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
Table of Contents

1. Materials and DNA/RNA oligonucleotides ................................................................. 3
   1.1. General materials ......................................................................................... 3
   1.2. DNA oligonucleotides: Primers and transcription templates ......................... 3
   1.3. RNA oligonucleotides: transcripts and synthetic RNAs ............................... 4
2. Experimental Procedures ......................................................................................... 6
   2.1. Synthesis of tenofovir diphosphate analogs ............................................... 6
   2.2. In vitro selection of ribozymes using Biotin-tenofovir-DP .......................... 17
   2.3. Characterization and application of FJ ribozymes ...................................... 17
       2.3.1 Single-turnover kinetic experiments ...................................................... 17
       2.3.2 Preparative tenofovir labeling and probing of the modification site .......... 18
       2.3.3 Application of FJ ribozymes for labeling of cellular RNA .................. 18
       2.3.4 Mutually orthogonal dual-color labeling of a synthetic transcript .......... 18
       2.3.5 Probing of the modification site on the target cellular RNA using reverse transcription 19
3. Results and Discussion ......................................................................................... 20
   3.1 Sequencing results of FJ ribozyme selection ............................................... 20
       3.1.1. FJ ribozyme sequences identified by Sanger sequencing ...................... 20
       3.1.2. FJ ribozyme sequences identified by Illumina sequencing ................... 20
   3.2 Supporting Figures S1-S17 ............................................................................... 22
   NMR spectra ......................................................................................................... 33
References .................................................................................................................. 59
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. (Diethoxy phosphoryl)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 (1H, 13C and 31P) were recorded using 400 MHz Bruker Avance III and Avance III HD spectrometers. Chemical shifts (1H and 13C) 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

<table>
<thead>
<tr>
<th>Description</th>
<th>5’-Sequence-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 Promoter</td>
<td>CTGTAATACGACTCACTATATA</td>
</tr>
<tr>
<td>Forward primer/2nd PCR fwd primer</td>
<td>TTGAAGGCTCAGTATGTCCTATAGTGAGTCGTATTACAG</td>
</tr>
<tr>
<td>Pool reverse primer</td>
<td>GGTAAAGGCTGCGGATCTACTAGTGGAGTGGCTATTACAG</td>
</tr>
<tr>
<td>Selection reverse primer</td>
<td>GGTAAAGGCTGCGGATCTACTAGTGGAGTGGCTATTACAG</td>
</tr>
<tr>
<td>1st PCR forward primer</td>
<td>CTTCACACAGCTCTACCCATCC</td>
</tr>
<tr>
<td>Cloning forward primer</td>
<td>TAAATAAAATAACTGTAATACGACTCACTATAGGACATACTGAGC</td>
</tr>
<tr>
<td>Fwd primer for substrate sequence</td>
<td>CTCGTAATACGACTCACTATAGGACATACTGAGC</td>
</tr>
<tr>
<td>E. coli SS rRNA forward primer</td>
<td>CTGTAATACGACTCACTATAGGACATACTGAGC</td>
</tr>
<tr>
<td>E. coli SS rRNA reverse primer</td>
<td>TGCCTGCGAGCTCTACCTCTCCGATGGGAGACACCCCACACTACCATCAGGCGCTACGGCGTTTACACCTGTTTCCGG</td>
</tr>
<tr>
<td>1st PCR NGS N8 UMI forward primer</td>
<td>TCTGCGGAGCTCTACCTCTCCGATGGGAGACACCCCACACTACCATCAGGCGCTACGGCGTTTACACCTGTTTCCGG</td>
</tr>
<tr>
<td>1st PCR NGS reverse primer</td>
<td>GGTGTGCTGCGGAGCTCTACCTCTCCGATGGGAGACACCCCACACTACCATCAGGCGCTACGGCGTTTACACCTGTTTCCGG</td>
</tr>
<tr>
<td>Indexed forward primer i5_S502 (Round 12)</td>
<td>GAACGAGAAGAGAGAGCTATCAGGAGATGCTCAGGAGATCCTCGTGGGCTGCGG</td>
</tr>
<tr>
<td>Indexed forward primer i7_N703 (Round 7)</td>
<td>CAAGCAGAAGAGAGAGCTATCAGGAGATGCTCAGGAGATCCTCGTGGGCTGCGG</td>
</tr>
<tr>
<td>RT primer 16S 344-361 (A325)</td>
<td>CCCACTGCTGCTCCCTCCG</td>
</tr>
<tr>
<td>RT primer 16S 413-430 (A383)</td>
<td>TACACACGCCAGCTCCCT</td>
</tr>
<tr>
<td>RT primer 16S 687-704 (A649)</td>
<td>TCTACGCTTCTCCGCTG</td>
</tr>
<tr>
<td>RT primer 23S 322-339 (A272)</td>
<td>AGGGGCTGCTCTCCGCTG</td>
</tr>
<tr>
<td>RT primer 23S 653-672 (A637)</td>
<td>GGTCTATACCCGCAACT</td>
</tr>
<tr>
<td>RT primer 23S 1599-1618 (A1572)</td>
<td>TGTGTCGCTTTGGGCTGCGG</td>
</tr>
</tbody>
</table>

ssDNA templates for

<table>
<thead>
<tr>
<th>5’-Sequence-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent substrate sequence</td>
</tr>
<tr>
<td>TM</td>
</tr>
<tr>
<td>TV1</td>
</tr>
<tr>
<td>TV2</td>
</tr>
<tr>
<td>Parent-AAG</td>
</tr>
<tr>
<td>Parent -UAG</td>
</tr>
<tr>
<td>Parent -CAG</td>
</tr>
<tr>
<td>Parent -GAA</td>
</tr>
</tbody>
</table>
### 1.3. RNA oligonucleotides: transcripts and synthetic RNAs

#### Pool and substrate sequences (modification site shown in bold)

<table>
<thead>
<tr>
<th>Description</th>
<th>5'-Sequence-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection pool</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC-3'</td>
</tr>
<tr>
<td>Parent substrate sequence</td>
<td>GACUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>TM</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>TM-UGAGC</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>5'TV1-3'parent</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>TV1-GAGC</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>Parent-AAG</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>Parent-UAG</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>Parent-CAG</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>Parent-GAA</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>Parent-GAU</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>Parent-GAC</td>
<td>GGCAUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>3'-aminohexyl RNA substrate</td>
<td>ACAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
<tr>
<td>5'-Hexynyl RNA substrate</td>
<td>Hexyne-GACUAUACUGAGGGCCUUCAACCUCCUCCAUUUGAGAGGC</td>
</tr>
</tbody>
</table>
SUPPORTING INFORMATION

RNA w two labeling sites (DLRNA)

GGACAUACUGA GCCUCAAUAUGAGUGUGCGAGGUAUAGUG

E. coli 5S rRNA
(In vitro transcribed)

GGUGCCUGCCCGCGGGGAAUGGGCCGGGUGUCCACUCCUCAGGGGAAACUGAAAGUGA
AACGCGAGAGCCAUUGGUGAUGUGGUGGUGUCACUAGUGGAAACUGACGCGA

Ribozymes (Recognition arms underlined)

<table>
<thead>
<tr>
<th>Name</th>
<th>5'-Sequence-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>FJ1</td>
<td>GGUUGAAGGCCACCCUCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJ8</td>
<td>GGUUGAAGGCCACCCUCAUAAAACUGAAGAUCCUUUGGCAAGGGUCUA</td>
</tr>
<tr>
<td>FJC1</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_5'TV1-3'par</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TM</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TV1</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TV1-GAGC</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TV1-3'par</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TV2</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TV2</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_TM-UGAGC</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_5S-A99</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_23S-A272</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_23S-A362</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_23S-A1572</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_16S-A383</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC1_16S-A649</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC9</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC9_5'SS-A99</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC9_23S-A272</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC9_23S-A637</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FJC9_23S-A1572</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A325</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A412</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A325</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A412</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A383</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A649</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A325</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A412</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A383</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A649</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A325</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A412</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A383</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A649</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A325</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A412</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A383</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A649</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A325</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A412</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A383</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
<tr>
<td>FH14_16S-A649</td>
<td>GGUUGAAGGCCACCCUAAAGAACAGUGCAAGGUGCGUACUGGAUGUCC</td>
</tr>
</tbody>
</table>
2. Experimental Procedures

2.1. Synthesis of tenofovir diphosphate analogs

**Scheme S1.** Synthesis of N6-biotin-tenofovir diphosphate (biotin-Ten-DP).

**Scheme S2.** a) Synthesis of tenofovir diphosphate, N6-modified analogs N6-hexynyl-Ten-DP and N6-azidohexyl-Ten-DP. b) Synthesis of fluorescently labeled tenofovir diphosphates by CuAAC.
List of abbreviations used in the synthetic procedures:
ACN = Acetonitrile, CDI = 1,1'-Carbonyldiimidazole, Boc = tert-Butyloxycarbonyl, Bu3N = Tri-n-butylamine, Cy5 = sulfo-cyanine-5, DCM = Dichloromethane, DMF = Dimethylformamide, DP = diphasphate, FAM = Fluorescein, h = hours, min = minutes, NHS = N-Hydroxysuccinimide, o.n. = overnight, RP = reversed-phase, RT = room temperature, TEAA = Triethylammonium acetate, TEAB = Triethylaminonium bicarbonate, Ten = tenofovir, TLC = Thin-layer chromatography, TMSBr = Bromotrimethylsilane, Ts = Tosyl.

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

\[
\begin{array}{c}
\text{N} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\end{array}
+ \text{K}_2\text{CO}_3, \text{DMF} \\
\text{RT, 16 h}
\]

\[
\begin{array}{c}
\text{N} \\
\text{N} \\
\text{N} \\
\text{Cl} \\
\text{O} \\
\text{O} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\end{array}
\]

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 K₂CO₃ (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 CHCl₃ and washed with saturated NaHCO₃ (4 x 25 ml) and brine (20 ml). The organic layer was dried over Na₂SO₄, 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.

\[1^H\text{ NMR (}400\text{ MHz, CDCl}_3\text{): } \delta 8.71 (s, 1H), 8.11 (s, 1H), 5.15 (s, 2H), 2.37 (s, 3H).\]
\[1^3C\text{ NMR (}101\text{ MHz, CDCl}_3\text{): } \delta 198.6, 152.1, 151.8, 151.2, 145.6, 131.1, 52.3, 27.2.\]
\[\text{HRMS (ESI+): Exact mass calculated for C}_8\text{H}_8\text{ClN}_4\text{O} [\text{M} + \text{H}]^+, 211.03811. \text{Found 211.03798.}\]

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

\[
\begin{array}{c}
\text{N} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\text{O} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\end{array}
\text{NaBH}_4 \text{ MeOH/ACN (2:1)} \rightarrow \text{0 °C, 40 min 93%}
\]

\[
\begin{array}{c}
\text{N} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\text{OH} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\end{array}
\]

Compounds 1 (780 mg, 3.7 mmol) was dissolved in 40 ml MeOH/ACN (2:1). The solution was cooled to 0 °C and NaBH₄ (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 NH₄Cl and the reaction volume was concentrated to ~25 ml. The reaction mixture was extracted with CHCl₃ (4 x 50 ml). The combined organic layers were washed with saturated NH₄Cl (30 ml), brine (30 ml), dried over Na₂SO₄, 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%).

\[1^H\text{ NMR (}400\text{ MHz, CDCl}_3\text{): } \delta 8.70 (s, 1H), 8.21 (s, 1H), 4.43 (dd, J = 14.0, 2.7 Hz, 1H), 4.31 (dqd, J = 8.0, 6.3, 2.7 Hz, 1H), 4.13 (dd, J = 14.1, 8.0 Hz, 1H), 3.13 (s, 1H), 1.32 (d, J = 6.3 Hz, 3H).\]
\[1^3C\text{ NMR (}101\text{ MHz, CDCl}_3\text{): } \delta 151.8, 151.7, 150.6, 146.4, 131.0, 66.0, 51.5, 20.8.\]
\[\text{HRMS (ESI+): Exact mass calculated for C}_8\text{H}_9\text{ClN}_4\text{NaO} [\text{M} + \text{Na}]^+, 235.03571. \text{Found 235.03518.}\]

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

\[
\begin{array}{c}
\text{N} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\text{P} \\
\text{OEt} \\
\text{OEt} \\
\text{OEt} \\
\text{OEt} \\
\text{OEt} \\
\end{array}
\text{Mg(OtBu)₂, DMF 70 °C, 8 h 87%}
\]

\[
\begin{array}{c}
\text{N} \\
\text{Cl} \\
\text{N} \\
\text{N} \\
\text{N} \\
\text{O} \\
\text{O} \\
\text{P} \\
\text{OEt} \\
\text{OEt} \\
\text{OEt} \\
\end{array}
\]

\[\text{HRMS (ESI+): Exact mass calculated for C}_8\text{H}_7\text{ClN}_4\text{NaO} [\text{M} + \text{Na}]^+, 235.03571. \text{Found 235.03518.}\]
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\textsubscript{t}Bu\textsubscript{2})\textsubscript{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%).

\textit{1H NMR} (400 MHz, CDCl\textsubscript{3}): \(\delta\) 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).

\textit{13C NMR} (101 MHz, CDCl\textsubscript{3}): \(\delta\) 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), 62.1, 48.6, 16.5 – 16.4 (m).

\textit{31P NMR} (162 MHz, CDCl\textsubscript{3}): \(\delta\) 20.66.

\textit{HRMS (ESI+)}: Exact mass calculated for C\textsubscript{13}H\textsubscript{21}ClN\textsubscript{4}O\textsubscript{4}P [M + H]+, 363.09835. Found 363.09836.

tert-butyl (6-((9-(2-((diethoxyphosphoryl)methoxy)propyl)-9\textit{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%).

\textit{1H NMR} (400 MHz, MeOD): \(\delta\) 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).

\textit{13C NMR} (101 MHz, MeOD): \(\delta\) 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), 63.8, 62.1, 49.1, 41.5, 41.3, 30.9, 30.5, 28.8, 27.7, 27.6, 16.7 – 16.6 (m).

\textit{31P NMR} (162 MHz, MeOD): \(\delta\) 21.99.

\textit{HRMS (ESI+)}: Exact mass calculated for C\textsubscript{24}H\textsubscript{44}N\textsubscript{6}O\textsubscript{6}P [M + H]+, 543.30545. Found 543.30534.

(((1-(6-((6-aminohexyl)amino)-9\textit{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\textsubscript{2}O. 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.

\textit{1H NMR} (400 MHz, D\textsubscript{2}O): \(\delta\) 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).

\textit{13C NMR} (101 MHz, D\textsubscript{2}O): \(\delta\) 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.
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\textsubscript{3}N (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\textsubscript{2}O). The reaction mixture was stirred overnight at RT before the solvent was evaporated. After coevaporation with H\textsubscript{2}O, 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\textsubscript{2}O, B: ACN, Gradient: 5–50% B in 50 min, Flow rate: 7 ml/min

\[^{1}H\text{NMR (400 MHz, CD\textsubscript{3}CN:D\textsubscript{2}O (1:1)):} \delta 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), 3.75 – 3.64 (m, 6H), 3.22 (d, J = 12.9 Hz, 2H), 3.07 – 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\text{NMR (162 MHz, CD\textsubscript{3}CN:D\textsubscript{2}O (1:1)):} \delta 15.70.

HRMS (ESI\textsuperscript{-}): Exact mass calculated for C\textsubscript{25}H\textsubscript{40}N\textsubscript{8}O\textsubscript{6}PS \[M - H]\textsuperscript{-}, 611.25346. Found 611.25471.

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

Phosphonate activation using CDI, followed by reaction with pyrophosphate was performed in analogy to published procedures.\textsuperscript{[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\textsubscript{3}NH)\textsubscript{2}H\textsubscript{2}P\textsubscript{2}O\textsubscript{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\textsubscript{2}O. The resulting residue was dissolved in 10 ml H\textsubscript{2}O 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 H₂O. 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% NaClO₄ 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

1H NMR (400 MHz, D₂O): 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).

31P NMR (162 MHz, D₂O): 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 C₂₅H₄₂N₈O₁₂P₃S [M - H]-, 771.18612. Found 771.18521.

Diethyl ((1-(6-amino-9H-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 NH₃ 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%).

1H NMR (400 MHz, CDCl₃): 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), 3.58 (dd, J = 13.6, 9.7 Hz, 1H), 1.29 (t, J = 7.1 Hz, 3H), 1.26 – 1.22 (m, 6H).

13C NMR (101 MHz, CDCl₃): 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).

31P NMR (162 MHz, CDCl₃): δ 20.83.

The MeO-substituted compound 8b was isolated as a side product (eluted first from the column, colorless oil, 68 mg, 0.190 mmol, 35%).

1H NMR (400 MHz, CDCl₃): 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).

13C NMR (101 MHz, CDCl₃): δ 161.0, 152.1, 151.9, 143.4, 121.1, 76.3 (d, J = 11.9 Hz), 63.5, 62.4 (d, J = 6.7 Hz), 62.3 (d, J = 6.5 Hz), 61.9, 54.2, 48.3, 16.5 – 16.3 (m).

31P NMR (162 MHz, CDCl₃): δ 20.72.

(((1-(6-amino-9H-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 Et₂O. 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/H₂O/Et₂N,
SUPPORTING INFORMATION

6.1:1), followed by a second purification by silica gel column chromatography (PrOH/H2O/conc. NH3, 7:2:1) to afford compound 9 as a white foam (63 mg, 0.22 mmol, 75%).

1H NMR (400 MHz, D2O): 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).

13C NMR (101 MHz, D2O): δ 155.1, 151.9, 148.8, 143.2, 117.7, 75.4 (d, J = 11.3 Hz), 66.2 (d, J = 5.8 Hz, 1H), 47.4, 16.1.

31P NMR (162 MHz, D2O): δ 13.92.

HRMS (ESI-): Exact mass calculated for C9H13N5O4P [M - H]-, 286.07106. Found 286.07109.

[1-(6-amino-9H-purin-9-yl)propan-2-yl]oxymethyl-[hydroxy(phosphonoxy)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, (Bu3NH)2H2P2O7 (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 AKTPrime 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:

λmax (UV-vis absorption in H2O): 292 nm

1H NMR (400 MHz, D2O): 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 (d, J = 13.3, 9.1 Hz, 1H), 3.75 (dd, J = 13.3, 9.5 Hz, 1H), 1.19 (d, J = 6.3 Hz, 3H).

31P NMR (162 MHz, D2O): δ 8.07 (d, J = 26.5 Hz), -11.00 (d, J = 19.9 Hz), -23.45 (t, J = 22.6 Hz).


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%).

λmax (UV-vis absorption in H2O): 259 nm

1H NMR (400 MHz, D2O): 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 (d, 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).

31P NMR (162 MHz, D2O): 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 C9H15N5O10P3 [M - H]-, 446.00373. Found 446.00372.

Diethyl (((1-(6-(hex-5-ynylamino)-9H-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%).

\[
\begin{align*}
\text{1H NMR (400 MHz, CDCl}_3\text{): } & \delta 8.36 (s, 1H), 7.88 (s, 1H), 5.76 (br s, 1H), 4.33 (dd, \quad J = 14.4, 2.9 Hz, 1H), 4.14 – 4.00 (m, 5H), 3.95 – 3.88 (m, 1H), 3.81 (dd, \quad J = 13.6, 9.1 Hz, 1H), 3.68 (br s, 2H), 3.57 (dd, \quad J = 13.8, 9.4 Hz, 1H), 2.25 (td, \quad J = 7.0, 2.6 Hz, 2H), 1.95 (t, \quad J = 2.9 Hz, 1H), 1.85 – 1.77 (m, 2H), 1.69 – 1.62 (m, 2H), 1.30 (t, \quad J = 7.1 Hz, 3H), 1.27 – 1.22 (m, 6H). \\
\text{13C NMR (101 MHz, CDCl}_3\text{): } & \delta 154.9, 153.0, 149.0, 140.9, 119.3, 83.9, 76.4 (d, \quad J = 12.0 Hz), 68.7, 63.6, 62.4 (d, \quad J = 6.5 Hz), 62.3 (d, \quad J = 6.5 Hz), 61.9, 48.1, 40.0, 28.8, 25.6, 18.1, 16.5 (d, \quad J = 6.8 Hz), 16.4 (d, \quad J = 5.4 Hz). \\
\text{31P NMR (162 MHz, CDCl}_3\text{): } & \delta 20.84. \\
\text{HRMS (ESI+): } \text{Exact mass calculated for C}_{19}\text{H}_{31}\text{N}_5\text{O}_4\text{P} [M + H]^+, 424.21082. \text{ Found 424.21083.}
\end{align*}
\]

11 ((1-(6-(hex-5-yn-1-ylamino)-9H-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 Et2O 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 (iPrOH/H2O/conc. NH3, 7:2:1) to afforded compound 12 as a white foam (78 mg, 0.22 mmol, 90%).

\[
\begin{align*}
\text{1H NMR (400 MHz, D}_2\text{O): } & \delta 8.08 (s, 1H), 8.05 (s, 1H), 4.26 (dd, \quad J = 14.6, 3.4 Hz, 1H), 4.12 (dd, \quad J = 14.7, 6.6 Hz, 1H), 3.92 – 3.84 (m, 1H), 3.61 (dd, \quad J = 13.1, 9.3 Hz, 1H), 3.46 – 3.40 (m, 3H), 2.26 (t, \quad J = 2.6 Hz, 1H), 2.16 (td, \quad J = 7.0, 2.6 Hz, 2H), 1.72 – 1.65 (m, 2H), 1.56 – 1.49 (m, 2H), 1.09 (d, \quad J = 6.3 Hz, 3H). \\
\text{13C NMR (101 MHz, D}_2\text{O): } & \delta 153.1, 150.7, 147.4, 142.7, 117.9, 85.5, 75.8 (d, \quad J = 11.9 Hz), 69.3, 64.8 (d, \quad J = 157.8 Hz), 47.7, 40.4, 27.5, 25.0, 17.3, 16.0. \\
\text{31P NMR (162 MHz, D}_2\text{O): } & \delta 15.73. \\
\text{HRMS (ESI-): Exact mass calculated for C}_{15}\text{H}_{21}\text{N}_5\text{O}_4\text{P} [M - H]^-, 366.13366. \text{ Found 366.13366.}
\end{align*}
\]


\[
\begin{align*}
\text{i) CDI, DMF, Bu}_3\text{N 18 h, RT} \\
\text{ii) (Bu}_3\text{NH})_2\text{H}_2\text{P}_2\text{O}_7, DMF overnight, RT } 21\%
\end{align*}
\]

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 (Bu3NH)2H2P2O7 (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 H2O. The residue was dissolved in 10 ml H2O 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

1H NMR (400 MHz, D2O): 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), 3.74 (dd, J = 13.3, 9.4 Hz, 1H), 3.58 (br s, 2H), 2.29 (t, J = 2.6 Hz, 2H), 1.82 – 1.75 (m, 2H), 1.65 – 1.58 (m, 2H), 1.11 (d, J = 6.3 Hz, 3H).

31P NMR (162 MHz, D2O): δ 8.64 (d, J = 26.6 Hz), -10.92 (d, J = 19.9 Hz), -23.36 (dd, J = 26.5, 19.9 Hz).


Diethyl (((1-(6-((6-azidohexyl)amino)-9H-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%).

1H NMR (400 MHz, CDCl3): 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).

13C NMR (101 MHz, CDCl3): 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, 28.6, 26.3, 16.3 (d, J = 8.1 Hz), 16.2 (d, J = 5.3 Hz).

31P NMR (162 MHz, CDCl3): δ 20.84.


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

In 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 Et2O, 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.

1H NMR (400 MHz, D2O): 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).

13C NMR (101 MHz, D2O): 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.

31P NMR (162 MHz, D2O): δ 19.65.

1H NMR (400 MHz, D2O): 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).

31P NMR (162 MHz, D2O): 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 C15H26N8O10P3 [M - H]-, 571.09902. Found 571.09905.

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 CuSO4. In another 1.5 mL reaction tube, 5 μL CuSO4 (100 mM in H2O; 1 eq) was mixed with 15 μL of a freshly prepared solution of sodium ascorbate (200 mM in H2O; 6 eq), and added to the first 1.5 mL reaction tube, to give a final volume of 50 μL (final DMF/H2O 3:2). The reaction mixture was stirred overnight at RT. Then, 1.2 ml ice-cold 2% NaClO4 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 AKTAprime plus, 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% ACN, Gradient: 5% B for 2.5 min, 5-20% B in 5 min, 20-22.5% B in 5 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 C39H41N9O16P3 [M - H]-; 984.18896. Found 984.18897.

In a small 1.5 mL reaction tube 500 nmol compound 13 (10 μL of 50 mM in H2O; 1 eq) were combined with 10 μL Sulfo-Cy5 azide (100 mM in H2O; 2 eq) and 60 μL DMF. In another 1.5 mL reaction tube, 5 μL CuSO4 (100 mM in H2O; 1 eq) were mixed with 15 μL of a freshly prepared solution of sodium ascorbate (200 mM in H2O; 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/H2O 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% NaClO4 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


**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%).

**1H NMR (400 MHz, CDCl3):** 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).

**13C NMR (101 MHz, CDCl3):** 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 C₉₃H₇₃BrNO₂ [M + H]⁺, 310.04372. Found 310.04457.
2-(6-azidohexyl)isoindoline-1,3-dione (20)

\[
\begin{array}{c}
\text{N} \\
\text{O} \\
\text{O} \\
\text{Br} \\
\text{N} \\
\text{O} \\
\text{N3}
\end{array} \\
\xrightarrow{\text{DMF, } \text{RT, overnight}} \\
\begin{array}{c}
\text{N} \\
\text{O} \\
\text{O} \\
\text{N3}
\end{array}
\]

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 NaN₃ (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 Na₂SO₄, 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, CDCl₃): \( \delta \) 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, CDCl₃): \( \delta \) 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 C₁₄H₁₆N₄NaO₂ \([\text{M + Na]}^+\), 295.11655. Found 295.11656.

6-azidohexan-1-amine (21)

\[
\begin{array}{c}
\text{N} \\
\text{O} \\
\text{N3}
\end{array} \\
\xrightarrow{\text{N₂H₄.H₂O, EtOH, RT, overnight}} \\
\begin{array}{c}
\text{N} \\
\text{O} \\
\text{N3}
\end{array}
\]

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 H₂O (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 H₂O (50 ml) followed by brine (50 ml). The DCM solution was dried over Na₂SO₄, followed by solvent evaporation to yield compound 21 as yellowish liquid (1.54 g, 10.83 mmol, 74%).

\(^1\)H NMR (400 MHz, CDCl₃): \( \delta \) 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, CDCl₃): \( \delta \) 51.4, 42.1, 33.6, 28.8, 26.6, 26.5.

HRMS (ESI\(^+\)): Exact mass calculated for C₆H₁₅N₄ \([\text{M + H]}^+\), 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. 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 MgCl₂ and 300 µM of Biotin-Tenofovir-diphosphate. 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 submitting 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. 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 MgCl₂ 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 \( Y = Y_{\text{max}}(1-\exp(-k_{\text{obs}}t)) \).

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'-AGC-3'. Optimal modification efficiency was obtained with sequences containing 5'-RAGCy-3' as the modification site context 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'-fluorescein 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 MgCl₂ and 500 µM of the N⁶-biotin-tenofovir substrate. The sample was incubated at 37°C overnight, followed by quenching via addition of 25 µL of high-dye solution. The sample was then 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 MnCl₂ 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 MgCl₂ 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 MgCl₂ and 200 µM, N⁶-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 MgCl₂ 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 each of these samples were then resolved on a 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 to 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 MnCl₂, 10 mM MgCl₂ 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 MgCl2 and the ribozyme’s cognate substrate. For FH14 type ribozyme N6-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. ³²P-Labeling of the primers
Primers (100 pmol) were dissolved in 10 µL of PNK buffer A including 5 µCi of γ-³²P-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 millQ water, resulting in a sample with 2000-3500 IPS.

2.3.5.3. Primer extension protocol
5 pmol of the ³²P-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 ³²P-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  TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ24 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ32 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ35 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ8  TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC
FJ11 TTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA 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(y) log2 \(x/y\)**
- >2-1-0=FJ8  AACCAGCCTACCATCCTTGAAGGC ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGTATGTCC  38.8  33325.9  9.74
- >5-4-1-0=FJ1  AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC  25.7 10479.4  8.67
- >8-2-0-1=FJC9  AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGTATGTCC  46.2  924.1  4.32
- >1-1-0=FJC1  AACCAGCCTACCATCCTTGAAGGC GAAACGTGTCACATAAGAAAACGGTAAACTAGCAAGTTCC CAGTATGTCC  130185 502593  1.95
- >48-7-1-0 (a)  AACCAGCCTACCATCCTTGAAGGC T CAGTATGTCC AACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTATGTCC  474.2 1586.4  1.74
- >28-5-1-0 (b)  AACCAGCCTACCATCCTTGAAGGC CGAATGCCACCGAACCGTATAATTGCCGCCTCCAATTTC CAGTATGTCC  3096.4 2437.6 -0.35
- >3-3-0-1=FJC3  AACCAGCCTACCATCCTTGAAGGC CGAATGCCACCGAACCGTATAATTGCCGCCTCCAATTTC CAGTATGTCC  39964.6 30094.3 -0.41
- >56-8-1-0  AACCAGCCTACCATCCTTGAAGGC ACTGGTGTTACAAGTAAACGCAACCTTAATTTTACGATCC CAGTATGTCC  4264.5 1258.8 -1.76
- >32-6-1-0 (c)  AACCAGCCTACCATCCTTGAAGGC ATCGGTGTAACATAAACGAAACCTTAGGTCCCTTGGCCCC CAGTATGTCC  34139.0 2206.6 -3.95
- >126-10-1-0  AACCAGCCTACCATCCTTGAAGGC ATAGTTTCACAATAGCCACTTTAAGCTTATCCAAAGCTCA CAGTATGTCC 11488.7  577.1 -4.32

\(\text{(a)}\) = extended substrate to complement sequence  
\(\text{(b)}\) reveals a fully complementary binding arm after 14 nt in the center of N40 core  
\(\text{(c)}\) related to FJ8
Partial alignment of N40 core:

**FJC3**

CGAA---TGCCACCG-AACCGTATAA-TTGCCGCCTCCAATTTC

**FJC9**

ACCGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC

**FJ8**

(407 sequences)

TYGAAAGCC AGCGTGTAAAAACGAAACCTTAGGACATTACCTTCCCC CAGTATGTCC

(10 sequences)

YTGAAGGC CACCCCTCATAAAAACTGAGATCTGTTGCGAAGGGGCGTCA CAGTATGTCC

SUPPORTING INFORMATION

**FJ1**

(158 sequences)

TTGAAGGC CACCCCTCATAAAAACTGAGATCTGTTGCGAAGGGGCGTCA CAGTATGTCC

**FJC9**

(10 sequences)

YTGAAGGC CACCCCTCATAAAAACTGAGATCTGTTGCGAAGGGGCGTCA CAGTATGTCC

Partial alignment of N40 core:

**FJC1**

GAAACGTC-YTACATAT++GAAAA-CGUTAATCCATAGGTC

**FJC3**

CGAAGAGACTAAATCCCTTAAGAACAGCTTAAAGTACCTTCCCC CAGTATGTCC

**FJC9**

ACCGGATATGTTGCACTACACTTTAGCGAATTGGGCATCC
3.2 Supporting Figures

**Figure S1.** Biotin-Tenofovir diphosphate is not a substrate for FH14. (3’-fluorescein-labeled parent substrate RNA; 200 μM N6-Biotin-ATP, 800 μM N6-Biotin-TenDP, with 40 mM MgCl2, in selection buffer. Overnight incubation at 37°C.

**Figure S2.** In vitro selection scheme, showing RNA substrate sequence in red, random region (N40) 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.
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. The minimum 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 MgCl₂, 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 MgCl₂, 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 MgCl₂, 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, 200 µM Biotin-ATP, with 40 mM MgCl2, in selection buffer, 37°C, timepoints are: 0, 1, 2, 4, 6, 22 h). b) Mg2+-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 Mg2+ concentration. At 5 mM Mg2+, only 7% labeled RNA is observed after 22h, but already at 10 mM Mg2+, the yield is above 60%.
Figure S12. Secondary structures 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 three 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 indicated 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'-32P-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).
Figure S16: HPLC chromatograms of compound 13, FAM azide and product 17 at 260 nm (above) and at 496 nm (below), respectively.
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 3 $^{31}$P NMR

Compound 4 $^1$H NMR
Compound 5 $^1$H NMR

Compound 5 $^{13}$C NMR
Compound 5 $^{31}$P NMR

Compound 6 $^1$H NMR
Compound 10 $^{31}$P NMR
Compound 12 $^{13}$C NMR

Compound 12 $^{31}$P NMR
Supporting Information

Compound 13 $^1$H NMR

Compound 13 $^{31}$P NMR

4Et$_3$NH$^+$
SUPPORTING INFORMATION

Compound 14 $^{31}$P NMR:

Compound 15 $^1$H NMR:
Compound 16 $^1$H NMR

Compound 16 $^{31}$P NMR
Compound 21 $^1$H NMR

Compound 21 $^{13}$C NMR
References