

Challenging artificial genetic systems: thymidine analogs with 5-position sulfur functionality

Heike A. Held¹ and Steven A. Benner^{1,2,*}

¹Department of Chemistry and ²Department of Anatomy and Cell Biology, University of Florida, Gainesville, FL 32611-7200, USA

Received April 8, 2002; Revised and Accepted July 5, 2002

ABSTRACT

Eight different polymerases, chosen from evolutionary families A (*Taq*, *Tfl*, *HotTub* and *Tth*) and B (*Pfu*, *Pwo*, *Vent* and *Deep Vent*), were examined for their ability to incorporate 5-position modified 2'-deoxyuridine derivatives that carry a protected thiol group appended via different linkers containing either three or four carbon atoms. This represents the first attempt to incorporate the thiol functionality into DNA via enzymatic synthesis. Each polymerase–substrate combination was evaluated using a hierarchy of increasingly more difficult challenges, starting with incorporation of a single derivative, proceeding to incorporation of two derivatives at adjacent sites and non-adjacent sites, then examining the ability of the polymerase to accept the derivative within the template, and concluding with a challenge involving PCR. The evaluation of thiol-bearing 2'-deoxyuridine derivatives was then extended to consider their chemical stabilities. Stability was found to be less than satisfactory when the thiol functionality has a 'propargylic' relationship to the unsaturation in the linker. The best polymerase–appendage combination used the polymerase from *Pyrococcus woesei* (*Pwo*) and the 5'-tBu-SS-CH₂-CH₂-C≡C- linker. This pair supported PCR amplification and therefore should have value in artificial *in vitro* selection experiments. Indeed, we discovered that *Pwo* and *Pfu* preferred the derivative triphosphate over TTP, the natural substrate, in competition studies. These studies confirm an earlier suggestion that membership of an evolutionary family of polymerases is a partial predictor of the ability of the polymerase to accept 5-modified 2'-deoxyuridines. Considerable differences are displayed by different members within a polymerase family, however. This remains curious, as the ability of the polymerase to replicate natural DNA with high fidelity and its propensity to

exclude unnatural analogs are presumed to be correlated.

INTRODUCTION

It has been two decades since the enzymic incorporation of functionalized nucleotides was first suggested to be useful to tag DNA with a functionality that it did not intrinsically carry (1,2). In the intervening years, enzymic incorporation of modified nucleotides has become the key to automated DNA sequencing (3). Also, increasing the functionality of nucleic acids has been suggested to be a key to enhancing the power of nucleic acids as catalysts (4). To implement a functionalized genetic molecule, the DNA alphabet was expanded some time ago from four to 12 letters (5,6) and modified standard nucleotides have been incorporated into *in vitro* evolution experiments (7–17). The ultimate goal, of course, is a combination of polymerases and unnatural/functionalized DNA-like molecules that support repeated copying and recopying with an efficiency and fidelity sufficient to support Darwinian evolution *in vitro*, and ultimately in cells.

While many polymerases are known to incorporate nucleoside analogs, others have proved to be quite idiosyncratic when challenged with nucleoside analogs (18), even though all polymerases are believed to be descendants from a single common ancestor (19). Many polymerases accept pyrimidine species carrying substituents at the 5-position of the nucleobase, at least to some extent. Crystal structures suggest that space is available in the active site for several polymerases to accommodate substituents at this position (20–22). However, successful expansion of the genetic alphabet requires incorporation of nucleoside analogs with high efficiency. Neither the crystal structure nor organic chemistry theory has sufficient resolution to predict which polymerase will accept which substituted pyrimidine derivative best.

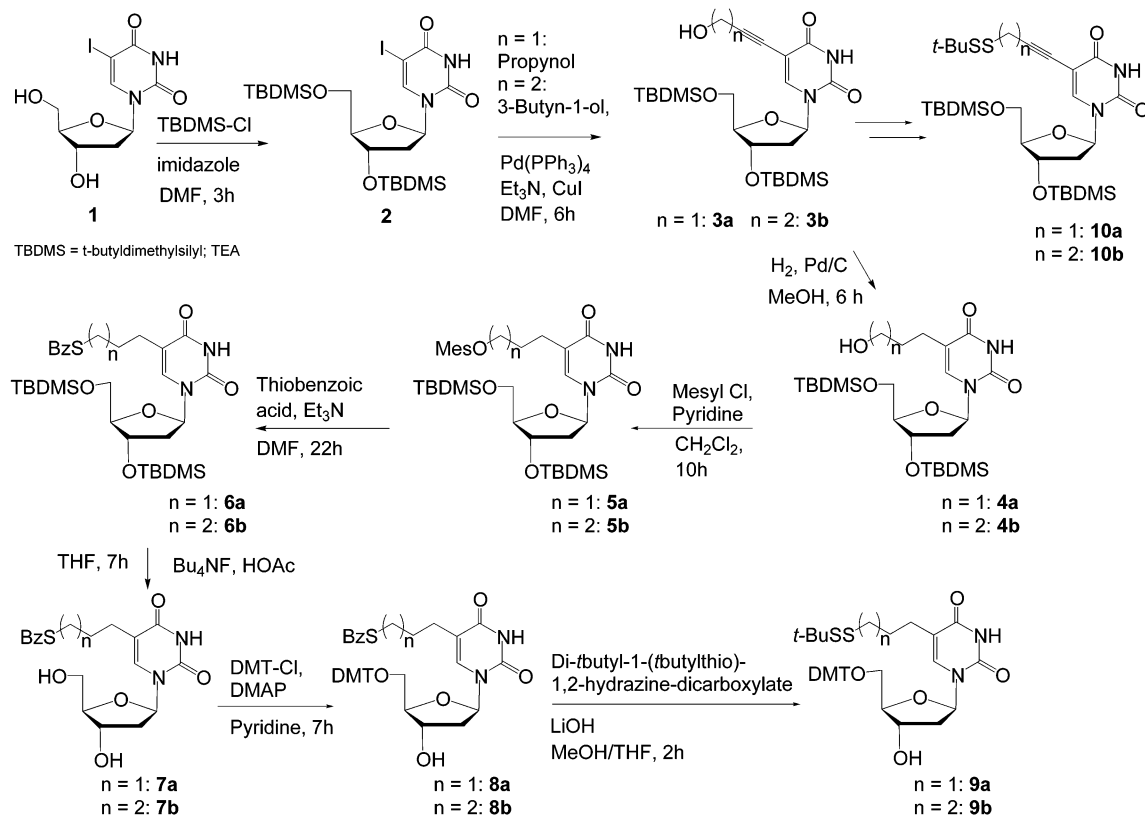
For these reasons, rational empiricism remains an essential part of any effort to generate an artificial genetic system that incorporates 5-position modified pyrimidines the best. Simple tests for nucleotide analog incorporation (such as incorporation of a single nucleotide analog in a primer extension reaction), while useful as a first level, need not, however, be good indicators of the suitability of a derivative for the

*To whom correspondence should be addressed at: Department of Chemistry, University of Florida, Box 117200, Gainesville, FL 32611-7200, USA.

Tel: +1 352 392 7773; Fax: +1 352 846 2095; Email: benner@chem.ufl.edu

Present address:

Heike A. Held, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA



Scheme 1.

long-term goal. Likewise, it is clear that simply challenging a polymerase with a particular analog in a PCR reaction is not a strategic way of developing our understanding of polymerases as they interact with unnatural substrates.

Rather, a set of graded challenges is needed to evaluate a polymerase–substrate combination for its suitability as part of an expanded genetic information system. This set might begin with simple primer extension experiments challenging the polymerase to incorporate a single nucleotide analog as a triphosphate. In further challenges, however, the polymerase must show its ability to incorporate more than one nucleotide analog, at both adjacent and non-adjacent sites, and to accept the analog in the template. Last, the polymerase might be challenged to amplify, in a PCR reaction, oligonucleotides containing the derivative.

Lee *et al.* recently reported a systematic study of the ability of a single polymerase (*Taq*) to accept 5-position modified 2'-deoxyuridines carrying four linkers, each with three carbon atoms and bearing an amino group (23,24). This work prompts us to report our work examining a range of polymerases for their ability to incorporate a series of 2'-deoxyuridine derivatives carrying a thiol functionality appended via different linking groups. Here, we exploit this hierarchy of challenges to evaluate the suitability of a variety of polymerase–analog combinations as part of an expanded alphabet.

MATERIALS AND METHODS

Nucleoside synthesis

Full experimental procedures describing the synthesis of the compounds used in this work are available in Supplementary

Material. The 2'-deoxyuridine derivatives carrying a 5-position appended functionality were prepared by palladium(0)-catalyzed coupling of the appropriate unit to 5-iodo-2'-deoxyuridine (Scheme 1). For example, 3',5'-bis-*O*-(*tert*-butyldimethylsilyl)-5-(3-hydroxy-propynyl)-2'-deoxyuridine **3a** was prepared in two steps by coupling 5-iodo-2'-deoxyuridine **1** with propargyl alcohol using a Pd(0) catalyst following a procedure adapted from Osborne *et al.* (25; Scheme 1). A portion of this material was reduced by catalytic hydrogenation to give 3',5'-bis-*O*-(*tert*-butyldimethylsilyl)-5-(3-hydroxypropyl)-2'-deoxyuridine **4a**. The free alcohol group was mesylated and then reacted with thiobenzoic acid to give the corresponding thiobenzoate **6a**. The silyl groups were removed, the 5'-hydroxyl group was reprotected as its dimethoxytrityl ether (compound **8a**), the thiobenzoate was cleaved and the free thiol was converted to the *tert*-butyl disulfide derivative **9a** by reacting with di-*t*-butyl-1-(*t*-butylthio)-1,2-hydrazinedicarboxylate. This gave a 2'-deoxyuridine derivative carrying a 5-position thiol attached via a four carbon saturated linker.

A corresponding series of transformations, taken from Osborne *et al.* (25), generated the disulfide **9b** between 5'-*O*-(4,4'-dimethoxytrityl)-5-(butanethiol) and *tert*-butyl disulfide. This gave a 2'-deoxyuridine derivative carrying a 5-position thiol attached via a three carbon saturated linker.

The corresponding mesylates were also prepared from the propynyl and butynyl intermediates and converted to the corresponding thiols protected as their disulfides with *tert*-butyl disulfide (**10a** and **10b**). This gave 2'-deoxyuridine derivatives carrying a 5-position thiol attached via three carbon and four carbon acetylenic linkers.

Table 1. Primer and templates employed for polymerase screening by primer extension and PCR

Pr1	3'-CA TGG GCC AAG CTT CTT CGG-5'
T1	5'-TCG TCA GTC <u>GGT ACC CGG TTC GAA GAA GCC</u> -3'
T2	5'-TCG TAA GTC <u>GGT ACC CGG TTC GAA GAA GCC</u> -3'
T3	5'-C ACA AAG ACA TCG TAA GTC <u>GGT ACC CGG TTC GAA GAA GCC</u> -3'
T4	5'-CC AS ² C GCA GS ² C <u>GGT ACC CGG TTC GAA GAA GCC</u> -3'
T5	5'-CC ATC GCA GTC <u>GGT ACC CGG TTC GAA GAA GCC</u> -3'
T6	5'-TGC AS ² S ² ACC AGT ACC CGG TTC GAA GAA GCC-3'
T7	5'-TGC ATT ATC <u>CGT ACC CGG TTC GAA GAA GCC</u> -3'
Pr2	5'-TGC CAT CTG CAT CCG TGG TTA CT-3'
Pr3	5'-CAT GCG TCA GTG TTC CAT GC-3'
T8	5'- <u>TGC CAT CTG CAT CCG TGG TTA CTC</u> GTA CCT GGT CAT ACG AAC TGG ACT CAC AGG TCA TGT TAG CGA TCC TAG ACG <u>CAT GGA ACA CTG ACG CAT G</u> -3'

Underlined positions indicate the primer binding regions of the templates.

The dimethoxytritylated nucleosides in CH₂Cl₂ were then reacted with *N,N*-diisopropylethylamine and chloro-*N,N*-diisopropylamino-β-cyanoethylphosphine to yield phosphoramidites following standard procedures.

Triphosphate synthesis

The 5'-triphosphates of the side chain modified 2'-deoxyuridine derivatives, designated **ds¹TP**, **ds²TP**, **ds³TP** and **ds⁴TP** (Fig. 1), were generated from the corresponding 5'-*O*-dimethoxytritylated nucleosides **9a**, **9b**, **10a** and **10b** by the method of Ludwig and Eckstein (26).

Oligodeoxyribonucleotides

Oligodeoxyribonucleotides containing **ds²** were synthesized on a PerSeptive Biosystems Expedite 8900 DNA synthesizer, employing standard β-cyanoethyl phosphoramidite chemistry. The oligonucleotides were purified by 15% PAGE and characterized by mass spectrometry on a MALDI TOF Voyager mass spectrometer (PerSeptive Biosystems). Standard oligonucleotides were purchased from Integrated DNA Technologies, Inc. as desalted crude products, or synthesized on a PerSeptive Biosystems Expedite 8900 DNA synthesizer and purified by PAGE. Labeling of primer Pr1 at its 5'-terminus with Redivue [γ-³²P]ATP (Amersham) was performed using T4 polynucleotide kinase (New England Biolabs).

A list of oligonucleotides used in this work can be found in Table 1.

Polymerases and reaction buffers

Pwo DNA polymerase was from Roche Molecular Systems. The 10× reaction buffer used with *Pwo* DNA polymerase (PCR buffer with MgSO₄ from Roche Molecular Systems) contained 100 mM Tris-HCl (pH 8.85 at 20°C), 20 mM MgSO₄, 250 mM KCl and 50 mM (NH₄)₂SO₄.

Vent[®] DNA polymerase and Deep Vent[®] DNA polymerase were from New England Biolabs. The 10× reaction buffer (ThermoPol reaction buffer from New England Biolabs) contained 200 mM Tris-HCl (pH 8.8 at 25°C), 20 mM MgSO₄, 100 mM KCl, 100 mM (NH₄)₂SO₄ and 1% Triton X-100.

*Pfu*Turbo[™] DNA polymerase was purchased from Stratagene. It was used with the 10× cloned *Pfu* reaction buffer from Stratagene [200 mM Tris-HCl (pH 8.8), 20 mM MgSO₄, 100 mM KCl, 100 mM (NH₄)₂SO₄, 1% Triton X-100 and 1 mg/ml nuclease-free BSA].

Tfl DNA polymerase and *Taq* polymerase (in storage buffer A) were from Promega. The 10× reaction buffer used with *Tfl* DNA polymerase was the *Tfl* DNA polymerase 10× reaction buffer from Promega [200 mM Tris-acetate (pH 9.0), 20 mM MgCl₂, 100 mM (NH₄)₂SO₄, 750 mM KOAc and 0.5% Tween 20]. *Taq* polymerase was used with the thermophilic DNA polymerase buffer, magnesium-free, from Promega [100 mM Tris-HCl (pH 9.0 at 25°C), 20 mM MgCl₂, 50 mM KCl and 1% Triton X-100]. The final MgCl₂ concentration in the assays with *Tfl* and *Taq* polymerases was adjusted to 2 mM.

Tth polymerase and HotTub[™] polymerase were purchased from Amersham/Pharmacia. The 10× *Tth* polymerase

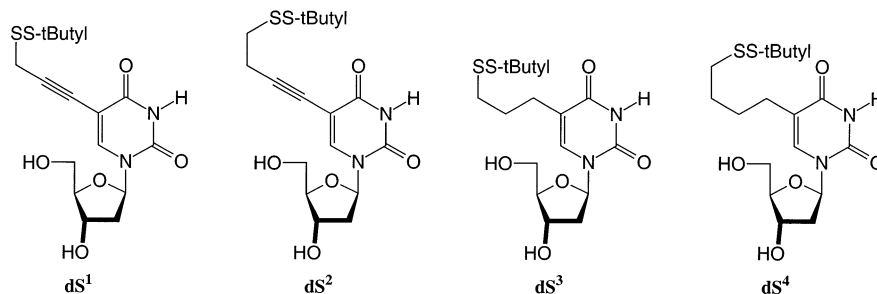


Figure 1. Structures of the substituted 2'-deoxyuridine derivatives **dS¹**, **dS²**, **dS³** and **dS⁴** modified with a 5-position side chain carrying a thiol group protected as the disulfide with *tert*-butyl thiol.

reaction buffer from Amersham was used in assays with *Th* polymerase. The final MgCl₂ concentration in the assays was adjusted to 2 mM. In assays with HotTub™ polymerase, the 10× HotTub™ reaction buffer from Amersham [500 mM Tris-HCl (pH 9.0), 20 mM MgCl₂ and 200 mM (NH₄)₂SO₄] was used.

Primer extension experiments

5'-³²P-labeled primer **Pr1** (15 pmol, final assay concentration 750 nM) was annealed to a template sequence (**T1-T7**) (18 pmol, final assay concentration 900 nM) in polymerase reaction buffer (1.25×, 16.5 μl) by heating the mixture for 2 min to 95°C and subsequently allowing the solution to cool over 2 h to room temperature. The triphosphates (18.8 nmol, final concentration 188 μM each) were added at room temperature, followed by the polymerase (1 U), to a final reaction volume of 20 μl. The reaction mixture was immediately incubated at 72°C. After 10 min, the reaction was stopped by addition of EDTA (1.25 equiv. with respect to [Mg²⁺]). The reaction products were separated by denaturing PAGE (7 M urea, 55–60°C) and visualized and quantified with a MolecularImager® (Bio-Rad, Hercules, CA).

PCR amplifications

PCR experiments were performed on a 50 μl scale in 0.2 ml thin-walled PCR tubes. Reaction mixtures contained primers **Pr2** and **Pr3** (50 pmol each, 1 μM), template **T8** (2.5 pmol, 50 nM), triphosphates (20 nmol each, 200 μM), the appropriate polymerase reaction buffer (at 1× concentration, see above) and a polymerase (1 U). In 'positive' control reactions, the triphosphates were the standard triphosphates TTP, dATP, dCTP and dGTP (Promega). In the actual test reactions, TTP was substituted with one of the modified triphosphates **dS²TP**, **dS³TP** or **dS⁴TP**. In 'negative' control reactions, water was added in lieu of TTP. The reaction mixtures were cycled through 15 (Family A polymerases) or 25 (Family B polymerases) cycles of amplification in a Stratagene Robocycler (1 min at 94°C, 2 min at 55°C and 2 min at 72°C). The mixtures were then stored on ice and the reactions were quenched by addition of EDTA (2 equiv. with respect to [Mg²⁺]). Analysis was performed by 2% agarose gel electrophoresis containing ethidium bromide. The products were visualized under UV light (254 nm). Gel images were recorded and analyzed with GelDoc® (Bio-Rad).

RESULTS

Four 2'-deoxyuridine derivatives bearing C⁵-mixed *tert*-butyl disulfide moieties (**dS¹**, **dS²**, **dS³** and **dS⁴**) were examined to learn whether they might be suitable as components of an artificial genetic system that introduces thiol functionality into oligonucleotides (Fig. 1). Both **dS³** and **dS⁴** are known previously in their 5'-dimethoxytritylated phosphoramidite forms (27), but not as triphosphates. As such, **dS³** and **dS⁴** have been incorporated into oligonucleotides by automated chemical synthesis, yielding thiol functionalized oligonucleotides through post-synthetic reduction of the disulfide moieties. They all share a *tert*-butyl blocked disulfide attached to the nucleoside through the C⁵-position of the heterocyclic base. They differed in the length and degree of saturation of the chain appending them to the heterocycle.

Two compounds are new to this work. One of them, **dS²**, carries the protected thiol on a butynyl linker. **dS²** differs from **dS⁴** because of its slimmer and more rigid linker. The propynyl *tert*-butyl disulfide substituted 2'-deoxyuridine, **dS¹**, proved to be insufficiently stable during the standard work-up of the synthetic oligonucleotides to be carried further. During synthesis of oligonucleotides containing **dS¹**, complex HPLC results were observed on crude materials after deprotection. This complexity arose even though trityl release indicated highly successful coupling. Simple TLC spot tests of the nucleoside in ammonia suggested that even after only 1 h at 55°C, **dS¹** nucleoside decomposed. UV spectra indicated that the propargyl linkage had disappeared. In contrast, both **dS³** and **dS⁴**, both lacking the 'propargylic' thiol, were found to be stable under these conditions.

We had noticed previously a lower level of instability with the analog carrying a propargylic amino group (T.R. Battersby, unpublished results). We suspect that this instability increases as the leaving group ability of the functionality increases and recommend that this consideration guide the design of other functionalized 5-position modified 2'-deoxyuridine derivatives.

Incorporation of dS^xTP opposite dA in a template

The simplest challenge for a polymerase is to place a single nucleotide site specifically into an oligonucleotide opposite a single nucleotide in a template. This is most easily implemented in a 'primer extension' (or 'run-off') experiment. A primer is annealed to a template oligonucleotide containing a single nucleotide that presents the challenge, and then

presented to a polymerase. Analysis of the products, generally by gel electrophoresis, determines how the polymerase met the challenge.

To study the incorporation of the uridine triphosphate analogs **ds^xTP**, templates **T1**, **T2** and **T3** were employed (Table 1). These oligonucleotides are composed of the four standard 2'-deoxynucleosides T, dA, dC and dG. **ds²**-containing templates **T4** and **T6** were utilized to analyze the behavior of polymerases when encountering templates containing 2'-deoxyuridine analogs carrying C⁵-position butynyl substituents. All five templates anneal to a common primer (Pr1).

In the simplest primer extension experiments studying incorporation of functionalized triphosphates **ds^xTP**, all standard nucleoside templates (**T1**, **T2** and **T3**; Table 1) were used. 5'-³²P-radiolabeled primer **Pr1** was annealed to an excess of template oligonucleotide. The primer-template complex was then incubated with triphosphates **ds^xTP**, dATP, dCTP and dGTP, and a polymerase. An analogous sample in which TTP took the place of the substituted triphosphate **ds^xTP** was prepared simultaneously. This 'positive' control reaction did not involve any modified nucleosides and therefore served as a reference for the performance of the polymerase. In a 'negative' control, water replaced the **ds^xTP** or TTP solution. After a defined incubation time, the reaction was stopped with EDTA and the reaction mixture was analyzed by denaturing PAGE.

The 'level one' challenge asks if a polymerase can incorporate a single non-standard nucleotide in a product DNA molecule opposite a single position in a template. To implement this challenge, **T1** and **Pr1** were used as a template-primer combination. Full-length product is formed by adding 10 nt to Pr1, with a single **ds^xTP** in the product introduced opposite a single A in the template. The reaction products were detected by PAGE.

Figure 2 shows a representative set of results in a 'two by two' polymerase-substrate matrix. The two polymerases were *Pwo* and *Vent* and the two substrates were the propynyl-linked **ds¹TP** and butynyl-linked **ds²TP** analogs.

Three features of the results are noteworthy. First, the major band observed with both polymerases and both **ds^xTP** substrates corresponded to full-length product. This indicated that both polymerases accepted both substrates under these conditions. Second, a small amount of an '*n* - 1' product was formed in addition to full-length product. As this was seen in the positive control, the '*n* - 1' product cannot indicate a deficiency in the interaction between the polymerase and the non-**ds^xTP**. Most likely, the '*n* - 1' product may have arisen because of a small amount of '*n* - 1' failure sequence in primer **Pr1** (Fig. 2, lane 1). Last, the full-length product incorporating **ds^xTP** migrated on the gel slightly more slowly than the analogous full-length product incorporating T. This is consistent, of course, with the modestly larger 'size' of an oligonucleotide containing **ds^xTP**.

In all subsequent experiments, this mobility shift was taken to indicate the presence of **ds¹** and **ds²** in the product strand. Additional confirmation that **ds¹TP** and **ds²TP** is indeed incorporated, and that the full-length product is not a result of misincorporation, was provided by the negative control. In the absence of **ds¹TP**, **ds²TP** and TTP, chain elongation stopped (<5% continued elongation) at the position of dA in the

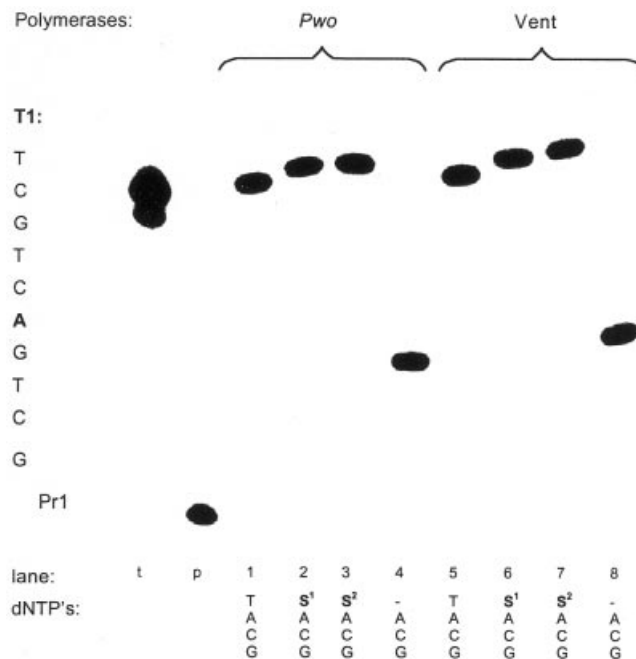


Figure 2. Primer extension reactions with 5'-³²P-labeled primer **Pr1** (p) and template **T1** (t), showing the synthesis of full-length product by these polymerases in the presence of **ds¹TP** and **ds²TP**, but not in their absence. Polymerases and triphosphates used in each reaction are noted below the lanes. Note the presence of *n* - 1 template (lane marked t). Denaturing 15% PAGE (urea, 60°C).

template (Fig. 2, lanes 4 and 8). This indicated that the polymerases could not form full-length product without 2'-deoxyuridine being present as a triphosphate. Together with the mobility shift, this makes compelling the case that full-length product, when it is observed, arises through incorporation of one of these nucleotides opposite dA in the template.

As chemical instability suggested that **ds¹** would never make a suitable analog, it was abandoned at this point. All eight polymerases (*Tfi*, *Tth*, *HotTub* and *Taq* from Family A and *Pwo*, *Pfu*, *Vent* and *Deep Vent* from Family B) were then challenged with **ds²**, **ds³** and **ds⁴**. To establish a metric for the evaluation of a polymerase-analog pair, the intensity of the band corresponding to full-length product was quantitated using the Molecular Imager™ software. The quantitative intensity was then compared to the intensity of the analogous band formed when TTP was replaced by **ds^xTP**. When an '*n* + 1' band was observed, as is frequently the case with Family A polymerases (28), this band was included as full-length product. The amount of full-length product formed by the positive control reaction was set at 100%.

Experiments then showed that **ds²TP**, **ds³TP** and **ds⁴TP** were all accepted in lieu of TTP by all eight polymerases tested (Fig. 3). Although most of the **ds³TP**/polymerase and **ds⁴TP**/polymerase combinations produced full-length product in amounts >75% of the positive control, larger amounts of abortion products were produced with **ds³TP** and **ds⁴TP** than with **ds²TP**. Surprisingly, the major abortion products formed by Family B polymerases were only 3 nt shorter than the full-length product and did not correspond to abortion at the position of the critical nucleotide dA, which would form

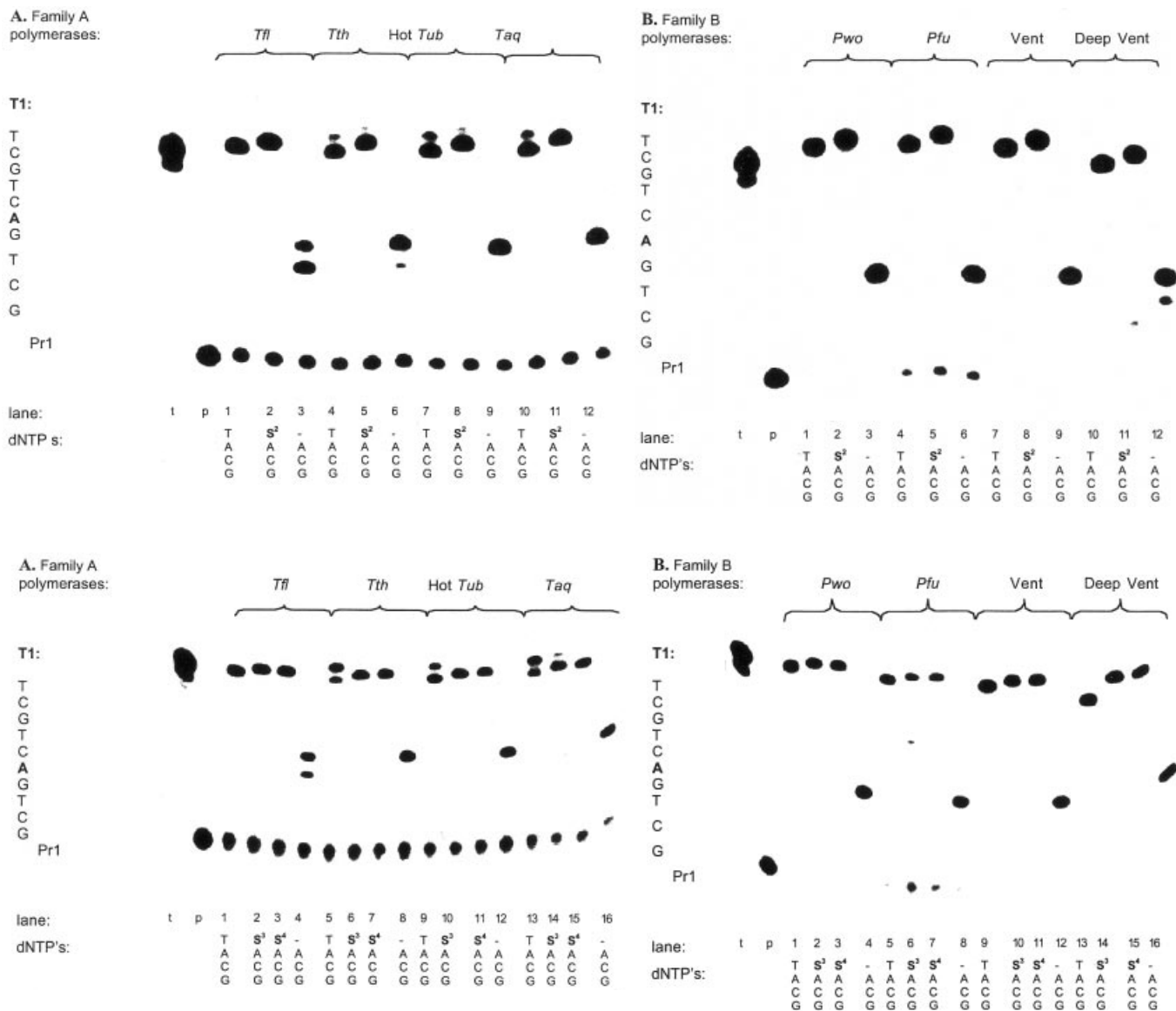


Figure 3. Primer extension reactions with 5'-³²P-labeled primer **Pr1** (p) and template **T1** (t), showing the synthesis of full-length product by these polymerases in the presence of TTP and **ds**²TP, **ds**³TP and **ds**⁴TP, but not in their absence. Polymerases and triphosphates used in each reaction are noted below the lanes. Note the presence of *n* - 1 template (lane marked t). Denaturing 17% PAGE (urea, 60°C).

products 5 nt shorter than the full-length product. It can be deduced from the shift of the bands of the abortion products that the functionalized nucleosides were probably incorporated before pausing occurred.

This set the stage for a broad survey of the suitability of polymerase-analog pairs that meet the level one challenge. It has been observed (our unpublished results) that the evolutionary family to which a polymerase belongs might be a predictor of its performance with a set of analogs. Therefore, we present data here according to their family (Figs 3 and 4).

It is worth noting that at this, the lowest level of challenge, all polymerases work with all substrates. This is the level of challenge that is met in a majority of experiments that examine the incorporation of non-standard nucleotide analogs into DNA via template-directed polymerase-catalyzed primer

extension. It is clear that performance at this level is, at best, a preliminary indicator of the suitability of the pair to support an artificial genetic system. We illustrated this by attempting to amplify an oligonucleotide using PCR with **ds**³TP or **ds**⁴TP replacing TTP and Vent as the polymerase. PCR produced little amplification under standard conditions (data not shown).

This result is hardly surprising. Copying the copies requires a polymerase that can, at the very least, accept a non-standard component in a template. Further, more than one non-standard component will occur in a real, artificial genetic system, possibly adjacent in the sequence. Polymerases must be able to copy these. In general, most polymerases will fail to PCR amplify oligonucleotides containing most nucleotide analogs.

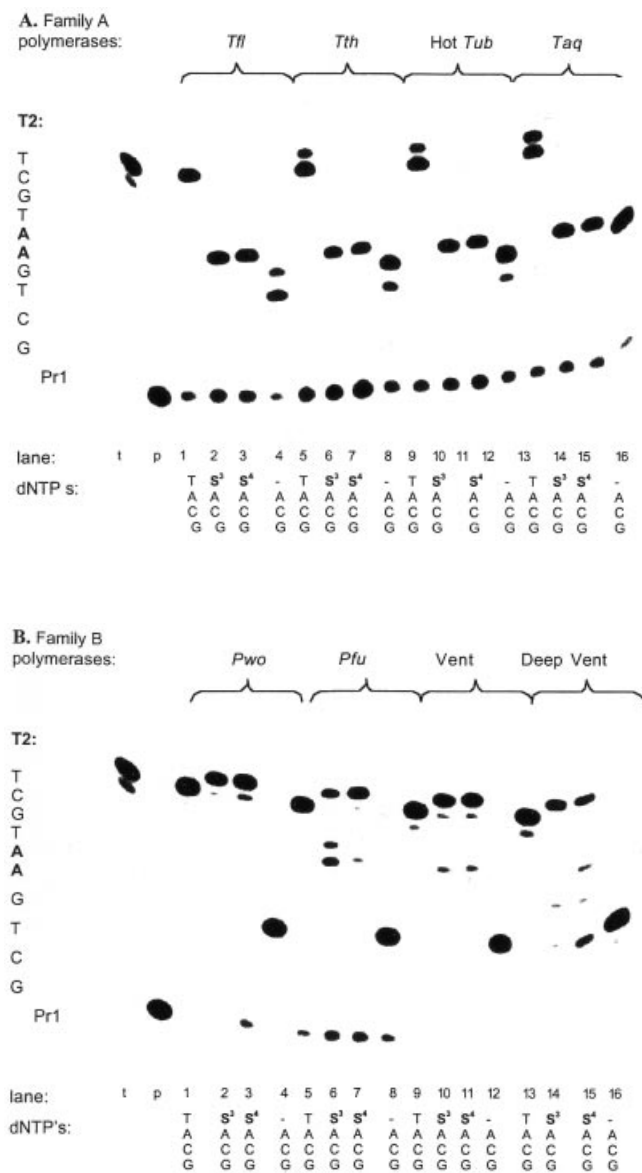


Figure 4. Primer extension reactions with 5'-³²P-labeled primer **Pr1** (p) and template **T2** (t). TTP, **ds³TP** and **ds⁴TP** were employed to replace TTP. Polymerases and triphosphates used in each reaction are noted. Denaturing 17% PAGE (urea, 60°C).

Unfortunately, a failed PCR does not provide sufficient information to say what went wrong. Therefore, intermediate level challenges are needed to evaluate and cull polymerase-analog systems for their suitability to support artificial genetics. At the very least, these will prevent time consuming experiments on polymerase-analog pairs that will not work. At best, these will provide diagnostic information to direct the experimentalist towards those systems that are most likely to work.

The next level of challenges: adjacent components and multiple components

Two of these situations were set up for the **ds^x** system. The first challenged the polymerase to incorporate two adjacent

analog components into a DNA product. This was achieved using template **T2**, which has two adjacent A residues. The second challenged the polymerase to incorporate multiple analog components at 'random' positions, both adjacent and non-adjacent, in a DNA product. This was achieved using template **T3**, which has eight A residues, including three in a row. The data are shown in Figure 4 (for consecutive incorporation) and Figure 5 (for multiple incorporation).

Several features of the results are striking. First, the ability to incorporate a single **ds^x** triphosphate was not predictive of the ability of a polymerase to incorporate two **ds²** nucleotides in adjacent positions. For example, all Family A polymerases incorporated a single **ds³** into a product (Fig. 3C), but failed to incorporate two consecutive **ds³** residues (Fig. 4).

Second, the data provided clear evidence that membership in a particular evolutionary family of polymerases was a predictor of the performance of a polymerase given a particular challenge. The only polymerase-analog systems that performed with results comparable to those obtained with TTP were samples containing **ds²TP** and Family B polymerases. A significant amount of a second band, probably the $n - 1$ product, was formed in the **ds²TP** reactions with template **T2**, but not in the corresponding reactions with TTP. Oddly, this phenomenon did not occur with template **T3**, which included the sequence of template **T2**. Family A polymerases generated significant amounts of full-length product containing **ds²**, but also clearly paused; it seemed as if the incorporation stopped after the **ds²TP**s had been incorporated (data not shown). It is interesting to note in this context that oligonucleotide **T4**, which contained two dAs separated by 4 nt, was utilized as a template without pausing by all eight polymerases in the presence of **ds²TP** as T derivative (see below).

Template **T3** was designed to identify polymerases that are capable of processing the functionalized triphosphates **ds²TP**, **ds³TP** and **ds⁴TP** with high efficiency. A total of eight modified triphosphates must be incorporated by a polymerase to form a full-length complement to **T3**. The decameric region adjacent to the primer-binding region of **T3** was identical to **T2**, which required incorporation of two consecutive modified triphosphates. In addition, **T3** posed the challenge of incorporating three successive modified triphosphates, as well as several individual modified triphosphates, in various sequence contexts.

Even with the demanding template **T3**, *Pwo*, *Pfu* and Deep Vent polymerases (Family B) generated substantial amounts of full-length product containing **ds²**, resulting in >75% of the product formed with TTP (Fig. 5B). The extent of the shift in the full-length product was remarkable. Of all tested polymerases, *Pwo* polymerase generated the cleanest product and the fewest failure products. Family A polymerases, in contrast, failed to yield any detectable full-length product in the presence of **ds²TP** as T analog (Fig. 5B).

With the alkanyl substituted nucleotides **ds³TP** and **ds⁴TP**, Family B polymerases were inefficient, while Family A polymerases completely failed to form fully elongated product with template **T2** (Fig. 4). Neither Family B nor Family A polymerases formed any full-length product with **ds³TP** and **ds⁴TP** and template **T3** under the assay conditions (data not shown).

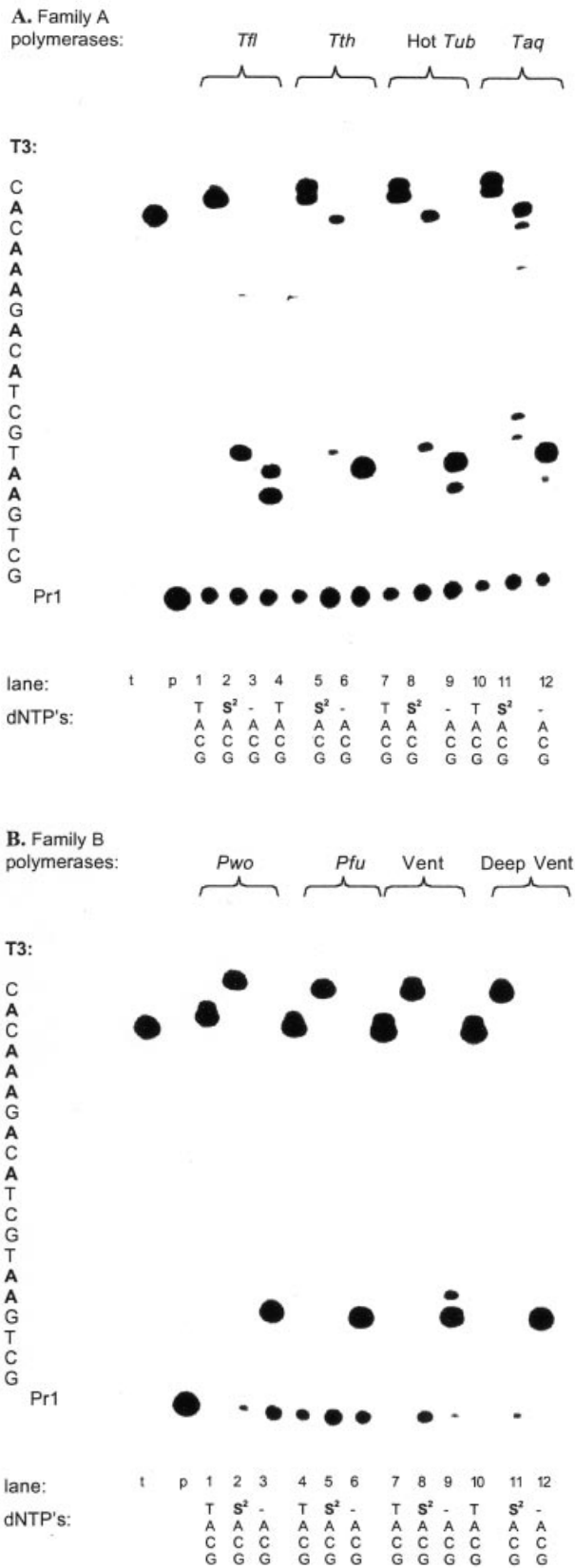


Figure 5. Primer extension reactions with 5'-³²P-labeled primer **Pr1** (p) and template **T3** (t). TTP and **dS²TP** were employed to replace TTP. Polymerases and triphosphates used in each reaction are noted. Denaturing 15% PAGE (urea, 60°C).

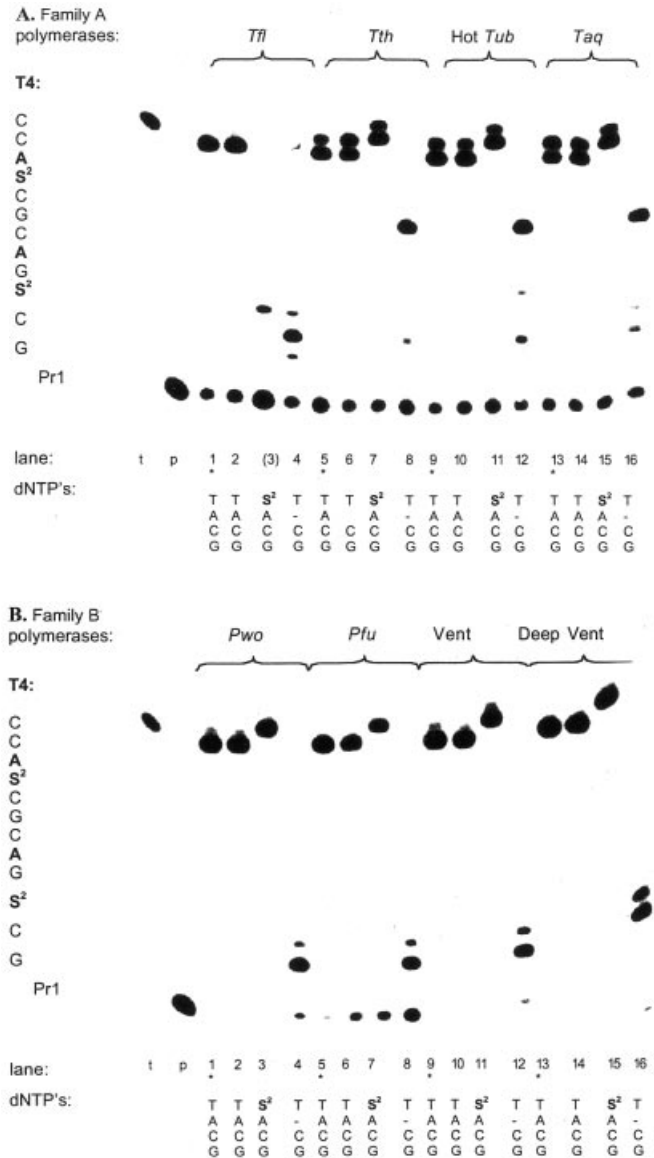


Figure 6. Primer extension reactions with 5'-³²P-labeled primer **Pr1** (p) and templates **T4** (t) and **T5** (marked with *). Polymerases and triphosphates used in each reaction are as noted. The result in lane 3 may be an artifact (see Fig. 7). Denaturing 15% PAGE (urea, 60°C).

Higher level challenges: non-standard components in the template

A second requirement for a non-standard analog to participate in an artificial genetics system is that it be accepted by a polymerase as a component of a template. Only the substituted thymidine derivative **dS²TP** survived intermediate challenges past simple primer extension reactions. The alkanyl substituted nucleotides **dS³TP** and **dS⁴TP** were not processed efficiently enough by the polymerases tested here. **dS¹** was eliminated due to its apparent chemical instability.

To implement this challenge for **dS²**, template **T4** was prepared containing **dS²**, and polymerases were challenged to incorporate dA opposite it (Fig. 6). Seven of the eight polymerases (the *Tfl* result is curious given the results in Fig. 7) readily incorporated dATP opposite **dS²** in template **T4** to

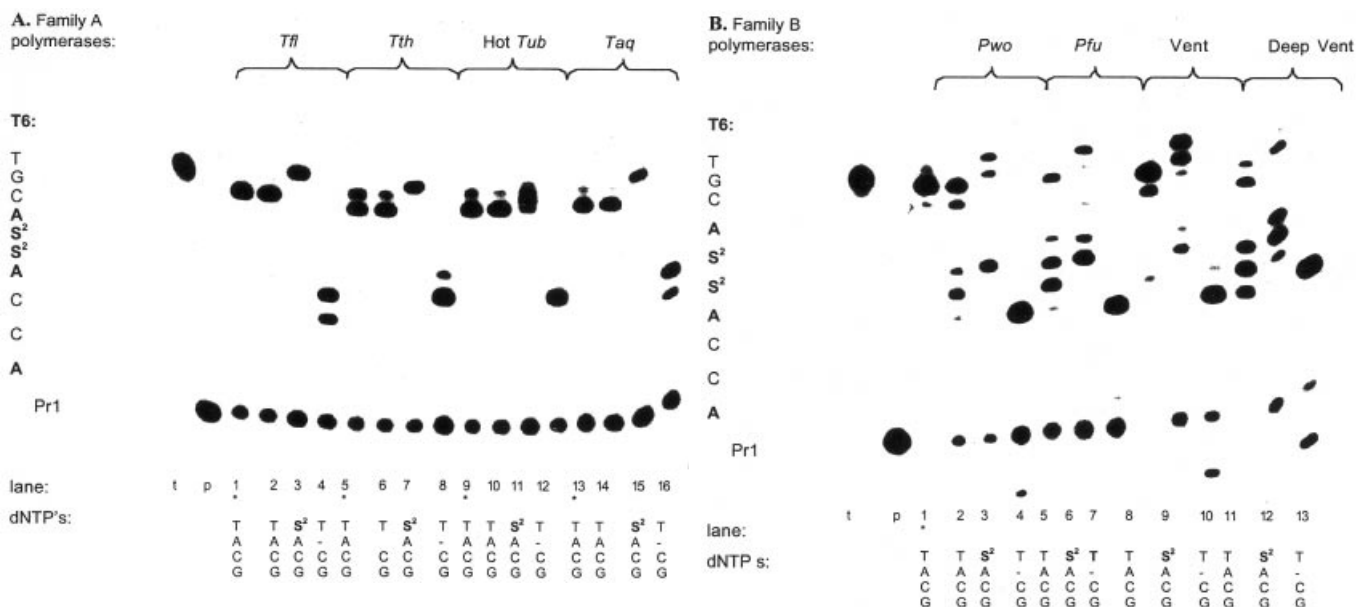


Figure 7. Primer extension reactions with 5'-³²P-labeled primer **Pr1** (p) and templates **T6** (t) and **T7** (marked with *). Polymerases and triphosphates used in each reaction are as noted. Denaturing 15% PAGE.

form full-length product. The longer $n + 1$ bands are characteristically seen with Family A polymerases (19). When TTP was employed as complement for dA in the template, the standard oligonucleotide product band migrated with the product band generated by the positive control sample. When **ds²TP** replaced TTP, full-length product containing two **ds²** nucleosides was generated and migrated slower than the product containing T.

In the absence of dATP, some full-length product (up to 10% of the amount formed in the presence of dATP) was observed with Family A polymerases. This might be due to contamination with residual [γ -³²P]ATP from the primer labeling procedure (requiring the DNA polymerase to incorporate a ribonucleoside triphosphate) or due to misincorporation. Misincorporation of dATP, dCTP or dGTP opposite dA was found to take place when the standard template **T1** was employed (Figs 3 and 4). Both observations are consistent with the known low fidelity of the relevant Family A polymerases (~ 1 error/ 1.3×10^5 nt incorporated) (29).

In the negative control samples with Family B polymerases, no full-length product was formed. With Vent (and slightly with Deep Vent), primer elongation continued beyond the first **ds²** nucleoside in the template by several nucleotides in the negative control (Fig. 7). This suggests that having **ds²** in the template may lower fidelity. This hypothesis needs to be examined in greater detail.

Further challenges: adjacent non-standard components in the template

The next challenge involved templates containing two adjacent non-standard components. For this purpose, template **T6** was used, carrying two adjacent **ds²** nucleosides, flanked by dA on each side. Template **T7**, a standard oligonucleotide of the same length and very similar sequence, was used for the positive control assays. All Family A polymerases formed full-length product in very high yields (Fig. 7A). However,

with all Family B polymerases, significant amounts of shorter products of various lengths were observed (Fig. 7B). *Pwo*, *Pfu* and Deep Vent even yielded <50% of the full-length product formed in the positive control sample. Nevertheless, very little misincorporation was seen in the negative control.

When **ds²TP** was offered with template **T6** instead of TTP, **ds²TPs** must be incorporated nearly across from the **ds²** nucleosides in the template, resulting in a duplex with four of the nucleotide analogs in close proximity. Only *Tfl* and *Tth* polymerases, both of Family A, were found to be able to generate this duplex with ease (Fig. 7A). *Pwo*, *Pfu* and Deep Vent polymerases paused significantly after encountering the two adjacent **ds²** nucleosides in the template (Fig. 7B). *Taq* and Deep Vent polymerases generated multiple products shorter than the full-length product, without a single predominant failure product.

Lengthening the time of incubation might resolve problems of pausing and slow incorporation, as long as there is little infidelity. Low infidelity in the negative control experiments was characteristic of Family B polymerases throughout these studies. Family A polymerases, in contrast, misincorporated other nucleotides when the complementary triphosphate was not present, especially with adjacent **ds^x** in the template (see Fig. 7). We propose that the ratio of incorporation to misincorporation in a 'minus' control may prove to be a more relevant predictor of the suitability of a polymerase-analog pair in an artificial genetics system. This hypothesis will need further examination.

Evaluation using competition experiments

To directly compare the efficiency of polymerase incorporation of **ds²TP** versus TTP, primer extension experiments were implemented in which **ds²TP** competes with TTP. Oligonucleotide **T1** was chosen as template, as it requires the incorporation of only a single dA complement and therefore enables a simple analysis of the experimental data.

Table 2. Results of primer extension experiments with direct competition between labeled TTP and unlabeled **ds**²TP for incorporation opposite dA by *Pwo* and Vent polymerase

	Radioactivity expected with no discrimination	<i>Pwo</i> polymerase		Vent polymerase	
		Radioactivity observed	Observed/expected	Radioactivity observed	Observed/expected
100% TTP*	100	100	100	100	100
75% TTP* + 25% ds ² TP	75	62	82	69	92
50% TTP* + 50% ds ² TP	50	35	70	46	92
25% TTP* + 75% ds ² TP	25	15	62	23	91

ds²TP and an [α -³²P]TTP/TTP mixture (TTP*) were offered to the polymerases in different ratios.

The assays were carried out under the conditions used in the polymerase screening experiments. The complex from primer **Pr1** (5'-fluorescein labeled) and template **T1** was incubated for 10 min at 72°C with either *Pwo* or Vent polymerase, the standard triphosphates, and [α -³²P]TTP (TTP*). The amount of radioactive product formed in the control reaction was set to 100%. In individual assays, 25, 50 and 75% of the TTP* solution in the control reaction was replaced by an equimolar **ds**²TP solution, and the decrease in the amount of radioactivity incorporated into full-length product was used as a measure of the ability of **ds**²TP to compete with TTP* in the elongation reaction. The product was visualized using the 5'-fluorescein label of the primer, ensuring that complete primer elongation had taken place in all samples.

If the polymerase made no discrimination between TTP* and **ds**²TP, T* and **ds**² would appear in the full-length product in a ratio identical to the ratio of their triphosphates. Thus, if TTP* and **ds**²TP were offered in a 1:1 ratio and the polymerase accepted either substrate equally well, the amount of radioactive product would be 50% of the amount of radioactive product formed in the presence of TTP* alone.

The amounts of radioactivity actually found in the product bands, relative to the samples without addition of **ds**²TP, are summarized in Table 2. Remarkably, for the two polymerases examined, the amounts of radioactivity found in the product bands indicate that both *Pwo* and Vent polymerase incorporate **ds**²TP preferentially to TTP opposite dA in template **T1**. *Pwo* polymerase discriminates even more strongly in favor of **ds**²TP than Vent.

We asked whether this result might be explained by a miscalculation of the concentration of the **ds**²TP solution, which was estimated by UV absorbance exploiting the Lambert-Beer Law. The extinction coefficient for **ds**²TP was taken as that of 5-(1-propynyl)-2'-deoxyuridine (Glen Research; molar $\epsilon_{260} = 3.2 \times 10^3$). Comparison of extinction coefficients of various alkynyl substituted deoxyuridine derivatives does not show a strong influence of the ω functionality of the alkynyl linker on the extinction coefficient (30). For example, ϵ_{260} is reported as $3.5 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ for linker $-\text{C}\equiv\text{C}-\text{R}$, R = $\text{CH}_2\text{CH}_2\text{OH}$, while ϵ_{260} is reported as $3.0 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ for linker $-\text{C}\equiv\text{C}-\text{R}$, R = CH_2CH_3 . This implies that the discrimination evident with *Pwo* was significantly greater than any plausible misjudgement of the concentration of **ds**²TP. The discrimination observed with Vent, however, was within these limits.

Tables 3 and 4 summarize the results of the primer extension experiments examining the enzymatic incorporation

of **ds**²TP, **ds**³TP and **ds**⁴TP opposite dA nucleosides. In general, the collection of data focuses on Family B, and *Pwo* polymerase in particular, as a suitable pair for **ds**²TP. These results provide the rationale for this pair as a candidate to support an artificial genetic system. It should be noted that a further understanding of the competition between these analogs could be gained by comparing the time courses of the primer extension reaction.

PCR amplification

Amplifications by PCR are much more exacting than a simple primer extension reaction, with very different demands on polymerases. In PCR, longer templates must be replicated in shorter times than in typical primer extension experiments. PCR amplification also puts higher demands on the fidelity of polymerases, because incorporation mistakes are amplified.

Pwo, *Pfu*, Vent, Deep Vent, *Tfl* and *Taq* polymerases were tested for their ability to PCR amplify the 94mer template sequence **T8**, utilizing **ds**²TP in place of TTP. The 94mer **T8** was used as a template in comparative studies of the polymerases. **T8** contained only the four standard nucleotides. **ds**² was introduced during the first cycle of PCR. In the following cycles, **ds**² was present as both triphosphate and in the template. **T8** was designed to contain only very few **ds**² nucleotides in the primer-binding regions generated after the first PCR cycle, in order to avoid potential annealing problems. The region between primers and primer-binding regions had a very high **ds**² content (30% after the first cycle), to make the test rigorous.

A set of three samples was prepared to test each polymerase. All three samples contained template **T8** and matching primers (**Pr2** and **Pr3**), the polymerase being tested and the standard triphosphates dATP, dCTP and dGTP. TTP was added to the 'positive' control sample (T sample). The TTP solution was substituted by **ds**²TP solution in a second sample (S sample). Finally, water replaced the T derivative solution in the 'negative' control sample.

In amplifications with **ds**²TP and *Pwo*, ~90% amplification product containing **ds**² (**T8_S**) was generated compared to the amount of amplification product containing T (**T8_T**) with TTP (Fig. 8). The amount of **T8_S** generated by Vent polymerase was ~55% of the amount of **T8_T** generated under the same conditions. For *Pfu* and *Tfl* polymerase the ratio **T8_S:T8_T** was even less, only ~30–35%. *Taq* did not form any detectable **T8_S** product at all.

With *Pwo* and *Pfu* polymerase PCR was performed over 25 cycles. Both the T and S products were obtained as sharp

Table 3. Summary of the influence of linker geometry and length on the incorporation of the functionalized nucleoside triphosphates **dS²TP**, **dS³TP** and **dS⁴TP** into DNA

Polymerase Polymerase family	<i>Tfl</i> A	<i>Tth</i> A	<i>HotTub</i> A	<i>Taq</i> A	<i>Pwo</i> B	<i>Pfu</i> B	Vent B	Deep Vent B
Template T1 (.....A....., single incorporation)								
dS²TP	+++	+++	+++	+++	+++	+++	+++	+++
dS³TP	+++	+++	+++	+++	+++	++	+++	+++
dS⁴TP	+++	+++	++	+++	+++	++	+++	+++
Template T2 (.....AA....., double adjacent incorporation)								
dS²TP	+	++	+++	++	+++	+++	+++	+++
dS³TP	-	-	-	-	++	+	++	+
dS⁴TP	-	-	-	-	++	++	++	+
Template T3 (...AA...A.A...AAA..., multiple incorporation)								
dS²TP	-	-	-	-	+++	+++	++	+++
dS³TP	-	-	-	-	-	-	-	-
dS⁴TP	-	-	-	-	-	-	-	-

The symbols estimate the amount of full-length product generated by a polymerase with **dS^xTP** relative to the amount of full-length product generated utilizing TTP in place of **dS^xTP**. Quantitation by PhosphorImager. +++, >75%; ++, 50–75%; +, 25–50%; -, <25%.

Table 4. Acceptance of **dS²**-containing templates by thermostable polymerases

Polymerase Polymerase family	<i>Tfl</i> A	<i>Tth</i> A	<i>HotTub</i> A	<i>Taq</i> A	<i>Pwo</i> B	<i>Pfu</i> B	Vent B	Deep Vent B
Template T4 (...AS ² ...A.S ² ..)								
dS²TP	+++	+++	+++	+++	+++	+++	++	+++
TTP	-	+++	+++	+++	+++	++	+++	+++
Template T6 (.....AS ² S ² A.....)								
TTP	+++	+++	+++	+++	+++	++	+	+++
dS²	+++	++	-	+	+	+	++	-

The symbols estimate the amount of full-length product generated by a polymerase relative to the amount of full-length product generated utilizing TTP and a template containing T in place of **dS²**. Quantitation by PhosphorImager. +++, >75%; ++, 50–75%; +, 25–50%; -, <25%.

bands. With Vent, *Tfl* and *Taq* amplification was stopped after 15 cycles. When more cycles were performed, non-specific amplification of the template occurred in the T samples, leading to longer products which were visible as a smear on the gel.

DISCUSSION

In the past decade, many laboratories have reported that polymerases incorporate nucleotide analogs, including those that carry appendages that protrude into the major groove. These have proven to be useful tools for the biotinylation of DNA, for DNA sequencing and for the addition of a variety of reporter groups and tags to DNA.

These results might suggest that incorporation of major groove functionalized nucleotides into DNA via template-directed polymerization must be easy. In fact, when this is done routinely within the context of a PCR experiment, the outcome is frequently unsatisfactory. This paper makes clear that this unsatisfactory outcome is not necessarily unexpected. In virtually all of these cases, polymerase incorporation was considered to be successful after only the simplest challenge is presented to the polymerase.

This does not, of course, detract from the utility of the tools. In DNA sequencing, of course, inefficient incorporation of a chain terminator is desired. Inefficient incorporation of a label can be tolerated in a range of biophysics experiments, where (for example) a fluorescence signal is sought and the signal from only a few labeled molecules is needed.

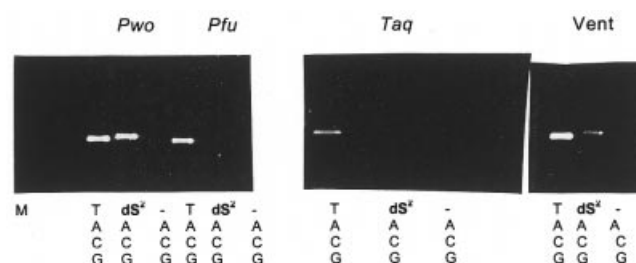


Figure 8. Letters beneath the lanes indicate the triphosphates that were present in the PCR incubation mixture. PCR products were resolved by 2% agarose gel electrophoresis, stained with ethidium bromide and visualized under UV light (254 nm). The marker, labeled M, is a 25 nt ladder from Life Sciences; the most intense band is at 125 bases. PCR cycling: 94°C, 1 min; 56°C, 2 min; 72°C, 2 min. All amplifications were done for 25 cycles, except for *Taq*, where 15 cycles were used to avoid 'smears' (presumably a broad product distribution), presumably arising from lower proof-reading. The intensity of fluorescence was integrated by the Molecular Imager software.

Only when the goal is an artificial genetic system must we be concerned about the highest level of incorporation. There is nothing particularly innovative about the strategy presented here to assess the specifications of a polymerase–analog pair. What is clear, however, is that a strategy of this type is important in developing artificial genetic systems based on them.

Many discoveries were made by this strategy. First, we were surprised to see the chemical instability of **dS¹**. This discovery should guide further development of artificial genetic systems.

Second, it was a surprise that two polymerases incorporated **ds²TP**, in some contexts, in preference to TTP. In the *Pwo/Pfu* sequence(s), however, it is interesting to note that an Ala residue replaces Ser411, which is highly conserved throughout most other Family B polymerases. Although no crystal structure is available for the *Pwo/Pfu* polymerase, crystal structures of other Family B polymerases allow homology modeling, which in turn suggests that the Ala411 side chain might lie in the position where a major groove substituent might bind. The Ala side chain is, of course, more hydrophobic than the Ser side chain. This difference might therefore explain why the *Pwo* polymerase so effectively accepts **ds²** in both the template and as a triphosphate. Realistically, however, chemical theory is inadequate (31) to enable rational design at this level of resolution, implying that explanations at this level of resolution must be viewed as being conjectural.

It was also surprising to see that single incorporation experiments were not necessarily good predictors of the outcome of multiple incorporation experiments. Much of the literature needs to be re-evaluated in the light of this result.

Further, we found further evidence that evolutionary classification is a valuable predictor for how a polymerase is likely to handle a non-standard substrate. Polymerases are classified into two evolutionary families according to their amino acid sequence similarity to *E.coli* DNA polymerase I (32). X-ray crystallography has shown that Family A and Family B polymerases possess analogous tertiary structures (20–22). They are, however, quite divergent in their amino acid sequences and it remains an open question as to whether the two folds arose by convergent or divergent evolution.

This statement may be general to other classes of nucleotide analogs. For example, it has been suggested that the ability of a polymerase to incorporate the nucleoside analogs *disoC* and *disoG* also correlates with membership of a polymerase family (33).

In this context, it is interesting to note that the Family B polymerases examined here all have a strong 3'→5' exonuclease (or 'proofreading') activity, providing replication with higher accuracy than the relatively error-prone Family A representatives, which all lack a 3'→5' exonuclease domain. The inherent proofreading activity of *Pwo* polymerase results in an increased fidelity of DNA replication that is (remarkably) >10-fold greater than that displayed by *Taq* polymerase (Roche Molecular Systems, *Pwo* polymerase product information). It is interesting to ask how this well known property of polymerases might correlate with their ability to accept with fidelity non-standard nucleotide analogs.

Finally, the *Pwo* and *Pfu* polymerases present a paradox. These enzymes were purchased from Roche Molecular Systems and Stratagene, respectively. The literature records their having the same sequences, despite their different names. Nevertheless, they behave quite differently in their interactions with non-standard nucleotide analogs, and these differences are reproducible. Inquiries to the commercial supply houses failed to obtain any information that resolved this paradox. We cannot say therefore whether the literature has confused the sequences of the genes, or whether the proteins are expressed or processed differentially from the same gene, or whether trace molecule 'effectors' in the two commercial preparations account for their different behaviors.

ACKNOWLEDGEMENTS

We are indebted to Dr Ajit P. Kamath, Dr Thomas R. Battersby and Prof. C. Ronald Geyer for helpful discussions. Scott McMillen, from the Protein Chemistry Core at the University of Florida, obtained the mass spectra of oligonucleotides. Dr Jodie V. Johnson from the Department of Chemistry obtained the mass spectra of the triphosphates. This work was partially supported by grants from the NIH (GM 54048), the NASA Astrobiology Institute and the Office of Naval Research (N00025-96-1-0362).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

REFERENCES

- Langer, P.R., Waldrop, A.A. and Ward, D.C. (1981) Enzymatic-synthesis of biotin-labeled polynucleotides. Novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. USA*, **78**, 6633–6637.
- Bergstrom, D.E. and Ruth, J.L. (1976) Synthesis of C-5 substituted pyrimidine nucleosides via organo-palladium intermediates. *J. Am. Chem. Soc.*, **98**, 1587–1589.
- Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A. and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science*, **238**, 336–341.
- Benner, S.A., Alleman, R.K., Ellington, A.D., Ge, L., Glasfeld, A., Leanz, G.F., Krauch, T., MacPherson, L.J., Moroney, S.E., Piccirilli, A.J. and Weinhold, E. (1987) Natural selection, protein engineering and the last riboorganism: rational model building in biochemistry. *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 53–63.
- Switzer, C.Y., Moroney, S.E. and Benner, S.A. (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.*, **111**, 8322–8323.
- Piccirilli, J.A., Krauch, T., Moroney, S.E. and Benner, S.A. (1990) Extending the genetic alphabet. Enzymatic incorporation of a new base pair into DNA and RNA. *Nature*, **343**, 33–37.
- Latham, J.A., Johnson, R. and Toole, J.J. (1994) The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. *Nucleic Acids Res.*, **22**, 2817–2822.
- Wiegand, T.W., Janssen, R.C. and Eaton, B.E. (1997) Selection of RNA amide synthases. *Chem. Biol.*, **4**, 675–683.
- Tarasow, T.M., Tarasow, S.L. and Eaton, B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54–57.
- Perrin, D.M., Garestier, T. and Hélène, C. (1999) Expanding the catalytic repertoire of nucleic acid catalysts: simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. *Nucleosides Nucleotides*, **18**, 377–391.
- Battersby, T.R., Ang, D.N., Burgstaller, P., Jurczyk, S.C., Bowser, M.T., Buchanan, D.D., Kennedy, R.T. and Benner, S.A. (1999) Quantitative analysis of receptors for adenosine nucleotides obtained via *in vitro* selection from a library incorporating a cationic nucleotide analog. *J. Am. Chem. Soc.*, **121**, 9781–9789.
- Teramoto, N., Imanishi, Y. and Ito, Y. (2000) *In vitro* selection of a ligase ribozyme carrying alkylamino groups on the side chains. *Bioconjugate Chem.*, **11**, 744–748.
- Sakthivel, K. and Barbas, C.F. (1998) Expanding the potential of DNA for binding and catalysis: highly functionalized dUTP derivatives that are substrates for thermostable DNA polymerases. *Angew. Chem. Int. Ed.*, **37**, 2872–2875.
- Dewey, T.M., Zyzniewski, M.C. and Eaton, B.E. (1996) RNA world: functional diversity in a nucleoside by carboxyamidation of uridine. *Nucleosides Nucleotides*, **15**, 1611–1617.
- Vaish, N.K., Fraley, A.W., Szostak, J.W. and McLaughlin, L.W. (2000) Expanding the structural and functional diversity of RNA. Analog uridine triphosphates as candidates for *in vitro* selection of nucleic acids. *Nucleic Acids Res.*, **28**, 3316–3322.

16. Matulic-Adamic,J., Daniher,A.T., Karpeisky,A., Haeberli,P., Sweedler,D. and Beigelman,L. (2000) Nucleoside triphosphates for *in vitro* selection of new catalytic RNAs. *Bioorg. Med. Chem. Lett.*, **10**, 1299–1302.
17. Ang,D.N. (1999) An alternative to the origins of life theories: amino acid-like DNA molecules capable of improved catalysis, PhD Thesis, University of Florida, Gainesville, FL.
18. Horlacher,J., Hottiger,M., Podust,V.N., Huebscher,U. and Benner,S.A. (1995) Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with non-standard hydrogen bonding patterns. *Proc. Natl Acad. Sci. USA*, **92**, 6329–6333.
19. Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) An attempt to unify the structure of polymerases. *Protein Eng.*, **3**, 461–467.
20. Franklin,M.C., Wang,J.M. and Steitz,T.A. (2001) Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell*, **105**, 657–667.
21. Rodriguez,A.C., Park,H.W., Mao,C. and Beese,L.S. (2000) Crystal structure of a pol alpha family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp.9 degrees N-7. *J. Mol. Biol.*, **299**, 447–462.
22. Kiefer,J.R., Mao,C., Braman,J.C. and Beese,L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304–307.
23. Lee,S.E., Sidorov,A., Gourelain,T., Mignet,N., Thorpe,S.J., Brazier,J.A., Dickman,M.J., Hornby,D.P., Grasby,J.A. and Williams,D.M. (2001) Enhancing the catalytic repertoire of nucleic acids: a systematic study of linker length and rigidity. *Nucleic Acids Res.*, **29**, 1565–1573.
24. Gourelain,T., Sidorov,A., Mignet,N., Thorpe,S.J., Lee,S.E., Grasby,J.A. and Williams,D.M. (2001) Enhancing the catalytic repertoire of nucleic acids. II. Simultaneous incorporation of amino and imidazolyl functionalities by two modified triphosphates during PCR. *Nucleic Acids Res.*, **29**, 1898–1905.
25. Osborne,S.E., Cain,R.J. and Glick,G.D. (1997) Structure and dynamics of disulfide cross-linked DNA triple helices. *J. Am. Chem. Soc.*, **119**, 1171–1182.
26. Ludwig,J. and Eckstein,F. (1989) Rapid and efficient synthesis of nucleoside 5'-O-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, **54**, 631–635.
27. Goodwin,J.T. and Glick,G.D. (1993) Incorporation of alkylthiol chains at C-5 of deoxyuridine. *Tetrahedron Lett.*, **34**, 5549–5552.
28. Clark,J.M. (1988) Novel non-template nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677–9686.
29. Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) PCR fidelity of *Pfu* DNA polymerase and other thermostable polymerases. *Nucleic Acids Res.*, **24**, 3546–3551.
30. Robins,M.J. and Barr,P.J. (1983) Nucleic acid related compounds. 39. Efficient conversion of 5-iodo to 5-alkynyl and derived 5-substituted uracil bases and nucleosides. *J. Org. Chem.*, **48**, 1854–1862.
31. Zepik,H. and Benner,S.A. (1999) Catalysts, anticatalysts and receptors for unactivated phosphate diesters in water. *J. Org. Chem.*, **64**, 8080–8083.
32. Braithwaite,D.K. and Ito,J. (1993) Compilation, alignment and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.*, **21**, 787–802.
33. Kodra,J.T. (1998) Chemistry and enzymology of an expanded genetic alphabet, dissertation no. 12916. Swiss Federal Institute of Technology, Zurich, Switzerland.
34. Ludwig,J. and Eckstein,F. (1989) Rapid and efficient synthesis of nucleoside 5'-O-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, **54**, 631–635.