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

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Organic &amp; Biomolecular Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>OB-ART-01-2015-000157.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>19-Feb-2015</td>
</tr>
</tbody>
</table>

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, Marière, Philippe; ISSB, Génopole genavenir6, Equipe Xénome, Herdewijn, Piet; Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven,
Sulfonate derived phosphoramidates as active intermediates in the enzymatic primer-extension of DNA

S. De\textsuperscript{a}, E. Groaz\textsuperscript{a}, L. Margamuljana\textsuperscript{a}, M. Abramov\textsuperscript{a}, P. Marlière\textsuperscript{b} and P. Herdewijn\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 desulfitation 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 fulfill 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. Two operons, *tauABCD* and *susuEADCB*, 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.

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.

**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. 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 BuOH-H2O 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 Et3N (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).

**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 mesopholic 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.
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) NaHCO3, H2O, reflux, 72 h; (b) TMP, DCC, Et3N, BuOH: H2O (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 \( \beta \)-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 hydrolysis. 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 phosphoramidazolide 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-[\( \text{N}-(\text{alkyl acid}) \)]-2-aminoethane sulfonic acid] phosphoramidates was prepared 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 \( \text{N}-(\text{ethyl acrylate}) \)-2-aminoethane sulfonic acid 19, following the nucleophilic displacement of the bromo atom in ethylbromoacetate by taurine, occurred only upon heating, whilst \( \text{N}-(\text{ethyl acrylate}) \)-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 thiotriethylammonium 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 \( \text{N}-(\text{ethyl acrylate}) \)-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 phosphoramidazolide-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.
Next, Tau-dTDP β-phosphoramide 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, 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

Finally, the synthesis of Tau-dTTP γ-phosphoramide 35 was accomplished through initial conversion of dTTP-Et3N salt 34 to thymidine-5'-trimetaphosphates by dehydration with DCC, 18 followed by ring opening of thymidine-5'-trimetaphosphate with an excess of taurine.

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 γ32P 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.

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

| Table 1. Overview of primer and templates sequences used in the incorporation experiments. |
|-------------------------------|---------------------------------|-----------------|----------------|-----------------|
| Primer | P1: 5'/Cy5/CAGGAAAAACGCTATGAC 3' | Template T1: 3' GTCTTGTTCGATCTGAAATAA 5' | 5' | Template T2: 3' GTCTTGTTCGATCTGTTTTTTT 5' |
| Template T1: 3' GTCTTGTTCGATCTGAAATAA 5' | | | |
| Template T3: 3' GTCTTGTTCGATCTGAGGAGG 5' | | | |
| Template T4: 3' GTCTTGTTCGATCTGCCCCC 5' | | | |
| Template T5: 3' GTCTTGTTCGATCTGACAAAA 5' | | | |

Chain elongation experiments

Scheme 5. Synthesis of sulfono-phosphoramidate nucleoside analogues 10-11 and 27-30. (a) Ethylbromo acetate, NaNHCO3, Dioxane/H2O, 70 °C, 24 h; (b) Ethyl acrylate, NaNH2, EtOH/H2O (1:1), r.t., 16 h; (c) dTMP-Et3N salt, DCC, Et3N, BuOH-H2O (4:1), 90 °C, 6 h; (d) Nucleoside-Imidazolide Na-Salt, Et3N, DMF, 35 °C, 168 h; (e) 0.4 M NaOH in MeOH/H2O (4:1 v/v), r.t., 3 h.

Scheme 6. Synthesis of taurine diphosphoramidate analogue 33 of thymidine. (a) Tris(tetra-n-butylammonium) hydrogen pyrophosphate, CH3CN, r.t., 24 h; (b) Taurine, DCC, Et3N, BuOH-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.

ARTICLE
Journal Name
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, 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 phosphoramidates 2 and 3, and incorporation efficiencies with the same polymerase were relatively higher for adenine analogous 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.

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 \( \Lambda \geq T = G > C \), 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 deoxyribo nucleoside 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 \( G > A \geq T \geq C \) 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 \( O^\prime \)-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, Taur-
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.

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 \( P+5 \) 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 sulfonyl or both carboxylate and sulfonyl 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 mesophytic 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 °C). 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 CH3CN/H2O gradients for RP-HPLC or TEAB -CH2SO3H, 2.33-2.27 (2H, H-2' and H-2''), 1.85 (d, J = 1.0 Hz, 3H, CH3-THy); 13C NMR (75 MHz, D2O) δ = 178.6 (-COH); 166.3 (C-4'), 151.5 (C-2'), 137.1 (C-6'), 85.5 (d, Jc,p = 9.3 Hz, C-5'), 84.7 (4C-1'), 70.9 (C-3'), 63.8 (d, Jc,p = 4.9 Hz, C-5'), 52.0 (d, Jc,p = 6.3 Hz, -CH2CH2SO3H); 31P NMR (121 MHz, D2O) δ = 8.0; HRMS for C12H20N3O10PS [M-H]- calcd.: 428.0534, found: 428.0531.

**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 BuOH/H2O 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 [DCC (0.34g, 1.64 mmol) in a 4:1 BuOH/H2O mixture (15 mL) at 90 ºC for 3 h. HRMS for C12H20N3O10PS [M-H]- calcd.: 428.0534, found: 428.0531.](31P NMR (121 MHz, D2O) δ = 8.0; HRMS for C12H20N3O10PS [M-H]- calcd.: 428.0534, found: 428.0531.)

**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 BuOH/H2O mixture (15 mL) at 90 °C for 3 h. HRMS for C12H20N3O10PS [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 BuOH/H2O mixture (14 mL) at 90 °C for 3 h. HRMS for C12H20N3O10PS [M-H]- calcd.: 428.0534, found: 428.0531.

**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 BuOH/H2O mixture (14 mL) at 90 °C for 3 h. HRMS for C12H20N3O10PS [M-H]- calcd.: 428.0534, found: 428.0531.

**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 °C). 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 CH3CN/H2O gradients for RP-HPLC or TEAB -CH2SO3H, 2.33-2.27 (2H, H-2' and H-2''), 1.85 (d, J = 1.0 Hz, 3H, CH3-THy); 13C NMR (75 MHz, D2O) δ = 178.6 (-COH); 166.3 (C-4'), 151.5 (C-2'), 137.1 (C-6'), 85.5 (d, Jc,p = 9.3 Hz, C-5'), 84.7 (4C-1'), 70.9 (C-3'), 63.8 (d, Jc,p = 4.9 Hz, C-5'), 52.0 (d, Jc,p = 6.3 Hz, -CH2CH2SO3H); 31P NMR (121 MHz, D2O) δ = 8.0; HRMS for C12H20N3O10PS [M-H]- calcd.: 428.0534, found: 428.0531.

**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 BuOH/H2O 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 (iProH2O; Et3N/NH3 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:H2O:NH3 20:1-1, v/v/v, 15:1:1, v/v/v, 10:1:1, v/v/v (Et3N was used instead of NH3 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 H2O + 2% CH3CN and 50 mmol TEAB in 50% H2O + 50% CH3CN) 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.
The triethylammonium salt of compound 6 (0.243 g, 53%) was obtained as a white solid (0.39 g, 74%) following the general procedure (A), starting from TMP-triethylammonium salt (0.286 g, 0.55 mmol), compound 14 (0.76 g, 2.73 mmol), triethylamino (0.91 mL, 6.55 mmol) and DCC (0.68 g, 3.28 mmol) in a 4:1 BuOH/H2O mixture (9 mL) at 90 °C for 5.5 h. 1H NMR (500 MHz, D2O) 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’), 1.48 (br s, 1H, H-4’), 4.05-3.99 (m, 2H, H-2’-H’-2”), 3.43-3.36 (m, 4H, CH2CH3SO3H), 3.17 (td, J = 8.0 Hz, 4H, CH2CH3SO3H), 2.40-2.37 (m, 2H, H-2’-H’-2”), 1.96 (sd, 3H, CH3-Tyr); 13C NMR (125 MHz, D2O) δ = 166.0 (C-4), 151.2 (C-2), 136.7 (C-6), 111.3 (C-5), 84.9 (d, dJCP = 9.1 Hz, C-4”), 84.4 (C-1”), 70.6 (C-3”), 63.5 (d, dJCP = 5.3 Hz, C-5”), 48.8 (d, dJCP = 2.9 Hz, CH2CH3SO3H), 44.5 (d, dJCP = 4.3 Hz, CH2CH3SO3H), 38.0 (C-2”), 33.0 (d, dJCP = 3.5 Hz, N-CH3), 11.1 (CH3-Tyr); 31P NMR (202 MHz, D2O) δ = 8.9; HRMS for C13H25N2O3PS2 [M+H]+ calcd.: 536.0415, found: 536.0418.

Thymidine-5’-(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 (0.42 g, 0.88 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 BuOH/H2O mixture (12 mL) at 90 °C for 5.5 h. 1H NMR (500 MHz, D2O) 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, CH2CH3SO3H), 3.08 (td, J = 7.0 Hz, 2H, CH2CH2SO3H), 2.58 (d, JCH = 9.5 Hz, 3H, N-CH3), 2.34-2.31 (m, 2H, H-2’-H’-2”), 1.89 (s, 3H, CH3-Tyr); 31C NMR (125 MHz, D2O) δ = 166.0 (C-4), 151.1 (C-2), 136.7 (C-6), 111.1 (C-5), 84.9 (d, dJCP = 9.1 Hz, C-4”), 84.3 (C-1”), 70.6 (C-3”), 63.5 (d, dJCP = 5.3 Hz, C-5”), 48.8 (d, dJCP = 2.9 Hz, CH2CH3SO3H), 44.5 (d, dJCP = 4.3 Hz, CH2CH3SO3H), 38.0 (C-2”), 33.0 (d, dJCP = 3.5 Hz, N-CH3), 11.1 (CH3-Tyr); 31P NMR (202 MHz, D2O) δ = 8.9; HRMS for C13H25N2O3PS2 [M+H]+ calcd.: 536.0415, found: 536.0418.

General procedure for taurine-amino acid syntheses 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 °C, 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 ethylamine (3 mL/mmol), and to this solution was added a solution of sodium vinyl sulfonate (6 mL/mmol) 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 H2N-aminoacid-Taurine-OH as a white crystalline solid.

H-Gly-Tau-OH (16a). Compound 16a was obtained as a white solid (0.32 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). 1H NMR (300 MHz, D2O) δ = 3.82 (s, 2H, CH2-Gly), 3.67 (s, J = 6.0 Hz, 2H, CH2CH2CH2SO3H), 3.13 (t, J = 6.8 Hz, 2H, CH2CH2CH2SO3H); 13C NMR (75 MHz, D2O) δ = 166.5 (CO-Gly), 49.1 (CH2CH2CH2SO3H), 40.1 (CH2-Gly), 34.8 (CH2CH2CH2SO3H); HRMS for C14H22N2O12S [M-H] + calcd.: 561.1293, found: 561.1292.

H-Asp-Tau-OH (18). To a stirred solution of 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, Et2N (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:H2O 100:0, v/v; 30:1, v/v; 20:1, v/v) to give compound 17 (2.89 g, 89%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ = 8.06 (t, J = 5.2 Hz, 1H, NH-Tau), 7.67 (d, J = 8.3 Hz, 1H, NH-Asp); 13C NMR (75 MHz, D2O) δ = 170.2 (CO-Asp), 169.8 (CO-Asp), 155.8 (COOH), 136.9 (1C of OCH2Ph), 136.1 (1C of OCH2Ph), 128.4 (Ar-C), 127.9 (Ar-C), 127.8 (Ar-C), 127.7 (Ar-C), 65.6 (2 x OCH2Ph), 51.5 (α-C-Asp), 50.1 (CH2CH2SO3H), 36.4 (β-C-Asp), 35.7 (CH2CH2SO3H); HRMS for C21H22N3O12S [M-] + calcd.: 463.1180, found: 463.1181.

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, Et2N (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:H2O/Et2N 20:1:5:0.5, v/v/v/v; 15:1:5:0.5, v/v/v/v).
Thymidine-5'-[Asp-Tau-OH] phosphoramidite 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'-phosphoramidite-sodium salt (0.15 g, 0.38 mmol), compound 18 (0.229 g, 0.951 mmol) and Et3N (0.41 mL, 3.04 mmol) in DMF (4 mL).1H NMR (500 MHz, D2O) 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, oCH2-Asp), 3.64-3.60 (m, 2H, CH2CH2SO3H), 3.13 (t, J = 7.0 Hz, 2H, CH2CH2SO3H), 2.71 (dd, J = 16.0, 5.4 Hz, 1H, βCH'-Asp), 2.60 (dd, J = 15.7, 5.1 Hz, 1H, βCH'-Asp), 2.40-2.37 (m, 2H, H-2' and H-2''), 1.95 (s, 3H, CH3-Thy); 13C NMR (125 MHz, D2O) δ = 177.7 (COH), 175.6 (d, 2JC,P = 6.6 Hz, oC=O-Asp), 166.0 (C-4), 151.2 (C-1'), 136.8 (C-3'), 111.1 (C-5), 85.1 (d, 2JC,P = 9.2 Hz, C-4'), 84.6 (C-1'), 70.8 (C-3'), 52.6 (d, 2JC,P = 4.6 Hz, C-5'), 49.0 (CH2CH2SO3H), 39.9 (d, 2JC,P = 3.5 Hz, βCH2-Asp), 38.1 (C-2'), 34.5 (CH3CH2SO3H), 11.1 (CH3-Thy); 31P NMR (121 MHz, D2O) (merged with Et3N); 13C NMR (75 MHz, D2O) δ = 6.6 (2JC,P = 9.4 Hz, C-4'), 84.6 (C-1'), 71.0 (C-3'), 63.9 (d, 2JC,P = 5.1 Hz, C-5'), 61.5 (OCH2CH2SO3H), 49.7 (CH2CH2SO3H), 48.5 (d, 2JC,P = 5.5 Hz, CH2CO2Et), 43.4 (d, 2JC,P = 4.7 Hz, CH3CH2SO3H), 38.5 (C-2'), 13.0 (OCH2CH2), 11.9 (CH3); 31P NMR (121 MHz, D2O) δ = 7.1; HRMS for C6H13NO5S [M-H]– calcd.: 210.0439. The triethylammonium salt of compound 21 was obtained as a white solid (123 mg, 52%) following the general procedure (B), starting from thymidine-5'-phosphoramidite-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). 1H NMR (300 MHz, D2O) δ = 7.69 (s, 1H, H-6), 6.37 (d, J = 7.1 Hz, 1H, H-1'), 4.58-4.54 (m, 1H, H-3'), 4.17-4.19 (m, 3H, H-4' and OCH2CH3), 4.05-3.98 (m, 2H, H-5' and H-5''), 3.85 (d, J = 10.6 Hz, 2H, CH2CO2Et), 3.48-3.37 (m, 2H, CH2CH2SO3H), 3.18-3.10 (m, 2H, CH2CH2SO3H) (merged with Et3N), 2.37-2.32 (m, 2H, H-2' and H-2''), 1.92 (s, 3H, CH3), 1.27-1.22 (m, 3H, OCH2CH2) (merged with Et3N); 13C NMR (75 MHz, D2O) δ = 173.5 (d, 2JC,P = 2.4 Hz, CO2Et), 171.0 (C-4), 155.1 (C-2), 136.5 (C-6), 111.5 (C-5), 85.0 (d, 2JC,P = 9.4 Hz, C-4'), 84.6 (C-1'), 71.0 (C-3'), 63.9 (d, 2JC,P = 5.1 Hz, C-5'), 61.6 (OCH2CH2), 49.7 (CH2CH2SO3H), 48.5 (d, 2JC,P = 5.5 Hz, CH2CO2Et), 43.4 (d, 2JC,P = 4.7 Hz, CH3CH2SO3H), 38.5 (C-2'), 13.0 (OCH2CH2), 11.9 (CH3); 31P NMR (121 MHz, D2O) δ = 7.1; HRMS for C7H15NO5S [M-H]– calcd.: 224.0598, found: 224.0592. 2'-Deoxyadenosine-5'-(Ethyl acetate)-2-aminoethanesulfonic acid phosphoramidate triethylammonium 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'-phosphoramidite-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). 1H NMR (300 MHz, D2O) δ = 8.40 (s, 1H, H-8), 8.13 (s, 1H, H-2'), 6.40 (d, 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 OCH2CH3), 3.70 (d, J = 10.5 Hz, 2H, CH2CO2Et), 3.41-3.28 (m, 2H, CH2CH2SO3H), 3.11-3.03 (m, 2H, CH2CH2SO3H), 2.91-2.74 (m, 1H, H-2'), 2.64-2.53 (m, 1H, H-2''), 1.13 (t, J = 7.1 Hz, 3H, OCH2CH2); 13C NMR (75 MHz, D2O) δ = 173.3 (d, 2JC,P = 2.8 Hz, CO2Et), 155.0 (C-6), 152.2 (C-2), 148.1 (C-4), 139.4 (C-8), 118.1 (C-5), 85.7 (d, 2JC,P = 9.3 Hz, C-4'), 83.3 (C-1'), 71.1 (C-3'), 63.8 (d, 2JC,P = 5.2 Hz, C-5'), 61.5 (OCH2CH3), 49.7 (CH2CH2SO3H), 48.4 (d, 2JC,P = 5.3 Hz, CH2CO2Et), 43.4 (d, 2JC,P = 4.6 Hz, CH3CH2SO3H), 38.6 (C-2'), 12.3 (OCH2CH2); 31P NMR (121 MHz, D2O) δ = 7.2; HRMS for C8H16N2O9P3S [M-H]– calcd.: 523.1017, found: 523.1019. 2'-Deoxyguanosine-5'-(Ethyl acetate)-2-aminoethanesulfonic acid phosphoramidate triethylammonium 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). 

1H NMR (300 MHz, D2O) 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, OCH3CH=), 3.96-3.93 (m, 3H, H-2' and H-5' and H-5''), 3.74 (d, J = 10.5 Hz, 2H, CH2CO2Et), 3.40-3.33 (m, 2H, CH2CH2SO3H), 3.01-2.96 (m, 2H, CH2CH2SO3H), 2.83-2.71 (m, 1H, H-2'), 2.50-2.43 (m, 1H, H-2''), 1.13 (t, J = 7.4 Hz, 3H, OCH3CH=).

2JC,P = 5.5 Hz, 1H, H-1'), 4.74-4.69 (m, 1H, H-1'), 4.23-4.22 (m, 1H, H-4'), 3.98-3.94 (m, 2H, OCH3CH=), 3.83-3.79 (m, 2H, CH2CH2SO3H), 3.19-3.15 (m, 2H, CH2CH2SO3H), 2.83-2.71 (m, 1H, H-2'), 2.50-2.43 (m, 1H, H-2''), 1.13 (t, J = 7.4 Hz, 3H, OCH3CH=).

2'-Deoxyadenosine-5'-O-[N-(ethyl acetate)-2- aminoethanesulfonic acid] phosphoramidate triethylammonium 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'-deoxyguanosine-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). 

1H NMR (300 MHz, D2O) 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'), 4.17-4.14 (m, 1H, H-3' and OCH3CH=), 4.03-3.98 (m, 2H, H-2' and H-5'), 3.82 (d, J = 10.7 Hz, 2H, CH2CO2Et), 3.44-3.35 (m, 2H, CH2CH2SO3H), 3.19-3.09 (unresolved m, 2H, CH2CH2SO3H, merged with Et3N), 2.43-2.21 (m, 2H, H-2' and H-2''), 1.27-1.22 (unresolved m, 3H, OCH3CH=, merged with Et3N); 13C NMR (75 MHz, D2O); 31P NMR (121 MHz, D2O) δ = 7.2; HRMS for C16H25N6O11PS [M-H]− calcd.: 539.0967, found: 539.0969.

Thymidine-5'-O-[N-(ethyl propionate)-2-aminoethanesulfonic acid] phosphoramidate triethylammonium 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 (1 mmol) in a 0.4M NaOH solution in MeOH/H2O (4:1v/v, 0.42 mL) and 1M TEAB (0.22 mL). 

1H NMR (300 MHz, D2O) 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, OCH3CH=), 3.93-3.89 (m, 2H, H-5' and H-5''), 3.32-3.23 (m, 2H, CH2CH2SO3H), 3.18-3.12 (unresolved m, 2H, CH2CH2SO3H merged with Et3N), 2.83-2.71 (m, 1H, H-2'), 2.43-2.21 (m, 2H, H-2' and H-2''), 1.27-1.22 (unresolved m, 3H, OCH3CH=, merged with Et3N); 13C NMR (75 MHz, D2O) δ = 174.2 (CO2Et), 155.2 (C-6), 152.4 (C-2), 148.4 (C-4), 139.5 (C-8), 118.3 (C-5), 85.6 (d, J2,C,P = 9.3 Hz, C-4'), 83.4 (C-1'), 70.8 (C-3'), 63.5 (d, J2,C,P = 5.0 Hz, C-5'), 61.2 (OCH3CH=), 49.9 (CH2CH2SO3H), 42.1 (d, J2,C,P = 4.7 Hz, C2H2CH2SO3H), 41.9 (d, J2,C,P = 4.6 Hz, C2H2CH2SO3H), 38.5 (C-2'), 33.5 (CH2CH2SO3H), 12.8 (OCH3CH=); 31P NMR (121 MHz, D2O) δ = 8.1; HRMS for C17H23N6O19P [M-H]− calcd.: 537.1174, found: 537.1168.

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

A solution of 2'-deoxyxynucleoside-5'-O-[N-(ethyl alkylate)-2-aminoethanesulfonic acid] phosphoramidate triethylammonium salt (1 mmol) in a 0.4M NaOH solution in MeOH/H2O (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% H2O + 2% CH3CN and 50 mmol TEAB in 98% H2O + 2% BuOH/H2O mixture (12 mL) at 90 °C for 6 h).

Thymidine-5'-O-[N-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate triethylammonium 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/H2O (4.1v/v, 0.42 mL) and 1M TEAB (0.22 mL).
The triethylammonium salt of compound 29 was obtained as a white solid (112 mg, 78%) following the general procedure (C), starting from compound 23 (130 mg, 0.185 mmol), 0.4M NaOH in MeOH/H2O (4:1 v/v, 2.77 mL) and 1M TEAB (1.4 mL). 1H NMR (500 MHz, D2O) δ = 8.00 (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, CH2CHOH), 3.46-3.40 (m, 2H, CH2CH2SO3H), 3.19-3.15 (m, 2H, CH2CH2SO3H), 2.46-2.41 (m, 1H, H-2′′), 2.34-2.29 (m, 1H, H-2′′′), 1.33 (C, CH2CH2SO3H) δ = 177.5 (CO2H), 165.5 (C-4), 157.0 (C-2), 150.9 (C-6), 134.4 (C-5), 83.5 (C-1′), 82.8 (C-3′), 71.1 (C-3′′), 71.0 (C-3′′′), 63.5 (d, J = 4.9 Hz, C-5′′′), 49.3 (d, J = 4.3 Hz, C2′′′), 37.9 (C-2′′′). 31P NMR (202 MHz, D2O) δ = 8.3; HRMS for C14H21N6O11PS [M-H]-: 509.0861, found: 509.0853.

Thymidine-5′-O-[N-(propionic acid)-2-aminoethanesulfonic acid] phosphoramidate triethylammonium 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/H2O (4:1 v/v, 3.04 mL) and 1M TEAB (1.50 mL). 1H NMR (500 MHz, D2O) δ = 7.78 (d, J = 1.2 Hz, 1H, H-6), 6.60 (dd, J = 7.9, 6.0 Hz, 1H, H-1′), 4.27-4.25 (m, 1H, H-3′), 3.86-3.85 (m, 2H, H-4′′ and H-5′′), 3.17-3.13 (m, 2H, CH2CH2SO3H), 3.10-3.05 (m, 2H, CH2CH2CH(OH)2), 2.60-2.57 (m, 2H, CH2CH2SO3H), 2.41-2.38 (m, 2H, CH2CH2COH), 2.14-2.03 (m, 1H, H-2′′ and H-2′′′), 1.81 (d, J = 1.2 Hz, 3H, CH3); 13C NMR (125 MHz, D2O) δ = 173.7 (CO2H), 163.9 (C-4), 150.6 (C-5), 138.1 (C-2), 120.9 (C-3′), 119.1 (C-3′′), 117.7 (C-5), 84.5 (d, J = 9.2 Hz, C-4′), 82.9 (C-1′), 70.6 (C-3′′), 63.5 (d, J = 4.8 Hz, C-4′′′), 49.3 (d, J = 3.7 Hz, CH2SO3H), 48.7 (CH2CH2SO3H), 42.7 (d, J = 4.0 Hz, CH2CH2SO3H), 38.3 (C-2′′), 231P NMR (202 MHz, D2O) δ = 8.3; HRMS for C16H26N3O12PS [M-H]-: 511.0745, found: 511.0743.

Thymidine-5′-O-[N-(acetic acid)-2-aminoethanesulfonic acid] phosphoramidate triethylammonium salt (31). The triethylammonium salt of compound 31 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/H2O (4:1 v/v, 3.48 mL) and 1M TEAB (1.85 mL). 1H NMR (500 MHz, D2O) 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, 2H, H-4′′ and H-5′′), 3.17-3.13 (m, 2H, CH2CH2SO3H), 3.10-3.05 (m, 2H, CH2CH2CH(OH)2), 2.60-2.57 (m, 2H, CH2CH2SO3H), 2.41-2.38 (m, 2H, CH2CH2COH), 2.14-2.03 (m, 1H, H-2′′ and H-2′′′), 1.81 (d, J = 1.2 Hz, 3H, CH3); 13C NMR (125 MHz, DMSO-d6) δ = 173.7 (CO2H), 163.9 (C-4), 150.6 (C-5), 138.1 (C-2), 120.9 (C-3′), 119.1 (C-3′′), 117.7 (C-5), 84.5 (d, J = 9.2 Hz, C-4′), 82.9 (C-1′), 70.6 (C-3′′), 63.5 (d, J = 4.8 Hz, C-4′′′), 49.3 (d, J = 3.7 Hz, CH2SO3H), 48.7 (CH2CH2SO3H), 42.7 (d, J = 4.0 Hz, CH2CH2SO3H), 38.3 (C-2′′), 231P NMR (202 MHz, D2O) δ = 8.3; HRMS for C16H26N3O12PS [M-H]-: 511.0745, found: 511.0743.
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-exchange 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%). ¹H NMR (300 MHz, D₂O) 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, CH₃). 13C NMR (125 MHz, D₂O) 100.53 (d, J = 10.0 Hz, C-4”), 84.5 (d, J = 10.2 Hz, C-5’), 84.2 (C-1’), 70.5 (C-3’), 64.6 (d, J = 20.0 Hz, C-5”), 51.8 (d, J = 20.2 Hz, C-5”). 31P NMR (121 MHz, D₂O) δ = 20.0 Hz, (α-P), -11.5 (d, J = 20.0 Hz, α-P). HRMS for C₉H₁₆N₂O₁₃P₂ [M-H]- calcd.: 508.0197, found: 508.0198. General Protocol for DNA polymerase reaction The primer P₁ was purchased from IDT, whilst all the templates T₁-5 were purchased from Eurogentec. Primer oligonucleotides were labelled 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 Thermotinder, 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 polymerases) 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’-5CAGGAAAACAGCTATGAC 3’ was annealed with template 3’-GTCCTTTGTCCGACTTGCAGAAA 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. 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 4.11 (m, 3H, H-4’, H-5’ and H-5”), 3.29-3.24 (m, 2H, CH₂CH₂SH2O₃H); 3.00-2.96 (m, 2H, CH₂CH₂SH2O₃H); 2.34-2.26 (m, 2H, H-2’ and H-2’”). 1.85 (s, 3H, CH₃); ¹⁳C NMR (125 MHz, D₂O) δ = 173.1 (C-4), 156.5 (C-2), 136.0 (C-6), 111.4 (C-5), 84.5 (d, J = 9.8 Hz, C-4”), 84.2 (C-1’), 70.3 (C-3’), 64.8 (d, J = 5.5 Hz, C-5”), 51.9 (d, J = 7.2 Hz, CH₂CH₂SH2O₃H), 42.1 (CH₂CH₂SH2O₃H), 37.9 (C-2”), 36.6 (CH₂CH₂SH2O₃H), 11.1 (CH₃); ³¹P NMR (202 MHz, D₂O) δ = -2.7 (d, J = 20.6 Hz, γ-P), -12.0 (d, J = 19.8 Hz, α-P), -23.3 (app t, J = 20.2 Hz, β-P); HRMS for C₁₂H₂₁N₇O₁₆P₂S [M-H]- calcd.: 587.9861, found: 587.9866.

Thymidine-5’-[β-N(2-aminooethanesulfonic acid)] diphenosphoridate triethylammonium salt (33). The triethylammonium salt of compound 33 was obtained as a white solid (95 mg, 69%) following the general procedure 2, starting from compound 32 (0.12 g, 0.17 mmol), taurine (0.085 g, 0.68 mmol), triethyamine (0.12 mL, 0.85 mmol) and DCC (0.175 g, 0.85 mmol) in a 4:1 BuOH/H₂O mixture (4 mL) at 85 °C for 4.5 h. ¹H NMR (500 MHz, D₂O) δ = 7.76 (s, 8H, H-6), 6.35 (app t, J = 7.0 Hz, 1H, H-1’), 4.15-4.14 (m, 1H, H-4’), 4.13-4.12 (m, 2H, H-2’ and H-2’”), 1.82 (s, 3H, CH₃); 13C NMR (125 MHz, D₂O) δ = 166.2 (C-4), 151.4 (C-2), 137.0 (C-6), 111.4 (C-5), 85.1 (d, J = 9.1 Hz, C-4”), 84.6 (C-1’), 70.7 (C-3’), 65.0 (d, J = 2.5 Hz, C-5’), 38.2 (C-2’), 11.3 (CH₃); ³¹P NMR (121 MHz, D₂O) δ = -10.9 (d, J = 20.0 Hz, β-P), -11.5 (d, J = 20.0 Hz, α-P). HRMS for C₁₀H₁₆N₂O₁₁P₂ [M-H]- calcd.: 401.0156, found: 401.0155. Thymidine-5’-[γ-N(2-aminooethanesulfonic acid)] triphenosphoridate 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 mixture of dry DMF and D₂O (8 mL) under an argon atmosphere and stirred for 3.5 h at room temperature. 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 0.022 g, 18%). ¹H NMR (500 MHz, D₂O) 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-
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 (VMAX and KM) 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/