CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.201000959

Solution Structure, Mechanism of Replication, and Optimization of an Unnatural Base Pair

Denis A. Malyshev,[a] Danielle A. Pfaff,[b] Shannon I. Ippoliti,[b] Gil Tae Hwang,[a, c] Tammy J. Dwyer,*[b] and Floyd E. Romesberg*[a]

Abstract: As part of an ongoing effort to expand the genetic alphabet for in vitro and eventual in vivo applications, we have synthesized a wide variety of predominantly hydrophobic unnatural base pairs and evaluated their replication in DNA. Collectively, the results have led us to propose that these base pairs, which lack stabilizing edge-on interactions, are replicated by means of a unique intercalative mechanism. Here, we report the synthesis and characterization of three novel derivatives of the nucleotide analogue d**MMO2**, which forms an unnatural base pair with the nucleotide analogue d**5SICS**. Replacing the *para*-methyl substituent of d**MMO2** with an annulated furan ring (yielding d**FMO**) has a dramatically negative effect on replication, while replacing it with a methoxy (d**DMO**) or with a thiomethyl group (d**TMO**) improves replication in both steady-state assays and during PCR amplification. Thus, d**TMO**–d**5SICS**, and especially d**DMO**–d**5SICS**, represent significant progress toward the expansion of the genetic alphabet. To elucidate the structure–activity relationships governing unnatural base pair replication, we determined the solution structure of duplex DNA containing the parental d**MMO2**–d**5SICS** pair, and also used this structure to generate models of the derivative base pairs. The results strongly support the intercalative mechanism of replication, reveal a surprisingly high level of specificity that may be achieved by optimizing packing interactions, and should prove invaluable for the further optimization of the unnatural base pair.

Keywords: DNA replication · genetic alphabet · intercalations · polymerase · unnatural base pairs

Introduction

The natural genetic alphabet relies on the selective pairing of the four natural nucleotides, which is governed by a combination of hydrogen-bonding[1] and shape complementarity.[2–4] However, a priori there is no reason that these forces should be unique in their ability to mediate base pairing. With the long-term goal of expanding the genetic code, we[5–15] and others[3,16–22] have explored the development of unnatural nucleotides bearing nucleobase analogues that pair through hydrophobic and packing forces. Among the most promising predominantly hydrophobic base pairs that we have identified is that formed between d**MMO2** and d**5SICS** (d**MMO2**–d**5SICS**, Figure 1a).[8] The d**MMO2**–d**5SICS** pair is synthesized (by insertion of each unnatural triphosphate opposite the other in the template)[23] and then extended (by insertion of the next correct dNTP) with relatively high efficiency and fidelity by diverse polymerases,[13] including the exonuclease deficient Klenow fragment (Kf) of *E. coli* DNA polymerase I.

The step that most limits the replication of DNA containing d**MMO2**–d**5SICS** is the insertion of d**MMO2**TP opposite d**5SICS** in the template. In general, we have found that the rates of insertion are most sensitive to triphosphate derivatization; thus our efforts to optimize d**MMO2**–d**5SICS** have focused on modification of d**MMO2** (with the goal of opti-

[a] D. A. Malyshev, Dr. G. T. Hwang, Prof. F. E. Romesberg
Department of Chemistry
The Scripps Research Institute
10550 North Torrey Pines Road
La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-7472
E-mail: floyd@scripps.edu

[b] D. A. Pfaff, S. I. Ippoliti, Prof. T. J. Dwyer
Department of Chemistry and Biochemistry
University of San Diego, 5998 Alcala Park
San Diego, CA 92110 (USA)
Fax: (+1) 619-260-2211
E-mail: tdwyer@sandiego.edu

[c] Dr. G. T. Hwang
Department of Chemistry
Kyungpook National University
Daegu 702-701 (Korea)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201000959.

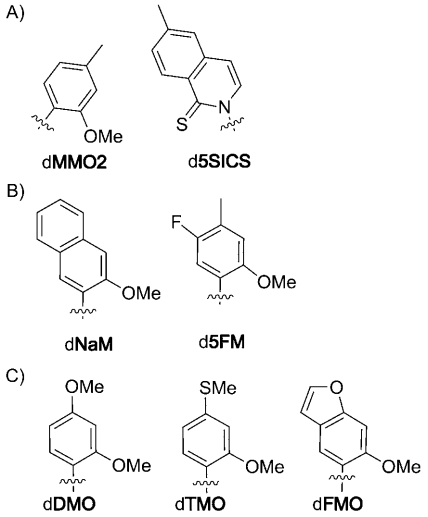

Figure 1. Unnatural nucleotides used in the current study. Only nucleobases are shown; sugars and phosphates have been omitted for clarity.

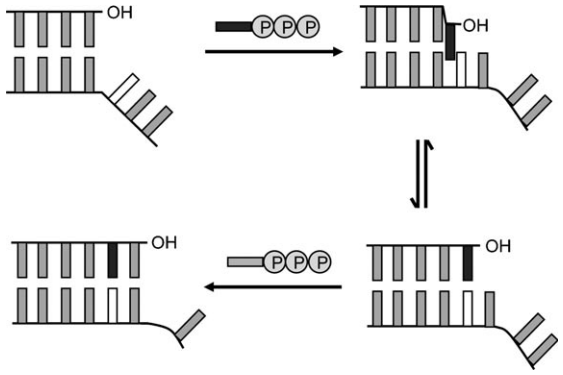

Figure 2. Intercalative model for unnatural base pair replication.[9] The unnatural nucleobase in the template is shown in white and the natural or unnatural nucleobase of the incoming dNTP is shown in black.

mizing the insertion of d**MMO2**TP opposite d**5SICS**). Because previous studies showed that the methoxy and sulfur substituents at the position *ortho* to the glycosidic linkage are essential for efficient extension of the nascent unnatural base pair,[6–8] our efforts focused specifically on *meta* and *para* derivatizations of d**MMO2**.[9,10] Indeed, we already demonstrated that both d**NaM**TP and d**5FM**TP (Figure 1b) are inserted opposite d**5SICS** with higher efficiency and fidelity than d**MMO2**TP, and that d**NaM**–d**5SICS** is sufficiently well recognized for expansion of the genetic alphabet in vitro.[9] However, we anticipate that one of the most interesting in vitro applications of an expanded genetic alphabet will be the use of an unnatural base pair to site specifically modify DNA or RNA in a format consistent with enzymatic synthesis, and one limitation of d**NaM** is that the second aromatic ring precludes derivatization at the position most commonly used to attach linkers (i.e., the C5 position of natural pyrimidines[24]). While other positions might be found to derivatize d**NaM** with linkers, modifications of d**MMO2** that improve replication without blocking the C5 position are desirable.

Previous kinetic[5–11] and structural[12] studies have prompted us to propose that the predominantly hydrophobic unnatural base pairs are replicated via a unique mechanism involving partial interstrand intercalation (Figure 2). In this mechanism, the unnatural triphosphates are recognized by at least partial intercalation of their nucleobases into the polymerase-bound template strand between the nucleobase of their cognate unnatural nucleotide and a flanking nucleobase. This mode of insertion likely results from the high hydrophobic packing and stacking potential of the unnatural nucleobases and the absence of interactions that favor edge-to-edge pairing, and also suggests that increased packing within the major groove underlies the more efficient insertion of d**NaM**TP and d**5FM**TP opposite d**5SICS**, relative to d**MMO2**TP. Importantly, the model also suggests that deintercalation is required to position the primer terminus appropriately for continued extension, which is also favored by hydrogen-bond formation between a polymerase-based donor and the *ortho* substituents of d**5SICS** and d**MMO2**,[6–8] explaining why they are essential for extension. Thus, a subtle balance between intercalation and deintercalation is required for the unnatural base pair to be synthesized and extended efficiently. (Despite the requirement of both intercalation and deintercalation, for simplicity, we refer to this mechanism as the "intercalative mechanism.").

To test the intercalative mechanism of replication and to continue our efforts to optimize the d**MMO2**–d**5SICS** unnatural base pair, we now report the kinetic and structural characterization of base pairs formed between d**5SICS** and three d**MMO2** derivatives that have been modified at the *meta* and/or *para* positions: d**DMO**, d**TMO**, and d**FMO** (Figure 1c). Complete steady-state kinetic analysis, as well as the efficiency and fidelity of PCR amplification show that derivatization with a thiomethyl, or especially with a methoxy substituent, significantly improves replication: d**TMO**–d**5SICS** and d**DMO**–d**5SICS** are replicated more efficiently than d**MMO2**–d**5SICS**. Unlike d**NaM**, the C5 position of d**TMO** and d**DMO** is available for derivatization, making them amenable for uses involving the site-specific modification of DNA. Surprisingly, the furan substituent of d**FMO** dramatically reduces the efficiency of each step of replication. To help understand these trends in replication, we also report the solution structure of d**MMO2**–d**5SICS** in duplex DNA, along with models of the derivative base pairs. The data strongly supports the intercalative model of replication and provides a rationale for the observed variations in the recognition of the d**MMO2** derivatives.

Results

Unnatural base pair design and evaluation: The d**DMO** and d**TMO** nucleotides were designed to probe the effects of heteroatom substitution in the developing major groove at the primer terminus while leaving the C5 position unsubsti-

Table 1. Second-order rate constants (k_{cat}/K_M) for Kf-mediated synthesis and extension of d**MMO2** or an analogue paired opposite d**5SICS** in the template.[a]

Sequence Context I 5′-dTAATACGACTCACTATAGGGAGA**Y** 3′-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC				Sequence Context II 5′-dTAATACGACTCACTATAGGGAGC**Y** 3′-dATTATGCTGAGTGATATCCCTCG**X**TCTAGGTTACGGCAGGATCGC			
X	**Y**	Synthesis [$M^{-1}min^{-1}$]	Extension [$M^{-1}min^{-1}$]	**X**	**Y**	Synthesis [$M^{-1}min^{-1}$]	Extension [$M^{-1}min^{-1}$]
dT	dA	3.2×10^{8}	1.7×10^{8}	dT	dA	3.3×10^{8}	3.1×10^{8}
d**5SICS**	d**MMO2**[b]	3.6×10^{5}	1.9×10^{6}	d**5SICS**	d**MMO2**[c]	4.8×10^{5}	3.2×10^{5}
d**5SICS**	d**DMO**	1.7×10^{6}	7.8×10^{6}	d**5SICS**	d**DMO**	5.0×10^{5}	1.6×10^{6}
d**5SICS**	d**TMO**	1.6×10^{6}	9.8×10^{5}	d**5SICS**	d**TMO**	not measured	not measured
d**5SICS**	d**FMO**	2.5×10^{4}	5.4×10^{4}	d**5SICS**	d**FMO**	not measured	not measured
d**5SICS**	d**5SICS**[b]	2.7×10^{4}	$<1.0\times10^{3}$[d]	d**5SICS**	d**5SICS**[c]	2.4×10^{5}	$<1.0\times10^{3}$[d]
d**5SICS**	dA[b]	2.2×10^{4}	1.0×10^{4}	d**5SICS**	dA[c]	5.9×10^{4}	3.6×10^{3}
d**5SICS**	dC[b]	$<1.0\times10^{3}$[d]	4.2×10^{3}	d**5SICS**	dC[c]	$<1.0\times10^{3}$[d]	$<1.0\times10^{3}$[d]
d**5SICS**	dG[b]	1.3×10^{5}	4.9×10^{3}	d**5SICS**	dG[c]	1.9×10^{5}	6.5×10^{2}
d**5SICS**	dT[b]	1.3×10^{4}	4.0×10^{5}	d**5SICS**	dT[c]	8.5×10^{3}	3.3×10^{5}

[a] In all cases "**X**" is the nucleotide in the template. For synthesis, "**Y**" corresponds to the triphosphate inserted, and for extension, it corresponds to the nucleotide at the nascent primer terminus. [b] Taken from reference [8]. [c] Taken from reference [9]. [d] Below the limit of detection.

tuted. In contrast, like d**NaM**, the C5 position of d**FMO** is substituted, and this analogue was designed to introduce a more rigidly positioned heteroatom, while simultaneously increasing the potential for nucleobase packing. Oligonucleotides containing the unnatural nucleotides were synthesized to act as templates or primers so that both synthesis and extension of each unnatural base pair could be evaluated independently. Kinetic analyses were performed under steady-state conditions using Kf, and second-order rate constants (efficiency, k_{cat}/K_M) were determined (individual k_{cat} and K_M values are reported in Supporting Information). PCR amplification was also performed to further evaluate the unnatural base pair and to gauge the potential practical utility. The unnatural base pairs are generally referred to as d**Y**–d**X**, when no strand context is implied, while d**Y**:d**X** is used to refer specifically to the strand context with d**Y** in the primer strand and d**X** in the template strand.

Unnatural base pair synthesis—insertion of dMMO2TP analogues opposite d5SICS: To characterize the effects of major groove substitution, we first characterized the rate at which the d**MMO2**TP analogues are inserted opposite d**5SICS** in the template (Table 1). For comparison, d**MMO2**TP itself is inserted with a second-order rate constant of $3.6\times10^{5}\,M^{-1}min^{-1}$.[8] We found that d**DMO**TP and d**TMO**TP are inserted fivefold more efficiently, resulting entirely from a decreased apparent K_M for unnatural triphosphate binding. However, insertion of d**FMO**TP opposite d**5SICS** is more than tenfold less efficient than insertion of d**MMO2**TP, due to changes in both the apparent k_{cat} and K_M. A complete characterization of mispair synthesis with d**5SICS** in the template was reported previously.[8]

Unnatural base pair synthesis—insertion of d5SICSTP opposite dMMO2 analogues: To characterize the recognition of the unnatural nucleotides in the template, we examined the efficiencies with which Kf inserts d**5SICS**TP (Table 2). For reference, d**5SICS**TP is inserted opposite d**MMO2** with an efficiency of $4.7\times10^{7}\,M^{-1}min^{-1}$, and it is inserted opposite itself in the template with an efficiency of $1.2\times10^{5}\,M^{-1}min^{-1}$. Insertion of natural dNTPs opposite d**MMO2** in the template is not detectable ($k_{cat}/K_M<1.0\times10^{3}\,M^{-1}min^{-1}$), except in the case of dATP, which is inserted with moderate efficiency ($k_{cat}/K_M=1.0\times10^{5}\,M^{-1}min^{-1}$).[8] We found that insertion of d**5SICS** opposite d**DMO** is threefold less efficient than opposite d**MMO2**. However, d**DMO** also directs the synthesis of the mispair with itself or that with dA less efficiently than does d**MMO2**, without significantly increasing the synthesis efficiencies of any of the other mispairs.

We found that insertion of d**5SICS**TP opposite d**TMO** is twofold less efficient than opposite d**MMO2**. Interestingly, the thiomethyl substituent significantly increases the rate at which dATP is inserted, while only slightly increasing the rates at which the other mispairs are synthesized. Surprisingly, incorporating the major groove oxygen atom as a restrained furan (i.e., d**FMO**), as opposed to a free methyl ether (i.e., d**DMO**), dramatically reduces the efficiency of d**5SICS**TP insertion (by 100-fold). While d**FMO** does not direct Kf to synthesize the self-pair ($k_{cat}/K_M<1.0\times10^{3}\,M^{-1}min^{-1}$), it does direct the relatively more efficient insertion of each natural triphosphate.

Unnatural base pair extension—extension of dMMO2 analogues paired opposite d5SICS: Efficient and high-fidelity replication of DNA containing the unnatural base pair also requires efficient continued primer elongation after incorporation of the unnatural nucleotide, and inefficient primer extension after incorporation of an incorrect nucleotide. We first examined the efficiencies with which Kf extends primers terminating with a d**MMO2** analogue paired opposite d**5SICS** or a natural nucleotide by insertion of dCTP opposite dG (Table 1). For comparison, Kf extends d**MMO2**:d**5SICS** (primer:template) with a second-order rate constant of $1.9\times10^{6}\,M^{-1}min^{-1}$. Changing the major groove substituent from the methyl group of d**MMO2** to the methoxy group of d**DMO** results in a fourfold increase in extension efficiency.

Table 2. Second-order rate constants (k_{cat}/K_M) for Kf-mediated synthesis and extension of d**5SICS** paired opposite d**MMO2** or an analogue in the template.[a]

Sequence Context I[b]				Sequence Context II[b]			
X	**Y**	Synthesis [$M^{-1}min^{-1}$]	Extension [$M^{-1}min^{-1}$]	**X**	**Y**	Synthesis [$M^{-1}min^{-1}$]	Extension [$M^{-1}min^{-1}$]
d**MMO2**	d**5SICS**[c]	4.7×10^{7}	6.7×10^{5}	d**MMO2**	d**5SICS**[d]	6.6×10^{7}	1.7×10^{6}
d**MMO2**	d**MMO2**[c]	1.2×10^{5}	5.3×10^{3}	d**MMO2**	d**MMO2**[d]	1.8×10^{6}	$<1.0\times10^{3}$[e]
d**MMO2**	dA[c]	1.0×10^{5}	4.6×10^{4}	d**MMO2**	dA[d]	1.7×10^{6}	1.1×10^{4}
d**MMO2**	dC[c]	$<1.0\times10^{3}$[e]	1.2×10^{6}	d**MMO2**	dC[d]	3.0×10^{3}	4.4×10^{5}
d**MMO2**	dG[c]	$<1.0\times10^{3}$[e]	$<1.0\times10^{3}$[f]	d**MMO2**	dG[d]	7.9×10^{3}	$<1.0\times10^{3}$[e]
d**MMO2**	dT[c]	$<1.0\times10^{3}$[e]	6.6×10^{5}	d**MMO2**	dT[d]	5.2×10^{3}	2.0×10^{6}
d**DMO**	d**5SICS**	1.5×10^{7}	2.6×10^{6}	d**DMO**	d**5SICS**	9.7×10^{7}	1.3×10^{6}
d**DMO**	d**DMO**	7.1×10^{4}	$<1.0\times10^{3}$[f]	d**DMO**	d**DMO**	1.6×10^{5}	$<1.0\times10^{3}$[e]
d**DMO**	dA	8.3×10^{4}	1.1×10^{5}	d**DMO**	dA	8.4×10^{5}	6.9×10^{3}
d**DMO**	dC	$<1.0\times10^{3}$[e]	1.3×10^{6}	d**DMO**	dC	$<1.0\times10^{3}$[e]	1.1×10^{6}
d**DMO**	dG	$<1.0\times10^{3}$[e]	$<1.0\times10^{3}$[e]	d**DMO**	dG	$<1.0\times10^{3}$[e]	$<1.0\times10^{3}$[e]
d**DMO**	dT	2.5×10^{3}	2.4×10^{5}	d**DMO**	dT	1.8×10^{3}	9.7×10^{4}
d**TMO**	d**5SICS**	2.7×10^{7}	1.9×10^{6}				
d**TMO**	d**TMO**	not measured	8.3×10^{3}				
d**TMO**	dA	8.5×10^{5}	5.3×10^{4}				
d**TMO**	dC	2.5×10^{3}[f]	2.2×10^{5}				
d**TMO**	dG	3.0×10^{3}	$<1.0\times10^{3}$[e]				
d**TMO**	dT	1.5×10^{3}	3.6×10^{4}				
d**FMO**	d**5SICS**	4.9×10^{5}	9.1×10^{3}				
d**FMO**	d**FMO**	$<1.0\times10^{3}$[e]	$<1.0\times10^{3}$[e]				
d**FMO**	dA	7.1×10^{5}	7.0×10^{3}				
d**FMO**	dC	7.5×10^{3}[f]	4.9×10^{3}				
d**FMO**	dG	9.0×10^{3}	$<1.0\times10^{3}$[e]				
d**FMO**	dT	2.2×10^{4}	5.2×10^{3}				

[a] In all cases "**X**" is the nucleotide in the template. For synthesis, "**Y**" corresponds to the triphosphate inserted, and for extension, it corresponds to the nucleotide at the nascent primer terminus. [b] For actual sequence see Table 1. [c] Taken from reference [8]. [d] Taken from reference [9]. [e] Below the limit of detection. [f] Kinetic parameters were calculated based on $n+2$ product, as dCTP is inserted slowly against the unnatural base pair and then efficiently against dG, the next nucleotide in the template.

However, the thiomethoxy and furanyl substituents result in approximately two- and 40-fold reduced efficiencies. Extension efficiencies of primers terminating with a natural nucleotide paired opposite d**5SICS** have been reported previously.[8]

Unnatural base pair extension—extension of d5SICS paired opposite dMMO2 analogues: We next examined unnatural base pair extension with the d**MMO2** analogues in the template paired opposite either the correct d**5SICS** nucleotide or one of the incorrect unnatural or natural nucleotides at the primer terminus (Table 2). For comparison, Kf extends primers terminating with d**5SICS** paired opposite d**MMO2** with an efficiency of $6.7\times10^{5}\,M^{-1}min^{-1}$. Kf does not efficiently extend primers terminating with the d**MMO2** self-pair or the dG:d**MMO2** mispair; however, the mispairs with dA, dT, and especially dC are extended more efficiently.[8] We found that primers terminating with d**5SICS** paired opposite d**DMO** are extended fourfold more efficiently than when paired opposite d**MMO2**. Primers terminating with the d**DMO** self-pair are extended less efficiently than those terminating with the d**MMO2** self-pair. As with d**MMO2**, primers terminating with dG paired opposite d**DMO** are not extended at a detectable rate, while primers terminating with dA are extended slightly faster, and primers terminating with dT or dC are extended slower. Extension of d**5SICS**:d**TMO** is threefold faster than extension of d**5SICS**:d**MMO2**. Again, primers terminating with dG paired opposite d**TMO** are not extended, while the dA:d**TMO** and dA:d**MMO2** mispairs are extended with similar efficiencies, and the mispairs with dT or dC paired opposite d**TMO** are extended an order of magnitude slower than when paired opposite d**MMO2**. Surprisingly, the d**5SICS**:d**FMO** pair is extended 70-fold less efficiently than d**5SICS**:d**MMO2**. Moreover, all of the mispairs between a natural nucleotide and d**FMO** are extended with rates slower than $7.8\times10^{3}\,M^{-1}min^{-1}$, revealing that both correct pairs and mispairs with d**FMO** in the template are extended only poorly by Kf.

dDMO–d5SICS replication as a function of sequence context: The steady-state kinetic data described above suggest that Kf recognizes d**DMO**–d**5SICS** better than d**MMO2**–d**5SICS** or the other derivatized unnatural base pairs. Because the practical utility of an unnatural base pair depends on its sequence-independent replication, we examined replication of d**DMO**–d**5SICS** in a second sequence context, hereafter referred to as sequence context II. In this context the unnatural nucleotide is positioned in the template between a 3′-dG and a 5′-dT (Table 1 and Table 2), as opposed to between a 3′-dT and a 5′-dG as in the context examined above, hereafter referred to a context I.

For comparison, Kf inserts d**MMO2**TP opposite d**5SICS** in sequence context II with the same efficiency as in context I ($\approx4\times10^{5}\,M^{-1}min^{-1}$).[9] We found that sequence context has a slightly larger effect on d**DMO**TP insertion, with a threefold lower efficiency in context II than in context I (Table 1). Thus, while d**DMO**TP is inserted opposite d**5SICS** in context I better than d**MMO2**TP, the two triphosphates are inserted with the same efficiency in context II. Sequence context also has a larger effect on the synthesis of d**5SICS**:d**DMO** than on that of d**5SICS**:d**MMO2**, in this case the efficiency of synthesis is increased more than sixfold, to the remarkable efficiency of $9.7\times10^{7}\,M^{-1}min^{-1}$, which is the most efficient rate for the synthesis of any unnatural base pair identified to date. In fact this efficiency is only marginally less than that for a natural base pair in the same sequence context ($3.3\times10^{8}\,M^{-1}min^{-1}$). While the efficiencies of mispairing with d**DMO** (i.e., self-pair formation)

and dA are also increased, they remain more than two-orders of magnitude less efficient, and the mispairs resulting from dCTP or dGTP insertion remain undetectable.

The effect of sequence context on the Kf-mediated extension of d**DMO**–d**5SICS** was also examined. For comparison, Kf extends d**MMO2**:d**5SICS** in context II approximately sixfold less efficiently than in context I.[9] However, it generally extends each mispair with lower efficiency, as well. We find that d**DMO**:d**5SICS** is also extended fivefold less efficiently in context II. In contrast, Kf extends d**5SICS**:d**MMO2** in context II threefold more efficiently than in context I, while it extends each mispair less efficiently, with the exception of dT:d**MMO2**, which it extends approximately threefold more efficiently.[9] We find that Kf extends the d**5SICS**:d**DMO** heteropair with similar efficiencies in both sequence contexts. Similar efficiencies were also observed for the extension of the mispairs with dG, dC, and dT paired opposite d**DMO** in the template, but surprisingly, extension of the mispair with dA is an order of magnitude less efficient in context II than in context I.

Generality of unnatural base-pair recognition: While derivatization of the nucleobase scaffold commonly results in large effects on the recognition of the nucleotide as a triphosphate, modifications to the templating nucleotide are typically less perturbative.[9,10] Thus, it is surprising that Kf recognizes d**FMO** in the template so poorly, relative to d**MMO2** or d**TMO**, both during unnatural base pair synthesis and extension. To determine whether this observation is specific for Kf, or whether it is inherent to the unnatural base pair itself, we characterized the ability of another A family polymerase, Taq, as well as a more diverged B family polymerase, exonuclease-negative Vent, to insert d**5SICS**TP opposite d**MMO2**, d**TMO**, or d**FMO** (Table 3).

Table 3. Second-order rate constant for incorporation of d**5SICS**TP against **X** in the template by different polymerases.[a]

Polymerase	**X**	k_{cat}/K_M [$M^{-1}min^{-1}$]
Kf(exo$^-$)	d**MMO2**	4.7×10^{7}[b]
Kf(exo$^-$)	d**TMO**	2.7×10^{7}
Kf(exo$^-$)	d**FMO**	4.9×10^{5}
Taq	d**MMO2**	3.5×10^{6}[c]
Taq	d**TMO**	6.4×10^{6}
Taq	d**FMO**	1.8×10^{4}
Vent(exo$^-$)	d**MMO2**	9.9×10^{6}[c]
Vent(exo$^-$)	d**TMO**	5.5×10^{6}
Vent(exo$^-$)	d**FMO**	1.1×10^{5}

[a] See the experimental section for experimental details. [b] Taken from reference [8]. [c] Taken from reference [13].

We found that Taq and Vent insert d**5SICS**TP opposite d**MMO2** with an efficiency of 3.5×10^{6} and $9.9\times10^{6}\,M^{-1}min^{-1}$, respectively.[13] These two polymerases insert the same triphosphate opposite d**TMO** in the template with similar efficiencies of $\approx6\times10^{6}\,M^{-1}min^{-1}$. However, just as observed with Kf, the efficiency of d**5SICS**TP insertion opposite d**FMO** by either Taq or Vent is greatly reduced relative to insertion opposite either d**MMO2** or d**TMO**. Thus, for all three polymerases, d**5SICS**TP is inserted opposite d**MMO2** and d**TMO** with similar efficiencies, but it is inserted opposite d**FMO** with an efficiency that is approximately two-orders of magnitude reduced. These results suggest that the factors disfavoring d**FMO** recognition are inherent to the unnatural base pair.

PCR amplification of DNA containing the unnatural base pairs: We recently showed that DNA containing d**MMO2**–d**5SICS** or d**NaM**–d**5SICS** in a variety of sequence contexts is PCR amplified with good efficiency and fidelity using multiple thermostable polymerases, including exonuclease-positive Deep Vent.[15] To further characterize the effects of the major groove modifications, the Deep Vent-mediated PCR amplification of DNA containing d**DMO**–d**5SICS**, d**TMO**–d**5SICS**, or d**FMO**–d**5SICS** was characterized to determine the amplification efficiency (fold amplification of strand) and fidelity (percentage of strands that retain the unnatural base pair per doubling) (Table 4 and Figure S2 in the Supporting Information; templates range in size from 134 to 149 nucleotides). As predicted by the steady-state data, d**DMO**–d**5SICS** is amplified with highest efficiency and fidelity, followed by d**TMO**–d**5SICS**, and then d**FMO**–d**5SICS**, which is replicated with lower efficiency and fidelity than is d**MMO2**–d**5SICS**. With template D1, where the unnatural base pair is flanked by a natural dG–dC and dA–dT, d**DMO**–d**5SICS** is amplified with virtually natural like efficiency and fidelity. To examine the sequence dependence of amplification, d**DMO**–d**5SICS** was further characterized with templates D2–D6 (see Supporting Information for full sequences). As expected, both efficiency and fidelity decreased slightly with increasing dG–dC content of the flanking DNA, as it does with natural sequences,[25,26] but the fact that it remained high in the randomized sequence context of duplex D6 suggests that the efficiencies and fidelities are generally reasonable in different sequence contexts.

Table 4. PCR efficiency and fidelity.[a]

Template	Base-pair amplified	Amplification	Fidelity[b]
D1[c]	d**MMO2**–d**5SICS**	224	99.4
D1	d**DMO**–d**5SICS**	397	99.8
D1	d**TMO**–d**5SICS**	364	99.1
D1	d**FMO**–d**5SICS**	121	91.9
D7[c]	dA–dT	556	–
D2	d**DMO**–d**5SICS**	69	95.7
D3	d**DMO**–d**5SICS**	130	97.9
D4	d**DMO**–d**5SICS**	78	90.7
D5[c]	d**MMO2**–d**5SICS**	25	97.1
D5	d**DMO**–d**5SICS**	12	94.4
D6[c]	d**MMO2**–d**5SICS**	52	92.9
D6	d**DMO**–d**5SICS**	109	97.6

[a] Conditions: 1 ng of the DNA template; dNTPs:d**X**TPs = 600:400 μM, 6 mM $MgSO_4$, 0.03 U μL^{-1} of the enzyme, 8 min extension, 14 cycles. [b] Calculated from sequencing data as described in Supporting Information. [c] Taken from reference [15].

Structures of the unnatural base pairs: To help elucidate the factors underlying unnatural base pair recognition, we determined the NMR structure of a 12-mer duplex containing d**5SICS** and d**MMO2** at the complementary positions 7 and 18 within the duplex (Figure 3, top). Resonance assignments for the duplex followed conventional NOESY based methods.[27] The NOESY and DQF-COSY spectra suggest that the unnatural base pair adopts a single, well defined structure, with only small distortions localized to the region of the unnatural base pair. Following standard protocols,[28] a family of 15 structures were generated and used to generate an average structure (Figure 3 bottom). In the average structure, both nucleobases of the unnatural base pair are positioned within the interior of a B-form duplex. Key cross-peaks in the NOESY spectra that support this conclusion include: d**5SICS**$_7$ HD to dC$_6$ H5, dC$_6$ H6, and dC$_6$ C1′H; d**MMO2**$_{18}$ CH$_3$ to dG$_{19}$ H8; d**5SICS**$_7$ CH$_3$ to A$_{17}$ H8; and d**MMO2**$_{18}$ CH$_3$/OCH$_3$ to dG$_{10}$ H8, in addition to cross-strand NOEs between d**5SICS**$_7$ HB and d**MMO2**$_{18}$ CH$_3$/OCH$_3$ and HH (see Figures S3 and S4 in the Supporting Information). However, slight distortions of the duplex, relative to a canonical B-form duplex, were apparent at the site occupied by the unnatural base pair. Specifically, relative to a canonical B-form duplex, the C1′–C1′ distance within the unnatural base pair is elongated ≈3 Å, and the nucleobases are inclined ≈10°, tilted ≈30°, and tipped ≈5°, with an increase in rise of ≈1.5 Å. The deoxyribose rings of the d**5SICS**$_7$–d**MMO2**$_{18}$ adopt a C2′-*endo* conformation, with an average sugar pucker (pseudorotation phase angle) of 137° (Figure S6 in the Supporting Information).

The structure clearly reveals that the unnatural nucleobases pair via partial interstrand intercalation (Figure 4 A). While d**5SICS**$_7$ stacks well with dT$_8$, it is not well packed with dC$_6$, and instead reaches across the duplex and partially intercalates into the opposite strand between d**MMO2**$_{18}$ and dA$_{17}$. Correspondingly, the nucleobase moiety of d**MMO2**$_{18}$ appears to stack rather poorly with both dA$_{17}$ and dG$_{19}$, and instead packs with d**5SICS**$_7$ from the opposite strand. This mode of pairing appears to induce an approximately 5 Å stagger of the d**5SICS**$_7$ nucleobase relative to that of d**MMO2**$_{18}$. The *ortho* sulfur and methoxy groups are oriented into the minor groove of the duplex, as predicted based on the expected *anti* geometry of the nucleotides, which is confirmed by cross-strand NOEs between d**MMO2**$_{18}$ CH$_3$/OCH$_3$ and d**5SICS**$_7$ HB, between d**5SICS**$_7$ HB and d**MMO2**$_{18}$ HH, as well as the absence of NOEs from HC, HD, and HE of d**5SICS**$_7$ to any proton of d**MMO2**$_{18}$. As further support of this nucleotide geometry, the aromatic protons giving rise to sequential NOEs between aromatic and C1′ protons along each strand include d**5SICS**$_7$ HE and d**MMO2**$_{18}$ HF. The methoxy group of d**MMO2**$_{18}$ rotates out of planarity with the aromatic ring to achieve favorable van der Waals contact with the polarizable sulfur group of d**5SICS**$_7$. This orientation necessarily places the ring substituted methyl groups of d**5SICS**$_7$ and d**MMO2**$_{18}$ in close con-

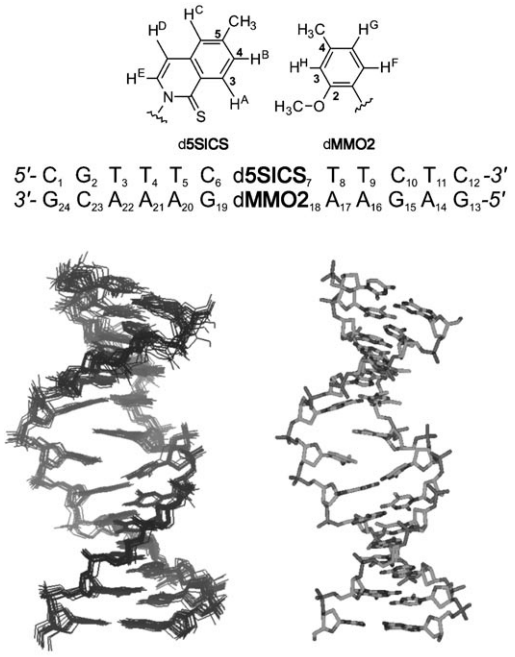

Figure 3. Top: Sequence of the duplex characterized by NMR and structure of d**5SICS**–d**MMO2** with atoms labeled. Family of structures (bottom left) and average structure (bottom right) generated from the NMR data.

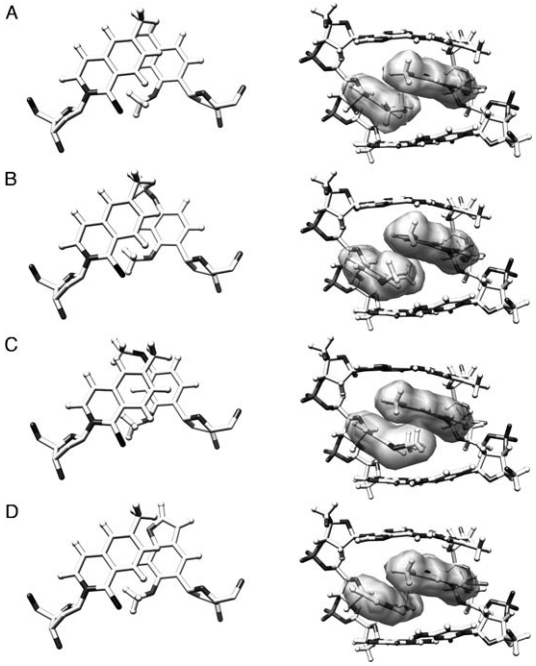

Figure 4. Structure of unnatural base pairs viewed along the helix axis (left; with sugar hydrogen atoms omitted for clarity) or from the major groove (right). A) The NMR structure of the d**MMO2**–d**5SICS** base pair, and B–D) the models generated from the parental base pair for d**DMO**–d**5SICS**, d**TMO**–d**5SICS**, and d**FMO**–d**5SICS**, respectively. A color version is provided in Supporting Information.

tact in the major groove. The sum of these interactions provides favorable hydrophobic packing, but drives the nucleobase of d**MMO2**$_{18}$ out of planarity with dT_{17} and dG_{19}. Within the d**5SICS**$_7$–d**MMO2**$_{18}$ pair, the aromatic rings are oriented such that C4 of d**5SICS**$_7$ is positioned nearly directly over C3 of d**MMO2**$_{18}$ (Figure 4A).

We next used the structure of d**5SICS**$_7$–d**MMO2**$_{18}$ as a starting point to model the structures of the derivative base pairs in the same 12-mer duplex. Suitable parameters for the derivative nucleotides (d**DMO**, d**TMO**, and d**FMO**) were generated using DFT calculations (B3LYP/6-31G*),[29] and then d**MMO2**$_{18}$ was replaced and the resulting duplex was subjected to unconstrained minimization for up to 5000 steps in the Sander module of AMBER,[30] until the energy converged (Figure 4B–D). Like the parental unnatural base pair, each derivative base pair shows a similar level of interstrand intercalation. While the increased major groove bulk of d**FMO**$_{18}$ appears to introduce some additional local perturbations, none of the structures predict significant distortions relative to the structure of DNA containing the parental base pair. The minor groove interactions between the methoxy and sulfur groups are conserved in all of the structures. While the overall structure of the base pairs in the major groove is also conserved, with the *para* substituent of the d**MMO2** analogue stacking against the methyl/aromatic portion of d**5SICS**$_7$, the models reveal differences in the stacking interactions that result from derivatization. The *para*-methoxy group of d**DMO**$_{18}$ appears to rotate so that the methyl group packs against the methyl group of d**5SICS**$_7$ and the oxygen lone pairs are oriented into the major groove. The increased size and hydrophobicity of the sulfur substituent of d**TMO** appears to preclude packing of the methyl group with d**5SICS**$_7$, and instead the sulfur atom packs with d**5SICS**$_7$ and the hydrophobic methyl group is oriented into the major groove. In contrast to d**DMO**, the cyclic aryl ether bond of d**FMO** is unable to rotate and thus the lone pairs of the oxygen atom are forced toward the methyl group of d**5SICS**$_7$. Furthermore, packing with the flanking dC_6–dG_{19} pair isolates this oxygen and precludes it from potentially engaging in stabilizing interactions with water or metal ions within the major groove.

Discussion

The effort to expand the genetic alphabet is predicated on the availability of an unnatural base pair that is well replicated and transcribed, and preferably also suitable for modification such that it may be used to enzymatically produce site-specifically modified DNA and/or RNA. The data reveal that d**DMO**–d**5SICS** is better replicated by Kf than is the parental base pair, d**MMO2**–d**5SICS**. In the steady-state experiments, d**MMO2**TP insertion opposite d**5SICS** limits replication, and the *ortho*-methoxy group of d**DMO** increases the rate of this step, at least in sequence context I. In the opposite strand context, where increases in efficiency are less critical (as it is already very efficient), d**5SICS**TP is inserted opposite d**DMO** slightly less efficiently than it is opposite d**MMO2** in sequence context I, but slightly faster in context II. In fact, the insertion of d**5SICS**TP opposite d**DMO** in context II is the most efficient reported for an unnatural base pair. Moreover, in both sequence contexts examined, extension of d**DMO**–d**5SICS** is more efficient than that of d**MMO2**–d**5SICS** by approximately a factor of four, except in the case of the extension with d**5SICS** in the primer in sequence context II, in which both unnatural base pairs were extended with similar efficiencies. In addition, no mispairs between the unnatural or natural nucleotides and d**DMO** are synthesized more efficiently than those with d**MMO2**, and in fact, most are synthesized less efficiently. Finally, the mispairs with d**DMO** are also generally extended less efficiently than those with d**MMO2**, except for the mispairs with dA in sequence context I and dC in sequence context II. These individual steps combine to make d**DMO**–d**5SICS** replication significantly higher fidelity than d**MMO2**–d**5SICS** (Table 5).

The improved recognition of d**DMO**–d**5SICS** relative to the other unnatural base pairs, including d**MMO2**–d**5SICS**, is also apparent in the PCR data. Importantly, the efficiencies and fidelities of d**DMO**–d**5SICS** amplification appear to be sufficient for in vitro applications.[31] For example, d**DMO**–d**5SICS** appears to be uniquely suited for the site specific labeling of DNA (and possibly RNA[11]) within a format compatible with PCR (or transcription). Along with analogous modifications of (d)**5SICS**, this should allow the

Table 5. Minimum single step and overall replication fidelities.

Primer[a] (**Y**)	Template[a] (**X**)	Minimum synthesis fidelity[b]		Minimum extension fidelity[b]		Minimum replication fidelity[c]	
		Context I	Context II	Context I	Context II	Context I	Context II
d**5SICS**	d**MMO2**	390 (d**MMO2**, dA)	37 (d**MMO2**, dA)	0.56 (dC)	0.85 (dT)	7100 (dA)	5900 (dA)
d**MMO2**	d**5SICS**	2.8 (dG)	2.0 (d**5SICS**, dG)	4.8 (dT)	1.0 (dT)	130 (dT)	56 (dT)
d**5SICS**	d**DMO**	180 (dA, d**DMO**)	120 (dA)	2 (dC)	1.2 (dC)	4300 (dA)	23000 (dA)
d**DMO**	d**5SICS**	13 (dG)	2.1 (d**5SICS**, dG)	20 (dT)	4.8 (dT)	2600 (dT)	280 (dT)
d**5SICS**	d**TMO**	32 (dA)	–[d]	8.6 (dC)	–[d]	1100 (dA)	–[d]
d**TMO**	d**5SICS**	12 (dG)	–[d]	2.5 (dT)	–[d]	300 (dT)	–[d]
d**5SICS**	d**FMO**	0.69 (dA)	–[d]	1.3 (dA)	–[d]	0.90 (dA)	–[d]
d**FMO**	d**5SICS**	0.19 (dG)	–[d]	0.14 (dT)	–[d]	0.26 (dT)	–[d]

[a] Primer and template sequences correspond to contexts I and II (see Table 1 and Table 2). [b] Minimum synthesis and extension fidelity represents the ratio of second-order rate constants for the synthesis or extension of the correct unnatural pair to the most efficiently synthesized mispair (shown in parentheses). [c] Minimum replication fidelity corresponds to the product of the minimum fidelities for synthesis and extension (relative to the most efficiently replicated mispair, shown in parentheses). [d] Not measured.

site-specific modification of DNA and RNA with two different functional groups, which should be useful for variety of in vitro applications, including SELEX with an expanded genetic alphabet,[32] as well as biophysical studies that rely on the modification of DNA with multiple biophysical probes.

The mechanism by which DNA polymerases replicate predominantly hydrophobic unnatural base pairs is of great interest for designing better base pairs, as well as for understanding the range of activities possible with these important enzymes. It has been suggested that shape complementarity is important;[2–4] however, it is critical to define in what context it is manifest (i.e., the mode of pairing). Shape complementarity is usually evoked within a natural, Watson–Crick-like mode of pairing, where two in-plane nucleobases interact in an edge-on manner. Each natural base pair thus adopts a similar shape that is thought to be uniquely well accommodated by DNA polymerases.[2–4] In contrast, the model proposed here (Figure 2) evokes a different mode of base pairing, where instead of interacting edge-to-edge, where little to no stabilization is available, the nucleobases partially interstrand intercalate during base pair synthesis, which is likely driven by their high stacking potential. However, extension of the nascent unnatural base pair requires deintercalation to position the primer terminus 3′-OH appropriately for continued elongation. While deintercalation is favored by a stabilizing hydrogen-bond between the polymerase and the *ortho* substituents of the nucleobase analogues,[33–38] the model emphasizes the balance of intercalation propensity that must be possessed by the pairing nucleobases: they must intercalate sufficiently for synthesis, but not so much that extension is inhibited. This model nicely explains a large body of previously reported kinetic[5–10] and structural data.[12]

The solution structure of the parental d**MMO2**–d**5SICS** pair, as well as the derivative model structures of the d**DMO**–d**5SICS**, d**TMO**–d**5SICS**, and d**FMO**–d**5SICS** pairs in duplex DNA supports the intercalative model of replication (Figure 2). The structures clearly reveal that the nucleotides are accommodated within a B-form duplex, adopt *anti*-orientations about their glycosidic bonds, and importantly, pair in an intercalative manner. The data further reveal that the stacking interface between the nucleobases is comprised of the methyl group and proximal portion of the associated aromatic ring of d**5SICS** and the *para* substituent of d**MMO2** or a d**MMO2** analogue. It should be emphasized that the structures suggest that the unnatural nucleobase analogues only partially intercalate, they do not fully insert into the opposite strand due to their size and the constraints imposed by the duplex (nonetheless, we refer to the interaction as intercalation for simplicity). Importantly, it is clear that the various substituents examined are predicted to be positioned within the stacking interface between the unnatural nucleobases, which accounts for their effects on replication. It should also be emphasized that the structural data is based on the analogues embedded within a duplex, and not at a primer terminus bound to a DNA polymerase. However, the fact that at least some of the specific interactions involved in base pair recognition are inherent to the base pair and not dependent on the polymerase supports the interpretation of the structure in terms of replication.

The structural models highlight the importance of how the different substituents affect the partitioning of the unnatural nucleobases between intercalated and deintercalated states, which appear to be required for synthesis and extension, respectively. In the intercalated state, the major groove substituents form a central part of the nucleobase packing interface, but upon deintercalation, these substituents are more solvent exposed in a more traditional-like major groove. The models suggest that the more efficient replication of d**DMO**–d**5SICS** results from an optimized balance of forces governing the stability of the intercalated and deintercalated states. Synthesis is likely favored by optimized packing interactions between the major groove methyl groups of d**DMO** and d**5SICS**. In addition, the structure adopted by d**DMO** orients the oxygen lone pairs toward the major groove, where upon deintercalation, they may engage in stabilizing interactions with proximal water molecules and/or metals, thus favoring unnatural base pair extension. While anisole is generally not a strong metal ligand or hydrogen-bond acceptor due to electron delocalization, both interactions are favored when the conjugation is disrupted by rotation,[39–42] as is observed in the modeled structure of d**DMO**–d**5SICS**. The increased substituent size of d**TMO** appears to induce subtle structural changes without any significant affect on replication. In contrast, the cyclic structure of d**FMO** appears to force the oxygen lone pairs directly into the hydrophobic interface between the nucleobases, which is likely destabilizing.[43–45] Moreover, if the furanyl oxygen is solvated as the free triphosphate,[46,47] then this stabilizing solvation will be lost upon insertion without being replaced with any other favorable interactions. Moreover, deintercalation is expected to force the hydrophobic methines further into the hydrophilic major groove, which is likely further destabilizing. Thus, with the aid of the structural models, the intercalative mechanism nicely explains the relatively large effects of the modifications on unnatural base pair synthesis and extension.

Conclusion

We have identified d**DMO**–d**5SICS** as an unnatural base pair that is better replicated than the parental d**MMO2**–d**5SICS** pair. In addition structural studies support an intercalative model of replication, as previously proposed based on kinetic data[9] and help to explain the observed effects of the various modifications. The intercalative mode of pairing is likely not limited to the analogues examined in the present work. Indeed, it is similar to that observed in the DNA zipper motif, where alternating natural nucleobases are interdigitated, as opposed to interacting in an edge-on manner.[48–56] Moreover, a similar mode of pairing has been observed previously by our group,[12] as well as by the Leumann group[57] with unnatural nucleotides bearing large aro-

matic nucleobase analogues. However, in these cases the extended aromatic surface area of the nucleobase analogues likely makes intercalation or extrusion from the duplex the only viable options. The intercalative mode of pairing observed between d**5SICS** and the d**MMO2** derivatives occurs despite their potential in-plane accommodation. It is likely that such an intercalative mode of pairing is common to all analogues that are incapable of engaging in stabilizing edge-on interactions. It is also possible that some mispairs between natural nucleotides may be synthesized in a similar manner. Regardless of its potential contribution to the replication of natural DNA, the elucidation of the intercalative mode of pairing should prove invaluable for the further optimization of the unnatural base pairs as well as for our understanding of the potential substrate repertoires of DNA polymerases in general.

Experimental Section

Synthetic methods: d**FMO** and d**TMO** were synthesized as described in Supporting Information and d**DMO** was synthesized as described previously.[7] Briefly, the corresponding arylbromides were lithiated and coupled to 3,5-di-(*tert*-butyldimethylsilyloxy)-2-deoxy-erythropentofuranose (Scheme S1 in the Supporting Information). After deprotection with TBAF, anomeric mixtures of nucleosides were obtained, and the β-anomer was purified by column chromatography. Nucleosides were converted to triphosphates by $POCl_3$ treatment in the presence of proton sponge, followed by reaction with tributylammonium pyrophosphate. Phosphoramidites of **FMO** and **TMO** were obtained from the free nucleosides by 5′ DMT protection and reaction with 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite. Oligonucleotides were synthesized by standard solid phase synthesis on controlled pore glass supports. Experimental details together with characterization of all nucleosides, phosphoramidites, oligonucleotides, and triphosphates are provided in the Supporting Information.

Kinetic assays: Primer oligonucleotides were 5′-radiolabeled with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP (Amersham Biosciences) and annealed to template oligonucleotides by heating to 95 °C followed by slow cooling. Reactions were initiated by adding a solution of 2×dNTP solution (5 μL) to a solution containing polymerase (0.15–1.23 nM) and primer template (40 nM) in reaction buffer (5 μL, see Supporting Information for details). After incubation at 25 °C (Kf) or 50 °C (Taq and Vent) for 3–10 min the reactions were quenched with 20 μL of loading dye (95% formamide, 20 mM EDTA, 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 software (Bio-Rad). Plots of k_{obs} versus triphosphate concentration were fit to a 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 K_M and k_{cat} were determined (see Tables S1–S4 in the Supporting Information). Representative raw kinetic data are shown in Figure S1 in the Supporting Information.

PCR amplification: DNA duplexes used as templates in PCR amplification reactions were synthesized as described previously.[15] PCR amplification of duplexes D1-D6 (see Table 4 in the main text for details and Supporting Information for sequences) was carried out in 1×ThermoPol reaction buffer (New England Biolabs) with the following modifications: $MgSO_4$ (6.0 mM), natural dNTP (0.6 mM), unnatural triphosphate (0.4 mM), primers (1 μM each, see Supporting Information for sequences), and 0.03 unit μL^{-1} of DeepVent (exo$^+$) in an iCycler Thermal Cycler (Bio-Rad) with a total volume of 25 μL under the following thermal cycling conditions: 94 °C, 30 s; 48 °C, 30 s; 65 °C, 8 min, 14 cycles. Upon completion, PCR products were purified utilizing the PureLink™ 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 the fidelity of unnatural base pair replication (see Supporting Information and reference[15] for details).

Structural studies: Lyophilized duplex DNA containing the d**MMO2**–d**5SICS** unnatural base pair was dissolved in buffer containing 10 mM sodium phosphate, pH 7.0, 100 mM NaCl, and 0.1 mM EDTA in 99.99% D_2O to a final analyte concentration of 2 mM. All NMR spectra were acquired at 25 °C to resolve as much cross peak overlap as possible on a Varian Inova 500 MHz spectrometer. Proton resonance assignments were made according to established procedures. NOESY spectra with a mixing time of 300 ms were collected with a spectral width of 5913 Hz, 2048 complex points in t_2 and 512 t_1 increments (zero filled to 2048 on processing); for each t_1 value 64 scans were averaged using a recycle delay of 2 s. The approach for computing the structure for the d**5SICS**–d**MMO2** duplex was patterned as described.[28,58] Forcefield parameters for d**5SICS** and d**MMO2** were calculated using Gaussian 98.[29] All energy minimization and restrained molecular dynamics (rMD) calculations were performed with the SANDER module of AMBER 9.[30] A total of 373 constraints were applied (including Watson–Crick hydrogen bonding constraints, 346 NMR-derived distance restraints and torsion restraints for each sugar moiety) during rMD. Structures of duplexes containing d**5SICS**:d**DMO**, d**5SICS**:d**TMO**, and d**5SICS**:d**FMO** shown in Figure 4 were modeled from the NMR determined d**5SICS**:d**MMO2** structure. Briefly, each nucleobase (**DMO**, **TMO**, **FMO**) was subjected to DFT calculations to obtain charge distribution and geometrical parameters. These analogues were used to replace d**MMO2** in the NMR structure and each duplex was then minimized (unconstrained) for up to 5000 steps in the Sander module of AMBER,[30] until the energy converged.

Acknowledgements

Funding for this work was provided by the National Institutes of Health (GM060005). The NMR facility at USD was established with a grant from the NSF-MRI program (0417731).

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
[2] M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10493–10495.
[3] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950–954.
[4] A. T. Krueger, E. T. Kool, *Curr. Opin. Chem. Biol.* **2007**, *11*, 588–594.
[5] D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586.
[6] C. Yu, A. A. Henry, F. E. Romesberg, P. G. Schultz, *Angew. Chem.* **2002**, *114*, 3997–4000; *Angew. Chem. Int. Ed.* **2002**, *41*, 3841–3844.
[7] S. Matsuda, A. M. Leconte, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 5551–5557.
[8] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343.
[9] Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 3246–3252.
[10] Y. J. Seo, F. E. Romesberg, *ChemBioChem* **2009**, *10*, 2394–2400.
[11] Y. J. Seo, S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 5046–5047.
[12] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 10466–10473.
[13] G. T. Hwang, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 14872–14882.
[14] A. M. Leconte, F. E. Romesberg in *Protein Engineering* (Ed.: C. K. a. U. L. RajBhandary), Springer, Berlin, **2009**, pp. 291–314.

[15] D. A. Malyshev, Y. J. Seo, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 14620–14621.
[16] T. Mitsui, A. Kitamura, M. Kimoto, T. To, A. Sato, I. Hirao, S. Yokoyama, *J. Am. Chem. Soc.* **2003**, *125*, 5298–5307.
[17] I. Hirao, *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627.
[18] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729–735.
[19] I. Hirao, T. Mitsui, M. Kimoto, S. Yokoyama, *J. Am. Chem. Soc.* **2007**, *129*, 15549–15555.
[20] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2008**, *37*, e14.
[21] M. Chiaramonte, C. L. Moore, K. Kincaid, R. D. Kuchta, *Biochemistry* **2003**, *42*, 10472–10481.
[22] K. Kincaid, R. D. Kuchta, *Nucleic Acids Res.* **2006**, *34*.
[23] Unnatural triphosphate insertion opposite an unnatural base in the template obviously results in the synthesis of the unnatural base pair, thus unnatural triphosphate insertion and unnatural base-pair synthesis in a specific primer:template strand context are used interchangeably throughout this manuscript.
[24] M. Hocek, M. Fojta, *Org. Biomol. Chem.* **2008**, *6*, 2233–2241.
[25] L. L. Hansen, J. Justensen in *PCR Primer: A Laboratory Manual* (Eds.: C. W. Diffenbach, G. S. Dveksler), Cold Spring Harbor Laboratory Press, Woodbury, **2003**, pp. 226–235.
[26] B. Arezi, W. Xing, J. A. Sorge, H. H. Hogrefe, *Anal. Biochem.* **2003**, *321*, 226–235.
[27] D. R. Hare, D. E. Wemmer, S. H. Chou, G. Drobny, B. R. Reid, *J. Mol. Biol.* **1983**, *171*, 319–336.
[28] J. A. Smith, L. Gomez-Paloma, D. A. Case, W. J. Chazin, *Magnet. Res. Chem.* **1996**, *34*, S147–S155.
[29] Gaussian 98, Revision A.7, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, **1998**.
[30] AMBER 9, D. A. Case, T. A. Darden, T. E. I. Cheatham, C. L. Simmerling, R. E. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. R. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, P. A. Kollman, University of California, San Francisco, **2006**.
[31] S. Klussmann, *The Aptamer Handbook: Functional Oligonucleotides and Their Applications*, Wiley-VCH, Weinheim, **2006**.
[32] A. D. Keefe, S. T. Cload, *Curr. Opin. Chem. Biol.* **2008**, *12*, 448–456.
[33] J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324.
[34] J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 1001–1007.
[35] Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225–1233.
[36] T. E. Spratt, *Biochemistry* **2001**, *40*, 2647–2652.
[37] A. S. Meyer, M. Blandino, T. E. Spratt, *J. Biol. Chem.* **2004**, *279*, 33043–33046.
[38] M. D. McCain, A. S. Meyer, S. S. Schultz, A. Glekas, T. E. Spratt, *Biochemistry* **2005**, *44*, 5647–5659.
[39] I. Nobeli, S. L. Yeoh, S. L. Price, R. Taylor, *Chem. Phys. Lett.* **1997**, *280*, 196–202.
[40] B. Reimann, K. Buchhold, H. D. Barth, B. Brutschy, P. Tarakeshwar, K. S. Kim, *J. Chem. Phys.* **2002**, *117*, 8805–8822.
[41] M. Becucci, G. Pietraperzia, M. Pasquini, G. Piani, A. Zoppi, R. Chelli, E. Castellucci, W. Demtroeder, *J. Chem. Phys.* **2004**, *120*, 5601–5607.
[42] J. W. Ribblett, W. E. Sinclair, D. R. Borst, J. T. Yi, D. W. Pratt, *J. Phys. Chem. A* **2006**, *110*, 1478–1483.
[43] R. A. Friedman, B. Honig, *Biophys. J.* **1995**, *69*, 1528–1535.
[44] K. S. C. Reid, P. F. Lindley, J. M. Thornton, *FEBS Lett.* **1985**, *190*, 209.
[45] M. Egli, S. Sarkhel, *Acc. Chem. Res.* **2007**, *40*, 197–205.
[46] C. Glidewell, C. M. Zakaria, G. Ferguson, *Acta Crystallogr. Sect. C* **1996**, *52*, 1305–1309.
[47] T. Nakanaga, F. Ito, *J. Phys. Chem. A* **1999**, *103*, 5440–5445.
[48] S. H. Chou, L. M. Zhu, B. R. Reid, *J. Mol. Biol.* **1994**, *244*, 259–268.
[49] W. Shepard, W. B. T. Cruse, R. Fourme, E. de La Fortelle, T. Prange, *Structure* **1998**, *6*, 849–861.
[50] S. H. Chou, Y. Y. Tseng, *J. Mol. Biol.* **1999**, *285*, 41–48.
[51] N. A. Spackova, I. Berger, J. Sponer, *J. Am. Chem. Soc.* **2000**, *122*, 7564–7572.
[52] S. H. Chou, K. H. Chin, *J. Mol. Biol.* **2001**, *314*, 139–152.
[53] S. H. Chou, K. H. Chin, *J. Mol. Biol.* **2001**, *312*, 753–768.
[54] S. H. Chou, K. H. Chin, A. H. J. Wang, *Nucleic Acids Res.* **2003**, *31*, 2461–2474.
[55] J. Kondo, S.-i. Umeda, K. Fujita, T. Sunami, A. Takenaka, *J. Synchrotron Radiat.* **2004**, *11*, 117–120.
[56] T. Sunami, J. Kondo, I. Hirao, K. Watanabe, K.-i. Miura, A. Takenaka, *Acta Crystallogr. Sect. D* **2003**, *60*, 90–96.
[57] Z. Johar, A. Zahn, C. J. Leumann, B. Jaun, *Chem. Eur. J.* **2008**, *14*, 1080–1086.
[58] D. A. Pfaff, K. M. Clarke, T. A. Parr, J. M. Cole, B. H. Geierstanger, D. C. Tahmassebi, T. J. Dwyer, *J. Am. Chem. Soc.* **2008**, *130*, 4869–4878.

Received: April 14, 2010
Published online: September 21, 2010

