Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2011, **9**, 111

www.rsc.org/obc

PAPER

3-Phosphono-L-alanine as pyrophosphate mimic for DNA synthesis using HIV-1 reverse transcriptase†

Shiqiong Yang,[a] Mathy Froeyen,[a] Eveline Lescrinier,[a] Philippe Marlière[b] and Piet Herdewijn*[a]

Received 9th August 2010, Accepted 29th September 2010
DOI: 10.1039/c0ob00554a

A series of sulf(on)ate and phosph(on)ate amino acid phosphoramidate analogues of deoxynucleotides were synthesized as potential substrates for HIV-1 reverse transcriptase. Taurine, L-cysteic acid, 3-phosphono-L-alanine, *O*-sulfonato-L-serine, and *O*-phospho-L-serine were investigated as leaving groups in an enzyme catalyzed DNA synthesis protocol. Among these analogues, the phosphonate congener performed best and 3-phosphono-L-alanine can be considered as an excellent mimic of the pyrophosphate (PPi) moiety of deoxyadenosine triphosphate, to be used in enzymatic synthesis of nucleic acids. During a single nucleotide incorporation assay the use of 3-phosphono-L-Ala-dAMP as substrate resulted in 95% conversion to a P + 1 strand in 60 min at 50 μM (a concentration 10 times less than found for L-Asp-dAMP) and with improved incorporation kinetics and less stalling. For the sequences investigated, the efficiency of the incorporation is base dependent and decreases in the order (A ≥ T = G > C). In all cases, the incorporation follows Watson–Crick rules.

Introduction

Nucleoside triphosphates are the building blocks for the enzymatic synthesis of nucleic acids, using polymerases as catalysts. In this reaction, a nucleoside monophosphate is incorporated in a growing nucleic acid chain and pyrophosphate functions as a leaving group. The mechanism of the polymerization reaction and the structural changes involved seem to be conserved between different polymerases.[1] Several mimics of nucleoside triphosphates have been tested as substrates for DNA polymerases. Most of these triphosphate mimics are obtained by modifying the bridging O-atoms of the triphosphate moiety[2] or using non-standard nucleobases.[3] In previous studies, we have demonstrated that the pyrophosphate moiety of nucleoside triphosphate can be replaced by leaving groups whose structures are not based on phosphate chemistry. For example, some selected amino acid derivatives could be considered as mimics of the pyrophosphate moiety of deoxyadenosine triphosphate.[4] Among these compounds, L-Asp-dAMP showed efficient incorporation in a primer-template assay and resulted in 90% conversion to a P + 1 strand in 60 min at 500 μM by HIV-1 RT. During this incorporation a P–N bond is cleaved. We have also shown that phosphodiesters are better substrates than phosphoramidates when it comes to polymerization of nucleotides using HIV-1 RT[5] and that the structure of the leaving group should not be limited to that of an α-amino acid.[5,6] However, the main problem when considering potential applications in biotechnology and medicine is that the kinetics obtained for incorporation of the nucleotide into a growing DNA chain using the alternative building blocks are poor. The V_{max} is mostly reasonably good, but the K_m value is generally 10^3–10^4 times higher when compared with dATP. Besides, chain termination was observed after incorporation of two or three nucleotides.[4] As the incorporation efficiency may be influenced by electrostatic interactions between substrate and enzyme, a series of sulf(on)ate and phosph(on)ate analogues (compounds **1–5**, Fig. 1) of L-Asp-dAMP have been synthesized as potential substrates for HIV-1 reverse transcriptase.

In a single nucleotide incorporation assay, 2′-deoxyadenosine-5′-(3-phosphono-L-alanine) phosphoramidate (**3**) is a considerably improved substrate, which resulted in 95% conversion to a P + 1 strand in 60 min at 50 μM (a concentration 10 times less than found for L-Asp-dAMP). It is found that compound **3** has improved incorporation kinetics and elongation capability. For example, the K_m for compound **3** is 78-fold higher than for the natural substrate dATP and the measured V_{max} is only 1.3-fold lower. Molecular modeling of compound **3** in the active site of HIV-RT revealed that two Mg^{2+} ions are tightly bound to the phosphonate group, the phosphoramidate group and to three catalytic carboxylate groups (Asp110A, Asp185A and Asp186A). The phosphonate group and the phosphoramidate group seem to mimic the γ- and α-phosphate groups of deoxynucleoside triphosphates, respectively. Lys65A and Arg72A are involved in the interactions with the phosphonate group and the carboxyl group, respectively. The carboxyl group functions as a mimic of the β-phosphate group.

[a]*Laboratory of Medicinal Chemistry, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium. E-mail: Piet.Herdewijn@rega.kuleuven.be; Fax: +32 16 337340; Tel: +32 16 337387*

[b]*Isthmus Sarl, 31 rue de Saint Amand, F-75015, Paris, France*

† Electronic supplementary information (ESI) available: NMR spectra of the sulf(on)ate and phosph(on)ate phosphoramidate analogues (compounds **1–8**). See DOI: 10.1039/c0ob00554a

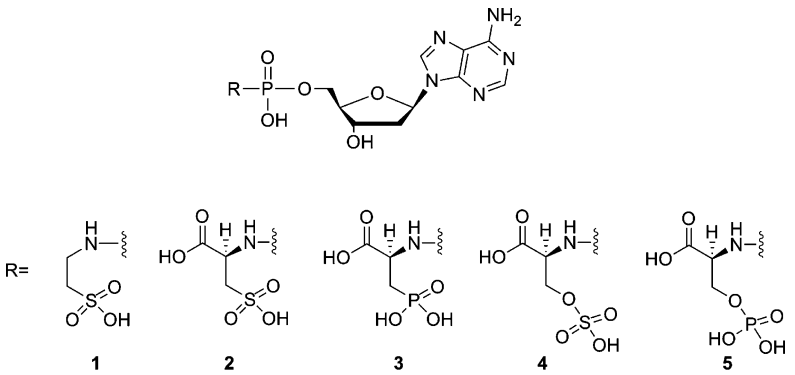
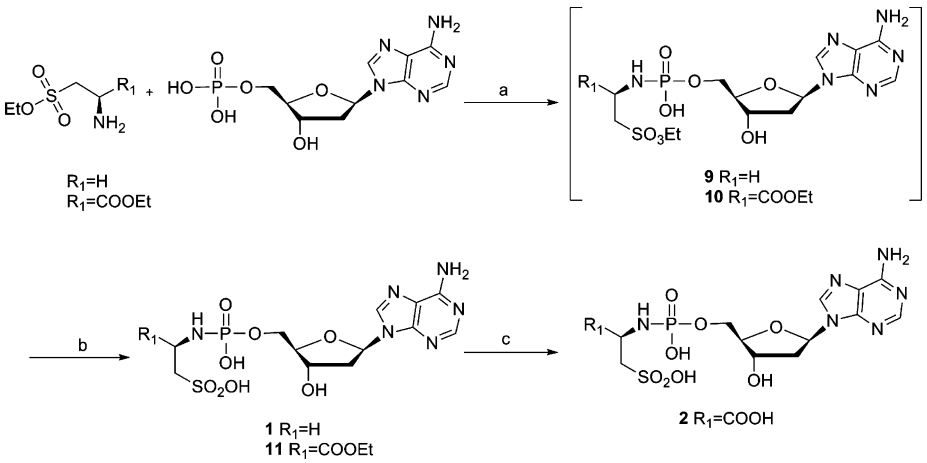


Fig. 1 Structures of phosphoramidate analogues of 2′-deoxyadenosine nucleotides.



Scheme 1 (a) EDAC, H_2O, r.t., 2–5 h; (b) 1.4 M K_2CO_3 in MeOH–H_2O 1 : 1, r.t., 2 h; (c) 0.4 M NaOH in MeOH–H_2O 1 : 1, r.t., overnight.

To further investigate the potential of 3-phosphono-L-alanine as a leaving group, compounds **6–8** (Fig. 2) were synthesized and evaluated. Among the three analogues, 3-phosphono-L-Ala-dGMP (**6**) and 3-phosphono-L-Ala-dTMP (**8**) are also good substrates resulting in 77% and 81% conversion to a P + 1 strand in 60 min at 50 μM in a single nucleotide incorporation assay. Nonetheless, the analogue **7** with a cytosine moiety is only incorporated at higher concentration. Apparently the efficiency of the leaving group is base-dependent. In all cases evaluated, the Watson–Crick base pairing rules are respected.

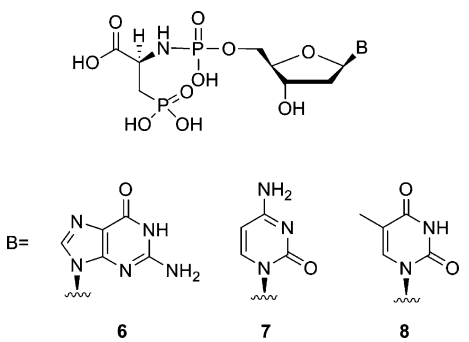
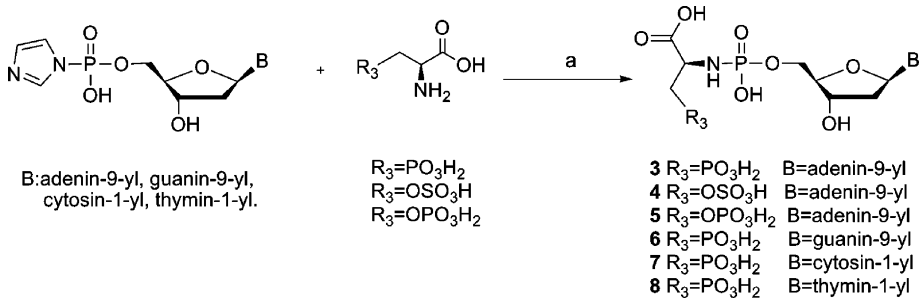


Fig. 2 Structures of phosphoramidate analogues of 2′-deoxynucleotides.

Results and discussion

The synthesis of Tau-dAMP (**1**) or L-Cys-dAMP (**2**) is shown in Scheme 1. Tau-dAMP (**1**) or L-Cys-dAMP (**2**) were synthesized from deoxyadenosine monophosphate (dAMP) and taurine or L-cysteic acid ethyl ester by an *N*-ethyl-*N*′-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDAC) mediated coupling, followed by a deprotection with base. The coupling and deprotection of taurine and L-cysteic acid was done in a one-pot reaction.

This coupling reaction was described before by F.-Q. Huang *et al.*,[7] in which they performed the EDAC-mediated coupling of adenosine 5′-monophosphate to diamines at room temperature in water. The DCC method (*t*-BuOH–H_2O 5 : 1, reflux, 4 h)[8] to synthesize the phosphoramidates needs more vigorous conditions than the EDAC coupling (H_2O, r.t., 2 h). Furthermore, because the protected sulfonic ester of taurine and L-cysteic acid (**9** and **10**) substituted dAMP were less stable than the carboxylic ester of protected L-Asp-dAMP, the initial attempts to separate and purify the sulfonic ester intermediates (**9** and **10**) failed and only trace amount of compounds were obtained. In addition, a sulfonamide side product (**12**) was obtained (Fig. 3). Apparently, the hydrolyzed sulfonic acid group could couple with another protected taurine in the presence of excess EDAC. Hydrolysis of compound **12** further generated compound **13**, which could be, likewise, isolated from the same reaction mixture. The formation



Scheme 2 Reagents and conditions: (a) *N*-ethylmorpholine, H_2O, ($ZnCl_2$), r.t., 1–3 d.

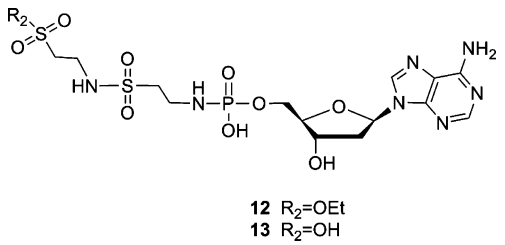


Fig. 3 Sulfonamide side products (**12** and **13**) obtained during the coupling reaction of dAMP and taurine ethyl ester.

of side products and the problem obtained with the purification of the sulfonic ester intermediates (**9** and **10**) can be avoided by carrying out this reaction in a one-pot process using a ratio of 3.3 : 1 (EDAC *vs.* dAMP). The yield of the desired compounds increases considerably when using this one-pot procedure.

The synthesis of compounds **3–8** is described in Scheme 2. As it could be a cumbersome procedure to selectively protect and deprotect the sulfate and phosph(on)ate amino acid derivatives during the preparation of **3–8**, these compounds were synthesized in a one-step reaction as described before by H. Sawai[9] for the formation of phosphodiesters. In this scheme, adenosine-5′-phosphorimidazolide is reacted with glycolic acid or lactic acid in 0.2 M *N*-ethylmorpholine buffer, pH 7.0 or 8.0, catalyzed by $Pb(NO_3)_2$ or $ZnCl_2$. After 1–10 d, 0.25 M EDTA disodium quench buffer is used to break down the nucleotide-metal complex. The reaction mixtures are subjected to purification by HPLC. It was postulated that divalent metal ions such as Pb^{2+}, Zn^{2+}, Mn^{2+} and Mg^{2+} neutralize the charge of the phosphate group and promote the nucleophilicity of the hydroxyl group of the substrate by coordination. When we tried similar reaction conditions for the synthesis of the phosphoramidates (**3–5**), the obtained compounds have an R_f value (TLC, *i*-PrOH–NH_3–H_2O 7 : 1 : 2) which is very close to that of EDTA and the compounds were difficult to purify. As the nucleophilicity of a primary amino group is stronger than the nucleophilicity of a hydroxyl group, we modified the reaction conditions by using more *N*-ethylmorpholine as base without adding any metal ions (the use of EDTA in the quenching procedure can be avoided). However, in the case of compounds **6–8**, $ZnCl_2$ was added to increase the reactivity of the nucleoside phosphorimidazolides[10] and the compounds were purified by column chromatography followed by preparative HPLC to remove the EDTA–Zn^{2+} complex.

The ability of the phosphoramidate analogues (**1–5**) to function as substrate for HIV-1 RT was investigated by the gel-based single

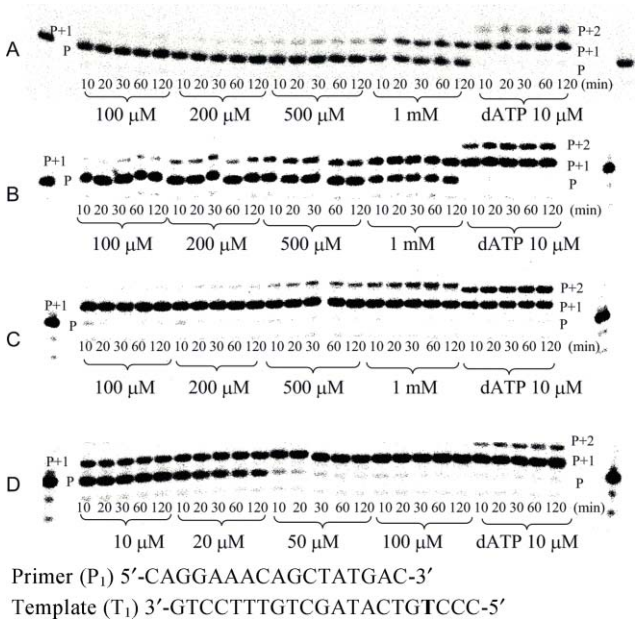


Fig. 4 Single incorporation of the primer of P_1T_1 (125 nM) by HIV-1 RT with phosphoramidate substrate concentrations and time intervals (min) as indicated, [HIV-1 RT] = 0.025 U μL^{-1}; A: incorporation of **1**, B: incorporation of **2**, C and D: incorporation of **3**, dATP (10 μM) is used as reference.

nucleotide incorporation assay (Fig. 4). The natural deoxyadenosine triphosphate (dATP) was used as reference.

Using identical conditions, Tau-dAMP (**1**), L-Cys-dAMP (**2**), *O*-sulfonato-L-Ser-dAMP (**4**), *O*-phospho-L-Ser-dAMP (**5**) were not good substrates for HIV-RT, showing 50%, 63%, 8%, 19% conversion to a P + 1 strand in 60 min at 1 mM, respectively. However, the use of 3-phosphono-L-Ala-dAMP (**3**) as substrate resulted in 95% conversion to a P + 1 strand in 60 min at 50 μM. When the 3-phosphono-L-Ala-dAMP (**3**) concentration was decreased to 10 μM 29% of the P + 1 strand was still formed at 10 min. The phosphonate group has an additional charge when compared with L-Asp and L-Cys, which may have a beneficial effect on metal chelation and catalysis. In contrast, *O*-sulfonato-L-Ser-dAMP (**4**) is a very poor substrate for HIV-1 RT in the polymerase reaction. The sulfate residue of *O*-sulfonato-L-Ser-dAMP (**4**) is unstable and is easily hydrolyzed to L-Ser-dAMP which has already been proven to be a poor substrate.[4] Although 3-phosphono-L-Ala-dAMP (**3**) and *O*-phospho-L-Ser-dAMP (**5**) both have four negative charges, the leaving group

ability of 3-phosphono-L-alanine is superior. This suggests that the phosphonate group is better accommodated in the active site of HIV-RT leading to a more productive complex than the *O*-phosphate group. These observations indicate that small changes[4] in the chemical constitution of the amino acid derivative can have a profound effect on its biochemical behavior.

The incorporation of more than one nucleotide was tested in a template dependent incorporation assay using HIV-1 RT (Fig. 5). Primer P_1 and template T_2 containing an overhang of seven thymidine residues and additional four nucleobases (GGAC) at the 5′-end were used.[11]

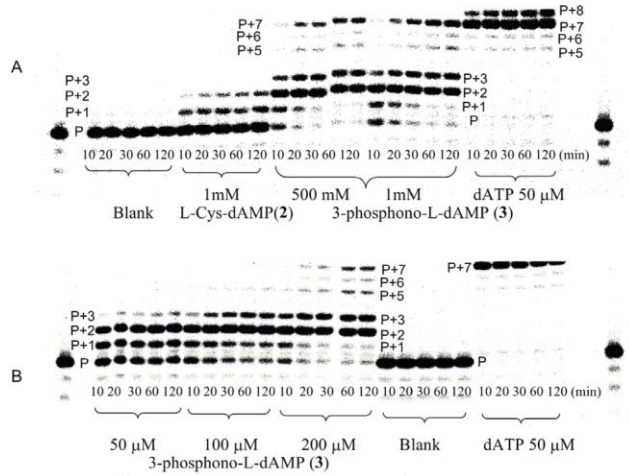


Fig. 5 Elongation of the primer of P_1T_2 (125 nM) by HIV-1 RT with phosphoramidate substrate concentrations and time intervals (min) as indicated, [HIV-1 RT] = 0.025 U μL^{-1}; A: incorporation of **2** and **3**, B: incorporation of **3**; blank: 125 nM primer/template (P_1T_2), [HIV-1 RT] = 0.025 U μL^{-1} and no nucleotide substrate, dATP (50 μM) is used as reference.

The gel electrophoresis experiments showed that L-Cys-dAMP (**2**) could extend a primer with one and two adenine nucleotides (P + 1 and P + 2 products). The use of 3-phosphono-L-Ala-dAMP (**3**), however, leads to full elongation. After 60 min of polymerase reaction, 3-phosphono-L-Ala-dAMP (**3**) leads to strand elongation to a P + 7 product (24%) at 1 mM, P + 7 (16%) at 500 μM, P + 7 (7%) at 200 μM, respectively. Although the primer has been elongated, the stalling effect is still present, demonstrated by the slow-down of the incorporation reaction resulting predominantly in the formation of P + 2 and P + 3 products. In the case of L-Asp-dAMP incorporation at 500 μM, only a trace amount of P + 6 products was obtained and the major product is the P + 2 elongated oligonucleotide. It is not clear how these non-canonical leaving groups influence the translocation step for processive DNA synthesis. A chemical footprinting assay of PPi analogues revealed that phosphonoformic acid can stabilize the 3′ end complex in a pre-translocated configuration leading to inhibition of DNA synthesis.[12a] High concentration of PPi (10 mM) also blocked DNA synthesis and results in the formation of P + 1 and P + 2 as major products.[12a] After formation of the first and the second templated oligonucleotides (*i.e.* P + 1 and P + 2), the 3′ end of the blocked primer could be shifted to an unproductive configuration.[12]

Table 1 Steady-state kinetics of single nucleotide incorporation into P_1T_1 by HIV-1 RT

Substrate	V_{max}/nM min^{-1}	K_m/μM	V_{max}/K_m ($\times 10^{-3}$ min^{-1})
dATP	12.50 ± 0.2623	1.010 ± 0.1107	12.37
3-Phosphono-L-Ala-dAMP (**3**)	9.842 ± 0.2572	78.90 ± 6.699	0.125

To investigate the influence of the leaving group on the incorporation reaction, we carried out a product inhibition experiment. The incorporation of 3-phosphono-L-Ala-dAMP (**3**) (50 μM, figure in the ESI†) was carried out in the presence of a 2-, 10- and 20-fold higher concentration of 3-phosphono-L-alanine. After 120 min of polymerase reaction, 8–11% less P + 1 product was obtained. The result indicates that 3-phosphono-L-alanine formed in the polymerization reaction may serve as an inhibitory agent, although the effect would be minor.

The presence of an excess of 3-phosphono-L-alanine also has an influence on the chain elongation reaction. When adding 1 mM of 3-phosphono-L-alanine to the reaction mixture containing 150 μM of compound **3**, P_1T_2 (125 nM) and [HIV-1 RT] = 0.025 U μL^{-1} for 120 min, 9% more P, 10% more P + 1 and 11% less P + 2 product was obtained (data not shown).

The efficiency of 3-phosphono-L-Ala-dAMP (**3**) was investigated by determining the kinetic parameters.[13] Identical enzyme concentration [0.0125 U μL^{-1}] was used for both dATP and 3-phosphono-L-Ala-dAMP (**3**). Steady-state kinetic analysis (Table 1) indicates that the K_m for 3-phosphono-L-Ala-dAMP (**3**) is 78-fold higher than for the natural substrate dATP, the measured V_{max} is only 1.3-fold lower than for dATP and the V_{max}/K_m ratio is 99 times lower than for dATP. These data indicate efficient nucleophilic displacement of 3-phosphono-L-alanine when the phosphoramidate is bound in the active site. The V_{max}/K_m value is superior to the value previously observed for L-Asp-dAMP, for which the V_{max}/K_m ratio is 1292 times lower than for dATP.[4]

To further investigate the property of 3-phosphono-L-Ala-dAMP (**3**), it was tested as a substrate for Taq DNA polymerase (Fig. 6). The primer extension by Taq DNA polymerase was less efficient than using HIV-1 RT.

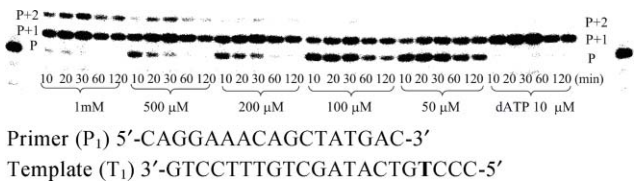
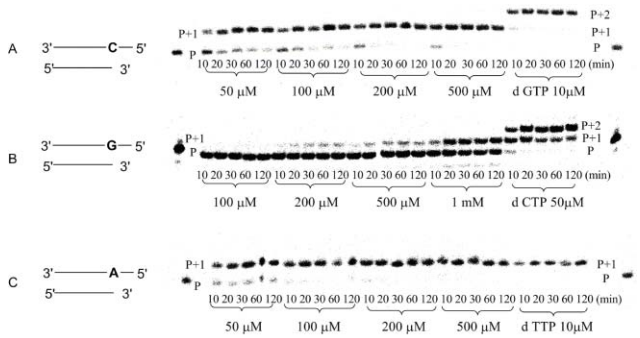


Fig. 6 Incorporation of the primer of P_1T_1 (125 nM) by Taq DNA polymerase with **3** as substrate; phosphoramidate concentrations and time intervals (min) are indicated; [Taq] = 0.025 U μL^{-1}; dATP (10 μM) incorporation is used as reference.

To investigate the influence of the base moiety on the incorporation efficiency, we have evaluated the leaving group ability of 3-phosphono-L-alanine using the phosphoramidate nucleotides with the other three nucleobases. Single nucleotide incorporation assays (Fig. 7) were carried out for compounds **6–8**. Efficient incorporation was observed for 3-phosphono-L-Ala-dGMP (**6**) and 3-phosphono-L-Ala-dTMP (**8**), which resulted in 77% and



Primer (P_1) 5′-CAGGAAACAGCTATGAC-3′
Primer (P_2) 5′-CAGGAAACAGCTATGACTGA-3′
Template (T_3) 3′-GTCCTTTGTCGATACTG**C**TTTT-5′
Template (T_4) 3′-GTCCTTTGTCGATACTGACT**G**AAAAA-5′
Template (T_5) 3′-GTCCTTTGTCGATACTG**A**CTGC-5′

Fig. 7 Single incorporation by HIV-1 RT with phosphoramidate substrate concentrations and time intervals (min) as indicated, [HIV-1 RT] = 0.025 U μL^{-1}; A: incorporation of **6** with P_1T_3, dGTP (10 μM) is used as reference; B: incorporation of **7** with P_2T_4, dCTP (50 μM) is used as reference; C: incorporation of **8** with P_1T_5, dTTP (10 μM) is used as reference.

81% conversion to a P + 1 strand in 60 min at 50 μM, respectively. In contrast, 3-phosphono-L-Ala-dCMP (**7**) was less efficient. A concentration of 1 mM was needed to reach 50% incorporation at 30 min. In agreement with this observation, the incorporation of dCTP is likewise less efficient than of dATP, dGTP and dTTP. It is not clear by now if this is a sequence specific effect of RT/primer-template interactions on substrate selection[14] or a general observation, and this phenomenon needs further investigation. Using 3-phosphono-L-alanine as a leaving group, the incorporation efficiency decreases in the order A ≥ T = G > C.

To confirm that the observed incorporation profile was due to true base pairing following canonical Watson–Crick rules, we carried out control experiments (figure in ESI†) with compounds **3**, **6** and **8** and mismatch sequences (A against A, C, G; G against G, T, A and T against C, T, G). After 120 min of the polymerase reaction at 1 mM substrate concentration, primer elongation was not observed in any of these cases.

Molecular modeling of 3-phosphono-L-Ala-dAMP in the active site of HIV-RT

The way in which 3-phosphono-L-Ala-dAMP (**3**), 3-phosphono-L-Ala-dGMP (**6**) and 3-phosphono-L-Ala-dTMP (**8**) could be accommodated in the active site of HIV-RT was investigated using molecular modeling and the Amber software.[15]

Fig. 8 shows the entering 3-phosphono-L-Ala-dAMP (**3**) in complex with the RT enzyme. The α-phosphate and leaving group are in the same position as the triphosphate in the original substrate thymidine triphosphate (TTP) complex, close to the 3′-OH group of the terminal primer residue. Also the 2 Mg^{2+} ions are comparable in position to their localization in the original TTP complex. They are tightly bound to both the phosphonate group of 3-phosphono-L-Ala-dAMP (**3**) (mimicking the γ phosphate group), the phosphoramidate group and to three Asp groups in the enzyme: Asp110A and Asp185A, which are widely conserved in polymerases, and Asp186A.[17] In the original X-ray structure,

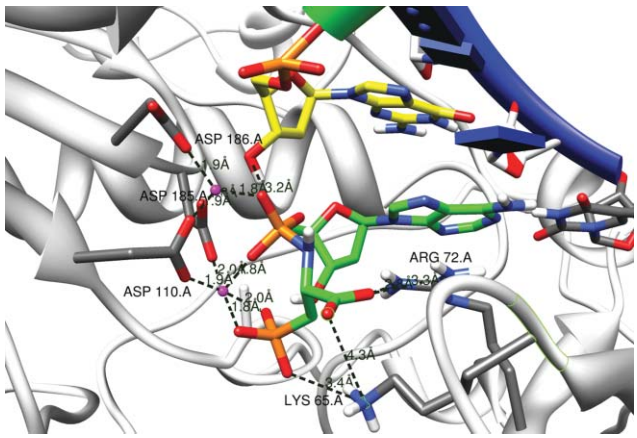


Fig. 8 Model structures of 3-phosphono-L-Ala-dAMP in the RT dNTP pocket. The primer strand is drawn with a green ribbon; the template strand has a blue ribbon. The residues Asp 110, 185 and 186 anchor the 2 Mg^{2+} ions (purple balls). Some distances between charged atoms are indicated. The first nucleic acid of the primer strand (yellow carbons) and the complementary residue to 3-phosphono-L-Ala-dAMP (**3**) are shown in stick representation. Possible stabilization of the carboxyl function and the phosphonate function in the leaving group by Arg 72 and Lys 65 is indicated. Figures are generated using Chimera.[16]

Asp186A is not connected to one of the Mg^{2+} ions. The side chain of Lys65A forms a salt bridge with the phosphonate group. Arg72A stabilizes the carboxyl function of the leaving group. Also Lys65A may interact with this carboxyl group *via* the formation of a salt bridge. Although the Lys65A.NZ carboxyl distance is 4.3 Å in the 3-phosphono-L-Ala-dAMP (**3**) model, such interaction may be possible. The Lys65A side chain could move near to the carboxyl group because of the flexibility of this side chain and no immediate steric hindrance is present for this group when moving closer. The possible binding mode of 3-phosphono-L-Ala-dNMP (N = G and T, respectively) is very similar and shown in the ESI.†

Conclusions

The enzymatic synthesis of DNA uses deoxynucleoside triphosphates as substrates. Little research has been done to investigate the role of the pyrophosphate leaving group in this process. One of the suggested functions of the release of pyrophosphate is that it provides the energy for translocation.[18] One way to study the role of the pyrophosphate moiety in the polymerization process is to replace the pyrophosphate group of deoxynucleoside triphosphates by other potential leaving groups, and analyze its influence on the polymerization process. Another important reason to study the substrate specificity of polymerases (such as reverse transcriptase) at the level of the leaving group is to develop analogues of nucleoside triphosphates that are direct substrates for the enzyme. 3-Phosphono-L-alanine is a remarkable leaving group for DNA synthesis catalyzed by HIV-RT. 3-Phosphono-L-Ala-dAMP is a better substrate for the polymerization process than the previously described L-Asp-dAMP (reflected by more successful DNA synthesis at 10-fold lower substrate concentration, and lower V_{max}/K_m value). However, DNA synthesis using this non-canonical leaving group leads to a stalling effect at P + 2 and P + 3, although at higher concentration the synthesis of P + 7 oligonucleotide can be observed. The primer extension is base

specific and the efficiency of incorporation for the primer sequence is A ≥ T = G > C. A modeling experiment shows two metal ions tightly binding to the leaving group and the carboxylic residues of the polymerase active site. The phosphonate group and the carboxyl group interact with Lys65A and Arg72A, respectively. As one of our interests is the development of a xenobiology[19] based on alternative nucleic acids and their building blocks, an additional level of interest is the potential metabolic accessibility[20] of the leaving group. 3-Phosphono-L-alanine is naturally occurring[20a] in the sea anemone *Zoanthus sociatus* and the protozoan *Tetrahymena pyriformis* and is not toxic.[20b] Biodistribution studies revealed that 3-phosphono-L-alanine is circulated in the human liver, intestine and spleen (no free phosphonic acid was detected and the acid was bound either to lipid or to protein).[20c] About 19% of 3-phosphono-L-alanine taken up by *Tetrahymena pyriformis* is accumulated as aminoethylphosphonic acid in the phospholipids.[20d] The catabolic pathway of 3-phosphono-L-alanine in *Tetrahymena* is the same as in rats.[20e] The observations that 3-phosphono-L-alanine is naturally occurring[20a] and nontoxic,[20b] further support its potential use as leaving group *in vivo*.

Experimental section

For all reactions, analytical grade solvents were used. Depending on sample amounts or availability, Bruker 600 MHz, Avance II 500 MHz or Avance 300 MHz spectrometers were used for ^{1}H NMR, ^{13}C NMR and ^{31}P NMR. All chemical shifts are listed in ppm. For sake of clarity of the NMR signal assignment, sugar protons and carbons are numbered with a prime. ^{31}P NMR chemical shifts are referenced to an external 85% H_3PO_4 standard ($\delta = 0.000$ ppm). UV spectra were recorded on a Varian Cary-300-Bio UV/Vis spectrophotometer and optical rotations were measured at 20 °C on a Perkin–Elmer 341 polarimeter at 589 nm. Mass spectrometry was performed on a quadrupole orthogonal acceleration time-of-flight mass spectrometer (Q-Tof-2, Micromass, Manchester, UK) equipped with a standard electrospray probe (*Z*-spray, Micromass, Manchester, UK). Samples were dissolved in acetonitrile–water (1 : 1 v/v) and infused with a flow rate of 5 μL min^{-1}. Electrospray capillary voltage was set to 3000 V and cone voltage to 30 V. Nitrogen gas was used for nebulisation and desolvation. Accurate masses were determined by coinfusion of the samples with leucine enkephalin (YGGFL) and recalibration of the spectrum using the peak at *m/z* 556.2771 as lock mass. Precoated aluminium sheets (MN ALUGRAM SIL G/UV$_{254}$ 20 × 20 cm) were used for TLC; the spots were examined with UV light. Column chromatography was performed on ICN silica gel 63–200, 60 Å. Preparative HPLC was performed on a Waters 1525–2487 system using Prep C18 5 μm column 19 × 150 mm at the flow rate of 3 mL min^{-1} by a gradient elution of acetonitrile and 50 mM triethylammonium bicarbonate (TEAB) buffer.

General procedure for the synthesis of compounds 1 and 2

The example described is for 2′-deoxyadenosine-5′-taurine phosphoramidate (compound **1**).

Taurine ethyl ester was synthesized according to the procedure described previously.[21] The crude product was directly used for the amidation reaction, because purification by flash chromatography on silica gel led to hydrolysis. 2′-Deoxyadenosine-5′-monophosphoric acid hydrate (40 mg, 0.11 mmol) and the crude taurine ethyl ester (81 mg, 0.53 mmol) were suspended in 1 mL water and stirred for 5 min under argon. Then, *N*-ethyl-*N*′-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDAC, 69 mg, 0.36 mmol) was added to the suspension and the reaction was continued to stir at room temperature. After 2 h, 1.4 M K_2CO_3 (MeOH–H_2O 1 : 1) 1.2 mL was added to the reaction mixture. The progress of the reaction was monitored by TLC ($CHCl_3$–MeOH–H_2O 5 : 4 : 0.5) and ^{31}P NMR until the disappearance of ester intermediate. The reaction mixture was neutralized by addition of 1 M triethylammonium acetate (TEAA). The solvent was evaporated to dryness in vacuum. The residue was purified by silica column chromatography eluting with *i*-PrOH–NH_3–H_2O (12 : 1 : 1, 10 : 1 : 1 to 9 : 1 : 1) to yield compound **1** (30 mg, 62%) as a white solid. The product was further purified by an ion exchange column (DEAE-Sephadex A-25) with a gradient of 1 M triethylammonium bicarbonate buffer (25 mg, 83.3%).

General procedure for the synthesis of compounds 3–8

The example described is for 2′-deoxyadenosine-5′-(3-phosphono-L-alanine) phosphoramidate (compound **3**).

Deoxyadenosine-5′-phosphorimidazolide was synthesized according to the standard procedures.[22] In a 25 ml flask, L-alanine-3-phosphono acid HCl salt[23] (60 mg, 0.29 mmol) were stirred in 0.5 mL *N*-ethyl morpholine for 5 min at room temperature, then deoxyadenosine-5′-phosphorimidazolide (240 mg, 0.63 mmol) and 6 mL 0.2 M *N*-ethyl morpholine buffer (pH 7.5) were added into the flask. The reaction mixture was continued to stir for 3 days at room temperature under argon, while the reaction process was monitored by TLC (*i*-PrOH–NH_3–H_2O 7 : 1 : 2) and ^{31}P NMR. The reaction mixture was concentrated to dryness in vacuum (30 °C bath) and the residue was purified by silica column chromatography eluting with *i*-PrOH–NH_3–H_2O (20 : 1 : 1, 9 : 1 : 1, 7 : 1 : 1 to 4.5 : 1 : 1) and yielding crude white solid (32 mg, 22.8%). The product was purified by preparative HPLC with a gradient of CH_3CN in 50 mM TEAB buffer (pH = 7.4) to yield compound **3** (10 mg, 31.3%).

2′-Deoxyadenosine-5′-taurine phosphoramidate (1). UV: (H_2O) λ_{max} = 260 nm (log ε = 4.18); $[\alpha]^{20}_{589} = -44.3°$ (c = 0.9, H_2O); ^{1}H NMR (300 MHz, D_2O): δ 8.43 (s, 1H, H-8), 8.18 (s, 1H, H-2), 6.46 (m, 1H, H-1′), 4.71 (m, 1H, H-3′), 4.26 (m, 1H, H-4′), 3.97 (m, 2H, H-5′ and H-5″), 3.16 (m, 2H, -$CH_2CH_2SO_3H$), 2.93 (m, 2H, -$CH_2CH_2SO_3H$), 2.85 (m, 1H, H-2′), 2.60 (m, 1H, H-2″); ^{13}C NMR (75 MHz, D_2O): δ 155.25 (C-6), 152.35 (C-2), 148.57 (C-4), 139.85 (C-8), 118.49 (C-5), 85.95 (d, ^{3}J(C, P) = 8.9 Hz, C-4′), 83.66 (C-1′), 71.29 (C-3′), 64.02 (d, ^{2}J(C, P) = 5.1 Hz, C-5′), 52.25 (d, ^{3}J(C, P) = 6.3 Hz, -$CH_2CH_2SO_3H$), 39.00 (C-2′), 36.93 (-$CH_2CH_2SO_3H$); ^{31}P NMR (121 MHz, D_2O): δ 8.16; HRMS for $C_{12}H_{19}N_6O_8PS$ $(M-H)^-$ calcd: 437.0650, found: 437.0634.

2′-Deoxyadenosine-5′-(L-cysteic acid) phosphoramidate (2). UV: (H_2O) λ_{max} = 260 nm (log ε = 4.18); $[\alpha]^{20}_{589} = -17.2°$ (c = 0.8, H_2O); ^{1}H NMR (500 MHz, D_2O): δ 8.50 (s, 1H, H-8), 8.24 (s, 1H, H-2), 6.51 (m, 1H, H-1′), 4.72 (m, 1H, H-3′), 4.26 (m, 1H, H-4′), 3.99 (m, 2H, H-5′ and H-5″), 3.87 (m, 1H, -C*H*COOH), 3.19 (m, 2H, -CH_2SO_3H), 2.82 (m, 1H, H-2′), 2.58 (m, 1H, H-2″); ^{13}C NMR (125 MHz, D_2O): δ 178.61 (d, ^{3}J(C, P) = 4.0 Hz,

-COOH), 155.14 (C-6), 152.22 (C-2), 148.35 (C-4), 139.57 (C-8), 118.20 (C-5), 85.73 (d, ^{3}J(C, P) = 9.3 Hz, C-4′), 83.20 (C-1′), 70.95 (C-3′), 63.57 (d, ^{2}J(C, P) = 4.7 Hz, C-5′), 54.38 (d, ^{3}J(C, P) = 6.2 Hz, -*C*H$_2$SO$_3$H), 54.22 (d, ^{2}J(C, P) = 6.9 Hz, -*C*HCOOH), 38.63 (C-2′); ^{31}P NMR (202 MHz, D$_2$O): δ 6.35; HRMS for $C_{13}H_{19}N_6O_{10}PS$ (M-H)$^-$ calcd: 481.0548, found: 481.0568.

2′-Deoxyadenosine-5′-(3-phosphono-L-alanine) phosphoramidate (3). UV: (H$_2$O) λ_{max} = 259 nm (log ε = 4.17); $[\alpha]^{20}_{589}$ = −22.8°(*c* = 0.6, H$_2$O); ^{1}H NMR (600 MHz, D$_2$O, 5 °C): δ 8.46 (s, 1H, H-8), 8.21 (s, 1H, H-2), 6.46 (m, 1H, H-1′), 4.64 (m, 1H, H-3′), 4.21 (m, 1H, H-4′), 3.90 (m, 2H, H-5′ and H-5″), 3.72 (m, 1H, -C*H*COOH), 2.76 (m, 1H, H-2′), 2.52 (m, 1H, H-2″), 1.97 (m, 1H, -C*H*$_2$PO$_3$H$_2$), 1.88 (m, 1H, -C*H*$_2$PO$_3$H$_2$); ^{13}C NMR (150 MHz, D$_2$O, 5 °C): δ 180.04 (d, ^{3}J(C, P) = 13.5 Hz, -COOH), 154.12 (C-6), 150.63 (C-2), 148.21 (C-4), 139.95 (C-8), 118.16 (C-5), 85.81 (d, ^{3}J(C, P) = 9.7 Hz, C-4′), 83.49 (C-1′), 71.05 (d, ^{4}J(C, P) = 4.9 Hz, C-3′), 63.72 (C-5′), 52.30 (d, ^{2}J(C, P) = 3.6 Hz, -*C*HCOOH), 38.86 (C-2′), 33.39 (d, ^{3}J(C, P) = 6.8 Hz, ^{1}J(C, P) = 129.4 Hz, -*C*H$_2$PO$_3$H$_2$); ^{31}P NMR (121 MHz, D$_2$O): δ 20.72 (-CH$_2$*P*O$_3$), 6.87 (*N*-*P*O$_3$); HRMS for $C_{13}H_{20}N_6O_{10}P_2$ (M-H)$^-$ calcd: 481.0643, found: 481.0654.

2′-Deoxyadenosine-5′-(*O*-sulfonato-L-serine) phosphoramidate (4). UV: (H$_2$O) λ_{max} = 260 nm (log ε = 4.18); $[\alpha]^{20}_{589}$ = −16.4°(*c* = 3.6, H$_2$O); ^{1}H NMR (300 MHz, D$_2$O): δ 8.50 (s, 1H, H-8), 8.25 (s, 1H, H-2), 6.51 (m, 1H, H-1′), 4.71 (m, 1H, H-3′), 4.28 (m, 1H, H-4′), 4.16 (m, 1H, -C*H*$_2$OSO$_3$H), 4.07 (m, 1H, -C*H*$_2$OSO$_3$H), 4.00 (m, 2H, H-5′ and H-5″), 3.76 (m, 1H, -C*H*COOH), 2.83 (m, 1H, H-2′), 2.60 (m, 1H, H-2″); ^{13}C NMR (75 MHz, D$_2$O): δ 177.10 (d, ^{3}J(C, P) = 4.1 Hz, -COOH), 154.38 (C-6), 151.08 (C-2), 148.33 (C-4), 140.01 (C-8), 118.28 (C-5), 85.90 (d, ^{3}J(C, P) = 9.2 Hz, C-4′), 83.54 (C-1′), 71.22 (C-3′), 70.85 (d, ^{3}J(C, P) = 15.7 Hz, -*C*H$_2$OSO$_3$H), 63.80 (d, ^{2}J(C, P) = 5.1 Hz, C-5′), 55.64 (d, ^{2}J(C, P) = 8.7 Hz, -*C*HCOOH), 38.83 (C-2′); ^{31}P NMR (121 MHz, D$_2$O): δ 6.54; HRMS for $C_{13}H_{19}N_6O_{11}P_1S_1$ (M-H)$^-$ calcd: 497.0497, found: 497.0471.

2′-Deoxyadenosine-5′-(*O*-phospho-L-serine) phosphoramidate (5). UV: (H$_2$O) λ_{max} = 259 nm (log ε = 4.17); $[\alpha]^{20}_{589}$ = −14.9°(*c* = 0.9, H$_2$O); ^{1}H NMR (600 MHz, D$_2$O, 5 °C): δ 8.46 (s, 1H, H-8), 8.18 (s, 1H, H-2), 6.47 (m, 1H, H-1′), 4.62 (m, 1H, H-3′), 4.22 (m, 1H, H-4′), 3.96 (m, 1H, H-5′), 3.92 (m, 2H, H-5″ and -C*H*$_2$OPO$_3$H$_2$), 3.82 (m, 1H, -C*H*$_2$OPO$_3$H$_2$), 3.63 (m, 1H, -C*H*COOH), 2.76 (m, 1H, H-2′), 2.50 (m, 1H, H-2″); ^{13}C NMR (150 MHz, D$_2$O, 5 °C): δ 179.12 (d, ^{3}J(C, P) = 4.1 Hz, -COOH), 155.24 (C-6), 152.24 (C-2), 148.41 (C-4), 139.52 (C-8), 118.17 (C-5), 85.97 (d, ^{3}J(C, P) = 9.6 Hz, C-4′), 83.23 (C-1′), 71.17 (C-3′), 67.16 (-*C*H$_2$OPO$_3$H$_2$), 63.79 (d, ^{2}J(C, P) = 4.2 Hz, C-5′), 56.87 (d, ^{2}J(C, P) = 7.5 Hz, -*C*HCOOH), 38.78 (C-2′); ^{31}P NMR (121 MHz, D$_2$O): δ 6.76 (*N*-*P*O$_3$), 0.28 (*P*O$_4$); HRMS for $C_{13}H_{20}N_6O_{11}P_2$ (M-H)$^-$ calcd: 497.0592, found: 497.0585.

2′-Deoxyguanosine-5′-(3-phosphono-L-alanine) phosphoramidate (6). UV: (H$_2$O) λ_{max} = 252 nm (log ε = 4.10); $[\alpha]^{20}_{589}$ = −37.1°(*c* = 0.5, H$_2$O); ^{1}H NMR (600 MHz, D$_2$O, 5 °C): δ 8.05 (d, 1H, H-8), 6.25 (m, 1H, H-1′), 4.60 (m, 1H, H-3′), 4.14 (m, 1H, H-4′), 3.86 (m, 2H, H-5′ and H-5″), 3.67 (m, 1H, -C*H*COOH), 2.74 (m, 1H, H-2′), 2.40 (m, 1H, H-2″), 1.91 (m, 1H, -C*H*$_2$PO$_3$H$_2$), 1.78 (m, 1H, -C*H*$_2$PO$_3$H$_2$); ^{13}C NMR (150 MHz, D$_2$O, 5 °C): δ 181.19 (d, ^{3}J(C, P) = 9.7 Hz, -COOH), 158.78 (C-6), 153.59 (C-2), 151.18 (C-4), 137.30 (C-8), 115.74 (C-5), 85.64 (d, ^{3}J(C, P) = 9.0 Hz, C-4′), 83.18 (C-1′), 71.24 (C-3′), 63.75 (d, ^{2}J(C, P) = 4.3 Hz, C-5′), 52.87 (d, ^{2}J(C, P) = 3.5 Hz, -*C*HCOOH), 38.26 (C-2′), 33.77 (dd, ^{3}J(C, P) = 7.5 Hz, ^{1}J(C, P) = 129.0 Hz, -*C*H$_2$PO$_3$H$_2$); ^{31}P NMR (202 MHz, D$_2$O): δ 21.00 (-CH$_2$*P*O$_3$), 7.01 (*N*-*P*O$_3$); HRMS for $C_{13}H_{20}N_6O_{11}P_2$ (M-H)$^-$ calcd: 497.0592, found: 497.0590.

2′-Deoxycytidine-5′-(3-phosphono-L-alanine) phosphoramidate (7). UV: (H$_2$O) λ_{max} = 274 nm (log ε = 3.95); $[\alpha]^{20}_{589}$ = +27.2°(*c* = 0.7, H$_2$O); ^{1}H NMR (500 MHz, D$_2$O, 5 °C): δ 8.05 (d, 1H, H-6), 6.27 (m, 1H, H-1′), 6.17 (d, 1H, H-5), 4.49 (m, 1H, H-3′), 4.15 (m, 1H, H-4′), 3.97 (m, 2H, H-5′ and H-5″), 3.70 (m, 1H, -C*H*COOH), 2.37 (m, 1H, H-2′), 2.29 (m, 1H, H-2″), 1.97 (m, 1H, -C*H*$_2$PO$_3$H$_2$), 1.85 (m, 1H, -C*H*$_2$PO$_3$H$_2$); ^{13}C NMR (150 MHz, D$_2$O, 5 °C): δ 180.61 (d, ^{3}J(C, P) = 14.2 Hz, -COOH), 170.41 (C-4), 158.15 (C-2), 141.86 (C-6), 95.80 (C-5), 85.65 (d, ^{3}J(C, P) = 9.3 Hz, C-4′), 85.56 (C-1′), 70.50 (C-3′), 63.50 (d, ^{2}J (C, P) = 4.0 Hz, C-5′), 52.59 (d, ^{2}J(C, P) = 4.5 Hz, -*C*HCOOH), 39.22 (C-2′), 33.71 (dd, ^{3}J(C, P) = 6.8 Hz, ^{1}J(C, P) = 136.5 Hz, -*C*H$_2$PO$_3$H$_2$); ^{31}P NMR (202 MHz, D$_2$O): δ 20.95 (-CH$_2$*P*O$_3$), 6.96 (*N*-*P*O$_3$); HRMS for $C_{12}H_{20}N_4O_{11}P_2$ (M-H)$^-$ calcd: 457.0531, found: 457.0511.

2′-Deoxythymidine-5′-(3-phosphono-L-alanine) phosphoramidate (8). UV: (H$_2$O) λ_{max} = 267 nm (log ε = 3.95); $[\alpha]^{20}_{589}$ = −23.5°(*c* = 0.6, H$_2$O); ^{1}H NMR (500 MHz, D$_2$O, 5 °C): δ 7.83 (s, 1H, H-6), 6.37 (m, 1H, H-1′), 4.54 (m, 1H, H-3′), 4.15 (m, 1H, H-4′), 3.96 (m, 2H, H-5′ and H-5″), 3.72 (m, 1H, -C*H*COOH), 2.19–2.30 (m, 2H, H-2′ and H-2″), 1.94 (m, 1H, -C*H*$_2$PO$_3$H$_2$), 1.92 (s, 3H, -C*H*$_3$), 1.85 (m, 1H, -C*H*$_2$PO$_3$H$_2$); ^{13}C NMR (125 MHz, D$_2$O, 5 °C): δ 173.55 (d, ^{3}J(C, P) = 12.7 Hz, -COOH), 166.08 (C-4), 151.19 (C-2), 136.77 (C-6), 111.17 (C-5), 85.12 (d, ^{3}J(C, P) = 2.6 Hz, C-4′), 84.34 (C-1′), 70.71 (C-3′), 63.68 (d, ^{2}J (C, P) = 8.1 Hz, C-5′), 51.70 (d, ^{2}J(C, P) = 3.6 Hz, -*C*HCOOH), 38.13 (C-2′), 28.17 (d, ^{1}J(C, P) = 129.6 Hz, -*C*H$_2$PO$_3$H$_2$), 11.10 (CH$_3$); ^{31}P NMR (202 MHz, D$_2$O): δ 21.11 (-CH$_2$*P*O$_3$), 6.75 (*N*-*P*O$_3$); HRMS for $C_{13}H_{21}N_3O_{12}P_2$ (M-H)$^-$ calcd: 472.0528, found: 472.0510.

2′-Deoxyadenosine-5′-(ethyl L-cysteate) phosphoramidate (11, intermediate for 2). TLC (*i*-PrOH–NH$_3$–H$_2$O 7 : 1 : 2, v/v): R_f 0.62; ^{1}H NMR (500 MHz, D$_2$O): δ 8.42 (s, 1H, H-8), 8.24 (s, 1H, H-2), 6.49 (m, 1H, H-1′), 4.80 (m, 1H, H-3′), 4.25 (m, 1H, H-4′), 4.05 (m, 2H, -OC*H*$_2$CH$_3$), 4.03 (m, 2H, H-5′ and H-5″), 3.85 (m, 1H, -C*H*COOH), 3.09 (m, 2H, -C*H*$_2$SO$_2$OH), 2.86 (m, 1H, H-2′), 2.62 (m, 1H, H-2″); ^{13}C NMR (125 MHz, D$_2$O): δ 173.50 (d, ^{3}J(C, P) = 4.6 Hz, -*C*OOEt), 155.22 (C-6), 152.37 (C-2), 148.40 (C-4), 139.52 (C-8), 118.27 (C-5), 85.60 (d, ^{3}J(C, P) = 16.2 Hz, C-4′), 83.26 (C-1′), 70.90 (C-3′), 63.74 (d, ^{2}J(C, P) = 8.1 Hz, C-5′), 62.02 (-OC*H*$_2$CH$_3$), 53.22 (d, ^{3}J(C, P) = 8.1 Hz, -*C*H$_2$SO$_3$H), 51.13 (-*C*HCOOEt), 38.45 (C-2′), 12.75 (-OCH$_2$*C*H$_3$); ^{31}P NMR (202 MHz, D$_2$O): δ 5.50; HRMS for $C_{15}H_{23}N_6O_{10}PS$ (M-H)$^-$ calcd: 509.0861, found: 509.0884.

Oligodeoxyribonucleotides

DNA oligonucleotides P_1, P_2, T_1, T_2, T_3, T_4 and T_5 were purchased from Sigma Genosys and Eurogentec. The concentrations were determined with a Varian Cary-300-Bio UV Spectrophotometer. The lyophilized oligonucleotides were dissolved in diethylpyrocarbonate (DEPC)-treated water and stored at −20 °C. The primer oligonucleotides were 5′-labeled with [γ-^{33}P] ATP (Perkin Elmer)

using T4 polynucleotide kinase (New England Biolabs) according to the standard procedures. Labeled oligonucleotides were further purified with Illustra™ Microspin™ G-25 columns (GE Healthcare).

DNA polymerase reactions

End-labeled primer was annealed to its template by combining primer and template at a molar ratio of 1 : 2 and heating the mixture to 70 °C for 10 min, followed by slow cooling to room temperature over a period of 2 h. For the single nucleotide incorporation of **1–5**, primer P_1 was annealed to template T_1. For the single nucleotide incorporation of **6**, **7** and **8**, P_1T_3, P_2T_4 and P_1T_5 were used, respectively. A series of 20 μL batch reactions were performed for the enzyme HIV-1 RT (Ambion, Inc.; 10 U μL^{-1} stock solution). The final mixture contained 125 nM primer-template complex, RT buffer (50 mM Tris-HCl, 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM spermidine, 10 mM DTT; pH 8.3), 0.025 U μL^{-1} HIV-1 RT, and different concentrations of phosphoramidate building blocks. In the control reaction with the natural nucleotide, 10 μM or 50 μM dATP, dGTP, dTTP, dCTP were used. Mixtures were incubated at 37 °C, and aliquots (2.5 μL) were removed and quenched after 10, 20, 30, 60, 120 min. Primer elongation study with HIV-1 RT was performed in a fashion similar to the single nucleotide incorporation experiments, in which P_1 and T_2 were used. The single nucleotide incorporation assay by Taq DNA polymerase (New England Biolabs) was similarly carried out at 75 °C in ThermoPol RB buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl, 2 mM $MgSO_4$, 0.1% Triton X-100; pH 8.8).

Steady-state kinetics of single nucleotide incorporation

To determine the kinetic parameters for the incorporation of 3-phosphono-L-Ala-dAMP (**3**) and a natural nucleoside dATP, a steady-state kinetics assay was carried out. The reaction was started by adding HIV-1 RT to P_1-T_1 complex, buffer, 3-phosphono-L-Ala-dAMP (**3**) and dATP. The final mixture (20 μL) contained 0.0125 U/μL HIV-1 RT, buffer, 125 nM primer-template complex, and various concentrations of 3-phosphono-L-Ala-dAMP and dATP. The range of concentrations for phosphoramidates was optimized according to a K_m value for the incorporation of an individual nucleotide. In the case of HIV-1 RT, reaction mixtures containing the enzyme in concentration (0.0125 U μL^{-1}) to attain 5–25% incorporation and appropriate substrate concentration (10–1000 μM used for 3-phosphono-L-Ala-dAMP (**3**) and 1–20 μM used for dATP) were incubated at 37 °C and run for 8–10 different time intervals (1–10 min). The incorporation velocities were calculated based on the percentage of single-nucleotide extension product (P + 1 band). The kinetic parameters (V_{max} and K_m) were determined by plotting V (nM min^{-1}) *versus* substrate concentration (μM) and fitting the data point to a nonlinear Michaelis–Menten regression using GraphPad Prism software.

Electrophoresis

All polymerase reactions (2.5 μL) were quenched by the addition of 10 μL of loading buffer (90% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol, and 50 mM ethylenediaminetetraacetic acid (EDTA)). Samples were heated at 75 °C for 5 min prior to analysis by electrophoresis for 2.5 h at 2000 V on a 0.4 mm 20% denaturing gel in the presence of a 100 mM Tris-borate, 2.5 mM EDTA buffer, pH 8.3. Products were visualized by phosphor imaging. The amount of radioactivity in the bands corresponding to the products of enzymatic reactions was determined with the imaging device Cyclone and the associated Optiquant image analysis software (Perkin–Elmer).

Molecular modelling

Electrostatic charges. Atomic electrostatic charges of the 3-phosphono-L-Ala-dAMP (**3**), 3-phosphono-L-Ala-dGMP (**7**) and 3-phosphono-L-Ala-dTMP (**8**) molecules, to be used in the Amber software package[15] were calculated from the electrostatic potential at the 6-31G* level using the package Gamess[24] and a RESP fitting procedure.[25] However because these molecules carry an electric charge of –4, the classical charge assignment procedure failed and was modified to keep the same atomic charges for the nucleosides as in the amber library nucleotide residues and a distributed charge of –3 on the leaving group.

Amber parameters. The force field parameters for amino acids and nucleic acids used in the amber simulations are those from the parm99 dataset.[26] The 3-phosphono-L-Ala-dNMP angle and bond parameters were generated *via* the antechamber program and taken from comparable bonds, or angles from the AMBER force field. In particular, the parameters for the P–N linkage and the P–C bond were added.

Model building. The modelling is based on the crystal structure of RT in complex with a trapped entering triphosphate (pdb structure file 1RTD).[14a] The geometry of the 3-phosphono-L-Ala-dAMP molecule was optimized in Gamess in the AM1 force field.[24] The technique of flexible superposition by changing dihedral angles (the χ angle and the angles in the P-bound leaving group) and optimizing the atomic overlap was used to fit the 3-phosphono-L-Ala-dNMP structure on the entering triphosphate TTP from the 1RTD structure. In the case of 3-phosphono-L-Ala-dNMP with N = A and G, the base was modified to an adenine or guanine and the complementary adenine base (E5) was changed into a thymine or cytosine by an inverse fitting procedure using Quatfit.[27] The two Mg^{2+} ions were kept in the original position.

Molecular dynamics simulations. Solvated Molecular Dynamics was used to verify the stability.[15] The complex was solvated in a truncated octahedron TIP3P water box.[28] 31 Na^+ counter-ions were then added to get an electrostatic neutral system. The water molecules and counter-ions were then allowed to relax their positions while keeping the solute fixed. Initially, for 20 ps, the system was heated up to 300 K with constant-T, constant-V conditions while constraining the position of the solute and using a Langevin temperature equilibration scheme. The MD was then continued for 200 ps at constant T and constant P to equilibrate the system further. The simulation temperature was 300 K. Molecular dynamics simulations were then initiated with most restraints removed (except restraints on the first bases of the template and primer) with periodic boundary conditions, using a cut-off distance of 8 Å for the non-bonded interactions and the particle-mesh-Ewald method for the summation of the Coulombic interactions,[29] MD time step = 0.002 ps. Finally all restraints were removed. Unfortunately, after removing all restraints, all systems became unstable, probably

because the high charge of –4 on the 3-phosphono-L-Ala–dNMP molecules. Therefore only partially restrained models can be shown (first template residues N5.E, C6.E and the 3′ end primer residue G22.F were restrained in their movement).

Acknowledgements

This work was financed by a grant from K. U. Leuven (GOA) and from EC (Orthosome). We are grateful to Dr Jef Rozenski for HRMS and Chantal Biernaux for editorial help. We are also indebted to Luc Baudemprez for running the NMR spectra. Shiqiong Yang is obliged to Chinese Scholarship Council for a Ph.D. fellowship.

Notes and references

1 A. J. Berdis, *Chem. Rev.*, 2009, **109**, 2862–2879; S. G. Sarafianos, K. Das, J.-P. Ding, P. L. Boyer, S. H. Hughes and E. Arnold, *Chem. Biol.*, 1999, **6**, R137–146; P. H. Patel, A. Jacobo-Molina, J.-P. Ding, C. Tantillo, A. D. Clark, R. Raag, R. G. Nanni, S. H. Hughes and E. Arnold, *Biochemistry*, 1995, **34**, 5351–5363.
2 (*a*) N. A. Boyle, V. K. Rajwanshi, M. Prhavc, G. Wang, P. Fagan, F. Chen, G. Ewing, J. L. Brooks, T. Hurd, T. W. Bruisce and P. D. Cook, *J. Med. Chem.*, 2005, **48**, 2695–2700; (*b*) B. A. Mulder, S. Anaya, P. Yu, K. W. Lee, A. Nguyen, J. Murphy, R. Willson, J. M. Briggs, X.-L. Gao and S. H. Hardin, *Nucleic Acids Res.*, 2005, **33**, 4865–4873.
3 H. Strobel, L. Dugué, P. Marlière and S. Pochet, *Nucleic Acids Res.*, 2002, **30**, 1869–1878; M. J. Lutz, H. A. Held, M. Hottiger, U. Hübscher and S. A. Benner, *Nucleic Acids Res.*, 1996, **24**, 1308–1313.
4 (*a*) O. Adelfinskaya and P. Herdewijn, *Angew. Chem., Int. Ed.*, 2007, **46**, 4356–4358; (*b*) O. Adelfinskaya, M. Terrazas, M. Froeyen, P. Marlière, K. Nauwelaerts and P. Herdewijn, *Nucleic Acids Res.*, 2007, **35**, 5060–5072; (*c*) M. Terrazas, P. Marlière and P. Herdewijn, *Chem. Biodiversity*, 2008, **5**, 31–39.
5 A. Giraut, N. Dyubankova, X.-P. Song and P. Herdewijn, *ChemBioChem*, 2009, **10**, 2246–2252.
6 I. Zlatev, A. Giraut, F. Morvan, P. Herdewijn and J-J. Vasseur, *Bioorg. Med. Chem.*, 2009, **17**, 7008–7014.
7 F.-Q. Huang, G.-C. Wang, T. Coleman and N. Li, *RNA*, 2003, **9**, 1562–1570.
8 T. W. Abraham, T. I. Kalman, E. J. McIntee and C. R. Wagner, *J. Med. Chem.*, 1996, **39**, 4569–4575.
9 H. Sawai, *Bull. Chem. Soc. Jpn.*, 1990, **63**, 692–696.
10 (*a*) A. R. Kore, M. Shanmugasundaram, I. Charles, A. V. Vlassov and T. J. Barta, *J. Am. Chem. Soc.*, 2009, **131**, 6364–6365; (*b*) K. Huang and P. A. Frey, *J. Am. Chem. Soc.*, 2004, **126**, 9548–9549.
11 A. Giraut, X.-P. Song, M. Froeyen, P. Marlière and P. Herdewijn, *Nucleic Acids Res.*, 2010, **38**, 2541–2550.
12 (*a*) B. Marchand, E. P. Tchesnokov and M. Götte, *J. Biol. Chem.*, 2007, **282**, 3337–3346; (*b*) L. Menéndez-Arias, *Virus Res.*, 2008, **134**, 124–146.
13 (*a*) M. S. Boosalis, J. Petruska and M. F. Goodman, *J. Biol. Chem.*, 1987, **262**, 14689–14696; (*b*) M. F. Goodman, S. Creighton, L. B. Bloom and J. Petruska, *Crit. Rev. Biochem. Mol. Biol.*, 1993, **28**, 83–126.
14 (*a*) H. Huang, R. Chopra, G. L. Verdine and S. C. Harrison, *Science*, 1998, **282**, 1669–1675; (*b*) G. J. Klarmann, R. A. Smith, R. F. Schinazi, T. W. North and B. D. Preston, *J. Biol. Chem.*, 2000, **275**, 359–366.
15 D. A. Case, T. E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
16 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
17 (*a*) T. A. Steitz, *Nature*, 1998, **391**, 231–232; (*b*) L. Yang, W. A. Beard, S. H. Wilson, S. Broyde and T. Schlick, *Biophys. J.*, 2004, **86**, 3392–3408.
18 Y. W. Yin and T. A. Steitz, *Cell*, 2004, **116**, 393–404.
19 P. Herdewijn and P. Marlière, *Chem. Biodiversity*, 2009, **6**, 791–808.
20 (*a*) J. S. Kittredge and R. R. Hughes, *Biochemistry*, 1964, **3**, 991–996; (*b*) N. G. Ternan, J. W. McGrath and J. P Quinn, *Appl. Environ. Microbiol.*, 1998, **64**, 2291–2294; (*c*) S. A. Tan and L. G. Tan, *Clin. Physiol. Biochem.*, 1989, **7**, 303–309; (*d*) J. D. Smith and J. H. Law, *Biochemistry*, 1970, **9**, 2152–2157; (*e*) A. Horigane, M. Horiguchi and T. Matsumoto, *Biochim. Biophys. Acta.*, 1979, **572**, 385–394.
21 T. J. Lehmann and J. W. Engels, *Bioorg. Med. Chem.*, 2001, **9**, 1827–1835.
22 J. Imai and P. F. Torrence, *J. Org. Chem.*, 1985, **50**, 1418–1426.
23 E. C. R. Smith, L. A. McQuaid, J. W. Paschal and J. DeHoniesto, *J. Org. Chem.*, 1990, **55**, 4472–4474.
24 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, Jr., *J. Comput. Chem.*, 1993, **14**, 1347–1363.
25 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.
26 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
27 Quatfit, CCL software archives, http://www.ccl.net/(accessed January 15, 2010).
28 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
29 T. E. Cheatham, III, J. L. Miller, T. Fox, T. A. Darden and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 4193–4194.