SUPPORTING INFORMATION

Supporting Information

69451 Weinheim, Germany

Repurposing antiviral drugs for orthogonal RNA-catalyzed labeling of RNA

Mohammad Ghaem Maghami, Surjendu Dey, Ann-Kathrin Lenz, Claudia Höbartner*

Abstract: *In vitro* selected ribozymes are promising tools for site-specific labeling of RNA. Previously known nucleic acid catalysts attached fluorescently labeled adenosine or guanosine derivatives via 2',5'-branched phosphodiester bonds to the RNA of interest. Here we report new ribozymes that use orthogonal substrates, derived from the antiviral drug tenofovir, and attach bioorthogonal functional groups, as well as affinity handles and fluorescent reporter units via a hydrolytically more stable phosphonate ester linkage. The tenofovir transferase ribozymes were identified by in vitro selection and are orthogonal to nucleotide transferase ribozymes. As genetically encodable functional RNAs, these ribozymes may be developed for potential cellular applications. Here, the orthogonal ribozymes addressed desired target sites in large RNAs *in vitro*, as shown by fluorescent labeling of *E.coli* 16S and 23S rRNAs in total cellular RNA.

DOI: 10.1002/anie.2020XXXXX

SUPPORTING INFORMATION

Table of Contents

SUPPORTING INFORMATION

1. Materials and DNA/RNA oligonucleotides

1.1. General materials

Unmodified oligonucleotides such as primers and transcription templates were ordered from Microsynth and subjected to PAGE purification as in our previous report.[1]

Dynabeads streptavidin T1 and speedbead neutravidin coated-magnetic particles were purchased from Thermo Fisher scientific and GE-healthcare respectively. (Diethoxyphosphoryl)methyl 4-methylbenzenesulfonate, 6-FAM azide and Sulfo-Cy5 azide were purchased from Fluorochem and Jena Bioscience, respectively. All other chemicals were purchased from Sigma-Aldrich or ABCR and used without further purification. HPLC grade solvents were purchased from VWR.

Silica gel plates coated with fluorescent indicator were used for thin layer chromatography (TLC) and the plates were visualized with UV light. Silica gel 60, 0.032-0.063 mm (230-450 mesh) was used for column chromatography.

NMR spectra (^{1}H, ^{13}C and ^{31}P) were recorded using 400 MHz Bruker Avance III and Avance III HD spectrometers. Chemical shifts (^{1}H and ^{13}C) were reported in parts per million (ppm) relative to TMS (δ = 0.00 ppm) and were referenced to residual protium in the solvent. Coupling constants (*J*) are reported in Hz with the following multiplicity designations: s (singlet), d (doublet), t (triplet), q (quartet), doublet of doublet (dd), m (multiplet), and br (broad).

High-resolution ESI mass spectra in positive or negative ion mode were acquired on a Bruker micrOTOF-Q III.

Anion exchange chromatography was performed on a GE Healthcare ÄKTAprime plus system, on DEAE Sephadex A-25 (GE-Healthcare), self-packed 3x20 cm column. *Detection wavelength:* 280 nm, *Solvent systems:* buffer A: 100 mM TEAB (pH = 7.5); buffer B: 1.2 M TEAB (pH = 7.5), *Flow rate:* 6 ml/min, *Gradient:* 0 –100% buffer B in 200 min.

Fluorescent imaging of the kinetic and activity assay gels were taken using a BioRad Chemidoc gel-documentation device.

1.2. DNA oligonucleotides: Primers and transcription templates

Description	5'-Sequence-3'
T7 Promoter	CTGTAATACGACTCACTATA
Forward primer/2nd PCR fwd primer	TTGAAGGCTCAGTATGTCCTATAGTGAGTCGTATTACA
Pool reverse primer	GGTAAGGTGGACATACTG-N40-GCCTTCAAGGATGGTAGGCTGG
Selection reverse primer	GGTAAGGTGGACATACTG
1st PCR forward primer	CTTCAACCAGCCTACCATCC
Cloning forward primer	TAAATAAAATAACTGTAATACGACTCACTATAGGACATACTGAGC
Fwd primer for substrate sequence deletion	CTGTAATACGACTCACTATAGGCCAGCCTACCATCC
E. coli 5S rRNA forward primer	CTGTAATACGACTCACTATAGGTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCATGCCG AACTCAGAAGTGAA
E. coli 5S rRNA reverse primer	TGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCCACACTACCATCGGCGCTACGGCGTTTCAC TTCTGAGTTCGG
1st PCR NGS N8 UMI forward primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNAACCAGCCTACCATCC
1st PCR NGS reverse primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAGGTCGACATACTG
Indexed forward primer I5_S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
Indexed reverse primer I7_N703 (Round 12)	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG
Indexed reverse primer I7_N704 (Round 7)	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG
RT primer 16S 344-361 (A325)	CCCACTGCTGCCTCCCGT
RT primer 16S 413-430 (A383)	TACAACCCGAAGGCCTTC
RT primer 16S 687-704 (A649)	TCTACGCATTTCACCGCT
RT primer 23S 322-339 (A272)	ACGGGGCTGTCACCCTGT
RT primer 23S 653-672 (A637)	GGTCTATACCCTGCAACT
RT primer 23S 1599-1618 (A1572)	TGTGTCGGTTTGGGGTACGA

ssDNA templates for	5'-Sequence-3'
Parent substrate sequence	TTGAAGGCTCAGTATGTCCTATAGTGAGTCGTATTACAG
TM	CCAGGAACTCGACGCACCCTATAGTGAGTCGTATTACAG
TV1	AACTTCCCTCTCATACACCTATAGTGAGTCGTATTACAG
TV2	GGTCCTTCTCCTGCGTGCCTATAGTGAGTCGTATTACAG
Parent-AAG	TTGAAGGCTTAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -UAG	TTGAAGGCTAAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -CAG	TTGAAGGCTGAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -GAA	TTGAAGGTTCAGTATGTCCTATAGTGAGTCGTATTACAG

Parent -GAU	TTGAAGGATCAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -GAC	TTGAAGGGTCAGTATGTCCTATAGTGAGTCGTATTACAG
TM-UGAGC	CCAGGAGCTCAACGCACCCTATAGTGAGTCGTATTACAG
5'TV1-3'parent	TTGAAGGCTCTCATACACCTATAGTGAGTCGTATTACAG
TV1-GAGC	AACTTCGCTCTCATACACCTATAGTGAGTCGTATTACAG
RNA w two labeling sites (DLRNA)	CATTAACTCGACGCATCCTTAATTGAAGGCTCAGTATGTCCTATAGTGAGTCGTATTACAG

ssDNA templates for	5'-Sequence-3'
FJ1	GGACATACTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ8	GGACATACTGGGGGAAGGTAATGTCCTAAGGTTTCGTTTTGTAACACCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC1	GGACATACTGGGAACTTGCTAGTTTACCGTTTTCTTATGTGACACGTTTCGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC3	GGACATACTGGAAATTGGAGGCGGCAATTATACGGTTCGGTGGCATTCGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ1_TM	GGGTGCGTCGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGTTCCTGGCCTATAGTGAGTCGTATTACAG
FJ1_TV1	GGTGTATGAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGGGAAGTTCCTATAGTGAGTCGTATTACAG
FJ1_TV2	GGCACGCAGGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGAAGGACCCCTATAGTGAGTCGTATTACAG
FJ1_TM-UGAGC	GGGTGCGTTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTCCTGGCCTATAGTGAGTCGTATTACAG
FJ1_5'TV1-3'Par	GGTGTATGAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ1_TV1-GAGC	GGTGTATGAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCGAAGTTCCTATAGTGAGTCGTATTACAG
FJ1_ AAG	GGACATACTATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ1_ GAA	GGACATACTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGACCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_TM	GGGTGCGTCGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGTTCCTGGCCTATAGTGAGTCGTATTACAG
FJC9_TV1	GGTGTATGAGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGGGAAGTTCCTATAGTGAGTCGTATTACAG
FJC9_TV2	GGCACGCAGGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGAAGGACCCCTATAGTGAGTCGTATTACAG
FJC9_ AAG	GGACATACTAGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_ UAG	GGACATACTTGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_ CAG	GGACATACTCGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_ GAA	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTACCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_ GAU	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTTCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_ GAC	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTCCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_5S-A99	TCCCCATGCGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGAGTAGGGAACCTATAGTGAGTCGTATTAC
FJ1_23S-A272	AGGAGCCCAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTGAATCCTATAGTGAGTCGTATTACAG
FJ1_23S-A362	ACATGCTGTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTCGATGCCTATAGTGAGTCGTATTACAG
FJ1_23S-A1572	TTCCAGGAAATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTCTAACCTATAGTGAGTCGTATTACAG
FJ1_16S-A383	AATGGGCGCATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTGATGCAGCCTATAGTGAGTCGTATTAC
FJ1_16S-A649	GATACTGGCATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTTGAGTCCTATAGTGAGTCGTATTAC
FJ1_16S-A860	GTGGCTTCCGGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTAACGCGTCCTATAGTGAGTCGTATTAC
FH14_16S-A325	ACTGGAACTGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGACACGGTCCTATAGTGAGTCGTATTACAG
FH14_16S-A412	GCGTGTATGATCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGAAGGCCTCCTATAGTGAGTCGTATTACAG
FH14_23S-A272	AGGAGCCCAGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGCCTGAATCCTATAGTGAGTCGTATTACAG
FH14_23S-A637	AGGGAAACCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTCTTAACCTATAGTGAGTCGTATTACAG
FH14_23S-A1572	TTCCAGGAAATCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGCCTCTAACCTATAGTGAGTCGTATTACAG
FH14_A34_DLRNA	GGATGCGTCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTTAATGCCTATAGTGAGTCGTATTACAG

1.3. RNA oligonucleotides: transcripts and synthetic RNAs

Pool and substrate sequences (modification site shown in bold)

Description	5'-Sequence-3'
Selection pool	GGACAUACUGAGCCUUCAACCAGCCUACCAUCCUUGAAGGC-N40-CAGUAUGUCCACCUUACC
Parent substrate sequence	GGACAUACUG**A**GCCUUCAA
TM	GGGUGCGUCG**A**GUUCCUGG
TV1	GGUGUAUGAG**A**GGGAAGUU
TV2	GGCACGCAGG**A**GAAGGACC
TM-UGAGC	GGGUGCGUUG**A**GCUCCUGG
5'TV1-3'parent	GGUGUAUGAG**A**GCCUUCAA
TV1-GAGC	GGUGUAUGAG**A**GCGAAGUU
Parent-AAG	GGACAUACUAAGCCUUCAA
Parent -UAG	GGACAUACUU**A**GCCUUCAA
Parent -CAG	GGACAUACUC**A**GCCUUCAA
Parent -GAA	GGACAUACUG**A**ACCUUCAA
Parent -GAU	GGACAUACUG**A**UCCUUCAA
Parent -GAC	GGACAUACUG**A**CCCUUCAA
3'-aminohexyl RNA substrate	ACAUACUG**A**GCCUUCAA-C6-NH2
5'-Hexynyl RNA substrate	Hexyne-GACAUACUG**A**GCCUUCAAAUA

SUPPORTING INFORMATION

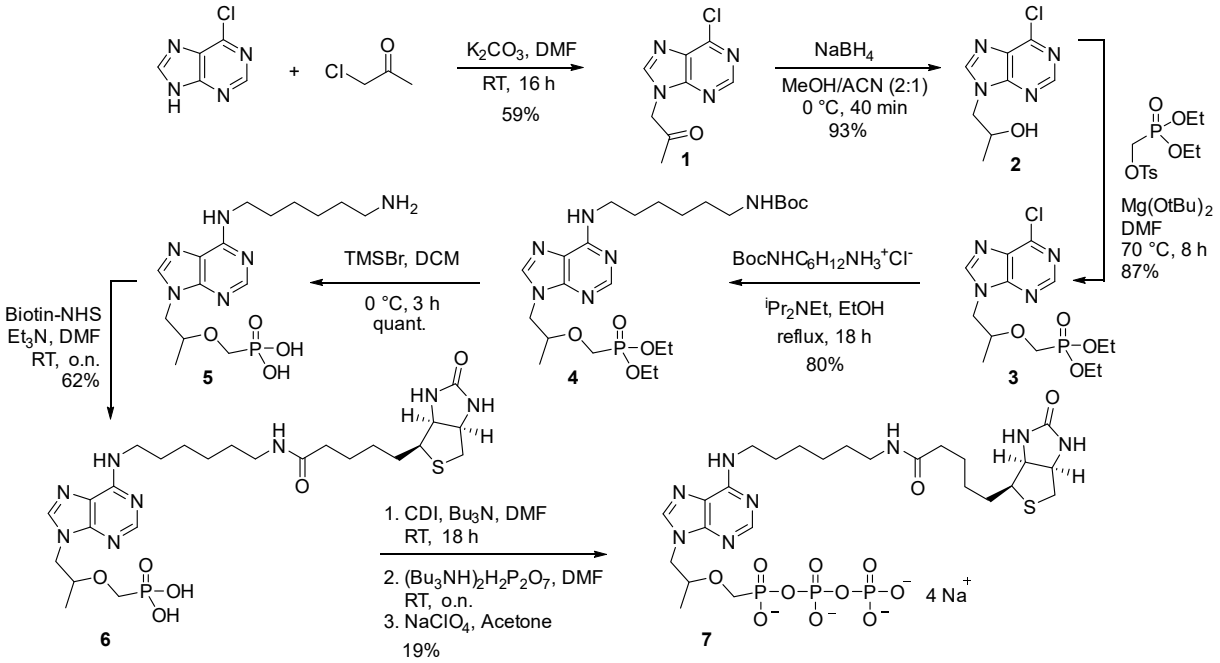
RNA w two labeling sites (DLRNA)	GGACAUACUG**A**GCCUUCAAUUAAGGAUGCGUCG**A**GUUAAUG
E. coli 5S rRNA (In vitro transcribed)	GGUGCCUGGCGGCCGUAGCGCGGUGGUCCCACCUGACCCCAUGCCGAACUCAGAAGUGA AACGCCGUAGCGCCGAUGGUAGUGUGGGGUCUCCCCAUGCGAG**A**GUAGGGAACUGCCAG GCAU

Ribozymes (Recognition arms underlined)

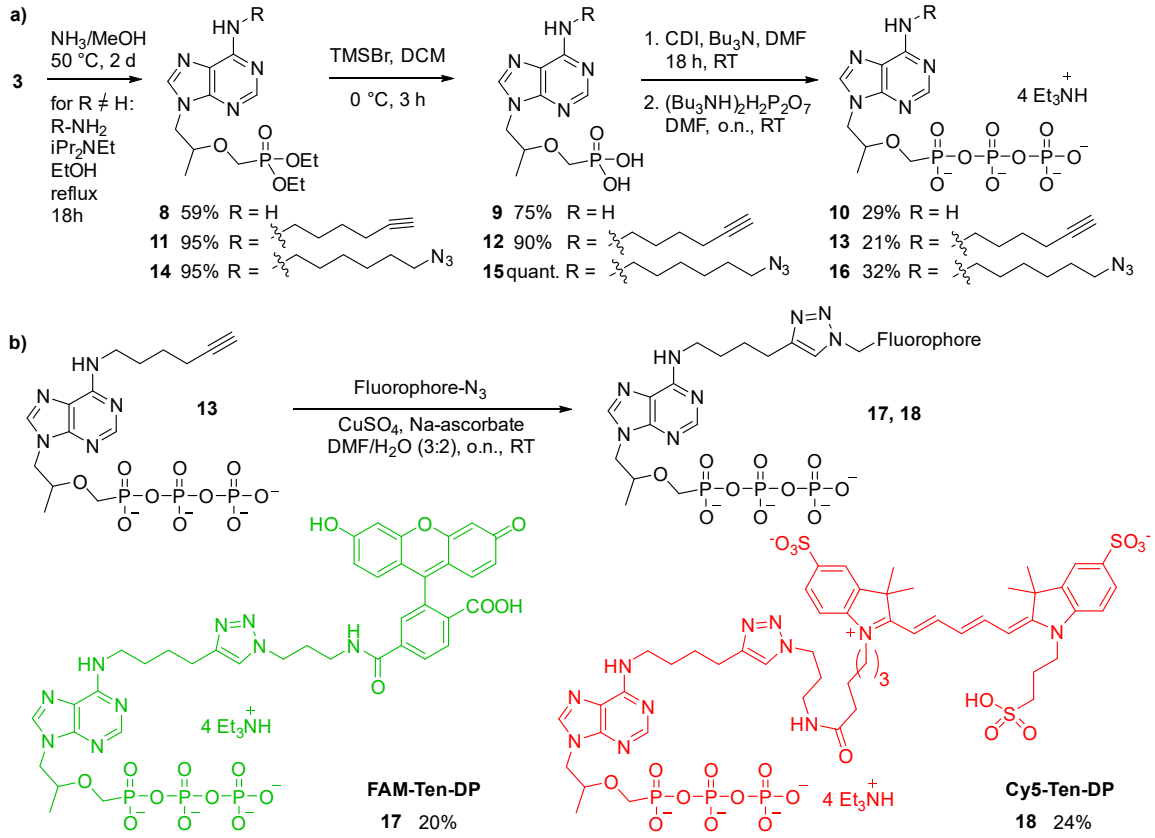
Name	5'-Sequence-3'
FJ1	GGUUGAAGGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUA CAGUAUGUCC
FJ8	GGUUGAAGGCGAAACGUGUCACAUAAGAAAACGGUAAACUAGCAAGUUCCCAGUAUGUCC
FJC1	GGUUGAAGGCCGAAUGCCACCGAACCGUAUAAUUGCCGCCUCCAAUUUC CAGUAUGUCC
FJC3	GGUUGAAGGCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCAGUAUGUCC
FJC9	GGCCAGGAACCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUA CGACGCACCC
FJ1_TM	GGAACUUCCCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUA CUCAUACACC
FJ1_TV1	GGGGUCCUUCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUA CCUGCGUGCC
FJ1_TV2	GGAACUUCCCACGGUGUUACAAAACGAAACCUUAGGACAUUACCUUCCCCCUCAUACACC
FJ1_TM-UGAGC	GGUUGAAGGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACUCAUACACC
FJ1_5'TV1-3'par	GGAACUUCGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACUCAUACACC
FJ1_TV1-GAGC	GGUUGAAGGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUAUAGUAUGUCC
FJ1_AAG	GGUUGAAGGUCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACAGUAUGUCC
FJ1_ GAA	GGUUGAAGGCACGGUGUUACAAAACGAAACCUUAGGACAUUACCUUCCCCCUCAUACACC
FJC9_TM	GGAACUUCCCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCUCAUACACC
FJC9_TV1	GGGGUCCUUCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCCUGCGUGCC
FJC9_TV2	GGUUGAAGGCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCUAGUAUGUCC
FJC9_ AAG	GGUUGAAGGCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCAAGUAUGUCC
FJC9_ UAG	GGUUGAAGGCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCGAGUAUGUCC
FJC9_ CAG	GGUUGAAGGUACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCAGUAUGUCC
FJC9_ GAA	GGUUGAAGGAACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCAGUAUGUCC
FJC9_ GAU	GGUUGAAGGGACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCAGUAUGUCC
FJC9_ GAC	GGAUUCAGGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACUGGGCUCCU
FJC9_5S-A99	GGUUCCCUACUCACGAGAUAUGUUGCACUACACUUUAGCGAAUUGGGCAUCCCGCAUGGGGA
FJ1_23S-A272	GGCAUCGAGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACACAGCAUGU
FJ1_23S-A362	GGUUAGAGGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUAUUUCCUGGAA
FJ1_23S-A1572	GGAUCAAGGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACAAUGUUCAG
FJ1_16S-A383	GGACUCAAGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUAUGCCAGUAUC
FJ1_16S-A649	GGACGCGUUAGCCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUACCGGAAGCCAC
FJ1_16S-A860	GGACCGUGUCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACAGUUCCAGU
FH14_16S-A325	GGAGGCCUUCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUCAUACACGC
FH14_16S-A412	GGUGCGGGACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUAACCCAACA
FH14_23S-A272	GGAUUCAGGCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACUGGGCUCCU
FH14_23S-A637	GGUUAAGACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGGUUUCCCU
FH14_23S-A1572	GGUUAGAGGCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUUUCCUGGAA
FH14_A34-DLRNA	GGCAUUAACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGACGCAUCC

2. Experimental Procedures

2.1. Synthesis of tenofovir diphosphate analogs



Scheme S1. Synthesis of N^6-biotin-tenofovir diphosphate (biotin-Ten-DP).

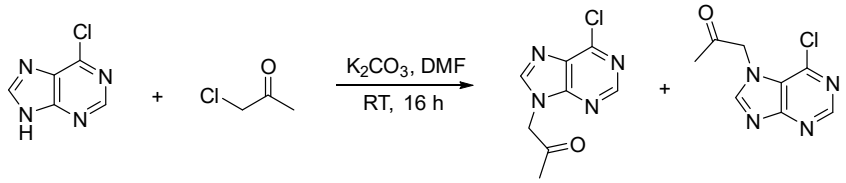


Scheme S2. a) Synthesis of tenofovir diphosphate, N^6-modified analogs N^6-hexynyl-Ten-DP and N^6-azidohexyl-Ten-DP. b) Synthesis of fluorescently labeled tenofovir diphosphates by CuAAC.

SUPPORTING INFORMATION

List of abbreviations used in the synthetic procedures:
ACN = Acetonitrile, CDI = 1,1'-Carbonyldiimidazole, Boc = *tert*-Butyloxycarbonyl, Bu_3N = Tri-*n*-butylamine, Cy5 = sulfo-cyanine-5, DCM = Dichloromethane, DMF = Dimethylformamide, DP = diphosphate, FAM = Fluorescein, h = hours, min = minutes, NHS = *N*-Hydroxysuccinimide, o.n. = overnight, RP = reversed-phase, RT = room temperature, TEAA = Triethylammonium acetate, TEAB = Triethylammonium bicarbonate, Ten = tenofovir, TLC = Thin-layer chromatography, TMSBr = Bromotrimethylsilane, Ts = Tosyl.

1-(6-chloro-9*H*-purin-9-yl)propan-2-one (1)



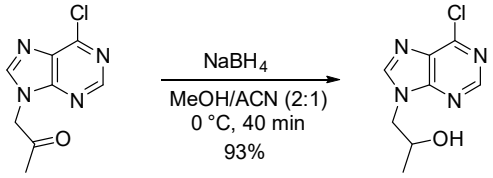
1 59% side product

Compound **1** was synthesized following a published procedure with slight modifications.[2]

To a solution of 6-chloropurine (1 g, 6.47 mmol) in 30 ml dry DMF, anhydrous K_2CO_3 (894 mg, 6.47 mmol) was added. The mixture was stirred for 30 min at RT, followed by addition of a solution of chloroacetone (0.57 ml, 7.12 mmol) in 10 ml dry DMF over 1.5 h. The reaction mixture was continuously stirred overnight at RT, then filtered through Celite, and the filtrate was evaporated to dryness. The residue was dissolved in 150 ml $CHCl_3$ and washed with saturated $NaHCO_3$ (4 x 25 ml) and brine (20 ml). The organic layer was dried over Na_2SO_4, and the solvent was evaporated. The crude product was purified by flash column chromatography (3% MeOH/DCM), affording compound **1** as a yellowish-white solid (804 mg, 3.82 mmol, 59%). A minor amount of 7-alkylated isomer was formed, which was separated by column chromatography.

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.71 (s, 1H), 8.11 (s, 1H), 5.15 (s, 2H), 2.37 (s, 3H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 198.6, 152.1, 151.8, 151.2, 145.6, 131.1, 52.3, 27.2.
HRMS (ESI$^+$): Exact mass calculated for $C_8H_8ClN_4O$ $[M + H]^+$, 211.03811. Found 211.03798.

1-(6-chloro-9*H*-purin-9-yl)propan-2-ol (2)

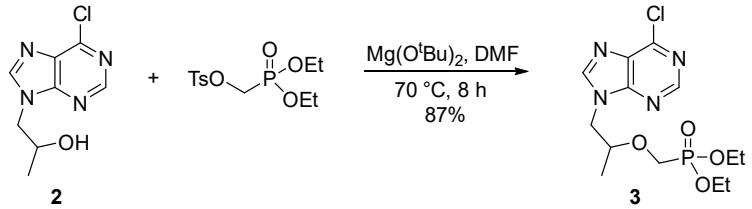


1 **2**

Compound **1** (780 mg, 3.7 mmol) was dissolved in 40 ml MeOH/ACN (2:1). Tthe solution was cooled to 0 °C and $NaBH_4$ (168 mg, 4.44 mmol) was added in three portions over 20 min. The mixture was stirred for another 20 min at 0 °C until TLC confirmed the complete consumption of the reactant. Excess of NaBH4 reagent was consumed by the addition of 20 ml saturated NH_4Cl and the reaction volume was concentrated to ~25 ml. The reaction mixture was extracted with $CHCl_3$ (4 x 50 ml). The combined organic layers were washed with saturated NH_4Cl (30 ml), brine (30 ml), dried over Na_2SO_4, and concentrated under reduced pressure. The crude product was purified by flash column chromatography (4-5% MeOH/DCM), to give the pure title compound **2** as a white solid (731 mg, 3.44 mmol, 93%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.70 (s, 1H), 8.21 (s, 1H), 4.43 (dd, *J* = 14.0, 2.7 Hz, 1H), 4.31 (dqd, *J* = 8.0, 6.3, 2.7 Hz, 1H), 4.13 (dd, *J* = 14.1, 8.0 Hz, 1H), 3.13 (s, 1H), 1.32 (d, *J* = 6.3 Hz, 3H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 151.8, 151.7, 150.6, 146.4, 131.0, 66.0, 51.5, 20.8.
HRMS (ESI$^+$): Exact mass calculated for $C_8H_9ClN_4NaO$ $[M + Na]^+$, 235.03571. Found 235.03518.

Diethyl (((1-(6-chloro-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonate (3)

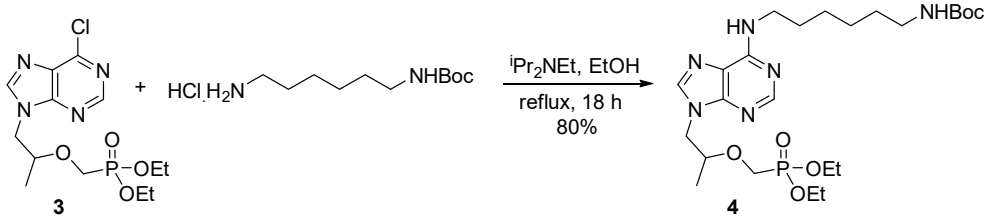


2 **3**

In a 25 ml Schlenk flask, a colorless solution of compound **2** (150 mg, 0.71 mmol) in 5 ml dry DMF was treated with $Mg(O^tBu)_2$ (363 mg, 2.13 mmol) at RT. The white suspension was stirred at 70 °C for 30 min. Then (diethoxyphosphoryl)methyl 4-methylbenzenesulfonate (344 mg, 1.07 mmol) was added in small portions over 10 min under argon flow. The reaction mixture was then stirred for 7 h at 70 °C. The solvent was evaporated to dryness, and the residue was put onto a pad of celite and washed with 20% MeOH/DCM. The filtrate was reduced in vacuo and the crude residue was purified by flash chromatography (1.5-3.5% MeOH/DCM) to afford compound **3** as a colorless oil (225 mg, 0.62 mmol, 87%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.73 (s, 1H), 8.31 (s, 1H), 4.47 (dd, *J* = 14.5, 2.8 Hz, 1H), 4.22 (dd, *J* = 14.5, 7.7 Hz, 1H), 4.11 – 4.05 (m, 2H), 4.04 – 3.98 (m, 2H), 3.97 – 3.91 (m, 1H), 3.85 (dd, *J* = 13.7, 8.9 Hz, 1H), 3.58 (dd, *J* = 13.7, 9.4 Hz, 1H), 1.32 – 1.29 (t, *J* = 7.1 Hz, 3H), 1.26 – 1.20 (m, 6H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 152.0, 151.8, 150.9, 146.5, 131.2, 75.9 (d, *J* = 11.2 Hz), 63.5, 62.4 (d, *J* = 6.5 Hz), 62.3 (d, *J* = 6.6 Hz), 61.8, 48.6, 16.5 – 16.4 (m).
^{31}P NMR (162 MHz, $CDCl_3$): δ 20.66.
HRMS (ESI$^+$): Exact mass calculated for $C_{13}H_{21}ClN_4O_4P$ $[M + H]^+$, 363.09835. Found 363.09836.

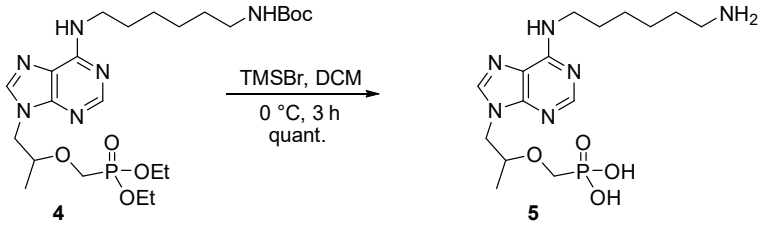
***tert*-butyl (6-((9-(2-((diethoxyphosphoryl)methoxy)propyl)-9*H*-purin-6-yl)amino)hexyl)carbamate (4)**



In a 50 ml Schlenk flask, a solution of compound **3** (205 mg, 0.57 mmol) and *N*-Boc-1,6-hexanediamine hydrochloride (358 mg, 1.41 mmol) in 12 ml absolute EtOH was treated with *N*,*N*-diisopropylethylamine (0.48 ml, 2.83 mmol). The solution was heated under reflux for 18 h until TLC confirmed the complete consumption of the reactants. The solvent was evaporated and the residue was purified by flash column chromatography (4-4.5% MeOH/DCM). The pure title compound **4** was obtained as a colorless sticky oil (246 mg, 0.45 mmol, 80%).

^{1}H NMR (400 MHz, MeOD): δ 8.24 (s, 1H), 8.07 (s, 1H), 4.35 (dd, *J* = 14.5, 3.1 Hz, 1H), 4.20 (dd, *J* = 14.5, 7.9 Hz, 1H), 4.06 – 3.90 (m, 6H), 3.71 (dd, *J* = 14.0, 9.7 Hz, 1H), 3.58 (br s, 2H), 3.03 (t, *J* = 6.8 Hz, 2H), 1.70 (p, *J* = 7.1 Hz, 2H), 1.50 – 1.36 (m, 15H), 1.27 – 1.18 (m, 9H).
^{13}C NMR (101 MHz, MeOD): δ 158.5, 156.1, 153.7, 149.8, 142.9, 120.0, 79.7, 77.6 (d, *J* = 12.3 Hz), 64.0 (d, *J* = 6.6 Hz), 63.9 (d, *J* = 6.6 Hz), 63.8, 62.1, 49.1, 41.5, 41.3, 30.9, 30.5, 28.8, 27.7, 27.6, 16.7 – 16.6 (m).
^{31}P NMR (162 MHz, MeOD): δ 21.99.
HRMS (ESI$^+$): Exact mass calculated for $C_{24}H_{44}N_6O_6P$ $[M + H]^+$, 543.30545. Found 543.30534.

(((1-(6-((6-aminohexyl)amino)-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonic acid (5)



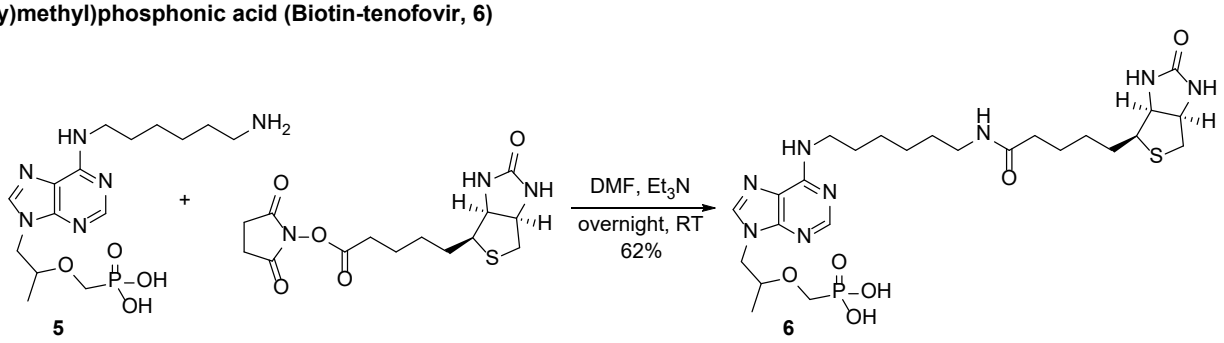
Deprotection of the phosphonate ethyl ester and Boc groups was performed according to modified published procedures.[3],[4]

In a 25 ml Schlenk flask, compound **4** (162 mg, 0.3 mmol) was dissolved in 6 ml dry DCM. The tube was placed in an ice bath, and TMSBr (0.24 ml, 1.8 mmol) was added dropwise to the solution. The reaction mixture was stirred for 3 h at 0 °C until TLC confirmed the complete consumption of compound **4**. Evaporation of the solvent afforded a yellowish sticky oil, which was co-evaporated with MeOH followed by Et_2O. The obtained yellowish-white solid was dissolved in 20 ml of water and extracted with DCM (4 x 3 ml). The aqueous fraction was lyophilized, and the title compound **5** was obtained as a white solid in quantitative yield.

^{1}H NMR (400 MHz, D_2O): δ 8.34 (s, 1H), 8.30 (s, 1H), 4.44 (dd, *J* = 14.9, 3.0 Hz, 1H), 4.26 (dd, *J* = 14.8, 7.8 Hz, 1H), 4.00 – 3.94 (m, 1H), 3.77 (dd, *J* = 13.8, 8.8 Hz, 1H), 3.57 – 3.51 (m, 2H), 2.94 (t, *J* = 7.6 Hz, 2H), 1.77 – 1.68 (m, 2H), 1.66 – 1.58 (m, 2H), 1.49 – 1.34 (m, 4H), 1.19 (d, *J* = 6.3 Hz, 3H).
^{13}C NMR (101 MHz, D_2O): δ 148.5, 147.0, 144.7, 144.1, 118.0, 76.5 (d, *J* = 11.6 Hz), 63.5 (d, *J* = 161.0 Hz), 48.4, 42.1, 39.3, 27.2, 26.5, 25.3, 25.2, 15.8.

^{31}P NMR (162 MHz, D_2O): δ 19.04.
HRMS (ESI$^-$): Exact mass calculated for $C_{15}H_{26}N_6O_4P$ [M - H]$^-$, 385.17586. Found 385.17591.
HRMS (ESI$^+$): Exact mass calculated for $C_{15}H_{28}N_6O_4P$ [M + H]$^+$, 387.19042. Found 387.18994.

(((1-(6-((6-(5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazole-4-yl)pentanamido)exyl)amino)-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonic acid (Biotin-tenofovir, 6)



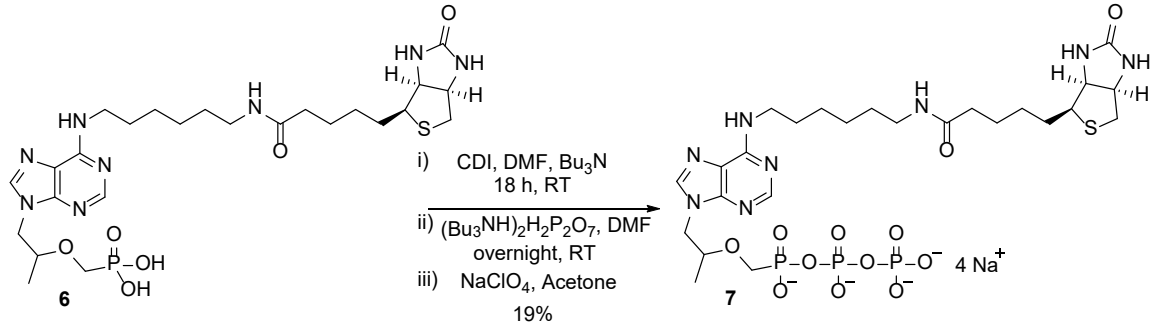
5 **6**

In a 10 ml Schlenk flask, compound **5** (40 mg, 0.104 mmol) was dissolved in 3 ml dry DMF. Biotin-NHS ester (42 mg, 0.124 mmol), was added and the reaction mixture was stirred for 15 min at RT before Et_3N (150 µl, 1.08 mmol) was added. The white suspension became clear after stirring for 30 min at RT, and the reaction progress was monitored by RP-TLC (using 20% MeCN/H_2O). The reaction mixture was stirred overnight at RT before the solvent was evaporated. After coevaporation with H_2O, the residue was dissolved in 30 ml water and extracted with DCM (3 x 3 ml). The aqueous fraction was lyophilized, and a yellowish-white solid was obtained. The crude product was purified by reversed-phase chromatography to give the title compound **6** as a white foam (40 mg, 0.065 mmol, 62%).

RP-purification procedure: *Performed on:* GE Healthcare ÄKTAprime plus, *Column:* Lobar 310-25 LiChroprep RP-18 (40-63 µm), Merck *Detection wavelength:* 280 nm, *Solvent systems:* A: H_2O, B: ACN, *Gradient:* 5–50% B in 50 min, *Flow rate:* 7 ml/min

^{1}H NMR (400 MHz, CD_3CN:D_2O (1:1)): δ 8.78 (s, 2H), 5.03 – 5.00 (m, 1H), 4.91 (dd, *J* = 14.7, 3.6 Hz, 1H), 4.84 (dd, *J* = 7.9, 4.5 Hz, 1H), 4.48 – 4.41 (m, 1H), 4.19 (dd, *J* = 12.9, 9.3 Hz, 1H), 4.10 – 3.99 (m, 2H), 3.75 – 3.64 (m, 6H), 3.42 (dd, *J* = 12.9, 5.0 Hz, 1H), 3.22 (d, *J* = 12.9 Hz, 1H), 2.70 (t, *J* = 7.3 Hz, 2H), 2.56 – 2.54 (m, 1H), 2.25 – 1.85 (m, 14H), 1.78 (t, *J* = 7.3 Hz, 4H), 1.63 (d, *J* = 6.3 Hz, 3H).
^{31}P NMR (162 MHz, CD_3CN:D_2O (1:1)): δ 15.70.
HRMS (ESI$^-$): Exact mass calculated for $C_{25}H_{40}N_8O_6PS$ [M - H]$^-$, 611.25346. Found 611.25471.

[((1-(6-((6-(5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazole-4-yl)pentanamido)hexyl)amino)-9*H*-purin-9-yl)propan-2-yl)oxy]methyl-[hydroxy(phosphonooxy)phosphoryl]oxy-phosphinic acid, sodium salt (Biotin-tenofovir-DP; 7)



6 **7**

Phosphonate activation using CDI, followed by reaction with pyrophosphate was performed in analogy to published procedures.[5]

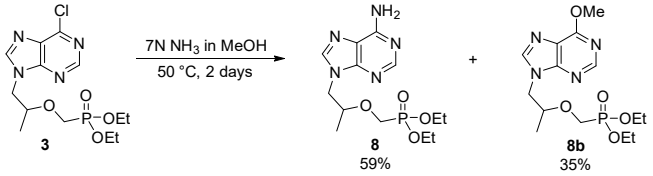
First, the tributylammonium salt of compound **6** was generated. A suspension of the free acid form of **6** (21 mg, 34 µmol) in 1 ml of dry DMF was treated with tributylamine (32 µl, 136 µmol) for 1 h at RT. The clear solution was evaporated, followed by co-evaporation with dry DMF (3 x 1 ml). Then, the tributylammonium salt of compound **6** was dissolved in 1 ml dry DMF, and a solution of CDI (28 mg, 170 µmol) in 0.5 ml of dry DMF was added dropwise. The reaction mixture was stirred for 18 h at RT. Excess CDI was quenched by the addition of MeOH (5.5 µl, 136 µmol) followed by stirring for 1 h at RT. Then, $(Bu_3NH)_2H_2P_2O_7$ (93 mg, 170 µmol) dissolved in 1 ml dry DMF was added. The resulting white suspension was stirred overnight at RT. The supernatant was separated by centrifugation and the precipitate (imidazolium pyrophosphate) was washed with dry DMF (2 x 0.2 ml). The combined DMF fractions were evaporated, followed by co-evaporation with H_2O. The resulting residue was dissolved in 10 ml H_2O and extracted with DCM (3 x 2 ml). The aqueous fraction was lyophilized, and the crude product was purified by anion exchange chromatography (on DEAE Sephadex, see general

methods, section 1.1). The product-containing fraction (55-65% buffer B) was evaporated, and then co-evaporated with H_2O. The residue was further purified by reversed-phase chromatography. The resulting triethylammonium salt of the title compound was obtained as a white foam, and was transformed to the sodium salt by precipitation from 20 ml 2% $NaClO_4$ in acetone. After centrifugation, the pellet was washed with acetone (3 x 5 ml), and dried in vacuo. The pure product **7** was obtained as a white solid (5.5 mg, 6.39 µmol, 19%).

RP-purification procedure: *Column:* Lobar 310-25 LiChroprep RP-18 (40-63 µm), Merck
Detection wavelength: 280 nm, *Solvent systems:* buffer A: 100 mM TEAB (pH = 7.5); buffer B: 100 mM TEAB in 80% ACN
Gradient: 10% B for 10 min, 10-25% B in 15 min, 25% B for 5 min, 25-35% B in 10 min, 35% B for 10 min, *Flow rate:* 5 ml/min

^{1}H NMR (400 MHz, D_2O): δ 8.25 (s, 2H), 4.47 – 4.42 (m, 2H), 4.30 – 4.25 (m, 2H), 4.05 – 3.99 (m, 1H), 3.87 – 3.68 (m, 2H), 3.54 (q, *J* = 7.1 Hz, 4H), 3.20 – 3.12 (m, 3H), 2.83 (dd, *J* = 13.1, 5.0 Hz, 1H), 2.63 (d, *J* = 13.0 Hz, 1H), 2.22 – 2.17 (m, 2H), 1.71 – 1.26 (m, 14H), 1.17 – 1.12 (m, 7H).
^{31}P NMR (162 MHz, D_2O): δ 8.72 (d, *J* = 26.4 Hz), -10.70 (d, *J* = 19.9 Hz), -23.11 (t, *J* = 21.5 Hz).
HRMS (ESI$^-$): Exact mass calculated for $C_{25}H_{42}N_8O_{12}P_3S$ [M - H]$^-$, 771.18612. Found 771.18521.

Diethyl (((1-(6-amino-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonate (Tenofovir diethyl ester, 8)

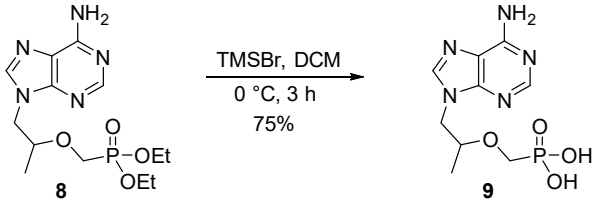


In a stainless steel autoclave, compound **3** (200 mg, 0.55 mmol) was treated with 25 ml of 7 N NH_3 in MeOH for 2 days at 50 °C. After solvent evaporation, the residue was directly purified by flash column chromatography (3-7.5% MeOH/DCM) to give the desired compound **8** as a white solid white solid (**8**, 112 mg, 0.326 mmol, 59%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.33 (s, 1H), 7.96 (s, 1H), 6.00 (br s, 2H), 4.35 (dd, *J* = 14.4, 3.0 Hz, 1H), 4.15 – 3.99 (m, 5H), 3.95 – 3.88 (m, 1H), 3.83 (dd, *J* = 13.6, 9.1 Hz, 1H), 3.58 (dd, *J* = 13.6, 9.7 Hz, 1H), 1.29 (t, *J* = 7.1 Hz, 3H), 1.26 – 1.22 (m, 6H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 155.5, 152.8, 150.1, 141.8, 119.1, 76.3 (d, *J* = 11.9 Hz), 63.6, 62.4 (d, *J* = 6.7 Hz), 62.3 (d, *J* = 6.7 Hz), 61.9, 48.2, 16.4 (d, *J* = 5.2 Hz), 16.3 (d, *J* = 5.2 Hz).
^{31}P NMR (162 MHz, $CDCl_3$): δ 20.83.
HRMS (ESI$^+$): Exact mass calculated for $C_{13}H_{23}N_5O_4P$ [M + H]$^+$, 344.14822. Found 344.14823.

The MeO-substituted compound **8b** was isolated as a side product (eluted first from the column, colorless oil, 68 mg, 0.190 mmol, 35%)
^{1}H NMR (400 MHz, $CDCl_3$): δ 8.51 (s, 1H), 8.07 (s, 1H), 4.40 (dd, *J* = 14.5, 3.0 Hz, 1H), 4.20 – 3.99 (m, 8H), 3.96 – 3.89 (m, 1H), 3.82 (dd, *J* = 13.7, 9.1 Hz, 1H), 3.57 (dd, *J* = 13.6, 9.6 Hz, 1H), 1.30 (t, *J* = 7.1 Hz, 3H), 1.28 – 1.20 (m, 6H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 161.0, 152.1, 151.9, 143.4, 121.1, 76.3 (d, *J* = 11.9 Hz), 63.5, 62.4 (d, *J* = 6.5 Hz), 62.3 (d, *J* = 6.5 Hz), 61.9, 54.2, 48.3, 16.5 – 16.3 (m).
^{31}P NMR (162 MHz, $CDCl_3$): δ 20.72.
HRMS (ESI$^+$): Exact mass calculated for $C_{14}H_{24}N_4O_5P$ [M + H]$^+$, 359.14788. Found 359.14783.

(((1-(6-amino-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonic acid (Tenofovir, 9)



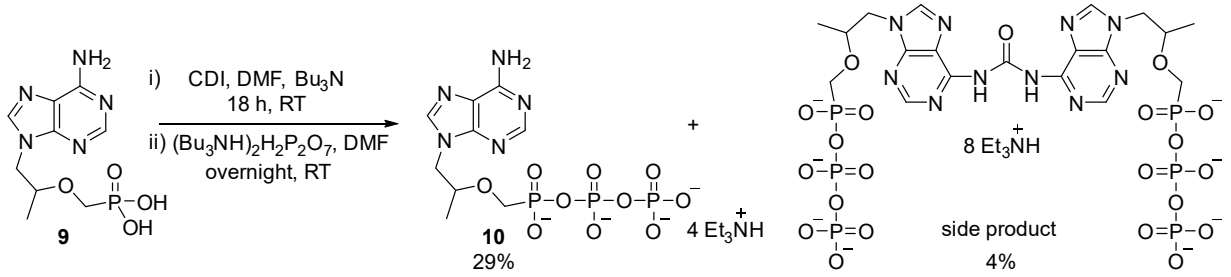
Unmodified Tenofovir was synthesized in analogy to previous reports.[3, 6]

In a 25 ml Schlenk flask, compound **8** (100 mg, 0.29 mmol) was dissolved in 6 ml dry DCM and treated with TMSBr (0.23 ml, 1.74 mmol) at 0 °C for 3 h. TLC confirmed the complete consumption of the reactant. The solvent was evaporated, and the resulting yellowish sticky oil was co-evaporated with MeOH, followed by Et_2O. The residue was dissolved in 20 ml of water and washed with DCM (4 x 3 ml). The aqueous fraction was lyophilized, and the crude product was purified by silica gel column chromatography (acetone/H_2O/Et_3N,

6:1:1), followed by a second purification by silica gel column chromatography (^{i}PrOH/H_2O/conc. NH_3, 7:2:1) to afford compound **9** as a white foam (63 mg, 0.22 mmol, 75%).

^{1}H NMR (400 MHz, D_2O): δ 8.14 (s, 1H), 8.02 (s, 1H), 4.25 (dd, *J* = 14.6, 3.9 Hz, 1H), 4.14 (dd, *J* = 14.6, 5.8 Hz, 1H), 3.92 – 3.84 (m, 1H), 3.50 (dd, *J* = 12.6, 9.3 Hz, 1H), 3.37 (dd, *J* = 12.6, 9.3 Hz, 1H), 1.03 (d, *J* = 6.3 Hz, 3H).
^{13}C NMR (101 MHz, D_2O): δ 155.1, 151.9, 148.8, 143.2, 117.7, 75.4 (d, *J* = 11.3 Hz), 66.2 (d, *J* = 153.0 Hz), 47.4, 16.1.
^{31}P NMR (162 MHz, D_2O): δ 13.92.
HRMS (ESI$^-$): Exact mass calculated for $C_9H_{13}N_5O_4P$ [M - H]$^-$, 286.07106. Found 286.07109.

[1-(6-amino-9*H*-purin-9-yl)propan-2-yl]oxymethyl-[hydroxy(phosphonooxy)phosphoryl]oxy-phosphinic acid, triethylammonium salt (Tenofovir-DP, 10)



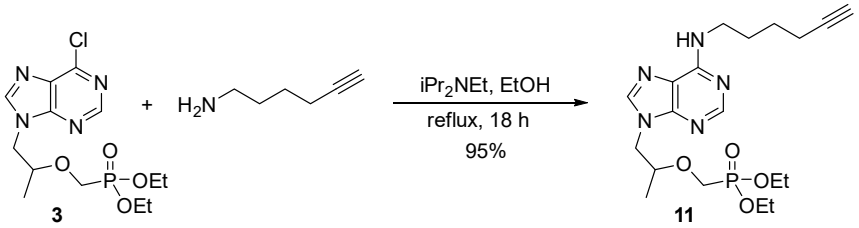
Compound **9** (25 mg, 86 µmol) was co-evaporated with dry DMF (3 x 1 ml), dissolved in 1 ml of dry DMF and treated with tributylamine (82 µl, 344 µmol) for 1 h at RT. To the clear solution, a solution of CDI (105 mg, 648 µmol) in 1 ml dry DMF was added dropwise, and the reaction mixture was stirred for 18 h at RT. Excess CDI was quenched by the addition of MeOH (22 µl, 557 µmol). After 1 h at RT, $(Bu_3NH)_2H_2P_2O_7$ (354 mg, 645 µmol) dissolved in 1.5 ml dry DMF was added, and the resulting white suspension was stirred overnight at RT. The precipitate was removed by centrifugation and the supernatant was evaporated. The residue was directly loaded onto an anion exchange column (see general method section). Two fractions were collected separately (first fraction: 60-70% buffer B, second fraction: 95-100% buffer B), and after evaporation further purified via reversed-phase chromatography. Compound **10** was obtained from the first eluting fraction as white foam (21.4 mg, 25 µmol, 29%). The side product from the second fraction was identified as the homo-coupled side product (5.8 mg, 3.5 µmol, 4%).

RP-purification procedure: *Performed on:* GE Healthcare ÄKTAprime plus, *Column:* Lobar 310-25 LiChroprep RP-18 (40-63 µm), Merck
Detection wavelength: 280 nm, *Solvent systems:* buffer A: 100 mM TEAB (pH = 7.5); buffer B: 100 mM TEAB in 80% ACN
Gradient: 5% B for 5 min, 5-15% B in 15 min, 15% B for 25 min, *Flow rate:* 5 ml/min

Compound **10**:
λ_{max} (UV-vis absorption in H_2O): 259 nm
^{1}H NMR (400 MHz, D_2O): δ 8.32 (s, 1H), 8.23 (s, 1H), 4.42 (dd, *J* = 14.7, 3.3 Hz, 1H), 4.24 (dd, *J* = 14.8, 5.9 Hz, 1H), 4.04 – 3.97 (m, 1H), 3.84 (dd, *J* = 13.3, 9.2 Hz, 1H), 3.74 (dd, *J* = 13.3, 9.5 Hz, 1H), 1.09 (d, *J* = 6.3 Hz, 3H).
^{31}P NMR (162 MHz, D_2O): δ 8.67 (d, *J* = 26.4 Hz), -10.93 (d, *J* = 19.2 Hz), -23.35 (t, *J* = 22.3 Hz).
HRMS (ESI$^-$): Exact mass calculated for $C_9H_{15}N_5O_{10}P_3$ [M - H]$^-$, 446.00373. Found 446.00372.

A second fraction was obtained from anion exchange chromatography, as a white foam after RP-purification which was identified as the homo-coupled side product (5.8 mg, 3.5 µmol, 4%).
λ_{max} (UV-vis absorption in H_2O): 292 nm
^{1}H NMR (400 MHz, D_2O): δ 8.92 (s, 1H), 8.82 (s, 1H), 4.61 (dd, *J* = 14.7, 2.9 Hz, 1H), 4.42 (dd, *J* = 14.8, 6.2 Hz, 1H), 4.12 – 4.05 (m, 1H), 3.89 (dd, *J* = 13.4, 9.1 Hz, 1H), 3.75 (dd, *J* = 13.2, 9.7 Hz, 1H), 1.19 (d, *J* = 6.3 Hz, 3H).
^{31}P NMR (162 MHz, D_2O): δ 8.70 (d, *J* = 26.5 Hz), -11.00 (d, *J* = 19.9 Hz), -23.45 (t, *J* = 22.6 Hz).
HRMS (ESI$^-$): Exact mass calculated for $C_{19}H_{28}N_{10}O_{21}P_6$ [M - 2H]$^{2-}$, 458.99336. Found 458.99559.

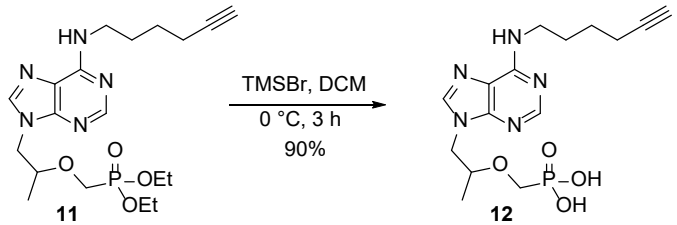
Diethyl (((1-(6-(hex-5-yn-1-ylamino)-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonate (Hexynyl-tenofovir diethyl ester, 11)



In a 50 ml Schlenk flask, compound **3** (200 mg, 0.55 mmol) and 5-Hexyn-1-amine (134 mg, 1.38 mmol) were dissolved in 15 ml absolute EtOH and treated with *N*,*N*-diisopropylethylamine (0.24 ml, 1.40 mmol). The solution was heated under reflux for 18 h. TLC confirmed the complete consumption of the reactant. The solvent was evaporated and the residue was purified by flash chromatography (3.5% MeOH/DCM). The pure product **11** was obtained as a colorless sticky oil (221 mg, 0.52 mmol, 95%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.36 (s, 1H), 7.88 (s, 1H), 5.76 (br s, 1H), 4.33 (dd, *J* = 14.4, 2.9 Hz, 1H), 4.14 – 4.00 (m, 5H), 3.95 – 3.88 (m, 1H), 3.81 (dd, *J* = 13.6, 9.1 Hz, 1H), 3.68 (br s, 2H), 3.57 (dd, *J* = 13.8, 9.4 Hz, 1H), 2.25 (td, *J* = 7.0, 2.6 Hz, 2H), 1.95 (t, *J* = 2.9 Hz, 1H), 1.85 – 1.77 (m, 2H), 1.69 – 1.62 (m, 2H), 1.30 (t, *J* = 7.1 Hz, 3H), 1.27 – 1.22 (m, 6H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.9, 153.0, 149.0, 140.9, 119.3, 83.9, 76.4 (d, *J* = 12.0 Hz), 68.7, 63.6, 62.4 (d, *J* = 6.5 Hz), 62.3 (d, *J* = 6.5 Hz), 61.9, 48.1, 40.0, 28.8, 25.6, 18.1, 16.5 (d, *J* = 6.8 Hz), 16.4 (d, *J* = 5.4 Hz).
^{31}P NMR (162 MHz, $CDCl_3$): δ 20.84.
HRMS (ESI^+): Exact mass calculated for $C_{19}H_{31}N_5O_4P$ $[M + H]^+$, 424.21082. Found 424.21083.

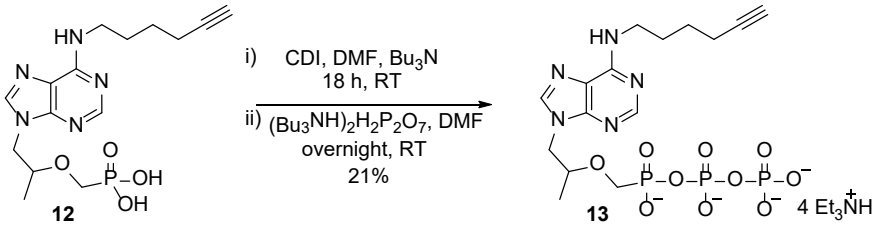
(((1-(6-(hex-5-yn-1-ylamino)-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonic acid (Hexynyl-tenofovir, 12)



In a 25 ml Schlenk flask, compound **11** (100 mg, 0.24 mmol) was dissolved in 6 ml dry DCM, and TMSBr (0.20 ml, 1.51 mmol) was added dropwise at 0 °C. The reaction was stirred for 3 h at 0 °C. TLC confirmed the complete consumption of the reactant. The solvent was evaporated and the resulting residue was co-evaporated with MeOH and Et_2O before it was dissolved in 20 ml water and extracted with DCM (4 x 3 ml). The aqueous fraction was lyophilized and purified by silica gel column chromatography (^{i}PrOH/H_2O/conc. NH_3, 7:2:1) to afforded compound **12** as a white foam (78 mg, 0.22 mmol, 90%).

^{1}H NMR (400 MHz, D_2O): δ 8.08 (s, 1H), 8.05 (s, 1H), 4.26 (dd, *J* = 14.6, 3.4 Hz, 1H), 4.12 (dd, *J* = 14.7, 6.6 Hz, 1H), 3.92 – 3.84 (m, 1H), 3.61 (dd, *J* = 13.1, 9.3 Hz, 1H), 3.46 – 3.40 (m, 3H), 2.26 (t, *J* = 2.6 Hz, 1H), 2.16 (td, *J* = 7.0, 2.6 Hz, 2H), 1.72 – 1.65 (m, 2H), 1.56 – 1.49 (m, 2H), 1.09 (d, *J* = 6.3 Hz, 3H).
^{13}C NMR (101 MHz, D_2O): δ 153.1, 150.7, 147.4, 142.7, 117.9, 85.5, 75.8 (d, *J* = 11.9 Hz), 69.3, 64.8 (d, *J* = 157.8 Hz), 47.7, 40.4, 27.5, 25.0, 17.3, 16.0.
^{31}P NMR (162 MHz, D_2O): δ 15.73.
HRMS (ESI^-): Exact mass calculated for $C_{15}H_{21}N_5O_4P$ $[M - H]^-$, 366.13366. Found 366.13366.

[1-(6-(hex-5-yn-1-ylamino)-9*H*-purin-9-yl)propan-2-yl]oxymethyl-[hydroxy(phosphonooxy)phosphoryl]oxy-phosphinic acid, triethylammonium salt (Hexynyl-tenofovir-DP, 13)

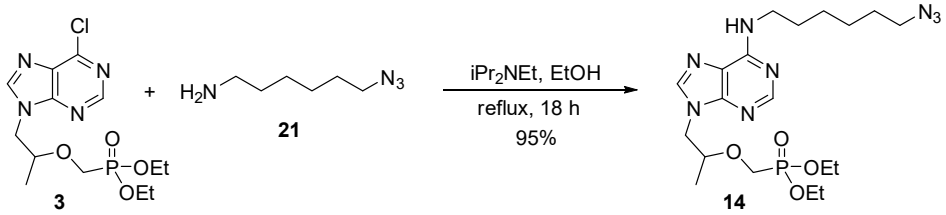


The free acid form of compound **12** (65 mg, 177 µmol) was suspended in 1 ml dry DMF and treated with tributylamine (168 µl, 708 µmol) for 1 h at RT to form a clear solution of the corresponding tributylammonium salt. The solvent was evaporated, followed by co-evaporation with dry DMF (3 x 1 ml). The tributylammonium salt of compound **12** was dissolved in 1.5 ml dry DMF and a solution of CDI (216 mg, 1.33 mmol) in 1.5 ml dry DMF was added dropwise. The reaction mixture was stirred for 18 h at RT before excess CDI was quenched by the addition of MeOH (46 µl, 1.15 mmol) and stirring was continued for 1 h at RT. Then $(Bu_3NH)_2H_2P_2O_7$ (730 mg, 1.33 mmol) dissolved in 2 ml dry DMF was added and the resulting white suspension was stirred overnight at RT. The precipitate was removed by centrifugation and the supernatant was evaporated, followed by co-evaporation with H_2O. The residue was dissolved in 10 ml H_2O and extracted with DCM (3 x 2 ml). The aqueous layer was lyophilized and the crude product was purified by anion exchange chromatography (see general method section). The product-containing fraction (60-70% buffer B) was collected and the solvent was evaporated. The final purification was achieved by reverse-phase chromatography to obtain compound **13** as a white foam (35.5 mg, 38 µmol, 21%).

RP-purification procedure: *Performed on:* GE Healthcare ÄKTAprime plus, *Column:* Lobar 310-25 LiChroprep RP-18 (40-63 µm), Merck *Detection wavelength:* 280 nm, *Solvent systems:* buffer A: 100 mM TEAB (pH = 7.5); buffer B: 100 mM TEAB in 80% ACN *Gradient:* 5% B for 5 min, 5-20% B in 5 min, 20% B for 10 min, 20-30% B in 5 min, 30% B for 20 min, *Flow rate:* 5 ml/min

^{1}H NMR (400 MHz, D_2O): δ 8.21 (s, 1H), 8.20 (s, 1H), 4.40 (dd, *J* = 14.8, 3.5 Hz, 1H), 4.25 (dd, *J* = 14.8, 5.7 Hz, 1H), 4.06 – 3.99 (m, 1H), 3.84 (dd, *J* = 13.3, 9.2 Hz, 1H), 3.74 (dd, *J* = 13.3, 9.4 Hz, 1H), 3.58 (br s, 2H), 2.29 (t, *J* = 2.6 Hz, 1H), 2.24 (td, *J* = 7.0, 2.6 Hz, 2H), 1.82 – 1.75 (m, 2H), 1.65 – 1.58 (m, 2H), 1.11 (d, *J* = 6.3 Hz, 3H).
^{31}P NMR (162 MHz, D_2O): δ 8.64 (d, *J* = 26.6 Hz), -10.92 (d, *J* = 19.9 Hz), -23.36 (dd, *J* = 26.5, 19.9 Hz).
HRMS (ESI^-): Exact mass calculated for $C_{15}H_{23}N_5O_{10}P_3$ [M - H]$^-$, 526.06633. Found 526.06663.

Diethyl (((1-(6-((6-azidohexyl)amino)-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonate (Azidohexyl-tenofovir diethyl ester, 14)



In a 50 ml Schlenk flask, compound **3** (150 mg, 0.41 mmol) and 6-azidohexan-1-amine **21** (150 mg, 1.05 mmol) were dissolved in 5 ml absolute EtOH and treated with *N*,*N*-diisopropylethylamine (0.18 ml, 1.05 mmol). Then the solution was refluxed for 18 h. TLC confirmed the complete consumption of the reactant. The solvent was evaporated and the residue was purified by flash chromatography (3.5% MeOH/DCM) to afford the pure product **14** as yellowish sticky oil (185 mg, 0.39 mmol, 95%).

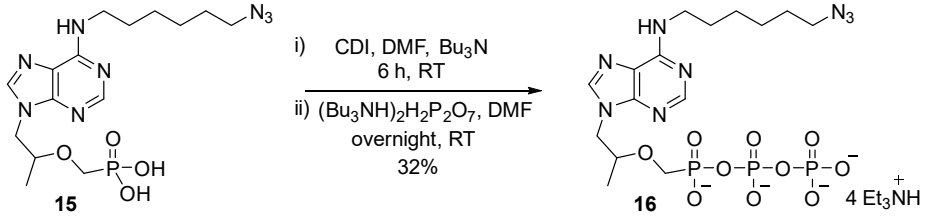
^{1}H NMR (400 MHz, $CDCl_3$): δ 8.36 (s, 1H), 7.88 (s, 1H), 5.77 (s, 1H), 4.32 (dd, *J* = 14.4, 3.0 Hz, 1H), 4.13 – 4.00 (m, 5H), 3.95 – 3.88 (m, 1H), 3.81 (dd, *J* = 13.6, 9.1 Hz, 1H), 3.64 – 3.54 (m, 3H), 3.25 (t, *J* = 6.9 Hz, 2H), 1.73 – 1.66 (m, 2H), 1.63 – 1.56 (m, 2H), 1.48 – 1.40 (m, 4H), 1.29 (t, *J* = 7.1 Hz, 3H), 1.27 – 1.22 (m, 6H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 154.7, 152.9, 148.9, 140.7, 119.0, 76.2 (d, *J* = 12.0 Hz), 63.5, 62.2 (d, *J* = 6.7 Hz), 62.1 (d, *J* = 6.7 Hz), 61.8, 51.1, 47.9, 40.4, 29.4, 28.6, 26.3, 16.3 (d, *J* = 8.1 Hz), 16.2 (d, *J* = 5.3 Hz).
^{31}P NMR (162 MHz, $CDCl_3$): δ 20.84.
HRMS (ESI^+): Exact mass calculated for $C_{19}H_{34}N_8O_4P$ [M + H]$^+$, 469.24351. Found 469.24351.

(((1-(6-((6-azidohexyl)amino)-9*H*-purin-9-yl)propan-2-yl)oxy)methyl)phosphonic acid (Azidohexyl-tenofovir, 15)

In a 25 ml Schlenk flask, compound **14** (100 mg, 0.21 mmol) was dissolved in 4 ml dry DCM and the solution was cooled in an ice bath. TMSBr (0.20 ml, 1.51 mmol) was added dropwise and the reaction mixture was stirred for 3 h at 0 °C. TLC confirmed the complete consumption of the reactant. The solvent was evaporated, co-evaporated with MeOH and Et_2O, before the resulting yellowish-white solid was dissolved in 20 ml water. After washing with DCM (4 x 3 ml), the aqueous fraction was lyophilized to afford compound **15** as a white solid in quantitative yield.

^{1}H NMR (400 MHz, D_2O): δ 8.24 (s, 1H), 8.22 (s, 1H), 4.32 (dd, *J* = 14.6, 3.0 Hz, 1H), 4.13 (dd, *J* = 14.8, 7.8 Hz, 1H), 3.87 – 3.79 (m, 1H), 3.68 (dd, *J* = 13.8, 8.8 Hz, 1H), 3.48 – 3.41 (m, 3H), 3.10 (t, *J* = 6.8 Hz, 2H), 1.65 – 1.50 (m, 2H), 1.42 – 1.32 (m, 2H), 1.31 – 1.13 (m, 4H), 1.06 (d, *J* = 6.3 Hz, 3H).
^{13}C NMR (101 MHz, D_2O): δ 148.2, 146.8, 144.6, 144.3, 117.4, 76.41 (d, *J* = 11.8 Hz), 63.08 (d, *J* = 161.8 Hz), 50.9, 48.5, 43.9, 42.2, 27.7, 27.3, 25.5, 15.8.
^{31}P NMR (162 MHz, D_2O): δ 19.65.
HRMS (ESI^-): Exact mass calculated for $C_{15}H_{24}N_8O_4P$ [M - H]$^-$, 411.16636. Found 411.16638.

[1-(6-((6-azidohexyl)amino)-9*H*-purin-9-yl)propan-2-yl]oxymethyl-[hydroxy(phosphonooxy)phosphoryl]oxy-phosphinic acid, triethylammonium salt (Azidohexyl-tenofovir-DP, 16)

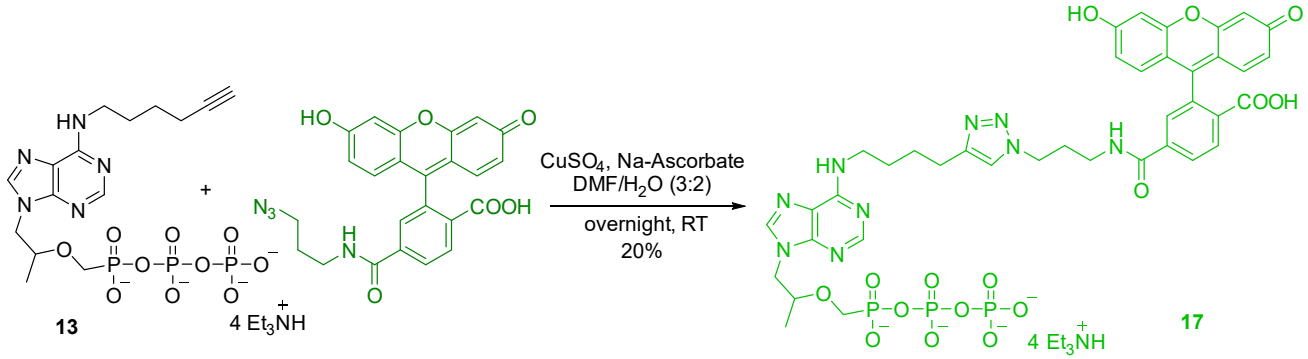


The free acid form of compound **15** (55 mg, 134 µmol) was suspended in 1 ml dry DMF and treated with tributylamine (128 µl, 536 µmol) for 1 h at RT. The clear solution of the resulting tributylammonium salt was evaporated, and dried by further co-evaporation with dry DMF (3 x 1 ml), before it was dissolved in 1.5 ml dry DMF. A solution of CDI (109 mg, 670 µmol) in 1.5 ml dry DMF was added dropwise, and the reaction mixture was stirred for 6 h at RT. Excess CDI was quenched by the addition of MeOH (22 µl, 536 µmol). After 1 h at RT, a solution of $(Bu_3NH)_2H_2P_2O_7$ (550 mg, 1 mmol) in 2 ml dry DMF was added, and the resulting white suspension was stirred overnight at RT. The supernatant was collected and separated from the precipitate (imidazolium pyrophosphate) by centrifugation. The solvent was evaporated, the residue was dissolved in 10 ml H_2O and extracted with DCM (3 x 2 ml). The aqueous solution was lyophilized and the crude product was purified by anion exchange chromatography (see general method section). The product-containing fraction (60-70% buffer B) was collected, the solvent was evaporated and the residue was subjected to purification by RP chromatography. The product **16** was obtained as a white foam (42.3 mg, 43 µmol, 32%).

RP-purification procedure: *Performed on:* GE Healthcare ÄKTAprime plus, *Column:* Lobar 310-25 LiChroprep RP-18 (40-63 µm), Merck *Detection wavelength:* 280 nm, *Solvent systems:* buffer A: 100 mM TEAB (pH = 7.5); buffer B: 100 mM TEAB in 80% ACN *Gradient:* 10-40% B in 30 min, 40% B for 10 min, 40-50% B in 10 min, 50% B for 10 min, *Flow rate:* 5 ml/min

^{1}H NMR (400 MHz, D_2O): δ 8.21 (s, 1H), 8.19 (s, 1H), 4.40 (dd, *J* = 14.7, 3.6 Hz, 1H), 4.25 (dd, *J* = 14.7, 5.6 Hz, 1H), 4.06 – 3.99 (m, 1H), 3.84 (dd, *J* = 13.3, 9.3 Hz, 1H), 3.74 (dd, *J* = 13.4, 9.3 Hz, 1H), 3.56 (br s, 2H), 3.27 (t, *J* = 6.8 Hz, 2H), 1.73 – 1.64 (m, 2H), 1.60 – 1.54 (m, 2H), 1.46 – 1.35 (m, 4H), 1.11 (d, *J* = 6.3 Hz, 3H).
^{31}P NMR (162 MHz, D_2O): δ 8.66 (d, *J* = 26.2 Hz), -10.36 (d, *J* = 20.2 Hz), -23.10 (t, *J* = 22.6 Hz).
HRMS (ESI$^-$): Exact mass calculated for $C_{15}H_{26}N_8O_{10}P_3$ [M - H]$^-$, 571.09902. Found 571.09905.

FAM-tenofovir-DP (17)

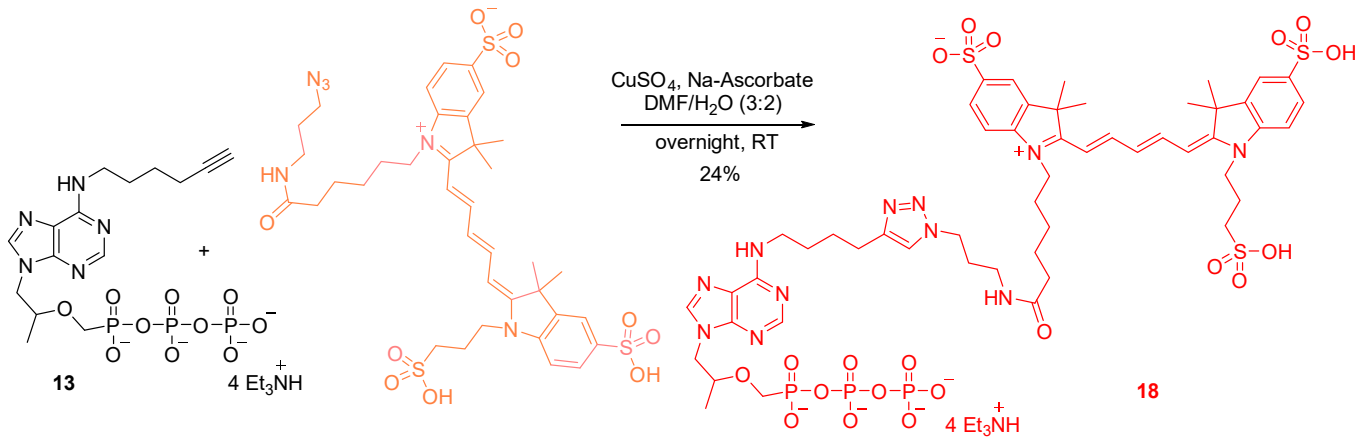


In a small 1.5 mL reaction tube 500 nmol compound **13** (10 µL of 50 mM in DMF; 1 eq) were combined with 10 µL 6-FAM azide (100 mM in DMF, 2 eq) and 10 µL DMF. In another 1.5 mL reaction tube, 5 µL $CuSO_4$ (100 mM in H_2O; 1 eq) was mixed with 15 µL of a freshly prepared solution of sodium ascorbate (200 mM in H_2O; 6 eq), and added to the first 1.5 mL reaction tube, to give a final volume of 50 µL (final DMF/H_2O 3:2). The reaction mixture was stirred overnight at RT. Then, 1.2 ml ice-cold 2% $NaClO_4$ in acetone was added and centrifuged (centrifuge at 4 °C, 15000 rpm for 30 min). The precipitate was dissolved in water and purified by reversed phase HPLC. The isolated product was lyophilized to obtained compound **17** as a red foam (100 nmol, 20%), which was dissolved in 100 µL water to prepare a 1 mM stock solution for further use.

RP-purification procedure: *Performed on:* GE Healthcare ÄKTAmicro, *Column:* Nucleosil 100-5 C18 column, 250 x 4.6 mm, Macherey-Nagel, *Detection wavelength:* 260 nm and 496 nm, *Solvent systems:* buffer A: 100 mM TEAA (pH = 7.0); buffer B: 100 mM TEAA in 80% CAN, *Gradient:* 5% B for 2.5 min, 5-20% B in 5 min, 20-22.5% B in 10 min, 22.5-50% B in 2.5 min, 50% B for 2.5 min, 50-100% B in 2.5 min, *Column oven temperature:* 40 °C, *Flow rate:* 1 ml/min

HRMS (ESI$^-$): Exact mass calculated for $C_{39}H_{41}N_9O_{16}P_3$ [M - H]$^-$, 984.18896. Found 984.18897.
Exact mass calculated for $C_{39}H_{40}N_9O_{16}P_3$ [M - 2H]$^{2-}$, 491.59084. Found 491.59459.

Cy5-tenofovir-DP (18)

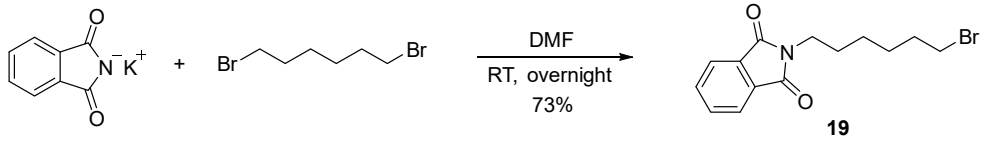


In a small 1.5 mL reaction tube 500 nmol compound **13** (10 µL of 50 mM in H_2O; 1 eq) were combined with 10 µL Sulfo-Cy5 azide (100 mM in H_2O, 2 eq) and 60 µL DMF. In another 1.5 mL reaction tube, 5 µL $CuSO_4$ (100 mM in H_2O; 1 eq) were mixed with 15 µL of a freshly prepared solution of sodium ascorbate (200 mM in H_2O; 6 eq), and the mixture was added to the first 1.5 mL reaction tube, to give a final volume of 100 µL (final DMF/H_2O 3:2). A small magnetic stir bar was put into the 1.5 mL reaction tube and the mixture was stirred overnight at RT. The solution was split into two 1.5 mL reaction tubes (2 x 50 µL) and 1.2 ml ice-cold 2% $NaClO_4$ in acetone was added to each tube. After centrifugation (at 4 °C, 15000 rpm for 30 min), the precipitate was dissolved in water and purified by reversed phase HPLC. The compound **18** was obtained as a dark blue foam (120 nmol, 24%), which was dissolved in 120 µL water to make a 1 mM stock concentration for further use.

RP-purification procedure: *Performed on:* GE Healthcare ÄKTAmicro, *Column:* Nucleosil 100-5 C18 column, 250 x 4.6 mm, Macherey-Nagel, *Detection wavelength:* 260 nm and 647 nm , *Solvent systems:* buffer A: 100 mM TEAA (pH = 7.0); buffer B: 100 mM TEAA in 80% CAN, *Gradient:* 5% B for 2.5 min, 5-20% B in 2.5 min, 20-40% B in 22.5 min, 40-100% B in 2.5 min, *Column oven temperature:* 40 °C, *Flow rate:* 1 ml/min

HRMS (ESI$^-$): Exact mass calculated for $C_{52}H_{70}N_{11}O_{20}P_3S_3$ $[M - 2H]^{2-}$, 678.65923. Found 678.65925.
Exact mass calculated for $C_{52}H_{69}N_{11}O_{20}P_3S_3$ $[M - 3H]^{3-}$, 452.10373. Found 452.10752.

2-(6-bromohexyl)isoindoline-1,3-dione (19)

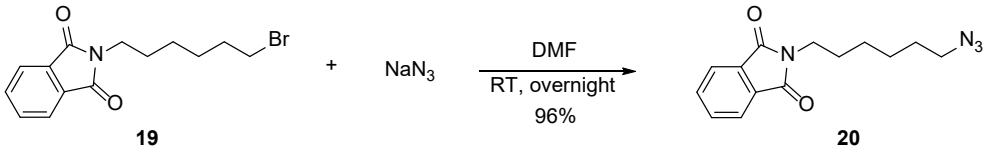


Compound **19** was synthesized according to a modified published procedure.[7]

In a 250 ml Schlenk flask, 1,6-Dibromohexane (15.6 ml, 101.41 mmol) was dissolved in 75 ml dry DMF, and potassium phthalimide (8 g, 43.19 mmol) was added in small portions over 30 min. The reaction mixture was stirred overnight at RT, and then the solvent was evaporated. The white residue was put onto celite and washed with 200 ml EtOAc. The filtrate was evaporated and the crude product was purified by flash chromatography (15% EtOAc/Hexane), affording compound **19** as a white solid (9.72 g, 31.34 mmol, 73%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.84 – 7.82 (m, 2H), 7.72 – 7.69 (m, 2H), 3.70 – 3.66 (m, 2H), 3.39 (t, *J* = 6.8 Hz, 2H), 1.88 – 1.81 (m, 2H), 1.72 – 1.65 (m, 2H), 1.51 – 1.44 (m, 2H), 1.40 – 1.32 (m, 2H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 168.4, 133.9, 132.1, 123.2, 37.8, 33.7, 32.6, 28.3, 27.7, 26.0.
HRMS (ESI$^+$): Exact mass calculated for $C_{14}H_{17}BrNO_2$ $[M + H]^+$, 310.04372. Found 310.04457.

2-(6-azidohexyl)isoindoline-1,3-dione (20)

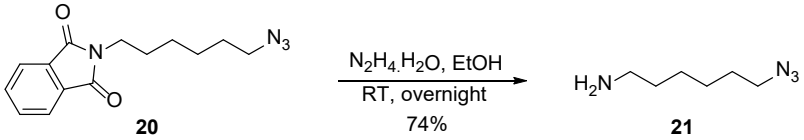


Compound **20** was synthesized in analogy to a published procedure.[7]

In a 100 ml round-bottom flask, compound **19** (9 g, 29.01 mmol) was dissolved in 30 ml DMF, followed by the addition of NaN_3 (5.66 g, 87.03 mmol) in portions over 1 h at RT. The reaction mixture was stirred overnight at RT. Then the solvent was evaporated and 150 ml DCM was added to the residue, followed by the extraction with water (4 x 50 ml) and brine (50 ml). The organic layer was dried over Na_2SO_4, and after evaporation of the solvent, the title compound **20** was obtained as colorless sticky liquid which turned solid (7.58 g, 27.84 mmol, 96%) on cooling with ice water. Compound **20** was used for the next reaction without further purification.

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.85 – 7.81 (m, 2H), 7.72 – 7.68 (m, 2H), 3.69 – 3.66 (m, 2H), 3.24 (t, *J* = 6.9 Hz, 2H), 1.72 – 1.64 (m, 2H), 1.62 – 1.55 (m, 2H), 1.45 – 1.32 (m, 4H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 168.4, 133.9, 132.1, 123.1, 51.3, 37.8, 28.6, 28.4, 26.3, 26.2.
HRMS (ESI$^+$): Exact mass calculated for $C_{14}H_{16}N_4NaO_2$ $[M + Na]^+$, 295.11655. Found 295.11656.

6-azidohexan-1-amine (21)



Compound **21** was synthesized according to a modified published procedure.[7]

In a 100 ml round-bottom flask, compound **20** (4 g, 14.69 mmol) was dissolved in 20 ml EtOH. To that solution, hydrazine hydrate (3.7 g) was added dropwise and the reaction was stirred overnight at RT. The reaction mixture turned to a white solid, which was crashed into small parts, followed by the addition of 75 ml EtOAc. After stirring for 1 h at RT, the reaction mixture was filtered and the white residue was washed with 200 ml EtOAc. The filtrate was evaporated and the remaining (yellowish liquid) was redissolved in 50 ml EtOAc, and treated with 50% aq. HCl. The clear solution turned cloudy after acidification. Then the acidic EtOAc solution was extracted with H_2O (4 x 15 ml). Aqueous fractions were combined and basified with concentrated aq. NaOH solution until the formation of liquid droplets was observed. The basified aq. solution was extracted with DCM (4 x 100 ml). The organic layers were combined and again extracted with H_2O (50 ml) followed by brine (50 ml). The DCM solution was dried over Na_2SO_4, followed by solvent evaporation to yield compound **21** as yellowish liquid (1.54 g, 10.83 mmol, 74%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 3.24 (t, *J* = 6.9 Hz, 2H), 2.67 (t, *J* = 7.0 Hz, 2H), 1.61 – 1.54 (m, 2H), 1.46 – 1.30 (m, 7H).
^{13}C NMR (101 MHz, $CDCl_3$): δ 51.4, 42.1, 33.6, 28.8, 26.6, 26.5.
HRMS (ESI$^+$): Exact mass calculated for $C_6H_{15}N_4$ $[M + H]^+$, 143.12912. Found 143.12917.

2.2. In vitro selection of ribozymes using Biotin-tenofovir-DP

The in vitro selection was performed in analogy to our previously described procedure for identification of FH ribozymes.[1] The conditions for selection of FJ ribozymes are summarized as follows: In the first round of selection, 3 nmol of the N40 pool mixed with ~150 pmol of the 3'-lucifer yellow labeled N40 pool was dissolved in 60 µL of the selection buffer (50 mM HEPES, pH 7.5, 120 mM KCl, 5 mM NaCl) including 40 mM $MgCl_2$ and 300 µM of Biotin-Tenofovir-diphosphate (**7**). Following incubation overnight at 37°C, the unreacted biotin-tenofovir substrate was removed using ethanol precipitation. The reactive species were then immobilized on 1 mg of neutravidin beads (pre-treated with *E.coli* tRNA), and the bulk of the unreactive species in the pool were washed away using denaturing wash buffer. The beads were then suspended in formamide elution buffer and placed at 95°C for 7 min. The supernatant of this step was then subjected to ethanol precipitation. The eluted RNA was reverse transcribed and subjected to 10 rounds of PCR amplification. The primers in this PCR step do not encompass the modification site, therefore providing equal amplification opportunity to the modified RNA as the unmodified inactive carry-over RNA. A 2nd PCR was then performed (18-30 rounds) with an alternative primer which restores the substrate sequence to its full length and adds the T7 promoter. The product of the 2nd PCR was then transcribed in vitro with T7 RNA polymerase to generate the RNA pool entering the next selection round.

In all the subsequent selection rounds the pool concentration was maintained at ~60 µM, including ~150 pmol of 3'-LY labeled RNA pool of the same round. In the following selection rounds the amount of RNA pool and the reaction volume used was adjusted based on the yield of the previous round and amount of beads used in the capture step was kept at 0.1 mg of the beads per 100 pmol of RNA pool. To prevent selection of neutravidin binders from the pool, the affinity matrix was alternated between neutravidin and streptavidin every two rounds. The frequency of the active variants was estimated by comparing the fluorescence intensity of the eluted RNA at the end of the round to that of the total pool prior to the capture step. First sign of enrichment became apparent at the end of the 4th selection round were ~ 0.44% of the pool was retained. A rise in the activity level to ~1.42 % in the next round further confirmed the success of the enrichment process. The incubation time was reduced at round 8 from overnight to 4 h which was further reduced to 1 h by the 12th round. A streptavidin gel shift assay on the outcome of the round 12 showed over 80% activity after overnight incubation. The pool also demonstrated trans-activity when the substrate sequence was removed from the pool and added as a separate entity to the reaction. At this point the pool was cloned (using a TOPO-TA cloning kit) and 10 representative colonies were subjected to Sanger sequencing. Two unique sequences were identified which were denoted as FJ1 and FJ8.

The pool was then further examined for additional variants by subjecting it to Next Generation Sequencing (NGS). To prepare the NGS libraries, RT-PCR samples from round 7 and round 12 were subjected to two consecutive PCR amplifications during which the unimolecular identifiers (UMIs) and indices were added. The sample from the last PCR was agarose gel purified and submitted to the Core Unit Systems Medicine (Uni Würzburg) for amplicon sequencing (NextSeq-500 HighOutput 75nt single end). New sequence variants were identified in the demultiplexed and trimmed fastq files (see below, section 3.1.2), three of which (denoted as FJC1, FJC3 and FJC9) were chosen for further characterization.

2.3. Characterization and application of FJ ribozymes

2.3.1 Single-turnover kinetic experiments

Kinetic experiments were performed as previously reported.[1] Briefly, a 10:1 ratio of the ribozyme : fluorescently labeled substrate sequence were dissolved in selection buffer resulting in 10 µM and 1 µM final concentrations of ribozyme and the substrate sequence (i.e. single-turnover conditions). The reactions were performed in the presence of 40 mM $MgCl_2$ and 300 µM of the Tenofovir-diphosphate analog substrate. The reaction was incubated at 37°C and 1 µL aliquots were taken from the experiment at desired time-points and quenched by adding to 4 µL of stop solution. Half of the resulting samples were resolved on 20% analytical denaturing PAGE for 1 h under constant power of 25 W for 25 cm plates and 35 W for 30 cm plates. For kinetic experiments in which different tenofovir analogs were tested the running time was increased to 1.5 h to achieve proper resolution. The gels were imaged on a Chemidoc device and band intensities were quantified by ImageLab software. The fraction of labeled product was calculated, and the timecourse fit to single-exponential kinetics according to $Y = Y_{max}(1-\exp(-k_{obs}t))$.

2.3.1.1 Target sequence requirement of FJ ribozymes

To investigate target sequence generality of FJ ribozymes, transition or transversion mutations were introduced on the original parent substrate sequence, initially outside the GAG modification context. FJ ribozymes were also designed with complementary arms to these sequences. Single turn-over kinetic experiments were then performed to determine the capacity of the ribozymes for modifying broader range of target sequences. Since FJ1 and FJ8 ribozymes were unable to modify the extensively mutated substrate sequence variants, a new set of mutants were designed to further investigate their modification context requirement. In one of the mutants the unchanged nucleotides was expanded by one nucleotide from each side of the GAG context, while the rest of the sequence was subjected to transition mutations, resulting in TM-UGAGC, in which UGAGC is identical to the parent substrate sequence. Another sequence was also designed in which the segment 5' to the GAG was transversion mutated while the 3'-segment was kept unchanged resulting in 5'TV1-3'parent sequence. Single-turnover kinetic experiments using these substrates determined the C nucleotide following the GAG context is essential. Further investigations determined the minimal modification context as 5'-**A**GC-3'. Optimal modification efficiency was obtained with sequences containing 5'-R**A**GCY-3' as the modification site context sequence.

2.3.1.2 Tenofovir-Diphosphate analog scope of FJ ribozymes

To determine nucleotide analog specificity of FJ ribozymes, unmodified tenofovir-diphosphate, N^6-hexynyl-tenofovir-diphosphate and N^6-azidohexyl-tenofovir diphosphate were employed in single turnover kinetic assays. None of the tested variants showed any dependency to the biotin moiety or the C6 linker as all the tested derivatives including the unmodified tenofovir-DP were efficiently ligated to the substrate sequence.

2.3.2. Preparative tenofovir labeling and probing of the modification site

2.3.2.1 Preparative modification of the substrate sequence with biotin-tenofovir

One nmol of 5'-flourescein labeled synthetic substrate sequence was mixed with 1.3 nmol of the FJ1 ribozyme in a total volume of 25 µL of selection buffer including 40 mM $MgCl_2$ and 500 µM of the N^6-biotin-TenDP substrate. The sample was incubated at 37°C overnight, followed by quenching via addition of 25 µL of high-dye solution. The sample was the resolved on a 20% denaturing PAGE and the band corresponding to the modified RNA was excised from the gel. The modified RNA product was then extracted from the gel and ethanol precipitated resulting in an isolated yield of 55%.

2.3.2.2 Probing of the modification site

Alkaline hydrolysis and T1 probing was used to locate the exact position of the modification site in the substrate sequence. For alkaline hydrolysis, 10 pmol of the modified or unmodified 5'-fluorescently labeled synthetic RNA substrate was incubated in 10 µL of 50, 25 or 10 mM NaOH and incubated at 95°C. Two µL aliquots were taken from each reaction after 1, 2.5 and 5 min of incubation and quenched by adding to 3 µL of stop solution and placing on ice. For T1 digestion, 10 pmol of the modified or unmodified RNA substrate were dissolved in 5 µL of 50 mM Tris pH= 7.5 containing 0.1 U/µL of RNase T1. The sample was placed at 37°C and quenched after 30 sec by adding 5 µL of stop solution and placing on ice.
2.5 µL of each of the timepoint samples and T1 probing samples were resolved on a 20% analytical denaturing PAGE for 1 h under constant power of 35 W and subjected to fluorescent imaging.

2.3.2.3. Debranching assay

5 pmol of 5'-fluorescein labeled substrate sequence modified using FJ1 or FH14 was dissolved in 5 µL of Dbr1 reaction buffer (50 mM Tris pH = 7.4, 25 mM NaCl, 2.5 mM DTT, 0.01% (V/V) Tween-20, 0.15% (V/V) glycerol) including 1 or 5 µM $MnCl_2$ and 0.5 ng/µL of recombinant Dbr1 (kindly provided by A. Hoskins, Univ. of Wisconsin). The sample was incubated at 37°C and 1 µL timepoint samples were taken at 0, 15, 30 and 60 min and quenched by adding to 4 µL of stop solution and placing in liquid nitrogen. 2.5 µL of each of these samples were then resolved on a 20% analytical denaturing PAGE for 1 h under constant power of 25W. The gel was then subjected to fluorescent imaging.

2.3.3 Application of FJ ribozymes for labeling of cellular RNA

Total cellular RNA from TOP10 E.coli cells was isolated as previously described.[1, 8]
For the experiments in which 16S and/or 23S rRNA were labeled, 20 ng of total cellular RNA from Top10 E. Coli cells (NEB) was mixed with 50 pmol of the corresponding ribozyme(s). The reaction was performed in 5 µL of the selection buffer including 40 mM $MgCl_2$ and 200 µM of the fluorescently labeled TenDP and/or ATP analogs. After 6h incubation at 37°C, the reaction was subjected to 2x isopropanol precipitation. The resulting pellet was washed using ice cold 70% ethanol. The dried pellet was dissolved in 2 µL of stop solution and 1 µL of milliQ water. The sample was then resolved on 1.3% agarose gel under 90 V for 45 min. The gel was then subjected to fluorescent imaging before and after staining with Sybr gold.

In case of 5S rRNA labeling using FJC9 ribozyme, 50 ng of the total cellular RNA was mixed with 100 pmol of the corresponding FJC9 ribozyme. Reaction was done in 5 µL of the selection buffer including 40 mM $MgCl_2$ and 200 µM, N^6-Fluorescein-TenDP. The reaction was incubated at 37°C for 6 h. Afterwards, the reaction was subjected to ethanol precipitation and the resulting pellet was dissolved in 2 µL of milliQ water and 2 µL of the stop solution. The sample was resolved on a 10% analytical denaturing gel, next to the fluorescein labeled in vitro transcribed 5S rRNA as size marker. The gel was then subjected to fluorescent imaging before and after staining with Sybr gold.

2.3.4 Mutually orthogonal dual-color labeling of a synthetic transcript

A synthetic substrate was designed by tandem fusion of the parent substrate sequence and one of its derivatives via a 4 nt linker. The 41 nt RNA was prepared by in vitro transcription with T7 RNA polymerase. 20 pmol of this substrate sequence was mixed with 50 pmol of each ribozyme in a total volume of 5 µL of selection buffer containing 40 mM $MgCl_2$ and 200 µM of either Cy5-TenDP or 6-FAM-ATP or both. The FJ1 ribozyme was targeted towards the parent substrate sequence while the FH14 was designed to target the mutated segment. Time-point samples (0.5 µL) were taken at 0, 5 h and after overnight incubation and quenched by adding to 99.5 µL of TEN buffer. The samples were then subjected to ethanol precipitation, and the pellet was dissolved in 5 µL of the stop solution. 2.5 µL of these samples were then resolved on 15% analytical denaturing PAGE for 45 min and subjected to dual-channel fluorescent imaging.
To confirm the specificity of each ribozyme for its cognate NTP analog and its target sequence an 8-17NG deoxyribozyme was designed the cleave the double-labeling RNA substrate asymmetrically resulting in a 24 nt fragment expected to harbor the label installed by FJ1 (Cy5-tenofovir branch) and a 17 nt fragment supposed to carry the label attached by FH14 (6-FAM-AMP branch). To perform the cleavage experiment, an aliquot (3.5 µL) of the labeling reactions explained above was subjected to ethanol precipitation. The pellet was then dissolved in 10 µL of 8-17NG reaction buffer containing 50 mM HEPES (pH= 7.5), 400 mM KCl, 100 mM NaCl, 10 mM $MnCl_2$, 10 mM $MgCl_2$ and 200 pmol of the corresponding 8-17NG deoxyribozyme. The reaction was then incubated at 37°C and 1.4 µL timepoint samples were taken at 0, 0.5 and 1 h and mixed with 3.57 µL of the stop solution. 2.5 µL of these samples were resolved on the same gel as the double labeling reaction.

2.3.5. Probing of the modification site on the target cellular RNA using reverse transcription

2.3.5.1. Modification of the cellular RNA for primer extension experiments

200 ng of *E. coli* total cellular RNA and 20 pmol of individual ribozymes were dissolved in a total volume of 10 µL of the selection buffer including 40 mM $MgCl_2$ and the ribozyme's cognate substrate. For FH14 type ribozyme N^6-aminohexyl-ATP- biotin (200 µM final concentration) was used, and for the FJ1 type ribozyme Biotin-Tenofovir-DP-biotin (**7**, 340 µM final concentration). The reactions were incubated at 37°C for 5 h and were subjected to ethanol precipitation. The pellet was dried and directly used in the primer extension reaction.

2.3.5.2. ^{32}P-Labeling of the primers

Primers (100 pmol) were dissolved in 10 µL of PNK buffer A including 5 µCi of γ-^{32}P-ATP, and 5 units of PNK. The reaction was incubated at 37°C for 1 h, followed by two rounds of ethanol precipitation. The pellet was dissolved in 10 µL of milliQ water, resulting in a sample with 2000-3500 IPS.

2.3.5.3. Primer extension protocol

5 pmol of the ^{32}P-labeled primer and 200 ng of the ribozyme-modified or non-modified total cellular RNA were dissolved in 5 µL of the annealing buffer (Tris 5 mM pH= 7.5, EDTA 0.1 mM). The sample was then placed at 95°C for 3 minutes, followed by incubation at room temperature for 10 minutes. 2 µL of the 5X First strand synthesis buffer (Invitrogen), 0.5 µL of 0.1 M DTT, 0.5 µL of 10 mM dNTP mix and 50 units of superscript III reverse transcriptase was then added to the reaction and the final volume was then adjusted to 10 µL by adding milliQ water. The reaction was then incubated at 55°C for 1 hour. 1 µL of 2 N NaOH was then added to the reaction and the sample was incubated for 5 minutes at 95°C. The reaction was then subjected to ethanol precipitation. The dried pellet was then dissolved in 10 µL of high-dye loading buffer. 2.5 µL of this sample was then resolved using 15%, analytical 45 cm long PAGE, under constant power of 45 W for 2 hours and 15 minutes. The gel was dried at 80°C under vacuum for 30 minutes and exposed to the phosphorus screen overnight. The screen was then scanned on a Typhoon phosphorus imager.
The sequencing reactions were performed by annealing 5 pmol of the ^{32}P-labeled primers to 200 ng of the non-modified total cellular RNA in 5 µL of the annealing buffer. for the A sequencing ladder ddTTP, was added to the reaction to a final concentration of 0.5 mM and dTTP at 0.05 mM. The rest of the dNTP were adjusted to a final concentration of 0.5 mM. For G sequencing ladder ddCTP was included in the reaction at a concentration of 0.5 mM and dCTP at 0.05 mM. The rest of the dNTPs had a final concentration of 0.5 mM. All the other reagents such as the buffer conditions and enzyme units were the same as the primer extension protocol.

3. Results and Discussion

3.1 Sequencing results of FJ ribozyme selection

3.1.1. FJ ribozyme sequences identified by Sanger sequencing.

Forty colonies were checked by colony-PCR, 10 plasmids were isolated and sequenced

FJ1 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ2 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ3 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ4 TGGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ24 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ32 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ35 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC

FJ8 YTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGTATGTCC
FJ11 TTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGTATGTCC

3.1.2. FJ ribozyme sequences identified by Illumina sequencing.

After quality control and trimming of illumina adapters the sequencing data were analyzed by fastaptamer.[9]

fastaptamer_count: FJ_Round 7: 4.789.062 total reads, 288.126 unique sequences
FJ_Round 12: 1.138.422 total reads, 66.443 unique sequences

fastaptamer_cluster: FJ_Round 12

Cluster	Unique	#Reads	RPM
1	1407	866354	761012.5
2	407	84781	74472.2
3	259	55156	48449.4
4	158	21523	18905.9
5	28	3697	3247.4
6	38	3943	3463.5
7	20	2733	2400.6
8	26	2401	2109.0
9	10	1268	1113.8
10	6	720	632.4

fastaptamer_compare: RPM(x) = Round 7 and RPM(y) = Round 12
First sequence of first ten clusters ordered by log2 (enrichment)

Cluster	sequence	RPM(x)	RPM(x)	log2 x/y
>2-**2**-1-0=FJ8	AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT	38.8	33325.9	**9.74**
>8-**4**-1-0=FJ1	AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA	25.7	10479.4	**8.67**
>82-**9**-1-0=FJC9	AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT	46.2	924.1	**4.32**
>1-**1**-1-0=FJC1	AACCAGCCTACCATCCTTGAAGGC GAAACGTGTCACATAAGAAAACGGTAAACTAGCAAGTTCC CAGT	130185	502593	1.95
>48-**7**-1-0 (a)	AACCAGCCTACCATCCTTGAAGGC T CAGTATGTC GACCTTCTGTCTCTTATACACATCTCCGAGCCCA	474.2	1586.4	1.74
>26-**5**-1-0 (b)	AACCAGCCTACCATCCCTGAAGGC CGAATGCCACCGAA CAGTATGTC GACCTTCTGTCTCTTATACAC	3096.4	2437.6	-0.35
>3-**3**-1-0=FJC3	AACCAGCCTACCATCCCTGAAGGC CGAATGCCACCGAACCGTATAATTGCCGCCTCCAATTTC CAGTA	39964.6	30094.3	-0.41
>56-**8**-1-0	AACCAGCCTACCATCCTTGAAGGC ACTGGTGTTACAAGTAAACGCAACCTTAATTTTACGATCC CAGT	4264.5	1258.8	-1.76
>32-**6**-1-0 (c)	AACCAGCCTACCATCCTTGAAGGC ATCGGTGTAACATAAACGAAACCTTAGGTCCCTTGGCCCC CAGT	34139.0	2206.6	-3.95
>126-**10**-1-0	AACCAGCCTACCATCCTTGAAGGC ATAGTTTCACAATAGCCACTTTAAGCTTATCCAAAGCTCA CAGT	11488.7	577.1	-4.32

(a) = extended complement to substrate sequence
(b) reveals a fully complementary binding arm after 14 nt in the center of N40 core
(c) related to FJ8

SUPPORTING INFORMATION

```
FJ8  (407 sequences)                                YTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGTATGTCC
>2-37939-33325.95-2-1-0    AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>5-18208-15994.07-2-2-1    AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAGTACCTTCCCC CAGT
>52-1677-1473.09-2-3-2     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CA
>91-893-784.42-2-4-2        ACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGTA
>97-855-751.04-2-5-1       AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCA CAGT
>101-840-737.86-2-6-3      AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAGTACCTTCCCC CA
>108-800-702.73-2-7-1      AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAATACCTTCCCC CAGT
>122-673-591.17-2-8-3      AACCAGGCCTACCATCCTTGAAGGCACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CA
>141-519-455.89-2-9-4      AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGAC  TACCTTCCCC CAGTAT
>153-468-411.10-2-10-1     AACCAGCCTACCATCCTTAAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>161-417-366.30-2-11-1     AACCAGCCTACCATCCCTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>162-415-364.54-2-12-1     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCCTCCCC CAGT
>167-404-354.88-2-13-3      ACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAGTACCTTCCCC CAGTA
>170-387-339.94-2-14-2     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAGTACCTTCCCA CAGT
>184-341-299.54-2-15-1     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAGACCTTAGGACATTACCTTCCCC CAGT
>198-300-263.52-2-16-1     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACAAAACCTTAGGACATTACCTTCCCC CAGT
>200-286-251.22-2-17-1     AACCAGCCTACCATCCTCGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>202-285-250.35-2-18-4     AACCAGGCCTACCATCCTTGAAGGCACGGTGTTACAAAACGAAACCTTAGGACAGTACCTTCCCC CA
>214-255-223.99-2-19-1     AACCAGCCTACCATCCTTGAAGGC ATGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>217-252-221.36-2-20-1     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACGAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>219-247-216.97-2-21-2     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAGAACCTTCCCC CAGT
>223-239-209.94-2-22-1     AACCAGCCTACCATCCTTGAAGGC ACGGTGCTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
>227-235-206.43-2-23-2     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC  AGTA
>228-232-203.79-2-24-1     AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAGCCTTAGGACATTACCTTCCCC CAGT
...
>2192-11-9.66-2-405-2      AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGAAAGTACCTTCCCC CAGT
>2192-11-9.66-2-406-3      AACCAGCCTACCCATCCTTGAAGGCACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CA
>2192-11-9.66-2-407-2      AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATT CCTTCCCC CAGTA

FJ1  (158 sequences)                                TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
>8-11930-10479.42-4-1-0    AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>63-1338-1175.31-4-2-1     AACCAGCCTACCATCCCTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>64-1316-1155.99-4-3-1     AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGT
>128-653-573.60-4-4-1      AACCAGCCTACCATCCTGGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>150-471-413.73-4-5-1      AACCAGCCTACCATCCTCGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>191-319-280.21-4-6-2       ACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTAT
>216-253-222.24-4-7-2      AACCAGGCCTACCATCCTTGAAGGCCCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGT
>230-228-200.28-4-8-2      AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAAACTGAAGATCCTTTGGCAAGGGTCTACAGT
>283-165-144.94-4-9-2      AACCAGCCTACCATCCCTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGT
>297-158-138.79-4-10-1     AACCAGCCTACCATCCTTGAAGGC CCACCCTTATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>302-156-137.03-4-11-1     AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTAAAGATCCTTTGGCAAGGGTCTA CAGTA
>347-127-111.56-4-12-1     AACCAGCCTACCATCCTTAAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>355-121-106.29-4-13-1     AACCAGCCTACCATCCTTGGAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>417-96-84.33-4-14-2       AACCAGCCTACCATCCTGGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGT
>417-96-84.33-4-15-1       AACCAGCCTACCATCCTTGAAGGC CCATCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>427-93-81.69-4-16-2       AACCAGCCTACCATCCT GAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTAT
>433-92-80.81-4-17-1       AACCAGCCTACCATCCTTGAAGGC CCACTCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>436-91-79.94-4-18-1       AACCAGCCTACCATCCTTGAAGGC CCACCCCCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA
>502-75-65.88-4-19-3       AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CA
>502-75-65.88-4-20-1       AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAGGATCCTTTGGCAAGGGTCTA CAGTA
...
>2192-11-9.66-4-156-1      AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCATTTGGCAAGGGTCTA CAGTA
>2192-11-9.66-4-157-2      AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CGGT
>2192-11-9.66-4-158-1      AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTC

FJC9  (10 sequences)
>82-1052-924.09-9-1-0      AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT
>701-50-43.92-9-2-1        AACCAGCCTACCATCCTTGAAGGC ACAAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT
>918-33-28.99-9-3-2        AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CA
>981-30-26.35-9-4-1        AACCAGCCTACCATCCTTGAAGGC ACTAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT
>1146-24-21.08-9-5-3       AACCAGGCCTACCATCCTTGAAGGCACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CA
>1391-19-16.69-9-6-1       AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCA CAGT
>1457-18-15.81-9-7-2        ACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGTA
>1520-17-14.93-9-8-1       AACCAGCCTACCATCCCTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT
>1927-13-11.42-9-9-1       AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACGCTTTAGCGAATTGGGCATCC CAGT
>2058-12-10.54-9-10-1      AACCAGCCTACCATCCTTAAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT
```

Partial alignment of N40 core:

```
FJC1  GAAACGTG-TCACATAA--GAAAA-CGGTAAACTAGCAAGTTCC
FJC3 CGAA---TGCCACCG-AACCGTATAA-TTGCCGCCTCCAATTTC
FJC9     ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC

FJ8  A-CGGTGTTACA---AAACGAAACCTTAGGACATTACCTTCCCC
FJC6 ATCGGTGTAACAT--AAACGAAACCTTAGG--TCCCTTGGCCCC
FJC8 ACTGGTGTTACAAGTAAACGCAACCTTA----ATTTTACGATCC
```

3.2 Supporting Figures

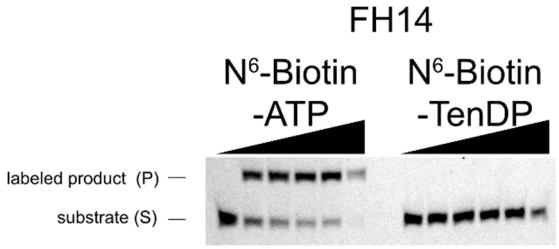


Figure S1. Biotin-Tenofovir diphosphate is not a substrate for FH14. (3'-fluorescein-labeled parent substrate RNA; 200 μM *N*^{6}-Biotin-ATP, 800 μM *N*^{6}-Biotin-Ten-DP, with 40 mM $MgCl_2$, in selection buffer. Overnight incubation at 37°C.

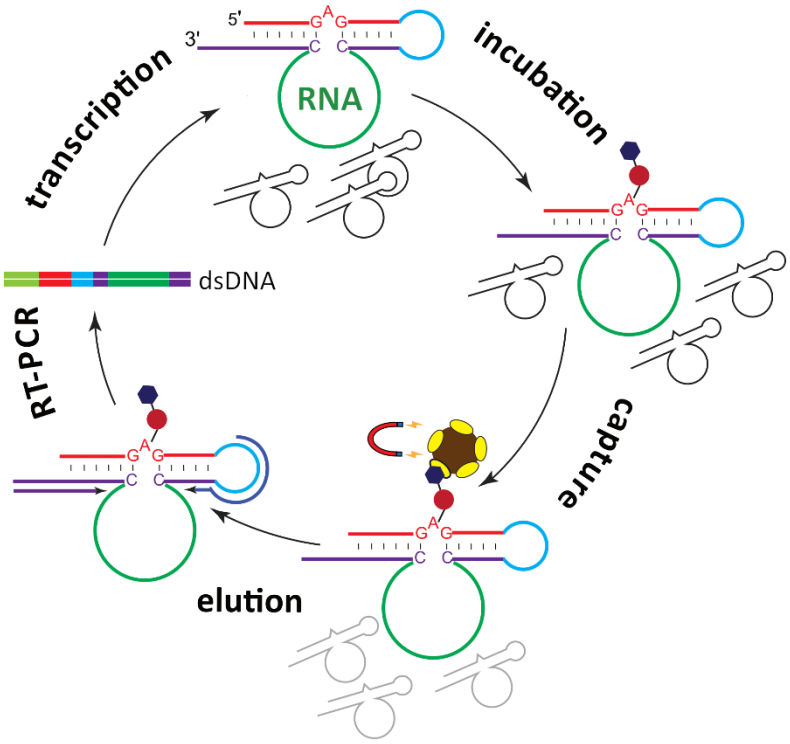
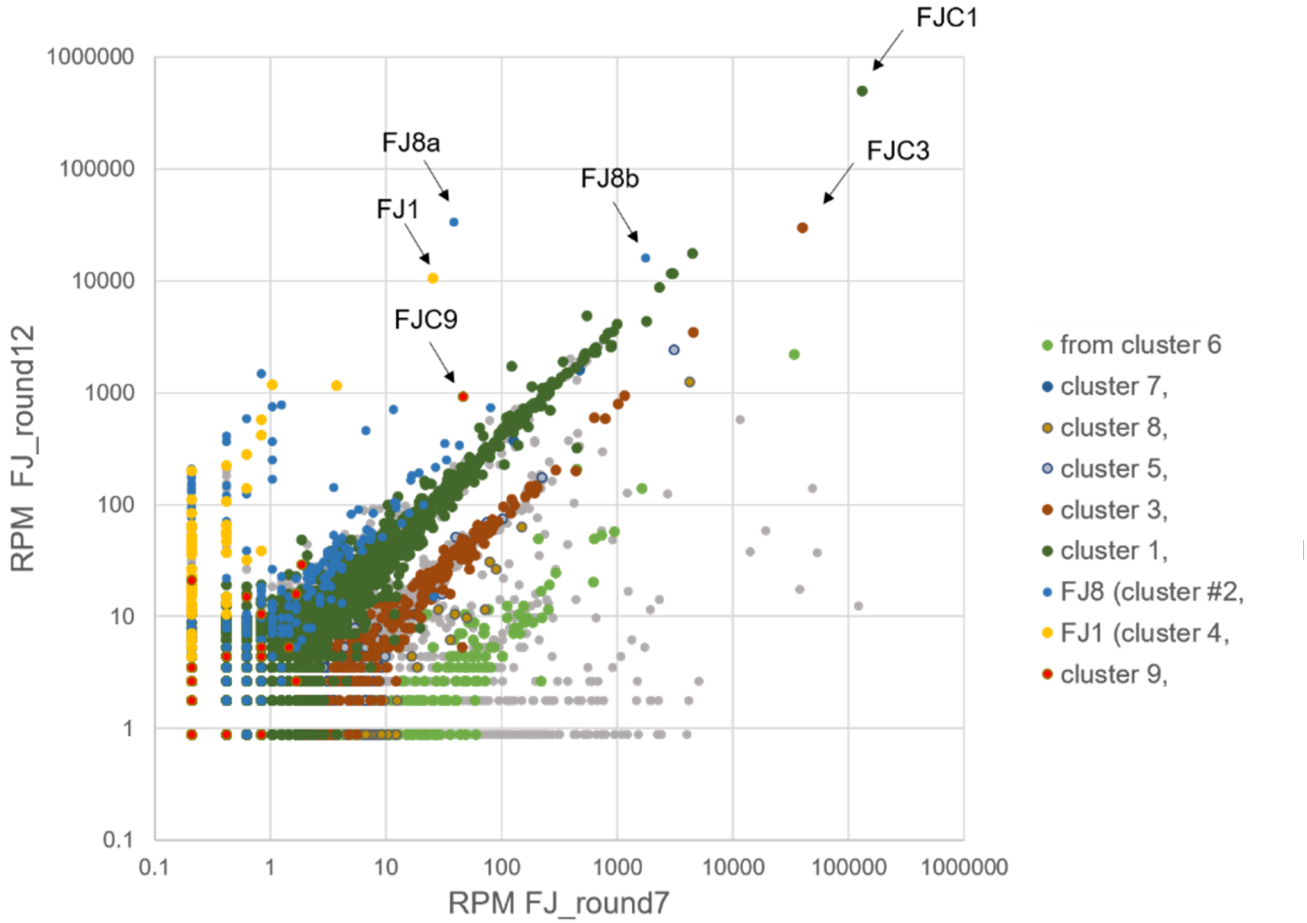


Figure S2. In vitro selection scheme, showing RNA substrate sequence in red, random region (N_{40}) in green, connecting loop in cyan, and in purple are the ribozyme binding arms complementary to the substrate sequence (flanking the bulged A). The primer binding sites for the first PCR are indicated. A second PCR generated the dsDNA template containing the T7 promoter (light green) necessary for transcription of the RNA.




```
FJC1 TTGAAGGC GAAACGTGTCACATAAGAAAACGGTAAACTAGCAAGTTCC CAGT
FJC3 CTGAAGGC CGAATGCCACCGAACCGTATAATTGCCGCCTCCAATTTC CAGTA
FJC9 TTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT

FJ8a TTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT
FJ8b TTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACAGTACCTTCCCC CAGT
```

Figure S3. Expanded version of Figure 2b with legend of color-coded clusters from fastaptamer_cluster FJ_round12. See also Section 3.1.2

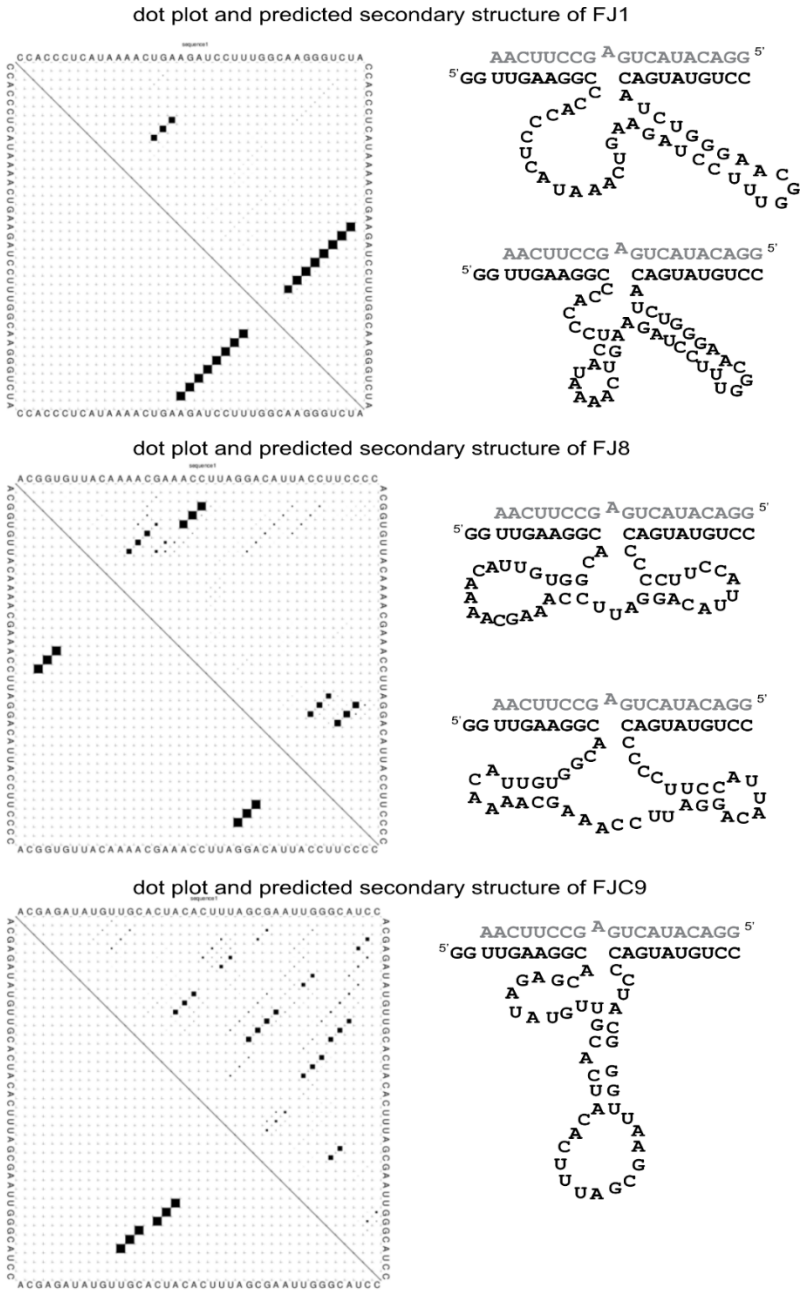


Figure S4. Secondary structures and dotplots of FJ1, FJ8 and FJC9. Dot plots were generated by Vienna RNA package.[10] The minium free energy secondary structure is drawn; for FJ1 and FJ8 one possible alternative secondary structure is also shown.

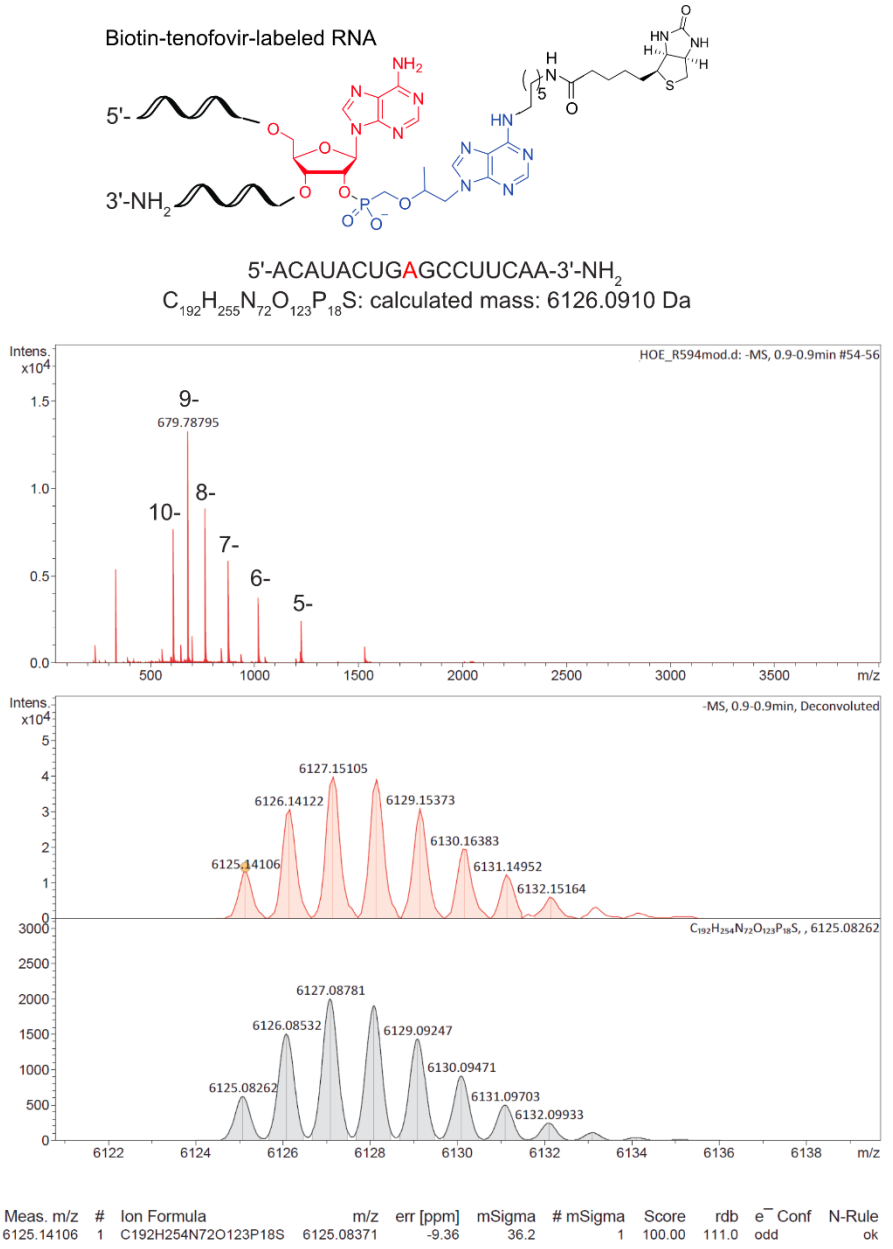


Figure S5. ESI mass spectrometric analysis of biotin-tenofovir-labeled 3'-aminohexyl-RNA produced by preparative labelling with FJ1. ESI-Mass spectrum (neg. mode) of the isolated RNA product; deconvoluted MS (red) and simulated mass spectrum (grey).

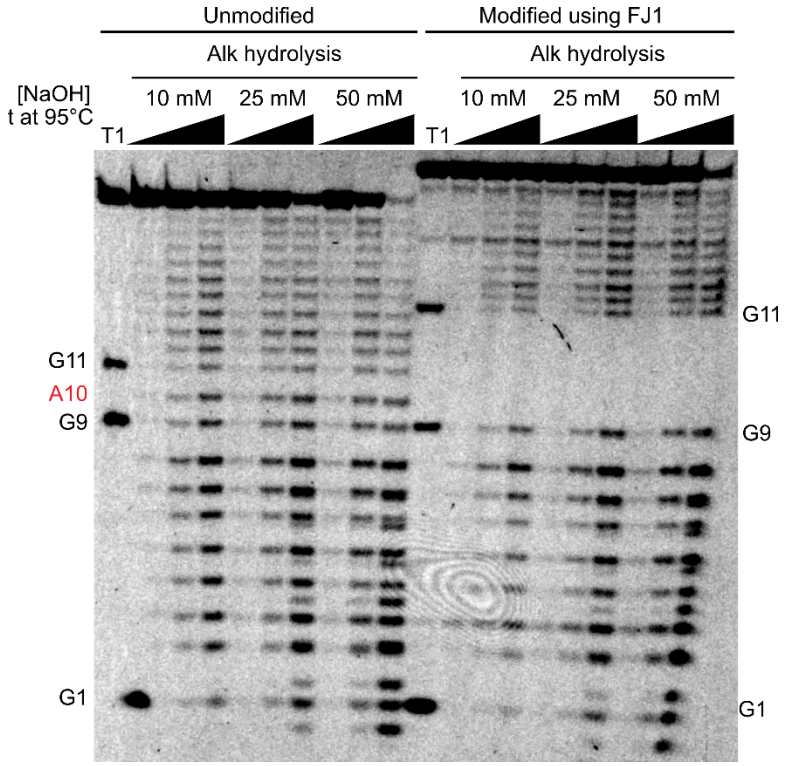


Figure S6. Alkaline hydrolysis and RNase T1 digestion of isolated FJ1 product for assignment of labeling site. Experimental details see section 2.3.2.2.

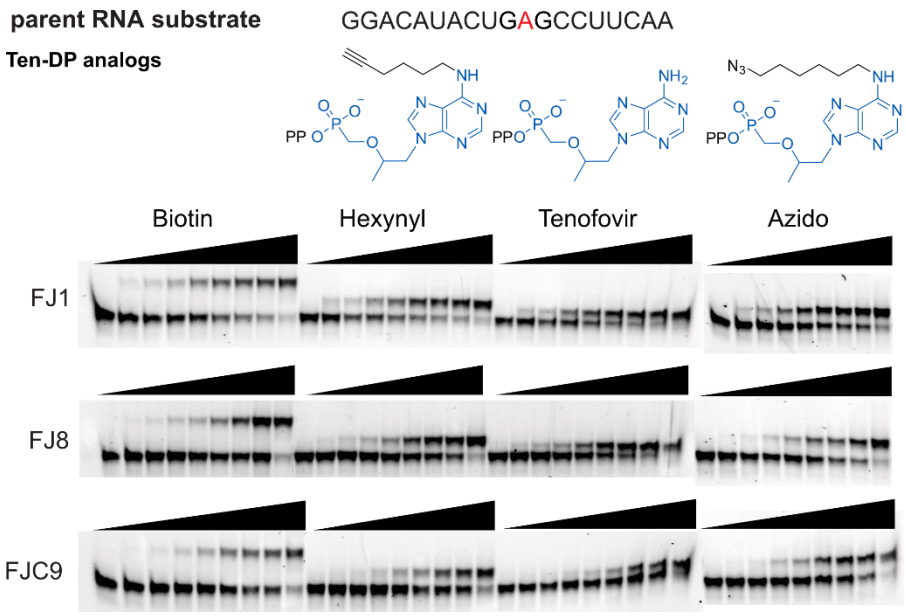


Figure S7. Gel images for data in Figure 4a. Small molecule reactivity with FJ1, FJ8 and FJC9 on 3'-fluorescein-labeled parent RNA. (300 µM Ten-DP analog, with 40 mM $MgCl_2$, in selection buffer, timepoints are: 0, 15, 30, 60 min; 2, 4, 6, 8, 22 h).

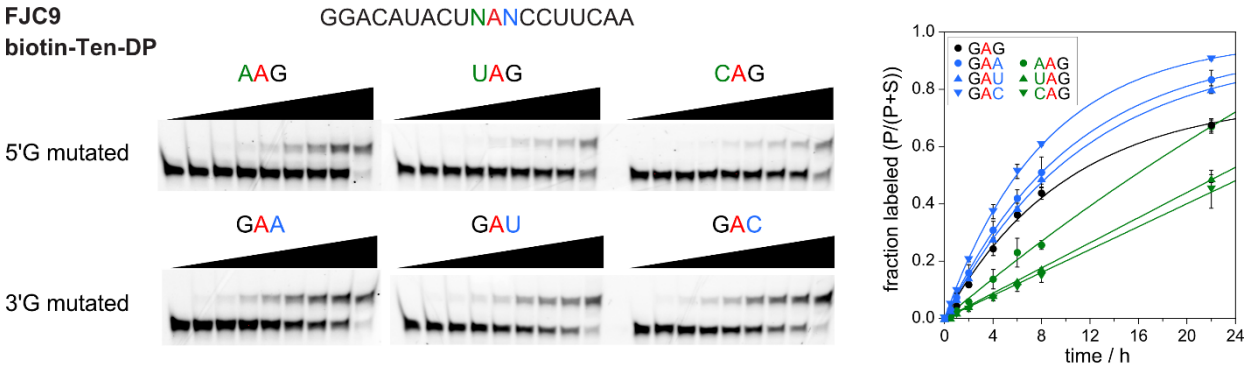


Figure S8. Substrate sequence mutations for FJC9. (3'-fluorescein-labeled RNAs, 300 µM Biotin-Ten-DP, with 40 mM $MgCl_2$, in selection buffer, 37°C, timepoints are: 0, 15, 30, 60 min; 2, 4, 6, 8, 22 h).

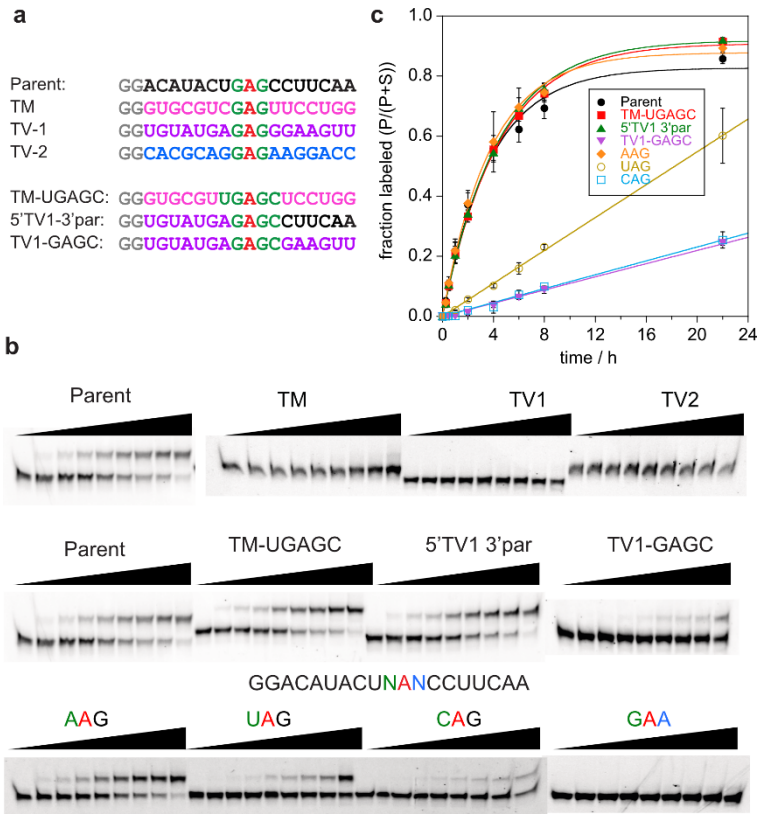


Figure S9. Substrate sequence scope for FJ1. Transition and transversion mutations are color coded in the sequence labels: pink: transition, purple: transversion-1, blue: transversion-2. Green: nucleotides that remained unchanged from the parent. 5'-GG in grey, since these are needed for transcription initiation. (3'-fluorescein-labeled RNAs, 300 µM Biotin-Ten-DP, with 40 mM $MgCl_2$, in selection buffer, 37°C, timepoints are: 0, 15, 30, 60 min; 2, 4, 6, 8, 22 h).

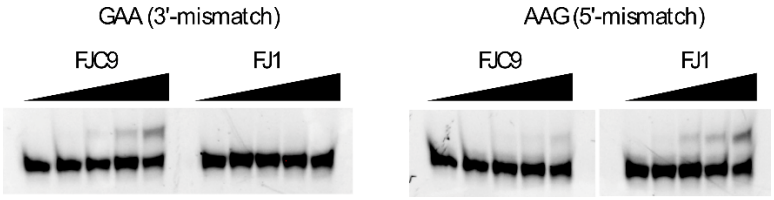


Figure S10. Mismatch effects on ribozyme labeling efficiency. The substrates were single point mutants with the central GAG motif changed to AAG or GAA. Labeling reactions were performed with parent FJ1 and FJC9 ribozymes, thus introducing 5' or 3' A-C mismatches. In both cases, the labeling efficiency was significantly reduced (below 15% after 22 h, compared to 90% in the matched case). This result indicates that there is little propensity for off-target labeling, i.e. both ribozymes require full complementarity to maintain efficient labeling.

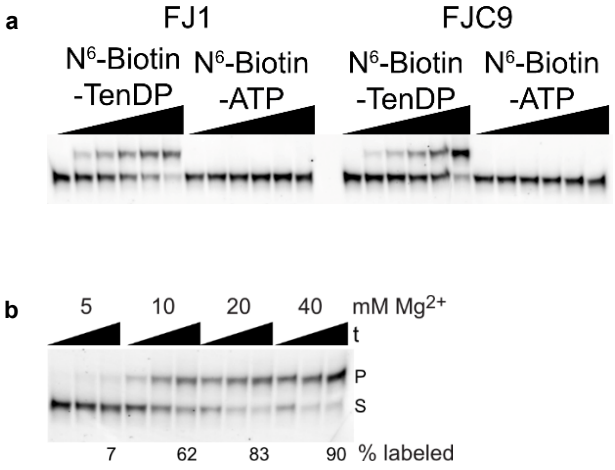


Figure S11. a) Orthogonality of FJC9 and FJ1 to FH14. FJ ribozymes do not show any RNA labelling activity with ATP. (3'-fluorescein-labeled parent RNA, 300 µM Biotin-Ten-DP, 200µM Biotin-ATP, with 40 mM $MgCl_2$, in selection buffer, 37°C, timepoints are: 0, 1, 2, 4, 6, 22 h). b) Mg^{2+}-dependence of FJ1 ribozyme. Reactions were performed with parent RNA and biotin-tenofovir-DP at 37°C for 1, 6, and 22 h. The ribozyme activity is dependent on Mg^{2+} concentration. At 5 mM Mg^{2+}, only 7 % labeled RNA is observed after 22h, but already at 10 mM Mg^{2+}, the yield is above 60%.

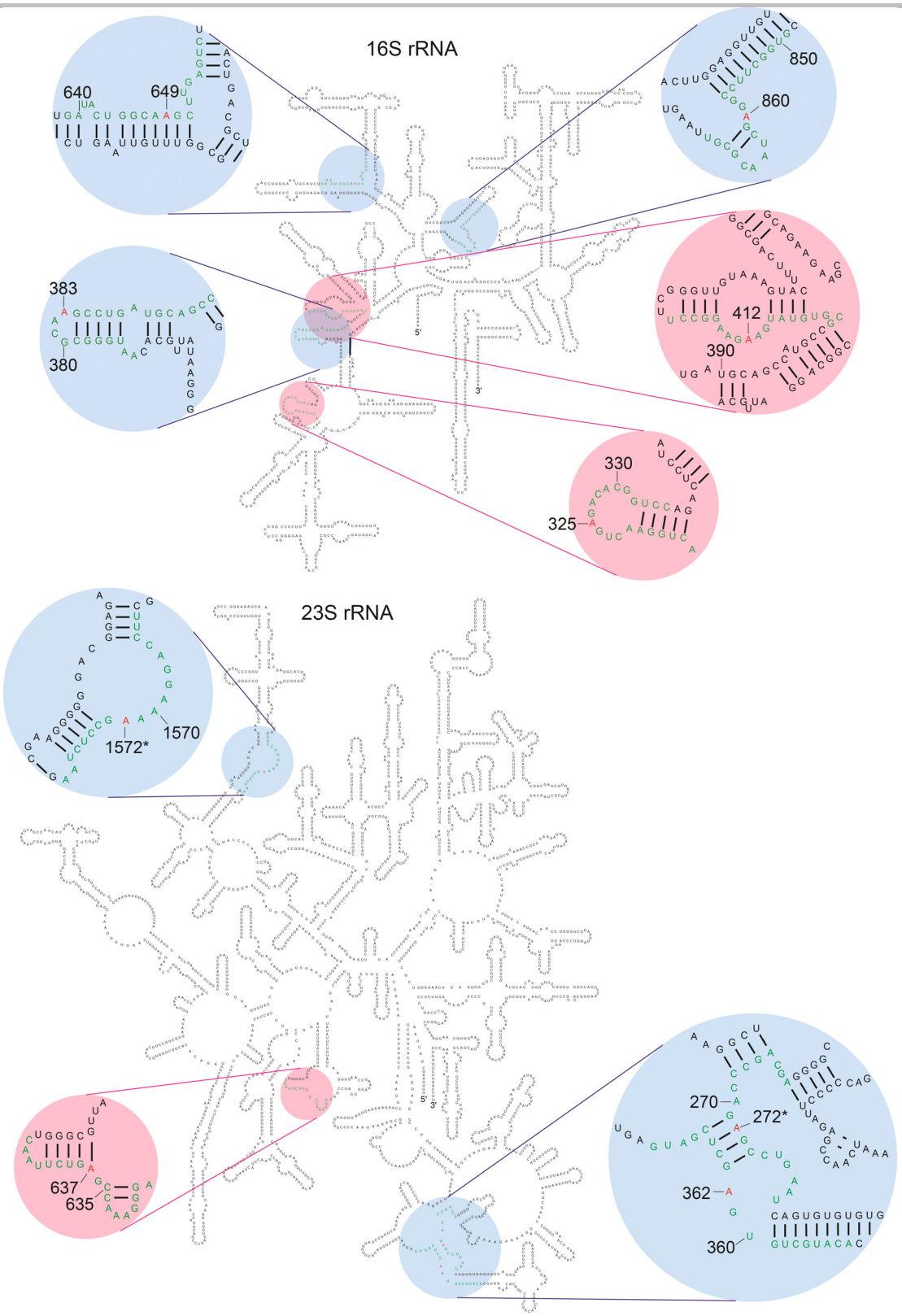


Figure S12. Secondary structures[11] of E.coli 16S and 23S rRNA with chosen labelling sites for FJ1 and FH14 ribozymes marked. Expanded regions show the sequence context on red background for FH14 and on blue background for FJ1. Positions A272 and A1572 in 23S marked with an * were targeted by both ribozymes. The target adenosine is shown in red, and the nucleotides hybridizing to the ribozyme binding arms are indicated in green.

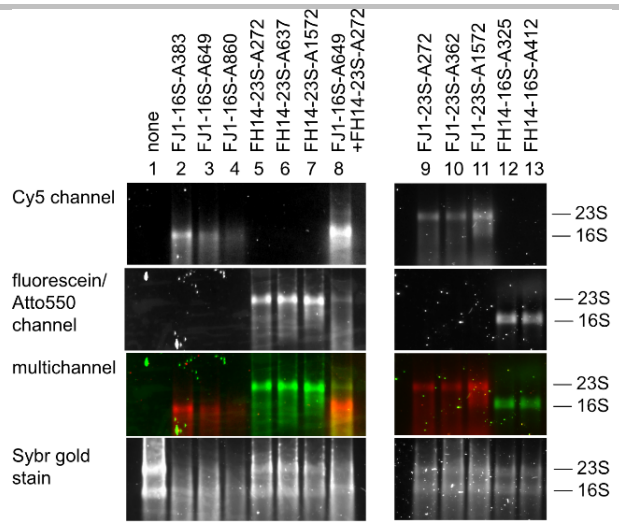


Figure S13. Labelling of E.coli 16S and 23S rRNA with FJ1 and FH14. Lane 1: unlabelled reference. lanes 2-4: FJ1 with Cy5-Ten-DP on three different positions of 16S rRNA, lanes 5-7: FH14 with Atto550-ATP on 3 different positions of 23S rRNA, lane 8: simultaneous labelling of 16S and 23S rRNA. lanes 9-11: FJ1 with Cy5-Ten-DP on three different positions of 23S rRNA, lane 12-13: FH14 with FAM-ATP on two different positions of 16S rRNA.

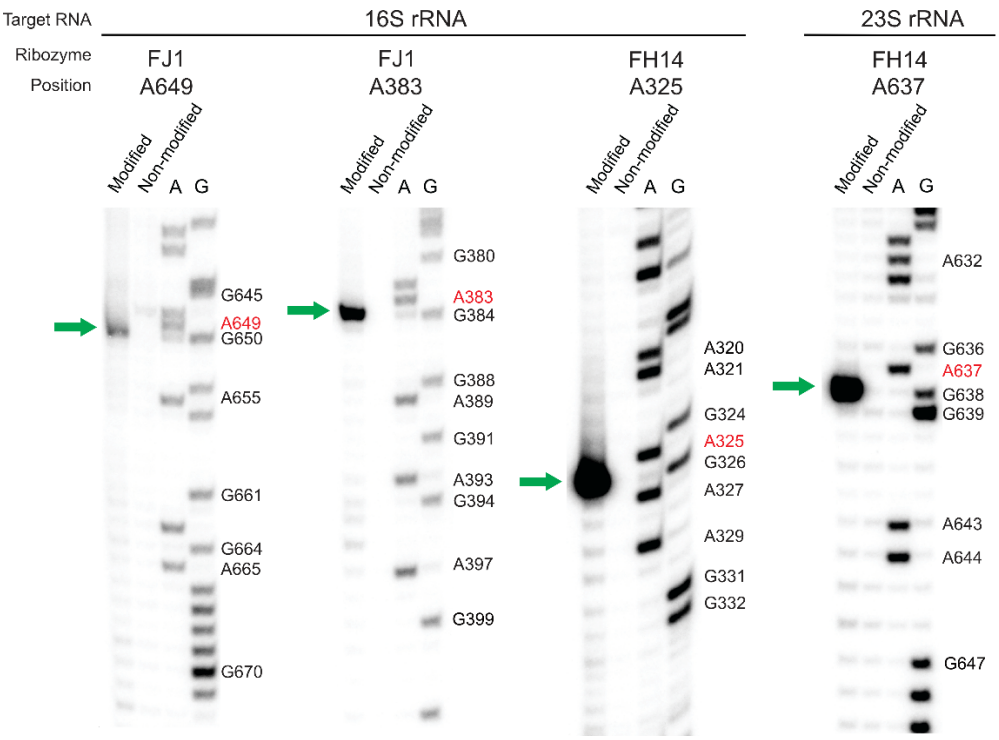


Figure S14. Primer extension analysis of FJ1 and FH14 labelling sites in 16S and 23S rRNA. As incdicated on top of each gel picture, E.coli RNA was labelled by FJ1 and FH14 ribozymes targeting positions A649, A383 and A325 in 16S rRNA and A637 in 23S rRNA, respectively, using cognate biotinylated tenofovir or ATP substrates. Primer extension assays on the modified RNA and the unmodified input RNA were performed with 5'-^{32}P-labeled primers and superscript III reverse transcriptase. The primer sequences are listed in section 1.2. Sequencing lanes (A, G) are indicated.

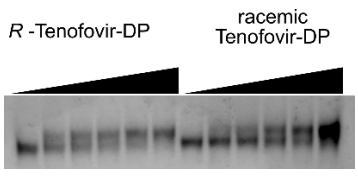
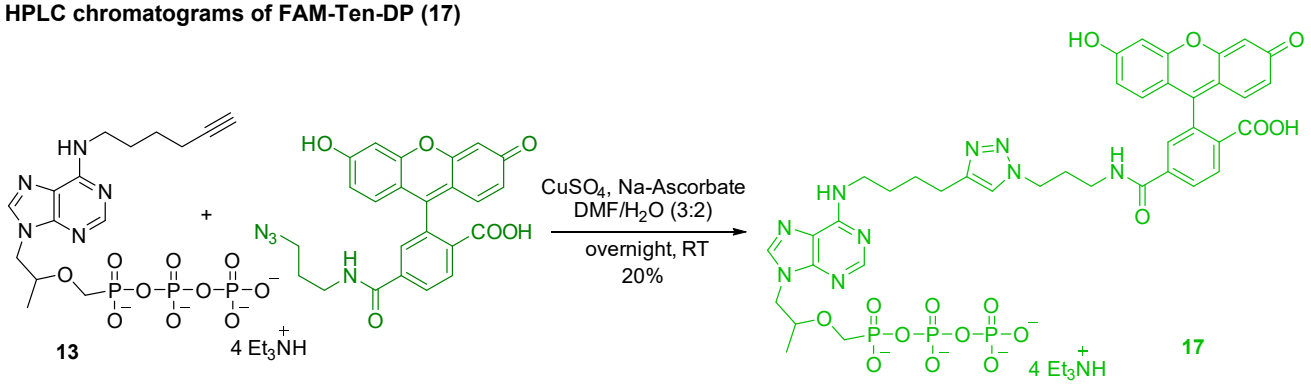
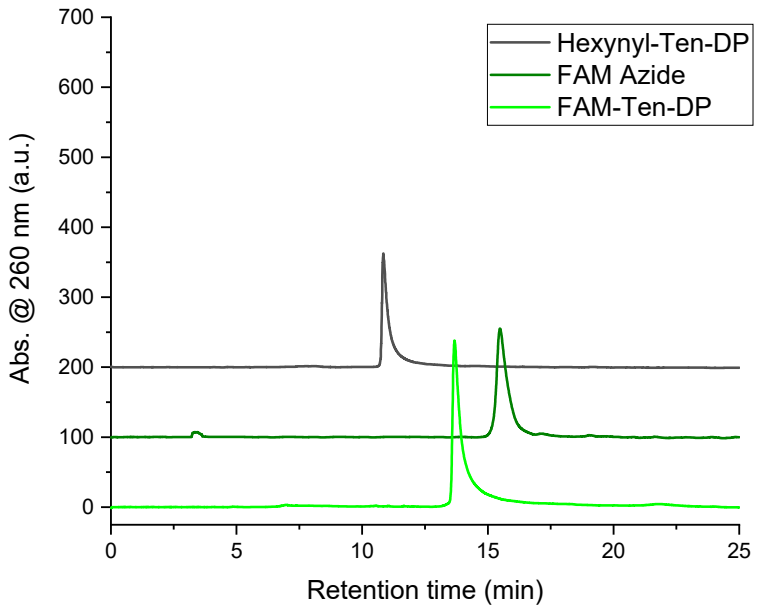


Figure S15. Comparison of *R*-enantiomer and racemic tenofovir diphosphate for RNA labelling reaction with FJ1 on parent RNA substrate. (last timepoint is 22h).

HPLC chromatograms of FAM-Ten-DP (17)





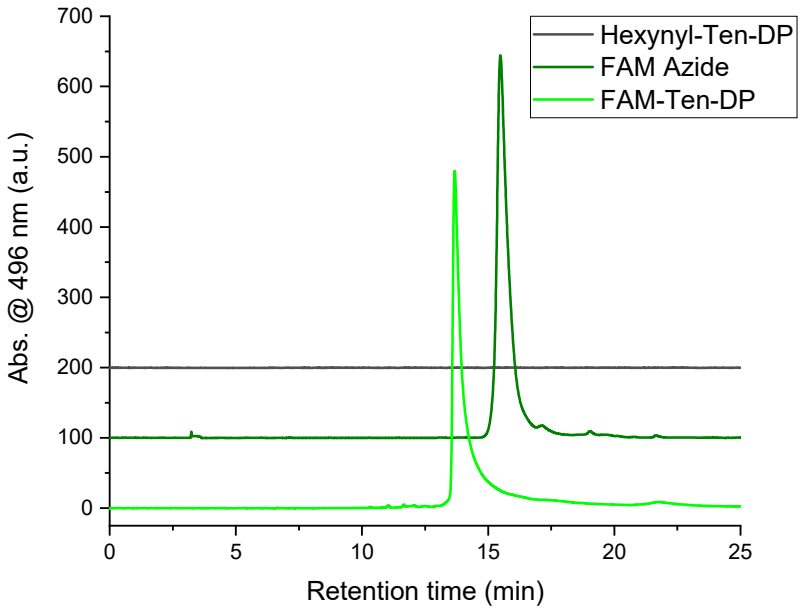
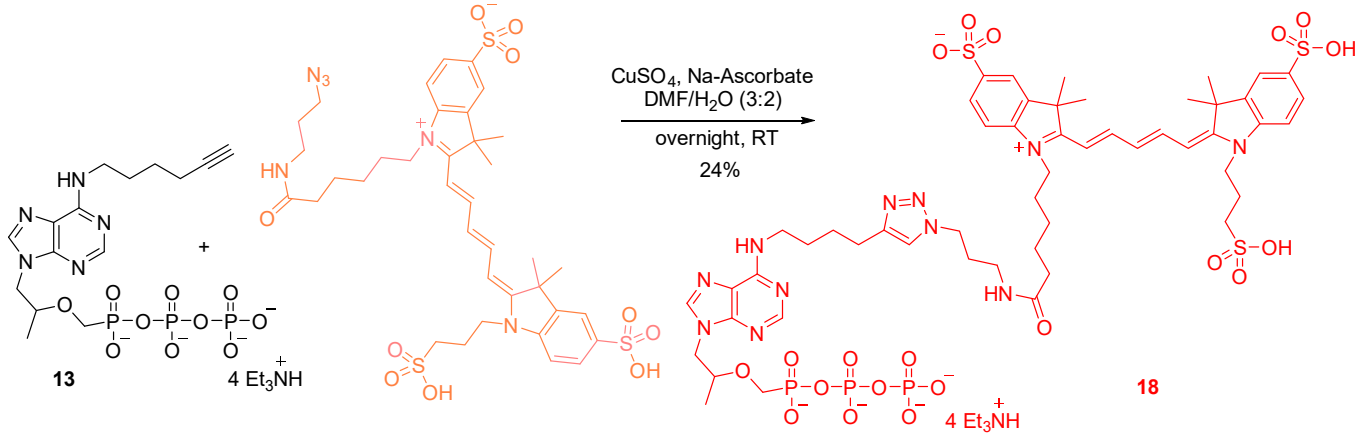


Figure S16: HPLC chromatograms of compound **13**, FAM azide and product **17** at 260 nm (above) and at 496 nm (below), respectively.

SUPPORTING INFORMATION

HPLC chromatograms of Cy5-Ten-DP (18)



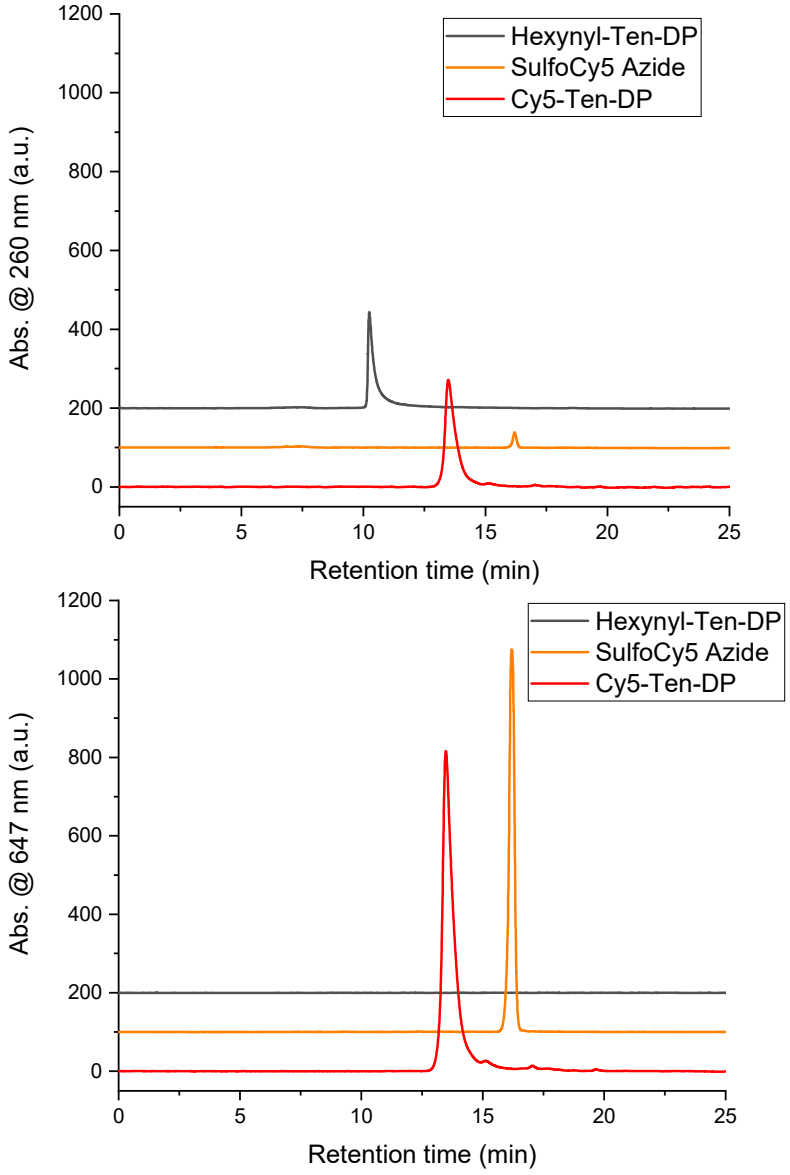
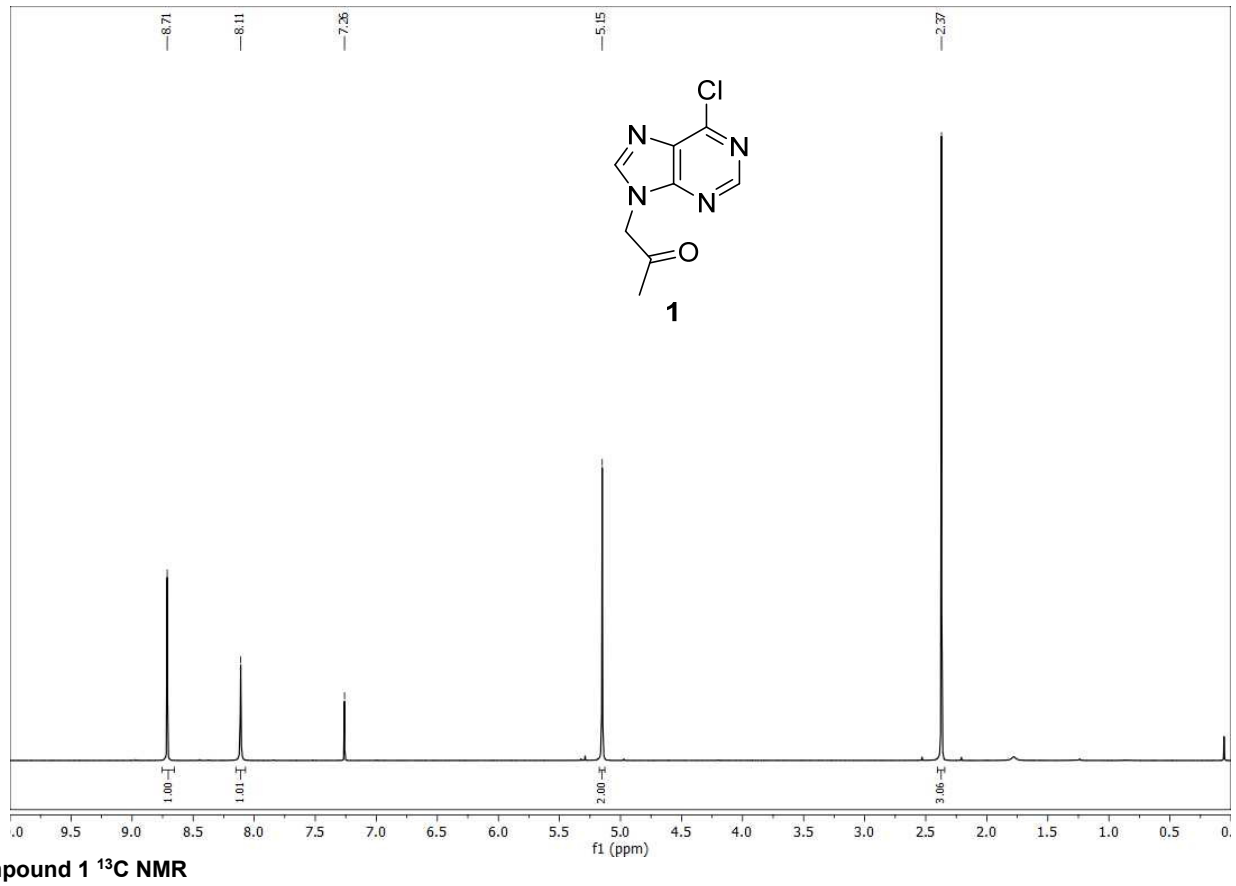


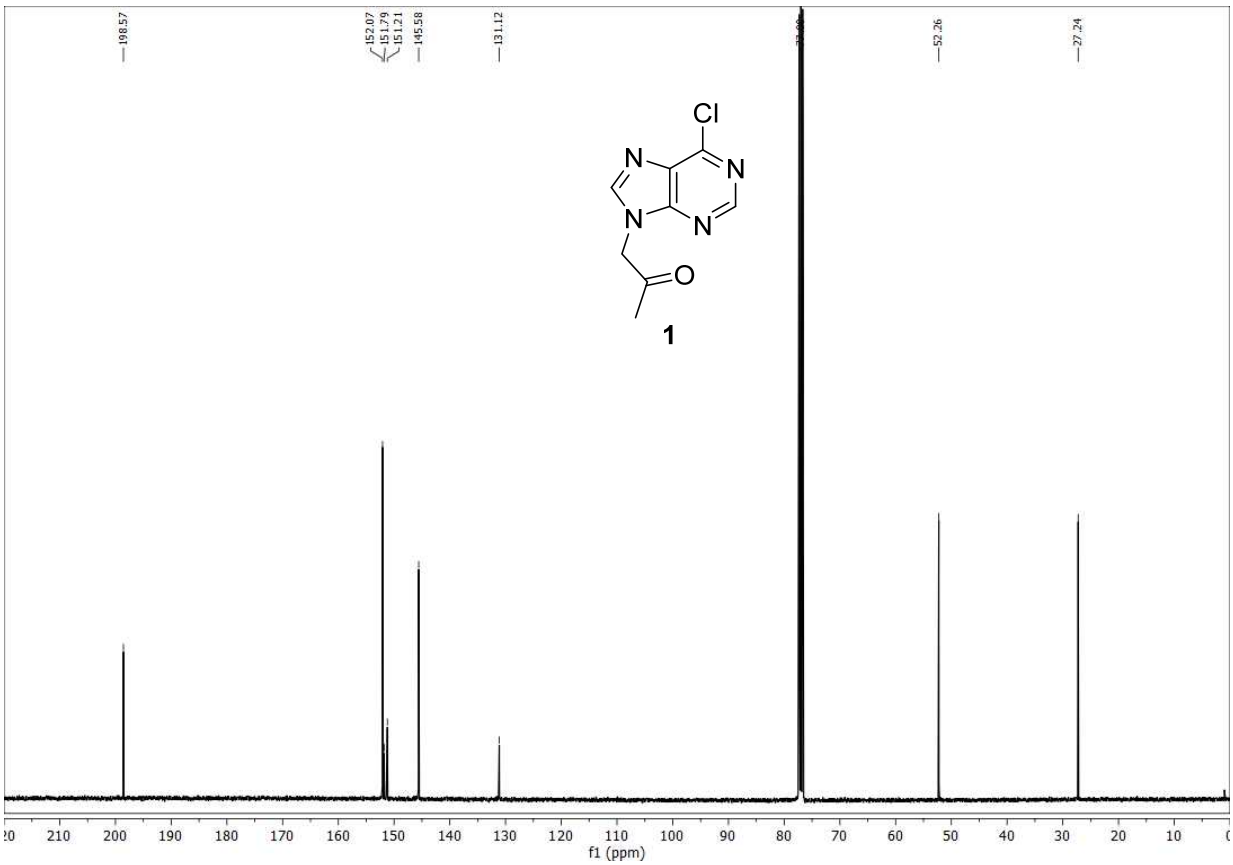
Figure S17: HPLC chromatograms of compound **13**, SulfoCy5 azide and product **18** at 260 nm (above) and at 647 nm (below), respectively.

NMR spectra

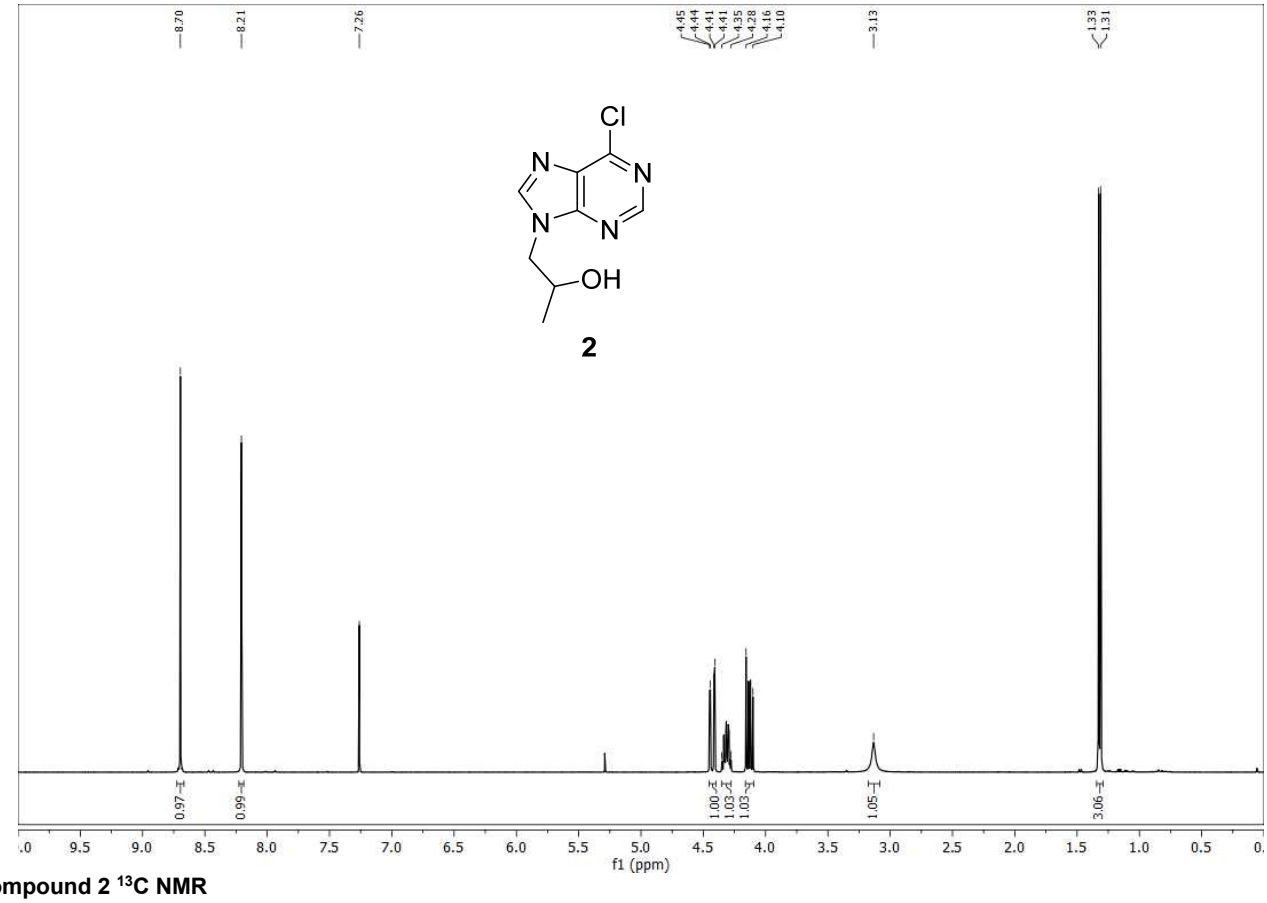
Compound 1 ^{1}H NMR



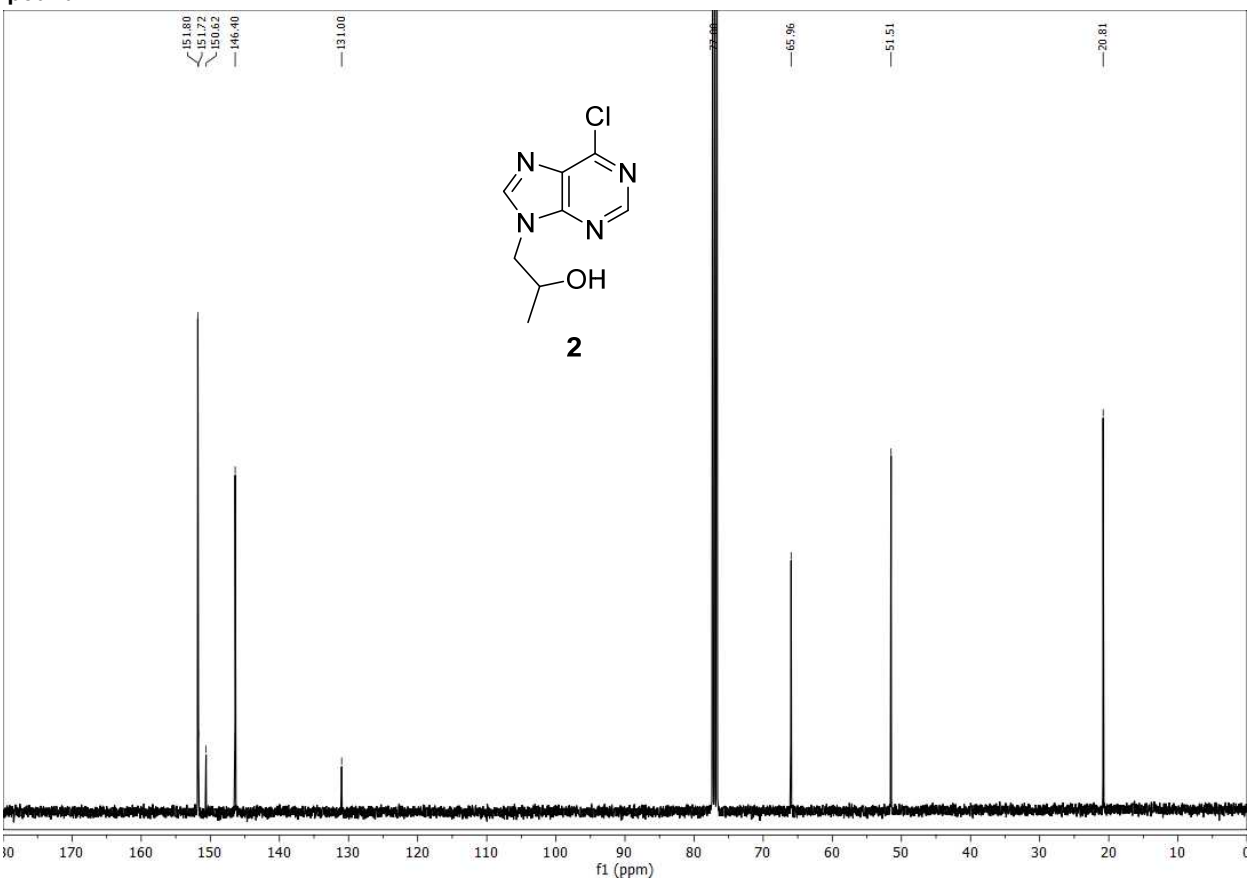
Compound 1 ^{13}C NMR



Compound 2 ^{1}H NMR

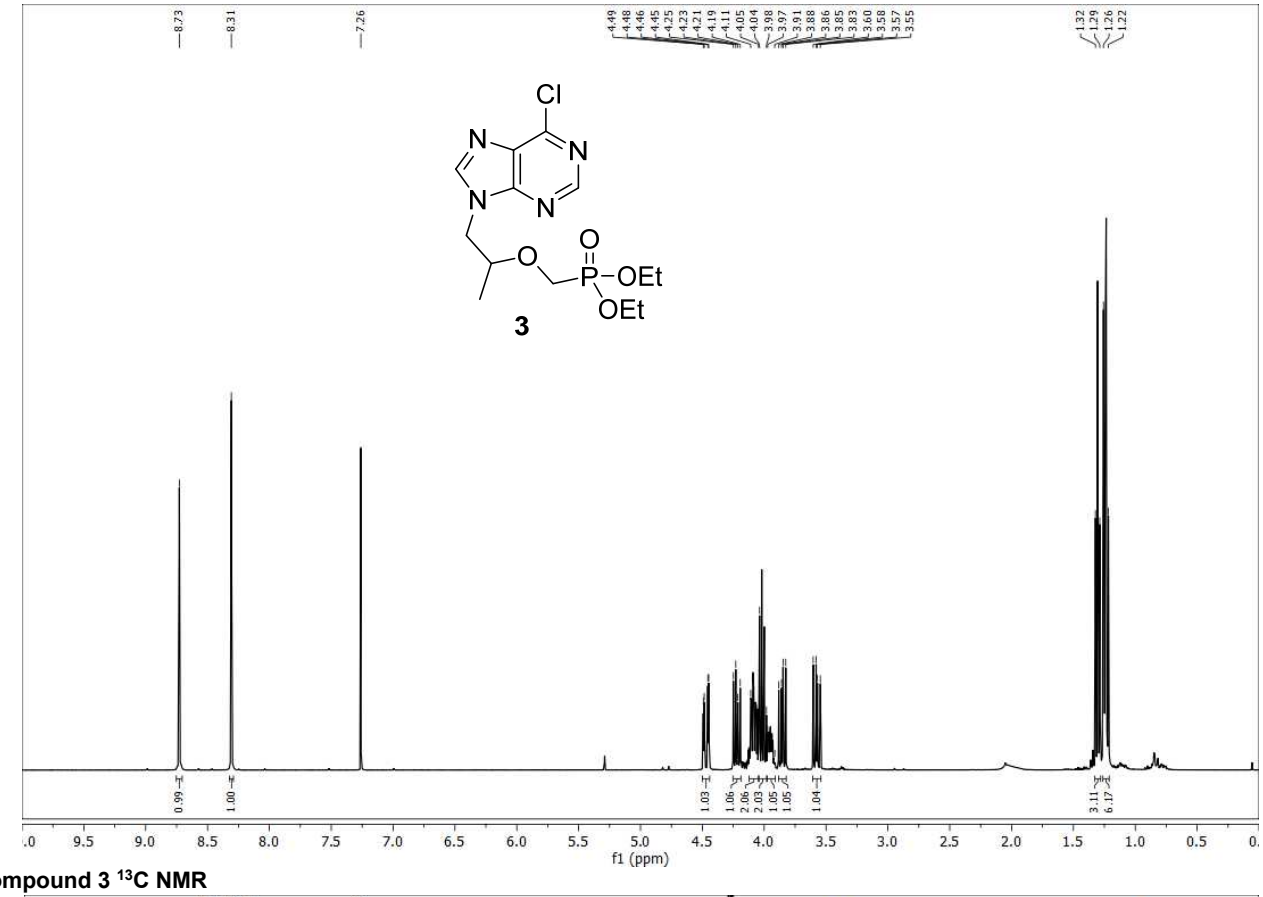


Compound 2 ^{13}C NMR

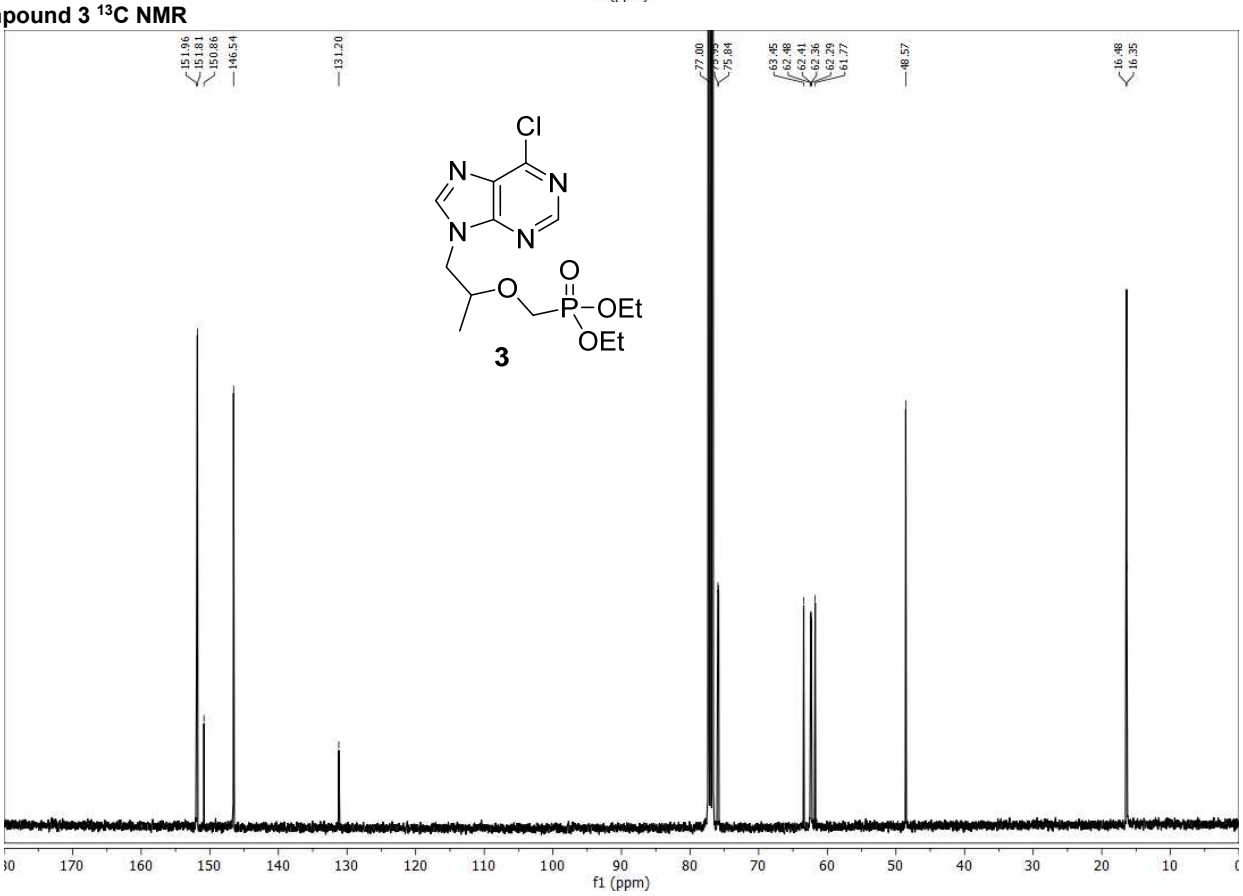


SUPPORTING INFORMATION

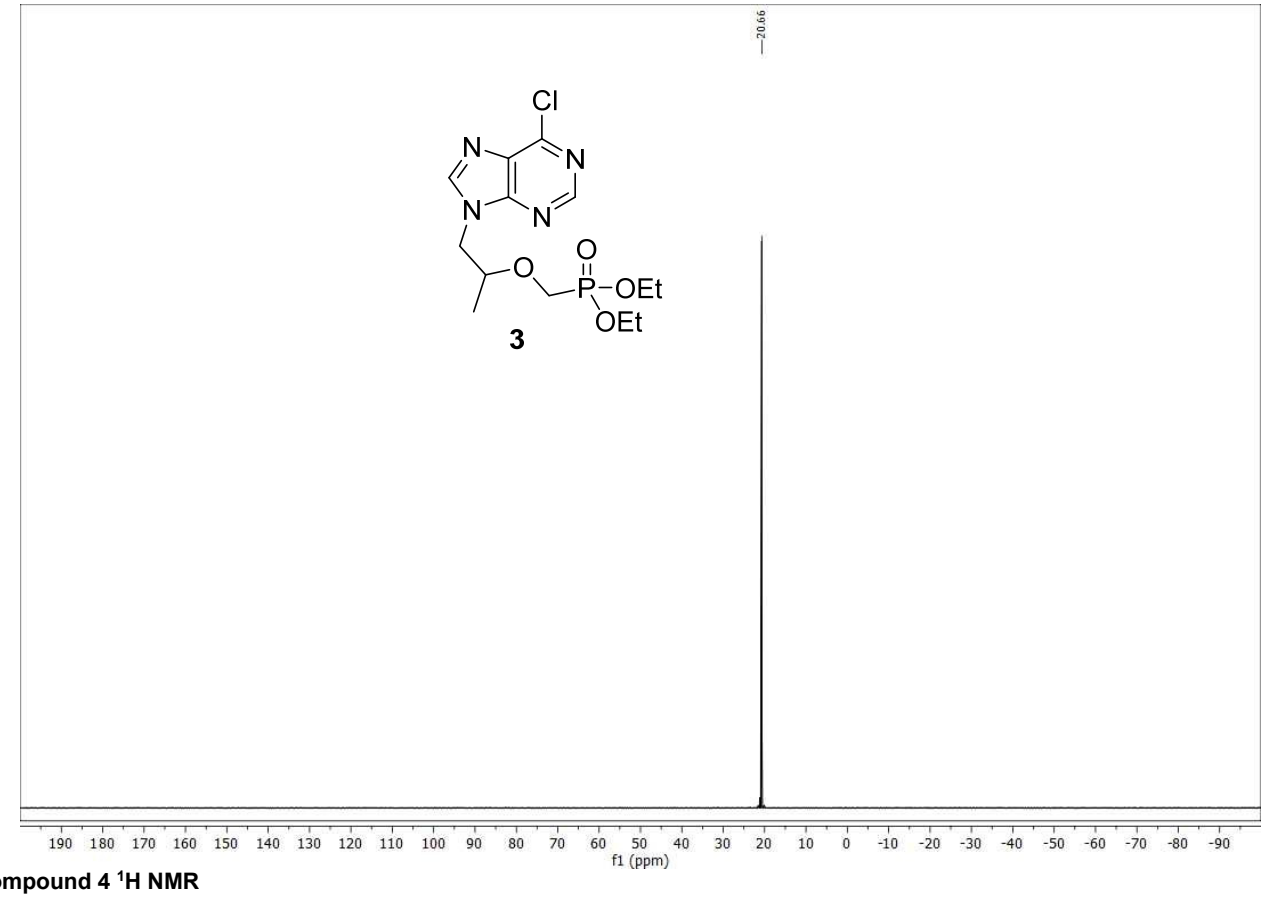
Compound 3 ^{1}H NMR



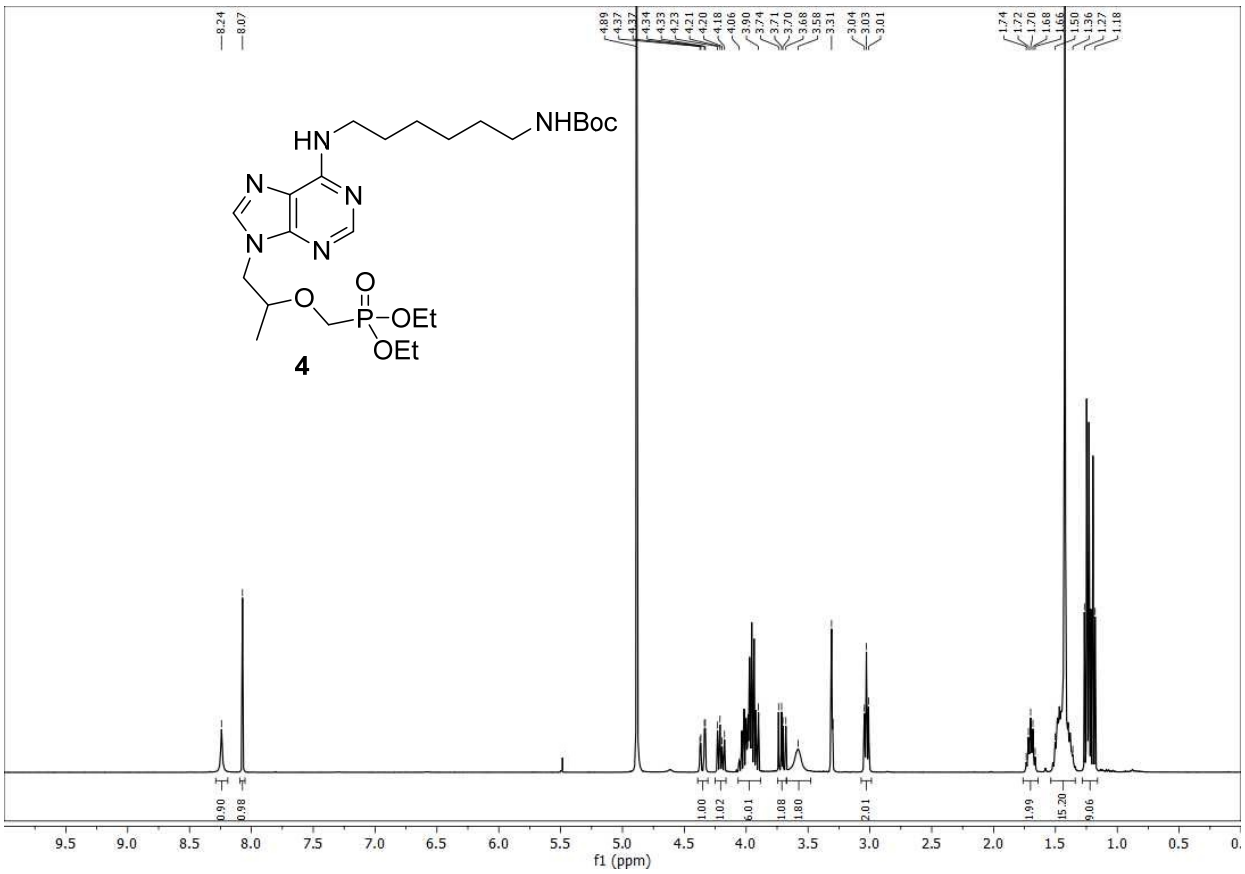
Compound 3 ^{13}C NMR



Compound 3 ^{31}P NMR

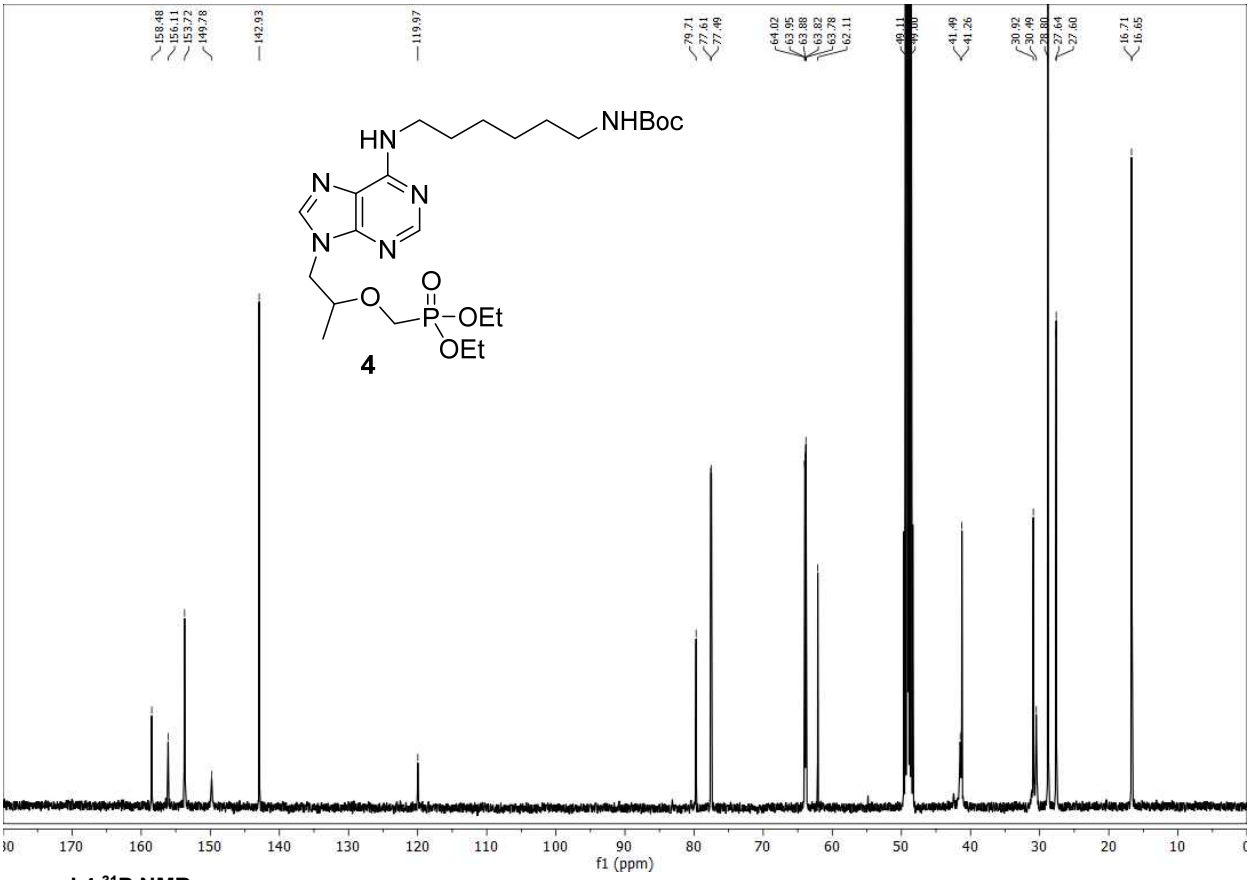


Compound 4 ^{1}H NMR

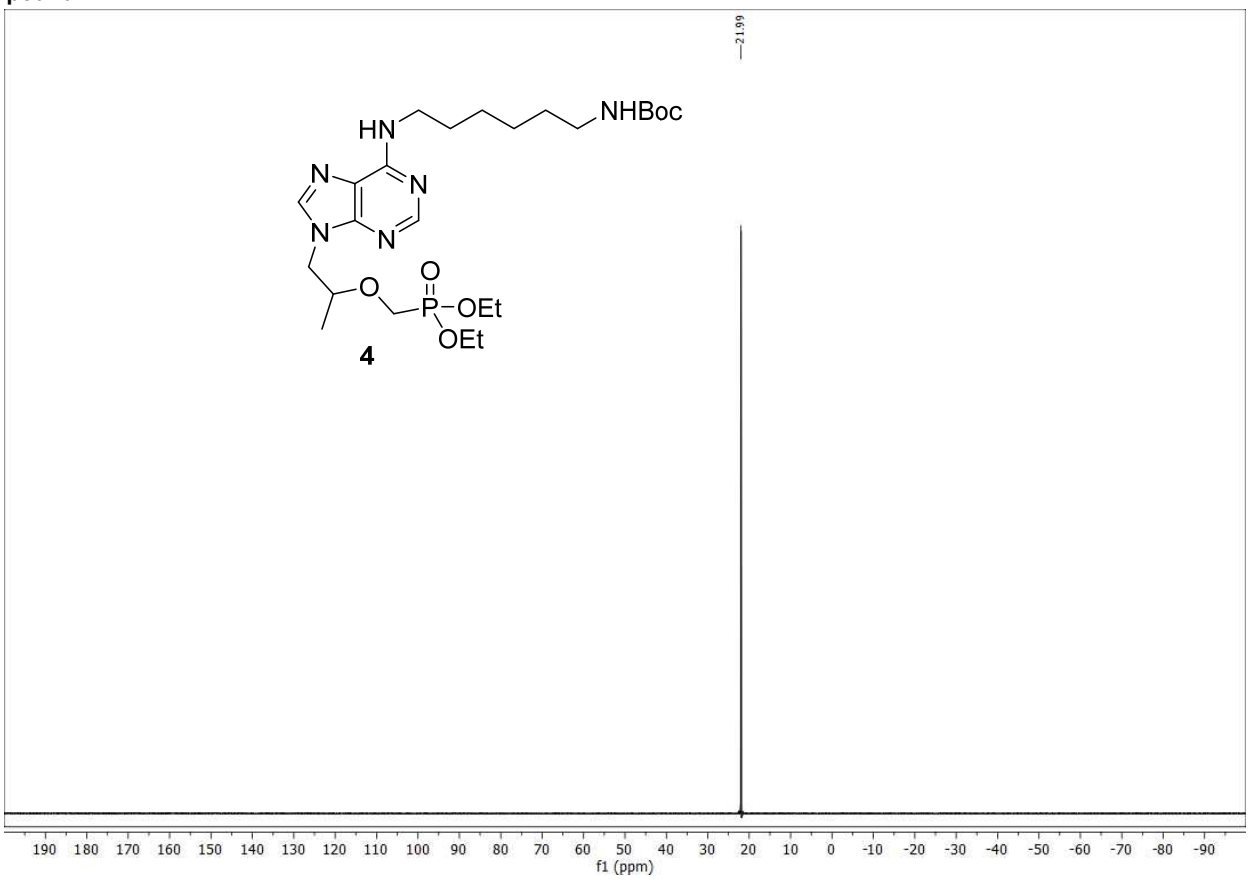


SUPPORTING INFORMATION

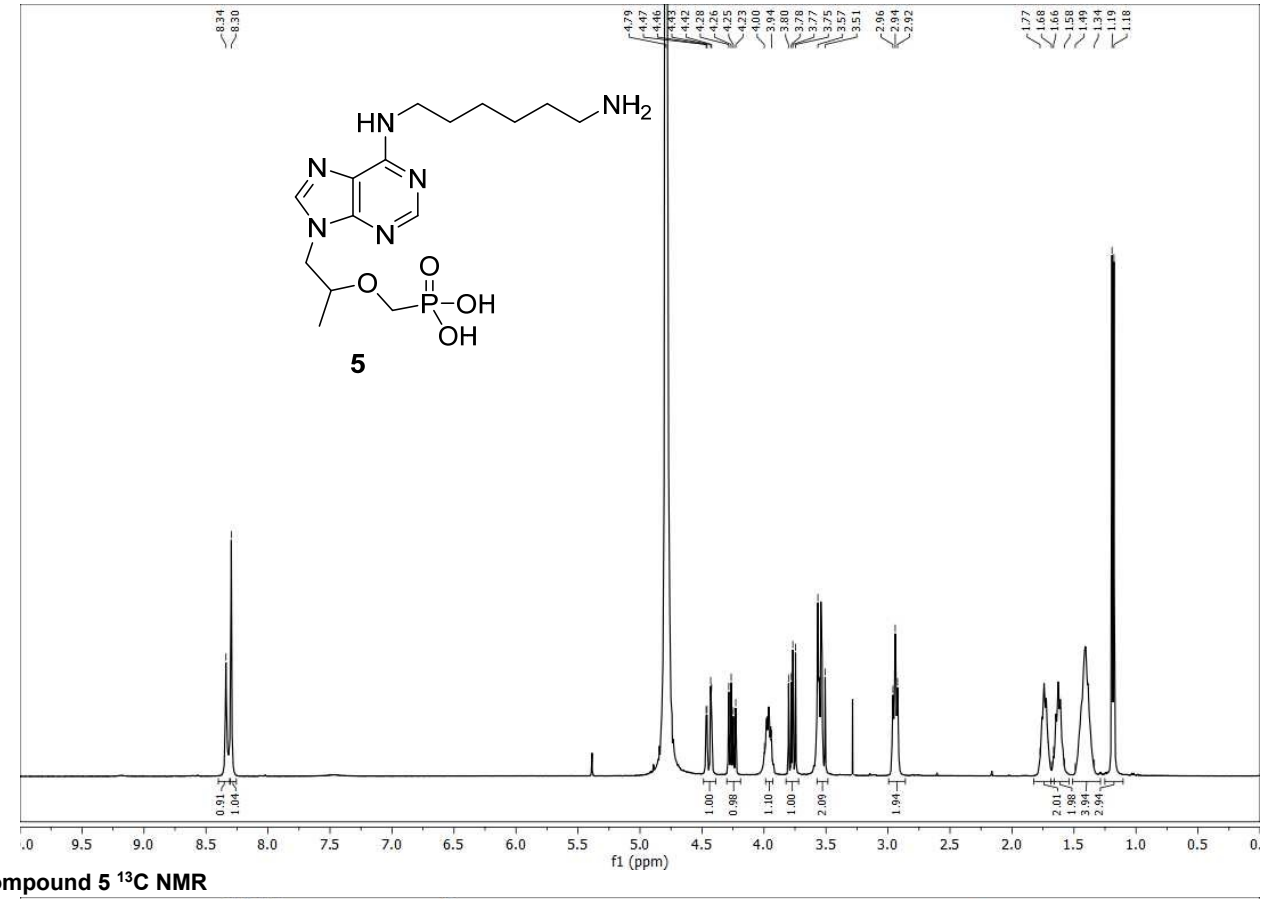
Compound 4 ^{13}C NMR



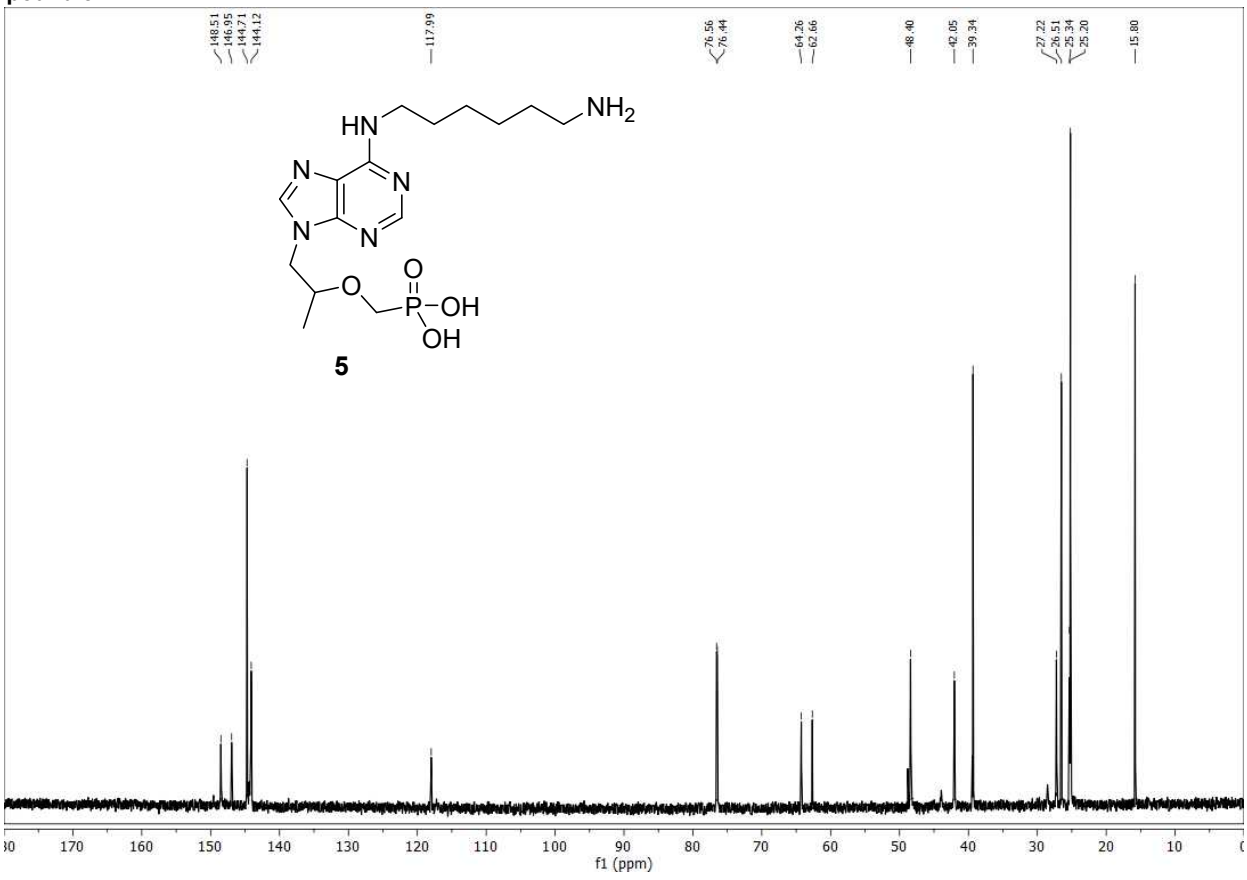
Compound 4 ^{31}P NMR



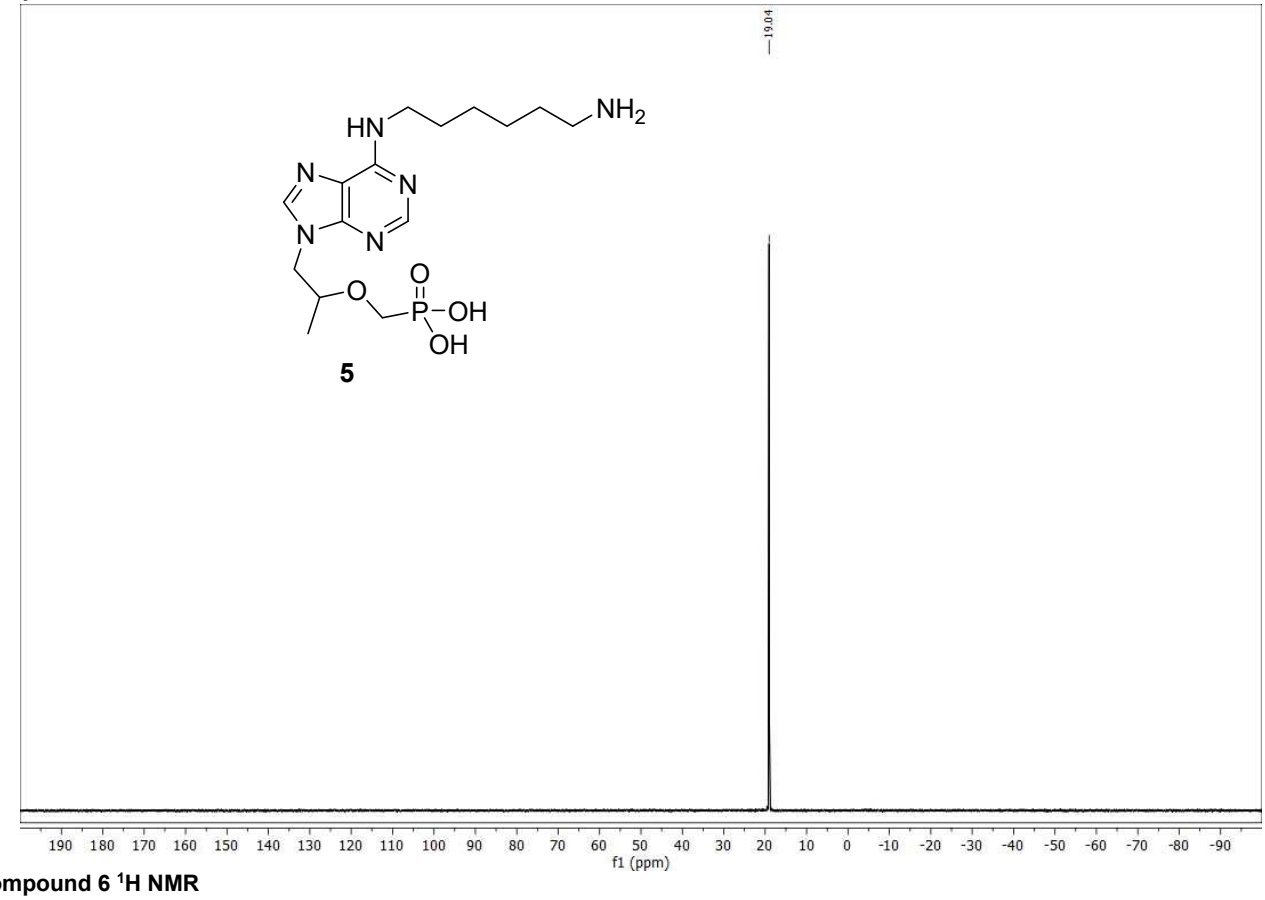
Compound 5 ^{1}H NMR



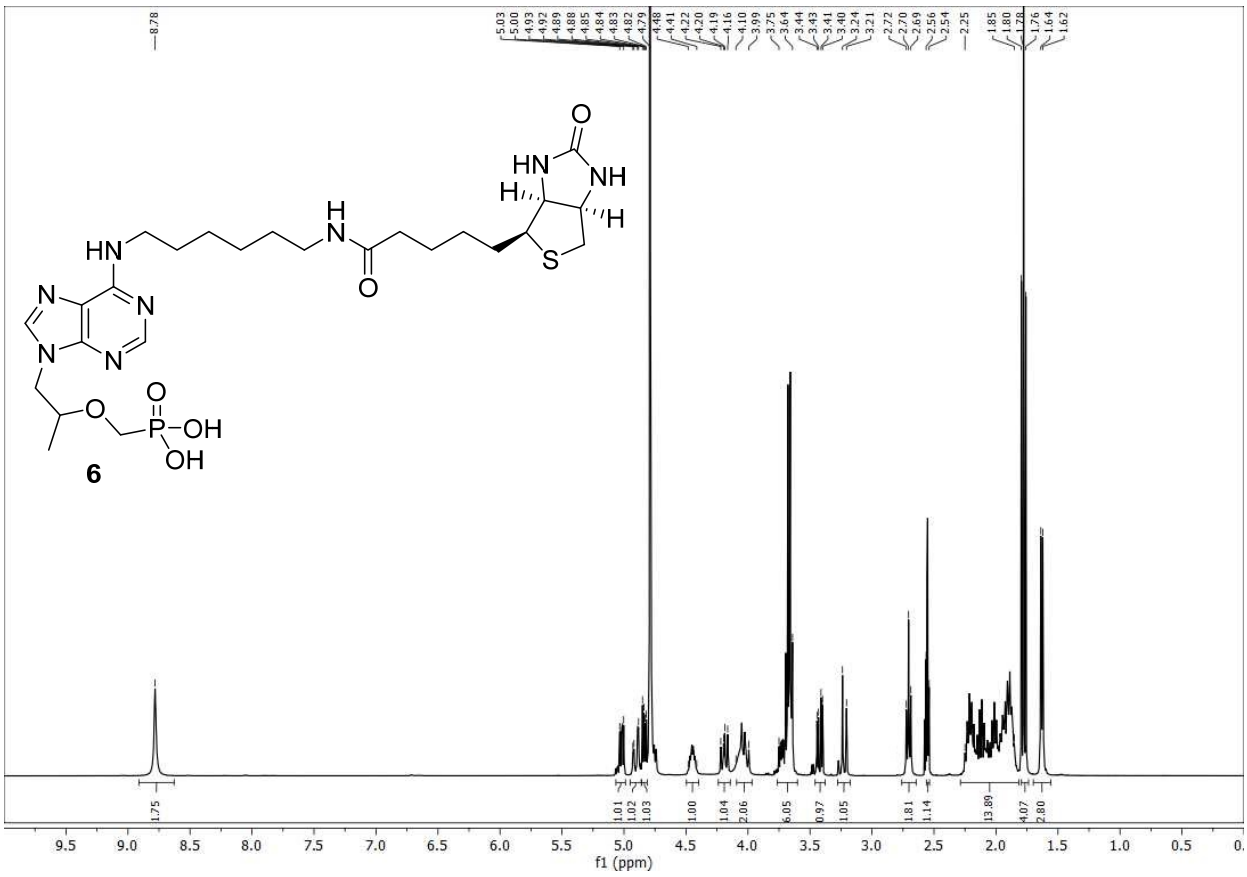
Compound 5 ^{13}C NMR



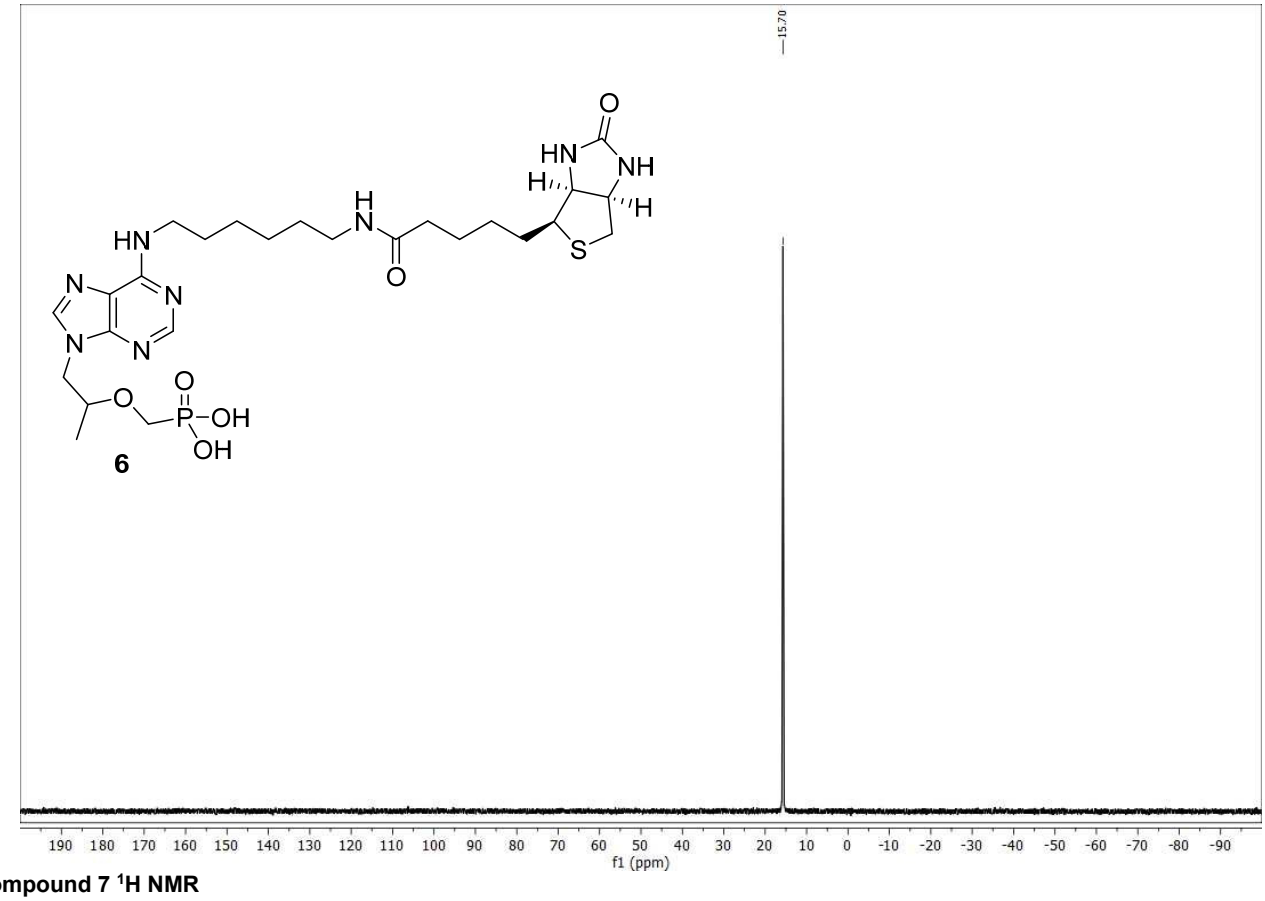
Compound 5 ^{31}P NMR



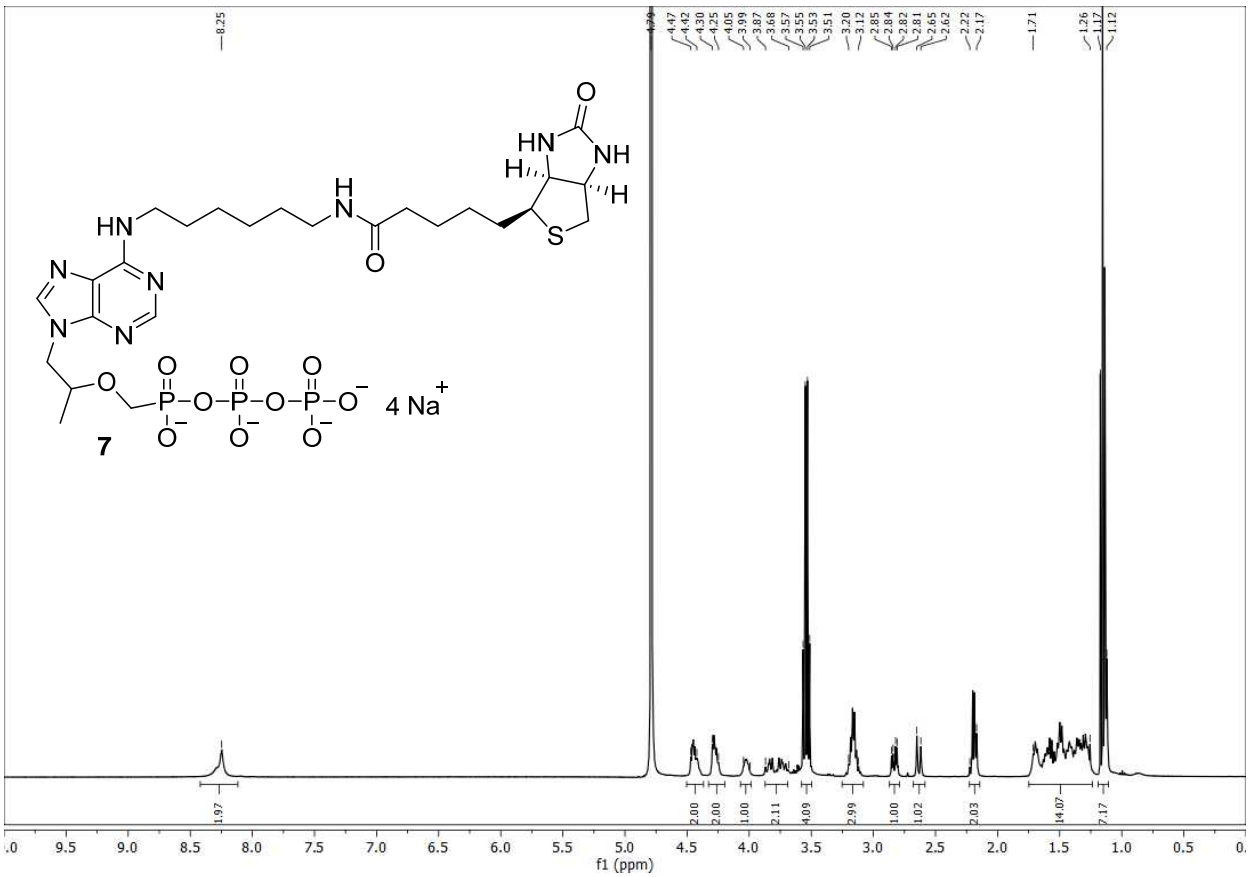
Compound 6 ^{1}H NMR



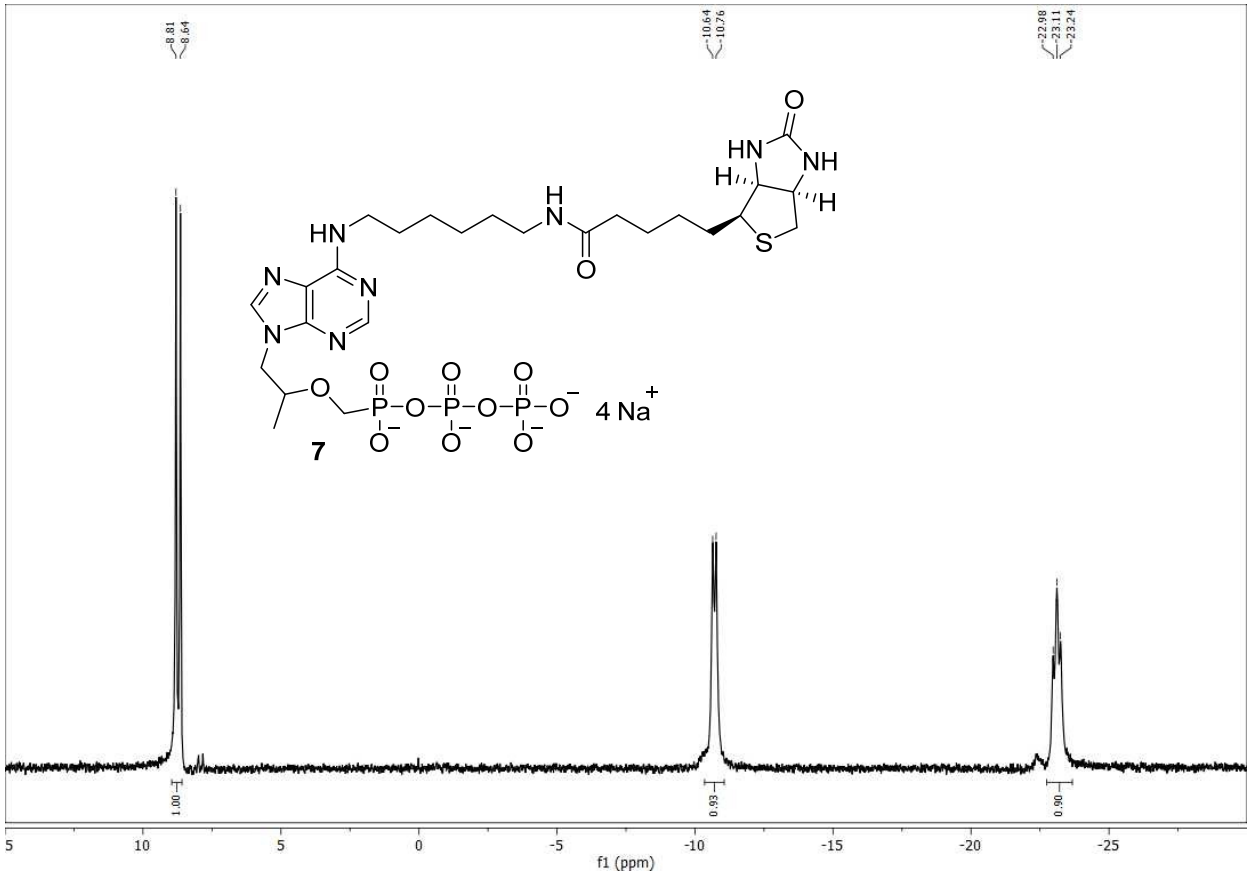
Compound 6 ^{31}P NMR



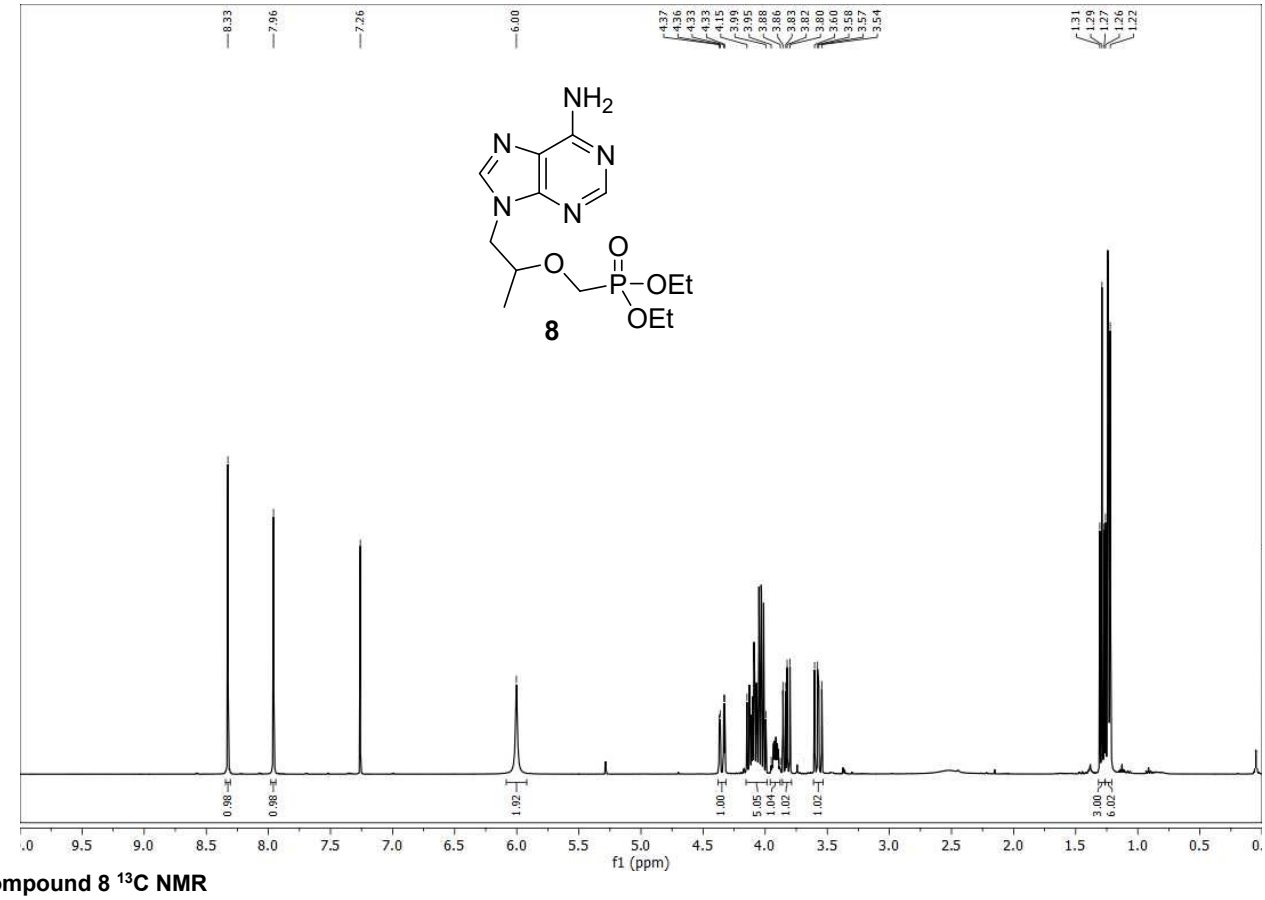
Compound 7 ^{1}H NMR



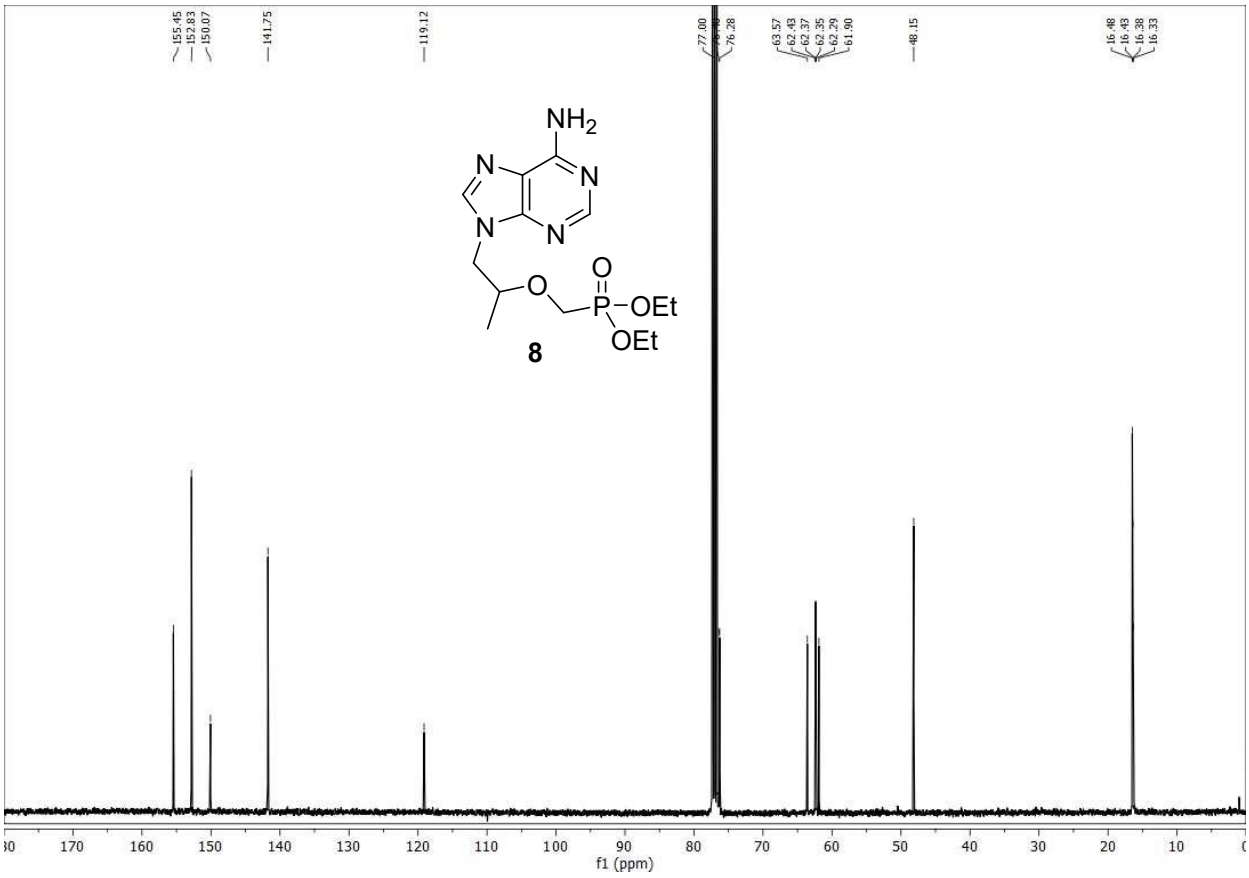
Compound 7 ^{31}P NMR



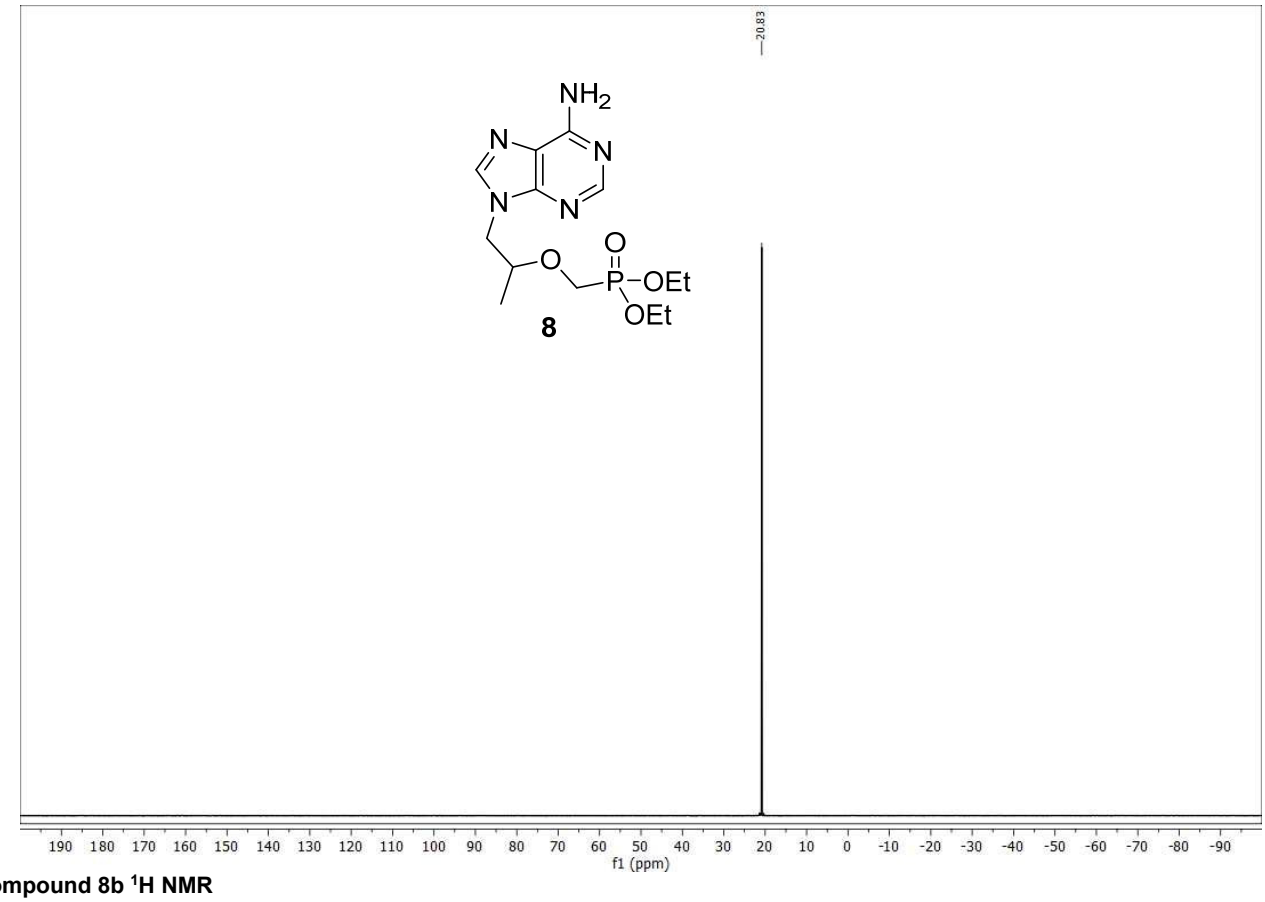
Compound 8 ^{1}H NMR



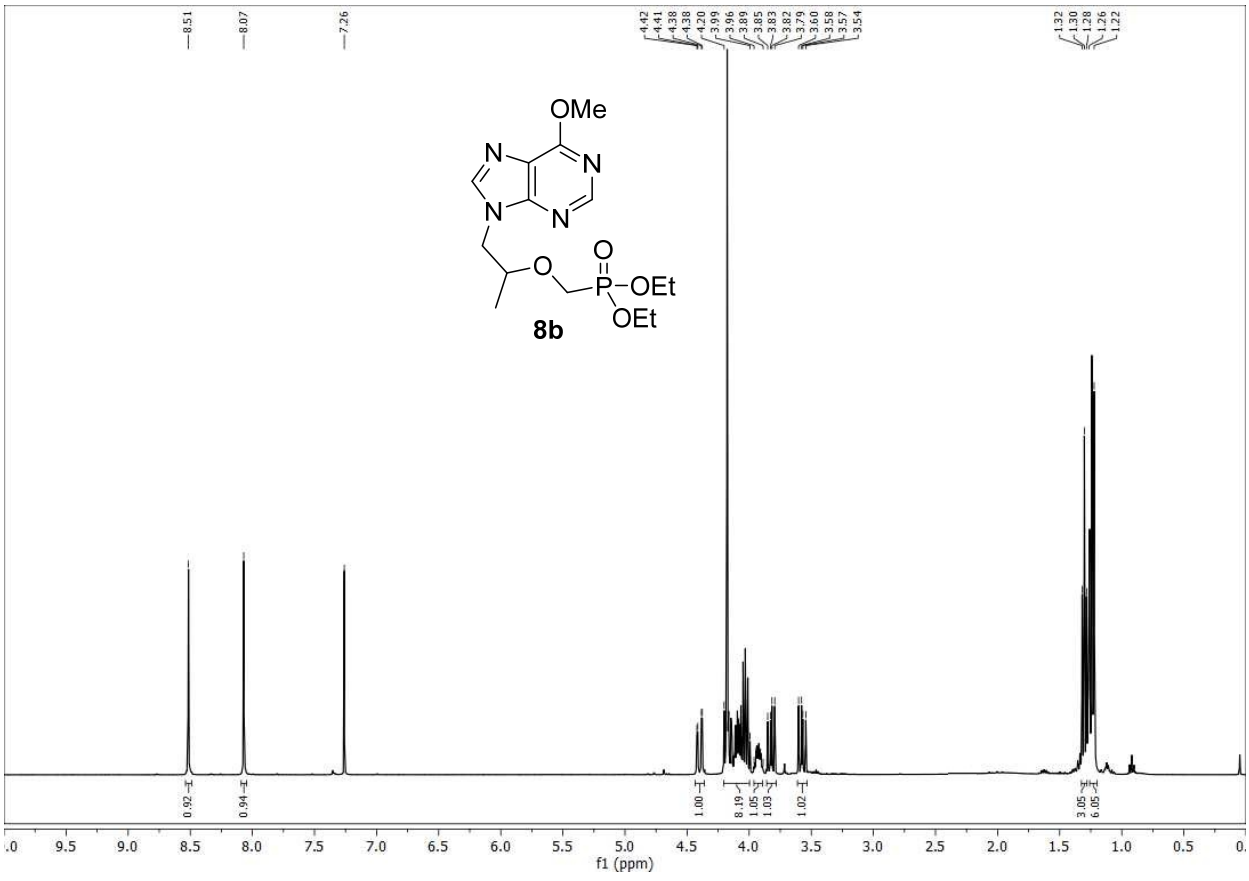
Compound 8 ^{13}C NMR



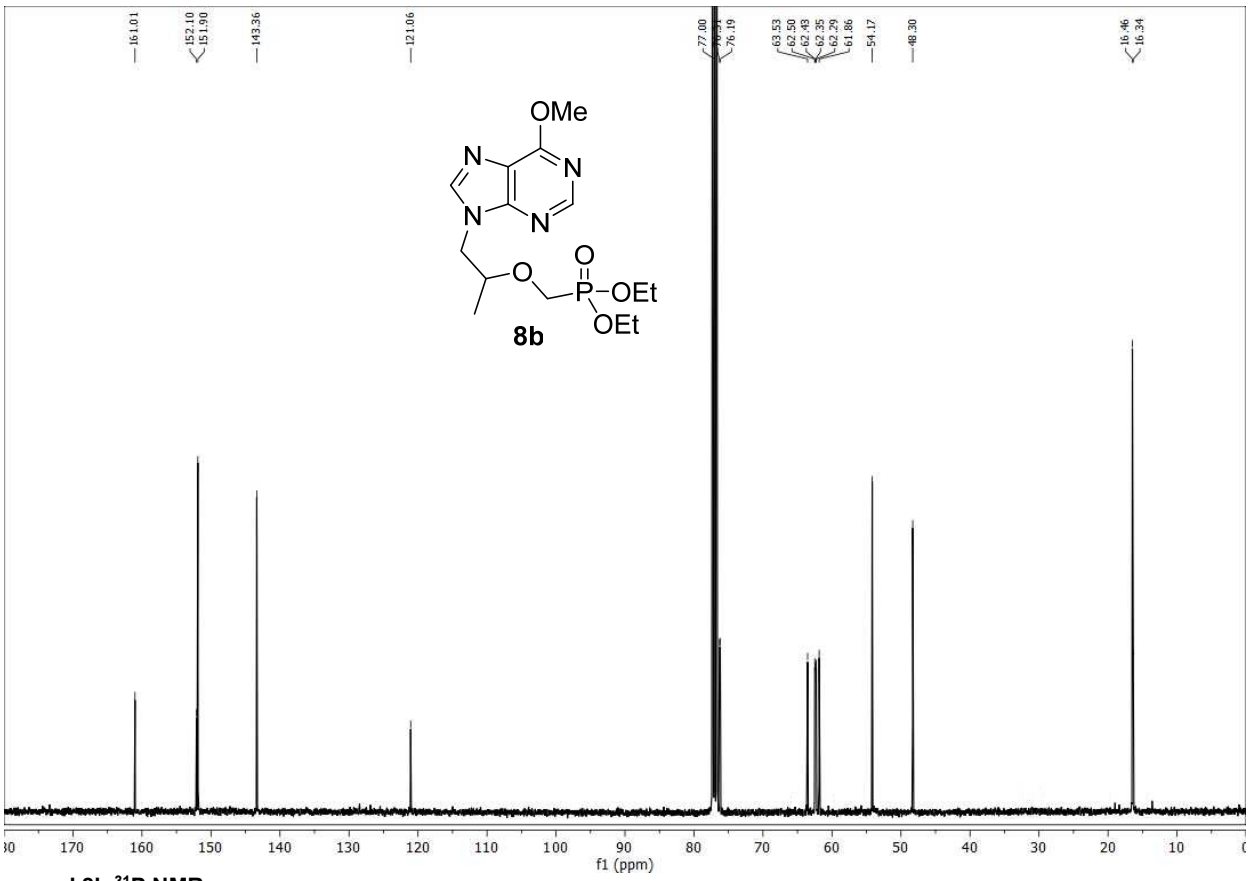
Compound 8 ^{31}P NMR



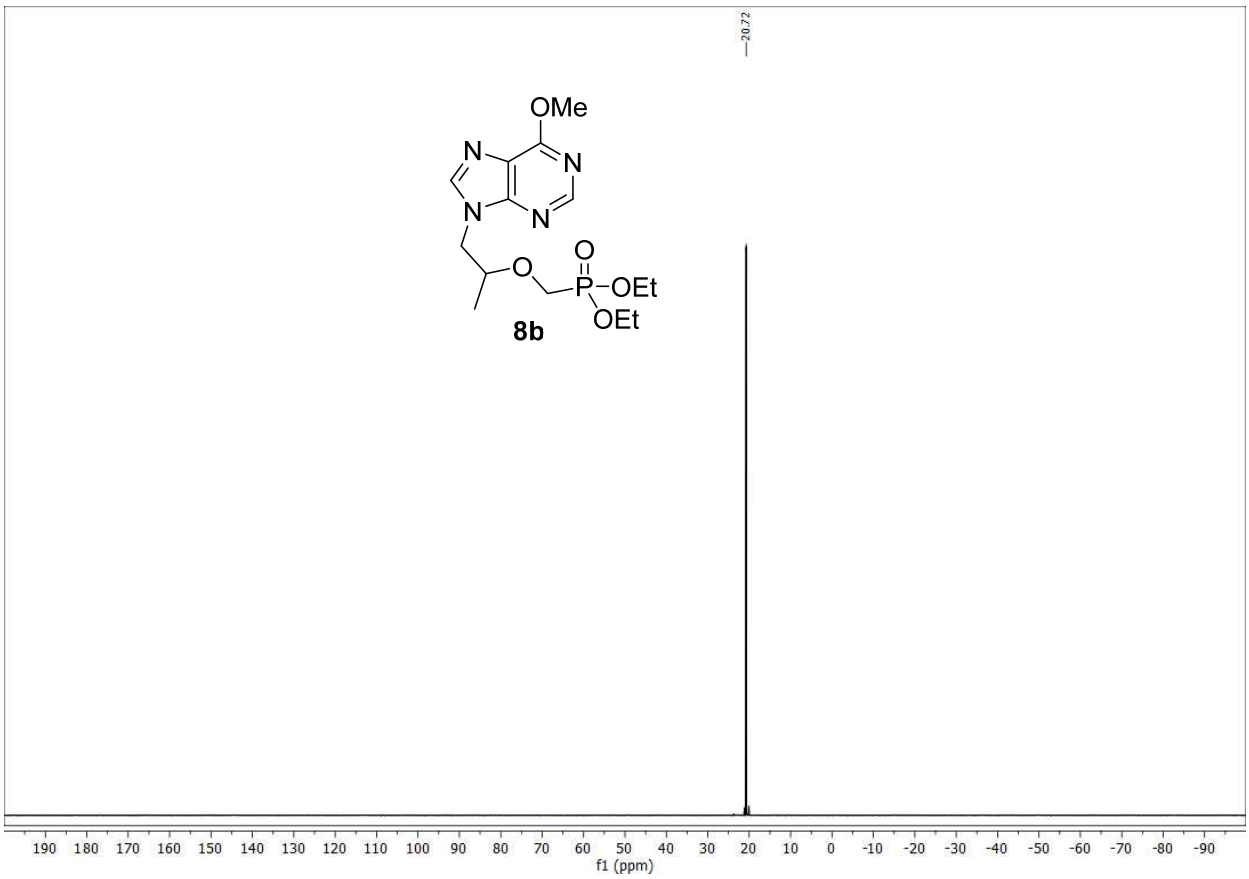
Compound 8b ^{1}H NMR



Compound 8b ^{13}C NMR

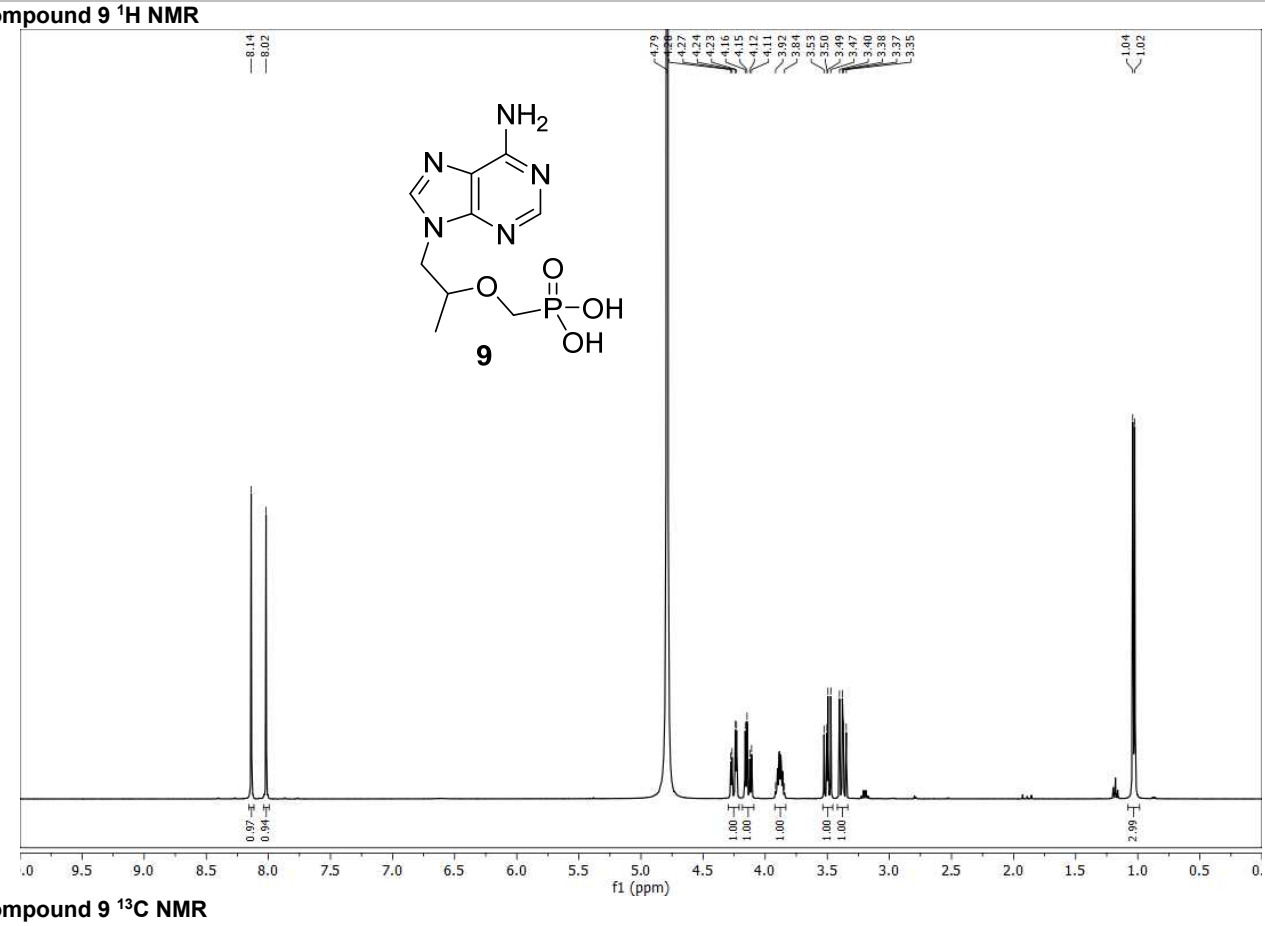


Compound 8b ^{31}P NMR

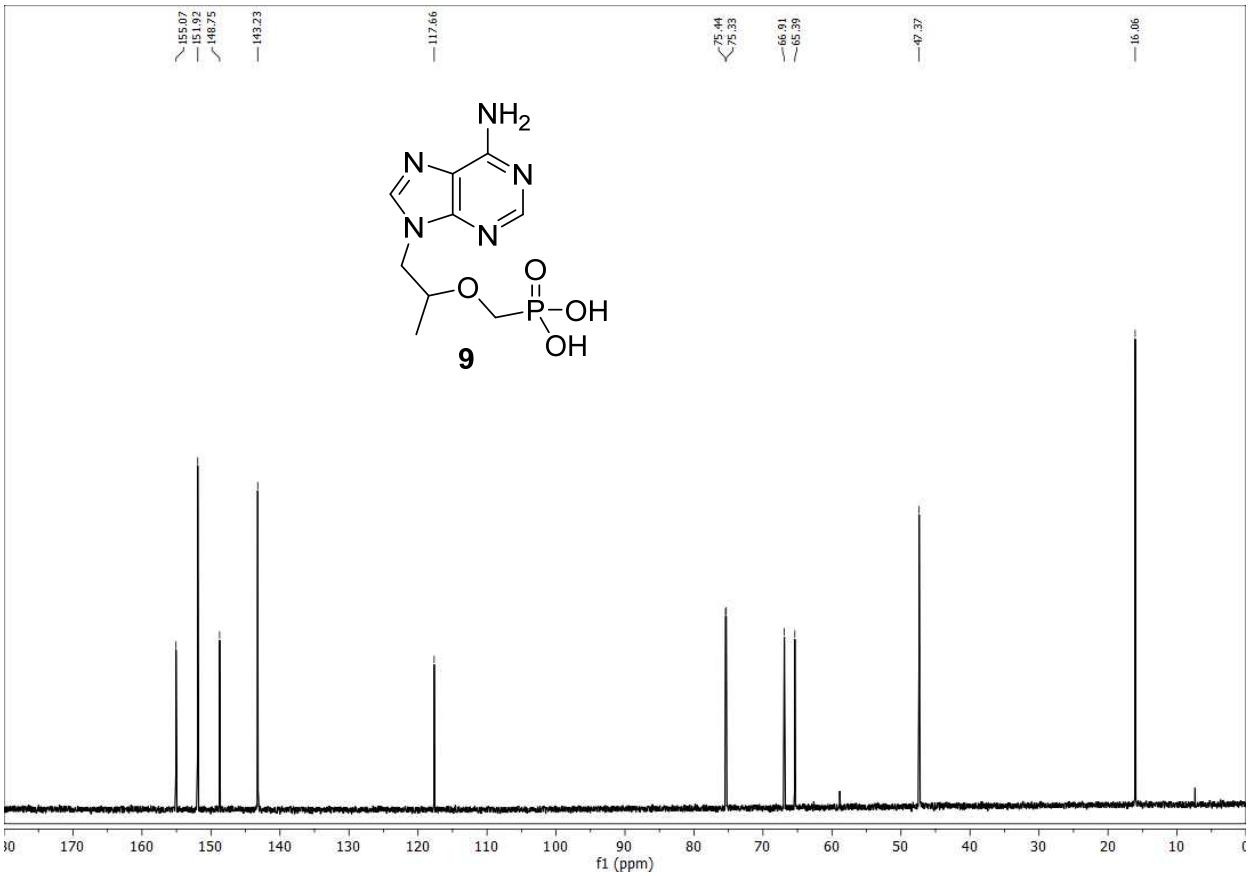


SUPPORTING INFORMATION

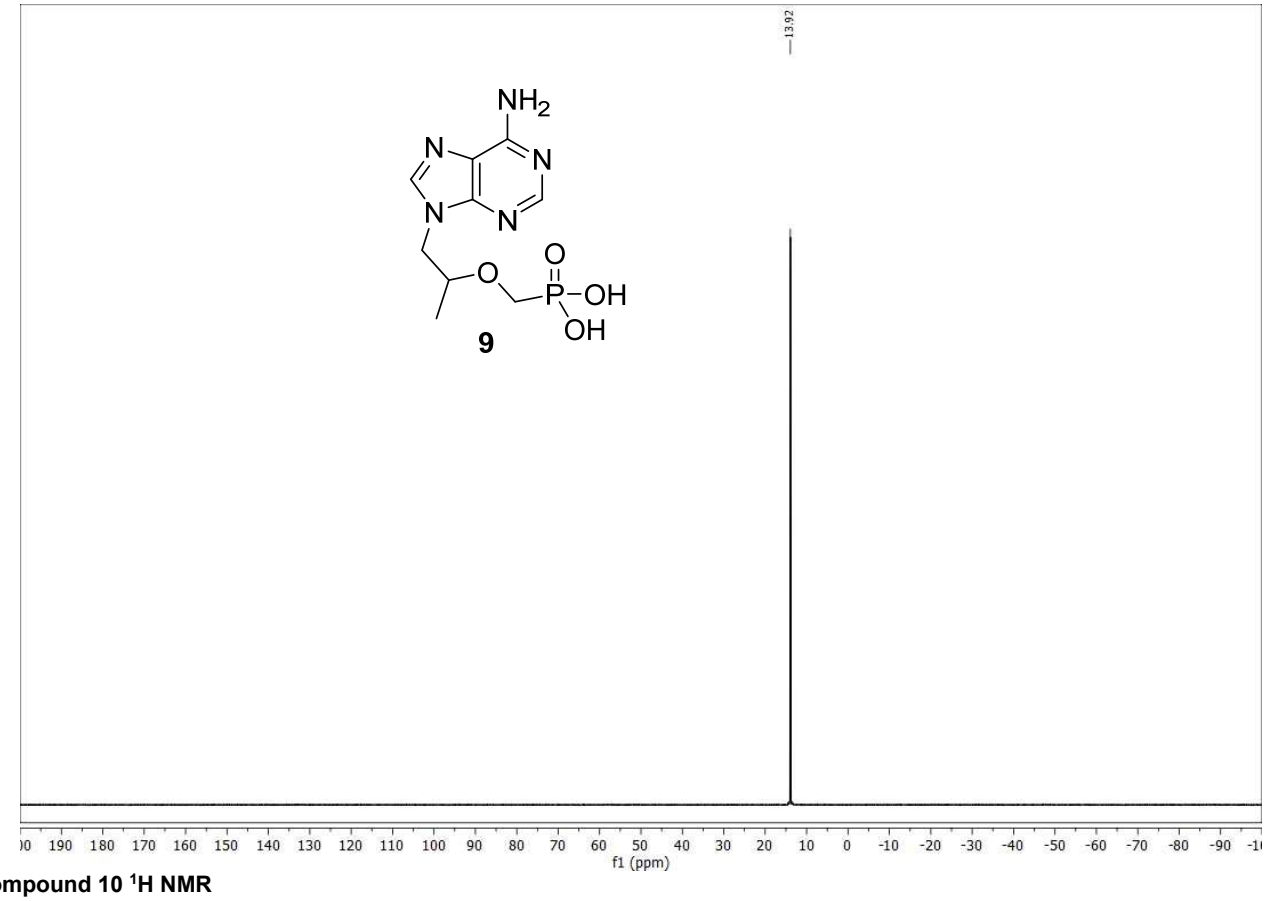
Compound 9 ^{1}H NMR



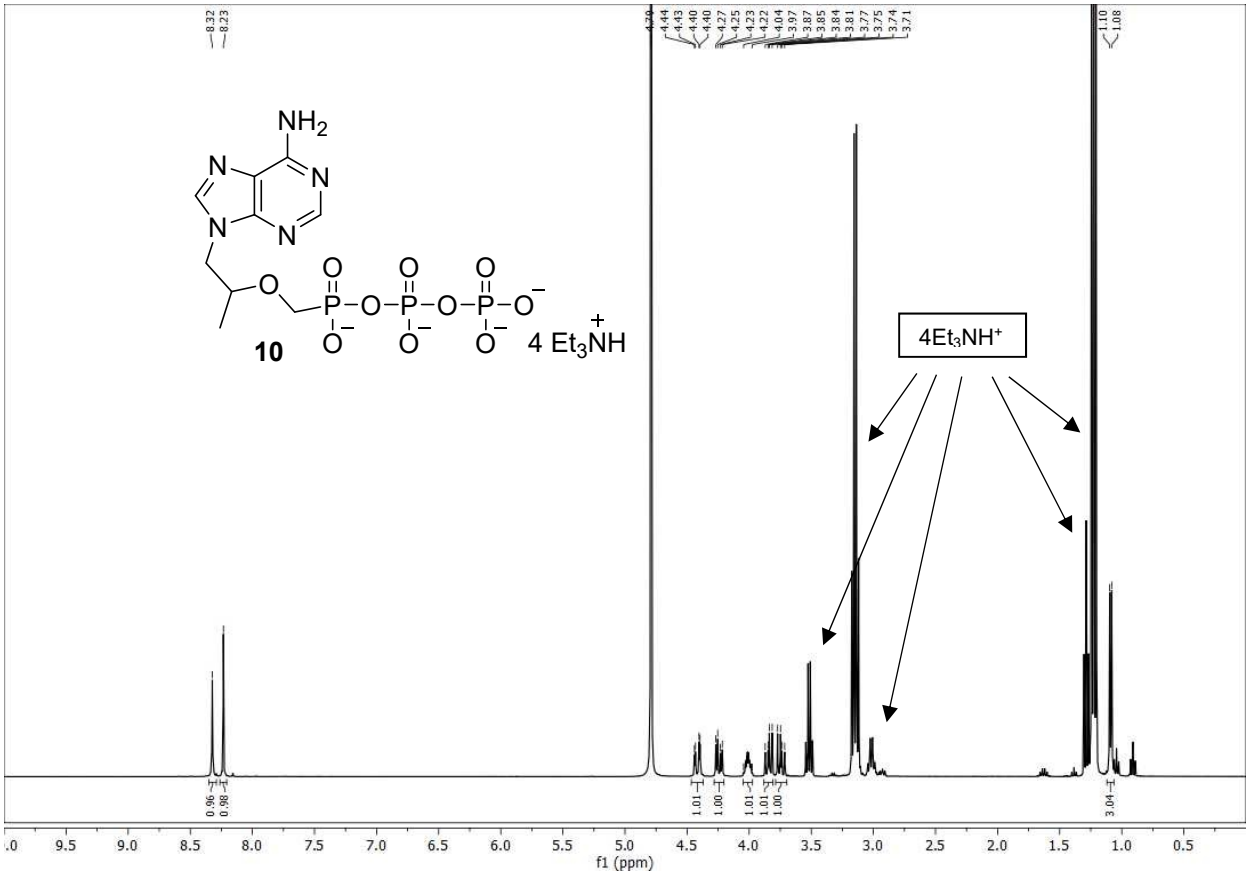
Compound 9 ^{13}C NMR



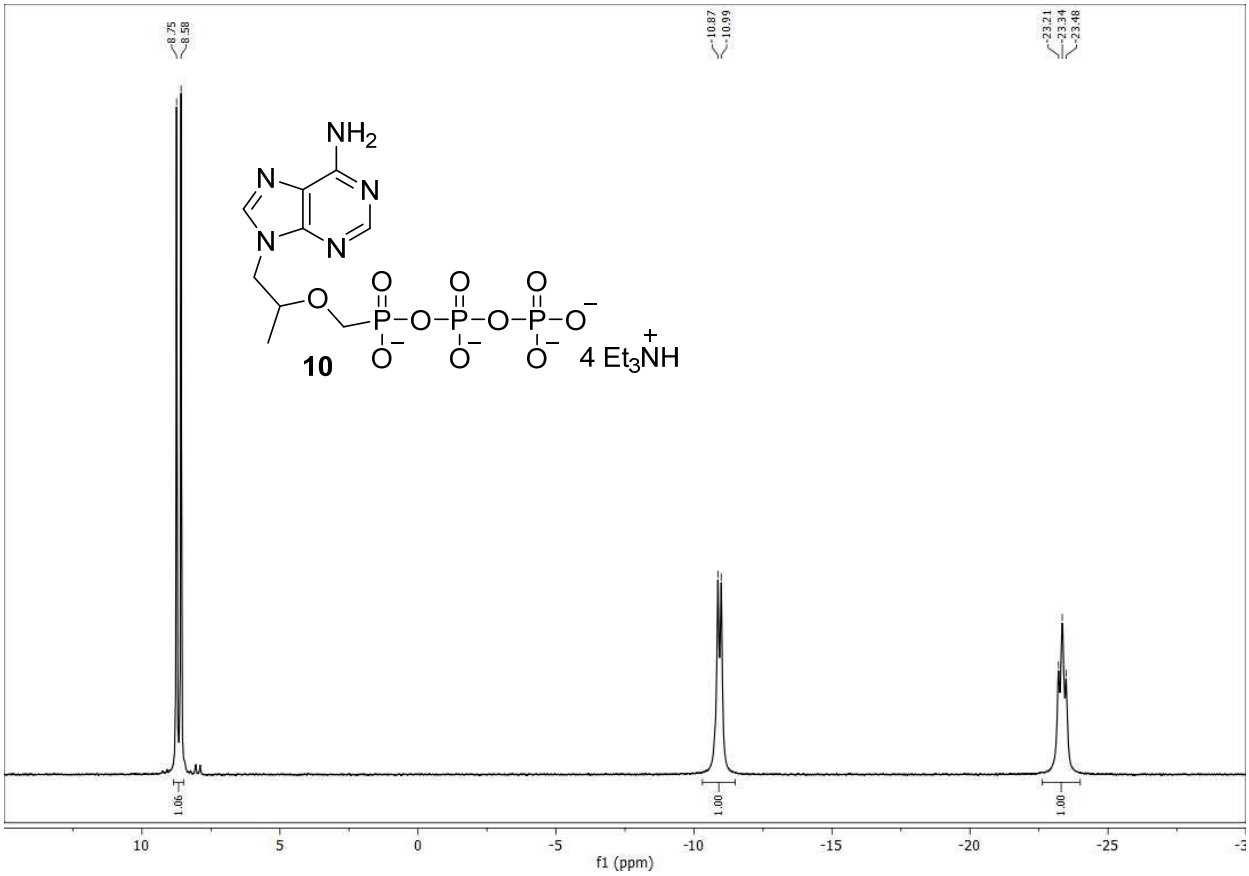
Compound 9 ^{31}P NMR



Compound 10 ^{1}H NMR

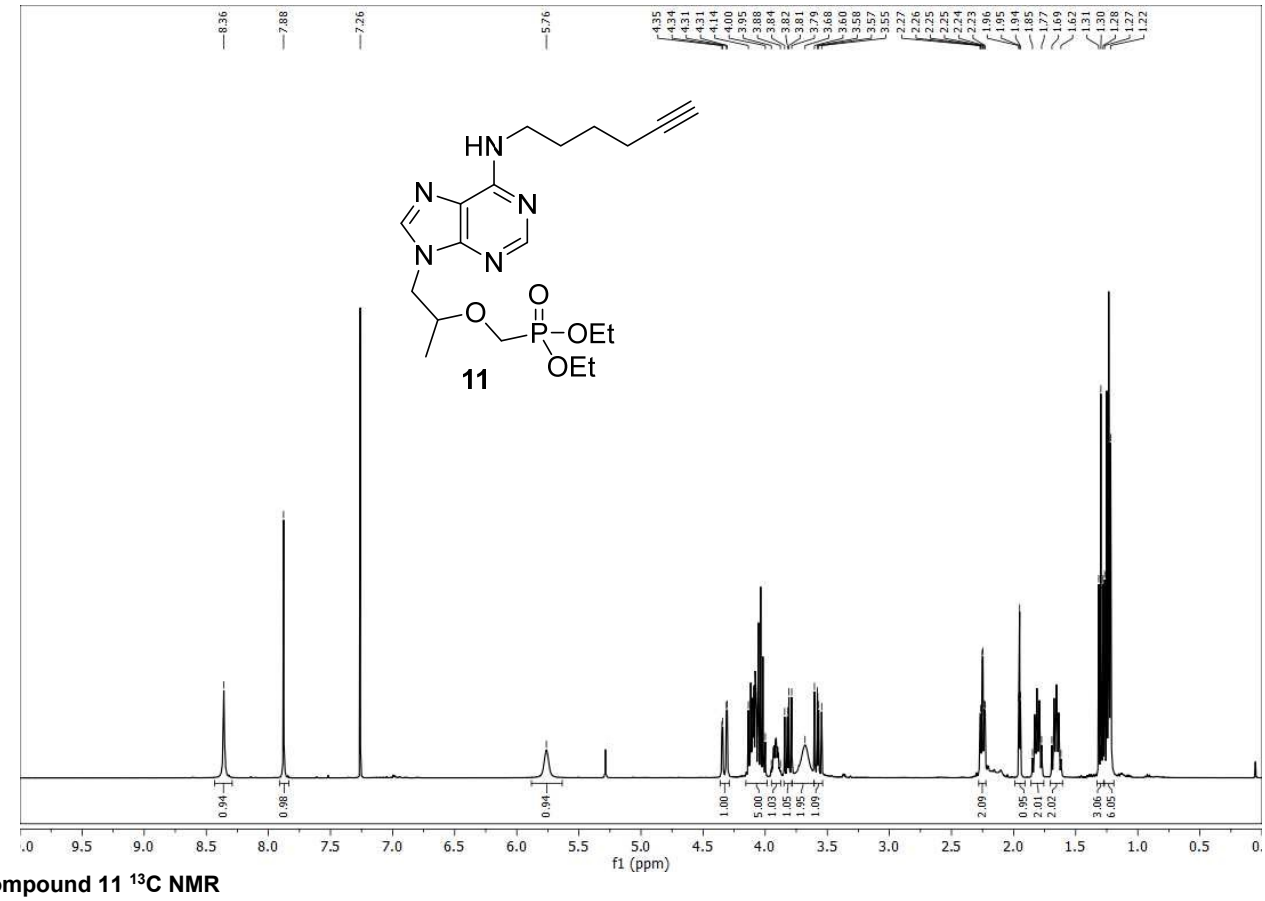


Compound 10 ^{31}P NMR

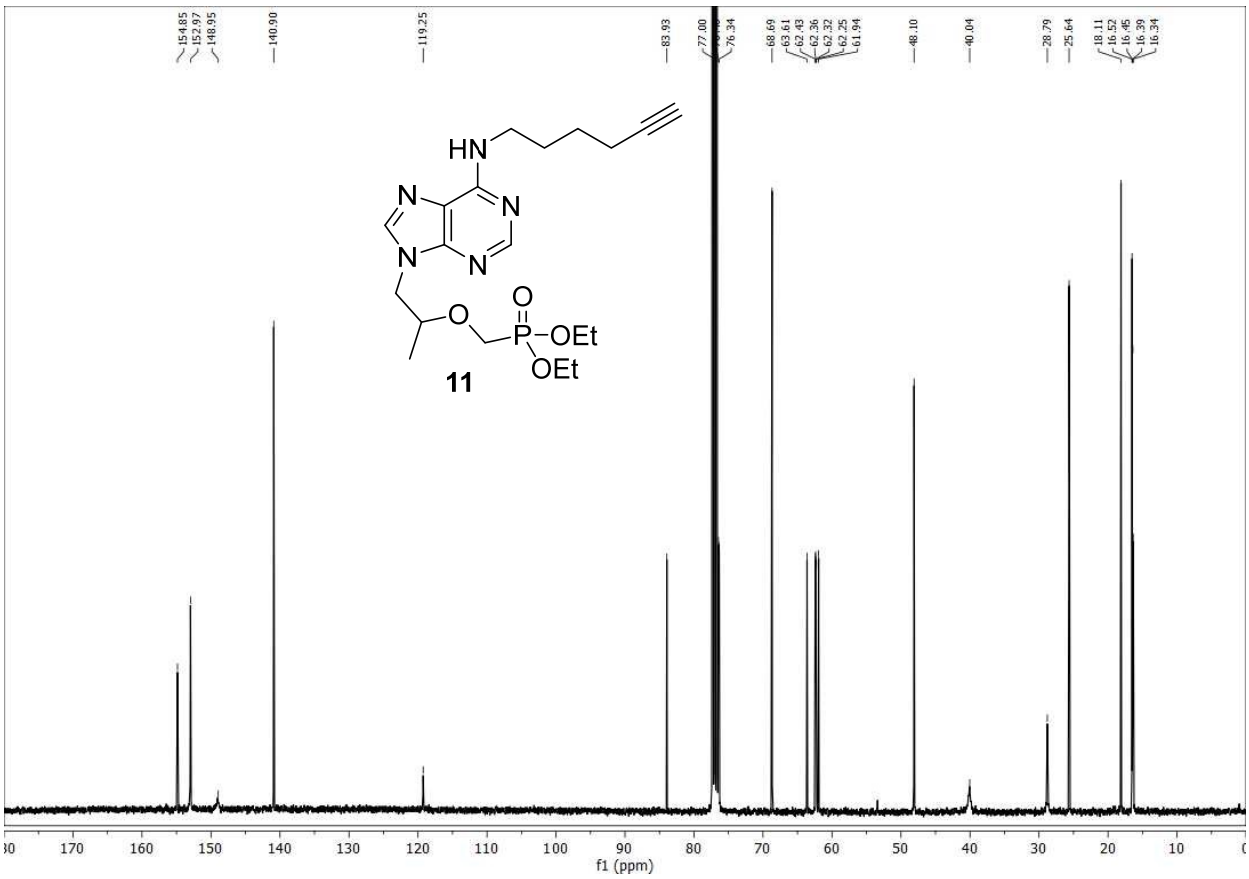


SUPPORTING INFORMATION

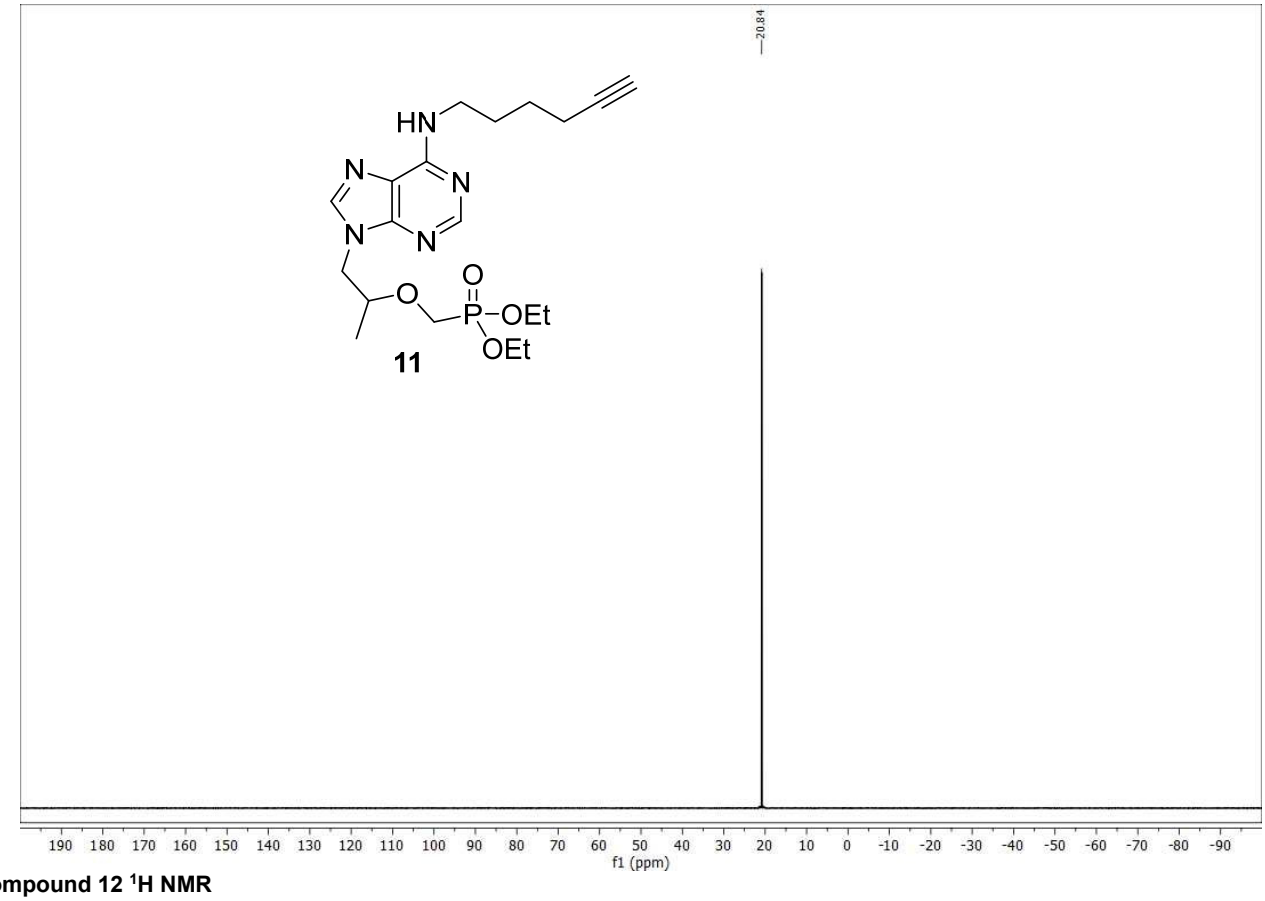
Compound 11 ^{1}H NMR



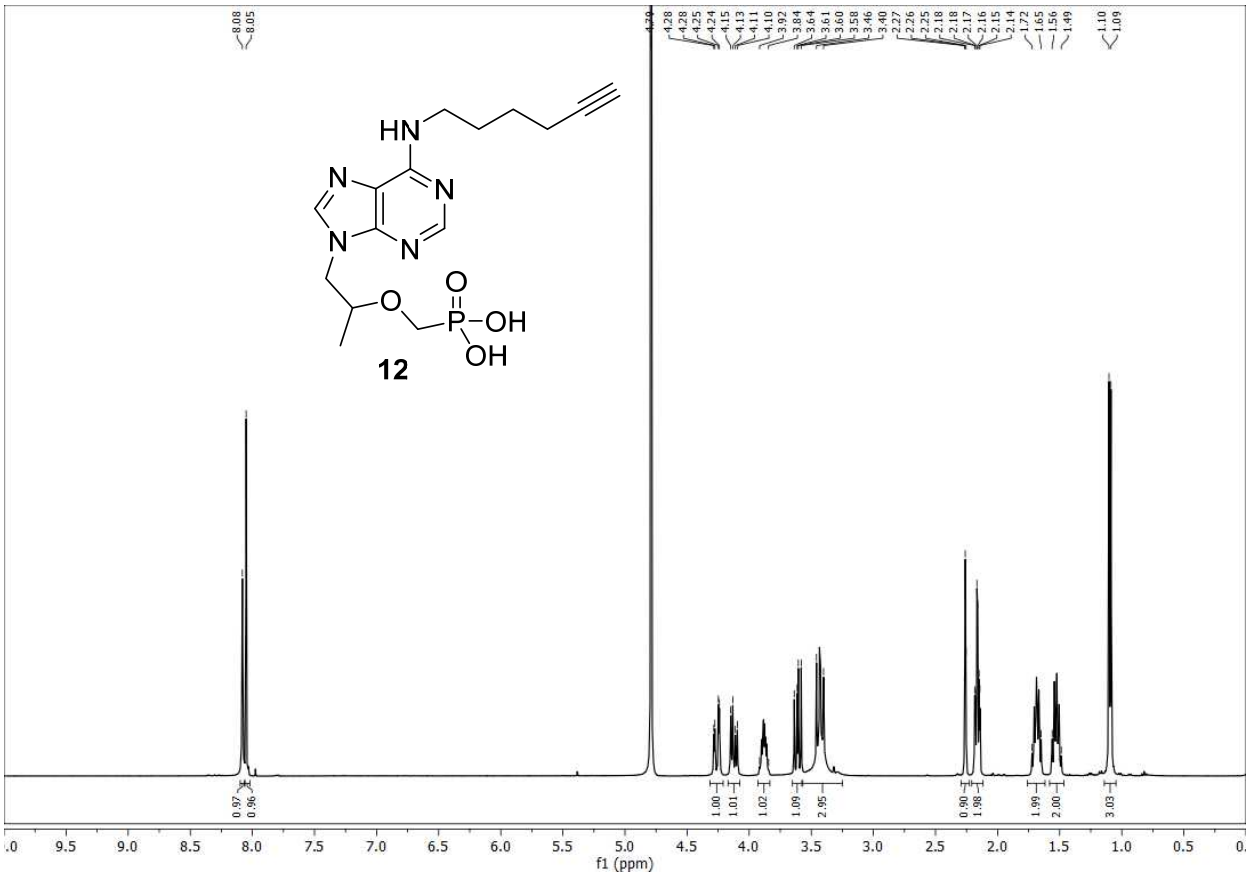
Compound 11 ^{13}C NMR



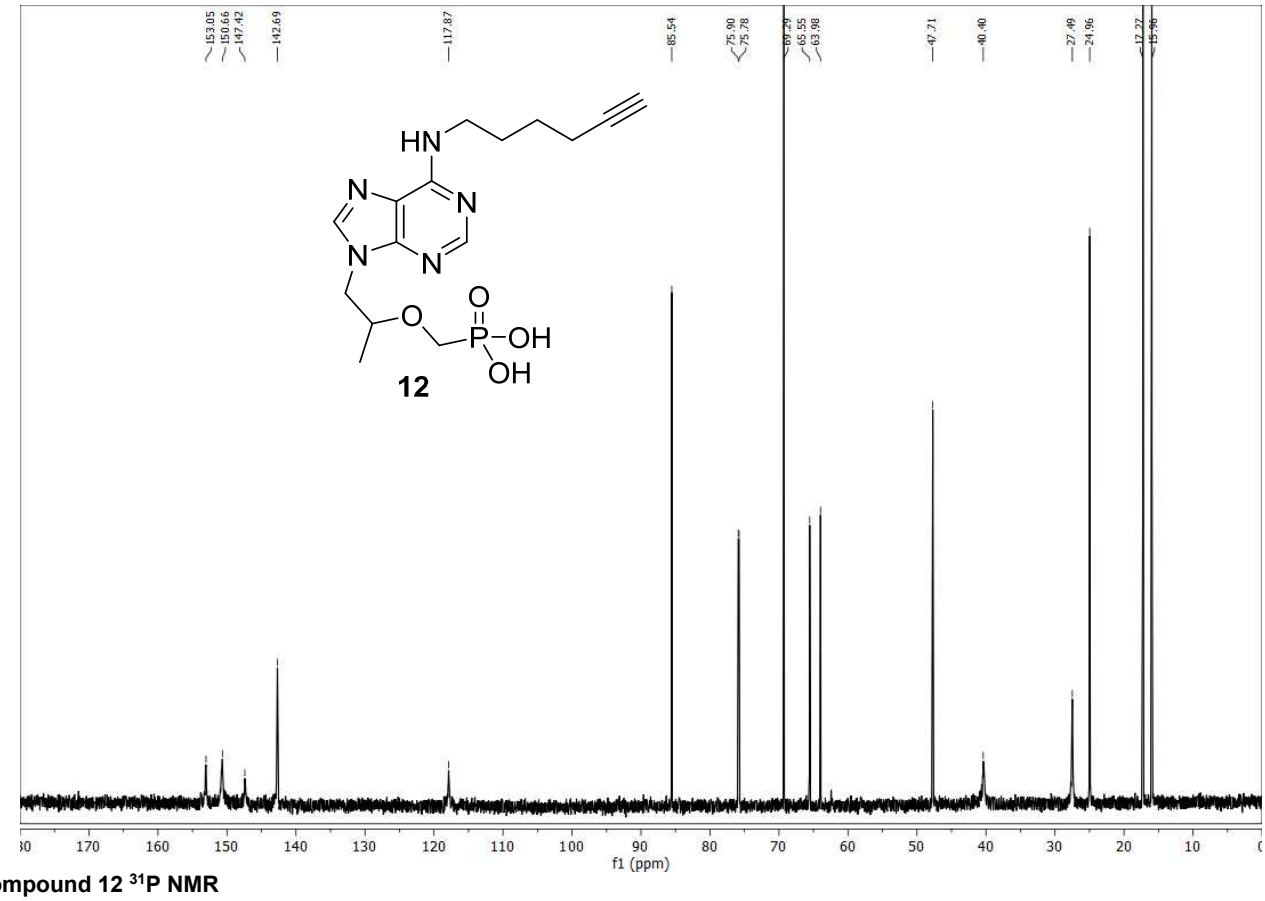
Compound 11 ^{31}P NMR



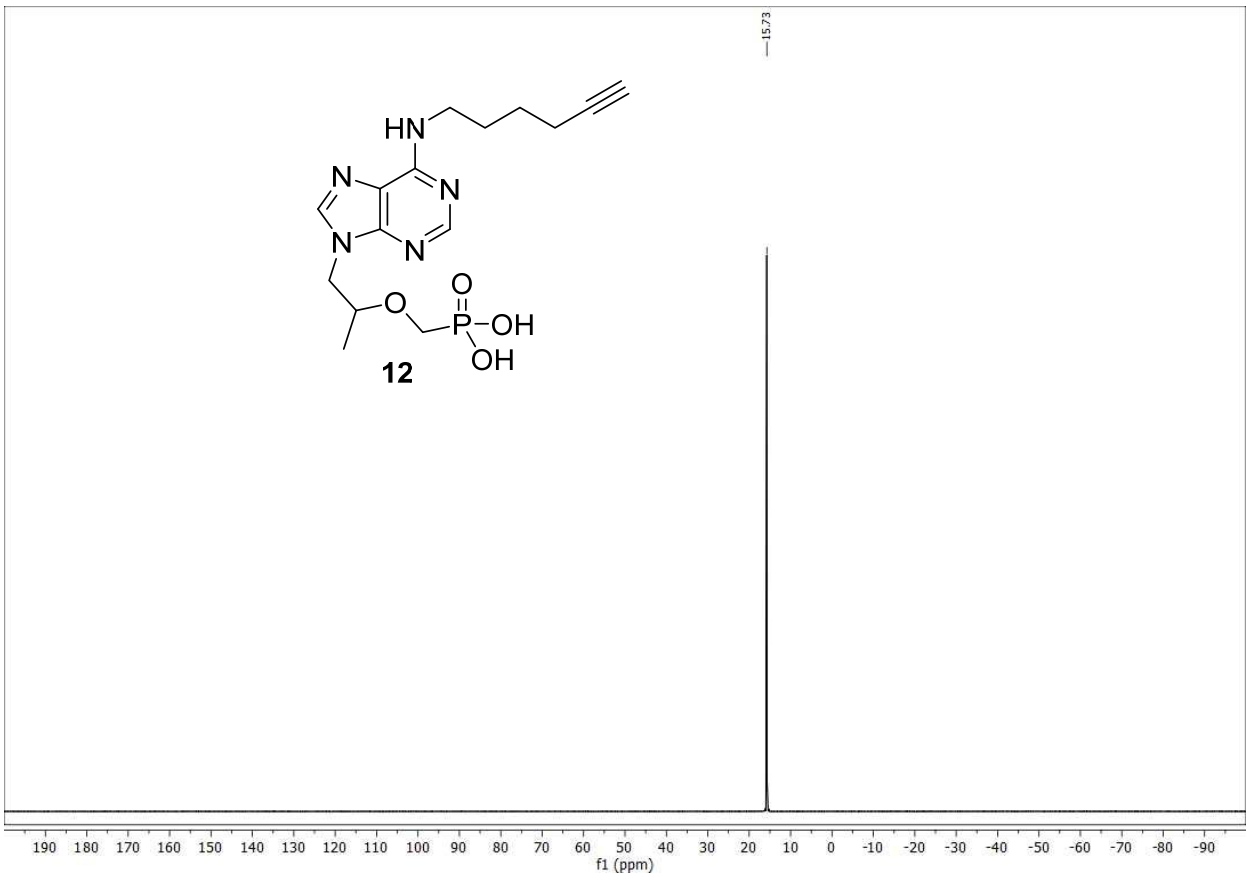
Compound 12 ^{1}H NMR



Compound 12 ^{13}C NMR

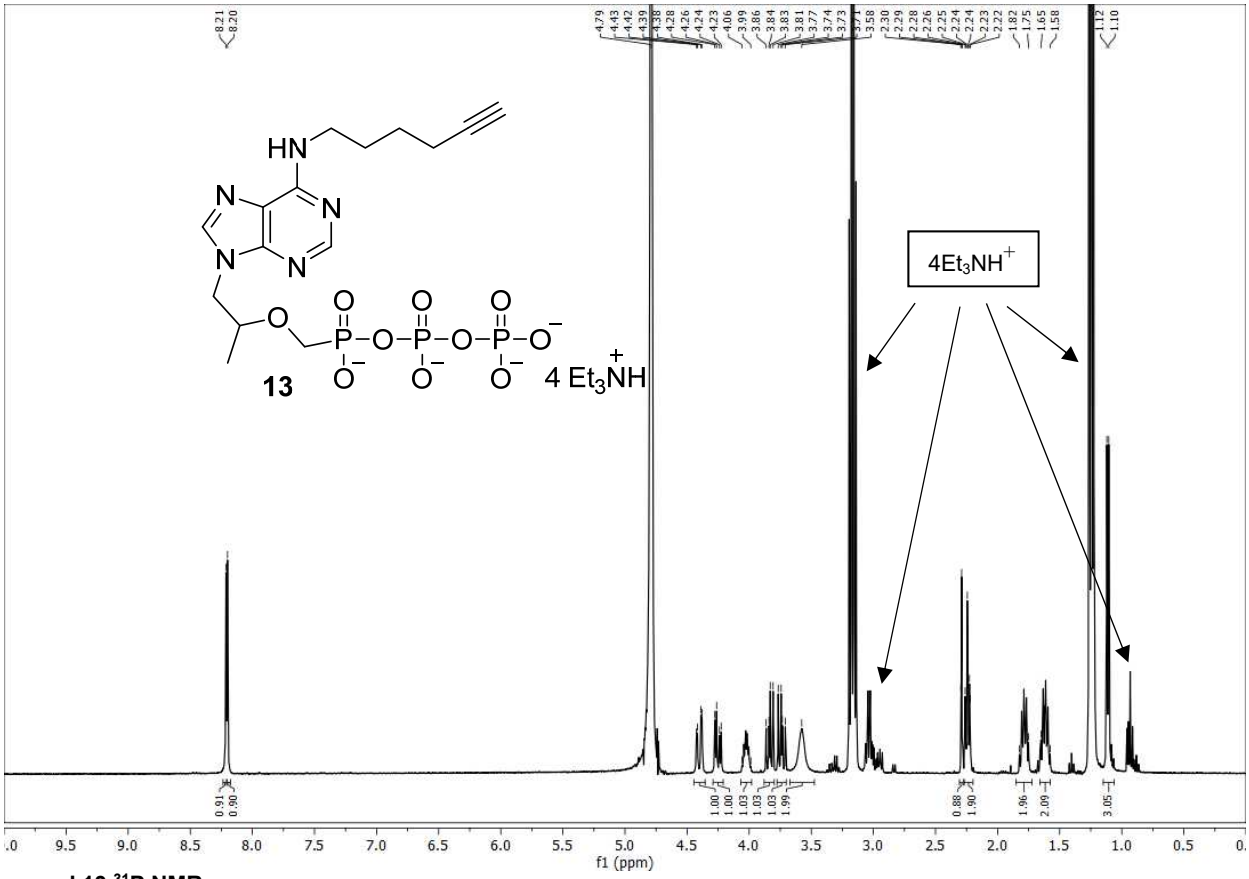


Compound 12 ^{31}P NMR

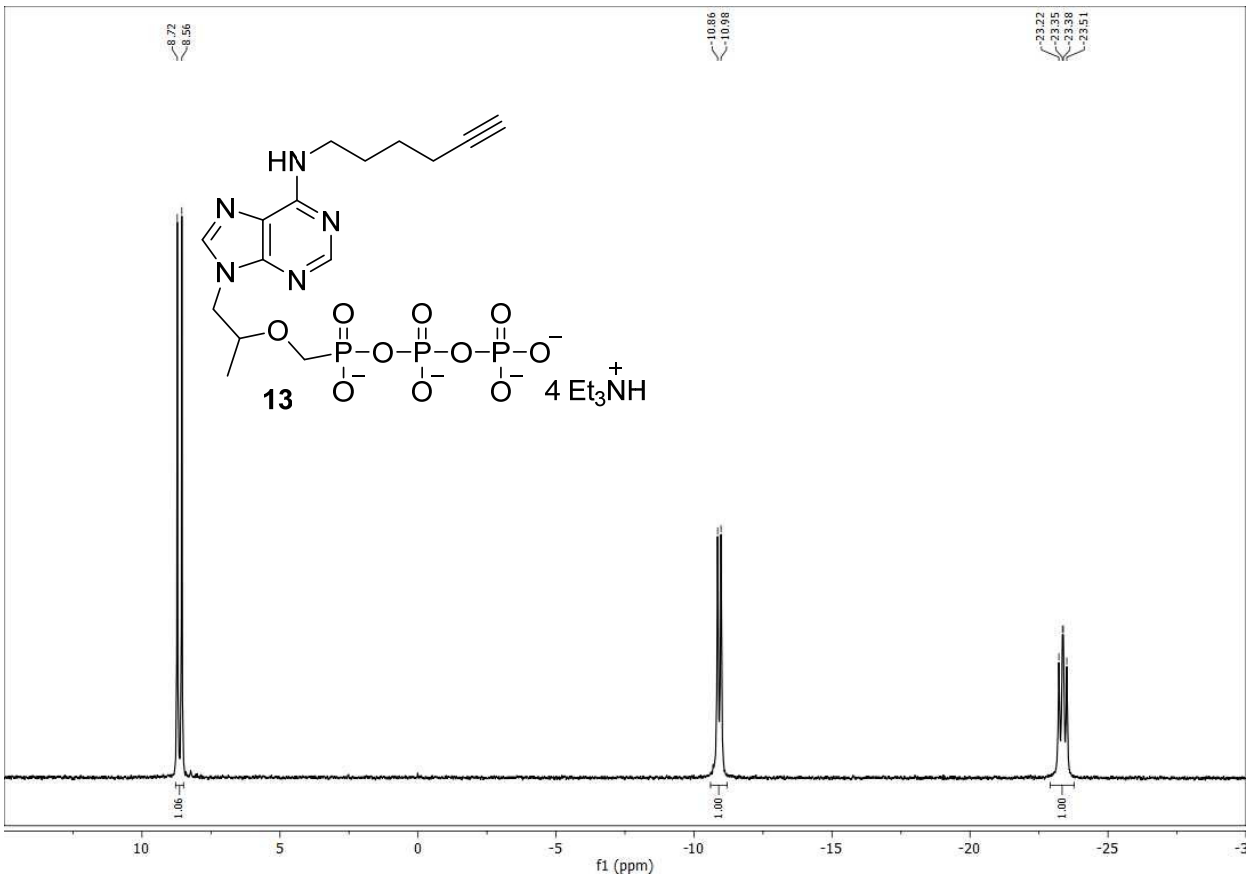


SUPPORTING INFORMATION

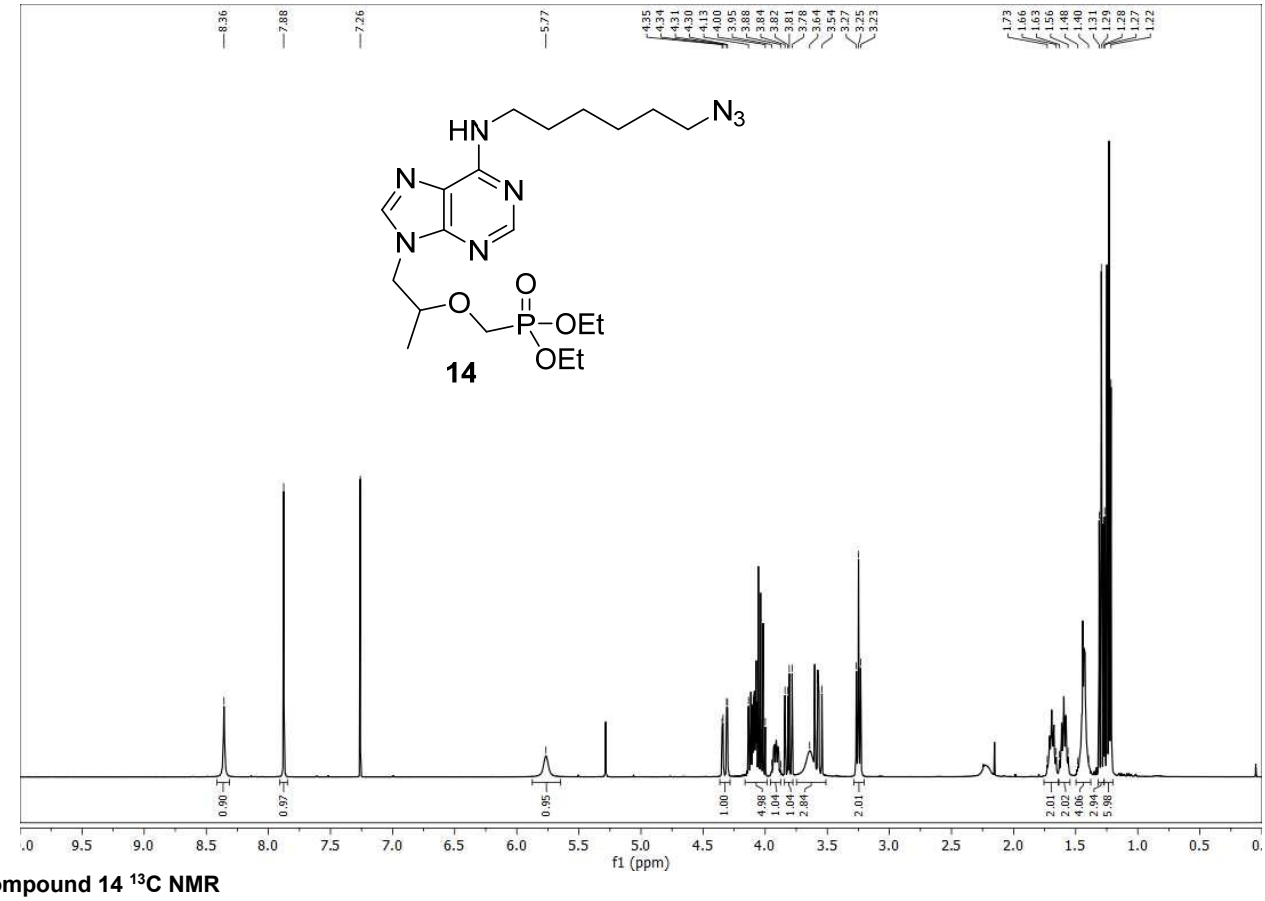
Compound 13 ^{1}H NMR



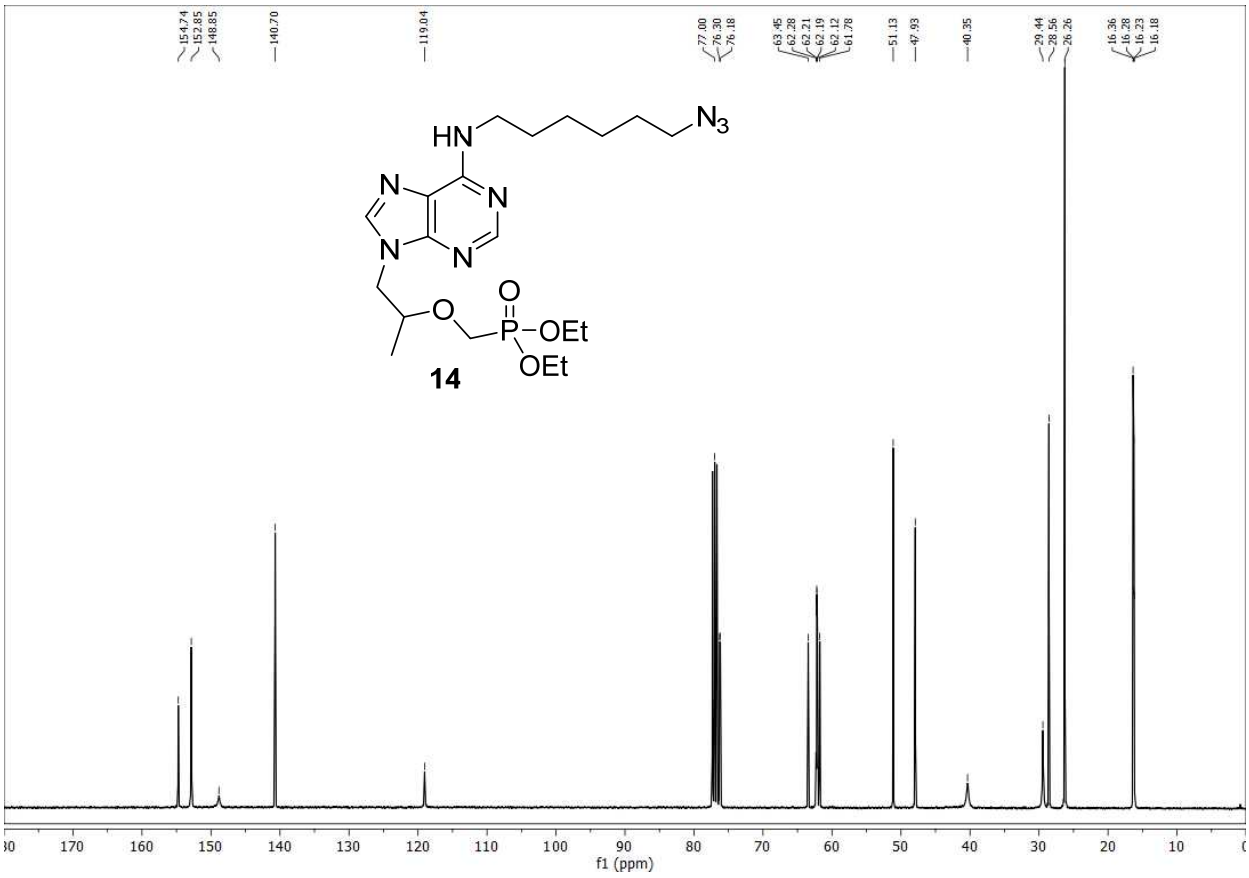
Compound 13 ^{31}P NMR



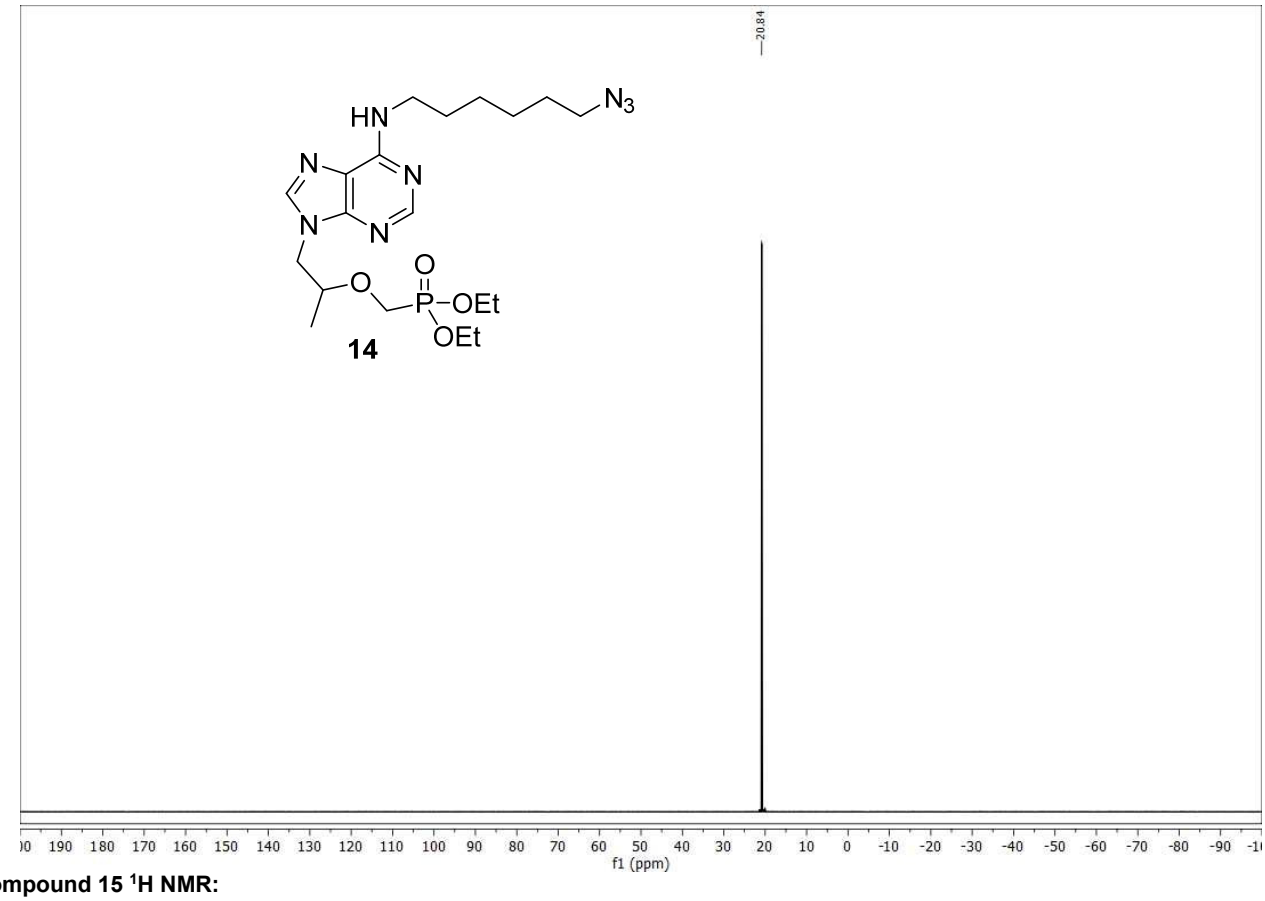
Compound 14 ^{1}H NMR



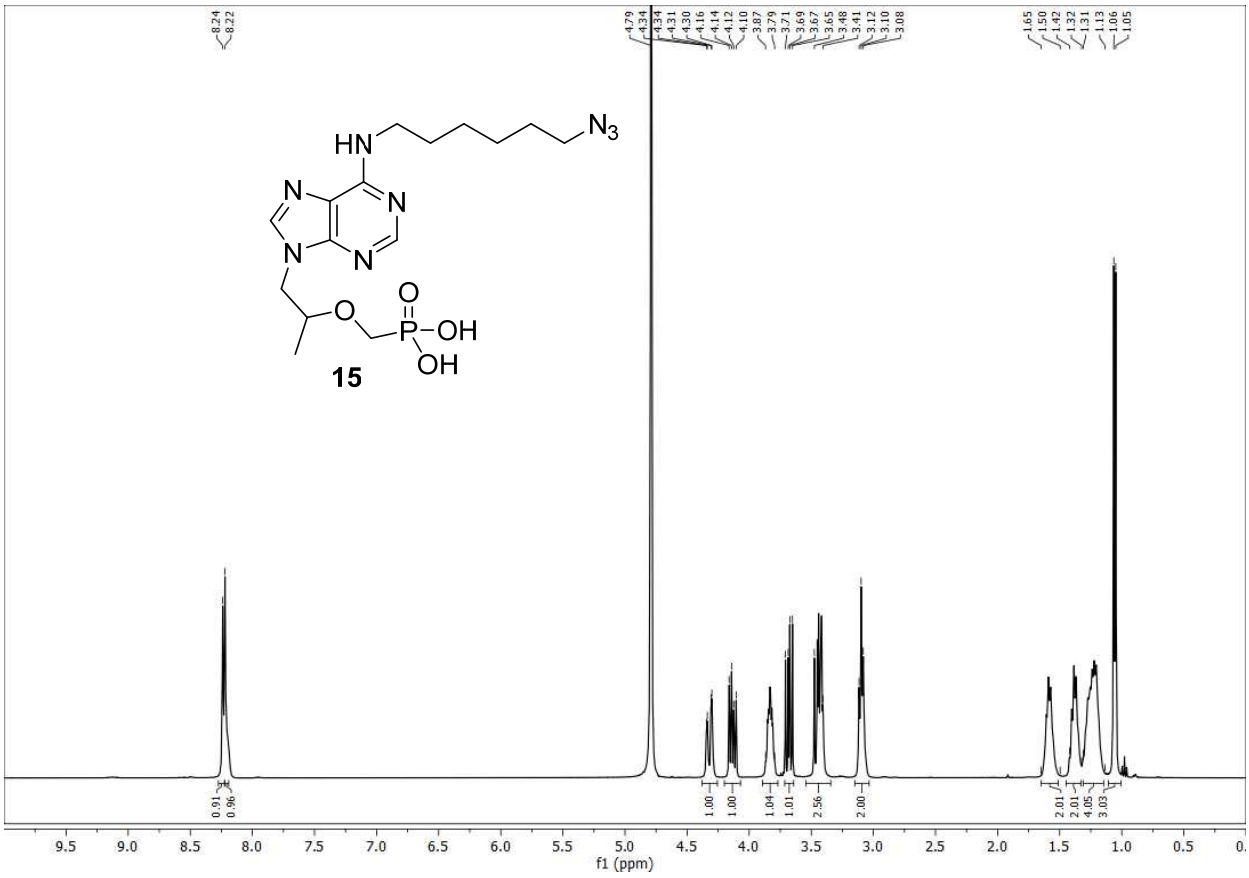
Compound 14 ^{13}C NMR



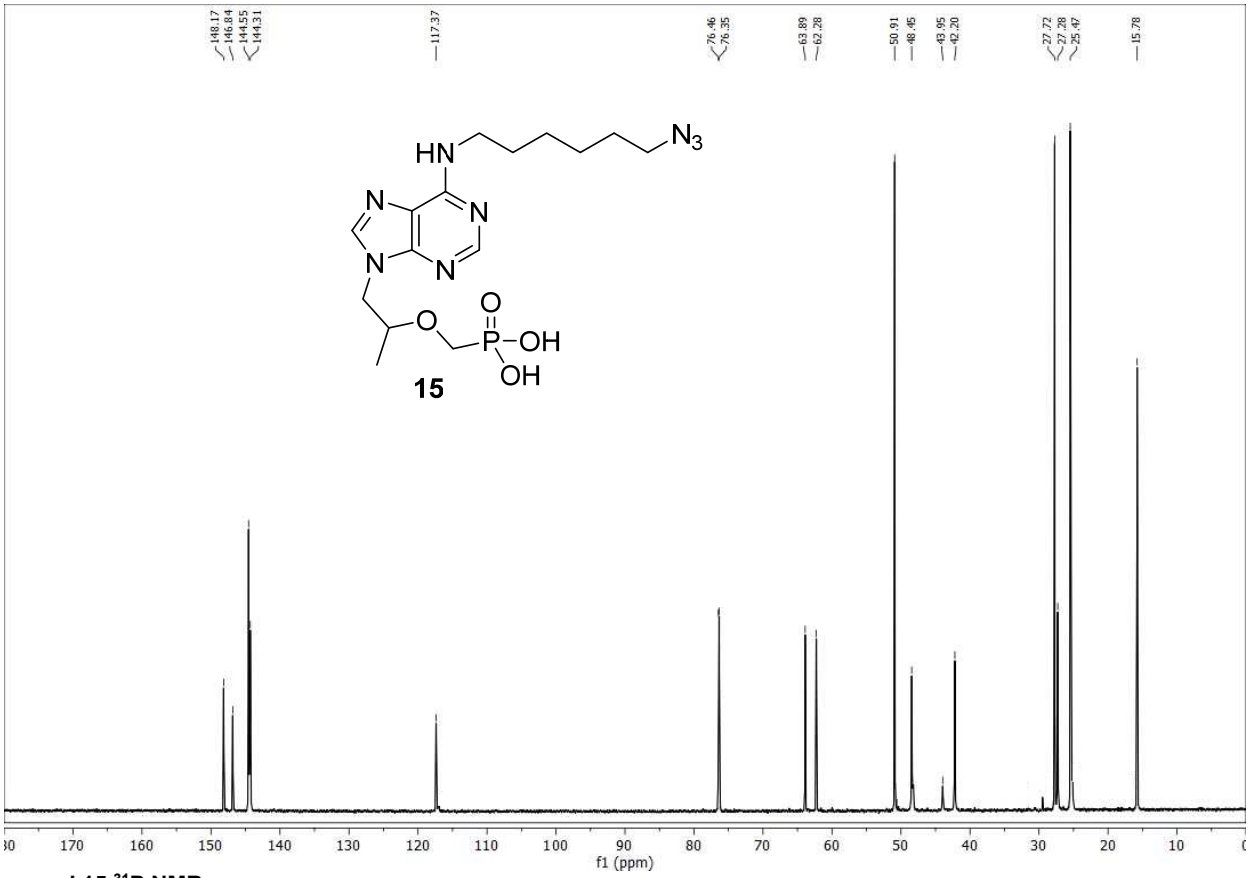
Compound 14 ^{31}P NMR:



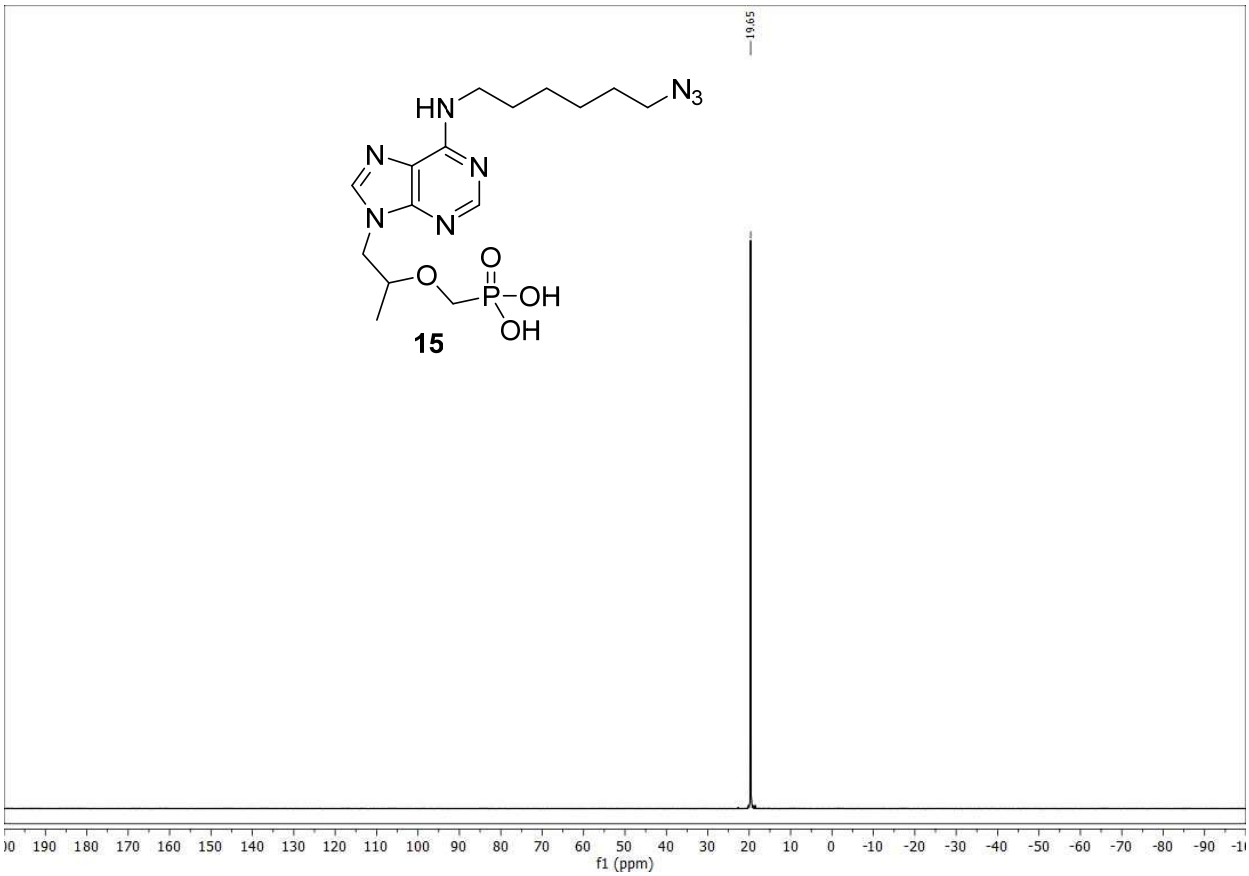
Compound 15 ^{1}H NMR:



Compound 15 ^{13}C NMR

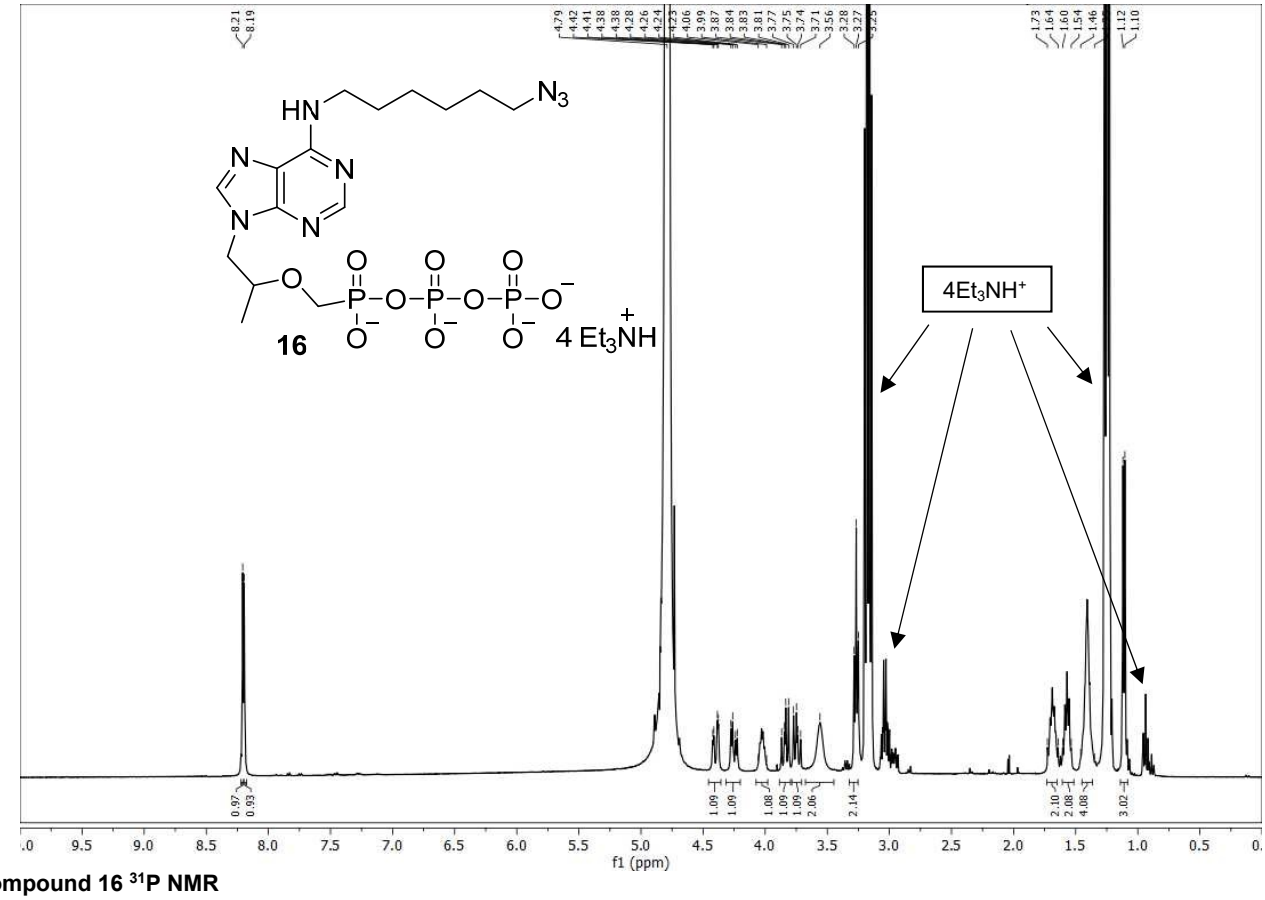


Compound 15 ^{31}P NMR

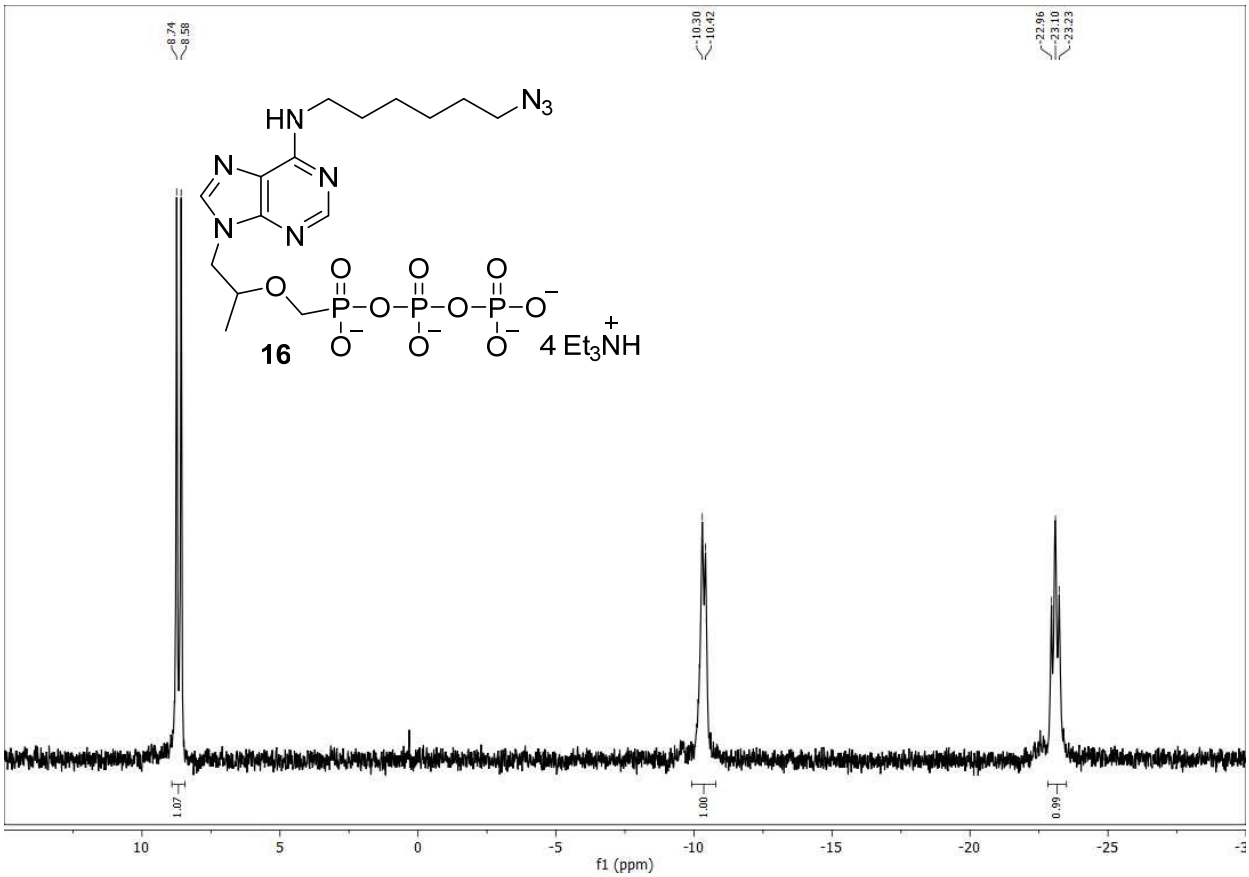


SUPPORTING INFORMATION

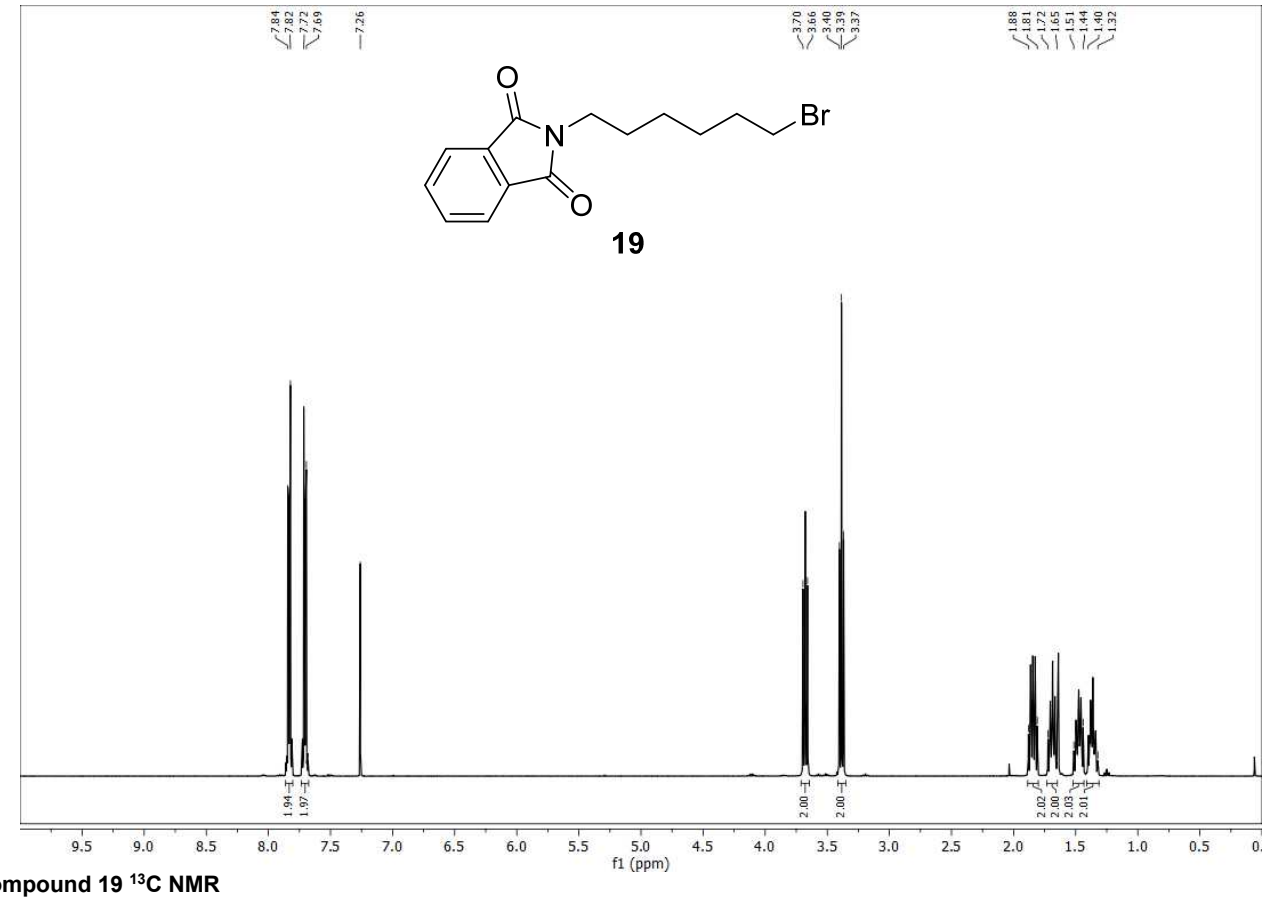
Compound 16 ^{1}H NMR



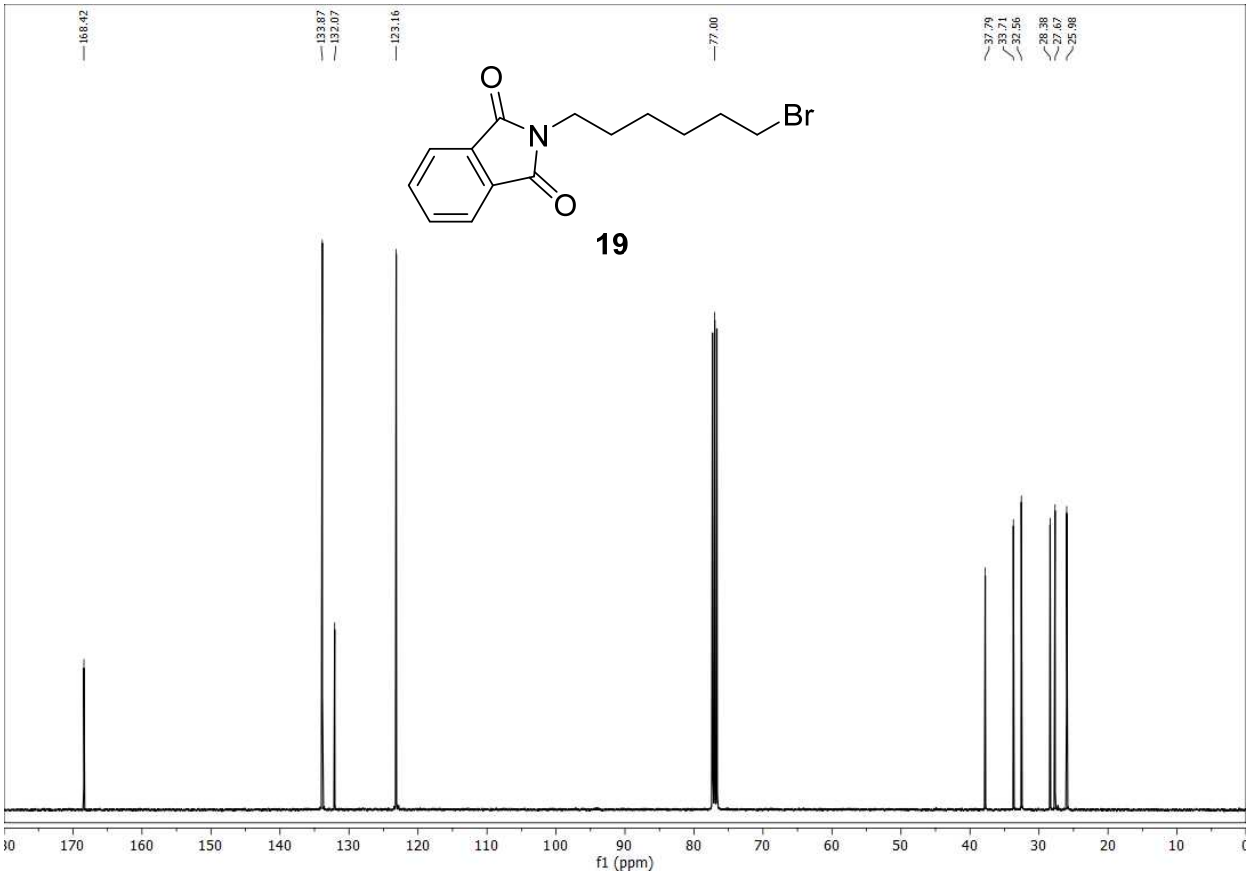
Compound 16 ^{31}P NMR



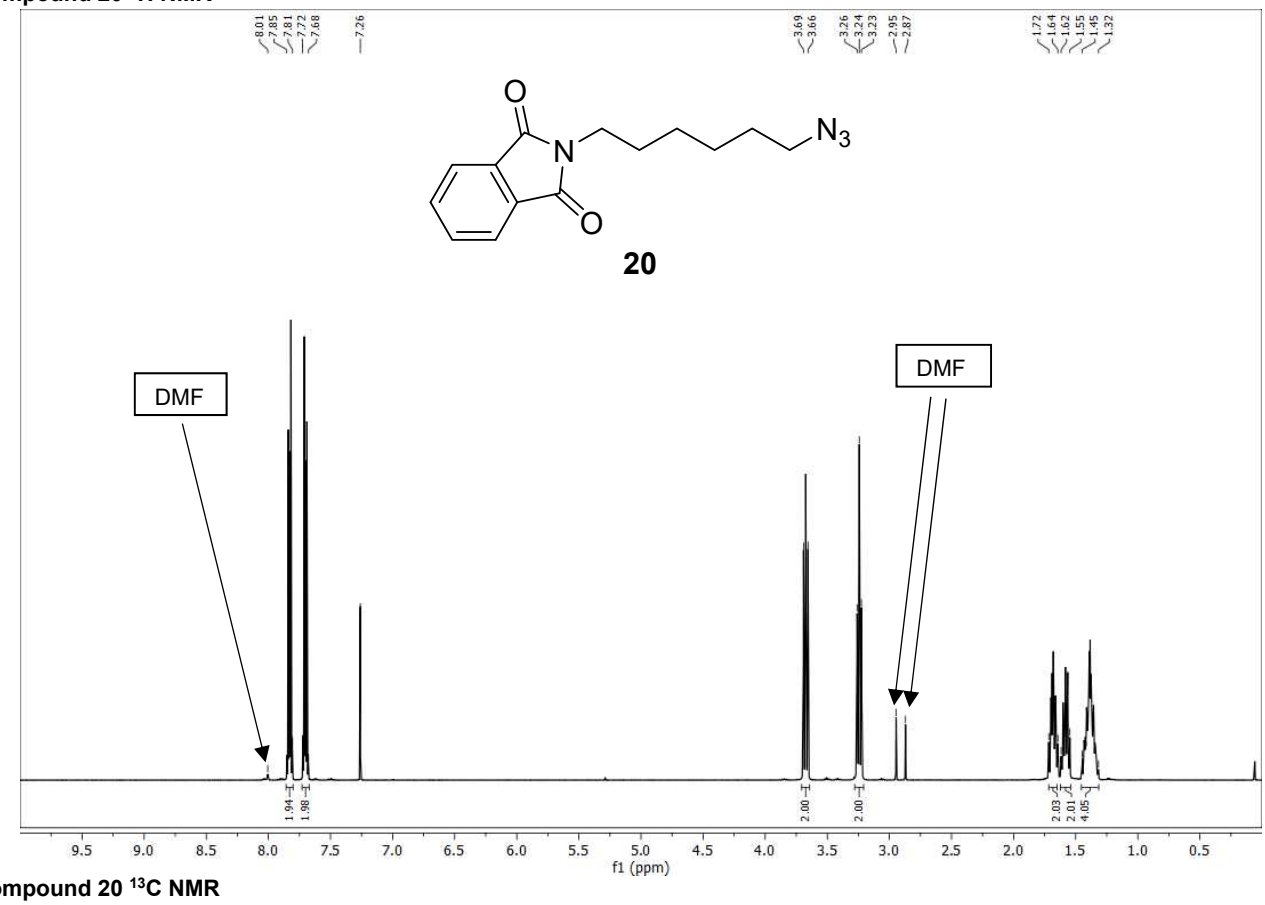
Compound 19 ^{1}H NMR



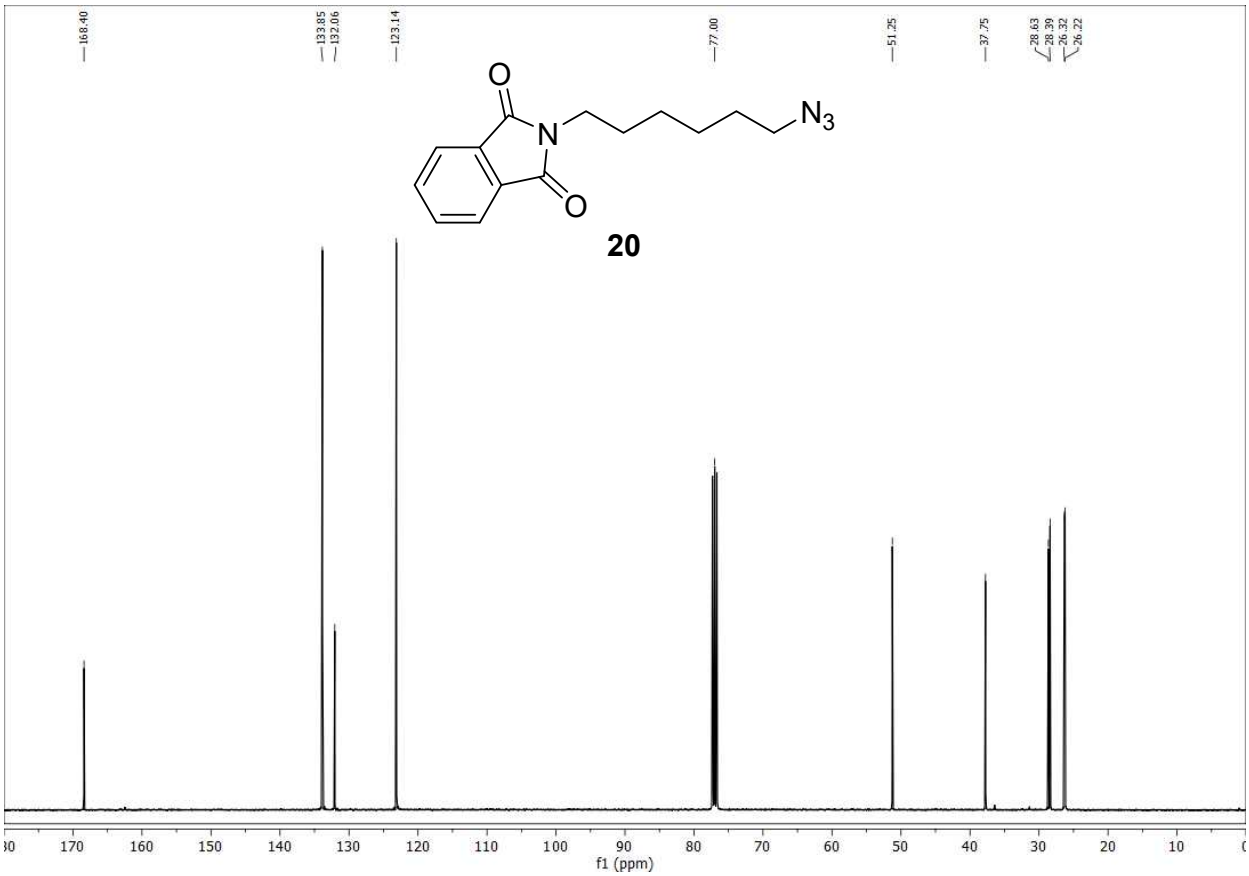
Compound 19 ^{13}C NMR



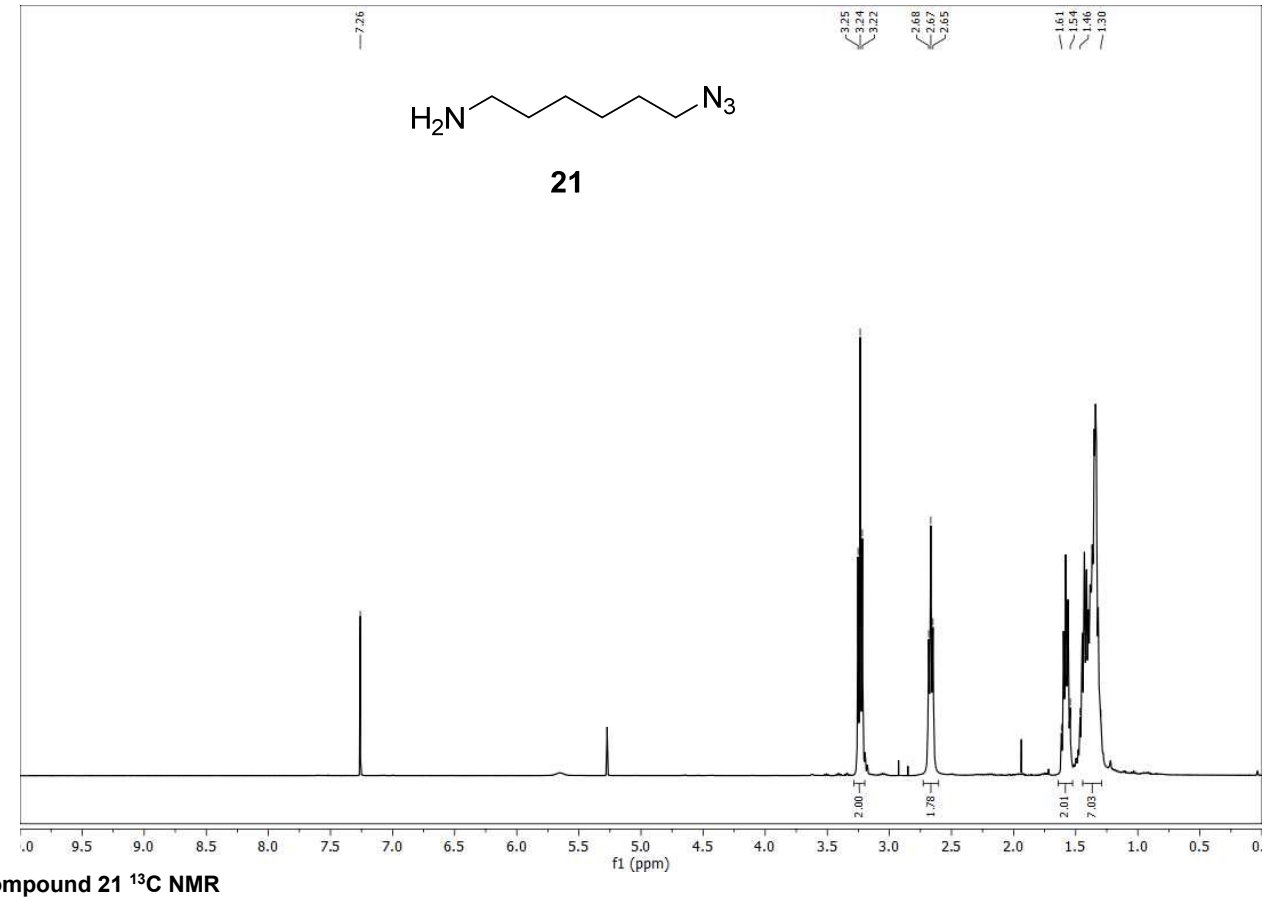
Compound 20 ^{1}H NMR



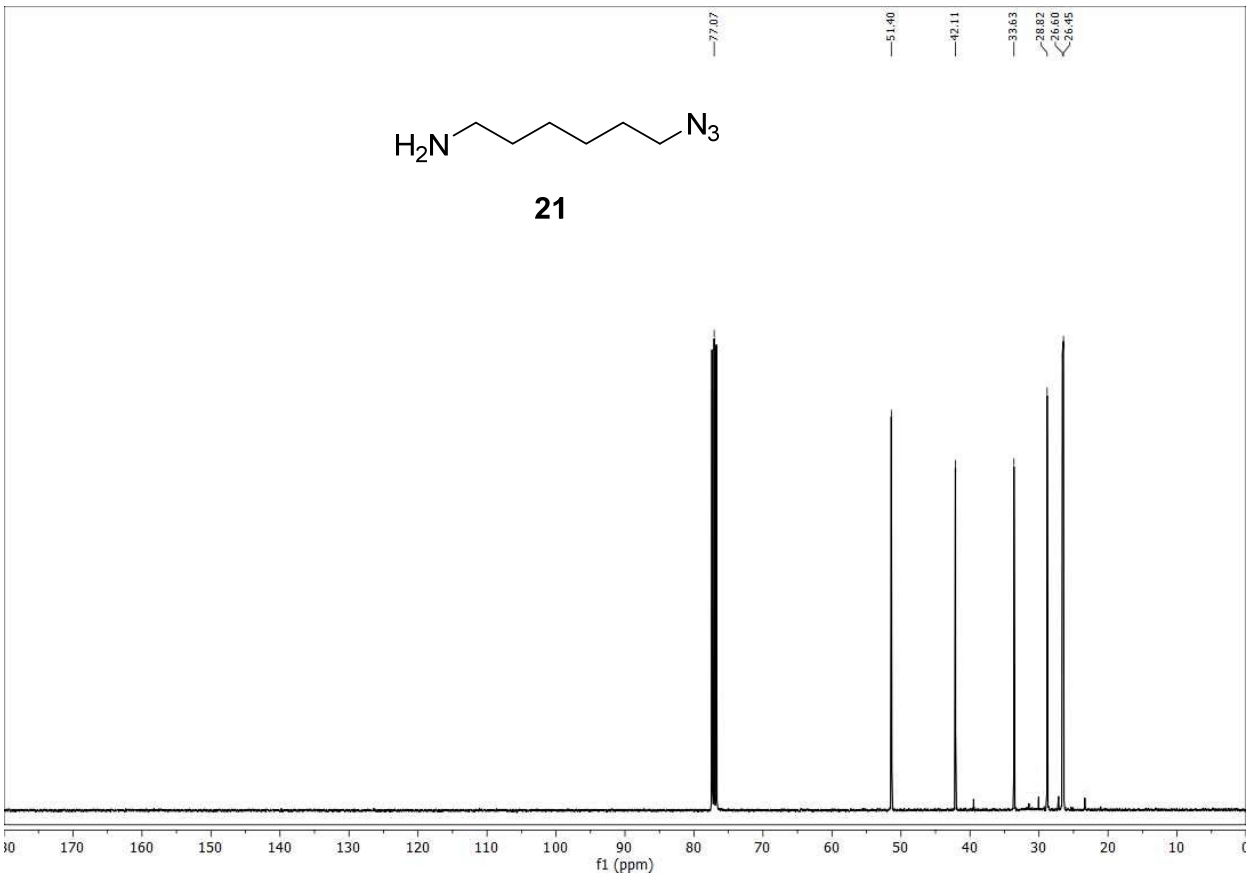
Compound 20 ^{13}C NMR



Compound 21 ^{1}H NMR



Compound 21 ^{13}C NMR



References

[1] M. Ghaem Maghami, C. P. M. Scheitl, C. Höbartner, *J Am Chem Soc* **2019**, *141*, 19546-19549.
[2] J. C. Kim, S. H. Kim, J. A. Kim, S. K. Choi, W. W. Park, *Arch Pharm Res* **1998**, *21*, 458-464.
[3] A. L. Jeffery, J. H. Kim, D. F. Wiemer, *Tetrahedron* **2000**, *56*, 5077-5083.
[4] Z. Liu, N. Yasuda, M. Simeone, R. A. Reamer, *J Org Chem* **2014**, *79*, 11792-11796.
[5] a) D. E. Hoard, D. G. Ott, *J Am Chem Soc* **1965**, *87*, 1785-1788; b) E. S. Simon, S. Grabowski, G. M. Whitesides, *Journal of Organic Chemistry* **1990**, *55*, 1834-1841.
[6] L. M. Schultze, H. H. Chapman, N. J. P. Dubree, R. J. Jones, K. M. Kent, T. T. Lee, M. S. Louie, M. J. Postich, E. J. Prisbe, J. C. Rohloff, R. H. Yu, *Tetrahedron Lett* **1998**, *39*, 1853-1856.
[7] C. M. Leenders, G. Jansen, M. M. Frissen, R. P. Lafleur, I. K. Voets, A. R. Palmans, E. W. Meijer, *Chemistry* **2016**, *22*, 4608-4615.
[8] M. B. Stead, A. Agrawal, K. E. Bowden, R. Nasir, B. K. Mohanty, R. B. Meagher, S. R. Kushner, *Nucleic Acids Res* **2012**, *40*, e156.
[9] K. K. Alam, J. L. Chang, D. H. Burke, *Mol Ther Nucleic Acids* **2015**, *4*, e230.
[10] http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi.
[11] 16S rRNA: https://rnacentral.org/rna/URS00000ABFE9/562; 23S rRNA: https://rnacentral.org/rna/URS00004B0F34/562;