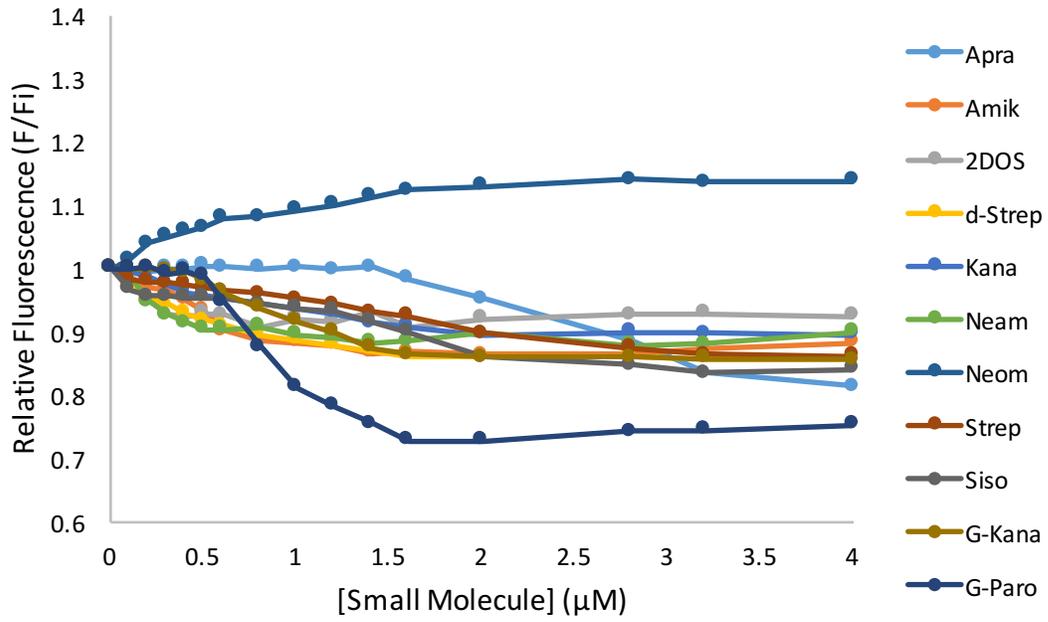
IL A



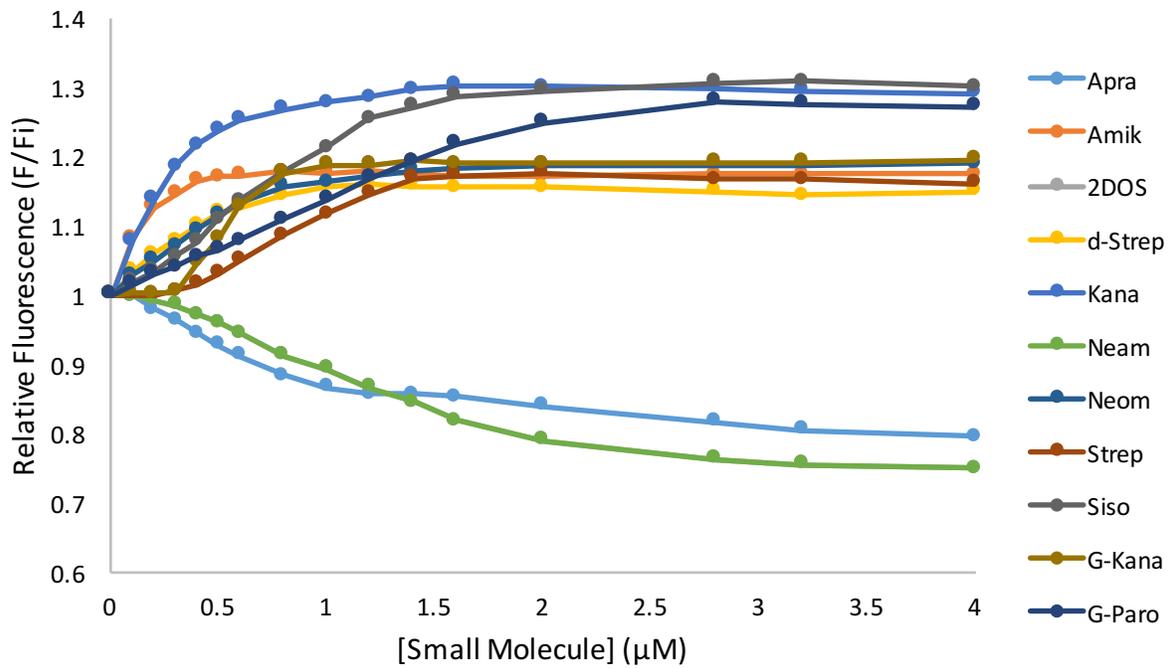
IL B



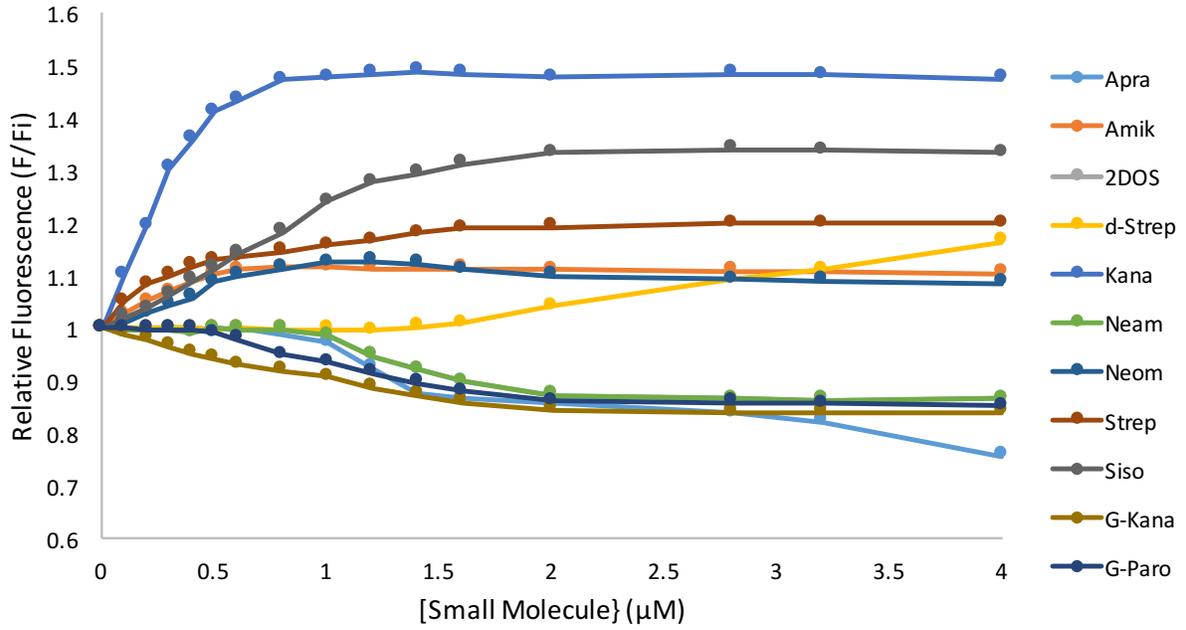
IL C



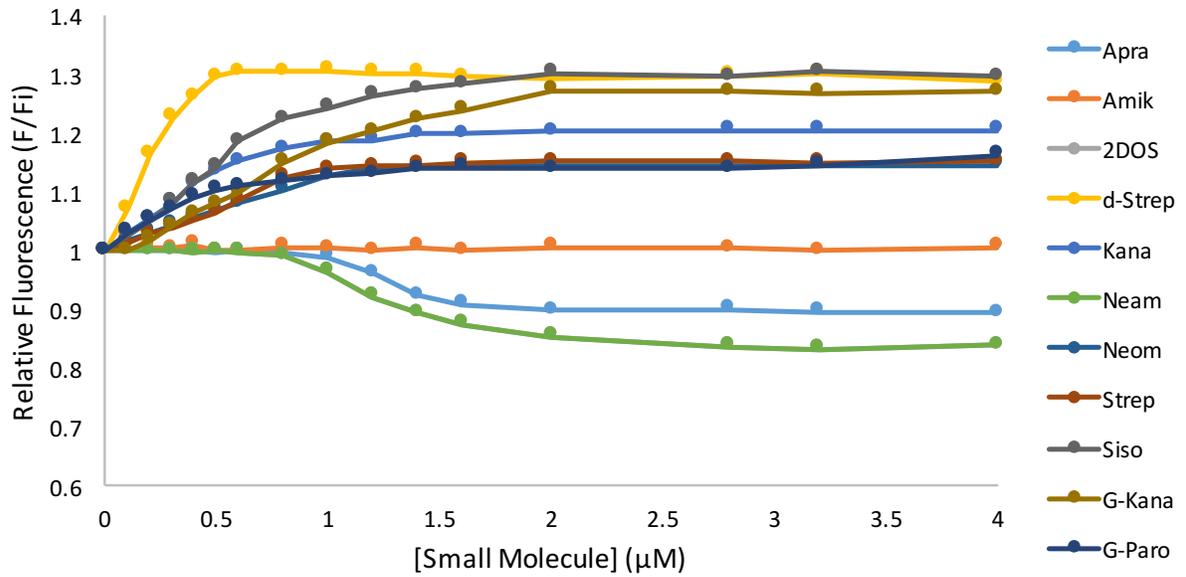
AIL A



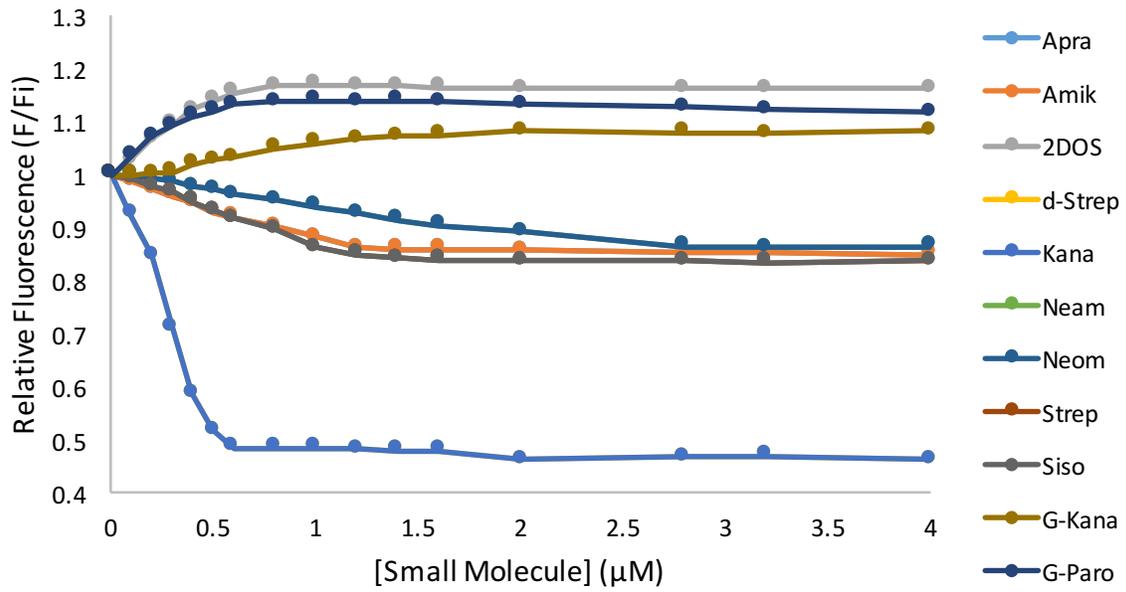
AIL B



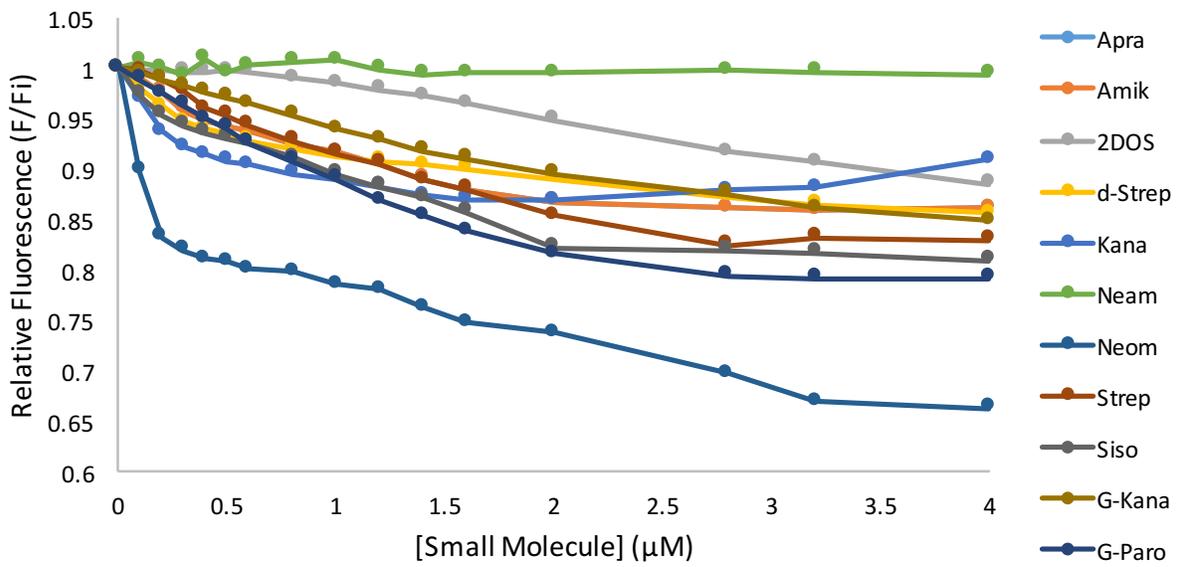
AIL C



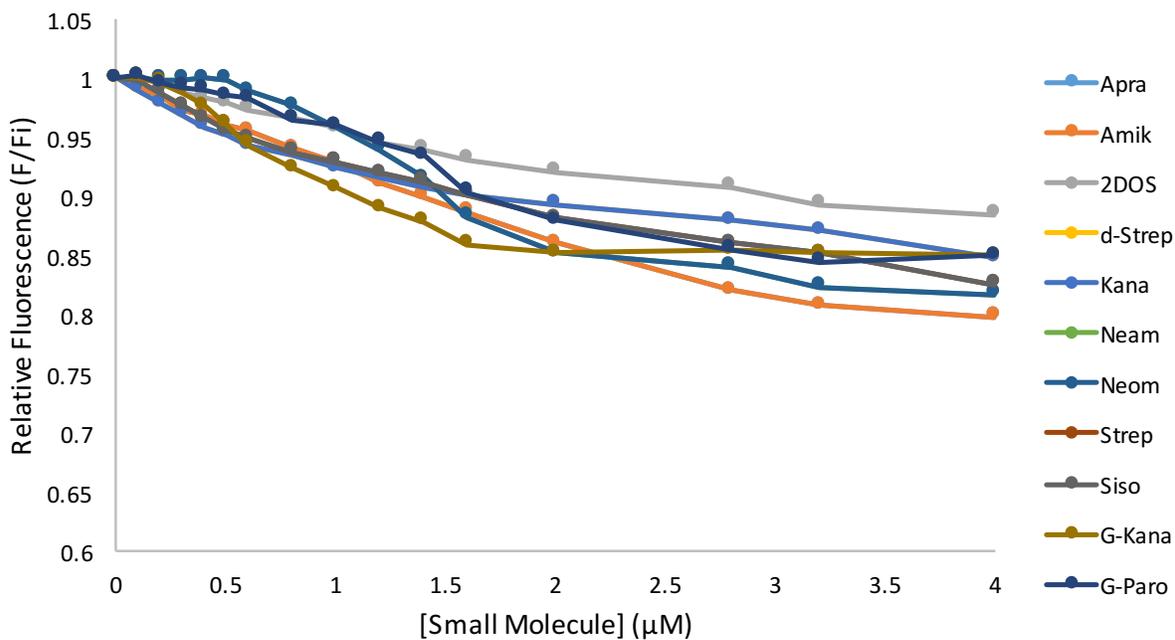
HP A



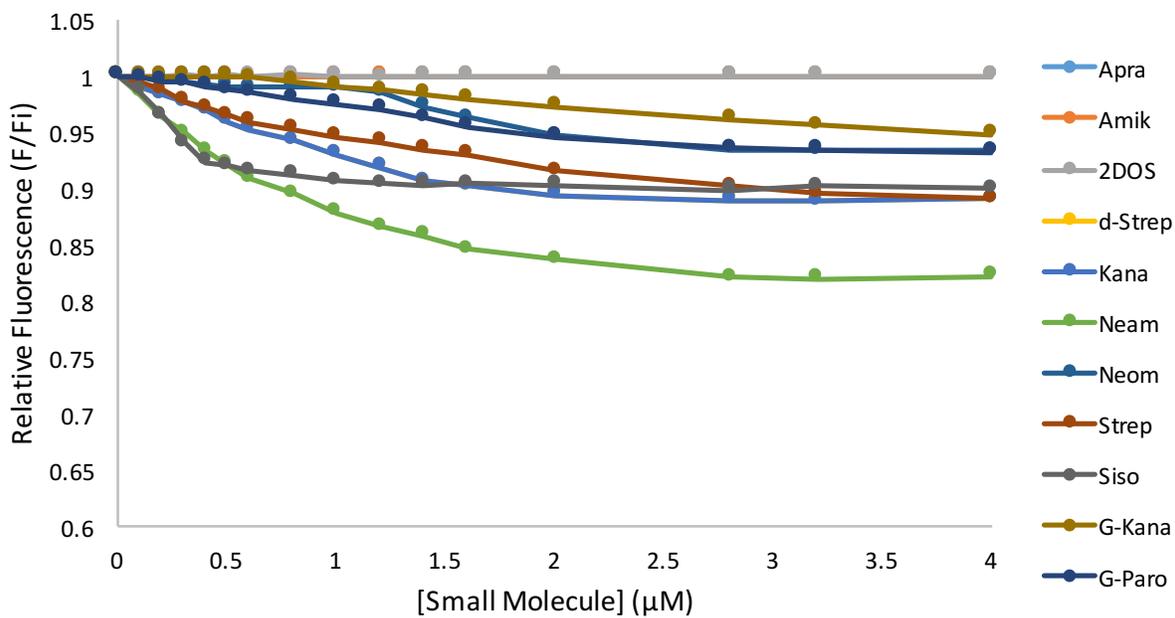
HP B



HPC



Stem A



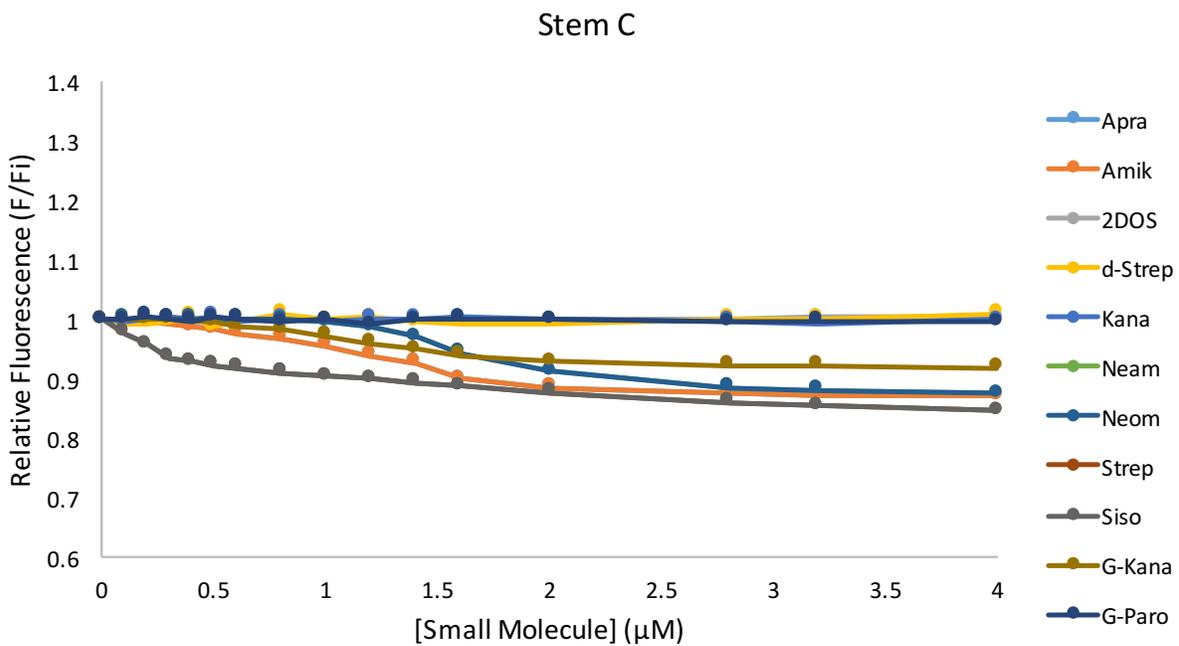
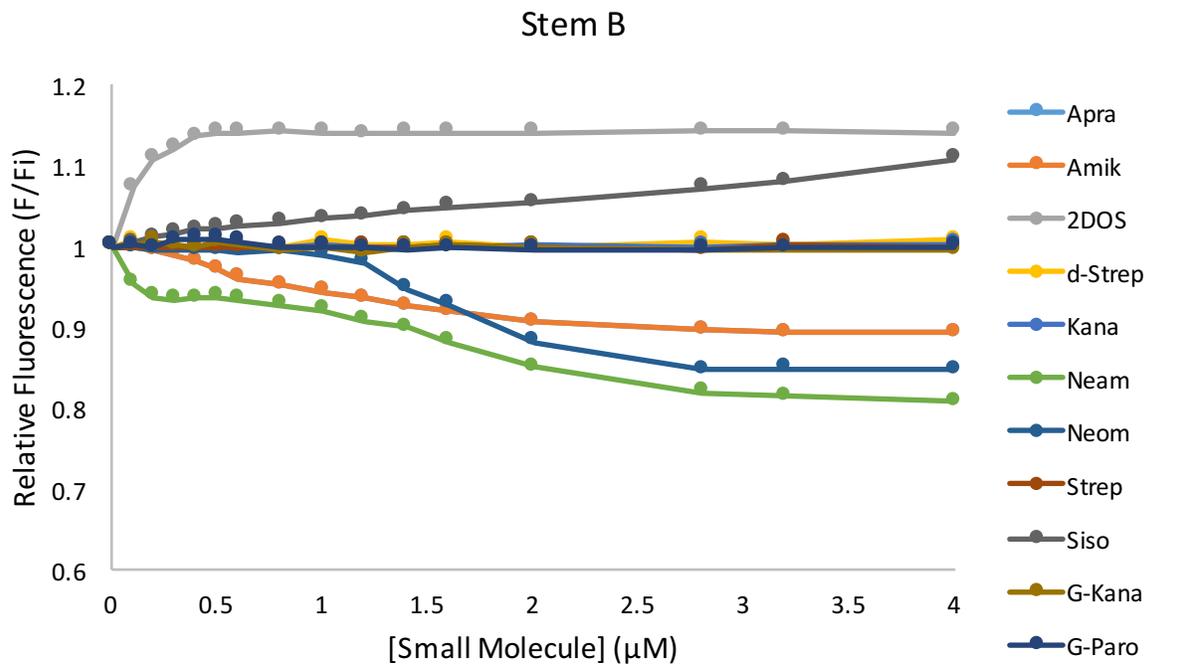


Figure S4-2. All RNA sequence titration curves to the small molecule receptors. Each graph shows a single RNA sequence and the receptors titration curves.

Table S4-1. Percent error of the titration curves. Each titration curve was run in experimental triplicate, and the standard deviation was calculated for each point. The averaged percent errors for each small molecule:RNA pair are shown below. Raw data is available upon request.

	AL A	AL B	AL C	Bg A	Bg B	Bg C	Bg D	Hp A	Hp B	Hp C	IL A	IL B	IL C	Stm A	Stm B	Stm C	TAR HP	TAR Bg
Amik	4.95	5.92	6.42	5.54	5.09	5.57	3.71	6.47	6.53	4.52	7.32	5.52	7.31	5.65	6.53	5.78	7.13	5.01
Apra	3.95	4.16	4.19	7.28	4.39	4.13	6.52	3.57	5.81	5.12	6.71	4.24	3.93	4.95	5.27	6.68	3.47	5.62
2DOS	7.56	4.36	4.39	3.49	4.56	5.26	4.68	3.91	4.49	4.06	4.09	4.61	6.28	5.24	6.97	4.88	5.36	4.44
D-Strep	4.22	3.57	7.35	5.23	5.27	5.02	7.45	3.88	5.16	5.63	5.07	7.05	6.56	5.82	6.09	6.41	3.65	4.78
Kana	5.06	5.99	5.04	5.61	4.27	4.56	5.48	5.18	4.40	4.91	5.33	6.01	4.74	5.72	6.55	6.07	4.81	3.86
Neam	4.96	7.02	5.88	4.94	3.68	4.28	5.03	3.73	3.45	2.95	6.92	4.46	5.15	4.69	5.51	6.33	3.55	6.42
Neom	7.75	4.00	5.43	3.67	5.33	5.57	4.52	6.19	7.41	10.30	3.93	7.53	7.22	4.13	6.47	5.88	4.27	6.71
Strep	7.63	4.86	7.43	5.53	5.43	6.13	7.62	4.75	4.89	7.13	5.05	4.24	6.68	5.36	6.87	5.07	6.18	4.06
Siso	4.76	7.57	3.52	6.12	4.13	7.81	5.58	4.93	7.32	7.25	6.34	5.16	3.64	7.26	6.10	5.52	4.92	4.26
Guan-Kana	5.00	4.66	3.72	7.62	3.91	7.39	6.48	5.87	3.81	5.87	7.11	6.88	6.85	7.77	4.42	6.72	3.37	5.04
Guan-Paro	4.81	4.48	7.28	5.28	4.92	5.46	3.56	6.75	4.81	8.51	7.58	5.81	4.24	6.62	4.09	6.28	3.92	4.06
Hygro	5.59	5.78	7.11	4.71	7.56	4.08	5.79	7.44	7.75	6.75	4.06	6.86	3.75	4.14	7.33	6.42	*	*
Tobra	4.58	3.91	7.32	8.98	7.11	6.97	5.04	5.32	6.12	5.75	5.93	7.33	4.51	5.46	4.87	4.37	*	*
Paro	5.82	3.58	6.45	5.24	5.78	5.76	4.55	3.45	5.34	5.84	5.07	5.74	7.82	5.22	4.37	4.65	*	*

S5. Training Set Principal Component Analysis and Loading Plots

Raw fluorescence data is inserted into the XLSTAT software (Addinsoft) in order to determine the principal components for the RNA training set. The first PCA plot had a predictive power of 78% using leave-one-out analysis (**Fig S5-1**) and it was determined hygromycin-B, tobramycin, and paromomycin could be removed and the predictive power of the PCA plot increased to 87% predictive power. Finally, the PCA with the guanidinylated aminoglycosides was found to have 100% predictive power for the clusters between RNA structure motifs. The loading factors indicate the importance of each aminoglycoside to the PCs from PC1-PC11 (**Table S5-1**). The loading plots for PC1 vs PC2, PC1 vs PC3, and PC2 vs PC3 allows the aminoglycoside importance for each PC compared to the other aminoglycosides (**Fig S5-2**).

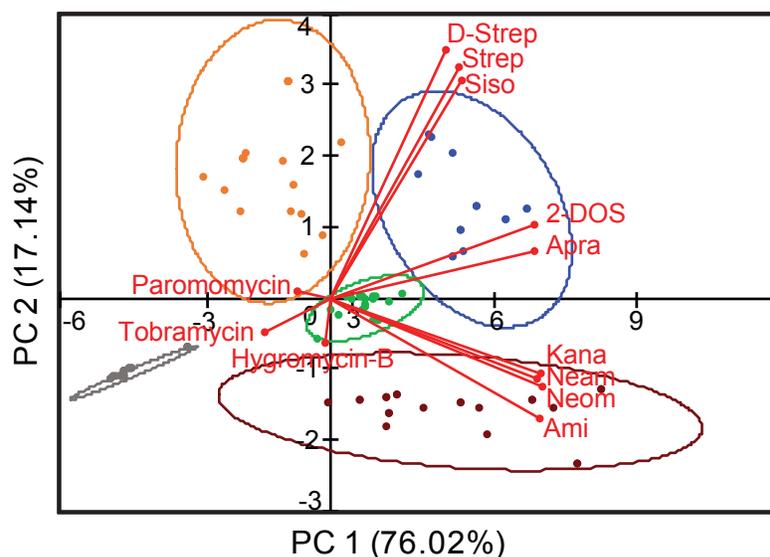


Figure S5-1. PCA plot of the RNA training set with all 12 commercially available aminoglycosides. Paromomycin, tobramycin, and hygromycin-B had the least effect on the principal components, and were removed to increase the predictive power from 78% to 87%.

Table S5-1: Loading factors of the final receptor library. The PCA gave 11 PC based on the data: 95% of the variance was explained by the first 3 PC's while 100% of the variance was explained by the first 10 PCs.

Loading Factors

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
2DOS	0.948	0.184	-0.177	-0.043	0.112	0.005	0.150	-0.039	0.008	-0.003	0.000
Amik	0.902	-0.390	-0.108	-0.013	-0.040	-0.118	-0.047	-0.060	-0.023	0.007	0.000
Apra	0.938	0.064	-0.230	0.068	-0.020	-0.132	-0.071	-0.018	0.027	-0.001	0.000
D-Strep	0.823	0.539	0.024	-0.182	0.156	-0.016	-0.064	0.052	-0.003	0.001	0.000
Guan-Kana	0.876	-0.337	0.155	0.224	0.202	0.037	-0.026	0.003	0.000	0.000	0.000
Guan-Paro	0.871	-0.020	0.447	-0.186	-0.002	-0.033	0.019	-0.004	-0.001	0.000	0.000
Kana	0.948	-0.299	-0.042	0.016	-0.101	0.006	-0.002	0.015	0.004	-0.023	-0.004
Neam	0.949	-0.289	-0.027	0.016	-0.104	0.047	0.024	0.023	0.015	0.029	-0.001
Neom	0.949	-0.291	-0.035	0.020	-0.107	0.024	0.014	0.018	0.006	-0.013	0.005
Siso	0.869	0.433	-0.001	-0.019	-0.035	0.224	-0.061	-0.035	-0.007	0.000	0.000
Strep	0.826	0.539	0.043	0.120	-0.058	-0.056	0.053	0.043	-0.026	0.003	0.000

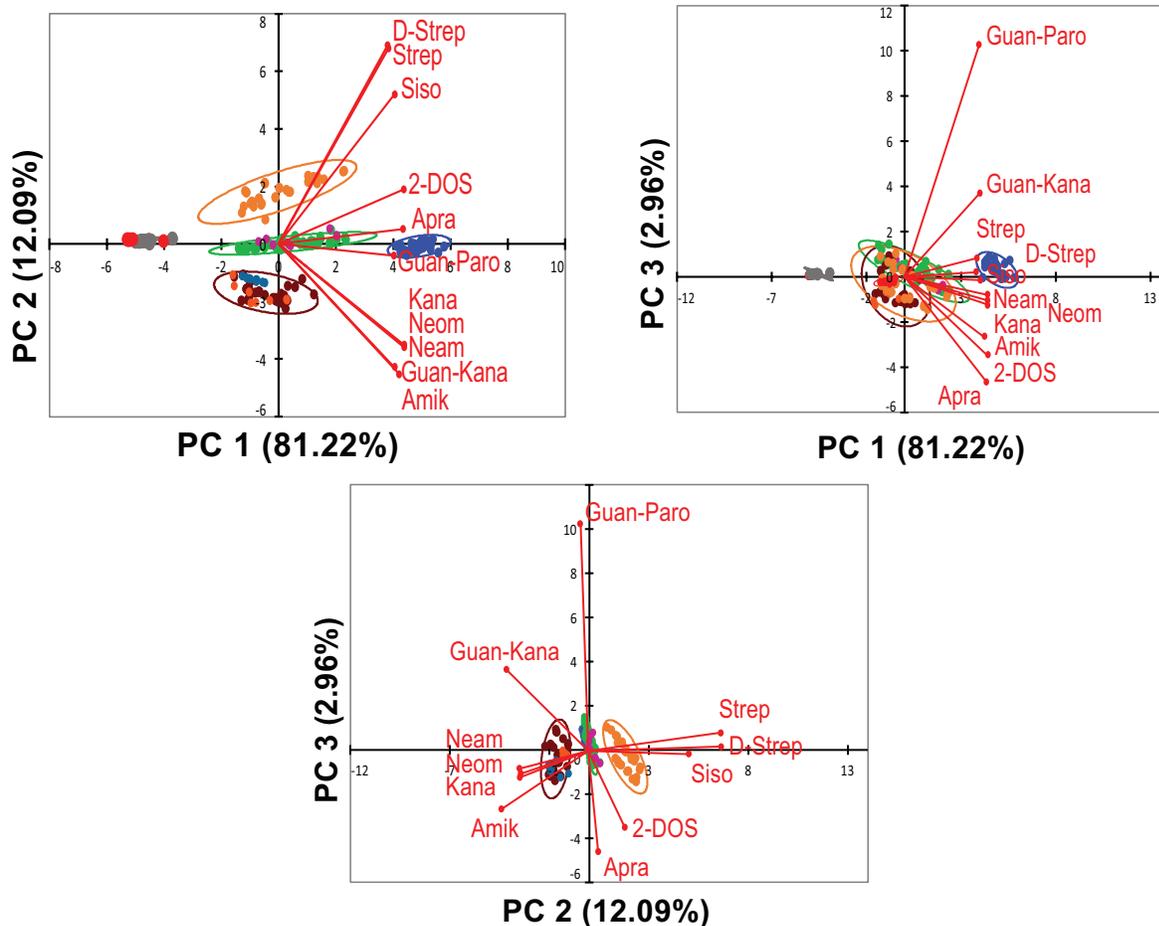


Figure S5-2. PCA plot and loadings for the training set with the expanded aminoglycoside library. All of the aminoglycosides are positively correlated with PC 1, while d-Strep and Strep are important for PC 2 and PC 3. The guanidinylated aminoglycosides nearly explained all the variance for PC 3.

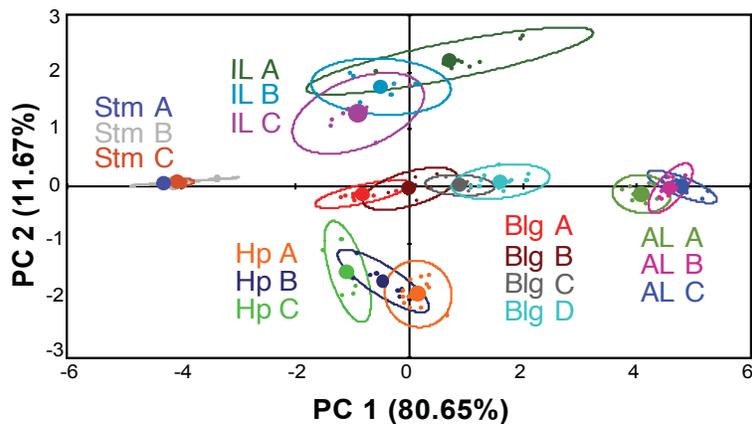


Figure S5-3. All RNA motifs separated with 95% confidence intervals. The centroid of each ellipse is shown with a large point.

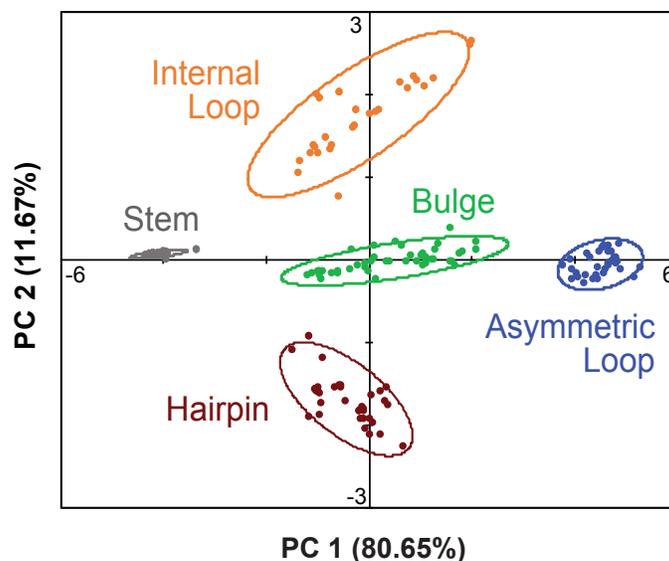


Figure S5-4. Removal of **2-DOS** and **Neam** had minimal change in the global PCA analysis. There is still 100% predictive power by LOOCV.

S6. TAR Structural Classification

Utilizing the solid phase synthesis from above, two different sites on TAR RNA were fluorescently tagged with BFU. The TAR RNA sequence used was: GGCAGAUC(U*)GAGCCUG(G*)GAGCUCUCUGCC; where the starred position are the two sites of modification, either within the bulge (U*) or the hairpin (G*). After synthesis, the fluorescence assay described above was run against the TAR RNA using the small molecule receptors. The graph below show the binding curves of the small molecule receptors to TAR RNA (**Figure S6-1**). The TAR RNA raw data was externally validated against the RNA training set principal components (**Fig S6-2**).

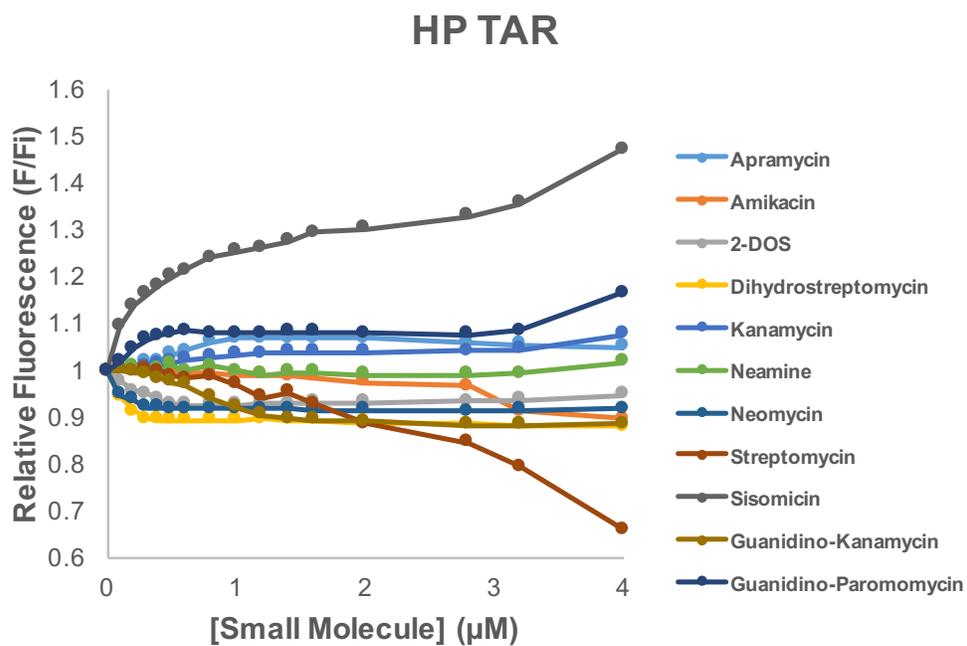
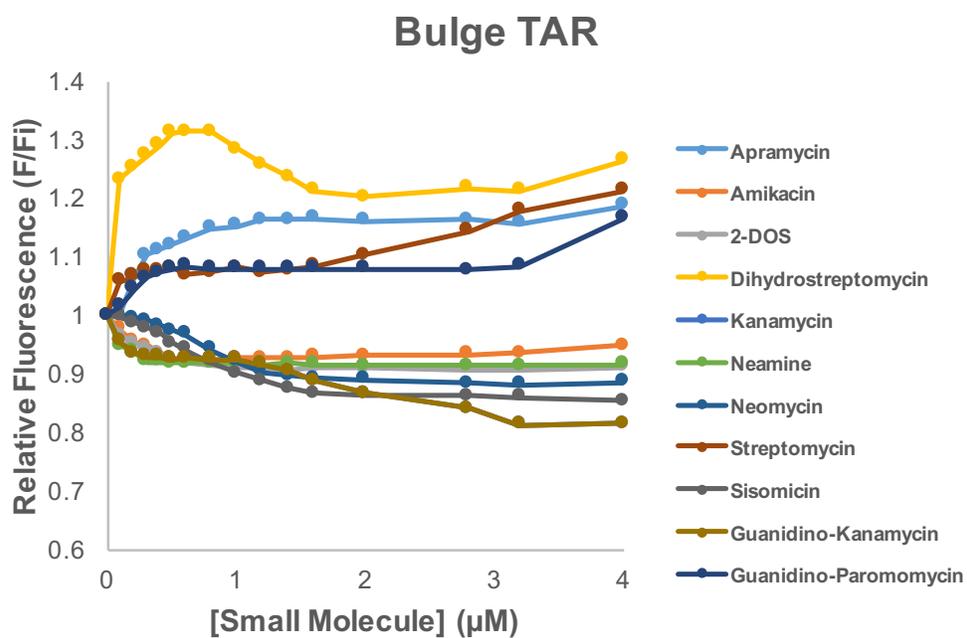


Figure S6-1. The titration curves of the small molecule receptors to bulge and hairpin TAR.

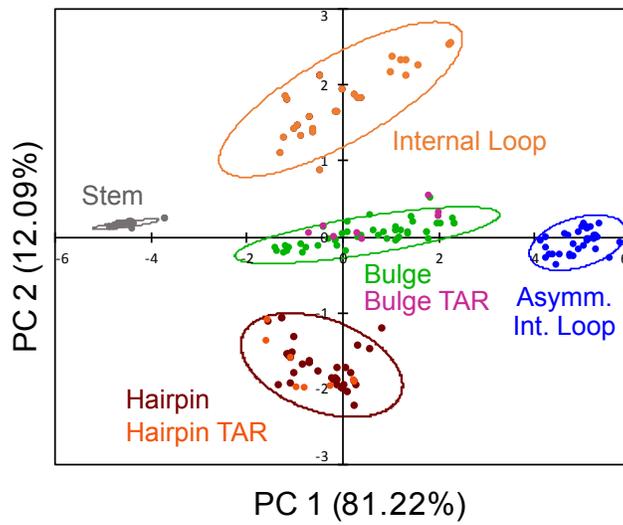


Figure S6-2. PCA plot of the RNA training set and the bulge and hairpin TAR RNA.

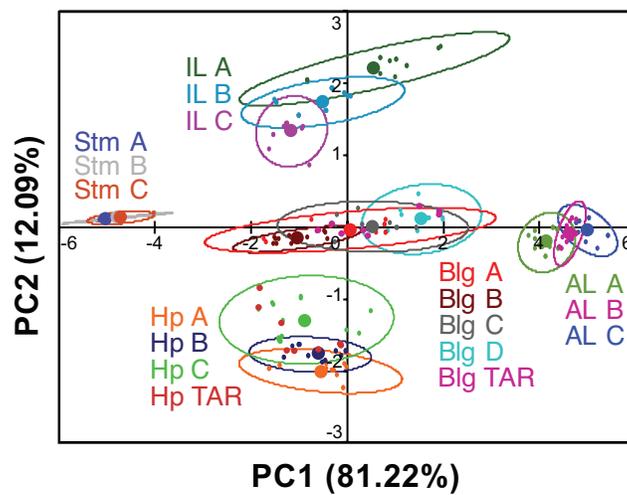


Figure S6-3. All RNA sequence separated into 95% confidence interval ellipses. The centroid of each ellipse is shown with a large point. The bulge TAR (Blg TAR) and hairpin TAR (Hp TAR) were externally validated and shown on the plot above.

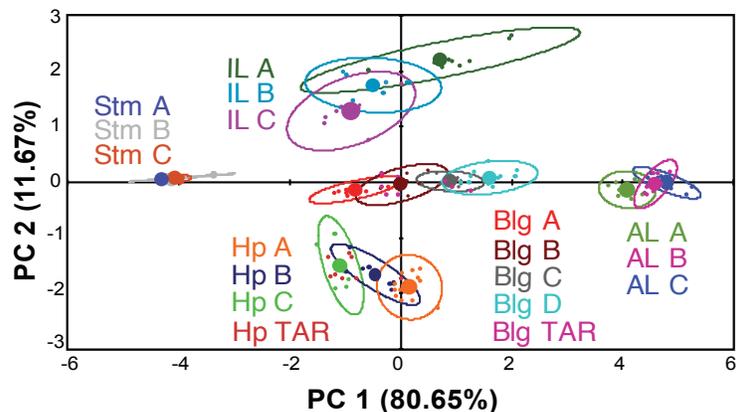


Figure S6-4. All RNA sequence separated into 95% confidence interval ellipses with **2-DOS** and **Neam** removed from the PCA analysis. The centroid of each ellipse is shown with a large point. The bulge TAR (Blg TAR) and hairpin TAR (Hp TAR) were externally validated and shown on the plot above.

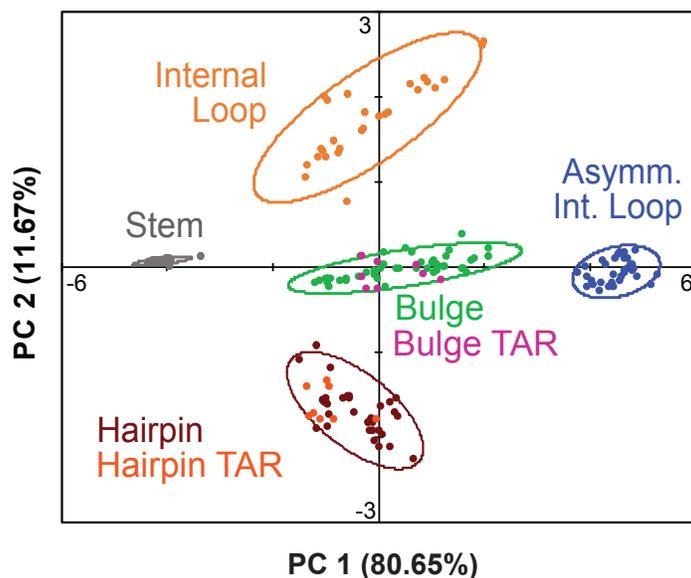


Figure S6-5. PCA plot of the RNA training set with **2-DOS** and **Neam** removed and the bulge and hairpin TAR RNA

57. Cheminformatic Analysis of Aminoglycoside Receptors

Receptors	MW	logD	HBA	HBD	RotB	tPSA	nStereo	nStereo/MW	Fsp3	relPSA	N	O	ArRings	RngLgl	SysRings	SysRR*	VWSA	logP	logS
Amik	585.61	-15.10	17	13	10	331.94	16	0.03	0.95	0.41	5	13	0	6	3	1.00	804.46	-3.22	-1.07
Apra	539.58	-12.41	16	11	6	283.64	17	0.03	1.00	0.38	5	11	0	6	3	1.33	743.33	-2.77	-1.01
2DOS	162.19	-5.34	5	5	0	112.73	5	0.03	1.00	0.48	2	3	0	6	1	1.00	234.57	-2.54	0.44
d-Strep	583.60	-12.18	19	13	9	339.59	15	0.03	0.90	0.43	7	12	0	6	3	1.00	795.74	-2.6	-0.91
G-Kana	652.67	-17.97	23	15	10	436.13	15	0.02	0.82	0.51	12	11	0	6	3	1.00	855.67	-2.88	-1.74
G-Paro	825.84	-22.03	29	18	14	539.22	19	0.02	0.82	0.50	15	14	0	6	4	1.00	1,076.74	-2.66	-1.91
Kana	484.50	-12.20	15	11	6	282.61	15	0.03	1.00	0.43	4	11	0	6	3	1.00	658.46	-3.1	-0.72
Neam	322.36	-9.69	10	8	3	203.46	10	0.03	1.00	0.45	4	6	0	6	2	1.00	450.62	-2.91	-0.5
Neom	614.65	-15.47	19	13	9	353.11	19	0.03	1.00	0.42	6	13	0	6	4	1.00	832.90	-2.81	-0.98
Siso	431.53	-11.37	11	7	6	193.49	10	0.02	0.89	0.30	5	6	0	6	3	1.00	647.01	-1.77	-1.81
Strep	581.58	-11.79	19	12	9	336.43	15	0.03	0.86	0.43	7	12	0	6	3	1.00	776.35	-2.37	-0.96

Table S7-1. The cheminformatic parameters used by Tan and co-workers were used to classify each of the small molecules.¹¹ The protonation state of each molecule was determined using Marvinview's pKa plugin; the most common species at pH 7.4 was used to determine cheminformatic parameters. Then an SDF file containing all of the small molecule structures was imported to an Instant Jchem database. 18 cheminformatic parameters were calculated using the IJchem built-in Chemical Terms calculators. The remaining two parameters, logP and logS, were calculated using the webservice AlogPS v. 2.1; since logP and logS concern neutral molecules, these values used the neutral state rather than the charged states used for the 18 IJchem parameters.

Table S7-2. Comparison of LOOCV and principal component variances depending on the removal of each aminoglycoside. Removal of a single aminoglycoside did not change the predictive power. Taking out two of **Neam**, **Neom**, or **Kana** has the only effect on the LOOCV analysis, with the predictive power lowering to 91.5-97.2%.

Removed	LOOCV	PC 1	PC 2	PC 3
None	100	81.22	12.09	2.96
2-DOS	100	80.40	12.15	3.80
Amik	100	81.78	9.93	4.11
Apra	100	80.70	11.08	3.99
d-Strep	100	80.25	11.24	3.99
G-Kana	100	82.50	10.92	2.62
G-Paro	100	82.17	12.30	2.66
Kana	100	80.22	11.27	4.02
Neam	100	80.92	10.78	4.13
Neom	100	81.69	9.99	4.10
Siso	100	80.38	11.59	3.95
Strep	100	80.22	11.28	4.03
Neom/Kana	97.2	80.77	10.31	4.40
Neam/Kana	94.3	79.98	11.11	4.46
Neam/Neom	91.5	82.38	8.95	4.56

Tanimoto Coefficients

	Amik	Apra	d-Strep	G-Kana	G-Paro	Kana	Neam	Neom	Siso	Strep
2-DOS	0.257	0.321	0.267	0.300	0.284	0.403	0.422	0.375	0.215	0.243
Amik	0.703	0.573	0.523	0.515	0.638	0.594	0.624	0.490	0.588	
Apra		0.713	0.626	0.657	0.798	0.762	0.835	0.557	0.653	
d-Strep			0.837	0.867	0.615	0.618	0.648	0.454	0.910	
G-Kana				0.947	0.744	0.692	0.705	0.469	0.763	
G-Paro					0.705	0.674	0.758	0.463	0.791	
Kana						0.926	0.931	0.545	0.561	
Neam							0.889	0.521	0.563	
Neom								0.536	0.591	
Siso									0.435	
Strep										0.435

Figure S7-1. Tanimoto coefficients calculated for the receptor library. A threshold of 0.85 was used to determine highly correlated receptors (dark green).

S8. NMR Spectra, HPLC Chromatograms, and LC-MS Analysis

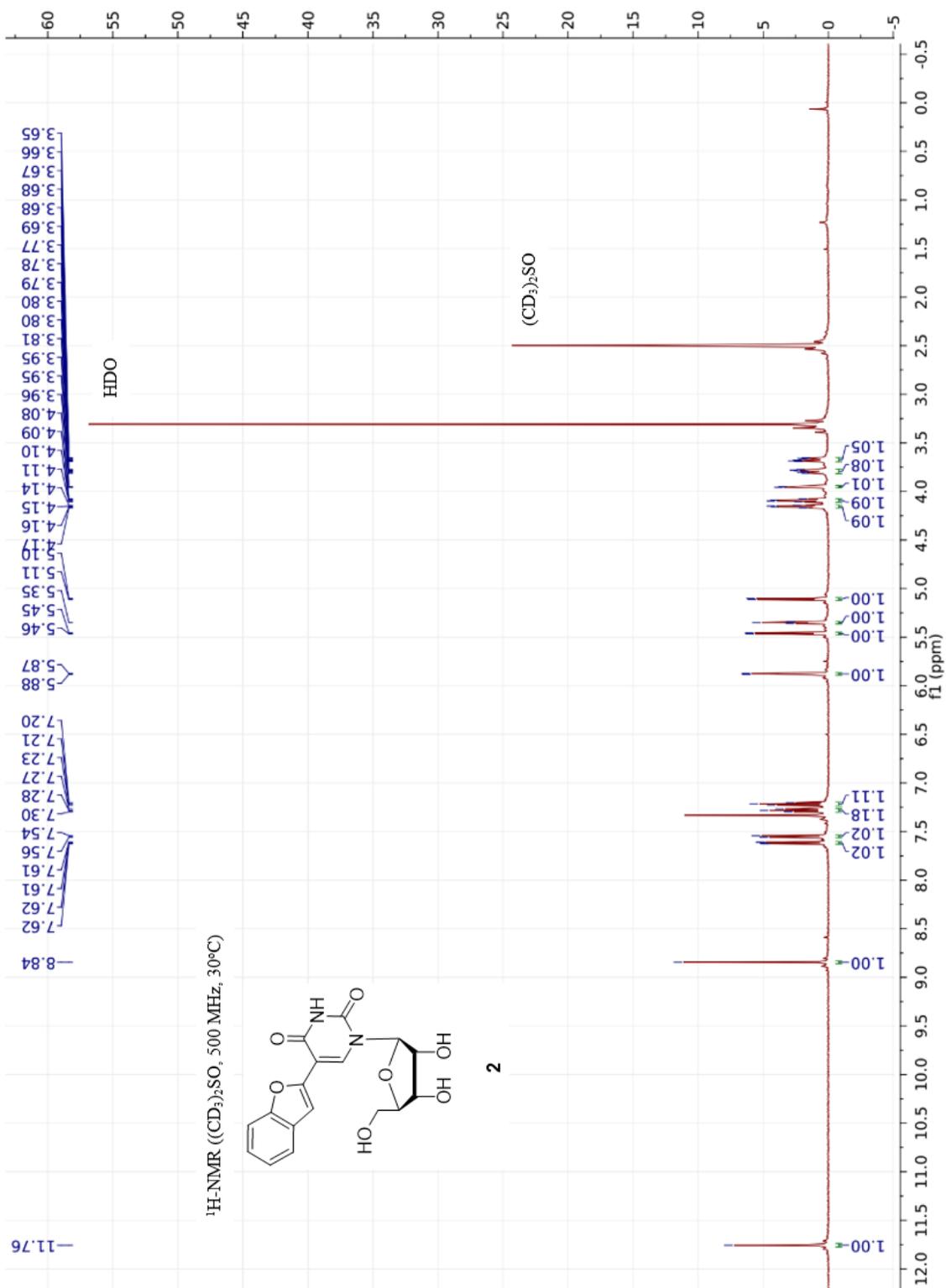


Figure S8-1. ¹H-NMR spectrum of **2**.

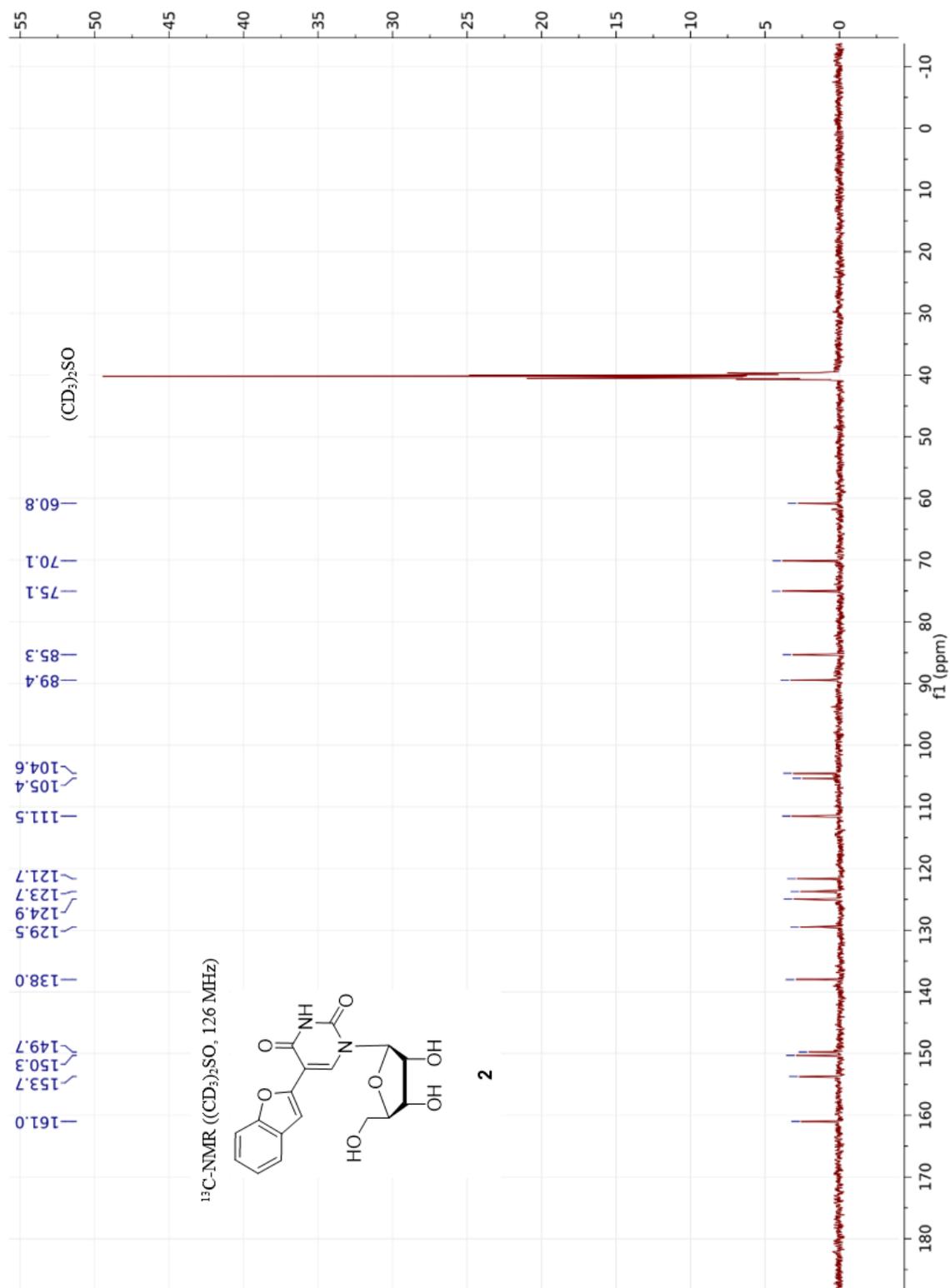


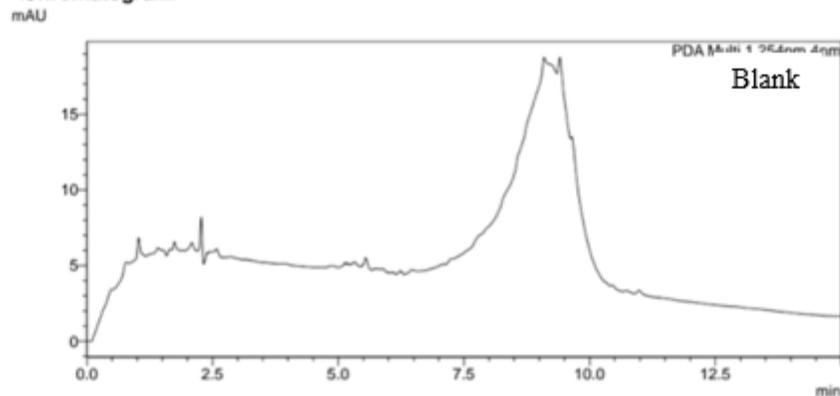
Figure S8-2. ¹³C-NMR spectrum of **2**.

SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : JEF-3-133-BFU-ACN-blank
 Sample ID : JEF-3-133-BFU-ACN-blank
 Data Filename : JEF-3-133-BFU-ACN-blank.lcd
 Method Filename : JEF-Grd95-05_Fast18min_PDA-ACN.lcm
 Batch Filename : JEF-3-133-ACN-BFU.lcb
 Vial # : 1-41
 Injection Volume : 10 uL
 Date Acquired : 3/29/2016 1:41:52 PM
 Date Processed : 3/29/2016 1:59:54 PM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>

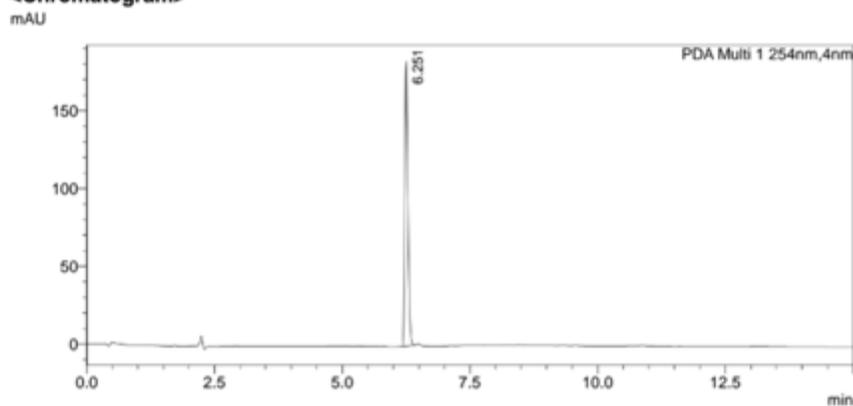


SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : JEF-3-133-BFU-ACN
 Sample ID : JEF-3-133-BFU-ACN
 Data Filename : JEF-3-133-BFU-ACN.lcd
 Method Filename : JEF-Grd95-05_Fast18min_PDA-ACN.lcm
 Batch Filename : JEF-3-133-ACN-BFU.lcb
 Vial # : 1-42
 Injection Volume : 10 uL
 Date Acquired : 3/29/2016 2:00:24 PM
 Date Processed : 3/29/2016 2:18:25 PM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>



<Peak Table>

PDA Ch1 254nm				
Peak#	Ret. Time	Area	Height	Area%
1	6.251	780567	181074	100.000
Total		780567	181074	100.000

Figure S8-3. HPLC analysis of 2.

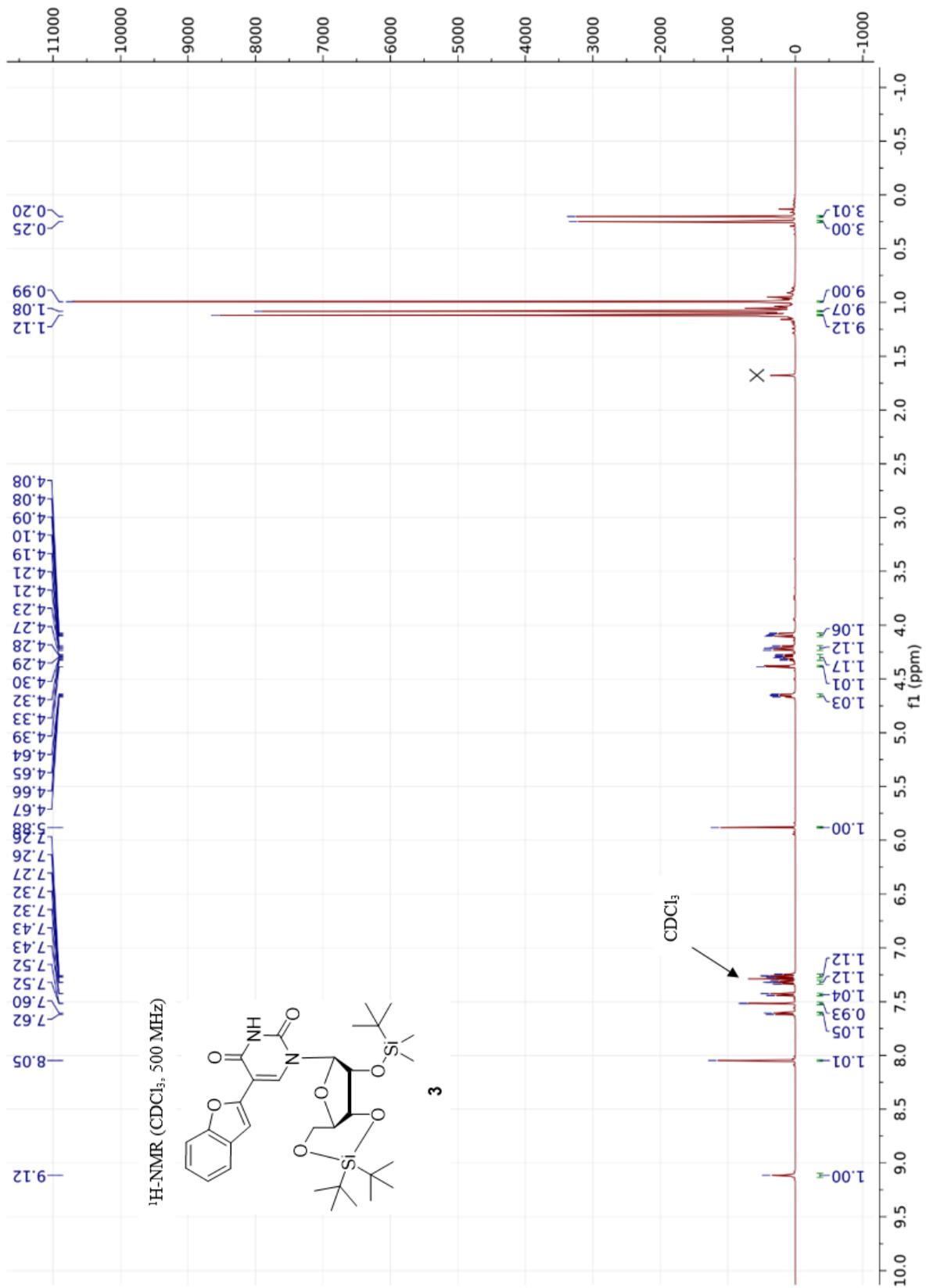


Figure S8-4. ¹H-NMR spectrum of **3**.

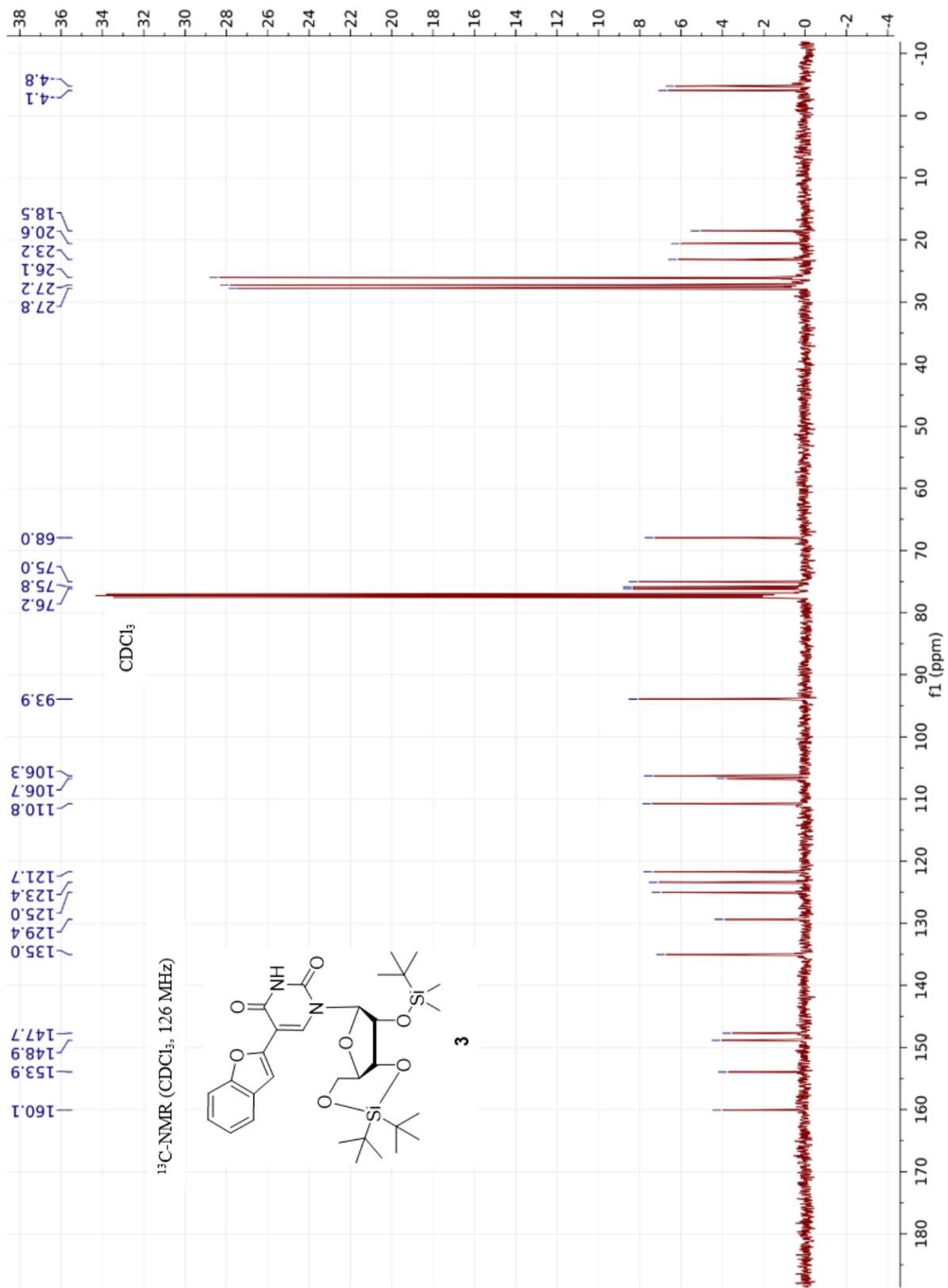


Figure S8-5. ¹³C-NMR spectrum of **3**.

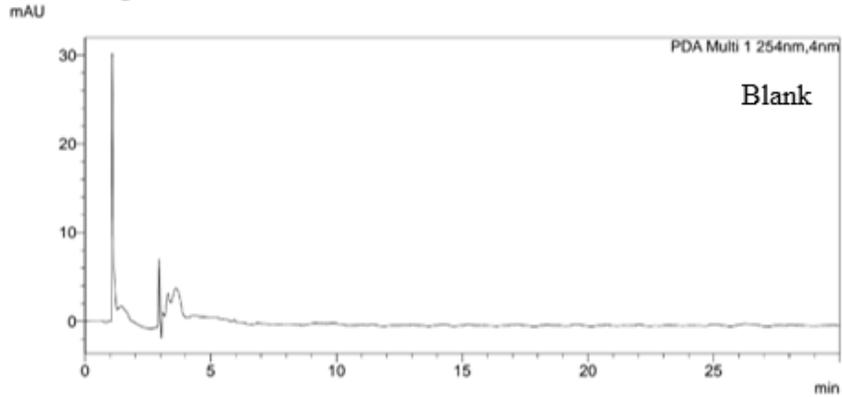


<Sample Information>

Sample Name : JEF-3-131-DTBFU-blank7
 Sample ID : JEF-3-131-DTBFU-blank7
 Data Filename : JEF-3-131-DTBFU-blank7.lcd
 Method Filename : JEF-Grd75-95_fast30min_PDA-ACN.lcm
 Batch Filename : JEF-3-131-DTBFU-4.lcb
 Vial # : 1-41
 Injection Volume : 10 uL
 Date Acquired : 3/31/2016 11:56:41 AM
 Date Processed : 3/31/2016 12:26:44 PM

Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>

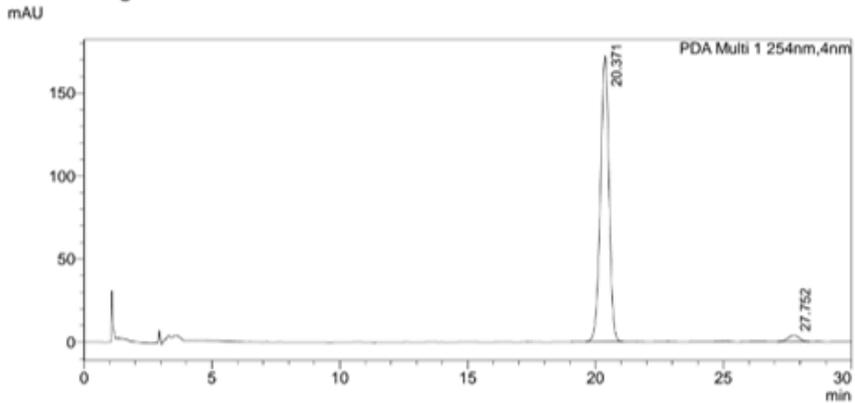


<Sample Information>

Sample Name : JEF-3-131-DTBFU-6
 Sample ID : JEF-3-131-DTBFU-6
 Data Filename : JEF-3-131-DTBFU-7.lcd
 Method Filename : JEF-Grd75-95_fast30min_PDA-ACN.lcm
 Batch Filename : JEF-3-131-DTBFU-4.lcb
 Vial # : 1-42
 Injection Volume : 10 uL
 Date Acquired : 3/31/2016 12:27:13 PM
 Date Processed : 3/31/2016 1:02:22 PM

Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

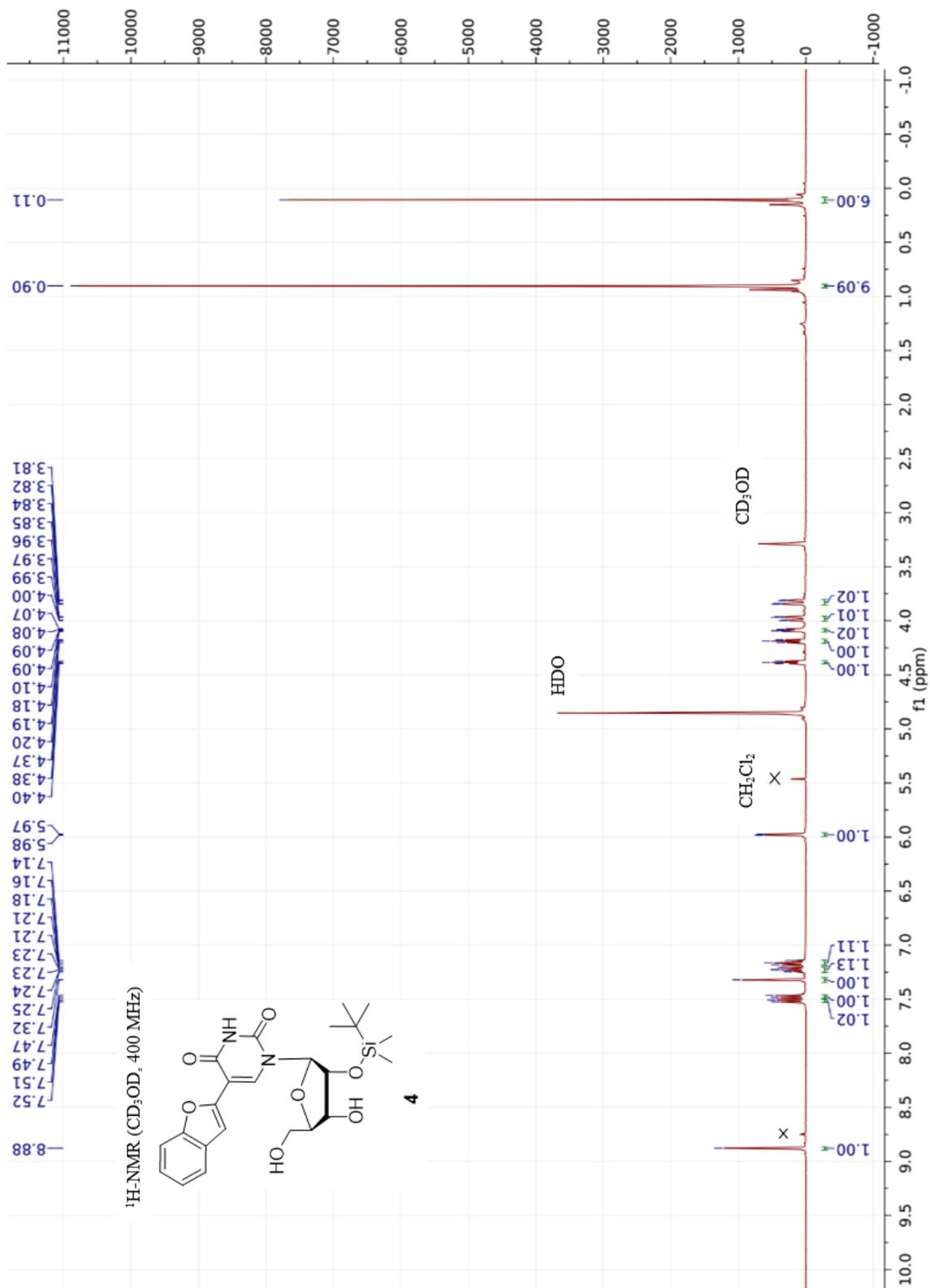
<Chromatogram>



<Peak Table>

Peak#	Ret. Time	Area	Height	Area%
1	20.371	4032938	172067	97.151
2	27.752	118253	3722	2.849
Total		4151191	175788	100.000

Figure S8-6. HPLC analysis of 3.



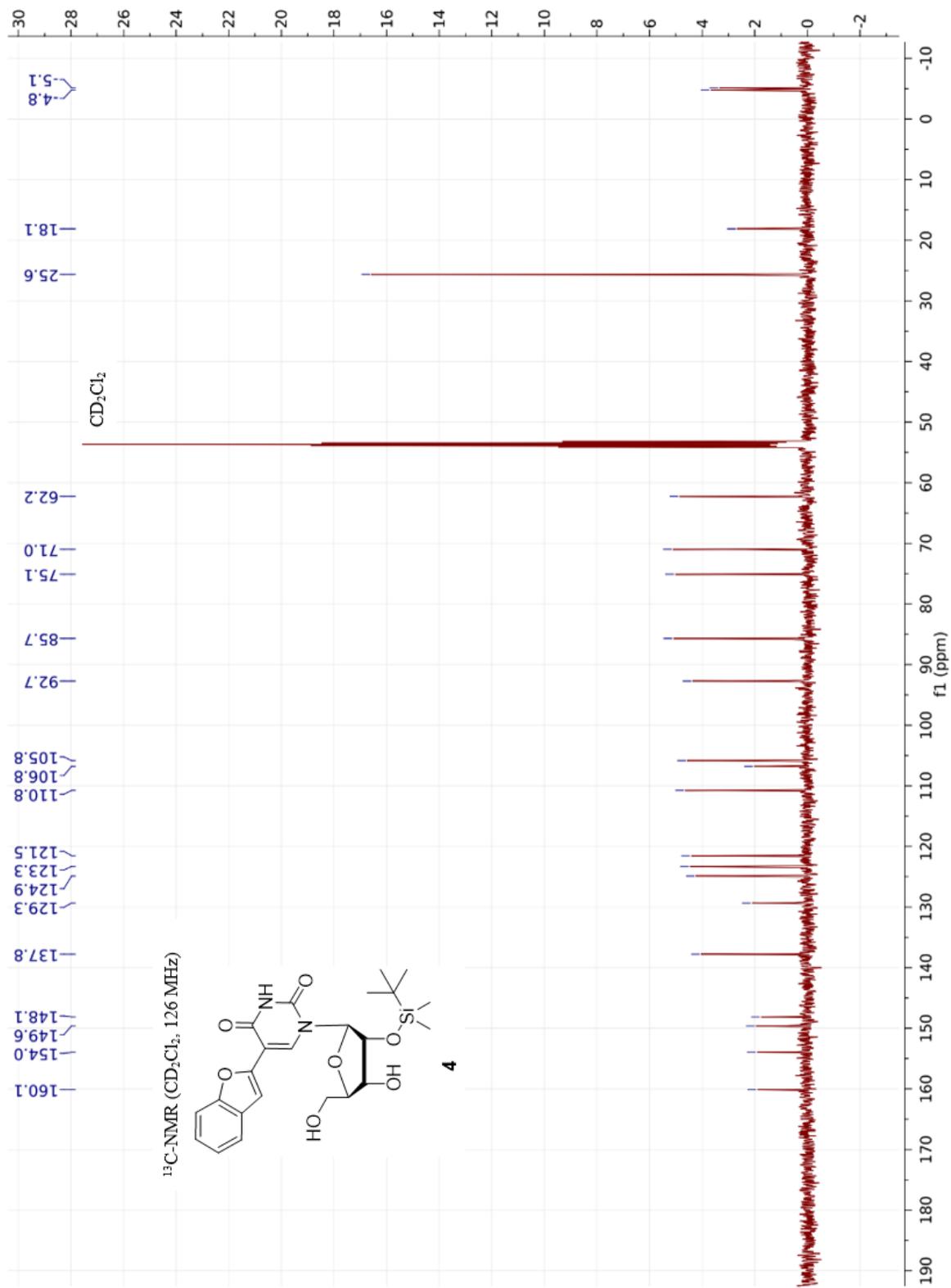


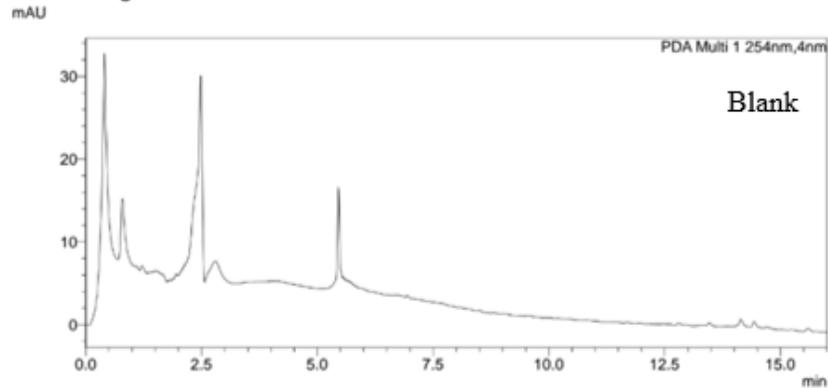
Figure S8-8. ¹³C-NMR spectrum of **4**.

SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : JEF-3-134-f34-41-blank
 Sample ID : JEF-3-134-f34-41-blank
 Data Filename : JEF-3-134-f34-41-blank.lcd
 Method Filename : JEF-Grd95-05_med16min_PDA-ACN.lcm
 Batch Filename : JEF-3-134-fraction-tests.lcb
 Vial # : 1-41
 Injection Volume : 10 uL
 Date Acquired : 3/30/2016 9:08:30 AM
 Date Processed : 3/30/2016 9:24:32 AM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>

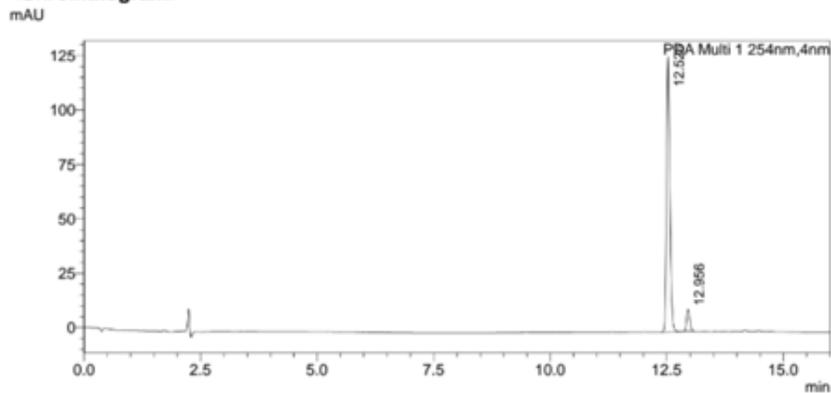


SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : JEF-3-134-f34-41
 Sample ID : JEF-3-134-f34-41
 Data Filename : JEF-3-134-f34-41.lcd
 Method Filename : JEF-Grd95-05_med16min_PDA-ACN.lcm
 Batch Filename : JEF-3-134-fraction-tests.lcb
 Vial # : 1-42
 Injection Volume : 10 uL
 Date Acquired : 3/30/2016 9:25:03 AM
 Date Processed : 3/30/2016 9:41:06 AM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>



<Peak Table>

Peak#	Ret. Time	Area	Height	Area%
1	12.526	638182	124395	93.099
2	12.956	47303	9807	6.901
Total		685485	134202	100.000

Figure S8-9. HPLC analysis of 4.

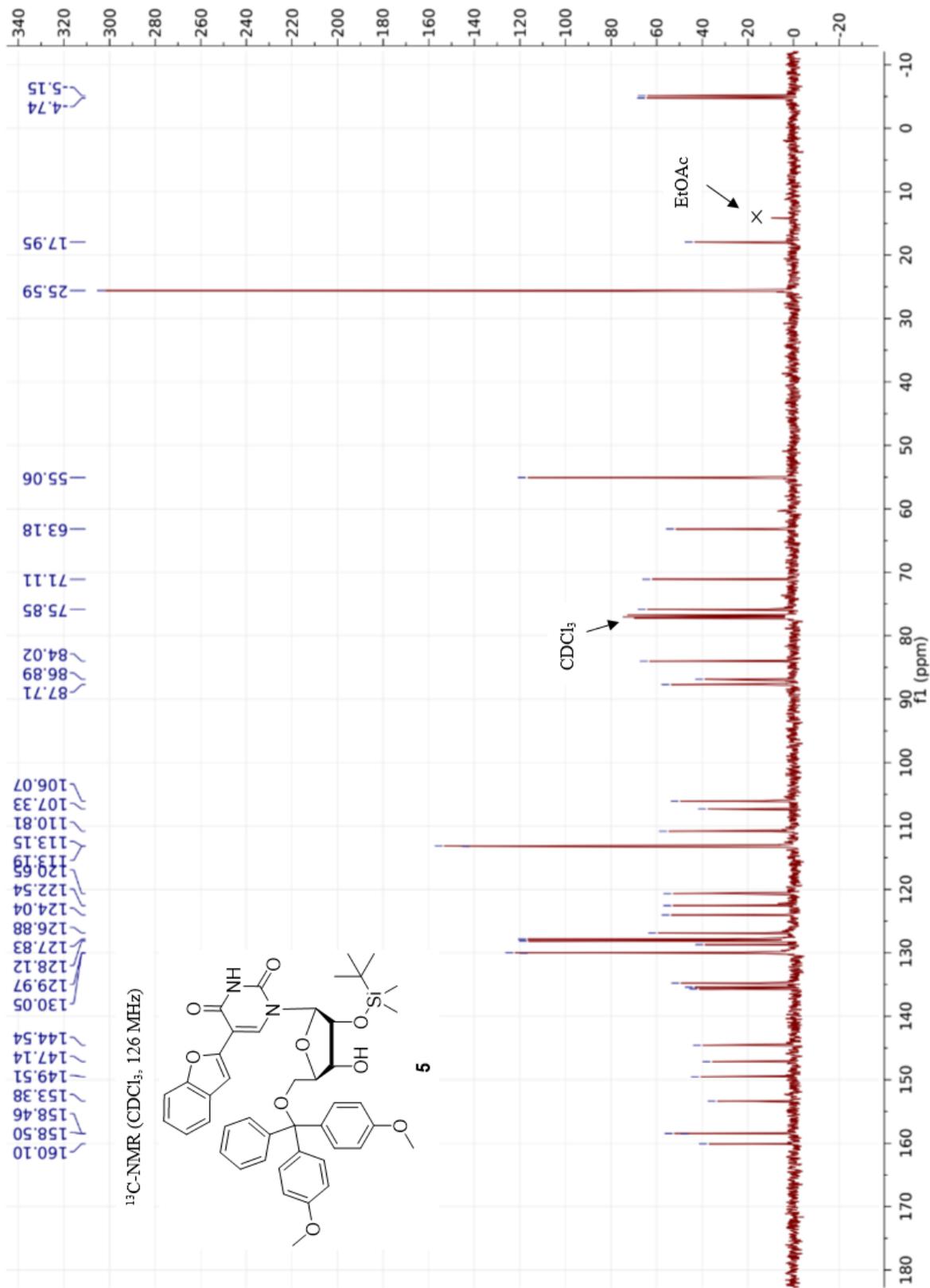
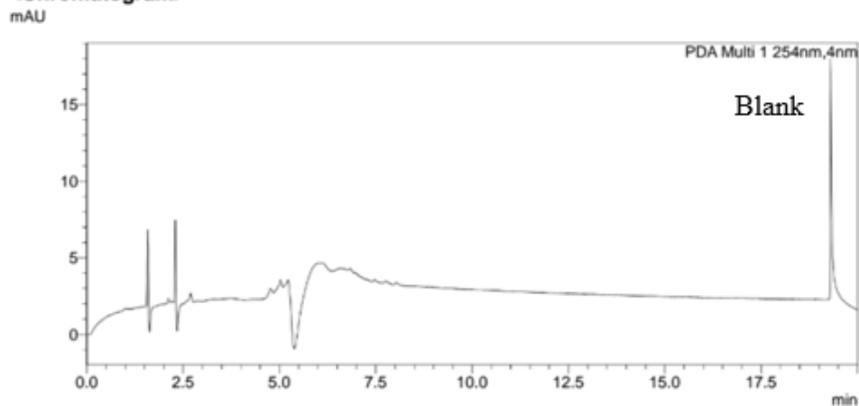


Figure S8-11. ¹³C-NMR spectrum of **5**.

<Sample Information>

Sample Name : JEF-3-138-DMT-BFU-blank5
 Sample ID : JEF-3-138-DMT-BFU-blank5
 Data Filename : JEF-3-138-DMT-BFU-blank5.lcd
 Method Filename : JEF-Grd95-05_Fast20min_PDA-ACN.lcm
 Batch Filename : JEF-3-138-DMT-BFU-5.lcb
 Vial # : 1-22
 Injection Volume : 10 uL
 Date Acquired : 4/5/2016 10:09:26 AM
 Date Processed : 4/5/2016 10:29:29 AM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

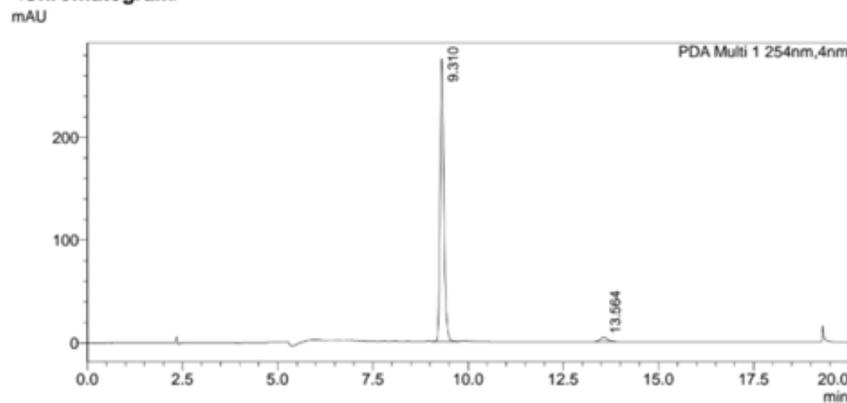
<Chromatogram>



<Sample Information>

Sample Name : JEF-3-138-DMT-BFU-5
 Sample ID : JEF-3-138-DMT-BFU-5
 Data Filename : JEF-3-138-DMT-BFU-5.lcd
 Method Filename : JEF-Grd95-05_Fast20min_PDA-ACN.lcm
 Batch Filename : JEF-3-138-DMT-BFU-5.lcb
 Vial # : 1-23
 Injection Volume : 10 uL
 Date Acquired : 4/5/2016 10:29:59 AM
 Date Processed : 4/5/2016 10:50:02 AM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>



<Peak Table>

Peak#	Ret. Time	Area	Height	Area%
1	9.310	1889628	272745	97.290
2	13.564	52641	4070	2.710
Total		1942269	276815	100.000

Figure S8-12. HPLC analysis of 5.

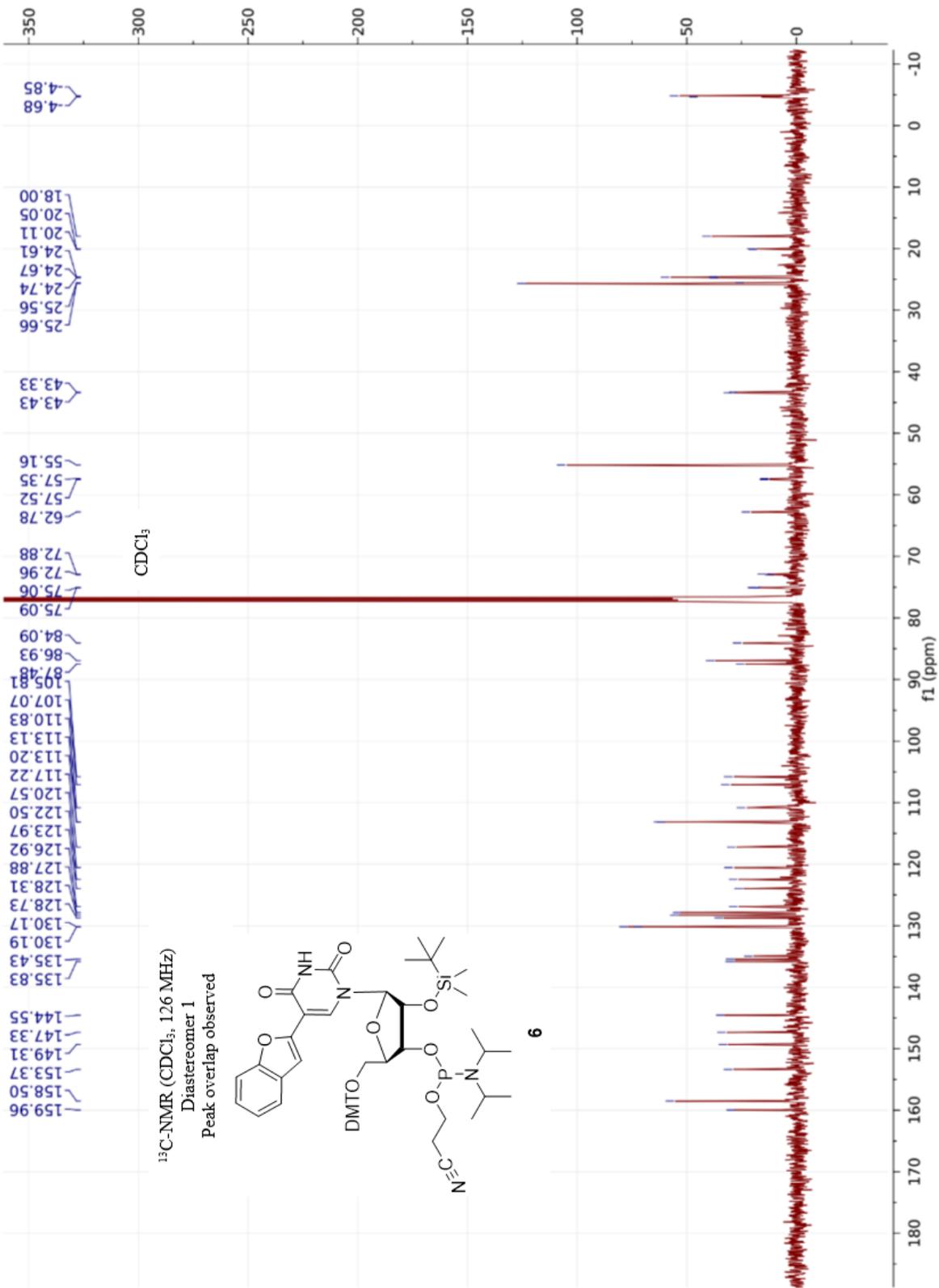
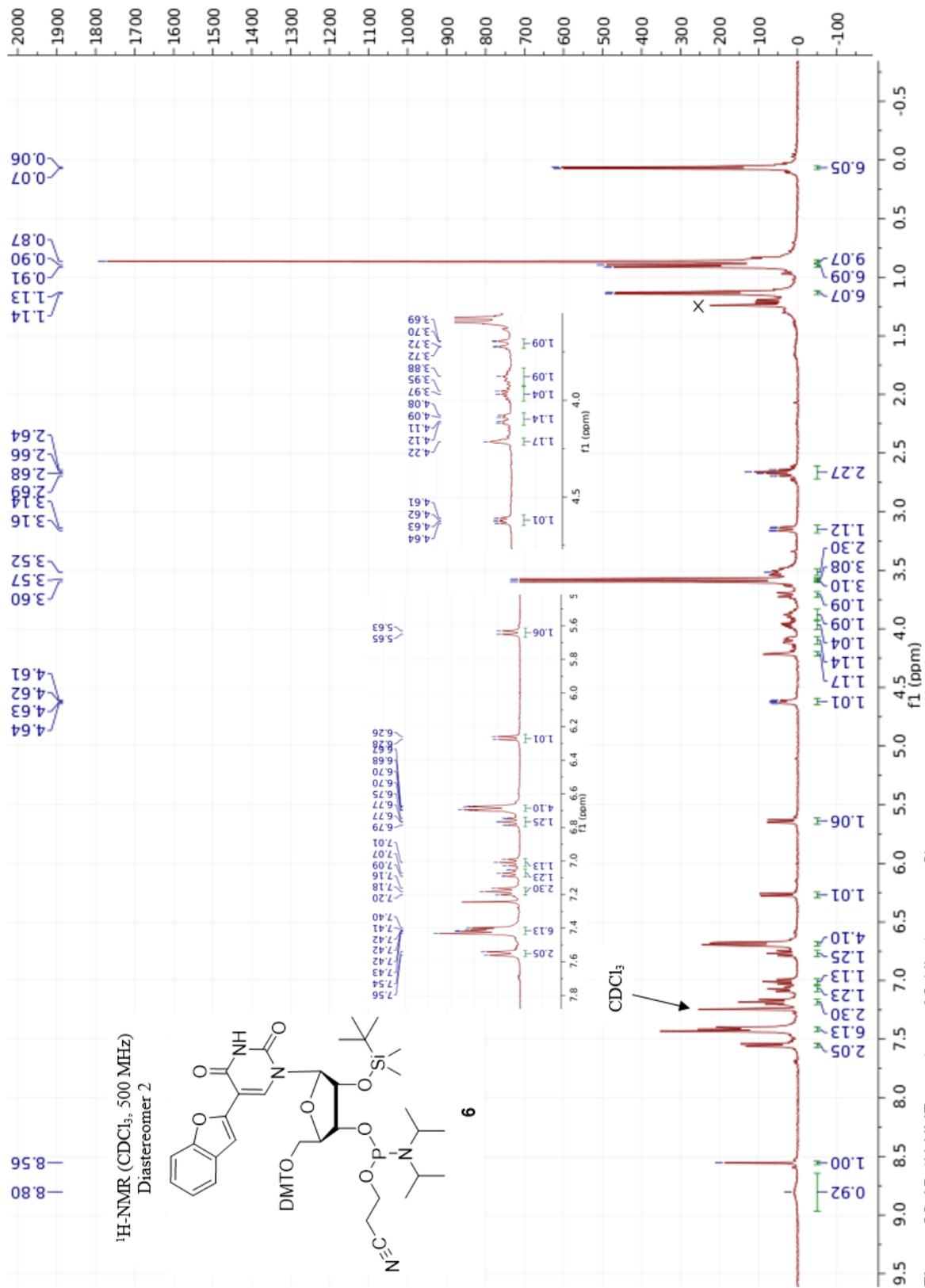


Figure S8-14. ¹³C-NMR spectrum of **6** (diastereomer 1). Peak overlap is observed.



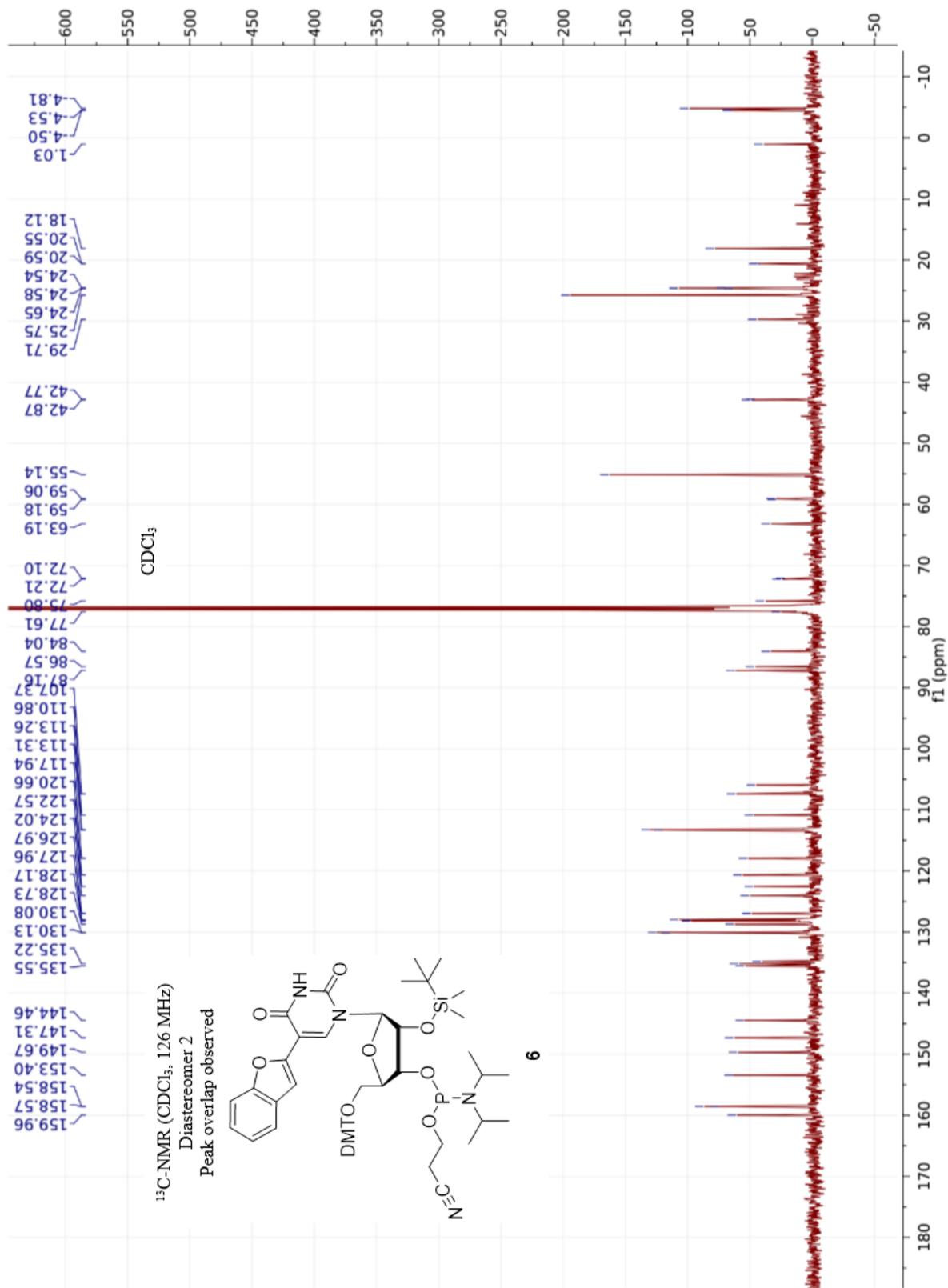


Figure S8-16. ¹³C-NMR spectrum of **6** (diastereomer 2). Peak overlap is observed.

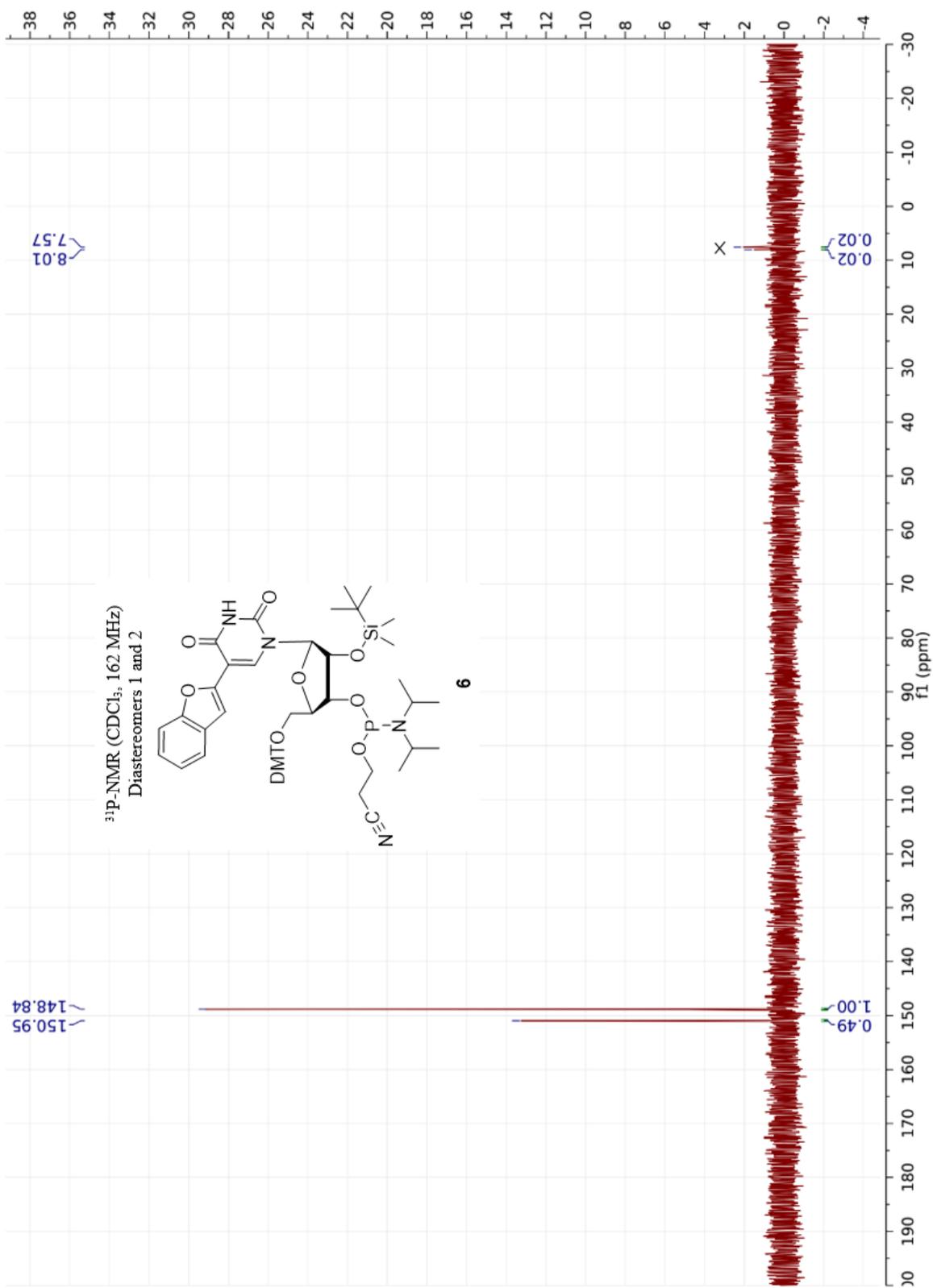


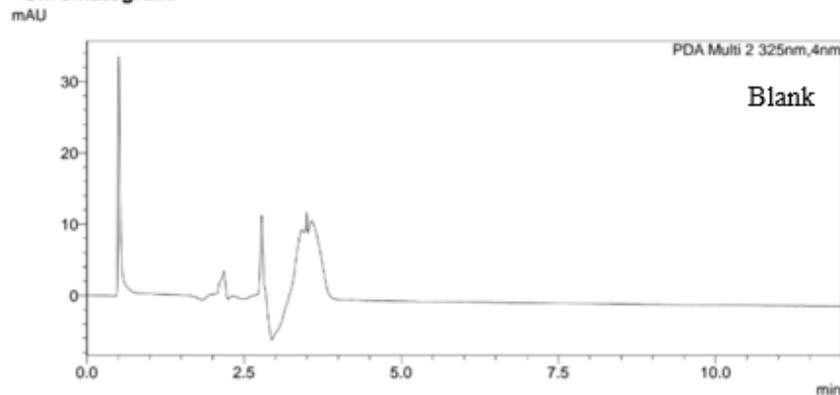
Figure S8-17. ³¹P-NMR spectrum of **6** containing both diastereomer 1 (150.95 ppm) and diastereomer 2 (148.84 ppm). Slight H-phosphonate impurity is observed at 8.01 ppm and 7.57 ppm.

SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : JEF-3-175-PhosBFU-Dias1-blank
 Sample ID : JEF-3-175-PhosBFU-Dias1-blank
 Data Filename : JEF-3-175-PhosBFU-Dias1-blank.lcd
 Method Filename : JEF-ISO-100-ACN_12min_PDA.lcm
 Batch Filename : JEF-3-175-PhosBFU-Dias1.lcb
 Vial # : 1-61
 Injection Volume : 10 uL
 Date Acquired : 7/13/2016 10:26:34 AM
 Date Processed : 7/13/2016 10:38:37 AM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>

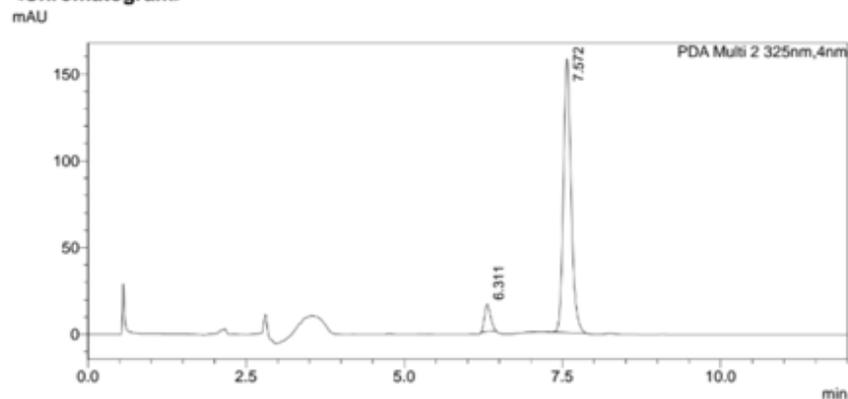


SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : JEF-3-175-F81
 Sample ID : JEF-3-175-F81
 Data Filename : JEF-3-175-F81.lcd
 Method Filename : JEF-ISO-100-ACN_12min_PDA.lcm
 Batch Filename : JEF-3-175.lcb
 Vial # : 1-62
 Injection Volume : 10 uL
 Date Acquired : 7/12/2016 2:10:42 PM
 Date Processed : 7/12/2016 3:39:42 PM
 Sample Type : Unknown
 Acquired by : chemist
 Processed by : chemist

<Chromatogram>



<Peak Table>

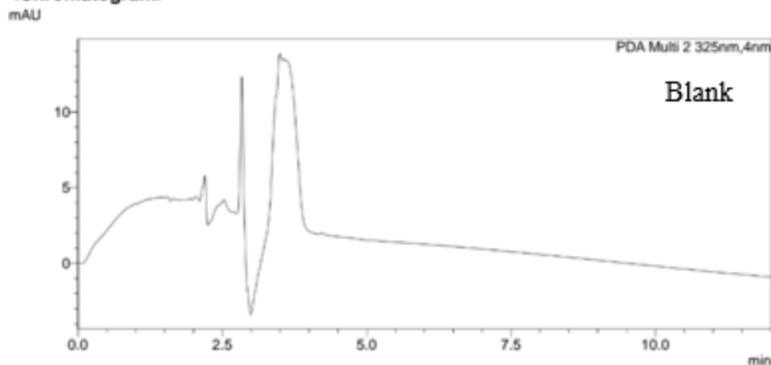
Peak#	Ret. Time	Area	Height	Area%
1	6.311	103777	15749	7.206
2	7.572	1336371	157352	92.794
Total		1440149	173101	100.000

Figure S8-18. HPLC analysis of 6, diastereomer 1.

<Sample Information>

Sample Name : JEF-3-141-PhosBFU-blank-1
 Sample ID : JEF-3-141-PhosBFU-blank-1
 Data Filename : JEF-3-141-PhosBFU-blank-1.lcd
 Method Filename : JEF-ISO-100-ACN_12min_PDA.lcm
 Batch Filename : JEF-3-141-PhosBFU-Diastereomers.lcb
 Vial # : 1-10 Sample Type : Unknown
 Injection Volume : 10 uL
 Date Acquired : 4/13/2016 1:18:48 PM Acquired by : chemist
 Date Processed : 4/13/2016 1:30:51 PM Processed by : chemist

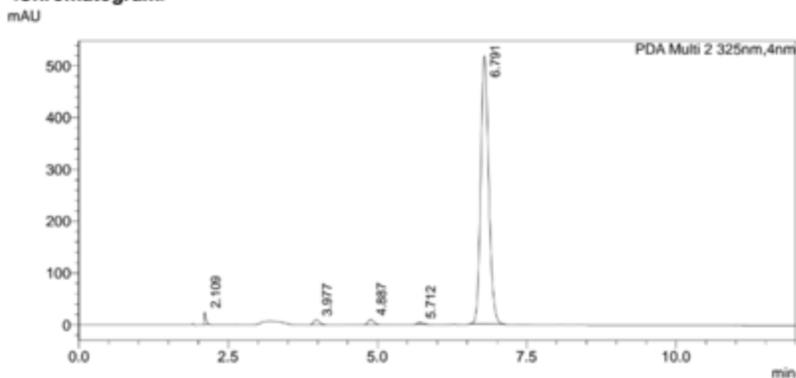
<Chromatogram>



<Sample Information>

Sample Name : JEF-3-141-PhosBFU-Dias2-F108-112
 Sample ID : JEF-3-141-PhosBFU-Dias2-F108-11
 Data Filename : JEF-3-141-PhosBFU-Dias2-F108-112.lcd
 Method Filename : JEF-ISO-100-ACN_12min_PDA.lcm
 Batch Filename : JEF-3-141-PhosBFU-Diastereomers.lcb
 Vial # : 1-12 Sample Type : Unknown
 Injection Volume : 10 uL
 Date Acquired : 4/13/2016 1:43:53 PM Acquired by : chemist
 Date Processed : 5/17/2016 9:17:08 AM Processed by : chemist

<Chromatogram>



<Peak Table>

Peak#	Ret. Time	Area	Area%
1	2.109	45107	0.897
2	3.977	65315	1.299
3	4.887	65079	1.295
4	5.712	19089	0.380
5	6.791	4831767	96.129
Total		5026356	100.000

Figure S8-19. HPLC analysis of 6, diastereomer 2.

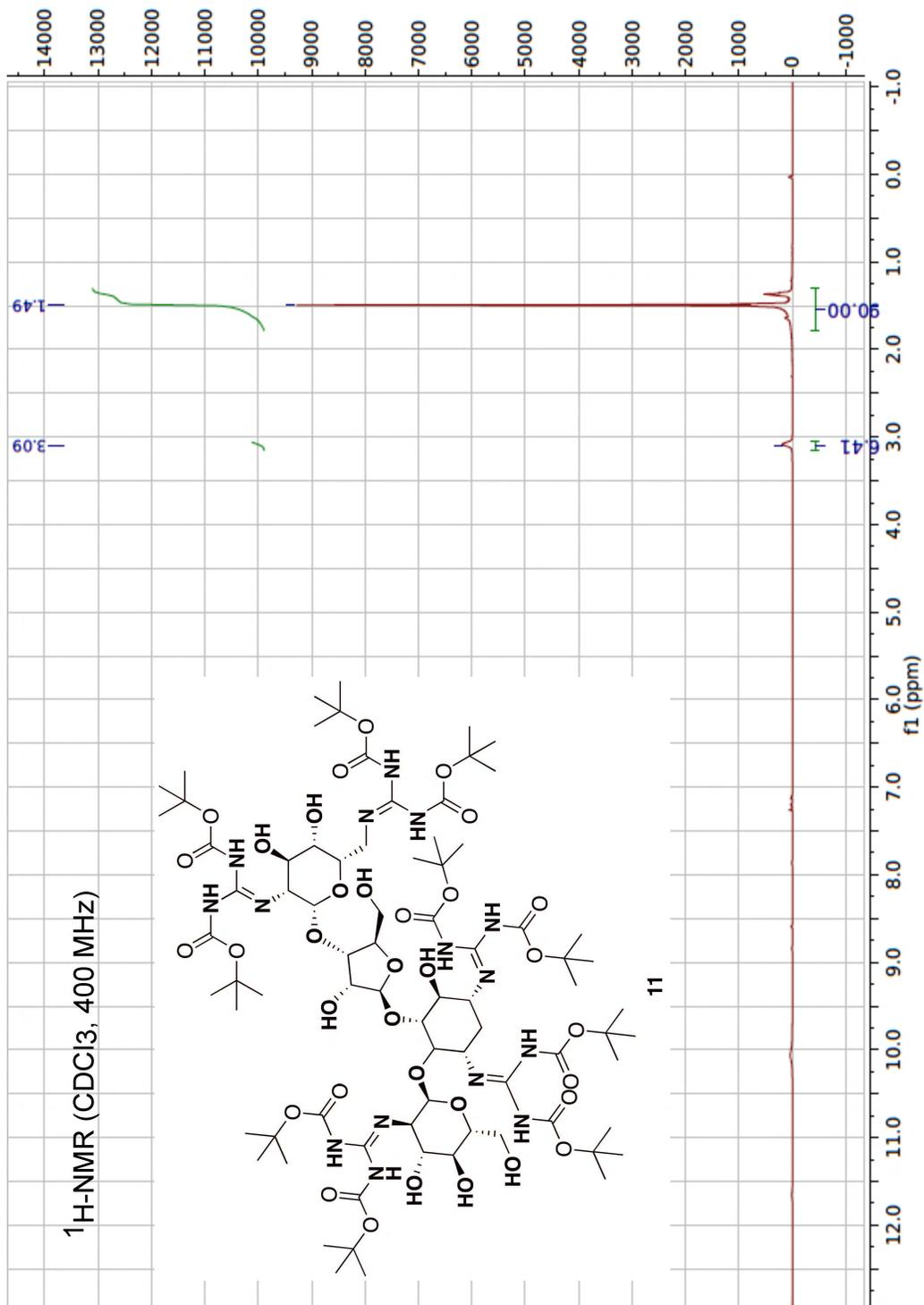


Figure S8-20. ¹H-NMR spectrum of 11.

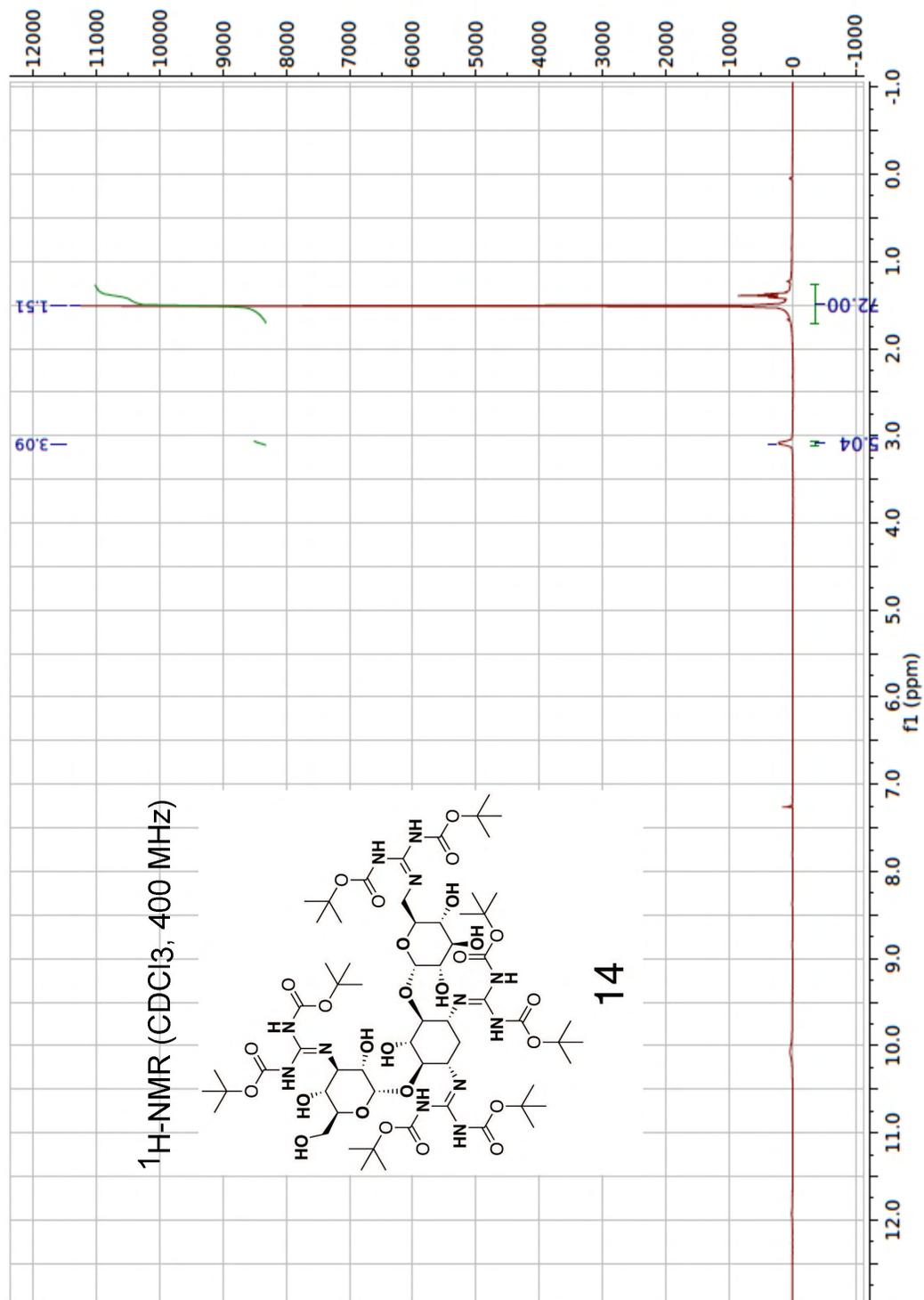


Figure S8-21. ¹H-NMR spectrum of **14**.

Analysis Name: 2016-07-20-0001.D
Method: 1-5-100-20MIN.M
Sample Name: MeOH Blank
Analysis Info: sample in meoh

Print Date: 7/21/2016 10:44:47 AM
Acq. Date: 7/20/2016 12:37:55 PM

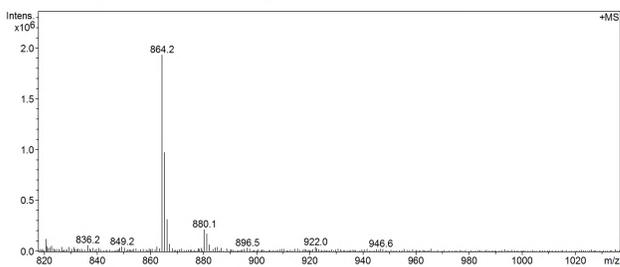
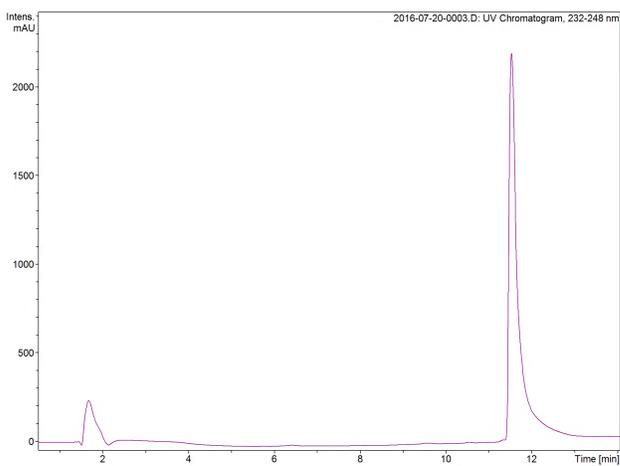
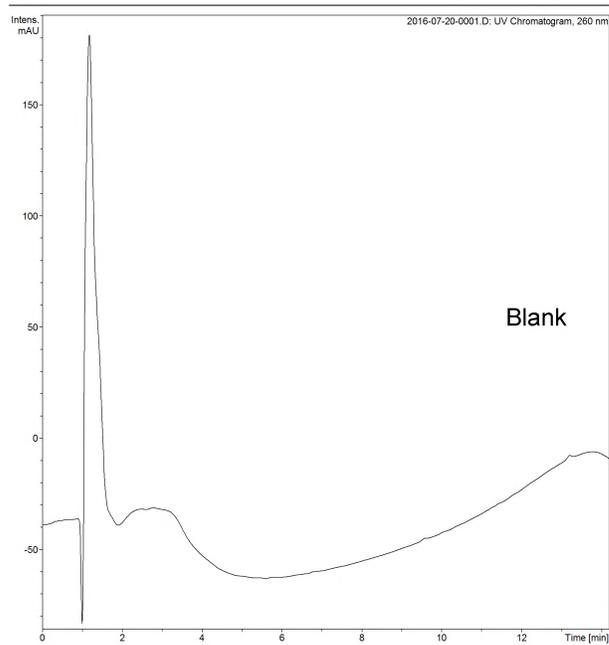


Figure S8-22. LC-MS analysis of 12.

Analysis Name: 2016-07-20-0001.D **Print Date:** 7/21/2016 10:44:47 AM
Method: 1-5-100-20MIN.M **Acq. Date:** 7/20/2016 12:37:55 PM
Sample Name: MeOH Blank
Analysis Info: sample in meoh

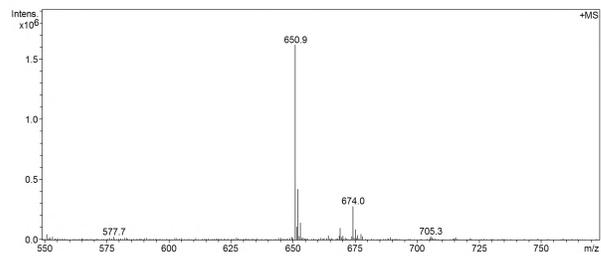
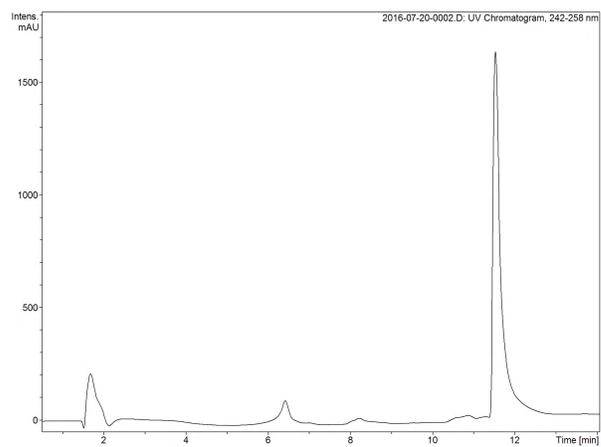
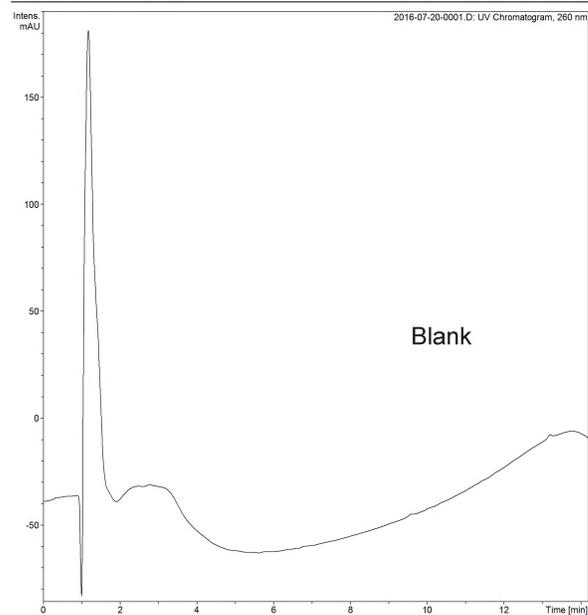


Figure S8-23. LC-MS analysis of 15.

References

- (1) Gallagher-Duval, S.; Herve, G.; Sartori, G.; Enderlin, G.; Len, C. *New J. Chem.* **2013**, 37, 1989.
- (2) Ghanty, U.; Fostvedt, E.; Valenzuela, R.; Beal, P. A.; Burrows, C. J. *J. Am. Chem. Soc.* **2012**, 134, 17643.
- (3) Efthymiou, T.; Krishnamurthy, R. In *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: 2001.
- (4) Meher, G.; Efthymiou, T.; Stoop, M.; Krishnamurthy, R. *Chem. Commun. (London)* **2014**, 50, 7463.
- (5) Luedtke, N. W.; Baker, T. J.; Goodman, M.; Tor, Y. *J. Am. Chem. Soc.* **2000**, 122, 12035.
- (6) Reuter, J. S.; Mathews, D. H. *BMC Bioinf.* **2010**, 11, 129.
- (7) Parisien, M.; Major, F. *Nature* **2008**, 452, 51.
- (8) Das, R.; Karanicolas, J.; Baker, D. *Nat. Methods* **2010**, 7, 291.
- (9) Lyskov, S.; Chou, F.-C.; Conchir, S.; Der, B. S.; Drew, K.; Kuroda, D.; Xu, J.; Weitzner, B. D.; Renfrew, P. D.; Sripakdeevong, P.; Borgo, B.; Havranek, J. J.; Kuhlman, B.; Kortemme, T.; Bonneau, R.; Gray, J. J.; Das, R. *PLoS ONE* **2013**, 8, e63906.
- (10) Chen, V. B.; Davis, I. W.; Richardson, D. C. *Protein Sci.* **2009**, 18, 2403.
- (11) Wenderski, T. A.; Stratton, C. F.; Bauer, R. A.; Kopp, F.; Tan, D. S. *Methods Mol. Biol. (N. Y.)* **2015**, 1263, 225.