A Parallel Synthesis Scheme for Generating Libraries of DNA Polymerase Substrates and Inhibitors

Heike Strobel,[a] Laurence Dugué,[a] Philippe Marlière,[b] and Sylvie Pochet*[a]

*We report a combinatorial approach aimed at producing in a single step a large family of nucleoside triphosphate derivatives that could be tested for their ability to be substrates for DNA polymerases. We propose as a unique triphosphate building block a nucleotide with a hydrazine function anchored to an imidazole ring. Condensation between the 5′-triphosphate derivative of 1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide (dY^{NH_2}TP) and any aldehyde or ketone, followed by reduction of the intermediate hydrazones dXmTP, resulted in the corresponding hydrazides (dXnTP). Following this scheme, a series of aldehydes having various aromatic parts yielded a number of adducts dY^{NHR}TP. Vent (exo$^-$) DNA polymerase is found to be able to catalyse the single incorporation of these bulky triphosphate derivatives. Subsequent extensions of the modified pairs with canonical triphosphates resulted mainly in abortive elongations at primer+2, except after the incorporation of dY^{NHben}TP and, to a lesser extent, dY^{NHphe}TP opposite C. These results illustrate the potential of this parallel synthetic scheme for generating new substrates or inhibitors of replication in a single step.*

KEYWORDS:

combinatorial chemistry · DNA polymerases · hydrazides · nucleoside triphosphates · parallel synthesis

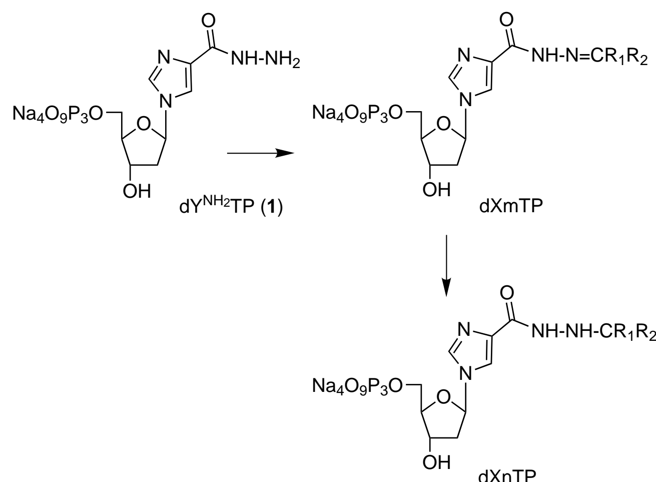
Introduction

Several research groups have recently taken up the challenge of finding new nucleoside molecular designs that can dupe polymerases into accepting and replicating them. Much attention has been focussed on research into new nucleobases that would allow the extension of the genetic alphabet,[1–3] hydrophobic nucleosides as shape analogues of the natural bases,[4] self-pairing motifs,[5] or ambiguous bases as mutagenic agents.[6–9] Although polymerases accept some synthetic modifications either in the template strand or in the incoming triphosphate, most of these modified base analogues suffer serious limitations with regard to enzyme recognition.

As a part of our effort to find new monomers accepted by DNA polymerases,[10–13] we wished to develop a parallel synthesis scheme aimed at producing a high number of nucleotides in a single step from a unique triphosphate building block.

Recently, we described the synthesis of a new nucleobase, imidazole-4-hydrazide (Y^{NH_2}), having the features of an ambiguous nucleoside.[14] We showed that the triphosphate (TP) derivative dY^{NH_2}TP is able to induce mutations during polymerase chain reactions (PCR) when it substitutes for deoxyadenosine triphosphate (dATP) or deoxyguanosine triphosphate (dGTP). In addition to its ambivalent pairing mode, the nucleoside contains a reactive hydrazino function that renders this nucleobase an attractive building block for a systematic exploration of this ambivalent motif.

The condensation between the 5′-triphosphate derivative of 1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide (**1**, dY^{NH_2}TP) and any aldehyde or ketone, followed by the in situ reduction of the resulting hydrazones (dXmTP), could generate in one step a large variety of nucleotides (dXnTP) (Scheme 1).



***Scheme 1.** Proposed synthetic route to a large family of nucleotides.*

These analogues could be used in elongation reaction assays catalysed by DNA polymerases in order to find monomers that have highly specific or ambiguous base-pairing schemes or to

[a] Dr S. Pochet, Dr H. Strobel, L. Dugué
Institut Pasteur
Unité de Chimie Organique (URA CNRS 2128)
28 rue du Docteur Roux
75724 Paris cedex 15 (France)
Fax: (+33) 1-4568-8404
E-mail: spochet@pasteur.fr

[b] Dr P. Marlière
Evologic SA
89 rue Henri Rochefort
91000 Evry (France)

 © 2002 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim 1439-4227/02/03/12 $ 20.00+.50/0

identify new inhibitors of DNA replication. As a result of our exploratory scheme, we report the discovery of a base that pairs like G although it doesn't contain a purine ring.

Results

Synthesis of nucleoside triphosphate derivatives from a unique building block

The 5′-triphosphate derivative of 1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide (dY^{NH_2}TP, **1**) was synthesised according to the previously reported procedure.[14] Our combinatorial scheme was assessed by condensation between the hydrazine function of **1** and a first series of aldehydes (Scheme 2). We chose six aldehydes with aromatic parts of various sizes (one – four rings) for the purpose of evaluating the effect of increased aromatic surface area on the recognition of these analogues by DNA polymerases. The phenyl moiety was linked to the aldehyde function with three chains of different length or flexibility. In order to increase the stability of the triphosphate derivatives under enzymatic polymerisation conditions, the coupling step was followed by the reduction of the resulting hydrazones (dXmTP) into the corresponding hydrazides (dXnTP) with sodium cyanoborohydride in phosphate buffer.

Treatment of **1** with benzaldehyde, *trans*-cinnamaldehyde, 3-phenyl-propionaldehyde and phenanthrene-9-carboxaldehyde, followed by the reduction step, afforded the hydrazide derivatives **2** (dY^{NHben}TP), **3** (dY^{NHtci}TP), **4** (dY^{NHpro}TP) and **6** (dY^{NHphe}TP), respectively (Scheme 2). However the reduction of the derivatives from the condensation between **1** and 2-quinoline aldehyde and between **1** and pyrene-1-carboxaldehyde could not not achieved, and the hydrazones **5** (dY^{NHchi}TP) and **7** (dY^{NHpyr}TP) were isolated, as confirmed by mass analysis (Table 1). The use of a higher concentration of sodium cyanoborohydride or sodium borohydride in dry CH_3CN failed to give the corresponding hydrazides. For further investigations we worked with the unreduced products **5** and **7**.

Table 1. Characteristics of compounds 1 – 7.

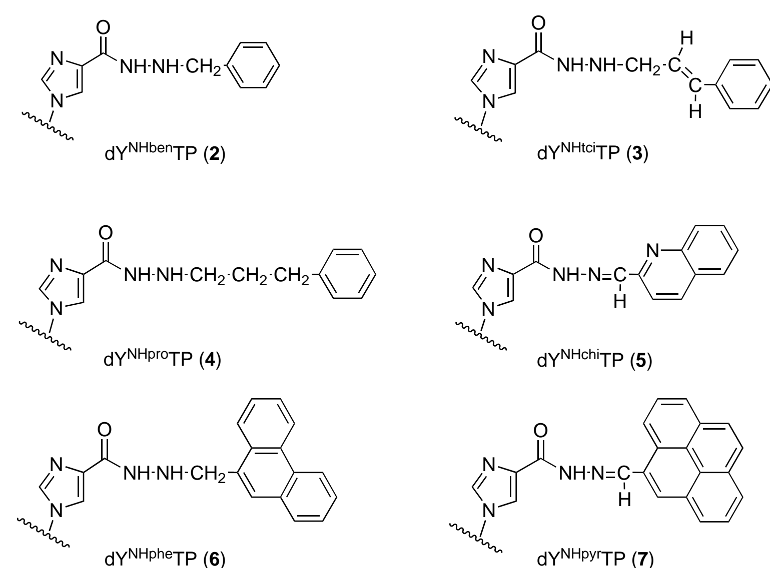
Compound	Aldehyde[a]	HPLC[b] eluent	HPLC[b] t_R [min]	MS $[M-H]^-$	Yield[c] %
1	none	5 – 25 %	4.5	481.0	
2	benzaldehyde	0 – 20 %	15.5	571.8	88
3	*trans*-cinnamaldehyde	0 – 55 %	15.5	597.0	40
4	3-phenyl-propionaldehyde	0 – 50 %	13.0	599.0	44
5	2-quinoline-aldehyde	0 – 50 %	9.0	620.8	49
6	phenanthrene-9-carboxaldehyde	0 – 75 %	7.0	671.0	38
7	pyrene-1-carboxaldehyde	0 – 50 %	9.5	693.0	18

[a] Reaction conditions between **1** and each aldehyde listed above are given in the Experimental Section. [b] Conditions of HPLC purification (gradients for elution and retention times t_R) are indicated for each compound. [c] Yields are calculated for the triphosphate derivatives taking the λ_{max} and ε of each aldehyde.

Recognition of the analogues dY^{NHR}TP by Vent (exo$^-$) DNA polymerase

In a previous study, we have selected the thermostable Vent (exo$^-$) DNA polymerase as being the most efficient for the incorporation of dY^{NH_2}TP (**1**).[14] Thus, under experimental conditions that minimised the frequency of base misincorporations, Vent (exo$^-$) polymerase incorporated one dY^{NH_2}MP (MP = monophosphate) in response to each of the four bases with a base preference for **1** that could be ranked as A ≈ T > G > C in the template. Extension by the canonical bases after the incorporation of **1** yielded the corresponding full-length product.

The capability of the analogues dY^{NHR}TP (**2** – **7**) to be incorporated into DNA was evaluated by primer extension reactions catalysed by Vent (exo$^-$) DNA polymerase. Their incorporation opposite each canonical base was monitored by the use of a primer (17-mer) labeled with ^{32}P at the 5′-end and annealed to four different templates (22-mers) with homopolymeric tails of five residues. The enzyme and hybrid concentrations were as determined for the incorporation of dY^{NH_2}TP, and the concentrations of the analogues **2** – **7** were increased from 10 μM to 1 mM. All dY^{NHR}TP derivatives were found to be substrates for Vent (exo$^-$) DNA polymerase. Compound **2** (dY^{NHben}TP) was incorporated with a better efficiency than the other derivatives. Figure 1 illustrates the incorporation of **2** as a function of time and nucleotide concentration opposite each canonical base. The elongation of the primer was complete when **2** was used at a concentration of 10 μM or 100 μM, opposite A, C or T in the template. In the incorporation assay with template $(G)_5$, the elongation of the primer was partial in the presence of compound **2** (70 %), even in the presence of deoxycytidine triphosphate (dCTP), under our experimental conditions. With compounds **3** (dY^{NHtci}TP), **6** (dY^{NHphe}TP) and **7**



*Scheme 2. Partial chemical structures of dY^{NHR}TP derivatives (**2** – **7**) synthesized from **1**.*

Figure 1. *Time-dependent incorporation of compound **2** opposite A, C, G and T. Concentrations used were 0.033 U μL⁻¹ of Vent (exo⁻) DNA polymerase, 100 nM primer-template and dNTPs as indicated. – indicates the blank reaction in the absence of dNTPs.*

(dY^{NHpyr}TP), concentrations of up to 100 μM were required for the complete elongation of the primer opposite A, C and T. The incorporation assay with template $(A)_5$ or template $(C)_5$ is shown in Figures 2 A, B and C for compounds dY^{NHtci}TP (**3**), dY^{NHphe}TP (**6**) and dY^{NHpyr}TP (**7**), respectively. Derivatives **4** (dY^{NHpro}TP) and **5** (dY^{NHchi}TP), which composed the third group, were the least accepted substrates: only 50 – 70% of the primer-templates $(A)_5$, $(C)_5$ or $(T)_5$ were elongated even in the presence of 1 mM of the analogues (data not shown). In the incorporation assay with template $(G)_5$, the elongation of the primer with the derivatives **3** – **7** was partial (data not shown). The incorporation of the analogues was ranked according to their efficiency: **2** > **3**, **6**, **7** > **4**, **5**.

In a second assay, the extension reaction after the incorporation of one Y^{NHR} residue at the 3′-extremity of the primer was investigated. Each duplex (corresponding to each canonical base in the template) was first incubated in the presence of one of the dY^{NHR}TP derivatives, and the four resulting hybrids were further elongated with the complementary canonical bases. Opposite A, T and G, the extensions stopped after the addition of one canonical base (primer+2) whichever analogue dY^{NHR}MP was first incorporated. As an example, results obtained from experiments with template $(A)_5$ are shown in Figure 3 A. The 22-mer (primer+5) detected in the case of elongation of the primer by compound **5** is attributed to the elongation by deoxythymidine triphosphate (dTTP) of the remaining 17-mer (Figure 3 A). Similar results (data not shown) were obtained with templates $(T)_5$ and $(G)_5$. Opposite C, full-length DNA synthesis (primer+5) was detected in the presence of compound **2** (dY^{NHben}TP; 100%) or **6** (dY^{NHphe}TP; 20%) followed by the addition of dGTP (Figure 3 B). In the presence of the other derivatives **4**, **5** and **7** with dGTP, the elongation stopped at primer+2, and the 22-mer products (primer+5) correspond to the elongation by natural nucleotides of the remaining 17-mer primer (Figure 3 B).

Discussion

The feasibility of the proposed exploratory scheme was demonstrated by the condensation between the 5′-triphosphate derivative of 1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide (**1**, dY^{NH_2}TP) and a first series of six aldehydes with aromatic rings of different sizes. Surprisingly, nucleotides with large adducts, such as compounds **6** (dY^{NHphe}TP) and **7** (dY^{NHpyr}TP), were tolerated as incoming triphosphates by Vent (exo⁻) DNA polymerase. Addition of one dY^{NHR}MP (**3** – **7**) to the primer strand was detected opposite each of the four canonical bases in the template with any preference order of the template base. But further addition of canonical nucleotides past the incorporation of the analogues resulted mainly in chain termination due to the inability of the polymerase to accommodate these bulky analogues in the nascent primer/template duplex.

Compound **2** (dY^{NHben}TP) gave an intermediate pattern compared to dY^{NH_2}TP (**1**) and other adducts (**3** – **7**): a) single

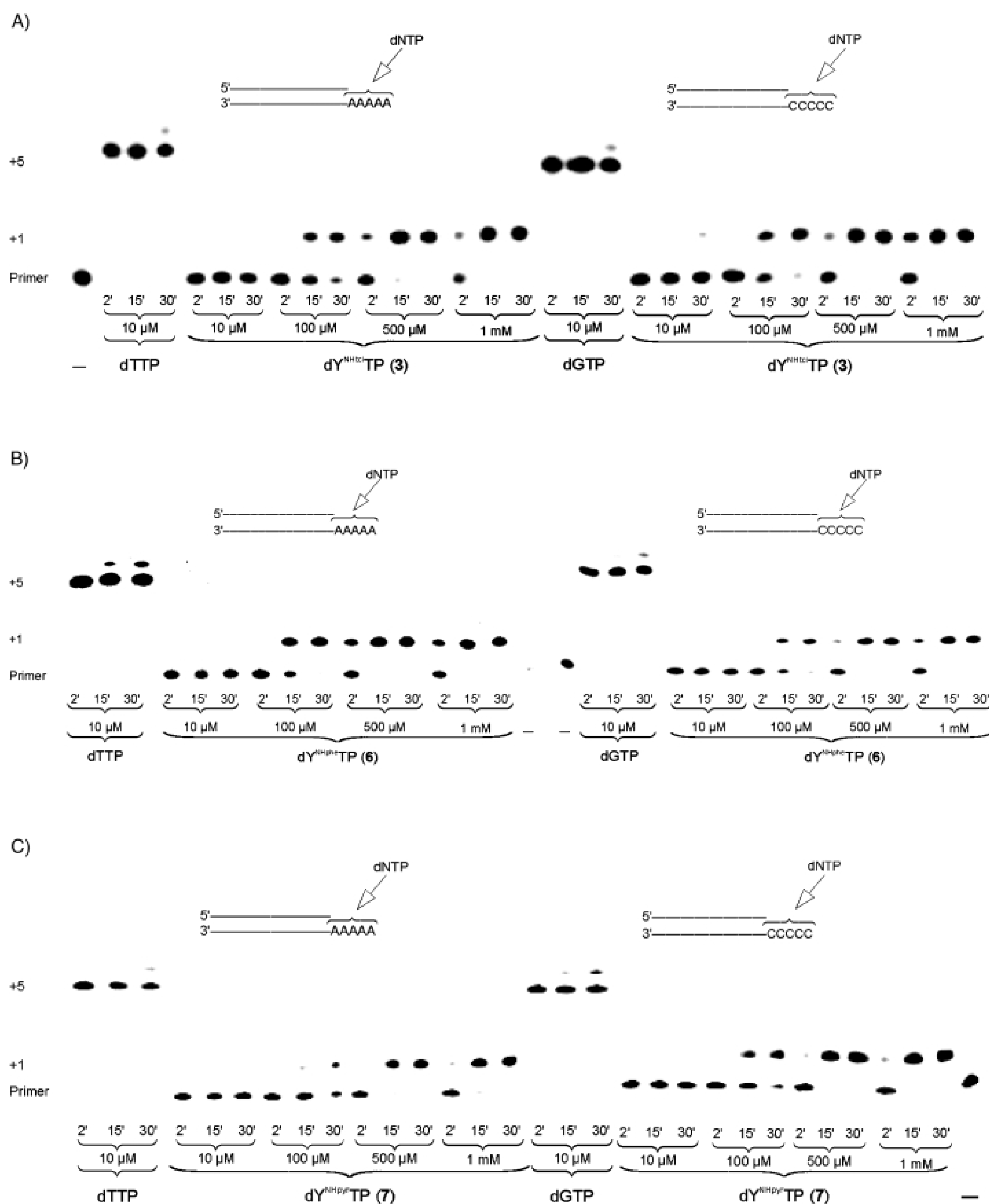


*Figure 2. Time-dependent incorporation of analogues **3** (A), **6** (B) and C) **7** (C). Concentrations were as given in the legend of Figure 1. – indicates the blank reaction in the absence of dNTPs.*

incorporation of **2** opposite A, T and C occurred with a similar efficiency, while **1** showed a preference for A and T in the template and the derivatives **3**–**7** were less accepted by Vent (exo$^-$) DNA polymerase; b) extension by canonical bases after the incorporation of **2** was observed in only one context (when dY^{NHben} paired with C) while extension after the incorporation of **1** occurred in any sequence context without chain termination.[14] The single incorporation efficiency of compound **2** compared that of compounds **3** and **4** could be explained by the reduced flexibility of the phenyl part in **2** that placed the ring within the double-helix structure.

The introduction of the benzyl group in dY^{NHben} precludes the free rotation around the carboxamide and glycosidic bonds of the modified imidazole when it is located at the 3′-extremity of the primer opposite a canonical base, compared to that observed with the other imidazole derivatives like dY[15] and dY^{NH_2}.[14] The base pairing of Y^{NHben} with C traced on the G:C pairing places the aromatic hydrophobic part in the minor

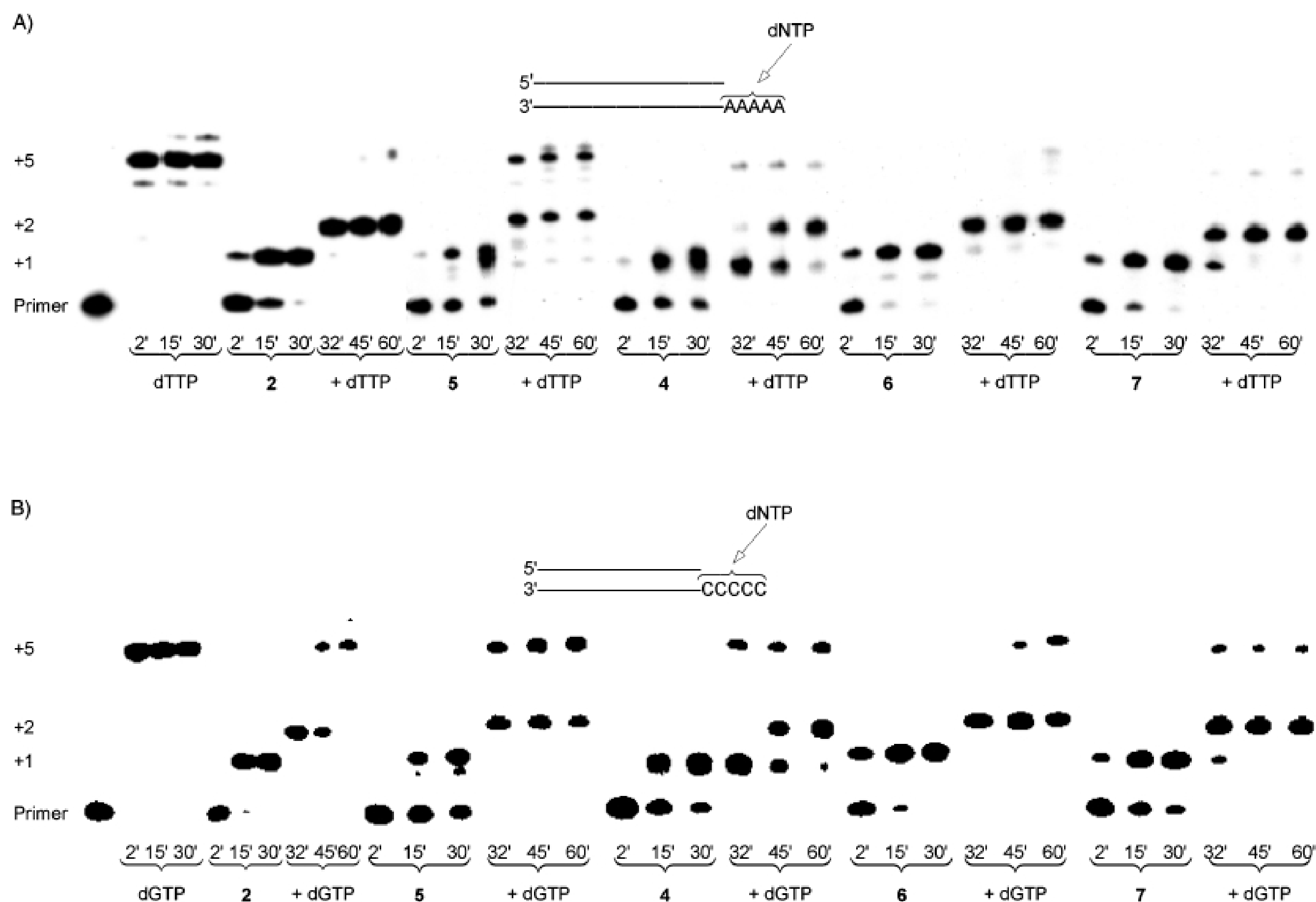
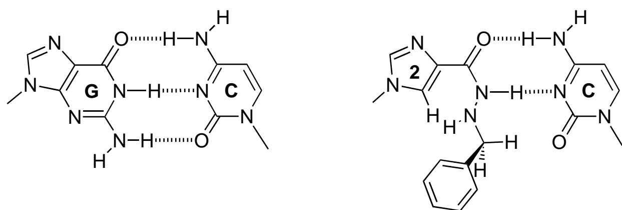


Figure 3. DNA synthesis with canonical or modified nucleoside triphosphates. Conditions: 0.033 UμL^{-1} of Vent (exo$^-$) DNA polymerase, 100 nM primer-template with a tail of five A (A) or C (B) residues, 100 μM canonical complementary dNTPs or 1 mM dY^{NH_2}TP followed by 100 μM canonical complementary dNTPs, time incubation as indicated. The first lane is the blank reaction in the absence of dNTPs.

groove of the duplex (Scheme 3). In the other base pairing patterns, the aromatic part and the hydrazide group are in the major groove, which presents a more hydrophilic environment than the interior of the helix.



Scheme 3. The 2:C base pair based on the G:C trace.

Such specificity in the recognition of azole derivatives by DNA polymerases is without precedent. Using our parallel scheme we found a new motif that is able to form a base pair with a canonical base and whose structure differs from the canonical model. Whether hydrogen bonds between Y^{NHben} and C or the hydrophobic context influence the polymerisation process remains to be addressed experimentally. Insertion of a deoxynucleoside triphosphate (dNTP) by a DNA polymerase is a multistep mechanism that involves a conformational rearrangement from an open to a closed ternary complex. Both the geometric fit of the incoming triphosphate and its ability to form Watson – Crick hydrogen bonds with the template nucleotide were found important for the stability of the closed ternary complex.[16] This combinatorial scheme could help in the rapid evaluation of shape and hydrogen-bond contributions through the use of these analogues.

A systematic synthesis of deoxyuridine triphosphate derivatives through the condensation of a 5-aminoalkyl-deoxyuridine triphosphate and various *N*-hydroxysuccinimide esters was reported with the purpose of DNA diversification.[17] The high reactivity of the hydrazine function and the huge diversity of commercially available aldehydes and ketones allow the scheme presented here to be considered as a general scheme for other nucleobases in order to provide numerous purine or pyrimidine analogues that are suitable for DNA diversification.

Materials and Methods

General methods: Electrospray mass spectra were collected on an electrospray triple quadruple spectrometer (API 365/PE SCIEX) working in the negative mode with CH_3CN/0.4% NH_4OH (50:50) or 0.5%

Et_3N in water (pH 10). Reagents and enzymes were purchased from the following sources: ultrapure dNTP solutions from Amersham Pharmacia Biotech, Vent (exo$^-$) DNA polymerase and *T4* polynucleotide kinase from New England Biolabs, sodium cyanoborohydride (3.0 g L^{-1} $NaBH_3CN$) in 20 mM sodium phosphate (pH 7.5) with 0.2 M NaCl from Sigma.

Reaction of the 5′-triphosphate derivtaive of 1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide with aldehydes: 1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-imidazole-4-hydrazide-5′-O-triphosphate (**1**) was synthesised as previously reported.[14] Condensation reactions were carried out by adding 100 mM stock solution of each aldehyde in methanol or ethanol (2.5 mL; 250 μmol) to compound **1** (50 μmol) in water (0.5 mL). After stirring for 2 h at 4 °C, reaction mixtures were evaporated to dryness and then treated overnight at 4 °C with sodium cyanoborohydride (3.0 g L^{-1} $NaBH_3CN$; 270 μmol) in 20 mM sodium phosphate (pH 7.5) with 0.2 M NaCl. The resulting solutions were reduced to a small volume, and insoluble material was removed by filtration. After freeze-drying, the crude products were purified by reverse-phase HPLC (C18 column) with a linear gradient of acetonitrile (A) in 10 mM triethylammonium acetate buffer (B) over 20 min. Elution was monitored by UV detection at 254 nm. Fractions containing pure products were lyophilised and passed through a cation-exchange resin (Dowex 50WX8, Na^+ form) to afford the triphosphates **2** – **7** as their sodium salts (18 – 88 % yield). Conditions of purification (retention times) and mass spectrometry data are given in Table 1. The hydrazones **5** and **7** were found stable since no decomposition product (reversible hydrolysis into **1**) could be detected after a long storage at −20 °C in water or after incubation at 50 °C for 60 min.

Primer extension reaction: 5′-End labelling of the primer (5′-CAGGAAACAGCTATGAC-3′, 1 μM) was performed in 70 mM tris(hydroxymethyl)aminomethane (Tris)/HCl (pH 7.6), 10 mM $MgCl_2$ and 5 mM 1,4-dithiotreitol (DTT) by addition of *T4* polynucleotide kinase (10 U) and [γ-^{32}P]-ATP (10 Ci mmol^{-1}) in a final volume of 50 μL. The mixture was incubated for 30 min at 37 °C, then for 15 min at 70 °C and finally stored at 4 °C. Annealing was realised in a final volume of 30 μL by incubating the appropriate 22-mer templates (6 μL; 10 μM), primer (2 μL; 10 μM) and ^{32}P-end-labeled primer (10 μL; 1 μM) for 15 min at 75 °C, followed by slowly cooling the mixture to room temperature over the course of 1 h. Time-dependent incorporations were carried out by incubating Vent (exo$^-$) DNA polymerase (0.033 U μL^{-1}) in the presence of 100 nM primer-template and dNTPs (concentrations as indicated in the legends of the figures) at 50 °C in ThermoPol reaction buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris/HCl (pH 8.8), 2 mM $MgSO_4$, 0.1 % Triton X-100). One unit catalyses the incorporation of 10 nmol of dNTP into an acid-insoluble material in 30 min at 75 °C in 1 × buffer by using 0.2 mg mL^{-1} of activated DNA. Elongation reactions were started by mixing diluted Vent (exo$^-$) DNA polymerase (6 μL) in 2 × polymerase buffer and annealed primer-template (4 μL; 600 nM) with dNTPs (2 μL). Aliquots of elongation reactions were taken at time intervals of 2, 15 and 30 min and quenched by adding an equal volume of the loading buffer (0.02 % (w/v) bromophenol blue, 0.02 % (w/v) xylene cyanol FF, 50 % (v/v) formamide, 50 % (v/v) water). Before loading onto a polyacrylamide gel (20 %, 7 M urea), samples were heated for 1 min at 75 °C. After electrophoresis, products were visualised by autoradiography.

This work was supported by grants from the Institut Pasteur and the Centre National de la Recherche Scientifique, and by a fellowship from the Alfred Kastner foundation to H.S. We are grateful to Prof. Pierre Potier for his encouragement during this work and Dr. Catherine Papanicolaou for fruitful comments.

[1] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33 – 37.
[2] M. J. Lutz, H. A. Held, M. Hottiger, U. Hubscher, S. A. Benner, *Nucleic Acids Res.* **1996**, *24*, 1308 – 1313.
[3] Y. Q. Wu, A. K. Ogawa, M. Berger, D. L. McMinn, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 7621 – 7632.
[4] E. T. Kool, *Biopolymers* **1998**, *48*, 3 – 17.
[5] S. Pochet, P. Marlière, *C. R. Acad. Sci. Paris* **1996**, *319*, 1 – 7.
[6] R. Nichols, P. C. Andrews, P. Zhang, D. E. Bergstrom, *Nature* **1994**, *369*, 492 – 493.
[7] M. Sala, V. Pezo, S. Pochet, S. Wain-Hobson, *Nucleic Acids Res.* **1996**, *24*, 3302 – 3306.
[8] M. Berger, Y. Wu, A. K. Ogawa, D. L. McMinn, P. G. Schultz, F. E. Romesberg, *Nucleic Acids Res.* **2000**, *28*, 2911 – 2914.
[9] D. Loakes, *Nucleic Acids Res.* **2001**, *29*, 2437 – 2447.
[10] S. Pochet, L. Dugué, *Nucleosides Nucleotides* **1998**, *17*, 2003 – 2009.
[11] J. J. Beaussire, S. Pochet, *Nucleosides Nucleotides* **1999**, *18*, 403 – 410.
[12] K. Vastmans, S. Pochet, A. Peys, L. Kerremans, A. Van Aerschot, C. Hendrix, P. Marlière, P. Herdewijn, *Biochemistry* **2000**, *39*, 12 757 – 12 765.
[13] K. Vastmans, M. Froeyen, L. Kerremans, S. Pochet, P. Herdewijn, *Nucleic Acids Res.* **2001**, *29*, 3154 – 3163.
[14] H. Strobel, L. Dugué, P. Marlière, S. Pochet, *Nucleic Acids Res.* **2002**, *30*, 1869 – 1878.
[15] S. Pochet, L. Dugué, A. Meier, P. Marlière, *Bioorg. Med. Chem. Lett.* **1995**, *5*, 1679 – 1684.
[16] L. Dzantiev, Y. O. Alekseyev, J. C. Morales, E. T. Kool, L. J. Romano, *Biochemistry* **2001**, *40*, 3215 – 3221.
[17] K. Sakthivel, C. F. Barbas III, *Angew. Chem.* **1998**, *110*, 2998 – 3002; *Angew. Chem. Int. Ed.* **1998**, *37*, 2872 – 2875.

Received: July 3, 2002 [F 448]