BIROn - Birkbeck Institutional Research Online

Liu, C. and Cozens, C. and Jaziri, F. and Rozenski, J. and Marechal, Amandine and Dumbre, S. and Pezo, V. and Marliere, P. and Pinheiro, Vitor B. and Groaz, E. and Herdewijn, P. (2018) Phosphonomethyl Oligonucleotides as Backbone-Modified Artificial Genetic Polymers. *Journal of the American Chemical Society* 140 (21), pp. 6690-6699. ISSN 0002-7863.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/22678/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Supporting Information

Phosphonomethyl Oligonucleotides as Backbone Modified Artificial Genetic Polymers

Chao Liu,[†] Christopher Cozens,[§] Faten Jaziri,[‡] Jef Rozenski,[†] Amandine Marechal,[§] Shrinivas Dumbre,[†] Valérie Pezo,[‡] Philippe Marlière,[‡] Vitor B. Pinheiro,^{§,||,*} Elisabetta Groaz,[†] and Piet Herdewijn^{†,‡,*}

[†]Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Herestraat 49, 3000 Leuven, Belgium

[§]University College London, Gower Street, London WC1E 6BT, UK

[‡]SSB, Genopole, CNRS, UEVE, Université Paris-Saclay, 5 rue Henri Desbruyères, 91030 Evry Cedex, France

^{||}Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet Street, WC1E 7HX, UK

Chemical synthesis of threosyl nucleoside *H*-phosphinates

1' α -(*N*⁶-Benzoyladenine-9-yl)-2'-*O*-(4-methoxytrityl)-3'-*O*-diisopropylphosphonomethyl-L-threose

(2a). To a solution of **1a** (223 mg, 0.43 mmol) in 3 mL of anhydrous THF were sequentially added dry pyridine (0.22 mL, 2.79 mmol), AgNO₃ (109 mg, 0.64 mmol), and MMTTrCl (199 mg, 0.64 mmol). The reaction mixture was stirred at room temperature overnight. It was then quenched with saturated aq. NaHCO₃ and extracted with ethyl acetate. The organic layer was washed with brine, dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The resulting crude residue was purified by column chromatography (100:1 to 100:2, CH₂Cl₂/Et₃N) to afford **2a** (301 mg, 89% yield) as a white foam. ¹H NMR (300 MHz, CDCl₃) δ 9.05 (s, 1H, NH), 8.89 (s, 1H, H-2), 8.33 (s, 1H, H-8), 8.04–8.00 (m, 2H, ArH), 7.65–7.16 (m, 15H, ArH), 6.78–6.70 (m, 2H, ArH), 6.49 (s, 1H, H-1'), 4.76–4.63 (m, 2H, CH(CH₃)₂), 4.43 (s, 1H, H-2'), 4.28 (d, J = 10.4 Hz, 1H, H-4a'), 4.16 (dd, J = 10.4, 2.9 Hz, 1H, H-4b'), 3.72 (s, 3H, OCH₃), 3.33–3.21 (m, 3H, PCH₂ and H-3'), 1.34–1.26 (m, 12H, CH(CH₃)₂); ¹³C NMR (75 MHz, CDCl₃) δ 164.4 (CO), 159.1 (Ar-C), 152.6 (C-2), 151.6 (C-4), 149.3 (C-6), 143.9, 143.8 (Ar-C), 142.7 (C-8), 134.9, 134.1, 132.6, 130.6, 128.9, 128.5, 128.1, 127.8, 127.5 (Ar-C), 122.3 (C-5), 113.5 (Ar-C), 89.5 (C-1'), 89.0 (Ph₃C), 85.0 (d, ³J_{P,C} = 11.9 Hz, C-3'), 82.3 (C-2'), 72.4 (C-4'), 71.3 (CH(CH₃)₂), 64.4 (d, ¹J_{P,C} = 169.7 Hz, PCH₂), 55.2 (OCH₃), 24.0 (CH(CH₃)₂); ³¹P NMR (121 MHz, CDCl₃): δ 18.6; HRMS: [M+H]⁺ calcd for C₄₃H₄₇N₅O₈P, 792.3157; found, 792.3159.

1' α -(Thymin-1-yl)-2'-*O*-(4-methoxytrityl)-3'-*O*-diisopropylphosphonomethyl-L-threose **(2b).**

Compound **2b** was obtained as a white foam (0.64 g, 90%) as described for **2a** starting from **1b** (0.43 g, 1.05 mmol). ¹H NMR (500 MHz, CDCl₃): δ 7.45–7.43 (m, 4H, ArH), 7.34 (s, 1H, H-6), 7.35–7.22 (m, 8H, ArH), 6.83–6.84 (m, 2H, Ar-H), 6.42 (d, J = 1.8 Hz, H-1'), 4.65–4.72 (m, 2H, CH(CH₃)₂), 4.07–4.10 (m, 2H, H-4a' and H-2'), 3.86 (dd, J = 10.6, 2.6 Hz, 1H, H-4b'), 3.78 (s, 3H, OCH₃), 3.07–3.19 (m, 2H, PCH₂), 2.81 (d, J = 2.5 Hz, 1H, H-3'), 1.89 (d, J = 1.2 Hz, 3H, Thy-Me), 1.24–1.31 (m, 12H, CH(CH₃)₂); ¹³C NMR (125 MHz, CDCl₃): δ 163.6 (C-4), 159.0 (Ar-C), 150.3 (C-2), 143.93, 143.73 (Ar-C), 137.1 (C-6), 135.0, 130.5, 128.5, 128.1, 128.0, 127.4, 113.4 (Ar-C), 111.4 (C-5), 90.6 (C-1'), 88.6 (Ph₃C), 84.5 (d, ³J_{P,C} = 14.6 Hz, C-3'), 82.3 (C-2'), 71.1 (CH(CH₃)₂), 70.5 (C-4'), 63.9 (d, ¹J_{P,C} = 170.8 Hz, PCH₂), 55.2 (OCH₃), 24.0 (CH(CH₃)₂), 12.4 (Thy-Me); ³¹P NMR (121 MHz, CDCl₃): δ 18.1; HRMS: [M+Na]⁺ calcd for C₃₆H₄₃N₂O₉PNa, 701.2598; found, 701.2596.

1' α -(Hypoxanthin-9-yl)-2'-*O*-(4-methoxytrityl)-3'-*O*-diisopropylphosphonomethyl-L-threose **(2c).**

Compound **2c** was obtained as a white foam (1.40 g, 47%) as described for **2a**, starting from **1c** (1.81 g, 4.35 mmol). ¹H NMR (600 MHz, CDCl₃) δ 8.16 (s, 1H, H-2), 8.05 (s, 1H, H-8), 7.45 (t, J = 7.8 Hz, 4H, ArH), 7.35–7.15 (m, 8H, ArH), 6.73 (d, J = 8.8 Hz, 2H, ArH), 6.23 (s, 1H, H-1'), 4.77–4.63 (m, 2H, CH(CH₃)₂), 4.43 (s, 1H, H-2'), 4.26 (d, J = 10.5 Hz, 1H, H-4a'), 4.13 (dd, J = 10.5, 2.9 Hz, 1H, H-4b'),

3.73 (s, 3H, OCH₃), 3.39–3.26 (m, 3H, PCH₂ and H-3'), 1.37–1.27 (m, 12H, CH(CH₃)₂); ¹³C NMR (150 MHz, CDCl₃) δ 159.1 (Ar-C), 159.0 (C-6), 148.8 (C-4), 144.9 (C-2), 143.8, 143.7 (Ar-C), 139.5 (C-8), 134.7, 130.5, 128.4, 128.3, 128.09, 128.06, 127.42, 127.39, 123.9 (Ar-C), 89.2 (C-1'), 88.8 (Ph₃C), 85.0 (d, ³J_{PC} = 12.5 Hz, C-3'), 82.4 (C-2'), 72.0 (C-4'), 71.4 (CH(CH₃)₂), 64.3 (d, ¹J_{PC} = 169.4 Hz, PCH₂), 55.2 (OCH₃), 24.0 (CH(CH₃)₂); ³¹P NMR (121 MHz, CDCl₃): δ 17.8; HRMS: [M+H]⁺ calcd for C₃₆H₄₂N₄O₈P, 689.2735; found, 689.2753.

1'α-(Adenin-9-yl)-2'-O-(4-methoxytrityl)-3'-O-diisopropylphosphonomethyl-L-threose (2d). A solution of **2a** (150 mg, 0.19 mmol) in 2 N NH₃ in MeOH (12 mL) was stirred at room temperature overnight. After removal of all the volatiles, the crude residue was purified by column chromatography (200:3:1, CH₂Cl₂/MeOH/Et₃N) to afford **2d** (114 mg, 88%) as a white foam. ¹H NMR (500 MHz, CDCl₃) δ 8.40 (s, 1H, H-2), 8.10 (s, 1H, H-8), 7.46–7.44 (m, 4H, ArH), 7.31 (d, *J* = 8.8 Hz, 2H, ArH), 7.25–7.15 (m, 6H, ArH), 6.70 (d, *J* = 8.8 Hz, 2H, ArH), 6.39 (s, 1H, H-1'), 5.86 (br s, 2H, NH₂), 4.75–4.66 (m, 2H, CH(CH₃)₂), 4.40 (s, 1H, H-2'), 4.23 (d, *J* = 10.4 Hz, 1H, H-4a'), 4.11 (dd, *J* = 10.4, 2.7 Hz, 1H, H-4b'), 3.71 (s, 3H, OCH₃), 3.26 (d, *J* = 9.6 Hz, 2H, PCH₂), 3.19 (d, *J* = 2.3 Hz, 1H, H-3'), 1.34–1.27 (m, 12H, CH(CH₃)₂); ¹³C NMR (125 MHz, CDCl₃) δ 158.9 (Ar-C), 155.4 (C-6), 152.9 (C-2), 149.8 (C-4), 143.9, 143.8 (Ar-C), 140.1 (C-8), 134.8, 130.6, 128.5, 128.4, 128.07, 128.05, 127.4, 127.3 (Ar-C), 119.0 (C-5), 113.4 (Ar-C), 89.2 (C-1'), 88.8 (Ph₃C), 84.9 (d, ³J_{PC} = 13.2 Hz, C-3'), 82.3 (C-2'), 71.9 (C-4'), 71.3, 71.2 (CH(CH₃)₂), 64.2 (d, ¹J_{PC} = 170.0 Hz, PCH₂), 55.1 (OCH₃), 24.0 (CH(CH₃)₂); ³¹P NMR (121 MHz, CDCl₃): δ 18.0; HRMS: [M+H]⁺ calcd for C₃₆H₄₃N₅O₇P, 688.2894; found, 688.2908.

[1' α-(Adenin-9-yl)-2' -O-(4-methoxytrityl)-L-threose-3' -O-yl]methyl-(H)-phosphinate (3a). To a 1 M solution of LiAlH₄ (0.58 mL, 0.58 mmol) in anhydrous THF (3 mL) was added dropwise TMSCl (0.074 mL, 0.58 mmol) at -78 °C. The reaction mixture was allowed to warm to room temperature and stirred for 2 h. After cooling again to -78 °C, a solution of **2d** (100 mg, 0.145 mmol) in anhydrous THF (1.5 mL) was added dropwise. After stirring for 10 min, the reaction was quenched with 1 mL of saturated aq. NaHCO₃ and then anhydrous MgSO₄ was added. The slurry was filtered through a pad of Celite and 1 mL of H₂O was added to the filtrate followed by NaHCO₃ (12.2 mg, 0.145 mmol) and 35% aqueous H₂O₂ (0.04 mL, 0.44 mmol). The mixture was stirred at room temperature overnight and evaporated under reduced pressure. The crude residue was purified by column chromatography (100:10:1 to 50:10:1, CH₂Cl₂/MeOH/Et₃N) to afford **3a** (88 mg, Et₃N salt, 74% yield) as a white foam. ¹H NMR (600 MHz, CDCl₃) δ 8.62 (s, 1H, H-8), 8.30 (s, 1H, H-2), 7.73 (br s, 2H, NH₂), 7.14 (d, *J* = 505.7 Hz, 1H, PH), 7.45 (t, *J* = 6.5 Hz, 4H, ArH), 7.31 (d, *J* = 8.7 Hz, 2H, ArH), 7.23–7.13 (m, 6H, ArH), 6.70 (d, *J* = 8.7 Hz, 2H, ArH), 6.42 (s, 1H, H-1'), 4.38 (s, 1H, H-2'), 4.22 (d, *J* = 10.4 Hz, 1H, H-4a'), 4.04 (d, *J* = 10.4 Hz, 1H, H-4b'), 3.70 (s, 1H, OCH₃), 3.25–3.03 (m, 2H, PCH₂), 2.93 (s, 1H, H-3'); ¹³C NMR (150 MHz, CDCl₃) δ 158.8 (Ar-C), 156.2 (C-6), 152.9 (C-2), 149.2 (C-4), 143.9, 143.9 (Ar-C), 140.7 (C-8), 135.0,

130.4, 128.5, 128.4, 127.97, 127.95, 127.3, 127.2 (Ar-C), 118.5 (C-5), 113.3 (Ar-C), 88.9 (C-1'), 88.7 (Ph₃C), 84.0 (d, ³J_{PC} = 13.6 Hz, C-3'), 82.8 (C-2'), 71.7 (C-4'), 69.6 (d, ¹J_{PC} = 106.7 Hz, PCH₂), 55.2 (OCH₃); ³¹P NMR (121 MHz, CDCl₃): δ 15.7; HRMS: [M+H]⁺ calcd for C₃₀H₃₁N₅O₆P, 588.2006; found, 588.2006.

[1'-α-(Thymin-1-yl)-2'-O-(4-methoxytrityl)-L-threose-3'-O-yl]methyl-(H)-phosphinate (3b).

Compound **3b** was obtained as a white foam (120 mg, 71%) as described for **3a**, starting from **2b** (147 mg, 0.22 mmol). ¹H NMR (600 MHz, CDCl₃): δ 7.59 (d, *J* = 1.3 Hz, 1H, H-6), 7.46–7.21 (m, 12H, ArH), 7.04 (d, *J* = 497.0 Hz, 1H, PH), 6.84–6.81 (m, 2H, ArH), 6.37 (d, *J* = 1.5 Hz, H-1'), 4.16–4.14 (m, 2H, H-4a' and H-2'), 3.83 (dd, *J* = 10.3, 2.7 Hz, 1H, H-4b'), 3.78 (s, 3H, OCH₃), 3.09–2.91 (m, 2H, PCH₂), 2.73 (d, *J* = 2.6 Hz, 1H, H-3'), 1.89 (d, *J* = 1.2 Hz, 3H, Thy-Me); ¹³C NMR (150 MHz, CDCl₃) δ 163.8 (C-4), 158.9 (Ar-C), 150.2 (C-2), 144.0, 143.9 (Ar-C), 138.1 (C-6), 135.2, 130.5, 128.63, 128.59, 128.03, 128.01, 127.3, 113.3 (Ar-C), 110.9 (C-5), 90.8 (C-1'), 88.5 (Ph₃C), 83.9 (d, ³J_{PC} = 12.2 Hz, C-3'), 82.4 (C-2'), 70.8 (C-4'), 69.6 (d, ¹J_{PC} = 107.3 Hz, PCH₂), 55.3 (OCH₃), 12.4 (Thy-Me); ³¹P NMR (121 MHz, CDCl₃): δ 16.6; HRMS: [M-H]⁻ calcd for C₃₀H₃₀N₂O₈P, 577.1745; found, 577.1749.

[1'-α-(Hypoxanthin-9-yl)-2'-O-(4-methoxytrityl)-L-threose-3'-O-yl]methyl-(H)-phosphinate (3c).

Compound **3c** was obtained as a white foam (0.90 g, 65%) as described for **3a**, starting from **2c** (1.37 g, 1.99 mmol). ¹H NMR (600 MHz, CDCl₃) δ 8.34 (s, 1H, H-8), 8.12 (s, 1H, H-2), 7.10 (d, *J* = 504.5 Hz, 1H, PH), 7.47–7.41 (m, 4H, ArH), 7.32–7.13 (m, 8H, ArH), 6.74–6.68 (m, 2H, ArH), 6.19 (s, 1H, H-1'), 4.47 (s, 1H, H-2'), 4.27 (d, *J* = 10.2 Hz, 1H, H-4a'), 4.09 (dd, *J* = 10.3, 3.0 Hz, 1H, H-4b'), 3.71 (s, 3H, OCH₃), 3.29–3.12 (m, 2H, PCH₂), 3.22 (d, *J* = 2.8 Hz, 1H, H-3'); ¹³C NMR (150 MHz, CDCl₃) δ 158.9 (Ar-C), 158.4 (C-6), 148.6 (C-4), 145.0 (C-2), 143.85, 143.80 (Ar-C), 140.0 (C-8), 134.9, 130.4, 128.39, 128.37, 128.02, 127.99, 127.34, 127.33, 123.8 (C-5), 113.3 (Ar-C), 89.3 (C-1'), 88.6 (Ph₃C), 84.3 (d, ³J_{PC} = 11.9 Hz, C-3'), 82.2 (C-2'), 72.1 (C-4'), 69.7 (d, ¹J_{PC} = 107.2 Hz, PCH₂), 55.2 (OCH₃); ³¹P NMR (121 MHz, CDCl₃): δ 15.2; HRMS: [M+H]⁺ calcd for C₃₀H₃₀N₄O₇P, 589.1846; found, 589.1858.

[1'-α-(Hypoxanthin-9-yl)-2'-O-(4-methoxytrityl)-L-threose-3'-O-yl]methyl-(H)-phosphinate (3d).

To a solution of **3a** (0.56 g, 0.81 mmol) in anhydrous MeOH (16 mL) were added Et₃N (0.68 mL, 4.88 mmol) and *N,N*-dimethylformamide dimethyl acetal (0.54 mL, 4.07 mmol). The reaction mixture was stirred at room temperature under an argon atmosphere overnight. The solvent was evaporated in vacuo and the resulting crude residue was purified by column chromatography (100:10:1 to 100:10:2, CH₂Cl₂/MeOH/Et₃N) to afford **3d** (0.42 g, Et₃N salt, 69% yield) as a white foam. ¹H NMR (600 MHz, CDCl₃) δ 9.00 (s, 1H, CHN(CH₃)₂), 8.61 (s, 1H, H-2), 8.44 (s, 1H, H-8), 7.08 (d, *J* = 503.4 Hz, 1H, PH), 7.46–7.44 (m, 4H, ArH), 7.32–7.30 (m, 2H, ArH), 7.23–7.12 (m, 6H, ArH), 6.71–6.68 (m, 2H, ArH), 6.41 (d, *J* = 0.7 Hz, 1H, H-1'), 4.42 (s, 1H, H-2'), 4.27 (d, *J* = 10.3 Hz, 1H, H-4a'), 4.08 (dd, *J* = 10.3, 3.0 Hz, 1H, H-4b'), 3.69 (s, 1H, OCH₃), 3.22 (s, 3H, NCH₃), 3.19 (s, 3H, NCH₃), 3.22–3.07 (m, 2H, PCH₂), 3.11

(d, $J = 2.7$ Hz, 1H, H-3'); ^{13}C NMR (150 MHz, CDCl_3) δ 159.2 (C-6), 158.8 (Ar-C), 158.4 ($\text{CHN}(\text{CH}_3)_2$), 152.5 (C-2), 151.6 (C-4), 143.90, 143.86 (Ar-C), 142.0 (C-8), 135.0, 130.5, 128.48, 128.46, 127.99, 127.96, 127.30, 127.25 (Ar-C), 125.1 (C-5), 113.3 (Ar-C), 89.1 (C-1'), 88.6 (Ph_3C), 84.2 (d, $^3J_{\text{PC}} = 11.9$ Hz, C-3'), 82.5 (C-2'), 71.9 (C-4'), 69.8 (d, $^1J_{\text{PC}} = 106.7$ Hz, PCH_2), 55.2 (OCH_3), 41.2 (NCH_3), 34.9 (NCH_3); ^{31}P NMR (121 MHz, CDCl_3): δ 15.6; HRMS: $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{33}\text{H}_{36}\text{N}_6\text{O}_6\text{P}$, 643.2428; found, 643.2442.

1' α -(*N*⁶-Benzoyladenine-9-yl)-3'-*O*-*tert*-butyldimethylsilyl-L-threose (5). To a solution of **4**¹ (5.70 g, 10.2 mmol) in 200 mL of THF/MeOH/H₂O (5:4:1) at 0 °C was added aq. NaOH (5.09 mL, 10.2 mmol). After stirring at 0 °C for 0.5 h, the reaction mixture was neutralized with acetic acid and the organic solvent was removed under reduced pressure. The remaining residue was partitioned between water and ethyl acetate. The organic layer was washed with brine, dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The crude residue was purified by column chromatography (1:1 to 1:3, $\text{CH}_2\text{Cl}_2/\text{EtOAc}$) to afford **5** (3.80 g, 82% yield) as a white foam. ^1H NMR (500 MHz, CDCl_3) δ 8.70 (s, 1H, H-2), 8.39 (s, 1H, H-8), 7.98 (d, $J = 7.6$ Hz, 2H, ArH), 7.54 (t, $J = 7.1$ Hz, 1H, ArH), 7.45 (t, $J = 7.7$ Hz, 2H, ArH), 6.22 (s, 1H, H-1'), 4.51 (s, 1H, H-2'), 4.36–4.32 (m, 2H, H-3' and H-4a'), 4.20–4.17 (m, 1H, H-4b'), 0.72 (s, 9H, $(\text{CH}_3)_3\text{C}$), 0.01 (s, 3H, SiCH_3), 0.12 (s, 3H, SiCH_3); ^{13}C NMR (125 MHz, CDCl_3) δ 164.9 (CO), 152.1 (C-2), 150.7 (C-4), 149.2 (C-6), 142.0 (C-8), 133.5, 132.7, 128.6, 127.8 (Ar-C), 122.4 (C-5), 91.5 (C-1'), 81.3 (C-2'), 76.8 (C-3'), 76.7 (C-4'), 25.5 ($\text{C}(\text{CH}_3)_3$), 17.8 ($\text{C}(\text{CH}_3)_3$), -5.15, -5.24 (SiCH_3); HRMS: $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{22}\text{H}_{30}\text{N}_5\text{O}_4\text{Si}$, 456.2061; found, 456.2057.

1' α -(*N*⁶-Benzoyladenine-9-yl)-2'-*O*-(4-methoxytrityl)-3'-*O*-*tert*-butyldimethylsilyl-L-threose (6). Compound **6** was obtained as a white foam (4.00 g, 50%) as described for **2a**, starting from **5** (5.00 g, 10.97 mmol), along with the recovery of 2.12 g (42%) of starting material. ^1H NMR (300 MHz, CDCl_3) δ 8.76 (s, 1H, H-2), 8.24 (s, 1H, H-8), 8.06–8.03 (m, 2H, ArH), 7.59–7.11 (m, 15H, ArH), 6.70–6.67 (m, 2H, ArH), 6.01 (s, 1H, H-1'), 4.57 (s, 1H, H-2'), 4.36–4.06 (m, 2H, H-3' and H-4'), 3.67 (s, 3H, OMe), 0.86 (s, 9H, $(\text{CH}_3)_3\text{C}$), 0.06 (s, 3H, SiCH_3), 0.01 (s, 3H, SiCH_3); ^{13}C NMR (75 MHz, CDCl_3) δ 164.7 (CO), 158.8 (Ar-C), 152.1 (C-2), 151.6 (C-4), 149.2 (C-6), 144.0, 143.9 (Ar-C), 142.0 (C-8), 134.6, 133.7, 132.4, 130.5, 128.6, 128.3, 128.1, 127.93, 127.86, 127.8, 127.2, 127.1 (Ar-C), 122.4 (C-5), 113.2 (Ar-C), 88.8 (C-1'), 88.5 (Ph_3C), 83.9 (C-2'), 76.8 (C-3'), 75.6 (C-4'), 54.9 (OMe), 25.6 ($\text{C}(\text{CH}_3)_3$), 17.7 ($\text{C}(\text{CH}_3)_3$), -4.7, -5.0 (SiCH_3); HRMS: $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{42}\text{H}_{46}\text{N}_5\text{O}_5\text{Si}$, 728.3263; found, 728.3277.

1' α -(*N*⁶-Benzoyladenine-9-yl)-2'-*O*-(4-methoxytrityl)-L-threose (7). To a solution of **6** (3.80 g, 5.22 mmol) in anhydrous THF (52 mL) was added triethylamine trihydrofluoride (1.70 mL, 10.44 mmol). The reaction mixture was stirred at room temperature for 48 h. After removal of all the volatiles in vacuo, the resulting residue was partitioned between ethyl acetate and water, the organic layer was washed with brine, dried over Na_2SO_4 , filtered, and concentrated under reduced pressure. The residue was purified by

column chromatography (4:1 to 1:2, CH₂Cl₂/EtOAc) to afford **7** (2.65 g, 83%) as a white solid. ¹H NMR (300 MHz, CDCl₃) δ 8.63 (s, 1H, H-2), 7.95–7.92 (m, 2H, ArH), 7.57 (s, 1H, H-8), 7.54–7.08 (m, 15H, ArH), 6.73–6.66 (m, 2H, ArH), 5.65 (s, 1H, H-1'), 4.71 (s, 1H, H-2'), 4.26–4.04 (m, 3H, H-3' and H-4'), 3.69 (s, 3H, OMe). ¹³C NMR (75 MHz, CDCl₃) δ 164.7 (CO), 158.8 (Ar-C), 151.6 (C-2), 150.0 (C-4), 149.7 (C-6), 143.9, 143.8 (Ar-C), 143.3 (C-8), 134.9, 133.3, 132.6, 130.2, 128.5, 128.2, 127.9, 127.8, 127.3 (Ar-C), 123.0 (C-5), 113.3 (Ar-C), 91.6 (C-1'), 88.4 (Ph₃C), 83.8 (C-2'), 76.7 (C-3'), 75.6 (C-4'), 55.0 (OMe). HRMS: [M+H]⁺ calcd for C₃₆H₃₂N₅O₅, 614.2398; found, 614.2410.

[1'α-(N⁶-Benzoyladenine-9-yl)-2'-O-(4-methoxytrityl)-L-threose-3'-O-yl]methyl-(H)-phosphinate (3e).

To a solution of **7** (2.00 g, 3.26 mmol) and tosyloxymethyl-(H)-phosphinate **8**² (1.06 g, 3.91 mmol) in anhydrous DMF (16 mL) was added NaH (60% suspension in mineral oil, 0.29 g, 7.17 mmol) at -10 °C. The reaction mixture was stirred at room temperature overnight. It was then neutralized with acetic acid at -10 °C and concentrated under reduced pressure. The oily residue was partitioned between CH₂Cl₂ and water, the water layer was further extracted with CH₂Cl₂, and the combined organic layers were dried over Na₂SO₄, filtered, and concentrated in vacuo. The residue was purified by column chromatography (8:1:1:1 to 4:1:1:1, EtOAc/acetone/EtOH/H₂O) on a silica gel column (pretreated with 1% Et₃N) to afford **3e** (1.90 g, Et₃N salt, 74% yield) as a white foam. ¹H NMR (300 MHz, DMSO-*d*₆) δ 11.24 (s, 1H, NH), 8.76 (s, 1H, H-2), 8.58 (s, 1H, H-8), 8.05 (d, *J* = 7.5 Hz, 2H, ArH), 7.65–7.11 (m, 15H, ArH), 6.86 (d, *J* = 486.1 Hz, 1H, PH), 6.69 (d, *J* = 8.8 Hz, ArH), 6.09 (s, 1H, H-1'), 4.39 (s, 1H, H-2'), 4.23–4.04 (m, 2H, H-4'), 3.60 (s, 3H, OMe), 3.53–3.40 (1H, H-3', overlapped with H₂O), 3.22–2.92 (m, 2H, PCH₂); ¹³C NMR (75 MHz, DMSO-*d*₆) δ 165.6 (CO), 158.4 (Ar-C), 151.7 (C-4), 151.4 (C-2), 150.0 (C-6), 144.0, 143.8 (Ar-C), 143.1 (C-8), 134.3, 133.4, 132.2, 130.2, 128.34, 128.28, 128.0, 127.9, 127.84, 127.81, 127.1, 127.0 (Ar-C), 124.6 (C-5), 113.2 (Ar-C), 88.4 (C-1'), 88.1 (Ph₃C), 83.6 (d, ³*J*_{PC} = 11.4 Hz, C-3'), 81.7 (C-2'), 72.0 (C-4'), 69.7 (d, ¹*J*_{PC} = 108.1 Hz, PCH₂), 54.9 (OMe); ³¹P NMR (121 MHz, DMSO-*d*₆): δ 11.5; HRMS: [M-H]⁻ calcd for C₃₇H₃₃N₅O₇P, 690.2123; found, 690.2113.

Chemical synthesis of deoxyribosyl nucleoside *H*-phosphinates

Scheme S1. Synthetic routes for the preparation of deoxyribosyl nucleoside *H*-phosphinates **S4a-e**, required for the solid-phase synthesis of dPhoNA containing oligonucleotides.

***N*⁶-Benzoyl-3'-*O*-(4-methoxytrityl)-2'-deoxyadenosine (**S2a**).** To a stirred slurry of *N*⁶-benzoyl-2'-deoxyadenosine (7.00 g, 19.7 mmol) in anhydrous pyridine (40 mL) was added *tert*-butylchlorodimethylsilane (50% in toluene, 6.83 mL, 19.7 mmol) at 0 °C. The reaction mixture was then stirred at room temperature overnight. After removal of all the volatiles under reduced pressure, the resulting residue was partitioned between EtOAc and saturated aq. NaHCO₃, the organic layer was washed with brine, dried over Na₂SO₄, filtered and concentrated in vacuo. The residue was purified by column chromatography (20:1 CH₂Cl₂/MeOH) to give *N*⁶-benzoyl-5'-*O*-*tert*-butyldimethylsilyl-2'-deoxyadenosine (7.80 g, 84%) as a white foam.

To a solution of *N*⁶-benzoyl-5'-*O*-*tert*-butyldimethylsilyl-2'-deoxyadenosine (5.00 g, 10.7 mmol) in dry THF (55 mL) were sequentially added pyridine (6.00 mL, 74.5 mmol), AgNO₃ (2.71 g, 16.00 mmol), and MMTTrCl (4.93 g, 16.0 mmol). After stirring at room temperature overnight, the reaction mixture was diluted with EtOAc and washed with saturated aq. NaHCO₃. The organic layer was washed with brine, dried over Na₂SO₄, filtered, and concentrated in vacuo. Next, anhydrous THF (100 mL) and Et₃N·3HF (4.58 mL, 28.1 mmol) were added to the residue obtained above. After stirring at room temperature for 24 h, the all the volatiles were removed under reduced pressure and the residue was dissolved in a mixture of EtOAc and saturated aq. NaHCO₃. The organic layer was washed with brine, dried over Na₂SO₄, filtered, and concentrated in vacuo. The crude residue was purified by column chromatography (2:1 to 0:1,

CH₂Cl₂/ EtOAc) to give **S2a** (5.50 g, 82%) as a white foam. ¹H NMR (500 MHz, CDCl₃) δ 9.35 (s, 1H, NH), 8.67 (s, 1H, H-2), 8.04 (s, 1H, H-8), 8.00–7.98 (m, 2H, ArH), 7.58–7.22 (m, 15H, ArH), 6.87–6.83 (m, 2H, ArH), 6.36 (dd, *J* = 9.9, 5.2 Hz, 1H, H-1'), 5.85 (dd, *J* = 11.5, 1.9 Hz, 1H, OH), 4.63 (d, *J* = 5.4 Hz, 1H, H-3'), 4.03 (s, 1H, H-4'), 3.78 (s, 3H, OMe), 3.72 (d, *J* = 13.2 Hz, 1H, H-5a'), 3.29 (dt, *J* = 12.9, 1.5 Hz, 1H, H-5b'), 2.72–2.62 (m, 1H, H-2a'), 1.75 (dd, *J* = 13.3, 5.4 Hz, 1H, H-2b'); ¹³C NMR (125 MHz, CDCl₃) δ 164.6 (CO), 158.8 (Ar-C), 151.9 (C-2), 150.6 (C-4), 150.2 (C-6), 144.6, 144.5 (Ar-C), 142.5 (C-8), 135.7, 133.4, 132.8, 130.5, 128.7, 128.44, 128.39, 128.02, 127.99, 127.8, 127.24, 127.22 (Ar-C), 124.5 (C-5), 113.3 (Ar-C), 88.6 (C-4'), 87.9 (C-1'), 87.7 (Ph₃C), 75.5 (C-3'), 63.0 (C-5'), 55.2 (OMe), 39.6 (C-2'); HRMS: [M+H]⁺ calcd for C₃₇H₃₄N₅O₅, 628.2554; found, 628.2564.

3'-O-(4-Methoxytrityl)-2'-deoxythymidine (S2b).³ To a solution of thymidine (10.0 g, 41.3 mmol) in dry pyridine (100 mL) was added pivaloyl chloride (5.6 mL, 45.4 mmol) at 0 °C. The reaction mixture was allowed to warm to room temperature and stirred overnight. 4-Methoxytrityl chloride (12.7 g, 41.3 mmol) and AgNO₃ (16.6 g, 53.7 mmol) were then added and the mixture was further stirred at room temperature for 24 h. It was then diluted with EtOAc and washed with saturated aq. NaHCO₃. The organic layer was dried over Na₂SO₄, filtered, and concentrated in vacuo. The residue was then dissolved in 300 mL of THF/ethanol (1:1, v/v) and 2 M aq. NaOH (150 mL, 300 mmol) was added. The reaction mixture was stirred at 65 °C for 3 h. After cooling to room temperature, NH₄Cl (17.0 g, 318 mmol) was added. The mixture was concentrated under reduced pressure to remove the organic phase, and the residual solution was extracted with EtOAc. The combined organic layers were washed with brine, dried over Na₂SO₄, filtered, and concentrated in vacuo. The residue was purified by column chromatography (3:1 to 1:1, CH₂Cl₂/EtOAc) to give **S2b** (18.9 g, 89%) as a white foam. ¹H NMR (300 MHz, CDCl₃) δ 9.64 (s, 1H, NH), 7.45 (d, *J* = 7.4 Hz, 4H, ArH), 7.35–7.17 (m, 9H, H-6 and ArH), 6.83 (d, *J* = 9.0 Hz, 2H, ArH), 6.20 (dd, *J* = 8.6, 5.8 Hz, 1H, H-1'), 4.37 (d, *J* = 5.8 Hz, 1H, H-3'), 3.98–3.93 (m, 1H, H-4'), 3.74 (s, 3H, OMe), 3.62 (d, *J* = 11.1 Hz, 1H, H-5a'), 3.29 (d, *J* = 11.8 Hz, 1H, H-5b'), 3.19 (br s, 1H, OH), 1.77 (d, *J* = 0.78 Hz, 3H, T-Me), 1.91–1.67 (m, 2H, H-2'); ¹³C NMR (75 MHz, CDCl₃) δ 164.1 (C-4), 158.6 (Ar-C), 150.4 (C-2), 144.5, 144.4 (Ar-C), 137.0 (C-6), 135.7, 130.4, 128.4, 128.3, 127.8, 127.1, 113.2 (Ar-C), 110.6 (C-5), 87.3 (Ph₃C), 86.6 (C-4'), 86.4 (C-1'), 74.3 (C-3'), 62.2 (C-5'), 55.1 (OMe), 38.6 (C-2'), 12.2 (Thy-Me).

3'-O-(4-Methoxytrityl)-2'-deoxyinosine (S2c). Compound **S2c** was obtained as a white foam (3.00 g, 48%) as described for **S2a**, starting from **S1c** (3.00 g, 11.9 mmol). ¹H NMR (300 MHz, CDCl₃) δ 12.98 (br s, 1H, NH), 8.23 (s, 1H, H-2), 7.93 (s, 1H, H-8), 7.50–7.21 (m, 12H, ArH), 6.85 (d, *J* = 8.9 Hz, 2H, ArH), 6.28 (dd, *J* = 9.6, 5.4 Hz, 1H, H-1'), 4.58 (d, *J* = 5.3 Hz, 1H, H-3'), 4.03 (s, 1H, H-4'), 3.77 (s, 3H, OMe), 3.68 (d, *J* = 11.8 Hz, 1H, H-5a'), 3.28 (d, *J* = 11.4 Hz, 1H, H-5b'), 2.57–2.47 (m, 1H, H-2a'), 1.78 (dd, *J* = 11.2, 5.2 Hz, 1H, H-2b'); ¹³C NMR (75 MHz, CDCl₃) δ 158.8 (Ar-C), 158.3 (C-6), 147.6 (C-4),

145.4 (C-2), 144.6, 144.5 (Ar-C), 139.8 (C-8), 135.7, 130.5, 128.5, 128.4, 128.0, 127.2 (Ar-C), 126.1 (C-5), 113.4 (Ar-C), 88.4 (C-4'), 87.7 (Ph₃C), 87.6 (C-1'), 75.4 (C-3'), 62.9 (C-5'), 55.2 (OMe), 40.2 (C-2'); HRMS: [M+Na]⁺ calcd for C₃₀H₂₈N₄O₅Na, 547.1952; found, 547.1952.

N⁶-Benzoyl-3'-O-(4-methoxytrityl)-5'-O-diethylphosphonomethyl-2'-deoxyadenosine (S3a). To a solution of **S2a** (5.40 g, 8.60 mmol) and diethylphosphonomethyl tosylate (5.55 g, 17.21 mmol) in dry THF (160 mL) was added NaH (60% in mineral oil, 1.03 g, 25.8 mmol) at -5 °C. The reaction mixture was allowed to warm to room temperature and stirred overnight. It was then quenched with saturated aq. NaHCO₃, extracted with EtOAc, dried over Na₂SO₄, filtered, and concentrated in vacuo. The residue was purified by column chromatography (20:1 CH₂Cl₂/MeOH) to give **S3a** (4.0 g, 60%) as a white foam. ¹H NMR (600 MHz, CDCl₃) δ 9.33 (s, 1H, NH), 8.79 (s, 1H, H-2), 8.43 (s, 1H, H-8), 8.02–7.98 (m, 2H, ArH), 7.58–7.24 (m, 15H, ArH), 6.87–6.85 (m, 2H, ArH), 6.66 (dd, *J* = 8.8, 5.8 Hz, 1H, H-1'), 4.49 (d, *J* = 5.6 Hz, 1H, H-3'), 4.16–4.00 (m, 4H, CH₂CH₃), 3.93 (br s, 1H, H-4'), 3.78 (s, 3H, OMe), 3.71–3.61 (m, 2H, PCH₂), 3.54 (dd, *J* = 10.5, 2.4 Hz, 1H, H-5a'), 3.27 (dd, *J* = 10.5, 3.0 Hz, 1H, H-5b'), 2.37–2.34 (m, 1H, H-2a'), 2.18–2.15 (m, 1H, H-2b'), 1.28–1.22 (m, 6H, CH₂CH₃); ¹³C NMR (150 MHz, CDCl₃) δ 164.6 (CO), 158.7 (Ar-C), 152.4 (C-2), 151.6 (C-4), 149.3 (C-6), 144.4 (Ar-C), 141.7 (C-8), 135.5, 133.6, 132.5, 130.4, 128.6, 128.4, 127.9, 127.7, 127.2 (Ar-C), 122.9 (C-5), 113.2 (Ar-C), 87.5 (Ph₃C), 85.3 (C-4'), 84.2 (C-1'), 75.1 (C-3'), 73.2 (d, ³*J*_{P,C} = 10.9 Hz, C-5'), 65.2 (d, ¹*J*_{P,C} = 167.9, PCH₂), 62.2 (CH₂CH₃), 55.1 (OMe), 40.2 (C-2'), 16.3 (CH₂CH₃); ³¹P NMR (121 MHz, CDCl₃): δ 21.4; HRMS: [M+H]⁺ calcd for C₄₂H₄₅N₅O₈P, 778.3000; found, 778.3003.

3'-O-(4-Methoxytrityl)-5'-O-diethylphosphonomethyl-2'-deoxythymidine (S3b). Compound **S3b** was obtained as a white foam (0.32 g, 75%) as described for **S3a**, starting from **S2b** (0.33 g, 0.64 mmol). ¹H NMR (300 MHz, CDCl₃) δ 8.89 (s, 1H, NH), 7.47–7.22 (m, 13H, H-6 and ArH), 6.85 (d, *J* = 9.0 Hz, 2H, ArH), 6.46 (dd, *J* = 9.3, 5.5 Hz, 1H, H-1'), 4.33 (d, *J* = 5.9 Hz, 1H, H-3'), 4.12–3.97 (m, 4H, CH₂CH₃), 3.83 (br s, 1H, H-4'), 3.80 (s, 3H, OMe), 3.77–3.58 (m, 2H, PCH₂), 3.57–3.50 (m, 1H, H-5a'), 3.13 (dd, *J* = 10.5, 2.6 Hz, 1H, H-5b'), 1.91 (s, 3H, Thy-Me), 1.97–1.87 (m, 1H, H-2a'), 1.81–1.71 (m, 1H, H-2b'), 1.28–1.20 (m, 6H, CH₂CH₃); ¹³C NMR (75 MHz, CDCl₃) δ 163.7 (C-4), 158.8 (Ar-C), 150.4 (C-2), 144.5, 135.73 (Ar-C), 135.65 (C-6), 130.4, 128.5, 128.0, 127.2, 113.3 (Ar-C), 111.2 (C-5), 87.6 (Ph₃C), 84.9 (C-4'), 84.8 (C-1'), 74.9 (C-3'), 73.2 (d, ³*J*_{P,C} = 10.5 Hz, C-5'), 65.4 (d, ¹*J*_{P,C} = 167.3, PCH₂), 62.2 (CH₂CH₃), 55.2 (OMe), 39.3 (C-2'), 16.5 (CH₂CH₃), 12.4 (Thy-Me); ³¹P NMR (121 MHz, CDCl₃): δ 20.6; HRMS: [M+Na]⁺ calcd for C₃₅H₄₁N₂O₉PNa, 687.2442; found, 687.2455.

3'-O-(4-Methoxytrityl)-5'-O-diethylphosphonomethyl-2'-deoxyinosine (S3c). Compound **S3c** was obtained as a white foam (0.50 g, 60%) as described for **S3a**, starting from **S2c** (0.65 g, 1.24 mmol). ¹H NMR (300 MHz, CDCl₃) δ 13.07 (br s, 1H, NH), 8.17 (s, 1H, H-2), 8.15 (s, 1H, H-8), 7.50–7.23 (m, 12H, ArH), 6.90–6.82 (m, 2H, ArH), 6.52 (dd, *J* = 8.6, 5.6 Hz, 1H, H-1'), 4.46 (d, *J* = 5.3 Hz, 1H, H-3'), 4.18–

4.01 (m, 4H, CH_2CH_3), 3.90 (s, 1H, H-4'), 3.79 (s, 3H, OMe), 3.75–3.60 (m, 2H, PCH_2), 3.53 (dd, $J = 10.1, 2.0$ Hz, 1H, H-5a'), 3.21 (dd, $J = 10.5, 2.8$ Hz, 1H, H-5b'), 2.35–2.26 (m, 1H, H-2a'), 2.16–2.10 (m, 1H, H-2b'), 1.32–1.22 (m, 6H, CH_2CH_3); ^{13}C NMR (75 MHz, CDCl_3) δ 159.1 (C-6), 158.9 (Ar-C), 148.8 (C-4), 144.9 (C-2), 144.5 (Ar-C), 138.6 (C-8), 135.7, 130.5, 128.5, 128.0, 127.3 (Ar-C), 124.6 (C-5), 113.4 (Ar-C), 87.6 (Ph_3C), 85.5 (C-4'), 84.3 (C-1'), 75.2 (C-3'), 73.2 (d, $^3J_{\text{P,C}} = 11.1$ Hz, C-5'), 65.3 (d, $^1J_{\text{P,C}} = 166.7$, PCH_2), 62.3 (CH_2CH_3), 55.2 (OMe), 40.4 (C-2'), 16.5 (CH_2CH_3); ^{31}P NMR (121 MHz, CDCl_3): δ 21.4; HRMS: $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{35}\text{H}_{40}\text{N}_4\text{O}_8\text{P}$, 675.2578; found, 675.2575.

3'-O-(4-Methoxytrityl)-5'-O-diethylphosphonomethyl-2'-deoxyadenosine (S3d). Compound **S3d** was obtained as a white foam (0.90 g, 90%) as described for **2d**, starting from **S3a** (1.20 g, 1.54 mmol). ^1H NMR (600 MHz, CDCl_3) δ 8.34 (s, 1H, H-2), 8.24 (s, 1H, H-8), 7.49–7.23 (m, 12H, ArH), 6.86–6.83 (m, 2H, ArH), 6.59 (dd, $J = 8.9, 5.7$ Hz, 1H, H-1'), 6.51 (br s, 2H, NH_2), 4.46 (d, $J = 5.7$ Hz, 1H, H-3'), 4.12–4.07 (m, 4H, CH_2CH_3), 3.92 (br s, 1H, H-4'), 3.78 (s, 3H, OMe), 3.70–3.62 (m, 2H, PCH_2), 3.50 (dd, $J = 10.4, 2.3$ Hz, 1H, H-5a'), 3.23 (dd, $J = 10.5, 2.9$ Hz, 1H, H-5b'), 2.35–2.30 (m, 1H, H-2a'), 2.11–2.07 (m, 1H, H-2b'), 1.28–1.22 (m, 6H, CH_2CH_3); ^{13}C NMR (150 MHz, CDCl_3) δ 158.8 (Ar-C), 155.7 (C-6), 153.0 (C-2), 149.7 (C-4), 144.5 (Ar-C), 138.9 (C-8), 135.7, 130.4, 128.5, 128.0, 127.2 (Ar-C), 119.4 (C-5), 113.3 (Ar-C), 87.5 (Ph_3C), 85.2 (C-4'), 83.9 (C-1'), 75.2 (C-3'), 73.4 (d, $^3J_{\text{P,C}} = 12.1$ Hz, C-5'), 65.4 (d, $^1J_{\text{P,C}} = 167.5$, PCH_2), 62.3 (CH_2CH_3), 55.2 (OMe), 40.2 (C-2'), 16.4 (CH_2CH_3); ^{31}P NMR (121 MHz, CDCl_3): δ 25.5; HRMS: $[\text{M}+\text{H}]^+$ calcd for $\text{C}_{35}\text{H}_{41}\text{N}_5\text{O}_7\text{P}$, 674.2738; found, 674.2751.

[2'-Deoxy-3'-O-(4-methoxytrityl)-5'-O-adenosinyl]methyl-(H)-phosphinate (S4a). Compound **S4a** was obtained as a white foam (1.65 g, 75%, Et_3N salt) as described for **3a**, starting from **S3d** (2.10 g, 3.12 mmol). ^1H NMR (300 MHz, CDCl_3) δ 8.75 (s, 1H, H-2), 8.23 (s, 1H, H-8), 7.83 (br s, 2H, NH_2), 7.50–7.23 (m, 12H, ArH), 7.18 (d, $J = 501.1$ Hz, 1H, PH), 6.85–6.82 (m, 2H, ArH), 6.67–6.62 (m, 1H, H-1'), 4.47 (d, $J = 5.4$ Hz, 1H, H-3'), 3.79 (s, 4H, OMe and H-4'), 3.50–3.00 (m, 4H, PCH_2 and H-5'), 2.55–2.36 (m, 1H, H-2a'), 2.12–2.06 (m, 1H, H-2b'); ^{13}C NMR (75 MHz, CDCl_3) δ 158.7 (Ar-C), 156.0 (C-6), 152.7 (C-2), 149.4 (C-4), 144.7 (Ar-C), 139.8 (C-8), 135.9, 130.5, 128.5, 127.9, 127.1 (Ar-C), 119.0 (C-5), 113.3 (Ar-C), 87.4 (Ph_3C), 85.7 (C-4'), 83.7 (C-1'), 75.9 (C-3'), 73.5 (d, $^3J_{\text{P,C}} = 13.6$ Hz, C-5'), 72.4 (d, $^1J_{\text{P,C}} = 107.5$, PCH_2), 55.2 (OMe), 40.8 (C-2'); ^{31}P NMR (121 MHz, CDCl_3): δ 20.7; HRMS: $[\text{M}-\text{H}]^-$ calcd for $\text{C}_{31}\text{H}_{31}\text{N}_5\text{O}_6\text{P}$, 600.2017; found, 600.2004.

[2'-Deoxy-3'-O-(4-methoxytrityl)-5'-O-thymidinyl]methyl-(H)-phosphinate (S4b). Compound **S4b** was obtained as a white foam (2.10 g, 67%, Et_3N salt) as described for **3a**, starting from **S3b** (3.00 g, 4.51 mmol). ^1H NMR (500 MHz, CDCl_3) δ 7.59 (d, $J = 1.2$ Hz, 1H, H-6), 7.47–7.20 (m, 12H, ArH), 7.13 (dt, $J = 497.4, 2.0$ Hz, 1H, PH), 6.85–6.81 (m, 2H, ArH), 6.50 (dd, $J = 8.4, 6.1$ Hz, 1H, H-1'), 4.38–4.36 (m, 1H, H-3'), 3.78 (s, 3H, OMe), 3.73 (s, 1H, H-4'), 3.50 (dd, $J = 10.5, 2.2$ Hz, 1H, H-5a'), 3.67–3.38 (m, 2H, PCH_2), 3.04 (dd, $J = 10.5, 2.5$ Hz, 1H, H-5b'), 2.00–1.90 (m, 2H, H-2'), 1.91 (s, 3H, Thy-Me); ^{13}C

NMR (125 MHz, CDCl₃) δ 164.2 (C-4), 158.5 (Ar-C), 150.7 (C-2), 144.53, 144.50 (Ar-C), 135.9 (C-6), 135.7, 130.3, 128.3, 127.7, 126.9, 113.1 (Ar-C), 110.7 (C-5), 87.2 (Ph₃C), 85.1 (C-4'), 84.7 (C-1'), 75.2 (C-3'), 72.8 (d, $^3J_{\text{P,C}} = 10.6$ Hz, C-5'), 72.0 (d, $^1J_{\text{P,C}} = 106.7$, PCH₂), 55.0 (OMe), 39.1 (C-2'), 12.2 (Thy-Me); ^{31}P NMR (121 MHz, CDCl₃): δ 16.7; HRMS: [M-H]⁻ calcd for C₃₁H₃₂N₂O₈P, 591.1902; found, 591.1901.

[2'-Deoxy-3'-O-(4-methoxytrityl)-5'-O-inosinyl]methyl-(H)-phosphinate (S4c). Compound **S4c** was obtained as a white foam (0.33 g, 70%, Et₃N salt) as described for **3a**, starting from **S3c** (0.43 g, 0.64 mmol). ^1H NMR (300 MHz, CDCl₃) δ 8.42 (s, 1H, H-2), 7.97 (s, 1H, H-8), 7.49–7.20 (m, 12H, ArH), 7.14 (d, $J = 501.6$ Hz, 1H, PH), 6.86–6.81 (m, 2H, ArH), 6.50 (dd, $J = 8.3, 5.9$ Hz, 1H, H-1'), 4.46 (d, $J = 4.7$ Hz, 1H, H-3'), 3.83 (s, 1H, H-4'), 3.78 (s, 3H, OMe), 3.53–3.33 (m, 3H, PCH₂ and H-5a'), 3.16–3.13 (m, 1H, H-5b'), 2.44–2.35 (m, 1H, H-2a'), 2.12–2.05 (m, 1H, H-2b'); ^{13}C NMR (75 MHz, CDCl₃) δ 158.8 (Ar-C), 158.1 (C-6), 148.4 (C-4), 145.6 (C-2), 144.6 (Ar-C), 138.9 (C-8), 135.7, 130.4, 128.5, 128.0, 127.2 (Ar-C), 124.0 (C-5), 113.3 (Ar-C), 87.5 (Ph₃C), 85.7 (C-4'), 83.9 (C-1'), 75.4 (C-3'), 73.0 (d, $^3J_{\text{P,C}} = 12.0$ Hz, C-5'), 71.8 (d, $^1J_{\text{P,C}} = 106.7$, PCH₂), 55.2 (OMe), 40.2 (C-2'); ^{31}P NMR (121 MHz, CDCl₃): δ 16.5. HRMS: [M-H]⁻ calcd for C₃₁H₃₀N₄O₇P, 601.1857; found, 601.1832.

[N⁶-(Dimethylamino)methylene-2'-deoxy-3'-O-(4-methoxytrityl)-5'-O-adenosinyl]methyl-(H)-phosphinate (S4d). Compound **S4d** was obtained as a white foam (0.70 g, 72%, Et₃N salt) as described for **3d**, starting from **S4a** (0.90 g, 1.28 mmol). ^1H NMR (300 MHz, CDCl₃) δ 8.95 (s, 1H, H-2), 8.56 (s, 1H, CHN(CH₃)₂), 8.52 (s, 1H, H-8), 7.50–7.21 (m, 12H, ArH), 7.14 (d, $J = 502.1$ Hz, 1H, PH), 6.86–6.83 (m, 2H, ArH), 6.66 (dd, $J = 8.9, 5.6$ Hz, 1H, H-1'), 4.48 (d, $J = 5.0$ Hz, 1H, H-3'), 3.83 (s, 1H, H-4'), 3.79 (s, 3H, OMe), 3.54–3.36 (m, 3H, PCH₂ and H-5a'), 3.25–3.14 (m, 7H, N-Me and H-5b'), 2.56–2.47 (m, 1H, H-2a'), 2.16–2.10 (m, 1H, H-2b'); ^{13}C NMR (75 MHz, CDCl₃) δ 159.3 (C-4), 158.7 (Ar-C), 158.1 (C-2), 152.4 (CHN(CH₃)₂), 151.8 (C-6), 144.7 (Ar-C), 141.0 (C-8), 135.9, 130.5, 128.5, 127.9, 127.1 (Ar-C), 125.9 (C-5), 113.2 (Ar-C), 87.4 (Ph₃C), 85.6 (C-4'), 84.0 (C-1'), 75.6 (C-3'), 73.2 (d, $^3J_{\text{P,C}} = 11.2$ Hz, C-5'), 72.1 (d, $^1J_{\text{P,C}} = 107.9$, PCH₂), 55.2 (OMe), 41.1 (N-Me), 40.0 (C-2'), 34.9 (N-Me); ^{31}P NMR (121 MHz, CDCl₃): δ 16.8; HRMS: [M+H]⁺ calcd for C₃₄H₃₈N₆O₆P, 657.2599; found, 657.2585.

[N⁶-Benzoyl-2'-deoxy-3'-O-(4-methoxytrityl)-5'-O-adenosinyl]methyl-(H)-phosphinate (S4e). Compound **S4e** was prepared as a white foam (0.85 g, 70%, Et₃N salt) as described for **3e**, starting from **S2a** (0.92 g, 1.47 mmol). ^1H NMR (300 MHz, DMSO-*d*₆) δ 11.23 (NH), 8.72 (s, 2H, H-2 and H-8), 8.05–8.03 (m, 2H, ArH), 7.59–7.24 (m, 15H, ArH), 6.93–6.90 (m, 2H, ArH), 6.51 (s, 1H, H-1'), 4.36 (s, 1H, H-3'), 3.71 (s, 4H, OMe and H-4'), 3.30–3.24 (m, 4H, PCH₂ and H-5'), 2.55–2.36 (m, 1H, H-2a'), 2.13–2.06 (m, 1H, H-2b'); ^{13}C NMR (75 MHz, DMSO-*d*₆) δ 165.7 (CO), 158.4 (Ar-C), 151.8 (C-4), 151.4 (C-2), 149.9 (C-6), 144.5, 144.4 (Ar-C), 143.0 (C-8), 135.3, 133.3, 132.3, 130.2, 128.4, 128.3, 128.1, 128.0, 127.0 (Ar-C), 125.0 (C-5), 113.4 (Ar-C), 86.9 (Ph₃C), 84.5 (C-4'), 83.7 (C-1'), 74.6 (C-3'), 72.3 (C-5'),

70.1 (d, $^1J_{\text{P,C}} = 110.3$, PCH_2), 55.0 (OMe), 38.2 (C-2'); ^{31}P NMR (121 MHz, $\text{DMSO}-d_6$) δ 11.3; HRMS: $[\text{M}-\text{H}]^-$ calcd for $\text{C}_{38}\text{H}_{35}\text{N}_5\text{O}_7\text{P}$, 704.2279; found, 704.2273.

Chemical synthesis of diphosphates of threosyl nucleoside phosphonates

Scheme S2. Synthetic route for the preparation of diphosphates of threosyl nucleoside phosphonates (PMTNpp) used for polymerase incorporation experiments and evolution.

1'- α -(Adenin-9-yl)-3'-O-phosphonomethyl-L-threose triethylammonium salt (S5a). To a solution of **S4a** (0.50 g, 1.20 mmol) and 2,6-lutidine (1.12 mL, 9.63 mmol) in dry CH_3CN (50 mL) was added bromotrimethylsilane (1.27 mL, 9.63 mmol) at 0 °C. The reaction mixture was stirred at room temperature overnight and quenched with a 1.0 M TEAB solution (3 mL). After removal of all the volatiles under reduced pressure, the resulting residue was partitioned between water and EtOAc/ether (1:1), the aqueous layer was lyophilized, and the residue was purified by chromatography on a silica gel column (10:1:0 to 10:5:1, $\text{CH}_2\text{Cl}_2/\text{MeOH}/1$ M TEAB) to give the crude product. A further purification step was carried out using preparative reverse phase HPLC using CH_3CN in a 0.05 M TEAB solution as eluent with a gradient ranging from 2 to 30%. The desired product **S5a** (Et_3N salt, 340 mg, 65%) was obtained as a white foam. ^1H NMR (300 MHz, D_2O): δ 8.37 (s, 1H, H-2), 8.11 (s, 1H, H-8), 5.98 (s, 1H, H-1'), 4.75 (s, 1H, H-2'), 4.26 (br s, 3H, H-3' and H-4'), 3.46–3.50 (m, 2H, PCH_2); ^{13}C NMR (75 MHz, D_2O): δ 154.6 (C-6), 151.9 (C-2), 147.8 (C-4), 140.3 (C-8), 117.6 (C-5), 89.0 (C-1'), 84.7 (d, $^3J_{\text{P,C}} = 10.5$ Hz, C-3'), 77.7 (C-2'), 72.3 (C-4'), 66.2 (d, $^1J_{\text{P,C}} = 152.4$ Hz, PCH_2); ^{31}P NMR (121 MHz, D_2O): δ 13.9. HRMS: $[\text{M}-\text{H}]^-$ calcd for $\text{C}_{10}\text{H}_{13}\text{N}_5\text{O}_6\text{P}$, 330.0609; found, 330.0615.

1'- α -(Thymin-1-yl)-3'-O-(phosphonomethyl)-L-threose triethylammonium salt (S5b). Compound **S5b** was obtained as a white foam (26 mg, 47%) as described for **S5a**, starting from **1b** (53 mg, 0.13 mmol).

^1H NMR (300 MHz, D_2O): δ 8.22 (br s, 1H, NH), 7.53 (s, 1H, H-6), 5.77 (s, 1H, H-1'), 4.37 (s, 1H, H-2'), 4.31 (d, $J = 9.8$ Hz, 1H, H-4a'), 4.13–4.18 (m, 2H, H-4b' and H-3'), 3.52 (d, $J = 8.4$ Hz, 2H, PCH_2), 1.81 (s, 3H, Thy- CH_3); ^{13}C NMR (75 MHz, D_2O): δ 166.3 (C-4), 151.3 (C-2), 137.7 (C-6), 110.5 (C-5), 91.0 (C-1'), 84.2 (d, $^3J_{\text{P,C}} = 10.4$ Hz, C-3'), 77.4 (C-2'), 72.5 (C-4'), 65.8 (d, $^1J_{\text{P,C}} = 149.6$ Hz, PCH_2), 11.3 (Thy- CH_3); ^{31}P NMR (121 MHz, D_2O): δ 14.2; HRMS: $[\text{M-H}]^-$ calcd for $\text{C}_{10}\text{H}_{14}\text{N}_2\text{O}_8\text{P}$, 321.0493; found, 321.0487.

1' α -(Cytosin-1-yl)-3'-O-phosphonomethyl-L-threose triethylammonium salt (S5c). Compound **S5c** was obtained as a white foam (110 mg, 57%) as described for **S5a**, starting from **S4b** (185 mg, 0.47 mmol). ^1H NMR (300 MHz, D_2O): δ 7.72 (d, $J = 7.7$ Hz, 1H, H-6), 5.95 (d, $J = 7.8$ Hz, 1H, H-5), 5.76 (s, 1H, H-1'), 4.39 (d, $J = 10.6$ Hz, 1H, H-4a'), 4.35 (s, 1H, H-2'), 4.19 (dd, $J = 10.6$ Hz, 3.4 Hz, 1H, H-4b'), 4.04 (d, $J = 3.1$ Hz, 1H, H-3'), 3.44–3.58 (m, 2H, PCH_2), ^{13}C NMR (75 MHz, D_2O): δ 164.2 (C-4), 154.9 (C-2), 142.6 (C-6), 94.8 (C-5), 92.3 (C-1'), 84.2 (d, $^3J_{\text{P,C}} = 11.9$ Hz, C-3'), 77.0 (C-2'), 73.4 (C-4'), 64.7 (d, $^1J_{\text{P,C}} = 157.4$ Hz, PCH_2); ^{31}P NMR (121 MHz, D_2O): δ 15.0; HRMS: $[\text{M-H}]^-$ calcd for $\text{C}_9\text{H}_{13}\text{N}_3\text{O}_7\text{P}$, 306.0497; found, 306.0496.

1' α -(Guanin-1-yl)-3'-O-phosphonomethyl-L-threose triethylammonium salt (S5d). To a solution of **S4c** (0.50 g, 0.90 mmol) in anhydrous CH_3CN (50 mL) was added 2,6-lutidine (0.84 mL, 7.22 mmol) and trimethylsilylbromide (0.95 mL, 7.22 mmol) at room temperature, and the mixture was stirred overnight. The volatiles were removed under reduced pressure and co-evaporated three times with anhydrous methanol to leave a crude residue. Subsequently, 2-mercaptoethanol (0.32 mL, 4.51 mmol) and NaOMe (1.0 M in MeOH, 4.51 mL, 4.51 mmol) were added to the above residue. The mixture was refluxed for 19 h, cooled, quenched with 5 mL of TEAB buffer, and evaporated in vacuo. The residue was partitioned between water and EtOAc, the aqueous phase was lyophilized, and the residue was purified by RP-HPLC using a gradient of CH_3CN in a 0.1 M TEAB buffer solution ranging from 2 to 30% to afford **S5d** (0.17 g, 42%) as a yellow solid. ^1H NMR (300 MHz, D_2O): δ 7.92 (s, 1H, H-8), 5.73 (d, $J = 3.9$ Hz, 1H, H-1'), 4.70 (1H, H-2', overlapped with H_2O), 4.12–4.21 (m, 3H, H-3' and H-4'), 3.50 (d, $J = 8.9$ Hz, 2H, PCH_2); ^{13}C NMR (75 MHz, D_2O): δ 167.3 (C-6), 160.6 (C-2), 151.1 (C-4), 136.0 (C-8), 117.1 (C-5), 87.7 (C-1'), 84.4 (d, $^3J_{\text{P,C}} = 8.9$ Hz, C-3'), 77.6 (C-2'), 71.2 (C-4'), 67.5 (d, $^1J_{\text{P,C}} = 148.7$ Hz, PCH_2); ^{31}P NMR (121 MHz, D_2O): δ 12.8. HRMS: $[\text{M-H}]^-$ calcd for $\text{C}_{10}\text{H}_{13}\text{N}_5\text{O}_7\text{P}$, 346.0558; found, 346.0554.

General procedure for the synthesis of diphosphates of threosyl nucleoside phosphonates. Morpholine (0.5 mmol) was added to a solution of phosphonic acid (**S5a-d**, 0.1 mmol) in a *t*-butyl alcohol-water mixture (1:1, 6 mL), and the resulting mixture was then heated to reflux. A solution of dicyclohexylcarbodiimide (0.5 mmol) in *t*-butyl alcohol (3 mL) was then added dropwise over a period of 3–5 h, and the mixture was further refluxed for 2 h. The solution was cooled to room temperature, and the solvent was evaporated under reduced pressure. The resulting residue was suspended in water (6 mL) and

filtered through a syringe filter. The filtrate was washed with diethyl ether (6 mL), and the aqueous phase was lyophilized. The resulting solid (**S6a-d**) was dissolved in DMSO (2 mL), and then bis(tributylammonium)pyrophosphate (0.3 mmol) was added. The reaction mixture was stirred at room temperature for 36 h. The solution was diluted with water, filtered, and directly purified by ion-exchange chromatography on a Sephadex DEAE A-25 column and eluted with a linear gradient of 0–1.0 M aq. triethylammonium bicarbonate (TEAB) buffer. The fractions containing the desired product were pooled and further purified by RP-HPLC (Gemini 10 μ m C18 250 $\times$ 21.2 mm column) using a linear gradient of acetonitrile from 0 to 25% in 0.1 M TEAB buffer. The relevant fractions were evaporated and then lyophilized to give the title compound as a white foam. Yields for all PMTpp products ranged from 30 to 40%.

1' α -(Adenin-9-yl)-3'-O-(diphosphate-phosphonomethyl)-L-threose triethylammonium salt (PMTApp). Yield 35%. ^1H NMR (300 MHz, D_2O): δ 8.35 (s, 1H, H-2), 8.16 (s, 1H, H-8), 6.01 (s, 1H, H-1'), 4.76 (s, 1H, H-2'), 4.25–4.38 (m, 3H, H-3' and H-4'), 3.77–3.92 (m, 2H, PCH_2); ^{13}C NMR (75 MHz, D_2O): δ 155.3 (C-6), 152.4 (C-2), 148.5 (C-4), 140.6 (C-8), 118.2 (C-5), 89.1 (C-1'), 84.6 (d, $^3J_{\text{P,C}} = 8.9$ Hz, C-3'), 77.7 (C-2'), 72.7 (C-4'), 65.0 (d, $^1J_{\text{P,C}} = 160.9$ Hz, PCH_2); ^{31}P NMR (121 MHz, D_2O): δ 7.63 (d, $J = 25.5$ Hz, P_α), -6.90 (d, $J = 20.4$ Hz, P_γ), -22.60 (t, $J = 23.2$ Hz, P_β); HRMS: $[\text{M-H}]^-$ calcd for $\text{C}_{10}\text{H}_{15}\text{N}_5\text{O}_6\text{P}_3$, 489.9935; found, 489.9932.

1' α -(Thymin-1-yl)-3'-O-(diphosphate-phosphonomethyl)-L-threose triethylammonium salt (PMTTpp). Yield 30%. ^1H NMR (300 MHz, D_2O): δ 7.55 (s, 1H, H-6), 5.77 (d, $J = 2.0$ Hz 1H, H-1'), 4.40 (s, 1H, H-2'), 4.34 (d, $J = 10.5$ Hz, 1H, H-4a'), 4.24 (d, $J = 4.0$ Hz, 1H, H-3'), 4.15 (dd, $J = 10.5$ Hz, 4.0 Hz, H-4b'), 3.73–3.87 (m, 2H, PCH_2), 1.82 (d, $J = 1.1$ Hz, 3H, Thy-Me); ^{13}C NMR (75 MHz, D_2O): δ 166.4 (C-4), 151.4 (C-2), 137.8 (C-6), 110.6 (C-5), 90.9 (C-1'), 84.2 (d, $^3J_{\text{P,C}} = 8.9$ Hz, C-3'), 77.5 (C-2'), 72.7 (C-4'), 64.7 (d, $^1J_{\text{P,C}} = 161.2$ Hz, PCH_2), 11.3 (Thy- CH_3); ^{31}P NMR (121 MHz, D_2O): δ 10.6 (d, $J = 25.4$ Hz, P_α), -4.26 (d, $J = 19.4$ Hz, P_γ), -19.71 (t, $J = 21.8$ Hz, P_β); HRMS: $[\text{M-H}]^-$ calcd for $\text{C}_{10}\text{H}_{16}\text{N}_2\text{O}_{14}\text{P}_3$, 480.9820; found, 480.9827.

1' α -(Cytosin-1-yl)-3'-O-(diphosphate-phosphonomethyl)-L-threose triethylammonium salt (PMTCpp). Yield 30%. ^1H NMR (300 MHz, D_2O): δ 7.73 (d, $J = 7.6$ Hz, 1H, H-6), 6.00 (d, $J = 7.6$ Hz, 1H, H-5), 5.81 (s, 1H, H-1'), 4.42 (d, $J = 10.2$ Hz, 1H, H-4a'), 4.35 (s, 1H, H-2'), 4.15–4.18 (m, 2H, H-4b' and H-3'), 3.65–3.82 (m, 2H, PCH_2); ^{13}C NMR (75 MHz, D_2O): δ 166.0 (C-4), 157.3 (C-2), 142.2 (C-6), 95.5 (C-5), 92.0 (C-1'), 84.4 ($^3J_{\text{P,C}} = 10.2$ Hz, C-3'), 77.5 (C-2'), 73.0 (C-4'), 64.5 ($^1J_{\text{P,C}} = 162.7$ Hz, PCH_2); ^{31}P NMR (121 MHz, D_2O): δ 7.72 (d, $J = 25.5$ Hz, P_α), -8.63 (d, $J = 19.6$ Hz, P_γ), -22.93 (dd, $J = 24.8$ Hz, 21.6 Hz, P_β); HRMS: $[\text{M-H}]^-$ calcd for $\text{C}_9\text{H}_{15}\text{N}_3\text{O}_{13}\text{P}_3$, 465.9823; found, 465.9826.

1' α -(Guanin-9-yl)-3'-O-(diphosphate-phosphonomethyl)-L-threose triethylammonium salt (PMTGpp). Yield 40%. ^1H NMR (300 MHz, D_2O): δ 7.97 (s, 1H, H-8), 5.80 (d, $J = 2.6$ Hz, 1H, H-1'),

4.78 (s, 1H, H-2'), 4.21-4.34 (m, 3H, H-3' and H-4'), 3.78-3.91 (m, 2H, PCH₂); ¹³C NMR (75 MHz, D₂O): δ 158.7 (C-6), 153.6 (C-2), 151.1 (C-4), 138.1 (C-8), 115.7 (C-5), 88.9 (C-1'), 84.6 (³J_{P,C} = 9.4 Hz, C-3'), 77.4 (C-2'), 72.2 (C-4'), 65.0 (¹J_{P,C} = 161.3 Hz, PCH₂); ³¹P NMR (121 MHz, D₂O): δ 7.87 (d, *J* = 26.0 Hz, P _{α}), -8.29 (d, *J* = 19.6 Hz, P _{γ}), -22.81 (dd, *J* = 25.1 Hz, 21.2 Hz, P _{β}); HRMS: [M-H]⁻ calcd for C₁₀H₁₅N₅O₁₃P₃, 505.9885; found, 505.9875.

Solid-phase synthesis of phosphonate containing oligonucleotides

Scheme S3. Synthesis cycle for the preparation of tPhoNA oligomeric sequences. An analogous methodology was employed for the generation of dPhoNA oligomers.

Table S1. Experimental Details of the Solid-Phase Protocol Utilized for PhoNA Synthesis.

Detritylation	3% DCA in DCM	6 min
Coupling	0.1 M <i>H</i> -phosphinate monomers in pyridine-acetonitrile (50:50, v/v); 0.4 M NEP in pyridine-acetonitrile (5:95, v/v)	10 min
Oxidation	CCl ₄ -methanol-Et ₃ N-MeIm (70:20:5:5; v/v)	30 min
Capping	5% phenoxyacetic anhydride in THF; 10% 1-methylimidazole in THF/pyridine (8:1, v/v)	3 min
End of cycle		
Deprotection	a) thiophenol-Et ₃ N-DMF mixture (23:32:45; v/v)	4 h
	b) 33% aqueous ammonia, room temperature	16 h

Figure S1. HPLC profiles of the dPhoNA oligomer Ttt ttt ttt ttt tt (14-mer) after incubation with 33% aqueous ammonia at room temperature or 55 °C for 16 h. Mono-Q 10/100 GL column (Pharmacia), 2 mL/min, Solvent A: 10 mM NaOH + 0.1 M NaCl (pH = 12.0), Solvent B: 10 mM NaOH + 0.9 M NaCl (pH = 12.0).

Incorporation experiments using commercially available DNA polymerases

General conditions for primer extension assays. Primer and templates were purchased from Integrated DNA Technologies (IDT). The primer oligonucleotide was annealed with the corresponding template oligonucleotides in a 1:2 molar ratio by heating the mixture at 75 °C for 5 min, followed by slow cooling

to room temperature. The DNA polymerization mixtures contained 125 nM of primer-template complex, 1x reaction buffer (supplied with the enzyme), different concentrations of building blocks (125, 500, and 1 mM) and either 0.01 U. μ L⁻¹ Vent (*exo-*) or 0.05 U. μ L⁻¹ Klenow (*-exo*) (New England Biolabs). The reactions were carried out at 75 or 37 °C, respectively, and aliquots were taken after 15, 30, and 60 min. In the control reaction, 50 μ M of natural deoxynucleoside triphosphates was used. All polymerase reactions were quenched by adding a double volume of gel loading buffer (90% formamide, 50 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol). Samples were heated at 90 °C for 5 min prior to separation on a 1 mm 15% denaturing polyacrylamide gel, and the gel bands were visualized using Ettan DIGE Imager (GE Healthcare).

Figure S2. Incorporation of PMTApp, PMTTpp, PMTCpp, and PMTGpp into P₁T₁, P₁T₂, P₁T₃, and P₁T₄ (Table S2), respectively, by (A) Klenow (*exo-*) fragment and (B) Vent (*exo-*) DNA polymerase.

Table S2. Overview of Primer and Template Sequences used in DNA Polymerase Primer Extension Reactions.

Substrate	Primer and templates ^a
	P ₁ : 5'- Cy5 -CTCGCTATGACACTCGTCTCTT-3'
PMTApp	T ₁ : 3'-GAGCGATACTGTGAGCAGAGAA TTTTTTT CAA-5'
PMTTpp	T ₂ : 3'-GAGCGATACTGTGAGCAGAGAA AAAAAA AGCC-5'
PMTCpp	T ₃ : 3'-GAGCGATACTGTGAGCAGAGAA GGGGG TCAA-5'
PMTGpp	T ₄ : 3'-GAGCGATACTGTGAGCAGAGAA CCCCC CTTAA-5'

^aRed letters in bold indicate the template overhang in each hybridized primer-template duplex.

Engineering of a tPhoNA polymerase by directed evolution of *Thermococcus gorgonarius* DNA polymerase (Tgo)

General Information. Enzymes were purchased from New England Biolabs (NEB; Ipswich, MA, USA) unless otherwise stated. MyTaq HS DNA Polymerase was acquired from Bioline (London, UK). All oligonucleotides were obtained from Integrated DNA Technologies (IDT; Coralville, IA, USA).

Point mutations. All polymerase engineering was carried out in plasmid pET29aΔ_TgoT – a variant of pET29a(+) lacking *rop* and harbouring a codon optimised TgoT gene (with a C-terminal hexahistidine tag) under the regulation of its T7 promoter. All point mutations and single-site diversification were made by inverse PCR (iPCR)⁴ using Q5 Hot Start High-Fidelity DNA Polymerase (NEB). PCRs were cleaned up using GeneJET PCR purification kit (Thermo Fisher Scientific, Waltham MA, USA), co-treated with 0.5 μL each DpnI (NEB) and T4 polynucleotide kinase (NEB) in 1x T4 DNA ligase buffer at 37 °C for 5-20 min to degrade the template and phosphorylate the PCR products. Next, 0.5 μL T4 DNA ligase was added and incubated for 30 min and 2 h at room temperature before phenol:chloroform extraction, ethanol precipitation, and subsequent transformation into T7 Express *LysY/T^r* cells (NEB).

Polymerase screening. Small scale screens were carried out in 96-well plates. Single residues were diversified by iPCR using small-intelligent libraries⁵ and 60 transformants picked into 96-well plates containing 100 μL 2xTY supplemented with 50 μg/mL kanamycin. After overnight culture at 37 °C (220 RPM), 5 μL from each well was transferred to a fresh 200 μL 2xTY supplemented with kanamycin and cultured for 3 h at 37 °C (220 RPM). Polymerase expression was induced with 1 mM IPTG (final concentration) for 2-4 h at 37 °C (220 RPM). The overnight plate was made up to 20% glycerol by addition of 0.5 volume 60% glycerol and stored frozen as a master plate. After expression, the cultures were harvested, supernatant removed, and resuspended in 20 μL 1x Thermopol buffer and transferred to a 96-well PCR plate. These cells were lysed (20 min at 80 °C in a PCR machine), clarified (10 min centrifugation at 3220 g, 4 °C), and the lysates were directly used for primer extension.

Polymerase expression and purification. TgoT (and variants generated) was expressed from pET29aΔ_TgoT in 2xTY supplemented with 50 μg/mL kanamycin. Cultures were grown to OD₆₀₀ 0.4–0.8 and induced with 1 mM IPTG for 2–4 h at 37 °C (220 RPM). Cells were harvested (30 min, 4000 g, 4 °C) after expression, the pellets were resuspended in 50 mM Tris.HCl pH 8, 50 mM NaCl, and centrifuged again to pellets. The washed pellets were either frozen or lysed directly.

For purification, the pellets were resuspended in a lysis buffer [50 mM NaH₂PO₄, 500 mM NaCl, 20 mM imidazole, 0.1% Nonidet P40 substitute (PanReac AppliChem; Barcelona, Spain), pH 8] and incubated at 70 °C for 20 min. Lysates were then sonicated using a probe sonicator (Soniprep 150, MSE UK Ltd;

London, UK) for 1 min at 20 Hz to fragment genomic DNA, and then clarified by centrifugation for 30–60 min at 20 000 *g* at 4 °C.

Lysates were purified by gravity flow chromatography using Super Nickel NTA Agarose Affinity Resin (Generon Ltd, UK) or 5 mL HisTrap FF columns (GE Life Sciences, UK) using a Bio-Rad NGC Discover Chromatography System (Bio-Rad Laboratories Ltd; Watford, UK). In both cases, the running buffer was 20 mM Tris.HCl pH 7.4, 500 mM NaCl, 20 mM imidazole and elution buffer 20 mM Tris.HCl pH 7.4, 500 mM NaCl, 350 mM imidazole for gravity flow (step elution) or 20 mM Tris.HCl pH 7.4, 500 mM NaCl, 500 mM imidazole for FPLC purification (gradient). The eluted DNA polymerases were concentrated and buffer exchanged using Amicon Ultra-4 mL Centrifugal Filters (EMD Millipore) with a 50 kD MWCO into 2x storage buffer (20 mM Tris.HCl pH 7.4, 200 mM KCl, 2 mM DTT, 0.2 mM EDTA). Finally, an equivalent volume of glycerol was added to make a final storage buffer of 10 mM Tris.HCl pH 7.4, 100 mM KCl, 10 mM DTT, 0.1 mM EDTA, 50% glycerol and polymerases were stored at -20 °C. Typically, only a single product was present by SDS-PAGE, and the polymerase concentration was determined by A280.

Mutiple site mutagenesis (Darwin Assembly). Plasmid pGDR11 encoding exo-KOD DNA polymerase (pGDR11_KOD) was the kind gift of J.C. Chaput (UC Irvine). To generate K.RT521K (KOD: V93E, D141A, E143A, A485L, I521L, E664K) we used Darwin Assembly as previously described.⁶ Briefly, one strand of pGDR11_KOD plasmid was digested through co-incubation with a nicking endonuclease (Nt.BspQI, NEB) and exonuclease III (NEB). Boundary oligonucleotides B_pGDR11_DA_F and pGDR11_DA_R_BsaI and inner mutagenic oligonucleotides (KOD_Leu75CTA, KOD_Leu463CTA, KOD_V93Q, KOD_A485L, KOD_I521L, KOD_E664K) were then phosphorylated using T4 polynucleotide kinase and annealed to the single stranded plasmid. A single step isothermal fill-in/ligation reaction was carried out by addition of 1x volume of 2x Darwin Assembly enzyme mix (0.05 U/μL Q5 High-Fidelity DNA polymerase (not hot start), 8 U/μL Taq DNA ligase, 2 mM NAD⁺, 0.4 mM each dNTP, 10% (w/v) PEG 8000, 2 mM DTT and 1x CutSmart buffer) for 15 min at 50 °C. The newly generated strand, containing all mutations, was isolated using Dynabeads MyOne Streptavidin C1 (4 μL beads, 15 min on a disc rotator at room temperature to capture), washed using 30 mM NaOH, PCR-amplified using KOD Xtreme DNA Polymerase (EMD Millipore) and cloned back into pGDR11 for expression.

Cloning and sequence analysis. For sequencing, cDNA was PCR amplified using primers outnest1(BsaI) and outnest2(BsaI) using MyTaq HS DNA Polymerase as described for tPhoNA reverse transcription and PCR. Products were gel-purified, digested with BsaI-HF (NEB) and cloned into a vector made from pSB1C3 using primers pSBvec_BsaI_fw and pSBvec_BsaI_rv. Cloning efficiency was

estimated at ~75000 : 1 (vector + insert : vector only ligations) and PCR screening suggested 95/95 colonies screened contained the correct insert.

Next, 96 colonies were sent for sequencing (48 each for T.521RT and K.RT521K reactions). The FASTA files were combined using the UNIX command “cat” (concatenate), reverse complemented using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), converted to a single line FASTA, trimmed and converted to upper case using PRINSEQ (<http://prinseq.sourceforge.net/>) and manually aligned using Microsoft Excel for Mac v16.11 or Jalview. Any sequences with no insert, any ambiguous base calls, lacking both expected mismatches or contiguous deletions of more than 4 nucleotides were omitted from analysis, resulting in 34 unique reads for T.RT521 (1938 nt of sequence data) and 37 unique reads for K.RT521K (2109 nt of sequencing data).

Large scale tPhoNA synthesis and purification. For sequencing experiments, tPhoNA was synthesized in reactions consisting 1x Thermopol buffer, 0.5 μ M primer b.F.out1tag1, 1 μ M template TempNv3, 50 μ M each PMTNpp, 0.25 μ M TgoT_EPFLH incubated for 1 min at 94 °C, 30 s at 50 °C, and 2 h at 65 °C. After synthesis, reactions were pooled and treated with 0.1U/ μ L exonuclease I (NEB) for 30 min at 37 °C.

In parallel with the tPhoNA synthesis, 4 μ L Dynabeads MyOne Streptavidin C1 beads (per pmol of biotinylated primer used) were blocked in 2x BWBST (20 mM Tris-HCl pH 7.4, 2 M NaCl, 0.2% v/v Tween-20, 2 mM EDTA). Blocked beads were placed on a magnetic stand, the supernatant removed, resuspended in 1 mL of 2x BWBST, added to the pooled synthesis reaction, and incubated on a disc rotator at room temperature overnight. Following the capture incubation, ~~beads were placed on a magnetic stand, allowed to form a pellet and the supernatant removed. After capture,~~ beads were isolated on a magnetic stand and washed twice in 200 μ L 37 °C 30 mM NaOH and once in 200 μ L EB-T (10 mM Tris-HCl pH 8.8, 0.1 mM EDTA, 0.01% Tween-20). ~~Following the final wash, the beads were~~ ~~washed twice in 500 μ L 30 mM NaOH, preheated to 37 °C, to remove unlabelled template and then~~ resuspended in 350 μ L 98% formamide and 10 mM EDTA, and incubated for 10 min at 99 °C to strip the nascent tPhoNA strand from the streptavidin beads. This was loaded directly onto denaturing (8 M urea) 10% acrylamide PAGE gels, with products visualised by SYBR Gold (Thermo Fisher Scientific) staining and eluted from the crushed gel slices overnight. Residual acrylamide was removed using Corning Costar Spin-X centrifuge tube filters (0.22 μ m cellulose acetate), then concentrated and washed using Vivaspin 500 centrifugal concentrators (Sartorius Stedim Biotech GmbH, Germany) with 10 kDa MWCO.

Exonuclease I sensitivity. DNA or tPhoNA TempN_MS2 were synthesized and purified as described for large-scale preparations ~~(see Materials and Methods)~~. Oligonucleotides at 0.1 μ M were incubated in 1x Thermopol buffer (NEB) with varying concentrations of exonuclease I (NEB) for 30 min at 37 °C (and 20 min 80 °C to inactivate the enzyme). Reactions were quenched with 2 volumes 98% formamide and 10

mM EDTA. Nucleic acids were separated by denaturing PAGE (8 M urea/20% acrylamide), stained with SYBR Gold (Thermo Fisher Scientific), imaged using a Typhoon FLA9500, and full-length oligonucleotides quantified using ImageJ.

Figure S3. Towards an efficient tPhoNA synthetase. (A) Rationale and impact of individual mutations introduced in the engineering of a tPhoNA synthetase. Estimate of the full-length tPhoNA synthesis refers to synthesis under forcing conditions, as shown in (B) and Figure S5. (B) Synthesis of tPhoNA from a TempN template by all engineered polymerases. Here, 10 μ L reactions in 1x Thermopol buffer containing 0.1 μ M primer Tag01F3, 0.2 μ M template TempN, 0.1 mM of each PMTpp, 1 mM $MnCl_2$, and 1 μ M polymerase were incubated for 1 min at 94 $^{\circ}C$, 5 min at 50 $^{\circ}C$, 4 h at 65 $^{\circ}C$ – forcing reaction conditions with long extensions, high polymerase and manganese

concentrations. The reaction conditions were chosen to enable engineering of intermediates to synthesize a full-length TempN but were too permissive for TgoT_EPFLH. Consequently, the observed fraction of full-length product is an underestimate of the enzyme activity. (C) Mutated residues in TgoT_EPFLH mapped onto a ternary complex of the homologous KOD DNA polymerase (PDB: 5OMF).⁷

Figure S4. Residual DNA polymerase activity of engineered tPhoNA synthetases. Mutations introduced to improve tPhoNA synthesis had in most cases a detrimental effect on the DNA polymerase activity of the engineered enzymes as seen by the reduction of full-length DNA synthesis at low dNTP concentrations. Reactions (10 μ L) were carried out with 0.1 μ M primer tag01F3, 0.1 μ M template TempN, 0.1 μ M of each polymerase, and were incubated with various concentrations of dNTPs for 1 min at 94 $^{\circ}$ C, 1 min at 50 $^{\circ}$ C, and 1 min at 65 $^{\circ}$ C. The introduction of L403P led to a drop in polymerase activity, which is expected to be the result of its proximity to the conserved A-motif (D404 – P410) and specifically the catalytic aspartate (D404). E664 has been previously described as an enhancer of XNA synthetase activity,⁸⁻¹⁰ and while E664H increased tPhoNA activity (Figure 1 and Figure S5), it had a significant impact on DNA synthesis. Finally, the introduction of L408F profoundly disrupted DNA synthesis, while significantly boosting tPhoNA synthesis. L408 is completely conserved in B-family DNA polymerases (A motif consensus DxxLYP).¹¹ Mutations H147E¹² and I521L¹³ have previously been described as enhancers of polymerase synthetase activity and did not affect or even rescue DNA synthesis, respectively.

Figure S5. Impact of the reaction conditions on the engineered mutants. Under forcing conditions (left panel: 1 μM polymerase, 1 mM MnCl_2 , 4 h extension), all engineered polymerases generated some full length tPhoNA (57 nt, mixed sequence). Decreasing Mn^{2+} and extension time (middle panel: 1 μM polymerase, 0.6 mM MnCl_2 , 1 h extension) reduced TgoT_P and TgoT_EP activity close to the parental polymerase (TgoT) activity. Clear gains could still be seen when the 664H and 521L mutations were added. Full-length product formation was clearly observed for TgoT_EPFLH (harboring the additional 408F mutation). The reduction in Mn^{2+} from 1 to 0.6 mM actually resulted in improved tPhoNA synthesis by TgoT_EPFLH. TgoT_EPFLH remained capable of full length synthesis under mild conditions, while the other enzymes failed to extend beyond ~15 incorporations (left panel: 0.2 μM polymerase, MnCl_2 absent, 1 h extension).

Figure S6. DNA and RNA-primed tPhoNA reverse transcription reactions. Because of the potential difference in hybridization efficiency between DNA, RNA, and tPhoNA sequences, we compared the impact of primer chemistry in RT reactions for both RTs available. Under the conditions tested [10 μ L reactions in 1x Thermopol buffer, containing 0.2 μ M DNA or RNA IR700-test7 (green), 0.1 μ M FAM-labelled tPhoNA TempN2 (red), 0.1 mM each dNTP, 30 nM polymerase], RT efficiency was high for both primer chemistries and strongly template-dependent, with full-length RT products observed (green). Reactions were made up without polymerase, annealed by heating to 95 $^{\circ}$ C for 5 min and cooling to 4 $^{\circ}$ C at 0.1 $^{\circ}$ C/s; the polymerase was then added and the mixture was incubated for 5 min at 50 $^{\circ}$ C and 1 h at 65 $^{\circ}$ C. Owing to the use of a 2-fold excess of primer over template, complete primer depletion was not expected.

Figure S7. Template preferences of TgoT_EPFLH. Given the reproducible stalling patterns observed in the tPhoNA synthesis of different mixed sequences, we investigated whether substrate incorporation of a particular PMTNpp or particular template sequences were interfering in tPhoNA synthesis. **(A)** Synthesis under forcing conditions in the absence of template or using templates designed to test multiple PMTApp (TempA) or PMTTpp (TempT) incorporations suggest a strong bias against multiple PMTTpp incorporations even under conditions where non-templated synthesis was detectable. **(B)** Systematic removal of di-PMTTpp incorporation sites between TempN, TempNv2, and TempNv3 clearly show a reduction of polymerase stalling during tPhoNA synthesis. **(C)** Template comparisons were also carried out as time courses to ensure that the differences observed were not the result of a single time point. Optimized reactions consisted of 0.1 μ M tag01F3, 0.2 μ M template, 50 μ M each PMT, 0.1 mM $MnCl_2$, and 0.2 μ M TgoT_EPFLH, and were incubated for 1 min at 94 $^{\circ}C$, 5 min at 50 $^{\circ}C$, and up to 80 min at 65 $^{\circ}C$. **(D)** Expected sequence in tPhoNA generated by each template. The DNA primer used in synthesis is shown in blue. Yellow boxes highlight TT doublet incorporations and red characters show base changes introduced to improve synthesis.

A.

Figure S8. tPhoNA exonuclease resistance. tPhoNA has previously been observed to be resistant to the 3'-5' exonuclease activity of snake venom phosphodiesterase.¹⁴ Here, we incubated identical sequences of DNA or tPhoNA with increasing quantities of exonuclease I (NEB) and observed a pronounced resistance.

Figure S9. Aggregate fidelity of tPhoNA synthesis and reverse transcription. The aggregate fidelity of information transfer from DNA to tPhoNA to DNA was assayed by synthesizing TempNv3(test7) in tPhoNA using TgoT_EPFLH and reverse transcribing it back to DNA using T.RT521 or K.RT521K. RT products were amplified, cloned, and a number of individual transformants sequenced. As previously described,¹⁵ sequences used in tPhoNA synthesis and RT included markers to ensure that sequenced information could only be generated by a combination of both processes. Location of the detected errors based on the RT used: T.RT521 (**A**) and K.RT521K (**B**). Insertions and deletions are shown as triangles and circles, respectively, above the expected sense sequence (tPhoNA sequence). Misincorporations are shown in orange below the sequence. It is not possible to unambiguously determine whether errors observed are a result of synthesis or RT misincorporations. A total of 34 (1938 nt of sequence data) or 37 (2109 nt of data) were analyzed for reactions carried out with T.RT521 (**C**) and K.RT521K (**D**), respectively, enabling the incorporation ratios in parts per thousand (‰) to be determined. The results suggest that early in both tPhoNA synthesis and RT reactions, polymerases tend to be more permissive to insertions and deletions. In addition, both systems suggest that A→G and T→G mutations are the most frequent either as a result of mismatch bypass or from misincorporations during synthesis (e.g. pmtG:dT that would account for A→G transitions) or RT (e.g. dC:pmtT that would account for T→G transversion).

Figure S10. Attenuated total reflectance (ATR) Fourier transform infrared (FTIR) spectroscopy of DNA and tPhoNA enzymatically synthesized oligonucleotides. This figure is representative of the ATR-FTIR spectra recorded on dried samples of enzymatically synthesized DNA and tPhoNA in the 1300-900 cm^{-1} range, where the deoxyribosephosphodiester backbone moiety of DNA displays well characterized absorption bands.^{16,17}

The frequency of the (PO_2^-) antisymmetric stretch (ν_{as}) of DNA has been reported to vary sensibly depending on the hydration level from 1225 (fully hydrated) to 1240 cm^{-1} (dry).^{16,18,19} (A) The two oligomers are single stranded with an identical sequence of 38 nucleotides. 19 nucleotides are DNA and the remaining 19 are either DNA (black) or tPho-NA (red). Hence, the tPhoNA sample is a chimera composed of 50% DNA and 50% tPho-NA. In the DNA (black), the (PO_2^-) ν_{as} is centered at 1225 cm^{-1} , although underlying contributions of higher frequency bands are evident. In the DNA-tPhoNA chimera spectrum (red), two bands can clearly be seen in this region with maxima at 1240 and 1208 cm^{-1} . (B) Subtracting the DNA component [defined a half the DNA spectrum in (A)] from the DNA-tPhoNA chimera spectrum generates a ‘tPhoNA-only’ IR spectrum (red), where only one band remains at 1205 cm^{-1} when compared to a DNA spectrum with an identical subtraction (black). The observed frequency change of the (PO_2^-) ν_{as} from 1225-1240 to 1205 cm^{-1} appeared as a consistent marker for the bond change from P-O-C in DNA to P-C-O in tPhoNA. Other differences in backbone IR absorptions were evident in the 1090-1030 cm^{-1} region with a shift of the (PO_2^-) symmetric stretching mode (ν_{s}) at 1085 cm^{-1} and (C-O) ribose stretch at 1062 cm^{-1} in DNA to, possibly, 1074 and 1051 cm^{-1} in tPho-NA.^{16,18,19} Two other remarkable differences when comparing the oligomers spectra were

the DNA band at 1009 cm^{-1} (assigned to a ribose ring mode) and the characteristic ribose-phosphodiester skeletal motion mode of DNA at 965 cm^{-1} (black traces in either A or B).^{16,18,19} They might have equivalents in the smaller 981 and 954 cm^{-1} bands of the tPhoNA spectrum (B, red) but stood out as distinct markers differentiating the two oligomer types. (C) tPhoNA *minus* DNA [red trace minus black trace in (B)], generating a calculated difference spectrum (blue).

In Vivo propagation assay of DNA-PhoNA-DNA chimeras

Table S3. Mass Analysis of Oligomer Sequences used for *in Vivo* Studies.^a

ON	Sequence	Calc.	Obs.
ON 4	5'-P-CTA GCG CCG <u>TGC</u> CAT GCA-3'	5537.9	5538.1
ON 15	5'-P-CTA GCG CCG <u>TIC</u> CAT GCA-3'	5522.9	5523.2
ON 16	5'-P-CTA GCG CCG <u>TIT</u> CAT GCA-3'	5537.9	5537.9
ON 17	5'-P-CTA GCG CCG <u>tGC</u> CAT GCA-3'	5551.9	5552.2
ON 18	5'-P-CTA GCG CCG <u>tiC</u> CAT GCA-3'	5550.9	5550.9
ON 19	5'-P-CTA GCG CCG <u>titt</u> CAT GCA-3'	5579.9	5579.9

^aModified units are bolded and underlined. tPhoNA and dPhoNA nucleotides are shown as capital blue and lowercase magenta letters, respectively.

Table S4. *In Vivo* Propagation of Mosaic DNA-tPhoNA-DNA Oligonucleotides.

Oligonucleotides ^a		N° of tPhoNA residues	<i>thyA</i> ⁺ ^b	<i>bla</i> ⁺ ^c	<i>thyA</i> ⁺ / <i>bla</i> ⁺ ^d
Wt sequence	5'-P-CTA GCG CCG TGC CAT GCA	0	22200	170000	0.13
Cys146 deleted	5'-P-CTA GCG CCG ΔΔΔ CAT GCA	0	0	350000	0
ON 4	5'-P-CTA GCG CCG <u>TGC</u> CAT GCA	1	21900	420000	0.052
ON 15	5'-P-CTA GCG CCG <u>TIC</u> CAT GCA	2	20300	860000	0.024
ON 16	5'-P-CTA GCG CCG <u>TIT</u> CATGCA	3	3200	380000	0.008

^aModified units are bolded and underlined (blue).

^b*ThyA*⁺ refers to the number of thymidine-prototrophic colonies.

^c*Bla*⁺ refers to the total number of transformant colonies (*ThyA*⁺ and *ThyA*⁻).

^d*ThyA*⁺/*Bla*⁺ refers to the ratio of thymidine-prototrophic colonies against the total number of transformant colonies.

Table S5. *In Vivo* Propagation of Mosaic DNA-dPhoNA-DNA Oligonucleotides.

Oligonucleotides ^a		N° of dPhoNA residues	<i>thyA</i> ⁺ ^b	<i>bla</i> ⁺ ^c	<i>thyA</i> ⁺ / <i>bla</i> ⁺ ^d
Wt sequence	5'-P-CTA GCG CCG TGC CAT GCA	0	11800	90000	0.13
Cys146 deleted	5'-P-CTA GCG CCG ΔΔΔ CAT GCA	0	0	220000	0
ON 17	5'-P-CTA GCG CCG <u>tGC</u> CAT GCA	1	11100	220000	0.05
ON 18	5'-P-CTA GCG CCG <u>tiC</u> CAT GCA	2	900	520000	0.0017
ON 19	5'-P-CTA GCG CCG <u>titt</u> CATGCA	3	100	170000	0.0006

^aModified units are bolded and underlined (magenta).

^b*ThyA*⁺ refers to the number of thymidine-prototrophic colonies.

^c*Bla*⁺ refers to the total number of transformant colonies (*ThyA*⁺ and *ThyA*⁻).

^d*ThyA*⁺/*Bla*⁺ refers to the ratio of thymidine-prototrophic colonies against the total number of transformant colonies.

Table S6. Oligonucleotides Used for Establishing tPhoNA as a Xenobiotic Genetic Material.

Oligonucleotide	Sequence (5'-3')	Function
TempN	TGGTCCAGCATCGTGAGATCGATTA CCGAACAGCACTACGTGGCTAAGTG CTTATCTCCTAGCTTAAACGGATCC G	Primer extension template
TempNv3	TGGTCCAGCATCGTGAGATCGATTA CCGATCAGCACTACGTGGCTAcGTG CTTATCTCCTAGCTTAAACGGATCC G	As TempN but TT incorporations removed (A70T, A87C in template)
TempN2	CCCTAGTTCTTCCTCTTCCCACGTG GCTAcGTGCTTATCTCCTAGCTTAA ACGGATCCG	Short version of TempN with TT doublet removed (A30C mutation in template), and with test7 binding site added for RT screening.
TempNv2 (test7)	AGCAGCACAGAGGTCAGATGaCCTA GTTCTTCCTCTTCCCTGGTCCAGCA TCGTGAGATCGATTACCGATCAGCA CTACGTGGCTAcGTGCTTATCTCCT AGCTTAAACGGAGCCGTGGCAAACG CTAATAAGGGG	As TempN but first TT doublet incorporation removed (A87C in template) and test7 binding site added for RT.
TempNv3 (test7)	aCCTAGTTCTTCCTCTTCCCTGGTC CAGCATCGTGAGATCGATTACCGAT CAGCACTACGTGGCTAcGTGCTTAT CTCCTAGCTTAAACGGAGCCG	As TempNv3 and with test7 binding site added for RT.
TempN_MS2nick	CGTGGCTACGTGCTTATCTCCgaag agcAACGGATCCG-biotin	Primer extension template to generate tPhoNA for IR, with 3'biotin for purification.
tagBspQI	CGGATCCGTTgctcttcGG	Primer for TempN MS2.
Tag1F3	Cy5-CGGATCCGTTTAAAGCTAGG	Tag1 primer with 5'Cy5.
IR700_test7	CCCTAGTTCTTCCTCTTCCC	Test7 primer with 5'IR700.
b.F.out1_tag1	Biotin(TEG)-T(FITC)- CCCTTATTAGCGTTTGCCACGGATC CGTTTAAAGCTAGG	Tag1 primer encoding outnest1 outnest PCR priming site. Dual biotin and FITC modified.
Outnest2_test7	AGCAGCACAGAGGTCAGATGaCCTA GTTCTTCCTCTTCCC	Test7 primer encoding outnest2 PCR priming site.
Outnest1	CCCCTTATTAGCGTTTGCCA	PCR primer
Outnest2	AGCAGCACAGAGGTCAGATG	PCR primer
Outnest1_BsaI	gagtcaggtctcCAGGACCCCTTAT TAGCGTTTGCCA	PCR primer pair with BsaI sites for cloning into pSB1C3.
Outnest2_BsaI	gagtcaggtctcTATTCAGCAGCAC AGAGGTCAGATG	
pSBvec_BsaI_fw	GAGTCAGGTCTCgaattccagaaa tcattccttagcg	PCR primer pair to make pSB1C3-based vector for cloning outnest1/2_BsaI-amplified PCR products.
pSBvec_BsaI_rv	GAGTCAGGTCTCttcctcgctcact gactcgctg	
B.pGDR11_DA_F_BsaI	Biotin(TEG)- CCCCTTATTAGCGTTTGCCAGGTCT CACTATGAGAGGATCTCACCATCAC	Boundary primers for Darwin Assembly, with 5' biotin on F primer for purification and 3' inverted dT on R primer to prevent extension / degradation.
pGDR11_DA_F_BsaI	CCTGCAGCCAAGCTTAATT'AGCT, GGAGACCCATCTGACCTCTGTGCTG CT(inverted-dT)	
KOD_Leu75CTA	GAAGAAGTTCCTaGGGAGACCAG	Inner mutagenic

KOD_Leu463CTA	CGAGCCTGCTaGGAGACCTCC	oligonucleotides for Darwin Assembly. KOD_Leu75CTA and KOD_Leu463CTA recode leucine codons to remove BsaI sites which may interfere with cloning. KOD_93Q is miscoded (actually V93E). KOD_A485L, KOD_521L and KOD_664K encode mutations as indicated.
KOD_93Q	CCGCAGGACGaaCCAGCGATAAG	
KOD_485L	CAGGCAGAGGttgATCAAGATCC	
KOD_521L	GAAGGGAGTACttAACGATGACCAT C	
KOD_664K	GGTGATCCACaaaCAGATAACGAG	
pGDR11_DA_vec_F BsaI	GAGTCAGGTCTCTAGCTGAGCTTGG AC	iPCR primers to make a vector for cloning of the mutated strand after Darwin Assembly.
pGDR11_DA_vec_R BsaI	GAGTCAGGTCTCCATAGTTAATTTTC TCCTCTTTAATGAATTC	

NMR spectra of intermediates and final compounds

References

- (1) Liu, C.; Dumbre, S. G.; Pannecouque, C.; Huang, C. S.; Ptak, R. G.; Murray, M. G.; De Jonghe, S.; Herdewijn, P. *J. Med. Chem.* **2016**, *59*, 9513-9531.
- (2) Kostov, O.; Pav, O.; Budesinsky, M.; Liboska, R.; Simak, O.; Petrova, M.; Novak, P.; Rosenberg, I. *Org. Lett.* **2016**, *18*, 2704-2707.
- (3) Winqvist, A.; Strömberg, R. *Eur. J. Org. Chem.* **2008**, *2008*, 1705-1714.
- (4) Stemmer, W. P. C.; Morris, S. K. *Biotechniques* **1992**, *13*, 214-220.
- (5) Tang, L. X.; Gao, H.; Zhu, X. C.; Wang, X.; Zhou, M.; Jiang, R. X. *Biotechniques* **2012**, *52*, 149-158.
- (6) Cozens, C.; Pinheiro, V. B. *Nucleic Acids Res.* **2018**, gky067-gky067.
- (7) Kropp, H. M.; Betz, K.; Wirth, J.; Diederichs, K.; Marx, A. *PLoS One* **2017**, *12*, 1-20.
- (8) Cozens, C.; Mutschler, H.; Nelson, G. M.; Houlihan, G.; Taylor, A. I.; Holliger, P. *Angew. Chem. Int. Ed.* **2015**, *54*, 15570-15573.
- (9) Cozens, C.; Pinheiro, V. B.; Vaisman, A.; Woodgate, R.; Holliger, P. *Proc. Nat. Acad. Sci. U S A* **2012**, *109*, 8067-8072.
- (10) Larsen, A. C.; Dunn, M. R.; Hatch, A.; Sau, S. P.; Youngbull, C.; Chaput, J. C. *Nat. Commun.* **2016**, *7*, 11235.
- (11) Ito, J.; Braithwaite, D. K. *Nucleic Acids Res.* **1991**, *19*, 4045-4057.
- (12) Kuroita, T.; Matsumura, H.; Yokota, N.; Kitabayashi, M.; Hashimoto, H.; Inoue, T.; Imanaka, T.; Kai, Y. *J. Mol. Biol.* **2005**, *351*, 291-298.
- (13) Taylor, A. I.; Pinheiro, V. B.; Smola, M. J.; Morgunov, A. S.; Peak-Chew, S.; Cozens, C.; Weeks, K. M.; Herdewijn, P.; Holliger, P. *Nature* **2015**, *518*, 427-430.
- (14) Renders, M.; Emmerechts, G.; Rozenski, J.; Krecmerova, M.; Holy, A.; Herdewijn, P. *Angew. Chem. Int. Ed.* **2007**, *46*, 2501-2504.
- (15) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S. Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. *Science* **2012**, *336*, 341-344.
- (16) Guchhait, B.; Liu, Y. L.; Siebert, T.; Elsaesser, T. *Structur. Dyn.* **2016**, *3*.
- (17) Siebert, T.; Guchhait, B.; Liu, Y. L.; Costard, R.; Elsaesser, T. *J. Phys. Chem. B* **2015**, *119*, 9670-9677.
- (18) Mello, M. L. S.; Vidal, B. C. *PLoS One* **2012**, *7*.
- (19) Wood, B. R. *Chem. Soc. Rev.* **2016**, *45*, 1980-1998.