Nucleic Acids Research, 2006, Vol. 34, No. 7 ***2037–2045***
doi:10.1093/nar/gkl049

Substituent effects on the pairing and polymerase recognition of simple unnatural base pairs

Gil Tae Hwang and Floyd E. Romesberg*

Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA

Received December 5, 2005; Revised December 27, 2005; Accepted February 22, 2006

ABSTRACT

As part of an effort to develop stable and replicable unnatural base pairs, we have evaluated a large number of unnatural nucleotides with predominantly hydrophobic nucleobases. Despite its limited aromatic surface area, a nucleobase analog scaffold that has emerged as being especially promising is the simple phenyl ring. Modifications of this scaffold with methyl and fluoro groups have been shown to impact base pair stability and polymerase recognition, suggesting that nucleobase shape, hydrophobicity and electrostatics are important. To further explore the impact of heteroatom substitution within this nucleobase scaffold, we report the synthesis, stability and polymerase recognition of nucleoside analogs bearing single bromo- or cyano-derivatized phenyl rings. Both modifications are found to generally stabilize base pair formation to a greater extent than methyl or fluoro substitution. Moreover, polymerase recognition of the unnatural base pairs is found to be very sensitive to both the position and nature of the heteroatom substituent. The results help identify the determinants of base pair stability and efficient replication and should contribute to the effort to develop stable and replicable unnatural base pairs.

INTRODUCTION

With the determination of the structure of the DNA double helix in 1953, it was assumed that the complementary hydrogen-bonding (H-bonding) of the natural nucleobases was the origin of selective duplex stability and replication. However, with the discovery that DNA replication does not absolutely require H-bond complementarity (1–4), it became apparent that this model is at least incomplete. The ability to use forces other than H-bonding to control the selective pairing of nucleobases during replication dramatically increases the range of possible unnatural base pairs. Such unnatural base pairs would have immediate biotechnology applications (5) and would also facilitate the longer term goal of expanding the genetic code (6,7). Toward the goal of developing unnatural base pairs, we have synthesized and characterized a large number of nucleotides bearing large and predominantly hydrophobic nucleobases, based on the isocarbostyril, napthyl and indole scaffolds, whose pairing should be mediated by hydrophobic interactions (Figure 1) (8–16). (Although many of the unnatural nucleobases are not actually basic, we refer to them as nucleobase analogs for simplicity.) We have systematically evaluated the stability of the resulting base pairs as well as analogs whose shape and electrostatic properties have been altered by methyl group and heteroatom substitution (11,15,17,18). From these studies, we have identified several self pairs (formed by pairing of identical analogs) and heteropairs (formed by pairing two different analogs) that rival natural base pairs in terms of thermal stability and selectivity against mispairing, despite not being able to form interbase H-bonds.

In addition to stable and selective pairing in duplex DNA, candidate unnatural base pairs must also be efficiently and selectively replicated by DNA polymerases. We have reported the kinetic analysis of how each of the nucleobases are recognized by the exonuclease deficient DNA polymerase I from *Escherichia coli* (Kf) and identified several self pairs and heteropairs that are efficiently synthesized (i.e. by insertion of the unnatural triphosphate opposite its cognate base in the template) (8–11,13,15–18). However, both selectivity against insertion of the natural triphosphates and unnatural base pair extension (i.e. continued elongation of the newly synthesized primer terminus) remained challenging. While large aromatic surface area increases the stability of the unnatural base pairs, it may also compromise synthesis fidelity and extension. This hypothesis is based on structural work suggesting that in duplex DNA, pairs formed between analogs with extended aromatic surface area interact via face-packing, where one analog intercalates into the opposite strand between its pairing nucleobase and an adjacent nucleobase (19,20). Such an interaction may not be optimal for discrimination against the

*To whom correspondence should be addressed. Tel: +1 858 784 7291; Fax: +1 858 784 7472; Email floyd@scripps.edu

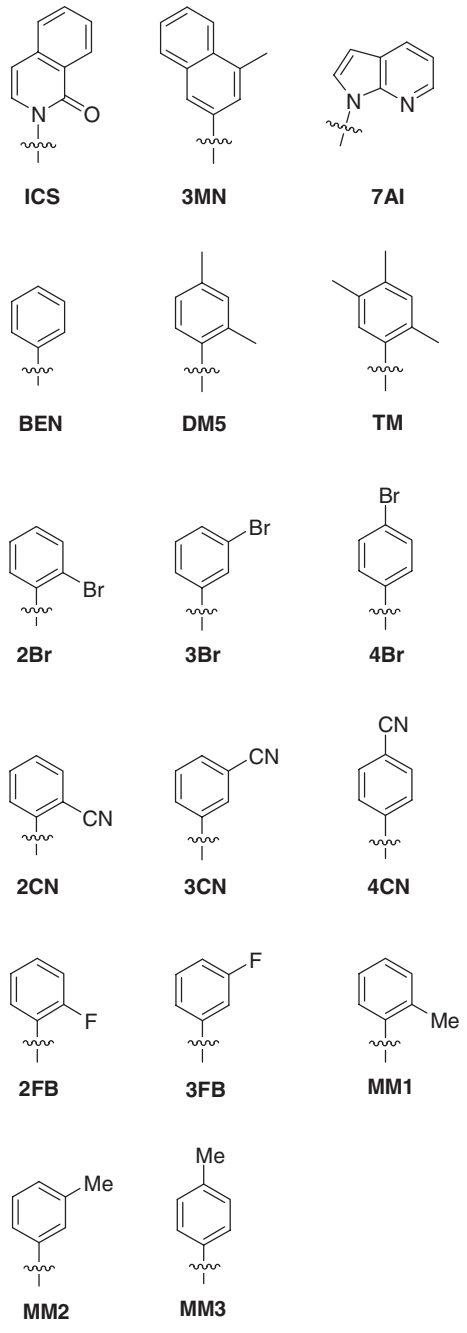


Figure 1. Unnatural nucleobases. Sugar and phosphate backbone have been omitted for clarity.

natural nucleobases during base pair synthesis, nor at the primer terminus where it may result in a structure that is not optimal for continued primer extension.

We thus became interested in examining nucleobase analogs that lack extended aromatic surface area but are optimized to interact edge-on in a fashion analogous to a natural Watson–Crick base pair. Surprisingly, with suitably chosen substituents, we have found that nucleotides bearing simple phenyl rings may form stable and selective pairs in duplex DNA and are also synthesized with reasonable efficiency and fidelity, despite having no H-bonding functionality and relatively little aromatic surface area. In addition, once synthesized at a primer terminus, several of the pairs are extended by Kf with efficiencies significantly greater than the pairs formed between the larger nucleobases. For example, the **3FB**, **DM5** and **TM** self pairs (Figure 1), identified during a systematic analysis of methyl and fluoro-derivatized phenyl ring nucleobase scaffolds are both synthesized and extended with reasonable efficiency by Kf [S. Matsuda and F. E. Romesberg, manuscript accepted; (18,21)]. These self pairs presumably interact in an edge-on manner and their more efficient extension suggests that the resulting interbase interface provides a primer terminus structure that is better recognized by DNA polymerases. However, discrimination against dATP and further optimization of the extension rates remain important for the development of these unnatural base pairs. Because adenine is the most hydrophobic of the natural nucleobases and the heteroatom-derivatized **3FB** self pair is extended about an order of magnitude more efficiently than either **DM5** or **TM**, we became interested further examining nucleobase hydrophobicity, dipole moment and polarizability.

To further explore base pair stability, discrimination against the natural dNTPs, and extension as a function of heteroatom substitution, we now report the synthesis and initial characterization of two series of nucleotide analogs bearing different heteroatom substituents. Like **3FB**, **DM5** and **TM**, these *C*-glycosides bear simple phenyl nucleobase analogs, but are derivatized with all possible single bromo or cyano substitutions. The substituents were designed to vary the size, dipole moment and polarizability of the nucleobase and for comparison with the analogous fluoro- and methyl-derivatized analogs. The data reveal that base pair stability and extension are sensitive to both the nature and the position of heteroatom substitution. When combined with the previous studies, these results suggest that the **3FB** self pair is efficiently extended due to a combination of steric and electrostatic effects, and that it might be optimized by modification at the 2- and/or 4-positions. These results should be useful for designing nucleobase analogs with a wide range of biotechnology applications as well as analogs with which to expand the genetic alphabet.

MATERIALS AND METHODS

Materials

Chemical reagents were purchased from Sigma–Aldrich and used without further purification, unless otherwise stated. All reagents for oligonucleotide synthesis were purchased from Glen Research. Oligonucleotides were synthesized using an Applied Biosystems Inc. 392 DNA/RNA synthesizer and purified using standard conditions. Concentrations were determined by ultraviolet (UV) absorption. T4 polynucleotide kinase and Klenow fragment exo- were purchased from New England Biolabs. [γ-^{33}P]ATP was purchased from Amersham Biosciences.

Thermal stability

The unnatural and natural nucleosides were incorporated into the complementary oligonucleotides d(GCGTAC**X**-CATGCG) and d(CGCATG**Y**GTACGC) at the positions labeled **X** and **Y**. For UV melting experiments the sample

[3 μM oligonucleotide, 10 mM PIPES buffer (pH 7.0), 100 mM NaCl and 10 mM $MgCl_2$] absorption was monitored at 260 nm from 20 to 80°C at a heating rate of 0.5°C per min, using a Cary 300 Bio UV/Vis spectrophotometer. Melting temperatures were determined from the derivative method using the Cary Win UV thermal application software.

Steady-state kinetics

The unnatural nucleobases were evaluated as substrates for Kf by measuring initial rates at which a [γ-^{33}P]-labeled primer, d(TAATACGACTCACTATAGGGAGAX) annealed to the 45mer template, d(CGCTAGGACGGCATTGGATCGYTCT-CCCTATAGTGAGTCGTATTA), was extended with varying concentrations of natural or unnatural nucleoside triphosphates. Each reaction included 40 nM primer-templates, 0.3–1.2 nM enzyme, 50 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT and 50 μg/ml acetylated BSA. The reactions were initiated by adding a 5 μl 2× dNTP solution to a 5 μl solution containing the polymerase and primer-template and incubating at 25°C for 5–12 min, and were then quenched with the addition of 20 μl of loading dye (95% formamide, 20 mM EDTA and sufficient amounts of bromophenol blue and xylene cyanole). The reactions were analyzed by polyacrylamide gel electrophoresis and a Phosphorimager (Molecular Dynamics) was used to quantify gel band intensities corresponding to the extended and unextended primer. The measured velocities were plotted against the concentration of dNTP and fit to the Michaelis–Menten equation to determine V_{max} and K_M. k_{cat} was determined from V_{max} by normalizing by the total enzyme concentration.

RESULTS

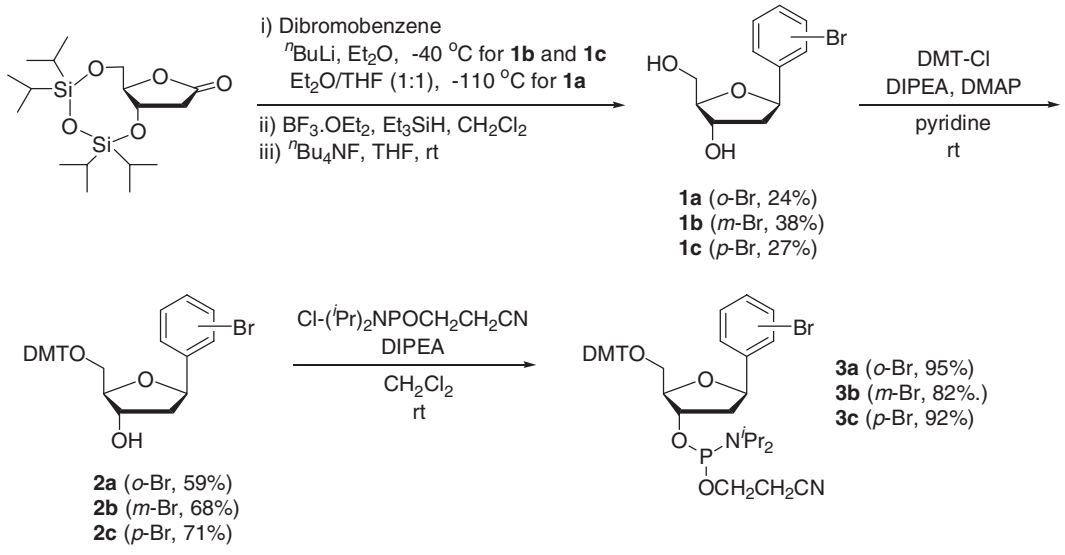
Design and synthesis of nucleobase analogs

To more systematically probe the contribution of hydrophobicity and electrostatics to the properties of the unnatural base pairs, we extended our studies of methyl and fluoro substituents to include bromo and cyano substituents (Figure 1). The benzene, toluene, fluorobenzene, bromobenzene and benzonitrile nucleobase analogs differ in terms of size, hydrophobicity, dipole moment and polarizability. The hydrophobicity of the nucleobase moieties increases in the order benzonitrile < benzene < fluorobenzene < toluene < bromobenzene (22,23); while the dipole moment and polarizability increase approximately in the order, benzene < toluene < bromobenzene < fluorobenzene < benzonitrile (24,25). The parental benzene scaffold is the smallest nucleobase analog, followed by fluorobenzene and bromobenzene, with toluene and benzonitrile being the largest. The varying contributions of nucleobase size, hydrophobicity, dipole moment and polarizability may thus be examined by comparing the stabilities and polymerase recognition of the different base pairs.

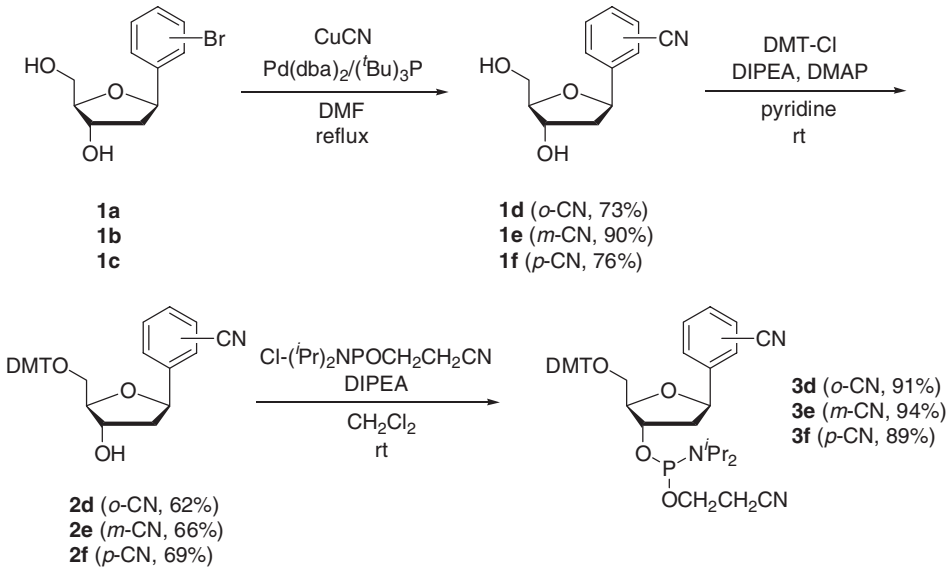
The bromo- and cyano-derivatized nucleotide analogs were synthesized as shown in Schemes 1 and 2. Briefly, tetraisopropyl disiloxane-protected nucleosides were synthesized by aryllithium coupling to the disiloxane-protected lactone, based on modified literature procedures (26,27). In the case of the 1,2-dibromobenzene analog, the reaction was carried out in a mixture of diethyl ether and THF at −110°C, above this temperature 1,2-bromophenyllithium elimination led to the formation of benzyne (28). Treatment with triethylsilane and strong Lewis acid resulted in reduction to give the protected aryl *C*-nucleoside in moderate yield. The β-anomer was the major product (26,27) which was purified from the α-anomer by flash column chromatography. Deprotection was achieved by treatment with *n*-TBAF. Cyano-derivatized analogs analogues **1d–f** were synthesized from **1a–c** using Pd-catalyzed cyanation of the aryl bromide (Scheme 2) (29). The assignment of β-stereochemistry at C1′ for all nucleosides was based on COSY and 1-D NOE experiments. The phosphoramidites for solid-phase DNA synthesis were obtained by standard dimethoxytritylation and phosphitylation.

Thermal stability and selectivity of unnatural base pairs

To evaluate the thermodynamic stability and selectivity of the unnatural base pairs in duplex DNA, all unnatural and



Scheme 1. Synthesis of bromo-derivatized nucleosides.



Scheme 2. Synthesis of cyano-derivatized nucleosides.

natural nucleosides were incorporated into the complementary oligonucleotides d(GCGTAC**X**CATGCG) and d(CGCATG**Y**GTACGC) at the positions labeled **X** and **Y**. These oligonucleotides were chosen in order to examine the effect of sequence context by comparing the effects of flanking pyrimidines (position **X**) and flanking purines (position **Y**). The unnatural nucleotides were examined both as self pairs (**X** = **Y**) and as heteropairs (**X** ≠ **Y**). The melting temperature (T_m) of each duplex was determined by thermal denaturation experiments (Table 1) and used as a measure of base pair stability. For comparison, the T_m for a duplex containing a natural base pair (**X**:**Y** = dA:dT) is 59.2°C.

Generally, the unnatural base pairs are surprisingly stable, showing T_m values that are only 3 to 6°C less stable than a dA:dT pair in the same sequence context, and significantly more stable than a typical mispair between two natural nucleotides (Table 1). However, as with other similarly substituted analogs, the T_m values show only a modest dependence on the specific substitution pattern of the nucleobase analog, with all self pairs and heteropairs showing T_m values between 53.0 and 56.0°C. The most stable pairs identified were the **2Br** self pair and the **2CN**:**2Br** and **4CN**:**3Br** heteropairs. Perhaps the pairs formed between 2-substituted analogs are stable due to efficient packing in the minor groove, while the 2°C decrease in stability of the **2CN** self pair suggests that two larger cyano groups are not well accommodated.

Generally, the stability of the unnatural base pairs show little dependence on sequence context, with similar stabilities regardless of being flanked by purines or pyrimidines. The most notable exception was the pair formed between **4CN** and **3Br**. When **3Br** is flanked by purines, it is one of the most stable base pair of the series (T_m = 56°C), but when flanked by pyrimidines it is one of the least stable (T_m = 53.1°C). This suggests that the polarizable bromine at the 3-position is well oriented to interact favorably with the aromatic surface area of the purine, but not pyrimidine nucleobases.

To examine the thermal orthogonality of the unnatural base pairs, we determined the duplex melting temperatures with the unnatural bases incorporated at position **X** and the natural nucleotides incorporated at position **Y** (Table 1). All of the mispairs are significantly less stable than even the least stable unnatural pair. This demonstrates that despite the only modest ability of these small nucleobase analogs to differentiate amongst themselves, they differentiate strongly between the hydrophobic nucleobases and the natural nucleobases. This

Table 1. T_m values for duplexes containing unnatural base pairs[a]

5′-d(GCGTAC**X**CATGCG)
3′-d(CGCATG**Y**GTACGC)

X	Y	T_m (°C)	X	Y	T_m (°C)
2Br	**2Br**	56.0	**2CN**	**2Br**	56.0
	3Br	55.1		**3Br**	53.0
	4Br	55.0		**4Br**	54.0
	2CN	55.0		**2CN**	54.1
	3CN	54.0		**3CN**	53.1
	4CN	55.0		**4CN**	54.0
	dC	46.0		dC	45.0
	dG	51.0		dG	49.0
	dA	51.0		dA	50.0
	dT	49.0		dT	47.1
3Br	**2Br**	54.1	**3CN**	**2Br**	54.0
	3Br	54.0		**3Br**	54.0
	4Br	54.1		**4Br**	55.1
	2CN	54.1		**2CN**	54.1
	3CN	53.1		**3CN**	54.0
	4CN	53.1		**4CN**	54.0
	dC	48.0		dC	48.0
	dG	50.0		dG	50.0
	dA	50.0		dA	51.1
	dT	49.0		dT	50.1
4Br	**2Br**	55.5	**4CN**	**2Br**	55.0
	3Br	54.1		**3Br**	56.0
	4Br	54.1		**4Br**	54.0
	2CN	55.1		**2CN**	54.0
	3CN	55.1		**3CN**	55.0
	4CN	53.1		**4CN**	54.0
	dC	51.1		dC	50.0
	dG	51.0		dG	50.0
	dA	51.0		dA	51.0
	dT	50.0		dT	49.0

[a]Uncertainty in values is less than 0.1°C. See Materials and Methods section for details.

likely results, at least in part, from the forced desolvation of the more hydrophilic natural bases.

The stabilities of the mispairs are also dependent on the specific substitution pattern of the unnatural nucleotide. The stabilities of the mispairs formed with the 4-substituted derivatives are the least sensitive to the specific natural nucleotide, ranging by only 1–2°C. The 2- and 3-substituted analogs show greater discrimination, with the purine mispairs being the most stable, followed by the mispair with dT, and finally by the mispair with dC, which is 1 to 5°C less stable than the other mispairs. The discrimination is greatest with the 2-substituted analogs, where mispair stability ranged over 5°C. While the stabilities generally parallel the hydrophobicity of the natural nucleobases, they also reflect the aromatic surface area of the natural base. In addition, the greater discrimination of **2Br** and **2CN** suggests that a substituent at the 2-position locks the nucleobase analog into an *anti*-orientation, which forces the substituent into the interbase interface, while the other analogs are free to rotate about the *C*-glycosidic bond and present the unsubstituted side of the phenyl ring to the pairing base. Nonetheless, mispairs with the natural nucleotides are significantly less stable than the unnatural pairs in all cases, demonstrating the thermal orthogonality inherent to even small hydrophobic nucleobases, as has been seen with other predominantly hydrophobic analogs (18,21).

Polymerase-mediated mispair synthesis

The nucleotides of any unnatural base pair must be not only thermally orthogonal, but also kinetically orthogonal. To characterize the kinetic discrimination against mispairing, we incorporated the unnatural analogs into template DNA and examined their ability to direct Kf to incorporate the natural dNTPs. In each case, we characterized the binding of the natural dNTP (K_M) and the turnover of the resulting ternary complex (k_{cat}). The ratio, k_{cat}/K_M, is the second order rate constant, and is a measure of catalytic efficiency. Generally, as has been observed with other predominantly hydrophobic nucleobases, dATP is inserted most efficiently opposite the analog in the template. However, the efficiency of dATP insertion is significantly less than that observed with other analogs, ranging from 7.8×10^3 $M^{-1}min^{-1}$ (for **2Br**) to 8.8×10^4 $M^{-1}min^{-1}$ (for **2CN**). The analogs that have bromo or cyano groups at either the 3- or 4-position template dATP insertion with a second order rate constant that varies only between 2×10^4 and 4×10^4 $M^{-1}min^{-1}$, while the same substituents at the 2-position have a more variable effect (Table 2).

None of the other natural dNTPs are inserted with more than moderate efficiencies, but the analogs do show interesting differences in discrimination. **2Br** templates the addition of dGTP, dTTP and dCTP relatively efficiently, with second order rate constants within 3-fold of that for dATP insertion. However, **2CN** directs the insertion of dGTP 7-fold less efficiently than dATP, dTTP an order of magnitude less efficiently than dGTP, and does not direct the insertion of dCTP at all. **3Br** does not direct Kf to insert any of the natural dNTPs with detectable efficiencies, while a cyano group at the same position results in an analog that templates dTTP insertion with modest efficiency, and to an even lesser extent dCTP, but not dGTP. **4Br** templates the insertion of the pyrimidine triphosphates with modest efficiencies of $\sim 3 \times 10^4$ $M^{-1}min^{-1}$, but not dGTP at a detectable rate, and the analog with a cyano group at the same position does not direct the insertion of any of the three other dNTPs at detectable rates.

Recognition of unnatural base pairs at a primer terminus by Kf DNA polymerase

To determine how bromo- and cyano-modifications impact unnatural base pair recognition by a DNA polymerase, we examined the efficiencies with which Kf extended primers that terminated with the unnatural base pairs (**X**:**Y**, where **X** is in the primer and **Y** is in the template) by incorporation of the next correct triphosphate, dCTP (Table 3). Thirty one of the thirty six unnatural base pairs are not extended with rates that are sufficient to detect ($<10^3$ $M^{-1}min^{-1}$). Of the five that were extended, three have **2CN** in the template strand, paired opposite either **4Br**, **4CN** or **3CN** in the primer strand. The **4Br**:**2CN** heteropair is most efficiently extended with a second order rate constant of 3.0×10^4 $M^{-1}min^{-1}$. The **4CN**:**2CN** and **3CN**:**2CN** pairs are extended slightly less efficiently, with virtually identical rate constants of $\sim 8.0 \times 10^3$ $M^{-1}min^{-1}$. The difference between extension rates results from a larger

Table 2. Incorporation of natural triphosphates opposite unnatural bases in the template[a]

5′-d(TAATACGACTCACTATAGGGAGA)
3′-d(ATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC)

X	TP	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)	X	TP	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
2Br	A	0.38 ± 0.03	49 ± 11	7.8×10^3	**2CN**	A	1.6 ± 0.1	18 ± 3	8.8×10^4
	C	0.14 ± 0.01	20 ± 4	7.0×10^3		C	nd[b]	nd[b]	$<1 \times 10^3$
	G	0.30 ± 0.01	115 ± 20	2.6×10^3		G	0.32 ± 0.03	27 ± 3	1.2×10^4
	T	0.35 ± 0.02	85 ± 9	4.1×10^3		T	0.16 ± 0.03	115 ± 12	1.4×10^3
3Br	A	0.25 ± 0.03	12 ± 2	2.2×10^4	**3CN**	A	1.3 ± 0.2	48 ± 4	2.7×10^4
	C	nd[b]	nd[b]	$<1 \times 10^3$		C	0.07 ± 0.01	58 ± 7	1.2×10^3
	G	nd[b]	nd[b]	$<1 \times 10^3$		G	nd[b]	nd[b]	$<1 \times 10^3$
	T	nd[b]	nd[b]	$<1 \times 10^3$		T	0.97 ± 0.11	124 ± 17	7.9×10^3
4Br	A	0.57 ± 0.11	15 ± 1	3.7×10^4	**4CN**	A	0.35 ± 0.03	20 ± 4	1.7×10^4
	C	0.17 ± 0.02	67 ± 10	2.6×10^3		C	nd[b]	nd[b]	$<1 \times 10^3$
	G	nd[b]	nd[b]	$<1 \times 10^3$		G	nd[b]	nd[b]	$<1 \times 10^3$
	T	0.24 ± 0.03	69 ± 9	3.5×10^3		T	nd[b]	nd[b]	$<1 \times 10^3$

[a]See Materials and Methods section for details.
[b]Reaction was too inefficient for k_{cat} and K_M to be determined independently.

Table 3. Rates of correct extension of unnatural self- and hetero-pairs[a]

5′-d(TAATACGACTCACTATAGGGAGA**X**)
3′-d(ATTATGCTGAGTGATATCCCTCT**Y**GCTAGGTTACGGCAGGATCGC)

X	Y	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)	X	Y	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
2Br	**2Br**	nd[b]	nd[b]	$<1 \times 10^3$	**2CN**	**2Br**	0.12 ± 0.01	120 ± 13	1×10^3
	3Br	nd[b]	nd[b]	$<1 \times 10^3$		**3Br**	nd[b]	nd[b]	$<1 \times 10^3$
	4Br	nd[b]	nd[b]	$<1 \times 10^3$		**4Br**	0.32 ± 0.07	84 ± 14	3.8×10^3
	2CN	nd[b]	nd[b]	$<1 \times 10^3$		**2CN**	nd[b]	nd[b]	$<1 \times 10^3$
	3CN	nd[b]	nd[b]	$<1 \times 10^3$		**3CN**	nd[b]	nd[b]	$<1 \times 10^3$
	4CN	nd[b]	nd[b]	$<1 \times 10^3$		**4CN**	nd[b]	nd[b]	$<1 \times 10^3$
3Br	**2Br**	nd[b]	nd[b]	$<1 \times 10^3$	**3CN**	**2Br**	nd[b]	nd[b]	$<1 \times 10^3$
	3Br	nd[b]	nd[b]	$<1 \times 10^3$		**3Br**	nd[b]	nd[b]	$<1 \times 10^3$
	4Br	nd[b]	nd[b]	$<1 \times 10^3$		**4Br**	nd[b]	nd[b]	$<1 \times 10^3$
	2CN	nd[b]	nd[b]	$<1 \times 10^3$		**2CN**	0.42 ± 0.05	59 ± 13	7.1×10^3
	3CN	nd[b]	nd[b]	$<1 \times 10^3$		**3CN**	nd[b]	nd[b]	$<1 \times 10^3$
	4CN	nd[b]	nd[b]	$<1 \times 10^3$		**4CN**	nd[b]	nd[b]	$<1 \times 10^3$
4Br	**2Br**	nd[b]	nd[b]	$<1 \times 10^3$	**4CN**	**2Br**	nd[b]	nd[b]	$<1 \times 10^3$
	3Br	nd[b]	nd[b]	$<1 \times 10^3$		**3Br**	nd[b]	nd[b]	$<1 \times 10^3$
	4Br	nd[b]	nd[b]	$<1 \times 10^3$		**4Br**	nd[b]	nd[b]	$<1 \times 10^3$
	2CN	1.2 ± 0.1	39 ± 5	3.0×10^4		**2CN**	0.45 ± 0.09	56 ± 8	8.0×10^3
	3CN	nd[b]	nd[b]	$<1 \times 10^3$		**3CN**	nd[b]	nd[b]	$<1 \times 10^3$
	4CN	nd[b]	nd[b]	$<1 \times 10^3$		**4CN**	nd[b]	nd[b]	$<1 \times 10^3$

[a]See Materials and Methods section for details.
[b]Reaction was too inefficient for k_{cat} and K_M to be determined independently.

k_{cat} for **4Br**:**2CN**, relative to the pairs formed with **4CN** or **3CN**. The position of the bromo substituent of **4Br**:**2CN** is apparently critical, since moving it to either the 3- or 2-position results in a pair whose extension is too slow to detect. However, the virtually identical k_{cat}'s and K_M's associated with extension of the **4CN**:**2CN** and **3CN**:**2CN** pairs suggest that the precise position of the cyano group at the primer terminus is less critical. However, moving the cyano group to the 2-position of the analog is not tolerated at the primer terminus, as Kf extends the **2CN** self pair with a rate that is too low to detect. Extension is more specific for the cyano group at the 2-position of the analog in the template, moving it to either the 3- or 4-position results in a pair whose extension could not be detected.

The remaining two unnatural base pairs that are extended with detectable rates also contained **2CN**, but in these cases **2CN** was located at the primer strand, paired opposite either **4Br** or **2Br** in the template strand. These termini were extended with second order rate constants of 3.8×10^3 $M^{-1}min^{-1}$ and 1.0×10^3 $M^{-1}min^{-1}$, respectively. The **2CN**:**4Br** pair is extended less efficiently than the three pairs with **2CN** in the template due a reduced k_{cat}, while the **2CN**:**2Br** pair is extended less efficiently due to both a reduced k_{cat} and an increased K_M. Moving the substituent of either the analog in the template or the primer resulted in loss of detectable polymerase recognition. Again, the pairs in the other strand context were extended with very different rates. **4Br**:**2CN** was extended significantly faster than **2CN**:**4Br** and **2Br**:**2CN** was not extended with a rate sufficient for characterization. In all, these results demonstrate that extension efficiency depends on both the specific substitution pattern of the nucleobase analog and the sequence context.

DISCUSSION

We (8–11,13,15–18) and others (30,31) have developed a wide variety of nucleotides that pair predominantly via hydrophobic packing. These, and other studies (1–4,32–35), have helped deconvolute the varying contributions of nucleobase shape, hydrophobicity and electrostatics to stability and polymerase-mediated replication. To further explore the determinants of base pair stability and replication, we examined nucleobase analogs bearing simple phenyl rings that are modified with bromo and cyano substituents. Characterizing the substituent effects is facilitated by comparing the data with that previously reported for **BEN**, the parental nucleoside possessing an unsubstituted phenyl ring, as well as for the analogs similarly derivatized with methyl and fluoro substituents [S. Matsuda and F. E. Romesberg, manuscript submitted; (18)] (Figure 1). (Note that the 4-fluoro analog has not been characterized.)

Thermodynamic and kinetic orthogonality

While hydrophobic nucleobase analogs are generally thermodynamically orthogonal, stable mispairs are often formed with dA, and kinetic orthogonality is consistently limited by the facile insertion of dATP opposite the analogs in the template. Thus, it is important that the determinants of mispair stability and synthesis be understood, especially relative to adenine.

Mispair stability. Generally, at the 2-position, fluoro and bromo substitution had larger effects than cyano or methyl group substitution. The presence of a halogen at the 2-position stabilizes the mispairs by 1.0 to 2.8°C, with the exception of the **2FB**:dG mispair, which is destabilized by 3.1°C, relative to the **BEN**:dG mispair. It is unclear why the two halogens have such large and opposite effects on the mispair with dG, while cyano or methyl group substitution has little effect.

At the 3- and 4-positions, bromo and cyano substituents stabilize each of the mispairs, which is in contrast to the effects observed with fluoro and methyl group modification, which are generally small. The nonspecific stabilization of all mispairs by the addition of a bromo or cyano substituent at these positions suggests that the stabilization results from either an increased dipole moment or increased polarizability. However, it is again interesting to note that, while methyl substitution at the 3-position has little effect, fluoro and bromo

substitution at the 3-position have opposite effects on the stability of the mispair with dG; **3Br**:dG is stabilized and **3FB**:dG is destabilized, relative to **BEN**:dG.

Mispair synthesis. Kinetic discrimination against the natural dNTPs is important for any candidate unnatural base pair. In the template, **BEN** is selective for dATP insertion. The efficiency results from a large k_{cat}, which is only 8-fold smaller than that for the synthesis of a natural base pair in the same sequence context. Relative to **BEN**, the addition of a bromo or cyano group consistently results in a decreased rate of insertion for each dNTP. The only exceptions are the insertion of dCTP opposite **4Br**, which was 2.6-fold faster than opposite **BEN**, and dGTP opposite **2CN,** which was 7-fold faster than opposite **BEN**. The decrease in dNTP insertion efficiency opposite the bromo- and cyano-derivatized analogs results predominantly from decreased k_{cat} values, which is most pronounced for dATP insertion opposite the analogs modified at the 2-position. The increased discrimination against dATP insertion with bromo or cyano modification at the 2-position is particularly notable, as modification at this position with either a fluoro or methyl group has little or the opposite effect, respectively. While it is often difficult to interpret steady-state kinetic data in terms of mechanism, this data suggests that the fluorine atom may be too small to have a significant effect in the developing transition state, while packing interactions mediated by the 2-methyl group stabilize the transition state, and electrostatic interactions mediated by the bromine atom or cyano group destabilize it. Regardless of its origins, the increased discrimination of **2Br** and **2CN** against dATP should be useful for the further development of more orthogonal analogs, as fidelity with respect to dATP insertion typically limits unnatural base pair synthesis fidelity.

Unnatural base pair stability and extension

In addition to being efficiently and selectively synthesized, an unnatural base pair must also be stable in duplex DNA and efficiently recognized when at the primer terminus so that efficient DNA synthesis may continue. The results of this study demonstrate that bromo and cyano substituents have significant effects on both base pair stability and extension that are again different from those that result from methyl or fluoro group modifications at analogous positions.

Stability of the unnatural base pairs. The stabilities of the bromo- and cyano-derivatized unnatural base pairs vary between 53 and 56°C. These pairs are significantly more stable than any mispair with a natural nucleotide, the greatest stability of which is 51°C. The most stable pairs were the **2Br** self pair and the **2Br**:**2CN** and **3Br**:**4CN** heteropairs. In each case, the bromine is positioned to pack on adjacent purines, where its polarizability may contribute to stable packing interactions. Despite the absence of H-bond donors/acceptors and significantly reduced aromatic surface area, these unnatural base pairs are only 3.2°C less stable than a natural dA:dT pair in the same sequence context. Moreover, this stability is ~2°C greater than any pair formed between analogs bearing a single methyl group or fluorine atom. The data suggest that increased size, dipole moment and polarizability may all contribute to base pair stability.

Kf-mediated extension of the unnatural base pairs. The replication of DNA containing unnatural base pairs is consistently limited by slow primer extension after incorporation of the unnatural nucleotide. For reference, Kf extends the **BEN** self pair by insertion of dCTP opposite dG with a k_{cat}/K_M of only 1.6×10^3 $M^{-1}min^{-1}$. Generally, base pairs formed between the bromo- and cyano-derivatives are not extended more efficiently. However, the pairs composed of an analog with a cyano group at the 2-position, especially in the template, are extended significantly more efficiently than **BEN**. The most efficiently extended pair, **4Br**:**2CN**, is extended by Kf with a second order rate constant of 3.0×10^4 $M^{-1}min^{-1}$, which is significantly more efficient than any pair formed between analogs with single methyl group substituents and only an order of magnitude less efficient than **3FB** self pair extension. Perhaps the increased efficiency of the base pairs with **2CN** in the template at the primer terminus reflects the at least marginal ability of the cyano group to accept a H-bond from the polymerase, as is thought to be important for the extension of a natural base pair (36,37). No analogs with a bromo or cyano group at the 3-position of the ring form pairs that are extended with detectable efficiency. Previous studies of methyl-derivatized analogs suggested that bulky substituents at the 3-position introduce eclipsing interactions within the interbase interface and result in a primer terminus structure that is not well recognized by the polymerase. However, as already mentioned, inclusion of fluorine at the 3-position results in a self pair that is relatively efficiently extended by Kf. Perhaps the smaller fluorine substituents are more optimal for interbase packing, while methyl, bromo and cyano substituents are too bulky. However, preliminary structural data suggests that the **3FB** self pair is dynamic, with each nucleobase adopting multiple conformations in duplex DNA, allowing for the possibility that a minor conformer might be efficiently extended, perhaps one with appropriately oriented dipoles or other electrostatic or packing interactions. Such electrostatic interactions may not be present with methyl substituents, and the size of the methyl, bromo and cyano substituents may disfavor the conformation required for extension.

Progress toward the expansion of the genetic alphabet

Efforts to develop hydrophobic base pairs were originally focused on nucleotides with nucleobase analogs that had extended aromatic surface area. However, these analogs are now thought to interact via face-packing, which may result in poor discrimination against the adenine nucleobase as well as a primer terminus geometry that is not well recognized by DNA polymerases. To circumvent this problem, we have turned our attention to nucleotides bearing smaller nucleobase analogs that should pack edge-on, in a manner topologically similar to a natural base pair, and thus might discriminate more strongly against adenine and form more natural-like primer termini.

In this study, bromo and cyano substitution of the small phenyl ring nucleobase scaffold was seen to have large effects on unnatural base pair and mispair stability. In fact, these substituents are significantly more stabilizing than either methyl or fluoro groups, suggesting that large dipole moments and polarizability are generally important. Thus, these modifications, especially bromo-substitution at the 2- and 3-positions, may be a generally useful tool for stabilizing

unnatural base pairs. This is particularly interesting from the perspective of expanding the genetic alphabet considering that these modifications also dramatically increase kinetic discrimination against dATP. In addition, several of the unnatural base pairs examined are extended more efficiently than the others, suggesting that their electrostatic properties may also be optimized for this critical step of replication. While the most efficiently extended pair, **4Br**:**2CN**, is still extended 10-fold less efficiently than the **3FB** self pair, the data nonetheless suggest that fluorine is not absolutely unique in its ability to facilitate extension.

A particularly exciting conclusion of these studies is that, if the various substituents act independently, it should be possible to use multiple substituents to further improve stability and replication. For example, analogs bearing 2- or 4-bromo and cyano substituents, in addition to a 3-fluoro substituent and perhaps a methyl group, might retain the kinetic orthogonality observed with the analogs characterized in this study, but like **3FB**, might also be more efficiently extended. Such analogs are currently being synthesized.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Funding was provided by the NIH (GM 60005) and the Korea Research Foundation (KRF-2004-214-C00204 TO G.T.H.). Funding to pay the Open Access publication charges for this article was provided by the NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Kool,E.T. (1998) Replication of non-hydrogen bonded bases by DNA polymerases: a mechanism for steric matching. *Biopolymers*, **48**, 3–17.
2. Morales,J.C. and Kool,E.T. (1998) Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nature Struct. Biol.*, **5**, 950–954.
3. Moran,S., Ren,R.X.-F. and Kool,E.T. (1997) A thymidine triphosphate shape analog lacking Watson–Crick pairing ability is replicated with sequence selectivity. *Proc. Natl Acad. Sci. USA*, **94**, 10506–10511.
4. Moran,S., Ren,R.X.-F., Rumney,S.I. and Kool,E.T. (1997) Difluorotoluene, a nonpolar isostere for thymine codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.*, **119**, 2056–2057.
5. Bittker,J.A., Phillips,K.J. and Liu,D.R. (2002) Recent advances in the *in vitro* evolution of nucleic acids. *Curr. Opin. Chem. Biol.*, **6**, 367–374.
6. Wang,L., Brock,A., Herberich,B. and Schultz,P.G. (2001) Expanding the genetic code of *Escherchia coli*. *Science*, **292**, 498–500.
7. Wang,L., Magliery,T.J., Liu,D.R. and Schultz,P.G. (2000) A new functional supressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incoporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.*, **122**, 5010–5011.
8. McMinn,D.L., Ogawa,A.K., Wu,Y., Liu,J., Schultz,P.G. and Romesberg,F.E. (1999) Efforts toward expansion of the genetic alphabet: recognition of a highly stable, self-pairing hydrophobic base. *J. Am. Chem. Soc.*, **121**, 11585–11586.
9. Ogawa,A.K., Wu,Y., Berger,M., Schultz,P.G. and Romesberg,F.E. (2000) Rational design of an unnatural base pair with increased kinetic selectivity. *J. Am. Chem. Soc.*, **122**, 8803–8804.
10. Ogawa,A.K., Wu,Y., McMinn,D.L., Liu,J., Schultz,P.G. and Romesberg,F.E. (2000) Efforts toward the expansion of the genetic alphabet: information storage and replication with unnatural hydrophobic base pairs. *J. Am. Chem. Soc.*, **122**, 3274–3287.
11. Wu,Y., Ogawa,A.K., Berger,M., McMinn,D.L., Schultz,P.G. and Romesberg,F.E. (2000) Efforts toward expansion of the genetic alphabet: optimization of interbase hydrophobic interactions. *J. Am. Chem. Soc.*, **122**, 7621–7632.
12. Berger,M., Ogawa,A.K., McMinn,D.L., Wu,Y., Schultz,P.G. and Romesberg,F.E. (2000) Stable and selective hybridization of oligonucleotides using unnatural hydrophobic bases. *Angew. Chem. Int. Ed.*, **39**, 2940–2942.
13. Tae,E.L., Wu,Y.Q., Xia,G., Schultz,P.G. and Romesberg,F.E. (2001) Efforts toward expansion of the genetic alphabet: replication of DNA with three base pairs. *J. Am. Chem. Soc.*, **123**, 7439–7440.
14. Berger,M., Luzzi,S.D., Henry,A.A. and Romesberg,F.E. (2002) Stability and selectivity of unnatural DNA with five-membered-ring nucleobase analogues. *J. Am. Chem. Soc.*, **124**, 1222–1226.
15. Matsuda,S., Henry,A.A., Schultz,P.G. and Romesberg,F.E. (2003) The effect of minor-groove hydrogen-bond acceptors and donors on the stability and replication of four unnatural base pairs. *J. Am. Chem. Soc.*, **125**, 6134–6139.
16. Henry,A.A., Yu,C. and Romesberg,F.E. (2003) Determinants of unnatural nucleobase stability and polymerase recognition. *J. Am. Chem. Soc.*, **125**, 9638–9646.
17. Yu,C., Henry,A.A., Romesberg,F.E. and Schultz,P.G. (2002) Polymerase recognition of unnatural base pairs. *Angew. Chem. Int. Ed.*, **41**, 3841–3844.
18. Henry,A.A., Olsen,A.G., Matsuda,S., Yu,C., Geierstanger,B.H. and Romseberg,F.E. (2004) Efforts to expand the genetic alphabet: identification of a replicable unnatural DNA self-pair. *J. Am. Chem. Soc.*, **126**, 6923–6931.
19. Henry,A.A. and Romesberg,F.E. (2003) Beyond A, C, G, and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.*, **7**, 727–733.
20. Brotschi,A., Mathis,G. and Leumann,C.J. (2005) Bipyridyl- and biphenyl-DNA: a recognition motif based on interstrand aromatic stacking. *Chem. Eur. J.*, **11**, 1911–1923.
21. Matsuda,S. and Romesberg,F.E. (2004) Optimization of interstrand hydrophobic packing interactions within unnatural DNA base pairs. *J. Am. Chem. Soc.*, **126**, 14419–14427.
22. Hansch,C., Leo,A. and Taft,R.W. (1991) A survey of Hammett substituent constants and resonance field parameters. *Chem. Rev.*, **91**, 165–195.
23. Hansch,C. and Leo,A. (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley & Sons Inc., NY.
24. Yaffe,D., Cohen,Y., Espinosa,G., Arenas,A. and Giralt,F. (2003) A fuzzy ARTMAP-based quantitative structure-property relationship (QSPR) for the henry's law constant of organic compounds. *J. Chem. Info. Comput. Sci.*, **43**, 85–112.
25. Kamlet,M.J., Abboud,J.L.M., Abraham,M.H. and Taft,R.W. (1983) Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π^*, α and β, and some methods for simplifying the generalized solvatochromic equation. *J. Org. Chem.*, **48**, 2877–2887.
26. Wichai,U. and Woski,S.A. (1998) An improved route to 1,2-dideoxy-β-1-phenyl-D-ribofuranose. *Bioorg. Med. Chem. Lett.*, **8**, 3465–3468.
27. Wichai,U. and Woski,S.A. (1999) Disiloxane-protected 2-deoxyribonolactone as an efficient precursor to 1,2-dideoxy-1-β-aryl-D-ribofuranoses. *Org. Lett.*, **1**, 1173–1175.
28. Chen,L.S., Chen,G.J. and Tamborski,C. (1980) The synthesis and reactions of ortho-bromophenyllithium. *J. Organomet. Chem.*, **193**, 283–292.
29. Ellis,G.P. and Romney-Alexander,T.M. (1987) Cyanation of aromatic halides. *Chem. Rev.*, **87**, 779–794.
30. Mitsui,T., Kimoto,M., Sato,A., Yokoyama,S. and Hirao,I. (2003) An unnatural hydrophobic base, 4-propynylpyrrole-2-carbaldehyde, as an efficient pairing partner of 9-methylimidazo[(4,5)-b]pyridine. *Bioorg. Med. Chem. Lett.*, **13**, 4515–4518.
31. Mitsui,T., Kimoto,M., Harada,Y., Sato,A., Kitamura,A., To,T., Hirao,I. and Yokoyama,S. (2002) Enzymatic incorporation of an unnatural base pair between 4-propynyl-pyrrole-2-carbaldehyde and 9-methyl-imidazo [(4,5)-b]pyridine into nucleic acids. *Nucleic Acids Res. Symp. Ser.*, 219–220.
32. Chiaramonte,M., Moore,C.L., Kincaid,K. and Kuchta,R.D. (2003) Facile polymerization of dNTPs bearing unnatural base analogues by DNA polymerase alpha and Klenow fragment (DNA polymerase I). *Biochemistry*, **42**, 10472–10481.

33. Paul,N., Nashine,V.C., Hoops,G., Zhang,P., Zhou,J., Bergstrom,D.E. and Davisson,V.J. (2003) DNA polymerase template interactions probed by degenerate isosteric nucleobase analogs. *Chem. Biol*., **10**, 815–825.
34. Adelfinskaya,O., Nashine,V.C., Bergstrom,D.E. and Davisson,V.J. (2005) Efficient primer strand extension beyond oxadiazole carboxamide nucleobases. *J. Am. Chem. Soc*., **127**, 16000–16001.
35. Zhang,X., Lee,I. and Berdis,A.J. (2005) The use of nonnatural nucleotides to probe the contributions of shape complementarity and π-electron surface area during DNA polymerization. *Biochemistry*, **44**, 13101–13110.
36. Spratt,T.E. (2001) Identification of hydrogen bonds between *Escherichia coli* DNA polymerase I (Klenow fragment) and the minor groove of DNA by amino acid substitution of the polymerase and atomic substitution of the DNA. *Biochemistry*, **40**, 2647–2652.
37. Morales,J.C. and Kool,E.T. (1999) Minor groove interactions between polymerase and DNA: more essential to replication that Watson–Crick hydrogen bonds. *J. Am. Chem. Soc*., **121**, 2323–2324.