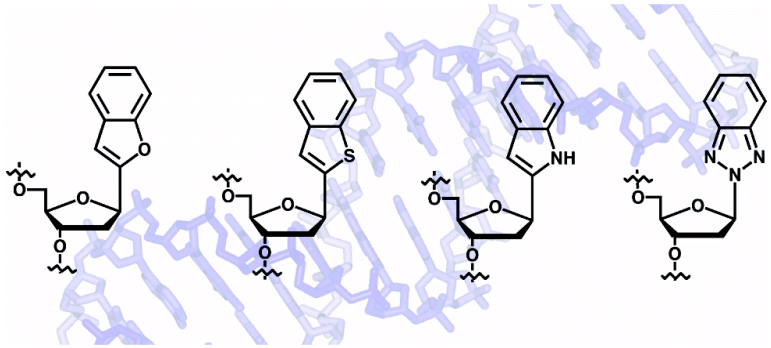
Article

The Effect of Minor-Groove Hydrogen-Bond Acceptors and Donors on the Stability and Replication of Four Unnatural Base Pairs

Shigeo Matsuda, Allison A. Henry, Peter G. Schultz, and Floyd E. Romesberg

J. Am. Chem. Soc., **2003**, 125 (20), 6134-6139• DOI: 10.1021/ja034099w • Publication Date (Web): 29 April 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 7 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 04/29/2003

The Effect of Minor-Groove Hydrogen-Bond Acceptors and Donors on the Stability and Replication of Four Unnatural Base Pairs

Shigeo Matsuda, Allison A. Henry, Peter G. Schultz, and Floyd E. Romesberg*

Contribution from the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received January 9, 2003; E-mail: floyd@scripps.edu

Abstract: The stability and replication of DNA containing self-pairs formed between unnatural nucleotides bearing benzofuran, benzothiophene, indole, and benzotriazole nucleobases are reported. These nucleobase analogues are based on a similar scaffold but have different hydrogen-bond donor/acceptor groups that are expected to be oriented in the duplex minor groove. The unnatural base pairs do not appear to induce major structural distortions and are accommodated within the constraints of a B-form duplex. The differences between these unnatural base pairs are manifest only in the polymerase-mediated extension step, not in base-pair stability or synthesis. The benzotriazole self-pair is extended with an efficiency that is only 200-fold less than a correct natural base pair. The data are discussed in terms of available polymerase crystal structures and imply that further modifications may result in unnatural base pairs that can be both efficiently synthesized and extended, resulting in an expanded genetic alphabet.

Introduction

Inter-nucleobase hydrogen bonding (H-bonding) and shape complementarity are important for the stability and faithful replication of DNA. However, it is not clear whether Watson−Crick base-pairing interactions are a unique solution for the storage of genetic information or if the specific forces that control base-pair stability in duplex DNA are the same as those that are important for correct base pairing during replication, where protein−DNA interactions are also critical. Evaluation of the stability and enzymatic synthesis of DNA containing unnatural nucleobase analogues provide insights into these issues[1−3] and is also useful for designing unnatural base pairs to supplement the genetic alphabet.[4−11] Unnatural base pairs with novel chemical functionality, such as reactive centers or redox-active moieties would expand the chemical potential of DNA and allow for the synthesis of novel polymers with a high degree of sequence and length control. An expanded genetic alphabet would also lay the foundation for the in vivo storage and retrieval of increased information, which might eventually be used to expand the genetic code of a living organism.[11,12]

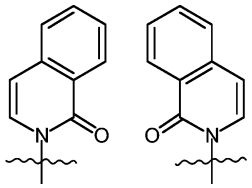


Figure 1. **ICS** self-pair.

In previous reports, we systematically evaluated a variety of self-pairs formed between a single unnatural nucleobase that is predominantly hydrophobic and does not bear an H-bonding pattern or shape that is complementary to the natural bases.[5−11] It should be emphasized that self-pairs, as opposed to hetero-pairs, do not limit efforts to expand the genetic alphabet, but rather facilitate the effort by minimizing the potential for mispairing. Significant progress was made with unnatural nucleosides derived from the isocarbostyryl (**ICS**) framework (Figure 1).[5−8] A variety of these self-pairs were identified, and these self-pairs are as stable or even more stable than natural base pairs in duplex DNA. Furthermore, several can also be synthesized by wild-type polymerases (i.e., enzymatic extension of a primer by incorporation of an unnatural dNTP opposite the unnatural base in the template) with reasonable efficiency and selectivity. However, after synthesis of any of the unnatural base pairs, the newly formed 3′-primer terminus cannot be efficiently extended by incorporation of the next correct natural

(1) Kool, E. T. *Chem. Rev.* **1997**, *97*, 1473−1487.
(2) Kool, E. T. *Biopolymers* **1998**, *48*, 3−17.
(3) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 954−959.
(4) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.
(5) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(6) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(7) Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2000**, *39*, 2940−2942.
(8) Yu, C.; Henry, A. A.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2002**, *41*, 3997−4000.
(9) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(10) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.
(11) Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.
(12) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537−539.

10.1021/ja034099w CCC: $25.00

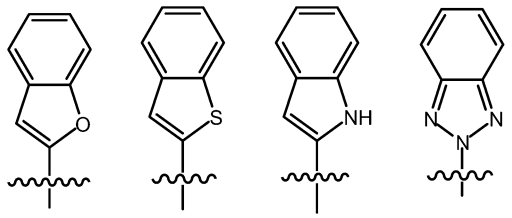


Figure 2. Unnatural bases **BFr**, **BTp**, **IN**, and **BTz** (from left to right).

dNTP. In general, the rate of extension is 3 or 4 orders of magnitude slower than natural base-pair extension. Thus, we have inferred that there are different and more-selective criteria governing extension than those critical for self-pair stability or synthesis. A thorough examination of the nucleobase properties that contribute to efficient extension is critical for the design of self-pairs that will expand the genetic alphabet.

It is commonly accepted that an important criterion for efficient base-pair extension is the presence of an H-bond acceptor positioned appropriately in the developing minor groove of the nascent DNA duplex.[13] Formation of an H-bond between the primer terminus and the polymerase may be critical for achieving the requisite geometry at the primer terminus for efficient continued synthesis. In this manner, the polymerase ensures that primers terminating in a correct base pair are efficiently extended while exonuclease activity is competitive with extension for primers terminating in a mispair. Although the ICS analogues have a carbonyl group that might act as an H-bond acceptor in the developing minor groove, interbase hydrophobic interactions within the self-pair may result in a 3′OH orientation that is less than optimal. To evaluate these issues, we synthesized unnatural nucleoside analogues with differently oriented H-bond acceptors and donors. To examine both H-bond donors and acceptors, while maintaining an aromatic glycosidic linkage, a [5 + 6] ring fusion was chosen as the unnatural nucleobase scaffolding (Figure 2). We now report the stability and replication properties of the self-pairs of benzofuran (**BFr**), benzothiophene (**BTp**), indole (**IN**), and benzotriazole (**BTz**).

Results and Discussion

The **BFr** and **BTp** nucleosides were synthesized as shown in Scheme 1. Lithiation of 2,3-benzofuran (**1a**) and benzothiophene (**1b**), followed by coupling with 3,5-*O*-((1,1,3,3-tetraisopropyl)disiloxanediyl)-2-deoxy-D-ribono-1,4-lactone,[14] afforded the corresponding hemiacetals, which were subsequently reduced with an excess of Et_3SiH and $BF_3{\cdot}OEt_2$ and then deprotected with *n*-TBAF to provide the desired free nucleosides **2a** (36%, 3 steps) and **2b** (19%, 3 steps), respectively. For the synthesis of the unnatural nucleoside (**10**) that has an **IN** base moiety, 1-(phenylsulfonyl)indole (**7**) was lithiated with *n*-butyllithium and added to 2-deoxy-3,5-*O*-(tetraisopropyldisiloxane-1,3-diyl)-D-erythropentofuranose.[15] This coupling reaction gave the corresponding diol in 35% yield and the subsequent ring-closure reaction, which was effected by treatment with 1,1′-azobis(*N*,*N*-dimethylformamide) and *n*-tributylphosphine, afforded the β-anomer of **8** in 57% yield. Deprotection of silyl groups by treatment with *n*-TBAF gave **9**, which was converted to the free nucleoside **10** with heating under basic conditions (K_2CO_3/MeOH/H_2O/reflux). The unnatural nucleoside that had a **BTz** base moiety was synthesized according to a previously reported procedure.[16] The phosphoramidites for solid-phase DNA synthesis were obtained by standard dimethoxytritylation and phosphitylation. The assignment of β-stereochemistry at C1′ for **BFr**, **BTp**, and **IN** was based on COSY and NOESY experiments. The corresponding triphosphates were obtained as described previously.[6] The oligonucleotides were synthesized on a DNA synthesizer, using standard solid-phase β-cyanoethyl phosphoramidite chemistry. After deprotection under basic conditions, the crude oligonucleotides were purified with gel electrophoresis and confirmed by matrix-assisted laser desorption ionization−time-of-flight mass spectroscopy.

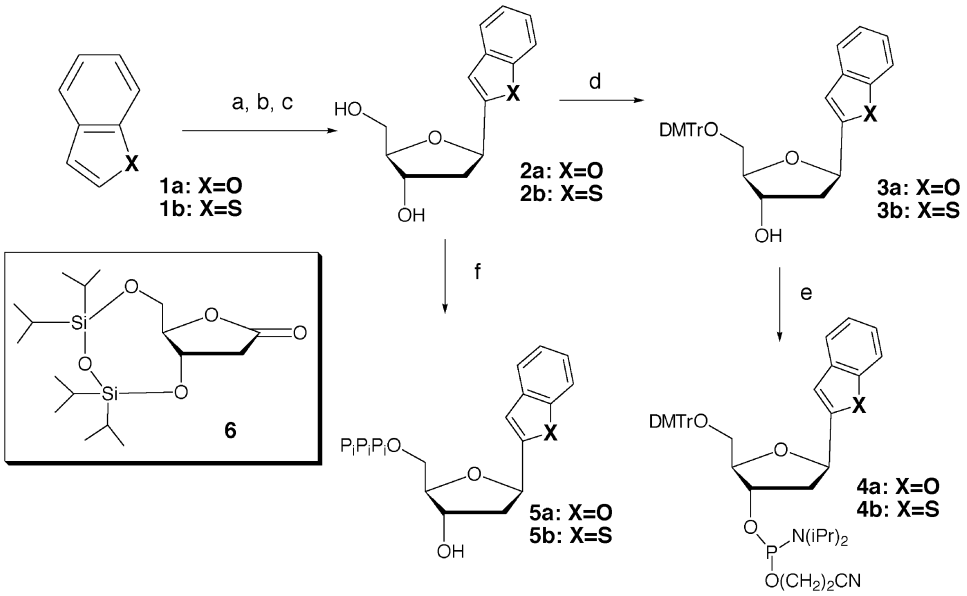
To determine whether the unnatural self-pairs affect the overall structure of duplex DNA, each was incorporated into the complementary oligonucleotides 5′-GCGATG**X**GTAGCG-3′ and 5′-CGCTAC**Y**CATCGC-3′ at positions **X** and **Y**, and circular dichroism (CD) spectra were measured (Figure 3). The spectra are similar and indicate a right-handed helix with a structure resembling that observed in the same duplex with a natural base pair (**X** = dG, **Y** = dC). The rotational strength of the duplex containing the **BTz** self-pair is most similar to that for the fully native natural duplex, whereas the duplexes containing the other self-pairs were slightly stronger. Thus, the unnatural bases do not appear to induce major structural distortions and are able to form self-pairs within the constraints of a B-form duplex. Therefore, any differences in base-pair stability or polymerase recognition do not appear to result from an inability of the self-pairs to be accommodated in duplex DNA, but rather must result from more local structural or electronic effects that are unique to each self-pair.

To evaluate the stability of the unnatural base pairs in duplex DNA, UV melting studies were performed using the same 13 mer oligonucleotides (Table 1). The duplex melting temperature (T_m) ranged from 59.2 °C (**X** = dA, **Y** = dT) to 61.8 °C (**X** = dC, **Y** = dG) for duplexes containing natural base pairs, and from 44.8 to 55.4 °C for duplexes containing mispairs between two natural bases. For duplexes containing unnatural self-pairs, the value of T_m was almost independent of the atom that was proposed to be located in the minor groove. The **BTp** self-pair was relatively more stable (52.2 °C), and **IN** and **BFr** self-pairs were relatively less stable (51.7 and 50.7 °C, respectively). In addition, neither the presence of an imine moiety nor the increased polarity/polarizability of **BTz** significantly affected duplex stability (T_m = 51.2 °C). This independence of T_m on the minor-groove group (ether, amine, thioether, or imine) implies that duplex formation does not strongly alter the solvation of the unnatural base pairs. This likely results from the minor-groove "spine of solvation",[17] which preserves the accessibility of minor-groove moieties to water upon duplex formation.

In all cases, the stability of the unnatural self-pair was greater than all possible mispairs in the sequence context that was examined. With the exception of those involving dC, mispairs

(13) Guo, M.-J.; Hildbrand, S.; Leumann, C. J.; McLaughlin, L. W.; Waring, M. J. *Nucleic Acids Res.* **1998**, *26*, 1863−1869.
(14) Wichai, U.; Woski, S. A. *Org Lett.* **1999**, *1*, 1173−1175.
(15) Yokoyama, M.; Ikeue, T.; Ochiai, Y.; Momotake, A.; Yamaguchi, K.; Togo, H. *J. Chem. Soc., Perkin Trans. 1* **1998**, 2185−2191.
(16) Kazimierczuk, Z.; Seela, F. *Helv. Chim. Acta* **1990**, *73*, 316−325.
(17) Saenger, W. *Principles of Nucleic Acid Structure*, 2nd ed.; Springer-Verlag: New York, 1983.

Scheme 1 [a]



[a] (a) *n*-BuLi, −78 °C, then **6**, THF; (b) BF_3·OEt_2/Et_3SiH/CH_2Cl_2, −78 °C; (c) *n*-TBAF/THF; (d) DMTrCl, DMAP, Et_3N, pyridine; (e) $NC(CH_2)_2OP(NiPr_2)_2$, 1*H*-tetrazole, CH_3CN, (f) $POCl_3$, *n*-Bu_3N, proton sponge, trimethyl phosphate, $(Bu_3NH)_3(H_4P_2O_7)$.

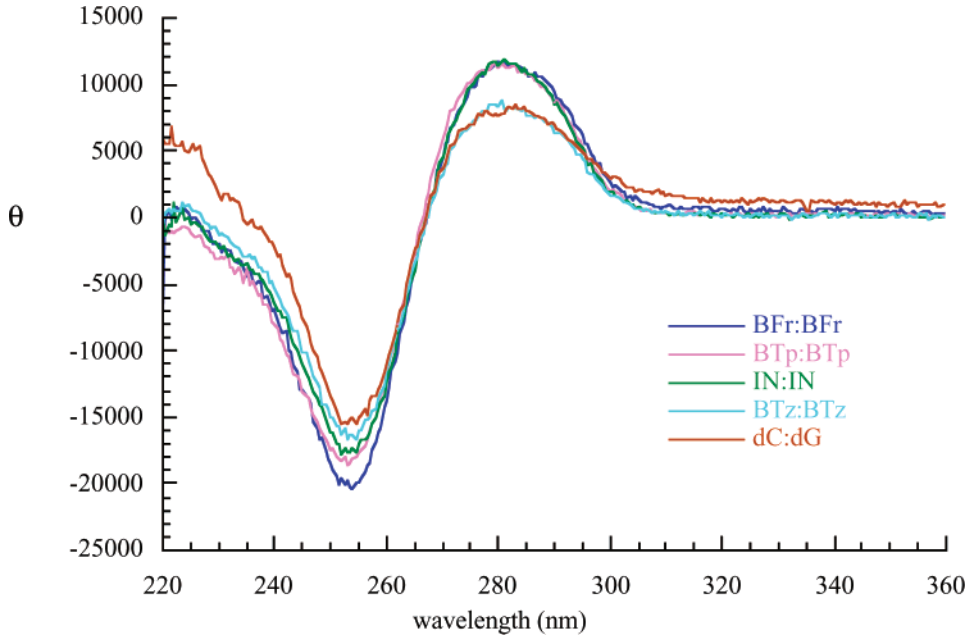


Figure 3. CD spectra of duplex DNA containing self-pairs or dG:dC. See text for details.

between an unnatural base and a natural base gave similar T_m values, differing by only 2.5 °C (47.3 °C, compared to 49.8 °C). In general, the mispairing of the unnatural bases with natural purine bases resulted in slightly more-stable duplexes. This likely reflects increased intrastrand aromatic stacking interactions. The mispairs with dC were significantly less stable, with T_m values of 44.0−47.0 °C. This may result from the diminished aromatic surface area and increased hydrophilicity of dC, relative to the other natural nucleobases.[18]

The unnatural nucleobases were also evaluated as substrates for the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I (KF). Steady-state kinetic constants were determined by measuring the initial velocities (<15% conversion) of radiolabeled primer extension reactions.[6] The reactions were analyzed by denaturing polyacrylamide gel electrophoresis and quantified with a PhosphorImager (Molecular Dynamics). The initial velocities were plotted versus [dNTP] and fit to the Michaelis−Menten equation to determine k_{cat} and K_M. Unnatural nucleobases were assayed both in template DNA and as triphosphates. The steady-state kinetic constants for single nucleotide incorporation are reported in Table 2.

The efficiency of self-pair synthesis was only modestly dependent on the unnatural base. Of the four self-pairs, **BTp** is the most efficiently synthesized with a specificity constant (k_{cat}/K_M) of 3.3×10^5 min^{-1} M^{-1}, 136 times less efficient than the synthesis of a dA:dT base pair in the same sequence context (~4.5×10^7 min^{-1} M^{-1}). The self-pair synthesis rate of **BFr** is 2.8×10^5 min^{-1} M^{-1}, 1.2 times lower than that of **BTp**. The

(18) Shih, P.; Pedersen, L. G.; Gibbs, P. R.; Wolfenden, R. *J. Mol. Biol.* **1998**, *280*, 421−430.

Table 1. Denaturation Temperatures for Duplex DNA Containing **BFr**, **BTp**, **IN**, and **BTz**[a]

5′ -dGCGTAC**X**CATGCG
3′ -dCGCATG**Y**GTACGC

X	Y	T_m, °C	X	Y	T_m, °C
BFr	**BFr**	50.7	**IN**	**IN**	51.7
BFr	C	44.0	**IN**	C	45.8
BFr	G	48.7	**IN**	G	49.7
BFr	T	47.7	**IN**	T	48.8
BFr	A	48.7	**IN**	A	49.8
BTp	**BTp**	52.2	**BTz**	**BTz**	51.2
BTp	C	47.0	**BTz**	C	44.3
BTp	G	49.1	**BTz**	G	48.7
BTp	T	48.2	**BTz**	T	47.3
BTp	A	48.2	**BTz**	A	48.3

[a] Determined in 10 mM PIPES, 10 mM $MgCl_2$, 100 mM NaCl, pH 7. For reference, T_m = 59.2 °C for X = dA, Y = dT and T_m = 61.8 °C for X = dC, Y = dG.

Table 2. Efficiency of Self-Pair Synthesis and Selectivity Opposite Unnatural Bases in the Template

5'-dTAATACGACTCACTATAGGGAGA
3'-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC

d**Y**TP

5'-dTAATACGACTCACTATAGGGAGA**Y**
3'-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC

X	Y	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} M^{-1})
BFr	**BFr**	19 ± 6	67 ± 5	2.8×10^5
BFr	A	1.5 ± 0.5	15 ± 5	1.0×10^5
BFr	C	0.4 ± 0.1	283 ± 105	1.6×10^3
BFr	G	0.28 ± 0.06	86 ± 13	3.2×10^3
BFr	T	1.3 ± 0.6	147 ± 18	9.1×10^3
BTp	**BTp**	19 ± 5	57 ± 30	3.3×10^5
BTp	A	2.9 ± 0.4	41 ± 5	7.0×10^4
BTp	C	nd[a]	nd[a]	$<1.0 \times 10^3$
BTp	G	0.17 ± 0.08	141 ± 15	1.2×10^3
BTp	T	4.9 ± 0.6	154 ± 56	3.2×10^4
IN	**IN**	10 ± 4	69 ± 5	1.4×10^5
IN	A	0.9 ± 0.2	27 ± 7	3.2×10^4
IN	C	nd[a]	nd[a]	$<1.0 \times 10^3$
IN	G	nd[a]	nd[a]	$<1.0 \times 10^3$
IN	T	0.9 ± 0.7	197 ± 56	4.6×10^3
BTz	**BTz**	16 ± 1	209 ± 93	7.5×10^4
BTz	A	4.8 ± 0.6	11 ± 1	4.2×10^5
BTz	C	1.0 ± 0.3	219 ± 44	4.7×10^3
BTz	G	1.6 ± 0.2	91 ± 8	1.8×10^4
BTz	T	3.8 ± 0.4	153 ± 21	2.5×10^4

[a] Rates too slow for determination of k_{cat} and K_M independently.

IN self-pair is synthesized with an efficiency (1.4×10^5 min^{-1} M^{-1}) that is reduced by a factor of 2.4, relative to that of the **BTp** self-pair. Finally, the **BTz** self-pair is the least efficiently synthesized, with a rate of 7.5×10^4 min^{-1} M^{-1}, which is 4.4 times lower than that of the **BTp** self-pair, and 600 times lower than that of natural synthesis. The moderate variation in self-pair synthesis rates demonstrates that, as in base-pair stability, the nature of the H-bond donor/acceptor at the position in the nucleobase α to the C-glycosidic bond is not critical for unnatural base-pair synthesis.

The crystal structures of two type I DNA polymerases are available (from *Bacillus stearothermophilus* and *Thermus aquaticus*).[19,20] Both polymerases are highly homologous to KF, in regard to sequence, structure, and function. The structures show that the heteroatom H-bond acceptor of the bound triphosphate interacts with the template base as well as a water molecule. The observed independence of the insertion rates on unnatural dNTP implies that these interactions are not significant for unnatural base-pair synthesis, or that they are sufficiently satisfied by bound water molecules.

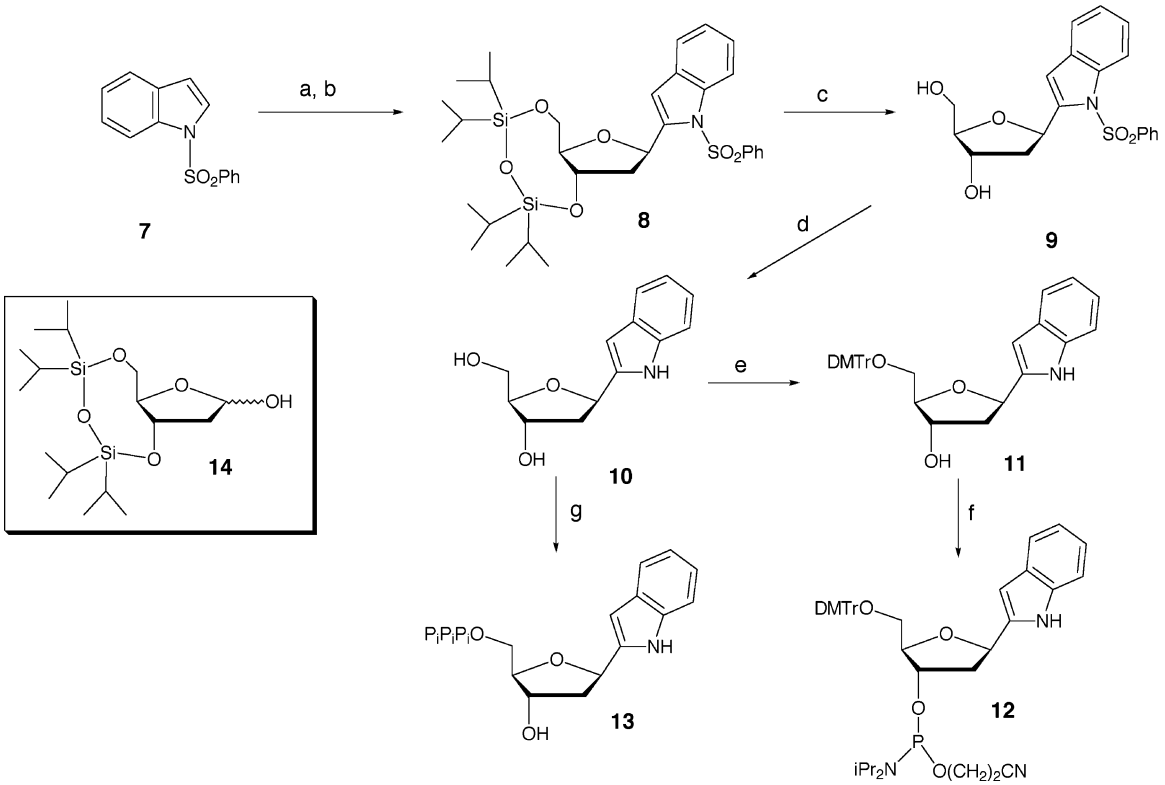
In three of the four cases, the unnatural self-pairs are synthesized selectively, relative to all possible mispairs, albeit with fidelities that are compromised, relative to the natural base pairs. For each unnatural base in the template, dATP is inserted most competitively, with rates varying by a factor of 10, from 3.2×10^4 min^{-1} M^{-1} (for **IN**) to 4.2×10^5 min^{-1} M^{-1} (for **BTz**). In the case of **BTz**, dATP is inserted 5.6 times faster than d**BTz**TP. These efficiencies result in part from K_M effects, because dATP is consistently the most tightly bound dNTP. This likely results both from the hydrophobicity of adenine,[18] which favors packing opposite the predominantly hydrophobic unnatural base, and from the heteroatom substitution of the purine scaffolding, which favors packing with the base at the primer terminus.[10,21,22]

The next most efficiently formed mispair is that resulting from dTTP insertion opposite each unnatural base in the template. The mispairs were synthesized, relative to the correct self-pairs, with 3-, 10-, 30-, and 31-fold reduced efficiency for the **BTz**, **BTp**, **IN**, and **BFr**, respectively. In the case of dTTP insertion, an order-of-magnitude gain in selectivity is observed in the **IN** and **BFr** bases, compared to that of the **BTz** base, and there is no significant difference between **IN** and **BFr**. Against each unnatural base in the template, dGTP is the next most efficiently incorporated, although less so than dATP or dTTP, and 4-, 88-, 140-, and 275-fold less efficiently than correct self-pair synthesis for **BTz**, **BFr**, **IN**, and **BTp**, respectively. In this case, a 20-fold gain in selectivity is observed for **BFr**, compared to **BTz**; a 35-fold gain is observed for **IN**, and a 70-fold gain is observed for **BTp**. Finally, dCTP is the triphosphate that is least efficiently incorporated opposite each unnatural base in the template, with self-pair synthesis selectivities of 16-, 140-, 175-, and 330-fold for **BTz**, **IN**, **BFr**, and **BTp**, respectively. In the case of dCTP insertion, the selectivities represent gains of 9- to 20-fold for **IN**, **BFr**, and **BTp**, compared to **BTz**.

The ability of KF to extend a primer that terminates at each of the unnatural self-pairs (by insertion of dCTP opposite dG) was also examined (Table 3). Unlike with stability and synthesis, the unnatural bases show a marked variation in extension rates, depending on the nature of the atom disposed toward the developing minor groove. Sulfur substitution results in a **BTp** self-pair that is extended with an efficiency of only 7.4×10^2 min^{-1} M^{-1}, which results from both a low k_{cat} value (0.17 min^{-1}) and a large K_M value (230 μM). Oxygen and NH substitution result in moderate increases in k_{cat} but have little effect on K_M. Remarkably, the **BTz** self-pair is extended with a k_{cat}/K_M value of 1.7×10^5 min^{-1} M^{-1}, which is 60−230 times more efficient than the other members of the family. The increase in rate is due to a large increase in k_{cat} (by a factor of 17−82) and a more modest decrease in K_M (by a factor of

(19) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304−307.
(20) Kim, Y.; Eom, S. H.; Wang, J.; Lee, D.-S.; Suh, S. W.; Steitz, T. A. *Nature* **1995**, *376*, 612−616.
(21) Bugg, C. E. *Solid-State Packing Patterns of Purine Bases*; Bergmann, E. D., Pullman, B., Eds.; The Israel Academy of Sciences and Humanities: Jerusalem, 1971; pp 178−204.
(22) Dahl, T. *Acta Chem. Scand.* **1994**, *48*, 95−106.

Scheme 2 [a]



[a] (a) *n*-BuLi, −78 °C, then **14**, THF; (b) 1,1′-azobis(*N*,*N*-dimethylformamide), *n*-Bu_3P, benzene; (c) *n*-TBAF/THF; (d) K_2CO_3, MeOH, H_2O, reflux; (e) DMTrCl, DMAP, Et_3N, pyridine; (f) $NC(CH_2)_2OP(NiPr_2)_2$, 1*H*-tetrazole, CH_3CN; (g) $POCl_3$, *n*-Bu_3N, proton sponge, trimethyl phosphate, $(Bu_3NH)_3(H_4P_2O_7)$.

Table 3. Correct Extension of Unnatural Self-Pairs[a]

```
5'-dTAATACGACTCACTATAGGGAGAX
3'-dATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC
                          ↓ dCTP
5'-dTAATACGACTCACTATAGGGAGAXC
3'-dATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC
```

unnatural self-pair	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\ M^{-1}$)
BFr	0.81 ± 0.05	294 ± 9	2.8×10^3
BTp	0.17 ± 0.04	230 ± 49	7.4×10^2
IN	0.4 ± 0.1	207 ± 30	1.8×10^3
BTz	14 ± 3	85 ± 20	1.7×10^5

[a] dCTP incorporation opposite G in the template. Extension with an incorrect natural dNTP was, in every case, $< 3 \times 10^2\ min^{-1}\ M^{-1}$.

2−3.5). This self-pair is extended only 200-fold less efficiently than a correct natural base pair. In the context of the high extension rate, it is interesting to recall that, although each self-pair appears to be accommodated in B-form duplex DNA, the CD spectrum of DNA containing the **BTz** self-pair was the most similar to that of fully natural DNA. Thus, the increased extension rate may have its origin in subtle structural effects that are also reflected in the duplex CD strength.

To help understand what limits the rate of **BTz** self-pair extension, we evaluated the α-thiotriphosphate elemental effect.[23] In general, KF-mediated extension of a primer that terminates at a correct pair between two canonical bases is rate-limited by a nonchemical step (which is proposed to be a protein conformational change,[23] although this has been recently challenged[24]), whereas extension of a mispair is limited by the rate of the bond formation step (nucleophilic displacement of pyrophosphate from the incoming dNTP by the primer 3′OH). The difference in the rate-limiting step for correct or mispair extension results in different elemental effects: k_{cat}(dNTP)/k_{cat}-(α(*S*)dNTP) is fully manifested for mispair extension (∼17) but partially masked for correct pair extension (∼2). To determine which step limits the rate of **BTz** self-pair extension, the kinetic experiments previously described with dCTP substrates were repeated with a mixture of *R* and *S* diastereomers of 5′-*O*-1-thiotriphosphate (Trilink Biotechnologies). Extension of a correct dA:dT base pair showed an α-thiotriphosphate elemental effect of 2.8, in good agreement with previous reports. Extension of the **BTz** self-pair showed an elemental effect of 34. This observation implies that nucleophilic attack of the 3′OH of the unnatural nucleotide at the primer terminus on the incoming dNTP is fully rate-limiting. Thus, even with the most efficiently extended unnatural base pair, the primer terminus is still not optimally oriented for continued synthesis. Importantly, this also implies that further self-pair derivatization that optimizes positioning of the 3′OH group should directly translate to more-efficient self-pair extension.

The polymerase crystal structures show that the nucleobase at the primer terminus packs against the nucleobase of the incoming triphosphate and, along with the template nucleobase, is tightly packed by hydrophobic side chains of the protein.[19,20] This binding pocket is thought to exert a steric selection against the aberrant structures of mispaired bases. The structures also reveal sequence-independent H-bonds between the polymerase (Arg573) and the purine N3 or pyrimidine O2 atoms of the

(23) Mizrahi, V.; Henrie, R. N.; Marlier, J. F.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1985**, *24*, 4010−4018.

(24) Showalter, A. K.; Tsai, M.-D. *Biochemistry* **2002**, *41*, 10571−10576.

natural bases at the primer terminus. Failure of the unnatural DNA base pair to engage the protein suitably via any of these interactions may result in a misaligned 3′OH primer terminus that is unable to act as an efficient nucleophile during the extension step. Although the overall structure and accommodation in B-form duplex DNA is similar for each self-pair, the **BTz** self-pair differs in at least two ways. **BTz** presents a natural-like endocyclic imine functionality in the developing minor grove at the primer terminus, as opposed to an ether (**BFr**), amine (**IN**), or thioether (**BTp**), to the developing minor groove at the primer terminus. The imine moiety may form a more favorable H-bond with Arg573 in KF and, therefore, orient the primer terminus for continued extension.[25] The overall dipole and polarizability are also significantly different for **BTz**, relative to the other unnatural nucleobases. This difference may alter packing with the adjacent primer base, or the incoming dNTP, in a manner favorable for extension.[10,21]

Further clarification of the physical origins of the increased rates for continued primer extension will require additional structure/function studies. For example, we have initiated NMR solution structure studies to address the orientation and packing of the unnatural nucleobases in duplex DNA. Nonetheless, the **BTz** self-pair shows extension rates that are reduced by a factor of only $\sim 10^2$, relative to natural DNA in the same sequence context. Primer extension after synthesis of the unnatural base pair is consistently the step that limits the synthesis of DNA containing unnatural bases. The sensitivity of this step to derivatization of the unnatural base suggests that modifications may be found that result in unnatural base pairs that can be both efficiently synthesized and extended, resulting in an expanded genetic alphabet.

Acknowledgment. Funding was provided by the National Institutes of Health (No. GM 60005 to F.E.R.), the Skaggs Institute for Chemical Biology (F.E.R. and P.G.S.), and the JSPS Research Fellowships for Young Scientists (S.M.).

Supporting Information Available: Details of synthetic procedures and characterizations (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA034099W

(25) (a) Monge, M. A.; Puebla, G.; Gallego, M. M. R.; Pardo, C.; Elguero, J. *J. Heterocycl. Chem.* **1992**, *29*, 499−502. (b) Buchet, R.; Dion, A. *J. Biomol. Struct. Dyn.* **1986**, *4*, 231−241.