Published in final edited form as:

Chemistry. 2012 January 23; 18(4): 1231–1239. doi:10.1002/chem.201102066.

Major Groove Substituents and Polymerase Recognition of a Class of Predominantly Hydrophobic Unnatural Base Pairs

Dr. Thomas Lavergne, Mr. Denis A. Malyshev, and Prof. Dr. Floyd E. Romesberg^a

Floyd E. Romesberg: floyd@scripps.edu

^aDepartment of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037 U.S.A., Fax: (+1) 858 784 7472

Abstract

Expansion of the genetic alphabet with an unnatural base pair is a long-standing goal of synthetic biology. We have developed a class of unnatural base pairs, formed between **d5SICS** and analogues of **dMMO2**, that are efficiently and selectively replicated by the Klenow Fragment (Kf) DNA polymerase. In an effort to further characterize and optimize replication, we report the synthesis of five new **dMMO2** analogues bearing different substituents designed to be oriented into the developing major groove and an analysis of their insertion opposite **d5SICS** by Kf and *Thermus aquaticus* DNA polymerase I (*Taq*). We also expand the analysis of the previously optimized pair, **dNaM-d5SICS**, to include replication by *Taq*. Finally, the efficiency and fidelity of PCR amplification of the base pairs by *Taq* or Deep Vent polymerases is examined. The resulting structure-activity relationship data suggest that the major determinants of efficient replication are the minimization of desolvation effects and the introduction of favorable hydrophobic packing, and that *Taq* is more sensitive than Kf to structural changes. In addition, we identify an analogue, **dNMO1**, that is a better partner for **d5SICS** than any of the previously identified **dMMO2** analogues with the exception of **dNaM**. We also find that **dNaM-d5SICS** is replicated by both Kf and *Taq* with rates approaching those of a natural base pair.

Keywords

DNA; Klenow fragment; *Taq*; PCR; Genetic Alphabet

Introduction

The four letter genetic alphabet is conserved throughout nature and is based on the complementary shape and hydrogen-bonding (H-bonding) of the natural purines and pyrimidines. For over two decades, efforts have been focused on the development of unnatural base pairs where pairing is mediated by orthogonal H-bonding patterns, and progress along this route continues.^[1-4] However, we^[5-16] and others,^[17-21] have demonstrated that hydrophobic and packing forces are also sufficient to underlie the efficient and selective replication of an unnatural base pair. If sufficiently well replicated, such unnatural base pairs could be used as part of an expanded genetic system and would allow for an increase in the information potential of a genome, but will likely be useful immediately for *in vitro* applications, such as the site-specific labeling of enzymatically synthesized DNA or RNA with novel functionality (*i.e.* fluorophores or reactive moieties) for *SELEX* (Systematic Evolution of Ligands by Exponential Enrichment)^[22] or nanomaterial applications.^[23]

Two of the most promising unnatural base pairs that we have identified are those formed by d5SICS and either dMMO2 (dMMO2-d5SICS) or dNaM (dNaM-d5SICS) (Figure 1A and B).^[5,6,9,13,14] Previous studies with the Klenow fragment of *E. coli* DNA polymerase I (Kf) and T7 RNA polymerase have demonstrated that d5SICS-dNaM is replicated and transcribed with greater efficiency than dMMO2-d5SICS.^[14] However, the dMMO2 scaffold is simpler and more atom-economical than dNaM. In addition, the dMMO2 scaffold provides more options for linker attachment, which is required for site-specific labeling of the DNA, including attachment at the site analogous to that widely used with the natural pyrimidines^[24] and not available with dNaM. Thus, our current efforts to expand the genetic alphabet include the continued evaluation of dNaM-d5SICS, with a particular focus on recognition by other polymerases that enable specific applications such as PCR, as well as to continue to generate structure-activity relationship (SAR) data that will facilitate the further optimization of dMMO2-d5SICS.

A limiting step of replicating DNA containing dMMO2-d5SICS is the incorporation of dMMO2TP opposite d5SICS. Because existing SAR data suggests that this step is more sensitive to modifications of the triphosphate than of the templating nucleobase, we have focused on derivatizing dMMO2TP. The SAR data also clearly reveal that the *ortho* methoxy group is absolutely required for replication (the methyl group facilitates triphosphate insertion and the oxygen atom facilitates continued primer extension once the unnatural nucleotide is incorporated at the terminus^[6]). In addition, the accumulated SAR data also suggest that *para* derivatization is more promising than *meta* derivatization.^[13,15] Initial efforts to identify beneficial modifications at the *para* position yielded dDMO (Figure 1B). While the dDMO-d5SICS unnatural base pair is replicated more efficiently than the parental dMMO2-d5SICS pair, it is still replicated less efficiently than dNaM-d5SICS.^[8]

To continue our efforts to optimize the dMMO2 scaffold, we now report the synthesis and characterization of five new derivatives bearing different moieties at the *para* position (Figure 1C). The specific modifications to the dMMO2 scaffold were selected to systematically vary the size, shape, electronic properties, and H-bonding potential of the nucleobase. The unnatural triphosphates were analyzed by determining the steady-state efficiencies with which they are inserted opposite d5SICS in a DNA template by Kf or by *Thermus aquaticus* DNA polymerase I (*Taq*). To further explore dNaM-d5SICS, we also characterized its synthesis in both strand contexts (*i.e.* insertion of dNaMTP opposite d5SICS and d5SICSTP opposite dNaM) by *Taq*. To facilitate the detailed comparison of all of the analogues in side-by-side experiments, the previously reported efficiencies of dNaM-d5SICS synthesis by Kf were re-determined, as were the insertion efficiencies of dMMO2TP and dDMOTP opposite d5SICS by Kf and *Taq*. Finally, each triphosphate was also analyzed by characterizing the PCR amplification of the corresponding unnatural base pair with d5SICS by *Taq* or Deep Vent polymerases. The results reveal important, and in some cases, polymerase-specific SAR data, and we find that dNMO1TP is better optimized than the other new analogues as a partner for d5SICS. The current analysis also reveals that dNaM-d5SICS is synthesized by Kf better than previously appreciated, and synthesized equally well by *Taq*, solidifying its position as the most promising unnatural base pair identified to date.

Results

Nucleotide design, synthesis, and evaluation

The unnatural nucleotides dAMO1, dAMO2, dAMO3, dPMO1, and dNMO1 (Figure 1C) were designed to position different substituents in the developing major groove during replication. Each substituent is attached via a single bond to the position *para* to the

glycosidic bond because SAR data suggest that rotational flexibility is important for efficient triphosphate incorporation^[8] (although dNaM is an interesting and incompletely understood exception), presumably by allowing for the optimization of the developing interactions with d5SICS and/or the DNA polymerase. The most simple derivative of the series, dAMO1, bears only an amino substituent at the *para* position. In dAMO2 and dAMO3, the amino moiety is modified with acetyl and trifluoroacetyl groups, which increase the size and reduce the ability of the substituent to donate electron density into the aromatic ring of the nucleobase scaffold. The nitro group of dNMO1 is a potent electron withdrawing substituent. In contrast, incorporating the nitrogen atom within the context of the pyrrolo ring of dPMO1 should have more modest electronic effects, but significantly increase the aromatic surface area of the scaffold.

The unnatural nucleotide derivatives were synthesized as shown in Scheme 1. Briefly, the modified nucleoside **3** was obtained in three steps via Heck coupling between the 2'-deoxyribose glycal **1** and the appropriately iodinated N-Cbz-protected anisidine **2**, followed by deprotection of the sugar moiety and selective reduction of the resulting 3' keto group. Acceptable yields of the Heck coupling product required the use of an electron withdrawing group to protect the aromatic amine. The major coupling product was the desired β -anomer, confirmed by NOE experiments, which was separated from the minor α -anomer by silica gel column chromatography. The hydroxyl groups were simultaneously protected with tetraisopropylidisiloxane groups, and the Cbz group was removed by hydrogenation. Compound **4** was used as a common precursor to introduce diversity and access each of the desired nucleotides. We found that even under mildly acidic conditions, compound **4** is prone to epimerize to a mixture of α - and β -anomers, requiring careful control of all reaction conditions.

Toward dNMO1, the aromatic amine of **4** was first oxidized using a potassium iodide-tert-butyl hydroperoxide catalyst, which cleanly converted the amine to the nitro group, while approaches based on *in situ* generation of the dimethyldioxirane or the use of methyltrioxorhenium/O₂ resulted in incomplete oxidation to the nitroso and hydroxylamine compounds, as well as the production of azoxy self coupled products. dNMO1(**5a**) was finally obtained by removal of the silyl protecting groups. Protected dPMO1 was obtained via acid-free condensation^[25] in aqueous 2,5-dimethoxytetrahydrofuran with microwave irradiation at 140 °C, which afforded the pure β -anomer, and the free nucleoside **5b** was obtained by removal of the silyl protecting group. Compounds **5c** and **5d** were obtained by acylation with acetic anhydride or trifluoroacetic anhydride respectively, followed by TBAF-mediated sugar deprotection.

Free nucleosides **5a - d** were converted to the corresponding triphosphates **6a - d** using Ludwig conditions^[26] and purified by anion exchange chromatography, followed by HPLC. dAMO1TP (**6e**) was obtained from dAMO3TP (**6d**) by ammonia-mediated deprotection of the amino group at room temperature. The purity of each triphosphate was confirmed by ³¹P NMR, HPLC, and MALDI-TOF (Supporting Information). The triphosphates of dMMO2, dDMO, dNaM and d5SICS were prepared as described previously,^[6,8,13] and the phosphoramidites of dNaM and d5SICS were prepared and incorporated in to DNA as described previously.^[13]

Each triphosphate was analyzed by examining its insertion opposite a correct or incorrect nucleotide in a DNA template by Kf and *Taq* under steady-state conditions, which provides a convenient assay to measure the overall rate at which product is formed.^[27] The most interpretable data is the second order rate constant (or efficiency, k_{cat}/K_M) relating the duplex-bound polymerase and free triphosphate to the rate limiting transition state for the multiple turnover reaction. Assays using Kf were performed at 25 °C, which is close to the

enzyme's optimal temperature of 37 °C. While *Taq* is a thermophilic polymerase, assays using *Taq* were performed at 50 °C to allow for the use of the same primer-template substrates used with Kf and in previous studies. Although it is approximately 20 °C below optimal, *Taq* does retain a significant level of activity at this temperature.^[28] Each analogue was also examined as a partner for d5SICS by PCR.

Efficiency of insertion of dMMO2TP, dDMOTP, and each natural dNTP opposite d5SICS

To gauge the efficiency of unnatural base pair synthesis, we first analyzed the insertion of dATP opposite dT in sequence context I (Table 1). In good agreement with previous data, we found that dATP was inserted opposite dT in the template with a k_{cat}/K_M of $7.7 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$. We then examined the insertion of dMMO2TP and dDMOTP opposite d5SICS under identical conditions. We found that dMMO2TP and dDMOTP are inserted with efficiencies of 3.8×10^5 and $2.2 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$, respectively. To determine fidelity, we measured the rate of insertion of each natural triphosphate opposite d5SICS in the same sequence context. We found that dATP, dGTP, and dTTP are inserted with efficiencies between 9×10^3 and $1.5 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$, while dCTP is not inserted at a detectable level ($k_{\text{cat}}/K_M < 1 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$). All of the data are in good agreement with previously reported data.^[6,8] We also determined the efficiency with which *Taq* inserts dATP opposite dT and dMMO2TP opposite d5SICS in sequence context I and observed values of $8.2 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$ and $9.7 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$, respectively, also in good agreement with previously reported data.^[5]

Insertion efficiencies of dMMO2TP derivatives opposite d5SICS

The insertion by Kf of each dMMO2 triphosphate derivative opposite d5SICS was characterized in sequence context I (Table 1). We found that the most simple derivative, dAMO1TP, is inserted opposite d5SICS with a second order rate constant of $8.6 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$. dAMO3TP is inserted with a similar efficiency ($1.1 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$), while dAMO2TP is inserted 6-fold less efficiently ($1.5 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$). However, dPMO1TP and dNMO1TP are inserted with higher efficiencies of 2.0×10^6 and $2.3 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$, respectively. The increased efficiency of dPMO1TP and dNMO1TP insertion opposite d5SICS by Kf results from both an increase in the apparent k_{cat} and a decrease in the apparent K_M .

We next characterized the ability of *Taq* to insert each dMMO2TP derivative opposite d5SICS in the same sequence context (Table 2). We found that *Taq* inserts dDMOTP with an efficiency of $1.6 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$, which is virtually identical to the efficiency of dMMO2TP insertion. We found that *Taq* inserts dAMO1TP, dAMO2TP, and dAMO3TP opposite d5SICS much less efficiently, with second order rate constants of $4.9 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$, $1.2 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$, and $1.6 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$, respectively. While insertion of dPMO1TP opposite d5SICS by *Taq* is also not very efficient ($3.2 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$), insertion of dNMO1TP is significantly more efficient ($4.3 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$).

Of the new dMMO2 analogues examined, the most promising in sequence context I are dPMO1TP, and especially dNMO1TP, thus we next determined whether their insertion efficiency by Kf opposite d5SICS is dependent on sequence context, using sequence context II (Table 3). Again, for comparison, we re-determined the efficiencies of dATP, dMMO2TP or dDMOTP insertion opposite their cognate base in the same sequence context. The natural base pair is synthesized with an efficiency of $5.8 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, while dMMO2TP is inserted opposite d5SICS with an efficiency of $1.9 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$; both values are in good agreement with previously reported data.^[13] Kf inserts dDMOTP opposite d5SICS with an efficiency of $3.9 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$. Interestingly, we found that insertion of both dPMO1TP

and dNMO1TP opposite d5SICS by Kf is approximately 4- and 8-fold more efficient than insertion of dDMOTP and dMMO2TP, respectively.

Efficiency and fidelity of dNaM-d5SICS synthesis

To provide a reference for how efficiently and selectively *Taq* synthesizes dNaM-d5SICS, we first measured the efficiency of Kf-mediated synthesis of dNaM-d5SICS and all possible mispairs in both strand contexts of sequence context I (Table 1). We found that d5SICSTP is inserted opposite dNaM with an efficiency of $2.1 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, in good agreement with previously reported data.^[13] However, we found that dNaMTP is inserted opposite d5SICS with an efficiency of $5.8 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$. While significantly greater than reported previously for this sequence context, it is similar to that reported for the same insertion in sequence context II.^[13] To confirm the rates for insertion in sequence context II, we reanalyzed the synthesis of dNaM-d5SICS in both strand contexts of sequence context II (Table 3). We found that d5SICSTP is inserted opposite dNaM and that dNaMTP is inserted opposite d5SICS with efficiencies of $1.2 \times 10^8 \text{ M}^{-1}\text{min}^{-1}$, and $2.8 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$, respectively, both in good agreement with the previously published data.

As already mentioned, the most efficiently inserted dNTP opposite d5SICS by Kf is dGTP, which we found to proceed with an efficiency of $1.5 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$, while dATP and dTTP are inserted less efficiently, dCTP is not inserted at detectable level, and d5SICSTP is inserted with an efficiency of $2.9 \times 10^5 \text{ M}^{-1}\text{min}^{-1}$ (Table 1). With dNaM in the template, the natural triphosphate most efficiently inserted by Kf is dATP, which proceeds with an efficiency of $1.8 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$ (Table 1). dTTP is also inserted, but with a reduced efficiency of $1.2 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$, while dGTP and dCTP are not inserted at detectable levels ($k_{\text{cat}}/K_{\text{M}} < 10^3 \text{ M}^{-1}\text{min}^{-1}$). dNaMTP is inserted opposite dNaM with an efficiency of $9.5 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$. The rates for the synthesis of these mispairs are all in good agreement with previously published data.^[13]

We next determined the efficiency and fidelity of dNaM-d5SICS synthesis by *Taq* in both strand contexts of sequence context I. Again, for comparison we first measured the efficiency with which *Taq* inserts dATP opposite dT in the same sequence context, which we found to be $8.2 \times 10^7 \text{ M}^{-1}\text{min}^{-1}$. We found that *Taq* inserts dNaMTP opposite d5SICS and d5SICSTP opposite dNaM with efficiencies of $7.6 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$ and $7.1 \times 10^6 \text{ M}^{-1}\text{min}^{-1}$, respectively. With d5SICS in the template, *Taq* inserts dGTP more efficiently than the other natural triphosphates, with an efficiency of $2.1 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$, followed by dATP and dTTP, with efficiencies of $3.5 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$ and $2.6 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$. *Taq* does not insert dCTP at a detectable rate ($< 10^3 \text{ M}^{-1}\text{min}^{-1}$). The most competitive mispair synthesized results from the insertion of d5SICSTP, which proceeds with an efficiency of only $1.8 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$. With dNaM in the template, *Taq* inserts dATP most efficiently, with a second order rate constant of $4.8 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$. dTTP is inserted with an efficiency of $2.4 \times 10^3 \text{ M}^{-1}\text{min}^{-1}$, while dGTP and dCTP are inserted with undetectable rates ($< 10^3 \text{ M}^{-1}\text{min}^{-1}$). dNaMTP is inserted opposite dNaM by *Taq* with an efficiency of $5.2 \times 10^4 \text{ M}^{-1}\text{min}^{-1}$.

Efficiency and fidelity of PCR amplification

To begin to examine unnatural base pair replication, which includes synthesis and extension in both strand contexts, we first explored the *Taq*-mediated PCR amplification of DNA containing d5SICS paired opposite either dNaM, dMMO2, dDMO, dNMO1, or dPMO1. *Taq* was employed despite its low fidelity for replication of the unnatural base pair, because it lacks exonuclease proofreading activity, to facilitate comparison with the steady-state kinetic data. In addition, the unnatural base pairs were incorporated into 134 nt DNA template, D6, in the middle of a six nucleotide randomized region.^[9] The randomized

template was selected to provide the strictest possible measure of fidelity as sequences with inherently low fidelity are expected to lose the unnatural base pair and then efficiently amplify. While it is not practical to characterize fidelity in these reactions due to the significant read through, sequencing traces clearly indicate that while **d5SICS-dNaM** is best replicated, amongst the other analogues, **dNMO1** is optimized for pairing opposite **d5SICS**, followed by **dDMO**, **dMMO2**, and **dPMO1** (Supporting Information, Fig. S28).

To examine amplification under more practical and high fidelity conditions we explored PCR using the exonuclease proficient Deep Vent polymerase of the same unnatural base pairs positioned in the middle of the 149 nt duplex referred to as D1, which was used previously to characterize the amplification of **dNaM-d5SICS** and **dMMO2-d5SICS**^[9] (Table 4 and Fig. S29). In this case fidelity was characterized as described previously.^[9] In agreement with previous results, **dNaM-d5SICS** is replicated with a remarkable fidelity of 99.7. Interestingly, **dNMO1-d5SICS**, **dDMO-d5SICS**, and **dMMO2-d5SICS** were also amplified with a similar fidelity, while as predicted by the steady-state kinetic data, **dPMO1-d5SICS** is amplified with a significantly lower fidelity.

Discussion

The identification **dMMO2-d5SICS** was a landmark in our efforts to identify an unnatural base pair, and early efforts directed toward its optimization yielded **dDMO-d5SICS** and in particular **dNaM-d5SICS**. While the former is replicated slightly better than **dMMO2-d5SICS**, **dNaM-d5SICS** is replicated significantly better, although whether this was unique to Kf, or a general property of the unnatural base pair was not known. In addition, the strategy of using nucleobases with little or no structural homology to their natural counterparts makes possible many different substituents, and it is unclear whether **dNaM-d5SICS** represents the best route to optimize **dMMO2-d5SICS**, especially considering the increased potential for linker modification of the more simple scaffolds.

In an effort to optimize the **dMMO2** scaffold for pairing with **d5SICS**, we examined derivatives with different substituents in place of the *para* methyl group. Specifically, five different substituents expected to alter the physicochemical properties of the nucleobases were examined, including amine, amide, trifluoroamide, nitro, and pyrrolo groups. The amine substituent is electron donating and expected to introduce a dipole along the C-N bond directed toward the aromatic ring. The amine group is also expected to form H-bonds with water molecules. While the amide groups are less electron donating, along with increased steric demands, they introduce increased H-bonding relative to the amine. The pyrrolo substituent reduces the electron donating ability of the nitrogen and adds steric bulk, but within the context of decreased H-bonding and increased stacking potential. In contrast, the nitro group is electron withdrawing and expected to introduce a significant dipole oriented along the carbon-nitrogen bond away from the aromatic ring. However, like the pyrrolo substituent, the nitro group is expected to decrease H-bonding, being only a moderate H-bond acceptor, and increase the ability of the nucleobase analog to stack with flanking nucleobases within the developing duplex.

With both Kf and *Taq*, the analogues examined clearly separate into two groups: **dAMO1TP**, **dAMO2TP**, and **dAMO3TP** are inserted less efficiently opposite **d5SICS** than **dMMO2TP**; and **dNMO1TP** and **dPMO1TP** are inserted more efficiently. For **dAMO1TP**, the decrease is 4-fold with Kf, while with *Taq* it is 20-fold, with the difference between the polymerases largely due to changes in the apparent k_{cat} ; the modification increases k_{cat} with Kf and decreases it with *Taq* while the K_M is similarly increased (~5-fold) with both enzymes. Because, the amino group is expected to increase the electron density of the nucleobase ring, the observed decrease in insertion efficiencies suggests that any favorable

increase in packing due to increased polarizability is offset by other deleterious factors, such as forced desolvation of the amino group, which is consistent with a similar effect on K_M observed with both enzymes. Modification of the amine with the acetyl group of dAMO2TP decreases insertion efficiency by both enzymes, relative to dMMO2TP, by a factor of 26 with Kf and more than a factor of 83 with *Taq*. The large reduction in efficiency with *Taq* results from a significant decrease in apparent binding, and an even larger decrease in turnover. Modification of the amino group with the trifluoroacetyl group of dAMO3TP also reduced the efficiency of insertion, but very differently with the two enzymes. With Kf, the decrease is only 4-fold relative to dMMO2TP, but it is more than 60-fold with *Taq*. The large decrease in recognition by *Taq* again results from both reduced binding and reduced turnover. Thus, insertion of these analogues appears to be limited by desolvation and steric or electrostatic clashes that are polymerase specific and generally more severe with *Taq*.

Relative to dMMO2TP, the behavior of dPMO1TP and dNMO1TP are very different from that of the other analogues with both Kf and *Taq*. With Kf, in both sequence contexts I and II, dPMO1TP and dNMO1TP are each inserted approximately 5-fold more efficiently, due to increased binding and increased turnover. In contrast, with in sequence context I, dPMO1TP is inserted by *Taq* 3-fold less efficiently than dMMO2TP, due to reduced binding, but dNMO1TP is inserted 4-fold more efficiently, due to a small increase in turnover and a slightly larger increase in apparent binding affinity. The data suggest that with Kf the presence of the nitro or the pyrrole group likely reduces the cost of desolvation relative to the amine or amide and also mediates favorable packing interactions with d5SICS or the polymerase in the developing transition state. As with the other analogues, *Taq* appears to be less accommodating to alterations in the size of the substituent, tolerating the nitro substituent, but not the pyrrole group.

The detailed kinetic analysis focused on the insertion of the different triphosphate analogue opposite d5SICS. However, replication requires base pair synthesis in the other strand context, the insertion of d5SICSTP opposite the analogue in template DNA, as well as the continued primer elongation in both strand contexts. The results of the PCR analysis suggest that the nitro and pyrrolo substituents of dNMO1 and dPMO1, respectively, do not significantly interfere with any of these other steps of replication. Moreover, the fidelities observed during PCR with *Taq* indicate that the improvement in steady-state triphosphate insertion characterized for dNMO1TP is also manifest as improved replication.

Generally, the data obtained with the five new analogues examined suggest that optimizing hydrophobicity, including reducing the cost of desolvation, and improving packing in the developing major groove are the most efficacious routes to optimization of the dMMO2 scaffold, although the interactions must be more carefully manipulated with *Taq* than with Kf, apparently due to a more discriminating active site. The importance of hydrophobicity and packing is also consistent with the remarkable insertion efficiency of dNaM opposite d5SICS. If supported by further study, these arguments suggest that, at least from the perspective of the *para* position of the dMMO2 scaffold, the developing major groove of the duplex within the polymerase active site is able to accommodate planar aromatic groups with favorable packing interactions, but the waters and metal ions found within the major groove of free duplex DNA are not yet available. Regardless of the underlying mechanism, the steady-state kinetic and PCR data indicate that dNMO1 is the most promising dMMO2 analogue yet identified.

Previous data collected for the Kf-mediated synthesis of dNaM-d5SICS suggested that its efficiency is strand and sequence context specific. However, this was based on a the efficiency of a single step in a single sequence context, the insertion of dNaMTP opposite d5SICS in sequence context I, which appeared to be significantly less efficient than that in

sequence context II, or than the insertion of d5SICSTP opposite dNaM in either sequence context. The efficiency of this reaction was re-evaluated in the current work and we found that the previous data underestimated the rate approximately 10-fold, which we attribute to the presence of an impurity. The corrected data places the efficiency of dNaMTP insertion on par with that of the others. Thus, the replication of dNaM-d5SICS does not appear to be strongly strand or sequence context dependent. The efficiency of both steps of unnatural base pair synthesis are within 4- to 13-fold that of a natural base pair. Moreover, relative to the most competitive mispair, the overall fidelity when both unnatural base pair synthesis and extension^[13] are combined is at least 10⁴. To our knowledge, this represents the most efficient and high fidelity replication reported to date for an unnatural base pair.

The synthesis of dNaM-d5SICS by Kf is remarkably efficient and selective, and it appears to be just as efficiently synthesized by *Taq*. *Taq* inserts both dNaMTP opposite d5SICS and d5SICSTP opposite dNaM only 10-fold less efficiently than a natural base pair in the same sequence context. Moreover, none of the natural dNTPs are inserted efficiently opposite either d5SICS and dNaM in the template, resulting in 150-fold or greater fidelities for this step alone. The efficient and selective recognition of dNaM-d5SICS by both Kf and *Taq* now allow us to conclude that the determinants of efficient replication are inherent to the nucleotides themselves.

Conclusion

The data reveal that reducing the cost of nucleotide desolvation and optimizing packing interactions within the developing major groove are promising routes to optimize the efficiency of polymerase-mediated insertion of the dMMO2TP analogues opposite d5SICS, and that other than dNaM, dNMO1 is the most promising analogue identified to date. While continued efforts toward the optimization of the dNMO1 scaffold are justified by its potential for accommodating linkers, the data reported herein for dNaM-d5SICS demonstrate just how challenging the identification of a more optimized base pair is likely to be. Thus, the data strongly suggest that efforts toward developing an unnatural base pair for *in vitro* applications should focus on the dNaM scaffold, including efforts to identify suitable sites for linker attachment to facilitate applications based on the site-specific modification of DNA or RNA. Such efforts are currently underway.

Experimental Section

General Methods

All reactions were carried out in oven-dried glassware under inert atmosphere, and all solvents were dried over 4 Å molecular sieves with the exception of tetrahydrofuran, which was distilled from sodium metal. All other reagents were purchased from Fisher or Aldrich. ¹H, ¹³C and ³¹P NMR spectra were recorded on Bruker DRX-600, DRX-500 or Varian Inova-300 spectrometers. High-resolution mass spectroscopic data were obtained on an ESI-TOF mass spectrometer (Agilent 6200 Series) at the TSRI Open Access Mass Spectrometry Lab and MALDI-TOF mass spectrometry (Applied Biosystems Voyager DE-PRO System 6008) from the TSRI Center for Protein and Nucleic Acid Research.

Synthetic Procedures and Characterizations

Compound 1—Compound 1 was synthesized according to the literature.^[29]

Compound 2—To a solution of *m*-anisidine (2 mL, 16.2 mmol, 1 equiv) in tetrahydrofuran (50 mL) at 0 °C was added NaHCO₃ (1.65 g, 19.4 mmol, 1.2 equiv). Benzyl chloroformate (2.8 mL, 19.4 mmol, 1.2 equiv) was then added dropwise under strong

stirring. The mixture was allowed to warm to room temperature over 20 min, stirred for an additional hour and then diluted with CH₂Cl₂. The organic layer was quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and evaporated. The residue was subjected to a short silica gel column chromatography with a step gradient of CH₂Cl₂ (20 – 100%) in hexane affording pure NH-Cbz *m*-anisidine. To a solution of the Cbz-protected *m*-anisidine in MeOH (120 mL) at -20 °C was added silver nitrate (4.48 g, 14.37 mmol) and iodine (3.65 g, 14.37 mmol). The mixture was stirred at -20 °C for 1 h, quenched by saturated aqueous Na₂S₂O₃ (80 mL) and filtered through a Celite pad. The filtrate was concentrated to 10% of its original volume and diluted with CH₂Cl₂ (200 mL). The organic layer was quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of CH₂Cl₂ (20 – 80%) in hexane. The desired compound **2** was obtained as white foam after evaporation of the solvent (4.04 g, 10.54 mmol, 65% yield). ¹H NMR (400 MHz, DMSO-*d*₆) δ 9.89 (s, 1H, NH), 7.60 (d, *J* = 8.5 Hz, 1H), 7.40 (m, 5H), 7.27 (d, *J* = 2.3 Hz, 1H), 6.90 (m, 1H), 5.16 (s, 2H), 3.76 (s, 3H). ¹³C NMR (100 MHz, CDCl₃) δ 158.66, 153.08, 139.53, 139.13, 135.79, 128.67, 128.49, 128.31, 112.37, 102.07, 78.04, 67.22, 56.33. HRMS (ESI⁺) *m/z* calcd for C₁₅H₁₅INO₃ (M+H)⁺ 384.0091, found 384.0097.

Compound 3—A mixture of palladium acetate (880 mg, 3.92 mmol, 0.15 equiv) and triphenylarsine (2 g, 3.92, 0.15 equiv) in dry dimethylformamide (150 mL) was stirred under argon atmosphere at room temperature for 20 min. To this mixture was added **2** (15 g, 39.2 mmol, 1.5 eq), **1** (9 g, 26.1 mmol, 1 eq) and tri-*n*-butylamine (9.3 mL, 39.2 mmol, 1.5 eq) in dimethylformamide (5 mL), and the resulting reaction mixture was stirred under nitrogen at 70 °C for 15 h. The mixture was cooled to 0 °C and 1 M tetrabutylammonium fluoride (44 mL, 44 mmol) in tetrahydrofuran was added and the mixture was stirred for 2 h while warming the reaction to room temperature. The reaction mixture was filtered through Celite and extracted with ethyl acetate and saturated aqueous NaHCO₃. The combined organic layers were dried (Na₂SO₄), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of ethyl acetate (5 – 30%) in CH₂Cl₂. The eluted product was dissolved in acetic acid (50 mL) and acetonitrile (50 mL). The solution was cooled to 0 °C, sodium triacetoxyborohydride (4.5 g, 21.23 mmol) was added, and the mixture was stirred for 1 h. The reaction mixture was diluted with ethyl acetate, quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of methanol (0 – 4%) in ethyl acetate. The desired compound **3** was obtained as white foam after evaporation of the solvent (6.58 g, 17.62 mmol, 45% yield). ¹H NMR (400 MHz, MeOD) δ 7.49 – 7.27 (m, 6H, H-ar, H-6), 7.26 – 7.17 (m, 1H, H-3), 6.95 (dd, *J* = 8.2, 2.0 Hz, 1H, H-5), 5.39 (dd, *J* = 10.2, 5.5 Hz, 1H, H-1'), 5.17 (s, 2H, OCH₂Ar (Cbz)), 4.35 – 4.24 (m, 1H, H-3'), 3.93 (td, *J* = 5.2, 2.7 Hz, 1H, H-4'), 3.81 (s, 3H, OCH₃), 3.74 – 3.59 (m, 2H, H-5', H-5''), 2.27 (ddd, *J* = 13.1, 5.6, 1.9 Hz, 1H, H-2'), 1.82 (ddd, *J* = 13.1, 10.3, 6.1 Hz, 1H, H-2''). ¹³C NMR (100 MHz, MeOD) δ 156.7 (C₂(C-OMe)), 154.4 (NHC=O), 139.1 (C₄(C-NHCbz)), 136.6 (C_q, Car), 128.1-127.6 (CH, Ar), 126.0 (C₆), 124.6 (C₁(c-sugar)), 110.2 (C₅), 101.2 (C₃), 87.1 (C₄'), 74.8 (C₁'), 73.0 (C₃'), 66.1 (OCH₂Ph), 62.7 (C₅'), 54.3 (OCH₃), 41.7 (C₂'). HRMS (ESI⁺) *m/z* calcd for C₂₀H₂₄NO₆ (M+H)⁺ 374.1598, found 374.1615.

Compound 4—To a solution of **3** (2.5 g, 6.7 mmol, 1 equiv) in dry pyridine (100 mL) was added 1,3-dichloro-1,1,3,3-tetraisopropylidisiloxane (2.6 mL, 8.1 mmol, 1.2 equiv) dropwise over 15 min. The reaction mixture was stirred for 2 h at room temperature under nitrogen atmosphere, concentrated two-fold and diluted with CH₂Cl₂. The organic layer was quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and

evaporated. The residue was dissolved in ethyl acetate (75 mL) and Et₃N (250 μ L, 1.76 mmol) was added. The resulting solution was treated with 10% Pd/C (150 mg) under H₂ atmosphere and allowed to stir until the presence of the starting material could no longer be detected by TLC (~1 h). The reaction mixture was filtered through Celite and the filtrate extracted with ethyl acetate and saturated aqueous NaHCO₃. The combined organic layers were dried (Na₂SO₄), filtered and evaporated. The residue was subjected to a silica gel column chromatography with a step gradient with ethyl acetate (0 – 5%) in CH₂Cl₂ containing 0.5% triethylamine. The desired compound **4** was obtained as a colorless oil after evaporation of the solvent (21 g, 4.28 mmol, 64% yield). ¹H NMR (500 MHz, CD₃CN) δ 7.10 (d, J = 8.1 Hz, 1H, H-6), 6.26 (d, J = 2.1 Hz, 1H, H-3), 6.18 (dd, J = 8.1, 2.1 Hz, 1H, H-5), 5.15 (m, 1H, H-1'), 4.47 (dt, J = 7.7, 5.6 Hz, 1H, H-3'), 4.11 (s, 2H, NH₂), 4.05 – 3.84 (m, 2H, H-5', H-5''), 3.75 – 3.68 (m, 4H, OCH₃, H-4'), 2.20 (ddd, J = 12.7, 7.3, 5.4 Hz, 1H, H-2'), 2.00 (m, 1H, H-2''), 1.13 – 1.00 (m, 28H, iPr). ¹³C NMR (125 MHz, CD₃CN) δ 158.3 (C₂), 149.6 (C₄), 128.0 (C₆), 120.2 (C₁), 107.1 (C₅), 98.6 (C₃), 86.3 (C₄'), 74.3 (C₁'), 74.0 (C₃'), 64.2 (C₅'), 55.8 (OCH₃), 42.5 (C₂'), 17.9–13.4 (iPr). HRMS (ESI⁺) m/z calcd for C₂₄H₄₄NO₅Si₂ (M+H)⁺ 482.2752, found 482.2748.

Compound 5a—To a solution of **4** (150 mg, 0.31 mmol, 1 equiv) and potassium iodide (5.1 mg, 0.031 mmol, 0.1 eq.) in CH₃CN (1.1 mL) was added a solution of 70% aqueous *tert*-Butyl hydroperoxide (160 μ L, 1.18 mmol, 3.8 equiv) dropwise over a period of 15 min and the mixture was stirred at 75 °C in the absence of light for 2 h. The mixture was quenched with saturated aqueous Na₂S₂O₃, washed with brine, extracted with ethyl acetate, dried (Na₂SO₄), filtered and evaporated. The resulting residue was dissolved in tetrahydrofuran (1.5 mL), 1 M tetrabutylammonium fluoride (0.5 mL, 0.5 mmol) in tetrahydrofuran was added and the mixture was stirred for 1 h at room temperature. The reaction mixture was diluted with ethyl acetate, quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of methanol (0 – 4%) in ethyl acetate. The desired compound **5a** was obtained as white foam after evaporation of the solvent (32 mg, 0.12 mmol, 38 %). ¹H NMR (500 MHz, CD₃CN) δ 7.82 (dd, J = 8.4, 2.2 Hz, 1H, H-5), 7.77 – 7.69 (m, 2H, H-6, H-3), 5.32 (dd, J = 10.1, 5.8 Hz, 1H, H-1'), 4.24 (dt, J = 5.2, 2.1 Hz, 1H, H-3'), 3.91 (s, 3H, OCH₃), 3.88 (td, J = 5.0, 2.6 Hz, 1H, H-4'), 3.63 – 3.59 (dd, J = 5.0, 0.9 Hz, 2H, H-5', H-5''), 2.32 (ddd, J = 13.0, 5.7, 2.0 Hz, 1H), 1.68 (ddd, J = 13.1, 10.1, 5.9 Hz, 1H). ¹³C NMR (125 MHz, CD₃CN) δ 157.4 (C₂), 149.0 (C₄), 140.3 (C₁), 127.1 (C₆), 116.6 (C₅), 106.1 (C₃), 88.5 (C₄'), 75.5 (C₁'), 73.9 (C₃'), 63.7 (C₅'), 56.8 (OCH₃), 42.9 (C₂'). HRMS (ESI⁺) m/z calcd for C₁₂H₁₆NO₆ (M+H)⁺ 270.0972, found 270.0984.

Compound 5b—A solution of **4** (250 mg, 0.52 mmol, 1 equiv) and 2,5-dimethoxytetrahydrofuran (170 mg, 1.3 mmol, 2.5 equiv) in H₂O (0.8 mL) was heated to 140 °C for 30 min in a microwave synthesizer (Biotage AB, Sweden). The reaction was allowed to cool and the resulting mixture was diluted with ethyl acetate, quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and evaporated. The resulting residue was dissolved in tetrahydrofuran (3 mL), 1 M tetrabutylammonium fluoride (1 mL, 1 mmol) in tetrahydrofuran was added and the mixture was stirred for 1 h at room temperature. The reaction mixture was diluted with ethyl acetate, quenched with saturated aqueous NaHCO₃, washed with brine, dried (Na₂SO₄), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of methanol (0 – 8%) in CH₂Cl₂. The desired compound **5b** was obtained as white foam after evaporation of the solvent (95 mg, 0.33 mmol, 63 %). ¹H NMR (600 MHz, CD₃CN) δ 7.52 (d, J = 8.0 Hz, 1H, H-6), 7.17 (m, J = 2.1 Hz, 2H, H-7), 7.02 (m, 2H, H-5, H-3), 6.29 (t, J = 2.2 Hz, 2H, H-8), 5.30 (dd, J = 10.2, 5.6 Hz, 1H, H-1'), 4.23 (m, 1H, H-3'), 3.88 (s, 2H, OCH₃), 3.83 (td, J = 5.0, 2.6 Hz, 1H, H-4'), 3.60 (t, J = 5.1 Hz, 2H), 3.21 (d, J = 3.7 Hz, 1H,

OH-3'), 2.92 (t, $J = 5.7$ Hz, 1H, OH-5'), 2.22 (m, 1H, H-2'), 1.76 (ddd, $J = 13.0, 10.3, 6.0$ Hz, 1H, H-2''). ^{13}C NMR (150 MHz, CD_3CN) δ 158.1 (C_2), 141.6 (C_4), 129.1 (C_6), 128.0 (C_1), 120.2 (C_7), 112.6 (C_5), 111.1 (C_8), 104.0 (C_3), 88.2 (C_4'), 75.4 (C_1'), 74.1 (C_3'), 63.8 (C_5'), 56.4 (OCH_3), 43.1 (C_2'). HRMS (ESI^+) m/z calcd for $\text{C}_{16}\text{H}_{20}\text{NO}_4$ ($\text{M}+\text{H}$) $^+$ 290.1387, found 290.1397.

Compound 5c—To a solution of **4** (250 mg, 0.52 mmol, 1 equiv) in CH_2Cl_2 (600 μL) was added triethylamine (125 μL , 0.89 mmol, 1.7 equiv) and then acetic anhydride dropwise (60 μL , 0.62 mmol, 1.2 equiv). The reaction mixture was stirred at room temperature for 20 min and then quenched with saturated aqueous NaHCO_3 . The organic layer was diluted with CH_2Cl_2 , washed with brine, dried (Na_2SO_4), filtered and evaporated. The resulting residue was dissolved in tetrahydrofuran (2 mL), 1 M tetrabutylammonium fluoride (1 mL, 1 mmol) in tetrahydrofuran was added and the mixture was stirred for 1 h at room temperature. The reaction mixture was diluted with ethyl acetate, quenched with saturated aqueous NaHCO_3 , washed with brine, dried (Na_2SO_4), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of methanol (0 – 12%) in CH_2Cl_2 . The desired compound **5c** was obtained as white foam after evaporation of the solvent (75 mg, 0.27 mmol, 51% yield). ^1H NMR (600 MHz, CD_3OD) δ 7.40 (d, $J = 8.3$ Hz, 1H, H-6), 7.33 (d, $J = 1.8$ Hz, 1H, H-3), 7.02 (dd, $J = 8.3, 1.8$ Hz, 1H, H-5), 5.38 (dd, $J = 10.2, 5.6$ Hz, 1H, H-1'), 4.27 (m, 1H, H-3'), 3.90 (td, $J = 5.2, 2.7$ Hz, 1H, H-4'), 3.80 (s, 3H, OCH_3), 3.65 (m, 2H, H-5', H-5''), 2.27 (ddd, $J = 13.1, 5.5, 1.7$ Hz, 1H, H-2'), 2.10 (s, 3H, COCH_3), 1.80 (ddd, $J = 13.2, 10.3, 6.0$ Hz, 1H, H-2''). ^{13}C NMR (150 MHz, CD_3OD) δ 171.5 ($\text{NHC}=\text{O}$), 157.8 (C_2), 140.1 (C_4), 127.2 (C_6), 127.2 (C_1), 112.7 (C_5), 103.7 (C_3), 88.5 (C_4'), 76.2 (C_1'), 74.4 (C_3'), 64.0 (C_5'), 55.7 (OCH_3), 43.1 (C_2'), 23.9 (COCH_3). HRMS (ESI^+) m/z calcd for $\text{C}_{14}\text{H}_{20}\text{NO}_5$ ($\text{M}+\text{H}$) $^+$ 282.1341, found 290.1351.

Compound 5d—To a solution of **4** (350 mg, 0.73 mmol, 1 equiv) in CH_2Cl_2 (2 mL) at 10 °C was added triethylamine (170 μL , 1.24 mmol, 1.7 equiv) and then trifluoroacetic anhydride dropwise (170 μL , 0.88 mmol, 1.2 equiv). The reaction mixture was stirred at room temperature for 20 min at 10 °C and then quenched with saturated aqueous NaHCO_3 . The organic layer was diluted with CH_2Cl_2 , washed with brine, dried (Na_2SO_4), filtered and evaporated. The resulting residue was dissolved in tetrahydrofuran (5 mL), 1 M tetrabutylammonium fluoride (1.9 mL, 1.9 mmol) in tetrahydrofuran was added and the mixture was stirred for 1 h at room temperature. The reaction mixture was diluted with ethyl acetate, quenched with saturated aqueous NaHCO_3 , washed with brine, dried (Na_2SO_4), filtered and evaporated. The residue was subjected to silica gel column chromatography with a step gradient of methanol (0 – 10%) in CH_2Cl_2 . The desired compound **5c** was obtained as white foam after evaporation of the solvent (200 mg, 0.60 mmol, 81% yield). ^1H NMR (600 MHz, CD_3CN) δ 9.25 (s, 1H, NH), 7.50 (d, $J = 8.2$ Hz, 1H, H-6), 7.27 (d, $J = 1.9$ Hz, 1H, H-3), 7.19 (m, 1H, H-5), 5.27 (dd, $J = 10.2, 5.6$ Hz, 1H, H-1'), 4.22 (m, 1H, H-3'), 3.82 (d, $J = 27.6$ Hz, 4H, H-4', OCH_3), 3.59 (d, $J = 4.9$ Hz, 2H, H-5', H-5''), 2.22 (m, 1H, H-2'), 1.72 (m, 1H, H-2''). ^{13}C NMR (150 MHz, CD_3CN) δ 157.5 (C_2), 156.2-155.4 ($\text{NHC}=\text{O}$), 136.9 (C_4), 129.7 (C_6), 127.3 (C_1), 119.8-114.1 (CF_3), 113.7 (C_5), 104.7 (C_3), 88.2 (C_4'), 75.4 (C_1'), 74.1 (C_3'), 63.9 (C_5'), 56.2 (OCH_3), 43.1 (C_2'). HRMS (ESI^+) m/z calcd for $\text{C}_{14}\text{H}_{17}\text{F}_3\text{NO}_5$ ($\text{M}+\text{H}$) $^+$ 336.1053, found 336.1066.

General Procedure for Triphosphate Synthesis

Proton sponge (1.3 equiv) and the free nucleoside derivative (1.0 equiv) were dissolved in dry trimethyl phosphate (40 equiv) and cooled to -15 °C under nitrogen atmosphere. Freshly distilled POCl_3 (1.3 equiv) was added dropwise and the resulting mixture was stirred at -10 °C for 2 h. Tributylamine (6.0 equiv) and a solution of tributylammonium pyrophosphate (5.0 eq.) in dimethylformamide (0.5 M) were added. Over 30 min, the reaction was allowed

to warm slowly to 0 °C and then was quenched by addition of 0.5 M aqueous Et₃NH₂CO₃ (TEAB) pH 7.5 (2 vol-equiv). The mixture was diluted two-fold with H₂O and the product was isolated on a DEAE Sephadex column (GE Healthcare) with an elution gradient of 0 to 1.2 M TEAB, evaporated, co-distilled with H₂O (3×). Additional purification by reverse-phase (C18) HPLC (0 – 35% CH₃CN in 0.1 M TEAB, pH 7.5) was performed.

Compound 6a—³¹P NMR (162 MHz, D₂O) δ -10.52 (d, J = 19.8 Hz, γ -P), -10.84 (d, J = 20.2 Hz, α -P), -22.92 (t, J = 20.1 Hz, β -P). MS (MALDI-TOF⁺, matrix : 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₂H₁₇NO₁₅P₃, 508.2; found, 508.3. $\epsilon(\lambda = 330 \text{ nm}) = 2200 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon(\lambda = 285 \text{ nm}) = 4800 \text{ M}^{-1} \text{ cm}^{-1}$

Compound 6b—³¹P NMR (162 MHz, D₂O) δ -10.22 (d, J = 19.8 Hz, γ -P), -10.75 (d, J = 20.1 Hz, α -P), -22.82 (t, J = 20.0 Hz, β -P). MS (MALDI-TOF, matrix: 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₆H₂₁NO₁₃P₃, 528.3; found 527.8. $\epsilon(\lambda = 283 \text{ nm}) = 5080 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon(\lambda = 256 \text{ nm}) = 10850 \text{ M}^{-1} \text{ cm}^{-1}$

Compound 6c—³¹P NMR (202 MHz, D₂O) δ -6.35 (d, J = 16.9 Hz, γ -P), -10.74 (d, J = 19.7 Hz, α -P), -22.38 (m, β -P). MS (MALDI-TOF⁺, matrix : 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₄H₂₁NO₁₄P₃, 520.2; found 519.3. $\epsilon(\lambda = 282 \text{ nm}) = 2700 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon(\lambda = 248 \text{ nm}) = 7500 \text{ M}^{-1} \text{ cm}^{-1}$

Compound 6d—³¹P NMR (162 MHz, D₂O) δ -10.38 (d, J = 19.8 Hz, γ -P), -10.87 (d, J = 20.2 Hz, α -P), -22.82 (t, J = 20.0 Hz, β -P). MS (MALDI-TOF⁺, matrix : 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₄H₁₈F₃NO₁₄P₃, 574.2; found 573.9. $\epsilon(\lambda = 285 \text{ nm}) = 3780 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon(\lambda = 251 \text{ nm}) = 7450 \text{ M}^{-1} \text{ cm}^{-1}$

Compound 6e—A solution of **6d** in aqueous ammonia (28% NH₃ w/v) (1 mL) was stirred for 1 h. Ammonia was removed by vacuum concentration (SpeedVac, 20 min) and the resulting solution was separated by reverse-phase (C18) HPLC (0 – 35% CH₃CN in 0.1 M TEAB, pH 7.5) providing pure compound **6e**. ³¹P NMR (162 MHz, D₂O) δ -9.21 (d, J = 20.0 Hz, γ -P), -10.85 (d, J = 20.4 Hz, α -P), -22.62 (t, J = 20.6 Hz, β -P). MS (MALDI-TOF, matrix: 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₂H₁₉NO₁₃P₃, 478.2; found 477.8. $\epsilon(\lambda = 284 \text{ nm}) = 2450 \text{ M}^{-1} \text{ cm}^{-1}$; $\epsilon(\lambda = 240 \text{ nm}) = 9250 \text{ M}^{-1} \text{ cm}^{-1}$

dNaMTP—dNaMTP was synthesized using the General Procedure described above starting from dNaM nucleoside (Berry & Associates Inc.). ³¹P NMR (162 MHz, D₂O) δ -8.96 (d, J = 19.4 Hz, γ -P), -10.81 (d, J = 20.1 Hz, α -P), -22.66 (t, J = 20.2 Hz, β -P). MS (MALDI-TOF⁺, matrix : 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₆H₂₀O₁₃P₃, 513.2; found 513.4. $\epsilon(\lambda = 230 \text{ nm}) = 75,000 \text{ M}^{-1} \text{ cm}^{-1}$

d5SICSTP—d5SICSTP was synthesized using the General Procedure described above starting from the d5SICS nucleoside (Berry & Associates Inc.). ³¹P NMR (162 MHz, D₂O) δ -10.11 (d, J = 20.0 Hz, γ -P), -11.05 (d, J = 20.2 Hz, α -P), -22.75 (t, J = 20.0 Hz, β -P). MS (MALDI-TOF⁺, matrix : 9-aminoacridine) (m/z): [M-H]⁺ calcd for C₁₅H₁₉NO₁₂P₃S, 530.3; found 529.3. $\epsilon(\lambda = 365 \text{ nm}) = 3,950 \text{ M}^{-1} \text{ cm}^{-1}$

Gel-Based Kinetic Assays

Primer oligonucleotides (Integrated DNA Technologies) were 5'-radiolabeled with T4 polynucleotide kinase (New England Biolabs) and [γ -³²P]-ATP (GE Biosciences) and annealed to template oligonucleotides^[13] by heating to 95 °C followed by slow cooling to room temperature. Reactions were initiated by adding a solution of 2× dNTP solution (5 μ L) to a solution containing polymerase (0.10 – 1.23 nM) and primer template (40 nM) in

reaction buffer (5 μ L); Klenow reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM DTT and 50 μ g/mL acetylated BSA) for Kf polymerase, ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 10 mM KCl, 2 mM MgSO_4 and 0.1% Triton X-100, pH 8.8) for Taq polymerase. After incubation at 25 $^\circ\text{C}$ (Kf) or 50 $^\circ\text{C}$ (Taq) for 3-10 min the reactions were quenched with 20 μ L of loading dye (95% formamide, 20 mM EDTA, and sufficient amounts of bromophenol blue and xylene cyanole). Reaction products were resolved by 15% polyacrylamide gel electrophoresis, and gel band intensities corresponding to the extended and unextended primers were quantified by phosphorimaging (Storm Imager, Molecular Dynamics) and Quantity One (BioRad) software. Plots of k_{obs} versus triphosphate concentration were fit to the Michaelis-Menten equation using the program Origin (Microcal Software) to determine V_{max} and K_{M} . k_{cat} was determined from V_{max} by normalizing by the total enzyme concentration. Each reaction was run in triplicate and standard deviations for both kinetic parameters were determined (see Tables 1 – 3). An example of the raw kinetic data is shown in Figure S27 in the Supporting Information.

PCR amplification

DNA templates D1 and D6 were synthesized as described previously.^[9] PCR amplification (see SI for details and sequences) was carried out starting with 0.1 ng of D6 or 1 ng of D1 (Taq or DeepVent, respectively) in 1 $\times$ ThermoPol reaction buffer with the following modifications: MgSO_4 adjusted to 6.0 mM, 0.6 mM or 0.7 mM each natural dNTP (Taq or DeepVent, respectively), 0.1 mM each unnatural triphosphate, 1 μ M each primer (see SI for sequences), and 0.03 unit/ μ L of Taq or 0.02 unit/ μ L of DeepVent (exo⁺) in an iCycler Thermal Cycler (Bio-Rad) under following thermal cycling conditions: 94 $^\circ\text{C}$, 30 s; 48 $^\circ\text{C}$, 30 s; 65 $^\circ\text{C}$, 4 min, 18 or 13 cycles (Taq or DeepVent, respectively). Upon completion, PCR products were purified utilizing the PureLinkTM PCR purification kit (Invitrogen), quantified by fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Invitrogen) and sequenced on 3730 DNA Analyzer (Applied Biosystems) to determine fidelity of unnatural base pair replication (see SI and Ref. [9] for details).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this work was provided by the National Institute for Health (GM060005). We wish to thank Dr. Phillip Ordoukhanian, director of The Center for Protein and Nucleic Acids Research at The Scripps Research Institute, for analytical support.

References

1. Horlacher J, Hottiger M, Podust VN, Hübscher U, Benner SA. Proc Natl Acad Sci USA. 1995; 92:6329–6333. [PubMed: 7541538]
2. Lutz MJ, Held HA, Hottiger M, Hübscher U, Benner SA. Nucleic Acids Res. 1996; 24:1308–1313. [PubMed: 8614635]
3. Piccirilli JA, Krauch T, Moroney SE, Benner SA. Nature. 1990; 343:33–37. [PubMed: 1688644]
4. Yang Z, Chen F, Chamberlin SG, Benner SA. Angew Chem Int Ed. 2010; 49:177–180.
5. Hwang GT, Romesberg FE. J Am Chem Soc. 2008; 130:14872–14882. [PubMed: 18847263]
6. Leconte AM, Hwang GT, Matsuda S, Capek P, Hari Y, Romesberg FE. J Am Chem Soc. 2008; 130:2336–2343. [PubMed: 18217762]
7. Leconte, AM.; Romesberg, FE. Protein Engineering. K, C.; RajBhandary, UL., editors. Springer-Verlag; Berlin: 2009. p. 291-314.

8. Malyshev DA, Pfaff DA, Ippoliti SI, Hwang GT, Dwyer TJ, Romesberg FE. *Chem Eur J*. 2010; 16:12650–12659. [PubMed: 20859962]
9. Malyshev DA, Seo YJ, Ordoukhanian P, Romesberg FE. *J Am Chem Soc*. 2009; 131:14620–14621. [PubMed: 19788296]
10. Matsuda S, Fillo JD, Henry AA, Rai P, Wilkens SJ, Dwyer TJ, Geierstanger BH, Wemmer DE, Schultz PG, Spraggon G, Romesberg FE. *J Am Chem Soc*. 2007; 129:10466–10473. [PubMed: 17685517]
11. Matsuda S, Leconte AM, Romesberg FE. *J Am Chem Soc*. 2007; 129:5551–5557. [PubMed: 17411040]
12. McMinn DL, Ogawa AK, Wu Y, Liu J, Schultz PG, Romesberg FE. *J Am Chem Soc*. 1999; 121:11585–11586.
13. Seo YJ, Hwang GT, Ordoukhanian P, Romesberg FE. *J Am Chem Soc*. 2009; 131:3246–3252. [PubMed: 19256568]
14. Seo YJ, Matsuda S, Romesberg FE. *J Am Chem Soc*. 2009; 131:5046–5047. [PubMed: 19351201]
15. Seo YJ, Romesberg FE. *ChemBioChem*. 2009; 10:2394–2400. [PubMed: 19722235]
16. Yu C, Henry AA, Romesberg FE, Schultz PG. *Angew Chem Int Ed*. 2002; 41:3841–3844.
17. Hirao I. *Curr Opin Chem Biol*. 2006; 10:622–627. [PubMed: 17035074]
18. Hirao I, Kimoto M, Mitsui T, Fujiwara T, Kawai R, Sato A, Harada Y, Yokoyama S. *Nat Methods*. 2006; 3:729–735. [PubMed: 16929319]
19. Hirao I, Mitsui T, Kimoto M, Yokoyama S. *J Am Chem Soc*. 2007; 129:15549–15555. [PubMed: 18027940]
20. Kimoto M, Kawai R, Mitsui T, Yokoyama S, Hirao I. *Nucleic Acids Res*. 2009; 37:e14. [PubMed: 19073696]
21. Mitsui T, Kitamura A, Kimoto M, To T, Sato A, Hirao I, Yokoyama S. *J Am Chem Soc*. 2003; 125:5298–5307. [PubMed: 12720441]
22. Keefe AD, Cload ST. *Curr Opin Chem Biol*. 2008; 12:448–456. [PubMed: 18644461]
23. Fritzsche, W.; Bier, F. *International Symposium on DNA-Based Nanodevices*; American Institute of Physics Conference Proceedings; 2008;
24. Vrabel M, Pohl R, Votruba I, Sajadi M, Kovalenko SA, Ernsting NP, Hocke M. *Org Biomol Chem*. 2008; 6:2852–2860. [PubMed: 18688477]
25. Wilson MA, Filzen G, Welmaker GS. *Tetrahedron Lett*. 2009; 50:4807–4809.
26. Ludwig J, Eckstein F. *J Org Chem*. 1989; 54:631–635.
27. Boosalis MS, Petruska J, Goodman MF. *J Biol Chem*. 1987; 262:14689–14696. [PubMed: 3667598]
28. Datta K, LiCata VJ. *Nucleic Acids Res*. 2003; 31:5590–5597. [PubMed: 14500822]
29. Larsen E, Jørgensen PT, Sofan MA, Pederson EB. *Synthesis*. 1994:1037–1038.

Figure 1.

A) dMMO2-d5SICS unnatural base pair. B) dNaM and dDMO. C) dMMO2 analogues synthesized and evaluated in this study. Only nucleobase moieties are shown; sugar and phosphate backbone are omitted for clarity.

Scheme 1. Conditions: (a) CBZ-Cl, NaHCO₃, THF, rt, 20 min; (b) I₂, Ag₂SO₄, MeOH, -20 °C, 1 h; (c) Pd(OAc)₂, AsPh₃, *n*Bu₃N, DMF, 70 °C, 15 h; (d) TBAF 1 M in THF, 0 °C → rt, 2 h; (e) NaBH(OAc)₃, AcOH, CH₃CN, 0 °C, 1 h; (f) 1,3-dichloro-1,1,3,3-tetraisopropylidisiloxane, pyridine, rt, 2 h; (g) 10% Pd/C, H₂, EtOAc, rt, 1 h; (h) KI, *t*-butyl hydroperoxide aq 70%, CH₃CN, dark, 75 °C, 2 h; (i) 2,5-dimethoxytetrahydrofuran, H₂O, microwave-140 °C, 30 min; (j) Ac₂O, Et₃N, CH₂Cl₂, rt, 20 min; (k) trifluoroacetic anhydride, Et₃N, CH₂Cl₂, 10 °C, 20 min; (l) TBAF 1 M in THF, 1 h; (m) proton sponge, POCl₃, PO(OMe)₃, -15 °C → -10 °C, 3 h then Bu₃N, (Bu₃NH)₂H₂P₂O₇ in DMF, -10 °C → 0 °C, 30 min then TEAB buffer (0.5 M), rt, 10 min; (n) NH₄OH 30%, rt, 1 h.

Table 1

Kinetic data for Kf-mediated insertion of triphosphates (dYTP) opposite d**5SICS** or d**NaM** in the template in sequence context I; insertion of dATP opposite dT is provided for comparison.

5'-d(TAATACGACTCACTATAGGGAGA)				
3'-d(ATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC)				
X	Y	k_{cat} (min ⁻¹)	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ (× 10 ⁵ M ⁻¹ min ⁻¹)
T	A	4.1 ± 0.3	0.0053 ± 0.0004	7700
5SICS	NaM	14.6 ± 0.8	0.25 ± 0.04	580
	DMO	27.0 ± 1.5	12.2 ± 1.5	22
	MMO2	13.7 ± 2.0	35.8 ± 0.6	3.8
	NMO1	50.1 ± 8.8	22.0 ± 3.5	23
	PMO1	41.7 ± 6.7	20.7 ± 2.6	20
	AMO1	15.3 ± 3.5	178 ± 62	0.86
	AMO3	13.4 ± 2.8	117 ± 30	1.1
	AMO2	7.0 ± 1.0	464 ± 100	0.15
	5SICS	12.6 ± 0.7	44.2 ± 11.7	2.9
	A	2.1 ± 0.3	52.4 ± 7.2	0.40
NaM	G	11.7 ± 1.3	75.5 ± 3.7	1.5
	C	n.d. ^[a]	n.d. ^[a]	< 0.01
	T	2.1 ± 0.4	230 ± 11	0.091
	5SICS	8.3 ± 1.1	0.039 ± 0.004	2100
	NaM	51.5 ± 3.4	5.4 ± 0.3	95
	A	24.7 ± 4.7	14.0 ± 0.8	18
	G	n.d. ^[a]	n.d. ^[a]	< 0.01
	C	n.d. ^[a]	n.d. ^[a]	< 0.01
	T	1.6 ± 0.2	129 ± 15	0.12

^[a] Below limits of detection.

Table 2

Kinetic data for *Taq*-mediated insertion of triphosphates (dYTP) opposite d**5SICS** or d**NaM** in the template in sequence context I; insertion of dATP opposite dT is provided for comparison.

5'-d(TAATACGACTCACTATAGGGAGA)				
3'-d(ATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC)				
X	Y	k_{cat} (min ⁻¹)	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ (× 10 ⁵ M ⁻¹ min ⁻¹)
T	A	2.6 ± 0.2	0.032 ± 0.002	820
5SICS	NaM	2.7 ± 0.4	0.35 ± 0.04	76
	DMO	3.3 ± 1.7	21.0 ± 0.5	1.6
	MMO2	3.4 ± 0.3	35.0 ± 6.0	0.97
	NMO1	5.5 ± 0.2	12.8 ± 3.4	4.3
	PMO1	3.6 ± 1.4	113 ± 25	0.32
	AMO1	0.83 ± 0.13	170 ± 31	0.049
	AMO3	0.35 ± 0.06	214 ± 28	0.016
	AMO2	0.45 ± 0.06	370 ± 36	0.012
	5SICS	3.4 ± 0.5	189 ± 16	0.18
	A	0.5 ± 0.05	147 ± 7	0.035
NaM	G	4.5 ± 0.1	216 ± 42	0.21
	C	n.d. [a]	n.d. [a]	< 0.01
	T	0.8 ± 0.1	303 ± 19	0.026
	5SICS	2.4 ± 0.6	0.34 ± 0.03	71
	NaM	7.9 ± 1.4	151 ± 9	0.52
	A	4.0 ± 0.8	84.0 ± 4.0	0.48
	G	n.d. [a]	n.d. [a]	< 0.01
	C	n.d. [a]	n.d. [a]	< 0.01
	T	1.0 ± 0.2	421 ± 71	0.024

[a] Below limits of detection.

Table 3

Kinetic data for Kf-mediated insertion of triphosphates (dYTP) opposite d**5SICS** or d**NaM** in the template in sequence context II; insertion of dATP opposite dT is provided for comparison.

5'-d(TAATACGACTCACTATAGGGAGC)				
3'-d(ATTATGCTGAGTGATATCCCTCGXTCTAGGTTACGGCAGGATCGC)				
X	Y	k_{cat} (min ⁻¹)	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ (× 10 ⁵ M ⁻¹ min ⁻¹)
T	A	0.95 ± 0.10	0.0017 ± 0.0001	5800
5SICS	NaM	11.4 ± 2.3	0.41 ± 0.01	280
	DMO	14.2 ± 1.1	35.8 ± 1.5	3.9
	MMO2	8.6 ± 0.2	46.6 ± 7.1	1.9
	NMO1	33.6 ± 7.0	23.0 ± 0.1	15
	PMO1	19.0 ± 1.6	13.5 ± 1.5	14
NaM	5SICS	5.8 ± 1.1	0.049 ± 0.003	1200

Table 4Fidelities of Deep Vent-mediated PCR amplification of unnatural base pairs. ^[a]

Base pair	Fidelity ^[b]
dNaM-d5SICS	99.7
dDMO-d5SICS	99.7
dMMO2-d5SICS	99.7
dPMO1-d5SICS	92.4
dNMO1-d5SICS	99.5

^[a] See text for experimental details.^[b] Fidelity is defined as % unnatural base pair retention per doubling, see Ref. [9].