Sulfonate derived phosphoramidates as active intermediates in the enzymatic primer-extension of DNA

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID:	OB-ART-01-2015-000157.R1
Article Type:	Paper
Date Submitted by the Author:	19-Feb-2015

Complete List of Authors:	De, Swarup; Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Groaz, Elisabetta; Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Margamuljana, Lia; Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Abramov, Mikhail; Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Marlière, Philippe; ISSB, Génopole genavenir6, Equipe Xénome, Herdewijn, Piet; Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven,

Journal Name

RSCPublishing

ARTICLE

Sulfonate derived phosphoramidates as active intermediates in the enzymatic primer-extension of DNA

Cite this: DOI: 10.1039/xoxxoooooox

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

S. De [a], E. Groaz [a], L. Margamuljana [a], M. Abramov [a], P. Marlière [b] and P. Herdewijn*[a,b]

Novel unnatural 5′-phosphoramidate nucleosides, capable of being processed as substrates by DNA polymerases for multiple nucleotide incorporations, have been designed. The mimics feature metabolites such as taurine and a broad range of aliphatic sulfonates coupled through a P-N bond to the 5′-phosphate position of deoxynucleotides, to allow binding interactions in the enzyme active site. The utility of all of the analogues as pyrophosphate mimics was demonstrated for the chain elongation of DNA, using both thermophilic and mesophilic microbial polymerases.

INTRODUCTION

In biological systems, phosphate esters and anhydrides are amongst the prevailing molecular fragments being released by metabolic transformations requiring good leaving groups, such as elimination and nucleophilic substitution reactions.[1] For instance, DNA polymerization, the polynucleotide synthesis reaction, is an enzymatically-catalysed process, in which a free 3′-hydroxyl nucleophilic group of a *primer*, base-paired to a DNA template, attacks the α-phosphate group of a nucleoside triphosphate with displacement and ensuing hydrolysis of pyrophosphate.[2] This biochemical mechanism might have emerged during evolution as a result of either functional selection or synthetic accessibility.

Our group has an ongoing program devoted to the diversification of nucleic acid metabolism by means of the systematic variation of both structural and enzymatic elements inherent to the genetic propagation of natural biopolymers.[3] One important goal is the identification of novel components that, although not favored by natural selection, may nevertheless exist in nature and follow biochemical pathways orthogonal to natural ones. This may provide an opportunity to implement and propagate *in vivo* a third type of information carrier molecule (XNA), whilst succeeding in the establishment of both genetic and energetic enclaves.

In the search for non-phosphate synthetic alternatives to pyrophosphate we have already demonstrated the leaving group properties of L-aspartic acid and other unnatural amino acids, in the DNA polymerase-catalyzed synthesis of DNA.[4] This selection of substrates was partly motivated by the structural and electronic properties necessary to consent accommodation and recognition within the enzyme catalytic site, but most importantly by their significance as non-toxic intermediates of metabolic steps in living cells.

In an extension of the substrate scope of the DNA polymerization reaction, we became interested in 5'-phosphoramidate deoxyribonucleoside analogues conserving the anionic residues for key binding interactions with polymerases in the form of aliphatic sulfonate-based functionalities.[5] The design of such unnatural mimics of triphosphates is defined by the abundance of simple organic molecules such as taurine (2-aminoethanesulfonic acid) and L-cysteic acid in natural environments and nutrients and their contribution as sources of mineral sulfur to microbial growth.[6] Organic sulfonates are ionized at physiological pH with very low pKa values, thus their cellular uptake cannot occur by passive diffusion, but are rather

taken up across bacterial membranes by complex transport systems and actively degraded by the action of oxygenolytic enzymes over desulfonation pathways. Typically, this process releases sulfite as degradative intermediate, which then enters the sulphur assimilation cycle. This is one of the essential conditions that a good leaving group candidate must fulfil in order for nucleotide incorporation to be irreversible.

It is well documented in the literature that under conditions of inorganic sulfate and cysteine starvation, various gene clusters are expressed in *E. coli*, allowing the alternative utilization of aliphatic sulfonates.[7] Two operons, *tau*ABCD[8] and ssuEADCB,[9] have been identified which encode for distinct ABC-type permeases, thus promoting uptake of taurine and other alkanesulfonates respectively, while the TauD gene expresses for the protein alpha-ketoglutarate-dependent dioxygenase responsible for the liberation of sulphite from taurine.[10] Therefore, given our interest in developing an uptake system for nucleotides in bacterial cells, the synthesis of these molecules was also deemed important in view of their potential use as substrates for cellular delivery within a sulfur-based nutritional selection system.

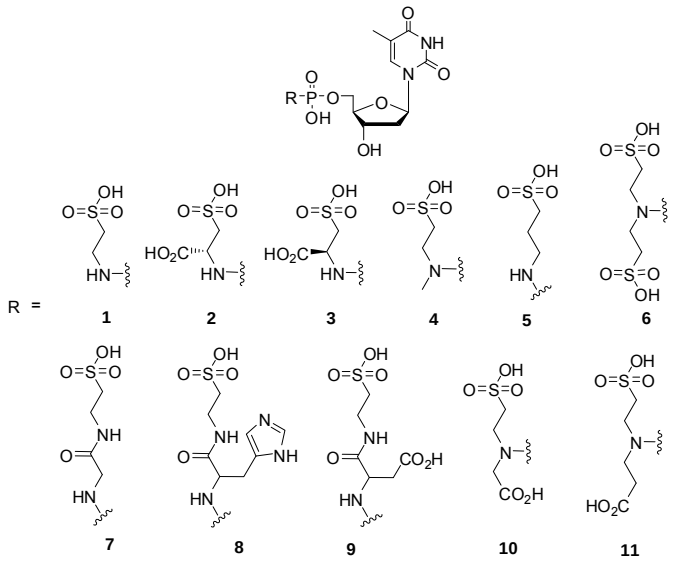


Figure 1. Overview of sulfonate-containing phosphoramidate analogues of thymidine.

Herein, we report novel straightforward syntheses of taurine, L-cysteic acid and a wide range of relevant sulfono-modified phosphoramidate nucleotide analogues (examples of TMP derivatives are shown in Figure 1), as well as an account of the efficiency with which various mesophilic and thermophilic microbial polymerases accept them as substrates.

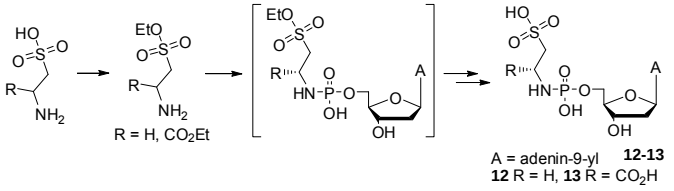
Along with the biochemical relevance, the knowledge of the extent to which molecular diversity can be manipulated but still accepted within the polymerase active site will provide fundamental parameters in the selection and design of new synthetic genetic variants.

RESULTS AND DISCUSSION

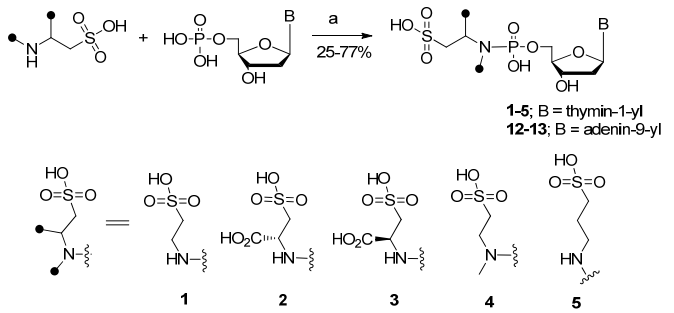
In an early communication, we established a synthetic approach to sulfonate-containing phosphoramidates Tau-dAMP **12** and L-Cys-dAMP **13**, which centred upon an EDAC-mediated coupling reaction between deoxyadenosine monophosphate (dAMP) and the relevant aminoalkylsulfonate motifs.[11] However, protection of both sulfonic and carboxylic acids as ethyl ester groups was necessary to prevent formation of undesired side-products (Scheme 1).

In addition, it was anticipated that in order to accomplish the synthesis of higher functionalized substrates, multiple selective protection and deprotection steps of acidic groups would be cumbersome, and most likely produce the novel phosphoramidate products in inadequate yields.

Surprisingly, the replacement in our existing method of EDAC by DCC in ^{t}BuOH-H_2O as coupling agent proved to be an alternative satisfactory procedure, even so in the presence of unprotected sulfonic acid functionalities. The resulting adenosine 5'-sulfono-phosphoramidates **12** and **13** were obtained in a single step with increased yields, provided that an excess of Et_3N (5 eq.) was added to the reaction mixture in order to mask the charges of the sulfonic acid groups. Similarly, sulfono-phosphoramidates **1**, **4** and **5** were isolated in good yields (72-77%) from thymidine monophosphate (TMP) and either taurine, *N*-methyl taurine or homo-taurine as starting materials (Scheme 2). This methodology was also applied to the preparation of L-Cys-dTMP **2** and D-Cys-dTMP **3** in the presence of both free sulfonic and carboxylic acids, albeit with reduced coupling efficiencies (25-30% yields) (Scheme 2).

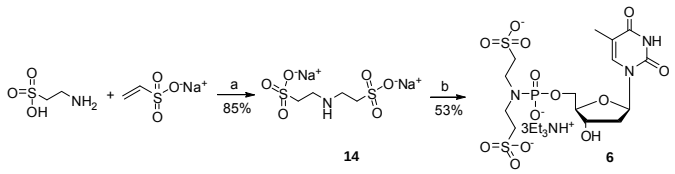


Scheme 1. Early route towards sulfono-phosphoramidates *via* ester protection of sulfonic acid.



Scheme 2. Synthesis of sulfono-phosphoramidate analogues of thimydine **1**-**5** and adenosine **12** and **13**. (a) TMP or dAMP (Na or Et_3N salt), DCC, Et_3N, ^{t}BuOH: H_2O (4:1), 90 °C, 4-6 h.

The same DCC-mediated protocol was also valuable to generate bis-sulfonate phosphoramidate **6** after coupling between the triethylammonium salt of TMP and the secondary amine *N*,*N*-di-2-aminoethanesulfonate **14**, afforded in turn from the Michael addition of taurine to sodium vinyl sulfonate (Scheme 3).



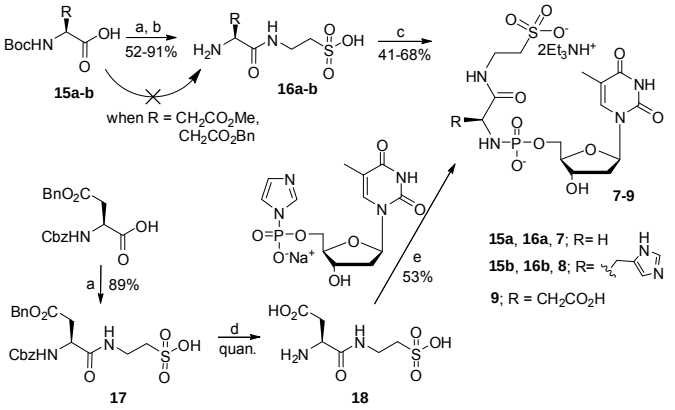
Scheme 3. Synthesis of thymidine phosphoramidate analogue **6**. (a) $NaHCO_3$, H_2O, reflux, 72 h; (b) TMP, DCC, Et_3N, ^tBuOH: H_2O (4:1), 90 °C, 5.5 h.

In order to provide further evidence to the geometry accepted by the active site,[12] we decided to prepare dipeptide-like phosphoramidate thymidine analogues bearing various amino acids interconnecting the taurine unit with the amidate bond, as shown in Scheme 4. Peptidic bonds can be degraded *in vivo* by hydrolase enzymes releasing amino acid and taurine.

The first pyrophosphate mimic of this series containing the neutral residue glycine, compound **16a**, was accessible through standard coupling conditions in the presence of DCC/NHS in a mixture of dioxane and water, followed by Boc-removal using IR-120B (H^+) resin.[13] When taurine was reacted with a non-neutral amino acid, the same two-steps sequence was found to be less rewarding, producing compound H-L-His-Tau-OH **16b** in modest yield. All attempts to induce coupling of taurine to aspartic acid, however, met with failure producing only undesired side products (uncharacterized), with both methyl and benzyl featuring as protecting groups of the β-carboxyl functionality. Thus, compound H-L-Asp-Tau-OH **18**, was alternatively obtained in two-steps starting from protected Z-L-Asp (OBn)-OH and taurine, followed by hydrogenolysis.
The final modified nucleosides were prepared under standard DCC coupling conditions, as outlined in Scheme 4. As the reaction between TMP and **18** led to the isolation of phosphoramidate **9** in low yield, which we believe to be due to the presence of a free carboxylic acid group in compound **18**, we opted for a different route for this latter substrate. The 5'-phosphate group was thus pre-activated by conversion into a better phosphoraimidazolide leaving group, which allows for enhanced nucleophilic attack of amines even under mild conditions.

N-Acetic acid-2-aminoethane sulfonic acid and *N*-propionic acid-2-aminoethane sulfonic can mimic respectively iminodiacetic acid (IDA) and iminodipropionic acid (IDP), previously reported by our group as very good pyrophosphate mimics for DNA polymerization.[14, 15]

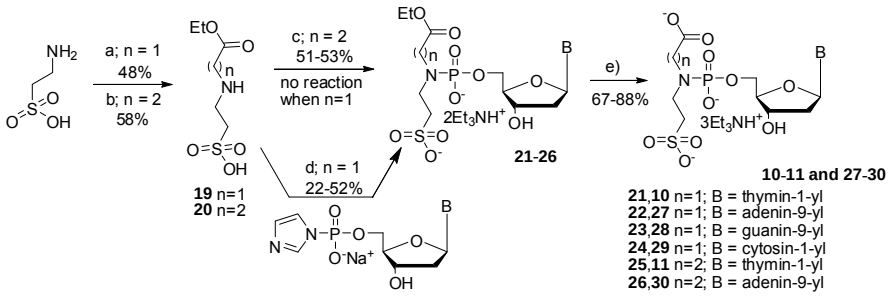
Therefore, a series of 2'-deoxynucleoside-5'-*O*-[*N*-(alkyl acid)-2-aminoethane sulfonic acid] phosphoramidates was prepared



Scheme 4. Synthesis of taurine-amino acid phosphoramidate nucleoside analogues **7**-**9**. (a) (i) NHS, DCC, THF, 0 °C to r.t., 2 h, (ii) Taurine Na-Salt, Dioxane-H_2O (1:1), r.t., 14 h; (b) IR-120B resin (H^+ form); (c) TMP-Na salt, DCC, Et_3N, ^tBuOH:H_2O (4:1), 90 °C, 4.5 h; (d) 10% Pd/C, H_2, EtOH-H_2O (5:1), r.t., 16 h; (e) Et_3N, DMF, 35 °C, 168 h.

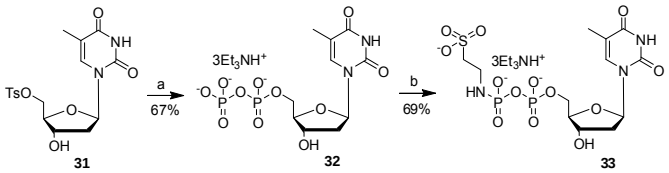
according to a three-step synthetic methodology (Scheme 5), starting from taurine without the need for protection of the sulfonic acid group. The formation of *N*-(ethyl acetate)-2-aminoethane sulfonic acid **19**, following the nucleophilic diplacement of the bromo atom in ethylbromoacetate by taurine, occurred only upon heating, whilst *N*-(ethyl acetate)-2-aminoethane sulfonic acid **20** resulted from the Michael addition of taurine to ethyl acrylate at ambient temperature over seven days.

As it might be speculated that DCC-mediated phosphoramidate coupling of secondary amines would prove less effective than that of primary amines, we thought it necessary to employ protected carboxylic ester derivatives in order to improve the reaction outcome in terms of isolated yields. Accordingly, reactions of TMP or dAMP thiethylammonium salts with secondary amine **20** were both reasonably efficient to yield respectively compound **25** and dAMP-analogue **26**. However, to our surprise, the secondary amine *N*-(ethyl acetate)-2-aminoethane sulfonic acid **19** failed to react under this conditions. Thus, for the synthesis of compounds **21**-**24** bearing all four nucleobases, we adopted the phosphoraimidazolide-based protocol described above, which delivered the T, A and C analogues **21**, **22**, and **24** in moderate yields (38-52%) and G analogue **23** in modest yield (22%). Ultimately, the deprotection of the ethyl ester group was carried out by treatment with a dilute solution of sodium hydroxide (0.4 M), and subsequent HPLC purification furnished the pure unprotected analogues **10**-**11** and **27**-**30** for incorporation studies.

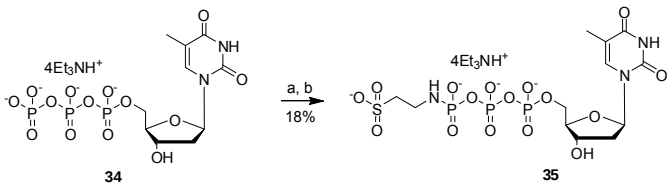


Scheme 5. Synthesis of sulfono-phosphoramidate nucleoside analogues **10**-**11** and **27**-**30**. (a) Ethylbromo acetate, $NaHCO_3$, Dioxane/H_2O, 70 °C, 24 h; (b) Ethyl acrylate, $NaHCO_3$, EtOH-H_2O (1:1), r.t., 168 h; (c) dTMP-Et_3N salt, DCC, Et_3N, ^tBuOH-H_2O (4:1), 90 °C, 6 h; (d) Nucleoside-Imidazolide Na-Salt, Et_3N, DMF, 35 °C, 168 h; (e) 0.4 M NaOH in MeOH:H_2O (4:1 v/v), r.t., 3 h.

Next, Tau-dTDP β-phosphoramidate **33** was prepared according to the DCC-mediated phosphoramidate coupling between taurine and the triethylammonium salt of thymidine diphosphate **32**. Compound **32** was derived in turn from 5'-*O*-tosyl derivative **31**,[16] upon treatment with tris(tetra-n-butylammonium) hydrogen pyrophosphate, acting as a nucleophile to displace the tosyl group as reported by Davisson *et. al*.[17]



Scheme 6. Synthesis of taurine diphosphoramidate analogue **33** of thymidine. (a) Tris(tetra-n-butylammonium) hydrogen pyrophosphate, CH_3CN, r.t., 24 h; (b) Taurine, DCC, Et_3N, ^tBuOH:H2O (4:1), 85 °C, 4.5 h.



Scheme 7. Synthesis of taurine triphosphoramidate analogue **35** of thymidine. (a) DCC, DMF, r.t., 3.5 h; (b) Taurine, MeOH/DMF, r.t., 6 h.

Finally, the synthesis of Tau-dTTP γ-phosphoramidate **35** was accomplished through initial conversion of dTTP-Et_3N salt **34** to thymidine-5'-trimetaphosphates by dehydration with DCC,[18] followed by ring opening of thymidine-5'-trimetaphosphate with an excess of taurine.

Chain elongation experiments

The individual competency of chemically-modified sulfono-phosphoramidate nucleosides **1-11**, **27-30**, **32-33** and **35** to perform as active intermediates in the DNA polymerization reaction was assessed by comparing the activities of different bacterial DNA polymerases, including Therminator, Vent (exo-) and the Klenow fragment (exo-) of *E. coli* DNA Polymerase I. The last two mutants lacking 3´ → 5´ proofreading activity were included in the study as preferred for affording high yield primer extension reactions in the presence of poor substrates, by suppressing simultaneous synthesis and editing.

Multiple nucleotides incorporation efficiencies in primer extension were estimated by means of a polyacrylamide gel-based template dependent assay, using a 5'-radiolabelled γ-^{33}P DNA primer annealed in turn to **T1-5** templates (Table 1). For each enzyme, reactions were also carried out with natural deoxynucleoside triphosphates (dNTP) as positive controls.

Table 1. Overview of primer and templates sequences used in the incorporation experiments.

Primer	P1: 5' /5Cy5/CAGGAAACAGCTATGAC 3'
Template	T1: 3' GTCCTTTGTCGATACTGAAAAA 5'
Template	T2: 3' GTCCTTTGTCGATACTGTTTTTTT 5'
Template	T3: 3' GTCCTTTGTCGATACTGGGGGG 5'
Template	T4: 3' GTCCTTTGTCGATACTGCCCCC 5'
Template	T5: 3' GTCCTTTGTCGATACTGCAAAA 5'

The first set of compounds to be examined as potential triphosphate mimics, were adenine sulfonate-phosphoramidates **12-13** and thymidine analogues **1-5**. Preliminary results of single nucleotide incorporation experiments for analogues Tau-dAMP **12** and L-Cys-dAMP **13** had shown modest incorporation efficiencies into a growing DNA chain, as catalyzed by HIV-1 reverse transcriptase[11] or family C *E. coli* Pol III.[19] However, it is recognized that phylogenetically distinct classes of enzymes

might show different substrate-specificities, despite sharing similar polymerization domains and mechanism. As evidence of a dynamic effect triggered by multiple incorporations on the progression of the reaction emerged in the course of previous studies,[19] all compounds were directly subjected to elongation tests.

While Vent (exo-) was not able to accept adenine analogues Tau-dAMP **12** and L-Cys-dAMP **13** or thymidine analogues Tau-dTMP **1**, L-Cys-dTMP **2** and D-Cys-dTMP **3**, 6% and 14% formation of the P + 1 strand was observed at 1mM (60 min) with Klenow (exo-) for cysteic acid-containing thymidine phophoramidates **2** and **3**, and incorporation efficiencies with the same polymerase were relatively higher for adenine analogues **12** and **13**, showing 5% (P + 2) and 6% (P + 4) respectively. However, the thermostable polymerase Therminator could use all those compounds as substrates, in line with the significant propensity of this replicative enzyme to accept modified nucleotides substrates. We have observed 75% (P + 1) and 6% (P + 2) for **1**, 68% (P + 1) and 6% (P + 2) for compound **2**, 70% (P + 1) and 16% (P + 2) for compound **3**, 18% (P+1) for compound **12** and complete conversion to (P + 1) strand, 58% of (P + 2), 28% of (P + 3), 9% of (P + 4) and up to 2% of (P + 5) strand for compound **13**.

Within the series of analogues containing thymine as nucleobase, *N*-methyl taurine derivative *N*-Me-Tau-dTMP **4** showed to be incorporated into the primer template duplex more efficiently than compounds bearing taurine, D- or L-cysteic acid, with 14% conversion up to (P + 2) strand using Vent (60 min, 1mM) and 17% conversion up to (P + 3) using Therminator (60 min, 1mM), whilst homotaurine-dTMP **5** showed only 6% and 20% conversion to (P + 1) strand with Vent and Therminator (60 min, 1mM) respectively. The leaving group properties of bis-taurine in derivative **6** was comparable to that of *N*-methyl taurine, showing formation of 15 % of (P + 2) and 2% of (P + 3) strand with Vent and 21 % of (P + 2) and 3% of (P + 3) strand with Therminator at 60 min (1mM).

The acceptance of taurine-amino acid containing derivatives Tau-Gly-dTMP **7**, Tau-His-dTMP **8** and Tau-Asp-dTMP **9** in primer extension reactions by Vent (exo-) and Therminator polymerases was disappointing, with the exception of aspartate analogue **9** which showed 17% P + 1 and 2% P + 2 after 60 min (1mM) with Therminator. Since the individual aminoacids composing those structural units are proved to be good leaving groups, it can be concluded that the increased linear size of those compounds does not allow accommodation in the enzymatic pocket, and the results obtained for aspartate analogue **9** might be due to induction from the β-carboxylic functionality.

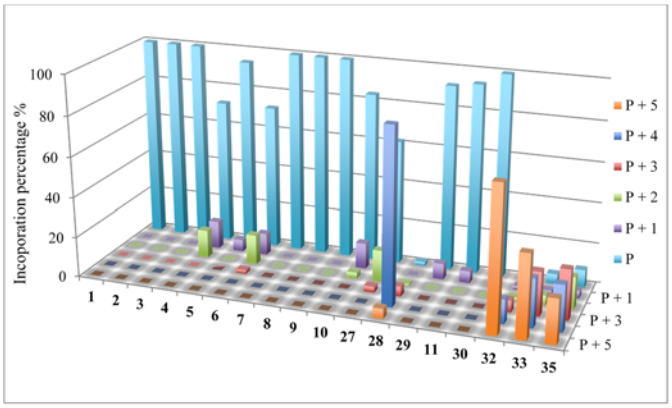


Figure 2. Chain elongation efficiency of compounds **1**-**11**, **27**-**30**, **32**-**33** and **35** by Vent (exo-) polymerase (0.01 U.µl-1), at 1mM conc. after 60 min.

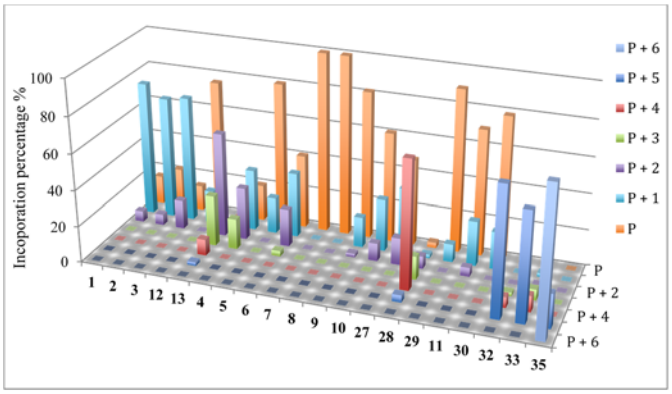


Figure 3. Chain elongation efficiency of compounds **1**-**13**, **27**-**30**, **32**-**33** and **35** by Therminator polymerase (0.01 U.µl-1), at 1mM conc. after 60 min.

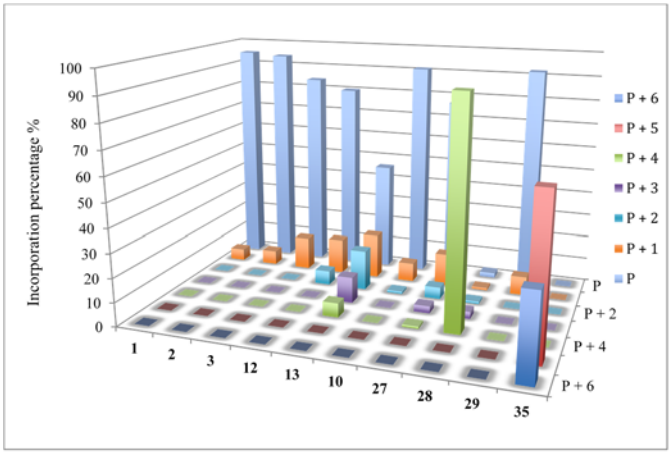


Figure 4. Chain elongation efficiency of compounds **1**-**3**, **10**, **12**-**13**, **27**-**30** and **35** by Klenow fragment polymerase (exo-) (0.05 U.µl-1), at 1mM conc. after 60 min.

The next analogues to be studied were the phosphoramidates derived from taurine-*N*-acetic acid and taurine-*N*-propionic acid. Previously, we observed a preference for incorporation for

purines over pyrimidines with various polymerases with the order A ≥ T = G > C, so we initially wished to compare thymine and adenine analogues. It was found that while thymidine analogues **10** and **11** were inserted with a comparable efficiency, taurine-*N*-acetic acid adenine nucleoside **27** was a better substrate for thermophilic polymerases than taurine-*N*-propionic acid adenine analogue **30**. In particular, for the thymidine analogue **10**, only 3%, 1% and up to 10% formation of P + 2 strand in 60 min at 1mM was observed with Vent, Klenow and Therminator respectively, whereas the incorporation properties of adenine analogue **27** were better than the thymidine analogue, showing up to 3% P + 3 strand with Vent, 15% of P + 2 with Therminator and 1% of P + 4 strand with Klenow. In view of these observations, it was considered important to additionally explore the influence of the other nucleobase moieties on the incorporation efficiency. To this end, we synthesized and investigated deoxynucleoside monophosphate analogues of G and C nucleobases **28** and **29**, modified at the 5'-position with the selected taurine-*N*-acetic acid as leaving group using Vent (exo-), Therminator and Klenow as polymerases. As anticipated, the cytidine analogue **29** showed the lowest efficiency, giving rise to the formation of 8% of P + 1 with Vent and Klenow and 10 % of (P + 1) only with Therminator. Surprisingly, guanidine analogue **28** was added into a growing DNA chain more efficiently than any other nucleotide, with a final base ranking order G > A ≥ T > C for this set of derivatives, showing formation of 87% of (P + 4) and up to 5% of full length (P + 5) strand with Vent, up to 93% of (P + 4) with Klenow and 71% of (P + 4) and 4% of (P + 5) strand with Therminator (Figure 5). The rationale behind the base-specificity of DNA polymerases is uncertain. In a distinct literature report, it has been shown that kinetics of incorporation could be affected by the diversification of the 3'-flanking neighbouring base on the template and between pairing of dCTP or dTTP opposite to O^6-methylguanine.[20]

Ultimately, the elongation ability of diphosphate analogue tau-dTDP **33** was excellent, giving up to 41% full-length chain (P + 5) formation at 1 mM in 60 min with Vent and 9 % full length chain (P + 5) formation even at 0.125 mM in 15 min with Therminator. These findings were thought to be due to the presence of the diphosphate group, thus as additional evidence we compared compound **33** with natural 5'-thymidine diphosphate **32** (dTDP). Enzymatic results pointed to a slightly reduced incorporation efficiency for tau-dTDP in comparison to dTDP. Unsurprisingly, the last compound to be evaluated, Tau-dTTP **35**, was found to be as efficient as the natural substrate dTTP using both Therminator and Klenow although less efficient in the presence of Vent (exo-), showing only 22% formation of P + 5 strand at 1mM concentration.

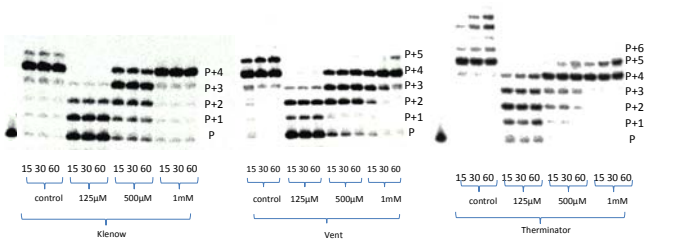


Figure 5. Profile of chain elongation of the compound **28** as substrate into the P1:T4 by three different DNA polymerase; Klenow (0.05 U.µl-1), Vent (0.01 U.µl-1) and Therminator (0.01 U.µl-1).

Table 2 Steady–state kinetics of single nucleotide incorporation into P1T5 by Klenow fragment (0.001 U.µl-1)

Substrate	V_{max} [nM.min^{-1}]	K_m [µM]	V_{max} / K_m [x 10^{-3} min^{-1}]
dGTP	58.10 ± 3.23	0.089 ± 0.015	652.80
Compound **28**	62.55 ± 3.30	67.34 ± 11.00	0.93

As the best results were seen with modified guanidine substrate **28** and Klenow fragment (exo-), incorporation and subsequent extension of compound **28** into primed-template P_1:T_5 were quantitatively determined, based on the Single Complete Hit model, in comparison to its natural nucleoside triphosphate (dGTP). The steady-state kinetic analysis, the results of which are summarized in Table 2, indicates a V_{max} value for compound **28** 1.07 fold higher than that of natural dGTP, but a K_M value 758 fold higher, with a V_{max}/K_m ratio of compound **28** 702-fold lower than the natural substrate. Although the relevant nucleobase (adenine) and DNA enzyme (HIV-1 RT) previously employed in the analysis of bio-isosteres iminodiacetic acid (IDA) and iminodipropionic acid (IDP) were different, we can nevertheless conclude that the leaving group properties of taurine-*N*-acetic acid is comparable to IDA (IDA-dAMP resulted 940-fold lower than the natural substrate) and slightly lower than IDP (IDP-dAMP was 83-fold lower than the natural substrate).

CONCLUSIONS

A novel series of modified deoxyribonucleosides bearing sulfonate or both carboxylate and sulfonate functional groups, connected through a phosphoramidate linker to the 5'-position were successfully prepared. Following multiple incorporation screening of various microbial polymerases (as reviewed in Figures **2**, **3** and **4**), the A family Klenow fragment exonuclease-free polymerase mutant was found to be the most effective polymerase, giving the best elongation rates, while B family thermostable Therminator was confirmed as possessing the broader substrate-specificity.

The development of new leaving groups for DNA polymerization can be a valuable tool for synthetic biology, for which the mesophylic Klenow polymerase from *E. coli* was tested and should aid the search for novel reagents able to perform DNA synthesis at high temperature for polymerase

chain reaction (PCR) for which thermostable DNA polymerases were tested.

EXPERIMENTAL

General information

For all reactions, analytical grade solvents were used. All moisture-sensitive reactions were carried out under an argon or a nitrogen atmosphere in oven-dried glassware (135 ºC). Reaction temperatures are reported as bath temperatures. Precoated aluminum sheets (254 nm) were used for TLC. Compounds were visualized with UV light (λ = 254 nm). Products were purified by flash chromatography on ICN silica gel 63-200, 60 Å. All final compounds were purified by preparative RP-HPLC (Xbridge™ Prep C18 5µm OBD 19 x 150 mm column) or Ion-exchange resin (Source 15 Q). All the elution methods use CH_3CN/H_2O gradients for RP-HPLC or TEAB-H_2O gradient for Ion-exchange. ^1H, ^{13}C and ^{31}P NMR spectra were recorded on Bruker Avance 300 MHz, 500 MHz or 600 MHz spectrometers. For all final compounds, ^1H and ^{13}C resonance assignments were made using 2D NMR correlation experiments (COSY, gHSQC and gHMBC). For sake of clarity, NMR signals of protons and carbons for sugar and base moieties are indicated with and without a prime, respectively. Chemical shifts were referenced to residual solvent signals at δ H/C 7.26/77.00 ($CDCl_3$), 3.31/49.10 (CD_3OD) and 2.50/39.50 (DMSO-d_6) relative to TMS as internal standard. Coupling constants are expressed in hertz (Hz). Splitting patterns are reported as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) and br (broad). High resolution mass spectra were acquired on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA). Samples were infused at 3 µL/min and spectra were obtained in positive or negative ionization mode with a resolution of 15000 (FWHM), using leucine enkephalin as lock mass.

General procedure for phosphoramidates synthesis (A)

The relevant 2'-deoxynucleoside-5'-monophosphate (1 eq.) and amino derivative (3-5 eq.) were suspended in a 4:1 ^{t}BuOH/H_2O mixture (~20 mL/mmol). Triethylamine (5-10 eq.) was then added to the suspension to facilitate dissolution, followed by DCC (2-5 eq.). The reaction mixture was heated at 90 ºC for 3-6 h. The reaction progress was monitored by TLC (iPrOH:H_2O:Et_3N/NH_3 10:2:2). Upon completion, the reaction mixture was cooled to room temperature and the solvent was removed by rotary evaporation. The residue was resuspended in water (100 ml), washed with diethyl ether (3 x 100 ml) and the aqueous phase was lyophilized. The resulting crude material was purified by column chromatography on silica gel using the following gradient IPA:H_2O:NH_3 20:1:1, v/v/v; 15:1:1, v/v/v; 10:1:1, v/v/v (Et_3N was used instead of NH_3 for carboxylic ester containing compounds, to avoid amide formation), to provide the desired nucleoside phosphoramidates as salts. Semi-preparative RP-HPLC (50 mmol TEAB in H_2O + 2% CH_3CN and 50 mmol TEAB in 50% H_2O + 50% CH_3CN) was employed for further purification to obtain the pure products (except for compounds **1**, **2**, **3** and **8**). The residues were lyophilized and freeze-dried repeatedly until constant mass.

Thymidine-5'-taurine phosphoramidate ammonium salt (1). The ammonium salt of compound **1** was obtained as a white solid (0.27 g, 72%) following the general procedure **(A)**, starting from 2'-deoxythymidine-5'-monophosphate (TMP) disodium-salt (0.3 g, 0.82 mmol), taurine (0.31 g, 2.46 mmol), triethylamine (0.57 mL, 4.09 mmol), DCC (0.51g, 2.46 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (15 mL) at 90 ºC for 3 h. ^1H NMR (300 MHz, D_2O) δ = 7.69 (d, J = 1.0 Hz, 1H, H-6), 6.27 (app t, J = 7.1 Hz, 1H, H-1'), 4.51-4.49 (m, 1H, H-3'), 4.09-4.07 (m, 1H, H-4'), 3.94-3.91 (m, 2H, H-5' and H-5''), 3.17-3.10 (m, 2H, -C$H_{2\alpha}$$CH_2SO_3H$), 2.97 (t, J = 6.6 Hz, 2H, -CH_2C$H_{2\beta}$$SO_3H$), 2.33-2.27 (m, 2H, H-2' and H-2''), 1.85 (d, J = 1.0 Hz, 3H, CH_3-Thy); ^{13}C NMR (75 MHz, D_2O) δ = 166.2 (C-4), 151.4 (C-2), 137.1 (C-6), 111.4 (C-5), 85.3 (d, $^3J_{C,P}$ = 8.9 Hz, C-4'), 84.7 (C-1'), 70.9 (C-3'), 63.8 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 52.0 (d, $^3J_{C,P}$ = 6.3 Hz, -CH_2C$_\beta$$H_2SO_3H$), 38.4 (C-2'), 36.7 (-C$_\alpha$$H_2CH_2SO_3H$), 11.3 (C$H_3$-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 8.0; HRMS for $C_{12}H_{20}N_3O_{10}PS$ [M-H]$^-$ calcd.: 428.0534, found: 428.0531.

Thymidine-5'-(L-cysteic acid) phosphoramidate ammonium salt (2). The ammonium salt of compound **2** was obtained as an off-white solid (0.086 g, 30%) following the general procedure **(A)**, starting from TMP disodium-salt (0.2 g, 0.55 mmol), L-cysteic acid (0.31 g, 1.64 mmol), triethylamine (0.76 mL, 5.46 mmol), DCC (0.34g, 1.64 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (14 mL) at 90 ºC for 3 h. ^1H NMR (300 MHz, D_2O) δ = 7.72 (d, J = 1.0 Hz, 1H, H-6), 6.27 (app t, J = 7.0 Hz, 1H, H-1'), 4.51-4.49 (m, 1H, H-3'), 4.09-4.07 (m, 1H, H-4'), 3.96-3.92 (m, 2H, H-5' and H-5''), 3.87-3.83 (m, 1H, C$H$$CO_2H$), 3.16 (d, J = 5.8 Hz, 2H, C$H_2$$SO_3H$), 2.31-2.25 (m, 2H, H-2' and H-2''), 1.86 (d, J = 1.0 Hz, 3H, CH_3-Thy); ^{13}C NMR (75 MHz, D_2O) δ = 178.6 (CO_2H), 166.3 (C-4), 151.5 (C-2), 137.1 (C-6), 111.5 (C-5), 85.5 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 84.6 (C-1'), 70.9 (C-3'), 63.7 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 54.5 (d, $^3J_{C,P}$ = 6.6 Hz, -CH_2SO_3H), 53.8 (d, $^2J_{C,P}$ = 6.1 Hz, CHCO_2H), 38.2 (C-2'), 11.3 (CH_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 6.1; HRMS for $C_{13}H_{20}N_3O_{12}PS$ [M-H]$^-$ calcd.: 472.0432, found: 472.0436.

Thymidine-5'-(D-cysteic acid) phosphoramidate ammonium salt (3). The ammonium salt of compound **3** was obtained as an off-white solid (0.072 g, 25%) following the general procedure **(A)**, starting from TMP disodium-salt (0.2 g, 0.55 mmol), D-cysteic acid (0.31 g, 1.64 mmol), triethylamine (0.76 mL, 5.46 mmol), DCC (0.34g, 1.64 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (14 mL) at 90 ºC for 3 h. ^1H NMR (300 MHz, D_2O) δ = 7.69 (d, J = 1.0 Hz, 1H, H-6), 6.26 (app t, J = 7.1 Hz, 1H, H-1'), 4.52-4.49 (m, 1H, H-3'), 4.09-4.07 (m, 1H, H-4'), 3.96-3.92 (m, 2H, H-5' and H-5''), 3.86-3.82 (m, 1H, C$H$$CO_2H$), 3.16 (d, J = 5.8 Hz, 2H, C$H_2$$SO_3H$), 2.31-2.25 (m, 2H, H-2' and H-2''), 1.86 (d, J = 1.0 Hz, 3H, CH_3-Thy); ^{13}C NMR (75 MHz, D_2O) δ = 178.0 (d, $^3J_{C,P}$ = 5.1 Hz, CO_2H), 166.3 (C-4), 151.5 (C-2), 137.2 (C-6), 111.4 (C-5), 85.4 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 84.7 (C-1'), 71.0 (C-3'), 63.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 54.5 (d, $^3J_{C,P}$ = 6.6 Hz, CH_2SO_3H), 53.5 (d, $^2J_{C,P}$ = 6.6 Hz, CHCO_2H), 38.2 (C-2'), 11.3 (CH_3-Thy); ^{31}P NMR (121 MHz, D_2O) δ = 6.1; HRMS for

HRMS for $C_{13}H_{20}N_3O_{12}PS$ [M-H]$^-$ calcd.: 472.0432, found: 472.0430.

2'-Deoxyadenine-5'-taurine phosphoramidate ammonium salt (12).[11] The ammonium salt of compound **12** was obtained as a white solid (0.45 g, 71%) following the general procedure (A), starting from dAMP (0.441 mg, 1.333 mmol), taurine (0.5 g, 4.00 mmol), triethylamine (1.48 mL, 10.66 mmol), DCC (0.82 g, 4.00 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (40 mL) at 90 ºC for 3 h.

2'-Deoxyadenine-5'-(L-cysteic acid) phosphoramidate ammonium salt (13).[11] The ammonium salt of compound **13** was obtained as a white solid (0.23 g, 32%) following the general procedure **(A)**, starting from dAMP (441.5 mg, 1.333 mmol), L-cysteic acid (338 mg, 1.999 mmol), triethylamine (1.48 mL, 10.66 mmol), DCC (0.82 g, 4.00 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (40 mL) at 90 ºC for 3 h.

Thymidine-5'-*O*-(*N*-methyltaurine) phosphoramidate triethylammonium salt (4). The triethylammonium salt of compound **4** was obtained as a white solid (0.39 g, 74%) following the general procedure **(A)**, starting from TMP-triethylammonium salt (0.43 g, 0.82 mmol), *N*-methyltaurine (0.57 g, 4.09 mmol), triethylamine (0.91 mL, 6.55 mmol) and DCC (1.01 g, 4.91 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (12 mL) at 90 ºC for 5.5 h.^{1}H NMR (500 MHz, D_2O) δ = 7.73 (s, 1H, H-6), 6.20 (app t, J = 7.0 Hz, 1H, H-1'), 4.53-4.52 (m, 1H, H-3'), 4.12-4.11 (m, 1H, H-4'), 3.94-3.91 (m, 2H, H-5' and H-5''), 3.33-3.27 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.08 (t, J = 7.0 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.58 (d, $^3J_{H,P}$ = 9.5 Hz , 3H, N-CH$_3$), 2.34-2.31 (m, 2H, H-2' and H-2''), 1.89 (s, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, D_2O) δ = 166.0 (C-4), 151.1 (C-2), 136.7 (C-6), 111.1 (C-5), 84.9 (d, $^3J_{C,P}$ = 9.1 Hz, C-4'), 84.3 (C-1'), 70.6 (C-3'), 63.5 (d, $^2J_{C,P}$ = 5.3 Hz, C-5'), 48.8 (d, $^3J_{C,P}$ = 2.9 Hz, CH$_2$CβH$_2$SO$_3$H), 44.5 (d, $^2J_{C,P}$ = 4.3 Hz, CαH$_2$CH$_2$SO$_3$H), 38.0 (C-2'), 33.0 (d, $^2J_{C,P}$ = 3.5 Hz, N-CH$_3$), 11.1 (CH$_3$-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 8.9; HRMS for $C_{13}H_{22}N_3O_{10}PS$ [M-H]$^-$ calcd.: 442.0691, found: 442.0689.

Thymidine-5'-(3-aminopropanesulfonic acid) phosphoramidate triethylammonium salt (5). The triethylammonium salt of compound **5** was obtained as a white solid (0.272 g , 77%) following the general procedure (A), starting from TMP-disodium salt (0.2 g, 0.55 mmol), 3-aminoethanesulfonic acid (0.23 g, 1.64 mmol), triethylamine (0.46 mL, 3.28 mmol) and DCC (0.45 g, 2.18 mmol) at 90 ºC for 4 h.^{1}H NMR (500 MHz, D_2O) δ = 7.76 (s, 1H, H-6), 6.32 (app t, J = 7.0 Hz, 1H, H-1'), 4.55-4.53 (m, 1H, H-3'), 4.13 (br s, 1H, H-4'), 3.99-3.92 (m, 2H, H-5' and H-5''), 22.91-2.86 (m, 4H, -$CH_{2\alpha}$CH$_2$CH$_2$SO$_3$H and -CH$_2$CH$_2$$CH_{2\gamma}SO_3$H), 2.39-2.28 (m, 2H, H-2' and H-2''), 1.89 (s, 3H, CH_3-Thy), 1.87-1.81 (m, 2H, -CH$_2$$CH_{2\beta}CH_2SO_3$H); ^{13}C NMR (125 MHz, D_2O) δ = 166.4 (C-4), 151.5 (C-2), 136.9 (C-6), 111.1 (C-5), 85.0 (d, $^3J_{C,P}$ = 9.1 Hz, C-4'), 84.3 (C-1'), 70.7 (C-3'), 63.4 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 48.2 (-CH$_2$CH$_2$$C_\gamma H_2SO_3$H), 39.4 (-$C_\alpha H_2CH_2CH_2SO_3$H), 38.1 (C-2'), 26.0 (d, $^3J_{C,P}$ = 6.8 Hz, -CH$_2$$C_\beta H_2CH_2SO_3$H), 11.2 ($CH_3$-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 8.5 ; HRMS for $C_{13}H_{22}N_3O_{10}PS$ [M-H]$^-$ calcd.: 442.0690, found: 442.0668.

***N,N*-2-Diaminoethanesulfonate-disodium salt (14)**. To a stirred solution of 2-aminoethane sulfonic acid (2.0 g, 16.0 mmol) and $NaHCO_3$ (1.34 g, 16.0 mmol) in water, was added an aqueous solution (25%) of sodium vinylsulfonate (6.72 mL, 15.2 mmol). The resulting mixture was heated to reflux for 3 d. The reaction mixture was then cooled and the solvent was removed by distillation under vacuum. The remaining solid was purified by column chromatography on silica gel (gradient: IPA:H_2O 50:1, v/v; 20:1, v/v; 10:1, v/v) to provide compound **14** (3.58 g, 85%) as a fluffy white solid. ^{1}H NMR (300 MHz, D_2O) δ = 3.56 (t, J = 6.6 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.32 (t, J = 6.6 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H); ^{13}C NMR (75 MHz, D_2O) δ = 46.2 (CH$_2$$C_\beta$H$_2SO_3$H), 43.3 ($C_\alpha$H$_2CH_2SO_3$H); HRMS for $C_4H_{11}NO_6S_2$ [M-H]$^-$ calcd.: 231.9955, found: 231.9954.

Thymidine-5'-*(N,N*-2-diaminoethanesulfonate) phosphoramidate triethylammonium salt (6). The triethylammonium salt of compound **6** (0.243 g, 53%) was obtained as a white solid following the general procedure (**A**), starting from TMP-triethylammonium salt (0.286 g, 0.55 mmol), compound **14** (0.76 g, 2.73 mmol), triethylamine (0.91 mL, 6.55 mmol) and DCC (0.68 g, 3.28 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (9 mL) at 90 ºC for 5.5 h. ^{1}H NMR (500 MHz, D_2O) δ = 7.80 (s, 1H, H-6), 6.37 (app t, J = 7.0 Hz, 1H, H-1'), 4.61-4.58 (m, 1H, H-3'), 4.18 (br s, 1H, H-4'), 4.05-3.99 (m, 2H, H-5' and H-5''), 3.43-3.36 (m, 4H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.17 (t, J = 8.0 Hz, 4H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.40-2.37 (m, 2H, H-2' and H-2''), 1.96 (s, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, D_2O) δ = 166.0 (C-4), 151.2 (C-2), 136.7 (C-6), 111.3 (C-5), 84.9 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 84.4 (C-1'), 70.6 (C-3'), 63.5 (d, $^2J_{C,P}$ = 5.4 Hz, C-5'), 49.5 (CH$_2$$C_\beta$H$_2SO_3$H), 41.7 (d, $^2J_{C,P}$ = 4.9 Hz, C_αH$_2$CH$_2$SO$_3$H), 38.1 (C-2'), 11.2 (CH$_3$-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 7.4 ; HRMS for $C_{14}H_{24}N_3O_{13}PS_2$ [M-H]$^-$ calcd.: 536.0415, found: 536.0418.

General procedure for taurine-amino acid synthons synthesis

DCC (1.1 eq.) was added to a solution of *N*-Boc-protected amino acid (1 eq.) and *N*-hydroxy succinimide (1.04 eq.) in dry THF (~4 mL/mmol) at 0 ºC, and the mixture was stirred for an additional 0.5 h at same temperature, then for 1.5 h at room temperature. The resulting DCU was filtered off, and the filtrate was concentrated *in vacuo* to afford the activated ester. The crude residue was then dissolved in dioxane (3 mL/mmol), and to this solution was added a solution of taurine sodium salt (1 eq.) in water (2 mL/mmol) at room temperature. The reaction mixture was stirred at room temperature for 14 h, followed by evaporation of the organic phase *in vacuo*. The aqueous layer was washed with ethyl acetate (3 x) and was then loaded on a column packed with IR-120B (H^+ form) resin. The aqueous eluent was evaporated to dryness *in vacuo* to give a crystalline powder, which was recrystallized from ethanol-water to give pure H_2N-aminoacid-Taurine-OH as a white crystalline solid.

H-Gly-Tau-OH (16a). Compound **16a** was obtained as a white solid (0.95 g, 91%) according to the general procedure, starting from DCC (1.29 g, 6.27 mmol), Boc-Gly-OH **15a** (1.0 g, 5.70 mmol), *N*-hydroxy succinimide (0.68 g, 5.93 mmol) in THF (20 mL) and taurine sodium salt (0.84 g, 5.70 mmol) in

dioxane-water (1:1, 30 mL). ^{1}H NMR (300 MHz, D_2O) δ = 3.82 (s, 2H, CH_2-Gly), 3.67 (t, J = 6.0 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.13 (t, J = 6.8 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H); ^{13}C NMR (75 MHz, D_2O) δ = 166.5 (CO-Gly), 49.1 (CH$_2$$C_\beta$H$_2SO_3$H), 40.1 ($CH_2$-Gly), 34.8 ($C_\alpha$H$_2CH_2SO_3$H); HRMS for $C_4H_{10}N_2O_4S$ [M-H]$^-$ calcd.: 181.0288, found: 181.0290.

H-His-Tau-OH (16b). Compound **16b** was obtained as a white solid (1.55 g, 52%) according to the general procedure, starting from DCC (2.58 g, 12.54 mmol), Boc-His-OH **15b** (2.92 g, 11.40 mmol), *N*-hydroxy succinimide (1.36 g, 11.86 mmol) in THF (44 mL) and taurine sodium salt (1.68 g, 11.40 mmol) in dioxane-water (1:1, 60 mL). ^{1}H NMR (300 MHz, D_2O) δ = 8.05 (s, 1H, His Ar-H), 7.17 (s, 1H, His Ar-H), 4.13 (t, J = 6.9 Hz, 1H, His-CH), 3.62-3.54 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.18 (d, J = 6.9 Hz, 2H, His-CH_2), 3.08-2.98 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H); ^{13}C NMR (75 MHz, D_2O) δ = 169.6 (CO-His), 135.4 (Ar C-His), 129.6 (Ar C-His), 116.9 (Ar C-His), 52.9 (αC-His), 49.0 (CH$_2$$C_\beta$H$_2SO_3$H), 34.8 ($C_\alpha$H$_2CH_2SO_3$H), 28.0 (βC-His); HRMS for $C_8H_{14}N_4O_4S$ [M-H]$^-$ calcd.: 261.0663, found: 261.0663.

Thymidine-5'-*N*-(Gly-Tau-OH)-phosphoramidate ammonium salt (7). The ammonium salt of compound **7** was obtained as a white solid (0.29 g , 68%) following the general procedure **(A)**, starting from TMP-disodium salt (0.3 g, 0.82 mmol), H-Gly-Tau-OH (0.49 g, 2.70 mmol), triethylamine (0.57 mL, 4.09 mmol) and DCC (0.78 g, 3.78 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (8 mL) at 90 ºC for 4.5 h. ^{1}H NMR (500 MHz, DMSO-d_6) δ = 8.19 (t, J = 5.4 Hz, 1H, NH-Gly), 7.82 (d, J = 1.0 Hz, 1H, H-6), 6.20 (dd, J = 7.8, 6.0 Hz, 1H, H-1'), 4.27-4.25 (m, 1H, H-3'), 3.86-3.65 (m, 1H, H-4'), 3.76-3.74 (m, 2H, H-5' and H-5''), 3.68 (br s, 1H, 3'-OH), 3.41-3.37 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.21 (d, J = 6.3 Hz, 2H, CH_2-Gly), 3.62 (t, J = 6.3 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.14-2.00 (m, 2H, H-2' and H-2''), 1.80 (d, J = 1 Hz, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ = 171.7 (d, $^3J_{C,P}$ = 5.1 Hz, CO-Gly), 163.9 (C-4), 150.6 (C-2), 136.4 (C-6), 109.9 (C-5), 86.2 (d, $^3J_{C,P}$ = 7.8 Hz, C-4'), 83.8 (C-1'), 71.2 (C-3'), 63.9 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 50.3 (CH$_2$$C_\beta$H$_2SO_3$H), 45.8 ($CH_2$-Gly), 39.2 (C-2'), 35.1 (C_αH$_2$CH$_2$SO$_3$H), 12.2 (CH$_3$-Thy); ^{31}P NMR (202 MHz, DMSO-d_6) δ = 5.1 ; HRMS for $C_{14}H_{23}N_4O_{11}PS$ [M-H]$^-$ calcd.: 485.0749, found: 485.0750.

Thymidine-5'-(*N*-His-Tau-OH)-phosphoramidate triethylammonium salt (8). The triethylammonium salt of compound **8** was obtained as white solid (0.26 g, 41%) following the general procedure **(A)**, starting from TMP-disodium salt (0.3 g, 0.82 mmol), H-His-Tau-OH (0.97 g, 3.68 mmol), triethylamine (0.57 mL, 4.09 mmol) and DCC (0.85 g, 4.09 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (10 mL) at 90 ºC for 4.5 h. ^{1}H NMR (500 MHz, D_2O) δ = 7.61 (d, J =1 Hz, 1H, H-6), 7.59 (d, J = 0.5 Hz, 1H, imidazole), 6.88 (s, 1H, imidazole), 6.32 (app t, J = 6.5 Hz, 1H, H-1'), 4.43-4.41 (m, 1H, H-3'), 4.04 (br s, 1H, H-4'), 3.81-3.75 (m, 3H, H-5', H-5'' and αCH-His), 3.54-3.47 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 2.93-2.91 (m, 2H, βCH_2-His), 2.90-2.88 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.28-2.18 (m, 2H, H-2' and H-2''), 1.84 (d, J = 1Hz, 3H, CH$_3$); ^{13}C NMR (125 MHz, D_2O) δ = 177.0 (CO-His), 173.9 (C-4), 157.0 (C-2), 137.9 (C-6), 136.9 (Ar C-His), 113.0 (C-5), 86.4 (d, $^3J_{C,P}$ =9.3 Hz, C-4'), 86.1 (C-1'), 72.5 (C-3'), 65.2 (d, $^2J_{C,P}$ = 4.2 Hz, C-5'), 57.1 (αCH-His), 50.6 (CH$_2$$C_\beta$H$_2SO_3$H), 40.0 (C-2'), 35.9 ($C_\alpha$H$_2CH_2SO_3$H), 32.2 (βCH$_2$-His), 13.5 (CH$_3$); ^{31}P NMR (202 MHz, D_2O) δ = 5.3; HRMS for $C_{18}H_{27}N_6O_{11}PS$ [M-H]$^-$ calcd.: 565.1123, found: 565.1126.

***N*-Cbz-Asp(OBn)-Tau-OH (17)**. Compound **17** was prepared according to a modification of the general procedure for taurine-amino acid synthon synthesis (ion-exchange chromatography was not required) starting from *N*-Cbz-Asp(OBn)-OH (2.5 g, 6.99 mmol), *N*-hydroxy succinimide (0.84 g, 7.27 mmol), DCC (1.59 g, 7.70 mmol) in THF (30 mL), and taurine (0.92 g, 7.35 mmol) and NaHCO$_3$ (1.17 g, 13.99 mmol) in dioxane-water (1:1, 60 mL). The crude product was resuspended in water, cooled in an ice-bath and acidified to pH = 5 with the addition of 1N HCl. The aqueous layer was then lyophilized and the crude residue was purified by column chromatography on silica gel (gradient: IPA:H$_2$O 100:0, v/v; 30:1, v/v; 20:1, v/v) to give compound **17** (2.89 g, 89%) as a white solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.06 (t, J = 5.2 Hz, 1H, NH-Tau), 7.67 (d, J = 8.3 Hz, 1H, NH-Asp), 7.37-7.30 (m, 10H, 2 x ArH of OBn), 5.08 and 5.02 (2 s, 2H, 2 x OCH_2Ph), 4.40-4.33 (m, 1H, αCH-Asp), 3.37-3.33 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 2.87-2.60 (m, 2H, βCH_2-Asp), 2.56 (t, J = 7.3 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H); ^{13}C NMR (75 MHz, DMSO-d_6) δ = 170.2 (βCO-Asp), 169.8 (αCO-Asp), 155.8 (OCONH), 136.9 (1C of OCH$_2$Ph), 136.1 (1C of OCH$_2$Ph), 128.4 (Ar-C), 128.3 (Ar-C), 127.9 (Ar-C), 127.8 (Ar-C), 127.7 (Ar-C), 65.6 (2 x OCH$_2$Ph), 51.5 (αC-Asp), 50.1 (CH$_2$$C_\beta$H$_2SO_3$H), 36.4 (βC-Asp), 35.7 (C_αH$_2$CH$_2$SO$_3$H); HRMS for $C_{21}H_{24}N_2O_8S$ [M-H]$^-$ calcd.: 463.1180, found: 463.1181.

H-Asp-Tau-OH (18). To a stirring solution of **17** (2.89 g, 2.44 mmol) in EtOH:H$_2$O (5:1, 80 mL) was added 10% Pd/C (0.5 g, 15% w/w) and the mixture was hydrogenated at atmospheric pressure using a balloon filled with H$_2$ for 16 h. The catalyst was then removed by filtration over a pad of Celite and the filtrate was concentrated under reduced pressure to give compound **18** as a white solid (1.49 g, quan.). The resulting solid was analysed for purity and used in the next step without any further purification. ^{1}H NMR (300 MHz, D_2O) δ = 4.22-4.18 (m, 1H, αCH-Asp), 3.66-3.60 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.11 (t, J = 6.7 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.84-2.64 (m, 2H, βCH$_2$-Asp); ^{13}C NMR (75 MHz, D_2O) δ = 177.7 (CO$_2$H), 170.9 (αCO-Asp), 52.5 (αC-Asp), 51.1 (CH$_2$$C_\beta$H$_2SO_3$H), 38.5 (βC-Asp), 37.0 (C_αH$_2$CH$_2$SO$_3$H); HRMS for $C_6H_{12}N_2O_6S$ [M-H]$^-$ calcd.: 239.0343, found: 239.0339.

General procedure for phosphoramidites synthesis starting from 2'-deoxynucleoside-5'-phosphorimidazolide sodium salts (B)

To a stirred solution of 2'-deoxynucleoside-5'-phosphorimidazolide sodium salt (1 eq.) and amino-acid (2.5-3.0 eq.) in DMF, Et$_3$N (5-10 eq.) was added and the mixture was stirred vigorously until total dissolution. The reaction was stirred at 35 °C for 7 days and then quenched by addition of water. The solvent was removed under reduced pressure and the resulting crude material was purified by column chromatography on silica gel (gradient: IPA/H$_2$O/Et$_3$N 20:1.5:0.5, v/v/v; 15:1.5:0.5,

v/v/v; 10:1.5:0.5, v/v/v) to provide the corresponding phosphoramidate as a trietylammonium salt which was further purified by RP-HPLC (50 mmol TEAB in 98% H_2O + 2% CH_3CN and 50 mmol TEAB in 50% H_2O + 50% CH_3CN) to give the pure product as a white solid. The isolated product was freeze-dried repeatedly until constant mass.

Thymidine-5'-(Asp-Tau-OH) phosphoramidate triethylammonium salt (9). The triethylammonium salt of compound **9** was obtained as a white solid (0.17 g, 53%) following the general procedure **(B)**, starting from thymidine 5'-phosphorimidazolide-sodium salt (0.15 g, 0.38 mmol), compound **18** (0.229 g, 0.951 mmol) and Et_3N (0.41 mL, 3.04 mmol) in DMF (4 mL). ^{1}H NMR (500 MHz, D_2O) δ = 7.81 (s, 1H, H-6), 6.39 (app t, J = 7.0 Hz, 1H, H-1'), 4.58-4.57 (m, 1H, H-3'), 4.20 (br s, 1H, H-4'), 4.03-4.01 (m, 2H, H-5', H-5''), 3.84-3.79 (m, 1H, αC*H*-Asp), 3.64-3.60 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.13 (t, J = 7.0 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.71 (dd, J = 16.0, 5.4 Hz, 1H, βC*H'*-Asp), 2.60 (dd, J = 15.7, 5.1 Hz, 1H, βC*H''*-Asp), 2.40-2.37 (m, 2H, H-2' and H-2''), 1.95 (s, 3H, CH_3-Thy); ^{13}C NMR (125 MHz, D_2O) δ = 177.7 (CO_2H), 175.6 (d, $^3J_{C,P}$ = 6.6 Hz, αCO-Asp), 166.0 (C-4), 151.2 (C-2), 136.8 (C-6), 111.1 (C-5), 85.1 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 84.4 (C-1'), 70.8 (C-3'), 63.6 (d, $^2J_{C,P}$ = 4.6 Hz, C-5'), 52.6 (αCH-Asp), 49.0 (CH$_2$C$_\beta$H$_2$SO$_3$H), 39.9 (d, $^3J_{C,P}$ = 3.5 Hz, βCH$_2$-Asp), 38.1 (C-2'), 34.5 (C$_\alpha$H$_2$CH$_2$SO$_3$H), 11.1 (CH$_3$-Thy); ^{31}P NMR (202 MHz, D_2O) δ = 5.8; HRMS for $C_{16}H_{25}N_4O_{13}PS$ [M-H]$^-$ calcd.: 543.0803, found: 543.0813.

***N*-(Ethyl acetate)-2-aminoethanesulfonic acid (19).** To a stirred solution of 2-aminoethane sulfonic acid (4.0 g, 31.96 mmol) in a 1:1 dioxane/H_2O mixture (80 mL), $NaHCO_3$ (2.68 g, 31.96 mmol) was added, followed by ethyl bromo acetate (3.75 mL, 35.16 mmol). The reaction mixture was then heated at 70 ºC for 24 h. The organic layer was removed under reduced pressure and the remaining aqueous layer was washed with ethyl acetate (2 x 100 mL) and neutralized with 1N HCl. The aqueous layer was then concentrated under reduced pressure and the residue was purified by column chromatography on silica gel (gradient: DCM/MeOH 9:1, v/v; 8:2, v/v; 7.5:2.5, v/v) to afford compound **19** (3.24 g, 48%) as a white solid. ^{1}H NMR (300 MHz, D_2O) δ = 4.32 (q, J = 7.1 Hz , 2H, CO$_2$CH_2CH$_3$), 4.05 (s, 2H, CH_2CO$_2$Et), 3.54 (t, J = 6.8 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.33 (t, J = 6.8 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 1.31 (t, J = 7.1 Hz, 3H, CO$_2$CH$_2$CH_3); ^{13}C NMR (75 MHz, D_2O) δ = 167.8, 64.1, 48.3, 47.1, 43.9, 13.8; HRMS for $C_6H_{13}NO_5S$ [M-H]$^-$ calcd.: 210.04416, found: 210.0439.

***N*-(Ethyl propionate)-2-aminoethanesulfonic acid (20).** To a stirred solution of 2-aminoethane sulfonic acid (4.0 g, 31.96 mmol) in a 1:1 EtOH/H_2O mixture (80 mL), $NaHCO_3$ (2.68 g, 31.96 mmol) was added, followed by ethyl acrylate (3.75 mL, 35.16 mmol). The reaction mixture was then stirred at room temperature for 7 days. Ethanol was removed under reduced pressure and the aqueous layer was washed with ethyl acetate (2 x 100 mL) and neutralized with 1N HCl. The aqueous layer was then concentrated under reduced pressure and the residue was purified by column chromatography on silica gel (gradient: DCM/MeOH 9:1, v/v; 8:2, v/v; 7:3, v/v) to afford compound **20** (4.17 g, 58%) as a white solid. ^{1}H NMR (300 MHz, D_2O) δ = 4.15 (q, J = 7.1 Hz , 2H, CO$_2$CH_2CH$_3$), 3.45 (t, J = 6.6 Hz, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.35 (t, J = 6.5 Hz, 2H, C$H_{2\alpha}$CH$_2$CO$_2$Et), 3.23 (t, J = 6.6 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.80 (t, J = 6.5 Hz, 2H, CH$_2$C$H_{2\beta}$CO$_2$Et), 1.20 (t, J = 7.1 Hz, 3H, CO$_2$CH$_2$CH_3); ^{13}C NMR (75 MHz, D_2O) δ = 172.1, 62.1, 46.1, 43.1, 42.9, 29.9, 12.9; HRMS for $C_7H_{15}NO_5S$ [M-H]$^-$ calcd.: 224.0598, found: 224.0598.

Thymidine-5'-*O*-[*N*-(ethylacetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (21). The triethylammonium salt of compound **21** was obtained as a white solid (123 mg, 52%) following the general procedure **(B)**, starting from thymidine-5'-phosphorimidazolide sodium salt (130 mg, 0.33 mmol), compound **19** (195 mg, 0.92 mmol), triethylamine (0.28 mL, 1.98 mmol) and DMF (4 mL). ^{1}H NMR (300 MHz, D_2O) δ = 7.69 (s, 1H, H-6), 6.37 (t, J = 7.1 Hz, 1H, H-1'), 4.58-4.54 (m, 1H, H-3'), 4.17-4.19 (m, 3H, H-4' and OCH_2CH$_3$), 4.05-3.98 (m, 2H, H-5' and H-5''), 3.85 (d, J = 10.6 Hz, 2H, CH_2CO$_2$Et), 3.48-3.37 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.18-3.10 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H) (merged with Et_3N), 2.37-2.32 (m, 2H, H-2' and H-2''), 1.92 (s, 3H, CH$_3$), 1.27-1.22 (m, 3H, OCH$_2$CH_3) (merged with Et_3N); ^{13}C NMR (75 MHz, D_2O) δ = 173.5 (d, $^3J_{C,P}$ = 2.4 Hz, CO_2Et), 171.0 (C-4), 155.1 (C-2), 136.5 (C-6), 111.5 (C-5), 85.0 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 84.6 (C-1'), 71.0 (C-3'), 63.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 61.6 (OCH$_2$CH$_3$), 49.7 (CH$_2$C$_\beta$H$_2$SO$_3$H), 48.5 (d, $^2J_{C,P}$ = 5.5 Hz, CH_2CO$_2$Et), 43.4 (d, $^2J_{C,P}$ = 4.7 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.5 (C-2'), 13.0 (OCH$_2$CH_3), 11.9 (CH$_3$); ^{31}P NMR (121 MHz, D_2O) δ = 7.1; HRMS for $C_{16}H_{26}N_3O_{12}PS$ [M-H]- calcd.: 514.0902, found: 514.0902.

2'-Deoxyadenosine-5'-*O*-[*N*-(ethylacetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (22). The triethylammonium salt of compound **22** was obtained as a white solid (47 mg, 38%) following the general procedure **(B)**, starting from 2'-deoxyadenosine-5'-phosphorimidazolide sodium salt (70 mg, 0.17 mmol), compound **19** (103 mg, 0.49 mmol), triethylamine (0.145 mL, 1.04 mmol) and DMF (3 mL). ^{1}H NMR (300 MHz, D_2O) δ = 8.40 (s, 1H, H-8), 8.13 (s, 1H, H-2), 6.40 (t, J = 6.7 Hz, 1H, H-1'), 4.72-4.68 (m, 1H, H-3'), 4.24-4.23 (m, 1H, H-4'), 4.05-3.95 (m, 4H, H-5',H-5'' and OCH_2CH$_3$), 3.70 (d, J = 10.5 Hz, 2H, CH_2CO$_2$Et), 3.41-3.28 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.11-3.03 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.91-2.74 (m, 1H, H-2'), 2.64-2.53 (m, 1H, H-2''), 1.13 (t, J = 7.1 Hz, 3H, OCH$_2$CH_3); ^{13}C NMR (75 MHz, D_2O) δ = 173.3 (d, $^3J_{C,P}$ = 2.8 Hz, CO_2Et), 155.0 (C-6), 152.2 (C-2), 148.1 (C-4), 139.4 (C-8), 118.1 (C-5), 85.7 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 83.3 (C-1'), 71.1 (C-3'), 63.8 (d, $^2J_{C,P}$ = 5.2 Hz, C-5'), 61.5 (OCH$_2$CH$_3$), 49.7 (CH$_2$C$_\beta$H$_2$SO$_3$H), 48.4 (d, $^2J_{C,P}$ = 5.3 Hz, CH$_2$CO$_2$Et), 43.4 (d, $^2J_{C,P}$ = 4.6 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.6 (C-2'), 12.3 (OCH$_2$CH$_3$); ^{31}P NMR (121 MHz, D_2O) δ = 7.2; HRMS for $C_{10}H_{14}N_5O_9PS$ [M-H]$^-$ calcd.: 523.1017, found: 523.1019.

2'-Deoxyguanosine-5'-*O*-[*N*-(ethyl acetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (23). The triethylammonium salt of compound **23** was obtained as an off-white solid (58 mg, 22%) following the general procedure **(B)**, starting from 2'-

deoxyguanosine-5'-phosphorimidazolide sodium salt (150 mg, 0.36 mmol), compound **19** (212 mg, 1.00 mmol), triethylamine (0.30 mL, 2.15 mmol) and DMF (4 mL). ^{1}H NMR (300 MHz, D_2O) δ = 7.99 (s, 1H, H-8), 6.26 (app t, *J* = 7.0 Hz, 1H, H-1'), 4.65-4.63 (m, 1H, H-3'), 4.17-4.15 (m, 1H, H-4'), 4.04 (q, *J* = 7.0 Hz, 2H, OC*H*$_2$CH$_3$), 3.96-3.93 (m, 2H, H-5' and H-5''), 3.74 (d, *J* = 10.5 Hz, 2H, C*H*$_2$CO$_2$Et), 3.40-3.33 (m, 2H, C*H*$_{2\alpha}$CH$_2$SO$_3$H), 3.01-2.96 (m, 2H, CH$_2$C*H*$_{2\beta}$SO$_3$H), 2.83-2.71 (m, 1H, H-2'), 2.50-2.43 (m, 1H, H-2''), 1.13 (t, *J* = 7.4 Hz, 3H, OCH$_2$C*H*$_3$); ^{13}C NMR (75 MHz, D_2O) *δ* = 173.4 (d, $^3J_{C,P}$ = 2.9 Hz, CO_2Et), 159.4 (C-6), 153.4 (C-2), 151.0 (C-4), 135.4 (C-8), 117.0 (C-5), 85.3 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 82.6 (C-1'), 71.2 (C-3'), 64.0 (d, $^2J_{C,P}$ = 5.4 Hz, C-5'), 61.6 (O*C*H$_2$CH$_3$), 49.8 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 48.5 (d, $^2J_{C,P}$ = 5.4 Hz, *C*H$_2$CO$_2$Et), 43.4 (d, $^2J_{C,P}$ = 4.6 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 38.4 (C-2'), 12.9 (OCH$_2$*C*H$_3$); ^{31}P NMR (121 MHz, D_2O) *δ* = 7.2; HRMS for $C_{16}H_{25}N_6O_{11}PS$ [M-H]$^-$ calcd.: 539.0967, found: 539.0969.

2'-Deoxycytidine-5'-*O*-[*N*-(ethyl acetate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (24). The triethylammonium salt of compound **24** was obtained as a white solid (130 mg, 47%) following the general procedure **(B)**, starting from 2'-deoxycytidine-5'-phosphorimidazolide sodium salt (150 mg, 0.39 mmol), compound **19** (234 mg, 1.11 mmol), triethylamine (0.33 mL, 2.37 mmol) and DMF (4 mL).^{1}H NMR (300 MHz, D_2O) δ = 7.94 (d, *J* = 7.6 Hz, 1H, H-6), 6.29 (app t, *J* = 6.8 Hz, 1H, H-1'), 6.08 (d, *J* = 7.5 Hz, 1H, H-5), 4.53-4.49 (m, 1H, H-3'), 4.17-4.10 (m, 3H, H-4' and OC*H*$_2$CH$_3$), 4.03-3.98 (m, 2H, H-5' and H-5''), 3.82 (d, *J* = 10.7 Hz, 2H, C*H*$_2$CO$_2$Et), 3.44-3.35 (m, 2H, C*H*$_{2\alpha}$CH$_2$SO$_3$H), 3.19-3.09 (unresolved m, 2H, CH$_2$C*H*$_{2\beta}$SO$_3$H, merged with Et$_3$N), 2.43-2.21 (m, 2H, H-2' and H-2''), 1.27-1.22 (unresolved m, 3H, OCH$_2$C*H*$_3$, merged with Et$_3$N); ^{13}C NMR (75 MHz, D_2O) *δ* = 173.4 (d, $^3J_{C,P}$ = 2.5 Hz, CO_2Et), 165.8 (C-4), 157.1 (C-2), 141.2 (C-6), 96.1 (C-5), 85.6 (C-1'), 85.4 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 70.7 (C-3'), 63.7 (d, $^2J_{C,P}$ = 5.2 Hz, C-5'), 61.6 (O*C*H$_2$CH$_3$), 49.7 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 48.6 (d, $^2J_{C,P}$ = 5.5 Hz, *C*H$_2$CO$_2$Et), 43.4 (d, $^2J_{C,P}$ = 4.7 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 39.3 (C-2'), 13.0 (OCH$_2$*C*H$_3$); ^{31}P NMR (121 MHz, D_2O) *δ* = 7.2; HRMS for $C_{15}H_{25}N_4O_{11}PS$ [M-H]$^-$ calcd.: 499.0905, found: 499.0898.

Thymidine-5'-*O*-[*N*-(ethyl propionate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (25). The triethylammonium salt of compound **25** was obtained as a white solid (0.204 g, 51%) following the general procedure **(A)**, starting from TMP-triethylammonium salt (0.286 g, 0.55 mmol), compound **20** (0.553 g, 2.46 mmol), DCC (0.563 g, 2.73 mmol), triethylamine (0.38 mL, 2.73 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (11 mL) at 90 °C for 6 h. ^{1}H NMR (500 MHz, D_2O) δ = 7.75 (s, 1H, H-6), 6.35 (app t, *J* = 7.1 Hz, 1H, H-1'), 4.58-4.56 (m, 1H, H-3'), 4.13-4.12 (m, 1H, H-4'), 4.09-4.03 (m, 2H, OC*H*$_2$CH$_3$), 3.96-3.90 (m, 2H, H-5' and H-5''), 3.42-3.35 (m, 2H, C*H*$_{2\alpha}$CH$_2$SO$_3$H), 3.34-3.10 (m, 2H, C*H*$_{2\alpha}$CH$_2$CO$_2$Et), 3.12-3.07 (m, 2H, CH$_2$C*H*$_{2\beta}$SO$_3$H), 2.63-2.51 (m, 2H, CH$_2$C*H*$_{2\beta}$CO$_2$Et), 2.41-2.30 (m, 2H, H-2' and H-2''), 1.90 (s, 3H, CH$_3$), 1.19 (t, *J* = 7.1 Hz, 3H, OCH$_2$C*H*$_3$); ^{13}C NMR (125 MHz, D_2O) δ = 174.2 (CO_2Et), 165.9 (C-4), 151.2 (C-2), 136.8 (C-6), 111.2 (C-5), 84.9 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 84.2 (C-1'), 70.5 (C-3'), 63.3 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 61.1 (O*C*H$_2$CH$_3$), 49.5 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 41.7 (d, $^2J_{C,P}$ = 4.2 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 41.5 (d, $^2J_{C,P}$ = 4.2 Hz, *C*$_\alpha$H$_2$CH$_2$CO$_2$Et), 37.9 (C-2'), 33.2 (CH$_2$*C*$_\beta$H$_2$CO$_2$Et), 12.6 (OCH$_2$*C*H$_3$), 11.1 (CH$_3$); ^{31}P NMR (202 MHz, D_2O) δ = 7.7; HRMS for $C_{16}H_{26}N_3O_{12}PS$ [M-H]- calcd.: 528.1058, found: 514.1060.

2'-Deoxyadenosine-5'-*O*-[*N*-(ethyl propionate)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (26). The triethylammonium salt of compound **26** was obtained as a white solid (237 mg, 53%) following the general procedure **(A)**, starting from dAMP (200 mg, 0.60 mmol), compound **20** (611.4 mg, 2.72 mmol), DCC (623.1 mg, 3.02 mmol), triethylamine (0.42 ml, 3.02 mmol) in a 4:1 ^{t}BuOH/H$_2$O mixture (12 mL) at 90 °C for 6 h. ^{1}H NMR (300 MHz, D_2O) δ = 8.42 (s, 1H, H-8), 8.21 (s, 1H, H-2), 6.46 (app t, *J* = 6.6 Hz, 1H, H-1'), 4.74-4.69 (m, 1H, H-3'), 4.23-4.22 (m, 1H, H-4'), 3.98-3.94 (m, 2H, OC*H*$_2$CH$_3$), 3.93-3.89 (m, 2H, H-5' and H-5''), 3.32-3.23 (m, 2H, C*H*$_{2\alpha}$CH$_2$SO$_3$H), 3.18-3.12 (unresolved m, 2H, C*H*$_{2\alpha}$CH$_2$CO$_2$Et merged with Et$_3$N), 3.06-2.99 (m, 2H, CH$_2$C*H*$_{2\beta}$SO$_3$H), 2.90-2.81 (m, 1H, H-2'), 2.64-2.55 (m, 1H, H-2''), 2.42-2.27 (m, 2H, CH$_2$C*H*$_{2\beta}$CO$_2$Et), 1.09 (t, *J* = 7.1 Hz, 3H, OCH$_2$C*H*$_3$); ^{13}C NMR (75 MHz, D_2O) δ = 174.2 (CO_2Et), 155.2 (C-6), 152.4 (C-2), 148.4 (C-4), 139.5 (C-8), 118.3 (C-5), 85.6 (d, $^3J_{C,P}$ = 9.3 Hz, C-4'), 83.4 (C-1'), 70.8 (C-3'), 63.5 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 61.2 (O*C*H$_2$CH$_3$), 49.9 (CH$_2$*C*$_\beta$H$_2$SO$_3$H), 42.1 (d, $^2J_{C,P}$ = 4.7 Hz, *C*$_\alpha$H$_2$CH$_2$SO$_3$H), 41.9 (d, $^2J_{C,P}$ = 4.6 Hz, *C*$_\alpha$H$_2$CH$_2$CO$_2$Et), 38.5 (C-2'), 33.5 (CH$_2$*C*$_\beta$H$_2$CO$_2$Et), 12.8 (OCH$_2$*C*H$_3$); ^{31}P NMR (121 MHz, D_2O) δ = 8.1; HRMS for $C_{17}H_{27}N_6O_{10}PS$ [M-H]- calcd.: 537.1174, found: 537.1168.

General deprotection procedure for 2'-deoxynucleoside-5'-*O*-[*N*-(alkyl acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salts (10-11 and 27-30) synthesis (C)

A solution of 2'-deoxynucleoside-5'-*O*-[*N*-(ethyl alkylate)-2-aminoethane sulfonic acid] phosphoramidate trietylammonium salt (1 mmol) in a 0.4M NaOH solution in MeOH/H$_2$O (4:1v/v, 15 mL) was stirred at room temperature for 3 h and then 1M TEAB buffer (8 mL) was added. All the volatiles were removed first by rotary evaporation (bath temp 10 °C) and then by lyophilization. The crude residue was purified by RP-HPLC (50 mmol TEAB in 98% H$_2$O + 2% CH$_3$CN and 50 mmol TEAB in 50% H$_2$O + 50% CH$_3$CN) to give pure 2'-deoxynucleoside-5'-*O*-[*N*,*N*-(alkyl acid)-2-aminoethane sulfonic acid] phosphoramidate trietylammonium salt after the product was freeze-dried repeatedly to constant mass.

Thymidine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (10). The triethylammonium salt of compound 10 was obtained as a white solid (18.5 mg, 84%) following the general procedure (C), starting from compound 21 (20 mg, 0.028 mmol), 0.4M NaOH in MeOH/H$_2$O (4:1v/v, 0.42 mL) and 1M TEAB (0.22 mL). ^{1}H NMR (600 MHz, D_2O) δ = 7.78 (s, 1H, H-6), 6.37 (app t, *J* = 7.1 Hz, 1H, H-1'), 4.54-4.52 (m, 1H, H-3'), 4.14-4.13 (m, 1H, H-4'),

4.02-4.01 (m, 2H, H-5' and H-5''), 3.57 (d, J = 10.6 Hz, 2H, CH_2CO$_2$H), 3.39-3.35 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.13-3.10 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.38-2.28 (m, 2H, H-2' and H-2''), 1.91 (s, 3H, CH$_3$); ^{13}C NMR (150 MHz, D2O) δ = 178.6 (d, $^3J_{C,P}$ = 3.9 Hz, CO$_2$H), 166.4 (C-4), 151.6 (C-2), 137.1 (C-6), 111.6 (C-5), 85.5 (d, $^3J_{C,P}$ = 9.4 Hz, C-4'), 84.7 (C-1'), 71.1 (C-3'), 63.9 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 49.7 (d, $^2J_{C,P}$ = 4.0 Hz, CH_2CO$_2$H), 49.1 (CH$_2$C$_\beta$H$_2$SO$_3$H), 43.1 (d, $^2J_{C,P}$ = 3.2 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.4 (C-2'), 11.4 (-CH3); ^{31}P NMR (202 MHz, D$_2$O) δ = 7.8; HRMS for C$_{16}$H$_{26}$N$_3$O$_{12}$PS [M-H]- calcd.: 486.0589, found: 486.0591.

2'-Deoxyadenosine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (27). The triethylammonium salt of compound 27 was obtained as a white solid (41 mg, 81%) following the general procedure (C), starting from compound 22 (46 mg, 0.063 mmol), 0.4M NaOH in MeOH/H$_2$O (4:1v/v, 1.0 mL) and 1M TEAB (0.5 mL). ^{1}H NMR (500 MHz, D$_2$O) δ = 8.41 (s, 1H, H-8), 8.09 (s, 1H, H-2), 6.39 (t, J = 6.8 Hz, 1H, H-1'), 4.66-4.63 (m, 1H, H-3'), 4.21-4.20 (m, 1H, H-4'), 4.01-3.92 (m, 2H, H-5' and H-5''), 3.50 (d, J = 9.0 Hz, 2H, CH_2CO$_2$H), 3.35-3.30 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.03-3.00 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.80-2.75 (m, 1H, H-2'), 2.57-2.52 (m, 1H, H-2''); ^{13}C NMR (125 MHz, D$_2$O) δ = 178.2 (d, $^3J_{C,P}$ = 4.7 Hz, CO$_2$H), 154.6 (C-6), 151.7 (C-2), 147.8 (C-4), 139.1 (C-8), 117.7 (C-5), 85.4 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 82.9 (C-1'), 70.6 (C-3'), 63.3 (d, $^2J_{C,P}$ = 4.8 Hz, C-5'), 49.3 (d, $^2J_{C,P}$ = 3.7 Hz, CH_2CO$_2$H), 48.7 (CH$_2$C$_\beta$H$_2$SO$_3$H), 42.7 (d, $^2J_{C,P}$ = 4.0 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.3 (C-2'); ^{31}P NMR (202 MHz, D$_2$O) δ = 8.1; HRMS for C$_{10}$H$_{14}$N$_5$O$_9$PS [M-H]$^-$ calcd.: 495.0705, found: 495.0703.

2'-Deoxyguanosine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (28). The triethylammonium salt of compound **28** was obtained as an off-white solid (42 mg, 67%) following the general procedure **(C)**, starting from compound **23** (58 mg, 0.078 mmol), 0.4M NaOH in MeOH/H$_2$O (4:1v/v, 1.17 mL) and 1M TEAB (0.6 mL). ^{1}H NMR (500 MHz, D$_2$O) δ = 7.78 (s, 1H, H-8), 6.25 (app t, J = 6.9 Hz, 1H, H-1'), 4.64-4.62 (m, 1H, H-3'), 4.18-4.16 (m, 1H, H-4'), 4.00-3.95 (m, 2H, H-5' and H-5''), 3.52 (d, J = 8.8 Hz, 2H, CH_2CO$_2$H), 3.37-3.31 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.07-3.04 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.80-2.75 (m, 1H, H-2'), 2.48-2.43 (m, 1H, H-2''), 1.91 (s, 3H, CH$_3$); ^{13}C NMR (125 MHz, D2O) δ = 178.3 (d, $^3J_{C,P}$ = 4.8 Hz, CO$_2$H), 158.3 (C-6), 153.2 (C-2), 150.8 (C-4), 136.9 (C-8), 115.5 (C-5), 85.3 (d, $^3J_{C,P}$ = 9.0 Hz, C-4'), 82.9 (C-1'), 70.8 (C-3'), 63.5 (d, $^2J_{C,P}$ = 4.9 Hz, C-5'), 49.3 (d, $^2J_{C,P}$ = 3.8 Hz, CH_2CO$_2$H), 48.7 (CH$_2$C$_\beta$H$_2$SO$_3$H), 42.7 (d, $^2J_{C,P}$ = 4.4 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 37.9 (C-2'); ^{31}P NMR (202 MHz, D$_2$O) δ = 8.1; HRMS for C$_{14}$H$_{21}$N$_6$O$_{11}$PS [M-H]- calcd.: 511.0654, found: 511.0651.

2'-Deoxycytidine-5'-*O*-[*N*-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (29). The triethylammonium salt of compound **29** was obtained as a white solid (112 mg, 78%) following the general procedure **(C)**, starting from compound **24** (130 mg, 0.185 mmol), 0.4M NaOH in MeOH/H$_2$O (4:1v/v, 2.77 mL) and 1M TEAB (1.4 mL). ^{1}H NMR (500 MHz, D$_2$O) δ = 8.00 (d, J = 7.5 Hz, 1H, H-6), 6.37 (app t, J = 6.8 Hz, 1H, H-1'), 6.16 (d, J = 7.5 Hz, 1H, H-5), 4.58-4.55 (m, 1H, H-3'), 4.21-4.20 (m, 1H, H-4'), 4.09-4.07 (m, 2H, H-5' and H-5''), 3.60 (dd, J = 8.9, 1.8 Hz, 2H, CH_2CO$_2$H), 3.46-3.40 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.19-3.15 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.46-2.41 (m, 1H, H-2'), 2.34-2.29 (m, 1H, H-2''); ^{13}C NMR (125 MHz, D$_2$O) δ = 178.3 (d, $^3J_{C,P}$ = 4.9 Hz, CO$_2$H), 165.5 (C-4), 157.0 (C-2), 140.9 (C-6), 96.0 (C-5), 85.3 (C-1'), 85.2 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 70.4 (C-3'), 63.3 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 49.5 (d, $^2J_{C,P}$ = 3.5 Hz, CH_2CO$_2$H), 48.8 (CH$_2$C$_\beta$H$_2$SO$_3$H), 42.8 (d, $^2J_{C,P}$ = 4.3 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 39.0 (C-2'); ^{31}P NMR (202 MHz, D$_2$O) δ = 8.1; HRMS for C$_{13}$H$_{21}$N$_4$O$_{11}$PS [M-H]- calcd.: 471.0592, found: 471.0594.

Thymidine-5'-*O*-[*N*-(propionic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (11). The triethylammonium salt of compound **11** was obtained as a white solid (164 mg, 88%) following the general procedure **(C)**, starting from compound **25** (170 mg, 0.23 mmol), 0.4M NaOH in MeOH/H$_2$O (4:1v/v, 3.48 mL) and 1M TEAB (1.85 mL). ^{1}H NMR (500 MHz, DMSO-d_6) δ = 7.78 (d, J = 1.2 Hz, 1H, H-6), 6.20 (dd, J = 7.9, 6.0 Hz, 1H, H-1'), 4.27-4.25 (m, 1H, H-3'), 3.86-3.85 (m, 1H, H-4'), 3.74-3.72 (m, 2H, H-5' and H-5''), 3.17-3.13 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.10-3.05 (m, 2H, C$H_{2\alpha}$CH$_2$CO$_2$H), 2.60-2.57 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.41-2.38 (m, 2H, CH$_2$C$H_{2\beta}$CO$_2$H), 2.14-2.01 (m, 2H, H-2' and H-2''), 1.81 (d, J = 1.2 Hz, 3H, CH$_3$); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ = 173.7 (CO$_2$H), 163.9 (C-4), 150.6 (C-2), 136.2 (C-6), 109.9 (C-5), 86.1 (d, $^3J_{C,P}$ = 7.6 Hz, C-4'), 83.8 (C-1'), 71.1 (C-3'), 64.0 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 51.2 (CH$_2$C$_\beta$H$_2$SO$_3$H), 44.0 (d, $^2J_{C,P}$ = 3.0 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 43.0 (d, $^2J_{C,P}$ = 3.3 Hz, C$_\alpha$H$_2$CH$_2$CO$_2$H), 39.6 (C-2', merged with DMSO), 35.7 (CH$_2$C$_\beta$H$_2$CO$_2$H), 12.1 (CH3); ^{31}P NMR (202 MHz, DMSO-d$_6$) δ = 6.3; HRMS for C$_{16}$H$_{26}$N$_3$O$_{12}$PS [M-H]- calcd.: 500.0745, found: 500.0746.

2'-Deoxyadenosine-5'-*O*-[*N*-(propionic acid)-2-aminoethanesulfonic acid] phosphoramidate trietylammonium salt (30). The triethylammonium salt of compound **30** was obtained as a white solid (142 mg, 86%) following the general procedure **(C)**, starting from compound **26** (150 mg, 0.202 mmol), 0.4M NaOH in MeOH/H$_2$O (4:1v/v, 3.04 mL) and 1M TEAB (1.50 mL). ^{1}H NMR (500 MHz, D$_2$O) δ = 8.38 (s, 1H, H-8), 8.10 (s, 1H, H-2), 6.39 (app t, J = 6.6 Hz, 1H, H-1'), 4.69-4.67 (m, 1H, H-3'), 4.20 (br s, 1H, H-4'), 3.91-3.89 (m, 2H, H-5' and H-5''), 3.30-3.24 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.17-3.07 (m, 2H, C$H_{2\alpha}$CH$_2$CO$_2$H, merged with Et$_3$N), 3.03-2.99 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.82-2.76 (m, 1H, H-2'), 2.59-2.54 (m, 1H, H-2''), 2.31-2.28 (t, 2H, J = 7.9 Hz, CH$_2$C$H_{2\beta}$CO$_2$H); ^{13}C NMR (125 MHz, D$_2$O) δ = 180.2 (CO$_2$H), 154.7 (C-6), 151.8 (C-2), 147.7 (C-4), 139.0 (C-8), 117.8 (C-5), 85.3 (d, $^3J_{C,P}$ = 9.2 Hz, C-4'), 83.0 (C-1'), 70.4 (C-3'), 63.1 (d, $^2J_{C,P}$ = 5.1 Hz, C-5'), 49.5 (CH$_2$C$_\beta$H$_2$SO$_3$H), 43.1 (d, $^2J_{C,P}$ = 4.2 Hz, C$_\alpha$H$_2$CH$_2$CO$_2$H), 41.2 (d, $^2J_{C,P}$ = 4.6 Hz, C$_\alpha$H$_2$CH$_2$SO$_3$H), 38.3 (C-2'), 36.8 (CH$_2$C$_\beta$H$_2$CO$_2$H); ^{31}P NMR (202 MHz, D$_2$O) δ = 8.3; HRMS for C$_{15}$H$_{23}$N$_6$O$_{10}$PS [M-H]- calcd.: 509.0861, found: 509.0853.

Thymidine-5'-diphosphosphate triethylammonium salt (32). To a stirred suspension of 5'-*O*-tosyl-thymidine **31** (0.20 g , 0.50 mmol) in dry acetonitrile (0.5 mL) was added tris(terta-n-

butylammonium) hydrogen phosphate (0.682 g, 0.75 mmol). The resulting solution was stirred at room temperature for 36 h. Upon reaction completion, the mixture was concentrated *in vacuo*, diluted with water (5 mL) and lyophilized. The crude product was purified by Source 15Q ion-exchance column chromatography from a gradient of 0-100% 1M TEAB solution. The product was freeze-dried repeatedly to constant mass, yielding **32** as a white triethylammonium salt solid (0.239 g, 67%). ^{1}H NMR (300 MHz, D_2O) δ = 7.75 (d, J = 1.0 Hz, 1H, H-6), 6.35 (app t, J = 7.0 Hz, 1H, H-1'), 4.65-4.61 (m, 1H, H-3'), 4.18-4.14 (m, 3H, H-4', H-5' and H-5''), 2.39-2.34 (m, 2H, H-2' and H-2''), 1.92 (d, J = 1.0 Hz, 3H, CH_3); ^{13}C NMR (75 MHz, D_2O) δ = 166.2 (C-4), 151.4 (C-2), 137.0 (C-6), 111.4 (C-5), 85.1 (d, $^3J_{C,P}$ = 9.1 Hz, C-4'), 84.6 (C-1'), 70.7 (C-3'), 65.0 (d, $^2J_{C,P}$ = 5.5 Hz, C-5'), 38.2 (C-2'), 11.3 (CH_3); ^{31}P NMR (121 MHz, D_2O) δ = -10.9 (d, $J_{P,P}$ = 20.0 Hz, β-P), -11.5 (d, $J_{P,P}$ = 20.0 Hz, α-P) HRMS for $C_{10}H_{16}N_2O_{11}P_2$ [M-H]$^-$ calcd.: 401.0156, found: 401.0155.

Thymidine-5'-[β-*N*-(2-aminoethanesulfonic acid)] diphosphoramidate triethylammonium salt (33). The triethylammonium salt of compound **33** was obtained as a white solid (95 mg, 69%) following the general procedure **(A)**, starting from compound **32** (0.12 g , 0.17 mmol), taurine (0.085 g, 0.68 mmol), triethylamine (0.12 mL, 0.85 mmol) and DCC (0.175 g, 0.85 mmol) in a 4:1 ^{t}BuOH/H_2O mixture (4 mL) at 85 ºC for 4.5 h.^{1}H NMR (500 MHz, D_2O) δ = 7.76 (s, 1H, H-6), 6.32 (app t, J = 7.0 Hz, 1H, H-1'), 4.61-4.59 (m, 1H, H-3'), 4.15-4.14 (m, 1H, H-4'), 4.13-4.12 (m, 2H, H-5' and H-5''), 3.28-3.23 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.07 (t, J = 7.0 Hz, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.38-2.27 (m, 2H, H-2' and H-2''), 1.89 (s, 3H, CH_3); ^{13}C NMR (125 MHz, D_2O) δ = 166.0 (C-4), 151.1 (C-2), 136.8 (C-6), 111.1 (C-5), 84.8 (d, $^3J_{C,P}$ = 10.0 Hz, C-4'), 84.2 (C-1'), 70.4 (C-3'), 64.6 (d, $^2J_{C,P}$ = 5.0 Hz, C-5'), 51.8 (d, $^3J_{C,P}$ = 7.5 Hz, CH$_2$CβH$_2$SO$_3$H), 37.9 (C-2'), 36.6 (C_αH$_2$CH$_2$SO$_3$H), 11.1 (CH_3); ^{31}P NMR (202 MHz, D_2O) δ = -2.6 (d, $J_{P,P}$ = 22.0 Hz, β-P), -11.7 (d, $J_{P,P}$ = 22.0 Hz, α-P); HRMS for $C_{12}H_{21}N_3O_{13}P_2S$ [M-H]$^-$ calcd.: 508.0197, found: 508.0198.

Thymidine-5'-[γ-*N*-(2-aminoethanesulfonic acid)] triphosphoramidate triethylammonium salt (35). DCC (0.153 g, 0.744 mmol) was added to a stirred solution of dTTP triethylammonium salt **34** (0.11 g, 0.124 mmol) in dry DMF (2 mL) under an argon atmosphere and stirred for 3.5 h at room temperature. The dTTP solution was then added dropwise to a solution of taurine (0.093 g, 0.744 mmol) in a 1:1 miture of dry methanol and dry DMF (16 mL) and the reaction mixture was left to stir for 6 h at room temperture. The reaction was then diluted with water (50 mL) and the precipitate was filtered off. The clear solution was collected and lyophilized. The crude residue was redissolved in 0.01M TEAB buffer and purified by Source-15 Q ion-exchange resin column chromatography. The products adhering to the column were eluted with a linear buffer gradient of 0 to 1 M TEAB. The desired product was isolated by using 0.6–0.8M TEAB and it was then further purified by HPLC to give compound **35** as a white triethylammonium salt solid (0.022 g, 18%).^{1}H NMR (500 MHz, D_2O) δ = 7.64 (s, 1H, H-6), 6.35 (app t, J = 7.0 Hz, 1H, H-1'), 4.62-4.60 (m, 1H, H-3'), 4.18-4.11 (m, 3H, H-4', H-5' and H-5''), 3.29-3.24 (m, 2H, C$H_{2\alpha}$CH$_2$SO$_3$H), 3.00-2.96 (m, 2H, CH$_2$C$H_{2\beta}$SO$_3$H), 2.34-2.26 (m, 2H, H-2' and H-2''), 1.85 (s, 3H, CH_3); ^{13}C NMR (125 MHz, D_2O) δ = 173.1 (C-4), 156.5 (C-2), 136.0 (C-6), 111.4 (C-5), 84.5 (d, $^3J_{C,P}$ = 9.8 Hz, C-4'), 84.2 (C-1'), 70.3 (C-3'), 64.8 (d, $^2J_{C,P}$ = 5.5 Hz, C-5'), 51.9 (d, $^3J_{C,P}$ = 7.2 Hz, CH$_2$CβH$_2$SO$_3$H), 42.1 (C_αH$_2$CH$_2$SO$_3$H), 37.9 (C-2'), 11.9 (CH_3); ^{31}P NMR (202 MHz, D_2O) δ = -2.7 (d, $J_{P,P}$ = 20.6 Hz, γ-P), -12.0 (d, $J_{P,P}$ = 19.8 Hz, α-P), -23.3 (app t, $J_{P,P}$ = 20.2 Hz, β-P); HRMS for $C_{12}H_{22}N_3O_{16}P_3S$ [M-H]$^-$ calcd.: 587.9861, found: 587.9866.

General Protocol for DNA polymerase reaction

The primer P_1 was purchased from IDT, whilst all the templates $T_{1\text{-}5}$ were purchased from Eurogentec. Primer oligonucleotides were labeled with 5' [γ-33P]-ATP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's protocol. The labeled primers were further purified using Illustra MicroSpin G-25 Column (GE Healthcare) and then 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 polymerisation mixtures contained 125 nM primer-template complex, 1X reaction buffer (supplied with the enzyme), different concentrations of building blocks (125 μM, 500 μM and 1 mM) and 0.01 U.μl^{-1} Therminator, 0.01 U.μl^{-1} Vent (-exo) or 0.05 U.μl^{-1} Klenow (-exo) polymerases (New England Biolabs). The reaction was performed either at 37 °C (mesophilic polymerase) or at 75 °C (thermophilic polymerases) and aliquots were taken after 15, 30 and 60 min. In the control reaction, 50 μM of the corresponding natural deoxynucleoside triphosphate was used. All the polymerase reactions were quenched by addition of a double volume of gel loading buffer (90% formamide, 50mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol). Samples were heated either at 75 °C (mesophilic polymerase) or at 90 °C (thermophilic polymerases) for 5 min prior to separation on a 0.4mm 20% denaturing polyacrylamide gel. The bands were then visualized using phosphorimaging (Perkin Elmer).

Steady-state kinetics of single nucleotide incorporation

Primer 5' /5Cy5/CAGGAAACAGCTATGAC 3' was annealed with template 3' GTCCTTTGTCGATACTG**CAAAA** 5' in a 1:2 molar ratio by heating the mixture at 75 °C for 5 min, followed by slow cooling to room temperature. A series of reactions with different enzyme concentrations and different time points were performed to obtain the optimum conditions satisfying the 'single completed hit' principle[21]. The final DNA polymerisation mixtures each contained 125 nM primer-template complex, 1X reaction buffer (supplied with the enzyme), building blocks (for modified building blocks the concentration ranged between 15 μM and 1 mM, whilst for natural nucleoside triphosphates the concentration ranged between 25 nM and 2 μM) and 0.001 U.μl^{-1} Klenow fragment (New England Biolabs). The reactions were performed at 37 °C and aliquots were taken after 30 seconds. All polymerase reactions were quenched by adding a double volume of gel loading buffer (90 % formamide, 50 mM EDTA and

0.05% bromophenol blue) and heated at 90 °C for 5 min. The samples were separated on a 1 mm 15% denaturing polyacrylamide gel and gel bands were visualized using Ettan DIGE Imager (GE Healthcare). The gel bands were then quantified using ImageQuant TL 1D version 7.0 (GE Healthcare) and the kinetics parameters (V_{MAX} and K_M) were determined by fitting the data to a non-linear Michaelis–Menten regression using GraphPad Prism Software version 5.0.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) /ERC Grant agreement n° ERC-2012-ADG_20120216/ 320683

NOTES AND REFERENCES

[a] Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium; E-mail: Piet.Herdewijn@rega.kuleuven.be; Tel. +32 16 337387

[b] ISSB, Génopole genavenir6, Equipe Xénome, 5 rue Henri Desbruères 91030 Evry Cedex, France.

Electronic Supplementary Information (ESI) available on NMR spectra of the compounds and kinetic experiment data of enzymatic incorporation of compound **28** and dGTP. See DOI: 10.1039/b000000x/. See DOI: 10.1039/b000000x/

1. F. H. Westheimer, Science, 1987, 235, 1173-1178.
2. C. M. Joyce and T. A. Steitz, Annu. Rev. Biochem., 1994, 63, 777-822.
3. P. Herdewijn and P. Marlière, Chem. Biodivers., 2009, 6, 791-808.
4. P. Herdewijn and P. Marlière, FEBS Lett., 2012, 586, 2049-2056.
5. S. Q. Yang, M. Froeyen, E. Lescrinier, P. Marlière and P. Herdewijn, Org. Biomol. Chem., 2011, 9, 111-119.
6. M. A. Kertesz, Fems Microbiol. Rev., 2000, 24, 135-175.
7. J. R. van der Ploeg, E. Eichhorn and T. Leisinger, Arch. Microbiol., 2001, 176, 1-8.
8. J. R. van der Ploeg, M. A. Weiss, E. Saller, H. Nashimoto, N. Saito, M. A. Kertesz and T. Leisinger, J. Bacteriol., 1996, 178, 5438-5446.
9. J. R. van der Ploeg, R. Iwanicka-Nowicka, T. Bykowski, M. M. Hryniewicz and T. Leisinger, J. Biol. Chem., 1999, 274, 29358-29365.
10. E. Eichhorn, J. R. vanderPloeg, M. A. Kertesz and T. Leisinger, J. Biol. Chem., 1997, 272, 23031-23036.
11. S. Yang, M. Froeyen, E. Lescrinier, P. Marliere and P. Herdewijn, Organic & Biomolecular Chemistry, 2011, 9, 111-119.
12. X.-P. Song, C. Bouillon, E. Lescrinier and P. Herdewijn, Chem. Biodivers., 2012, 9, 2685-2700.
13. K. Ienaga, K. Higashiura, Y. Toyomaki, H. Matsuura and H. Kimura, Chem. Pharm. Bull. (Tokyo), 1988, 36, 70-77.
14. A. Giraut, X.-p. Song, M. Froeyen, P. Marlière and P. Herdewijn, Nucleic Acids Res., 2010, 38, 2541-2550.
15. X.-P. Song, C. Bouillon, E. Lescrinier and P. Herdewijn, Chembiochem, 2011, 12, 1867-1879.
16. W. Zhang and Z. Huang, Org. Lett., 2011, 13, 2000-2003.
17. V. J. Davisson, D. R. Davis, V. M. Dixit and C. D. Poulter, J. Org. Chem., 1987, 52, 1794-1801.
18. S. Martic, M. Labib, D. Freeman and H.-B. Kraatz, Chem. Eur. J., 2011, 17, 6744-6752.
19. A. Giraut, R. Abu El-Asrar, P. Marliere, M. Delarue and P. Herdewijn, Chembiochem, 2012, 13, 2439-2444.
20. B. Singer, F. Chavez, M. F. Goodman, J. M. Essigmann and M. K. Dosanjh, Proc. Natl. Acad. Sci. USA, 1989, 86, 8271-8274.
21. S. Creighton, Bloom, L. B., Goodman, M. F., Method. Enzymol, 1993, 232-256.