Optimization of Interstrand Hydrophobic Packing Interactions within Unnatural DNA Base Pairs

Shigeo Matsuda and Floyd E. Romesberg*

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

Received May 9, 2004; E-mail: floyd@scripps.edu

Abstract: As part of an effort to expand the genetic alphabet, we have evaluated a large number of predominantly hydrophobic unnatural base pairs. We now report the synthesis and stability of unnatural base pairs formed between simple phenyl rings modified at different positions with methyl groups. Surprisingly, several of the unnatural base pairs are virtually as stable as a natural base pair in the same sequence context. The results show that neither hydrogen-bonding nor large aromatic surface area are required for base pair stability within duplex DNA and that interstrand interactions between small aromatic rings may be optimized for both stability and selectivity. These smaller nucleobases are not expected to induce the distortions in duplex DNA or at the primer terminus that seem to limit replication of larger unnatural base pairs, and they therefore represent a promising approach to the expansion of the genetic alphabet.

1. Introduction

In an effort to develop a third base pair to expand the genetic alphabet, we have synthesized and characterized a large number of nucleotides bearing unnatural nucleobase analogues.[1−10] Just as with their natural counterparts, these unnatural nucleotides must pair stably and selectively within duplex DNA. In the case of natural DNA, thermal stability and selectivity result from a combination of intrastrand packing and interstrand hydrogen bonding (H-bonding) interactions. Correspondingly, thermal stability and selectivity of DNA containing unnatural base pairs may be mediated by either intra- or interstrand interactions. Nucleobase H-bonding interactions have been modified[11−15] and optimized,[16,17] and their intrastrand packing has been improved
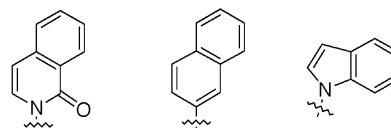

Figure 1. Isocarbostyril (left), naphthyl (middle), and indol (right) scaffolds.

by increasing their surface area.[18,19] Increased surface area has been achieved by the addition of alkyl, alkynyl, or aryl substituents at the C7 and C5 positions of purines and pyrimidines, respectively.[18−22] Similarly, we have evaluated a variety of large aromatic nucleobase analogues, based on the isocarbostyril, naphthyl, and indole scaffolds (Figure 1). (Although many of the unnatural nucleobases are not actually basic, we refer to them as nucleobase analogues for simplicity.) Generally, these nucleobases form unnatural self-pairs (pairs formed between two identical unnatural nucleotides) and heteropairs (pairs formed between different unnatual nucleotides) with reasonable stability and good selectivity, relative to the natural base pairs.[1−10] Either self-pairs or heteropairs may be used for expansion of the genetic code.

Efficient replication by a DNA polymerase is also required of an unnatural base pair. Unfortunately, we have generally found that nucleobases with large hydrophobic surface area are poor substrates for replication.[1−10] While the unnatural triphosphates tend to be efficiently inserted opposite their designed partner in the template, the nascent unnatural primer terminus

(1) Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2000**, *39*, 2940−2942.
(2) Berger, M.; Luzzi, S. D.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1222−1226.
(3) Henry, A. A.; Yu, C.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 9638−9646.
(4) Matsuda, S.; Henry, A. A.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 6134−6139.
(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) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(8) Tae, E. L.; Wu, Y. Q.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.
(9) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.
(10) Yu, C.; Henry, A. A.; Romesberg, F. E.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 3841−3844.
(11) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322−8323.
(12) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37.
(13) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 2, 10489−10496.
(14) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333.
(15) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.

(16) Lin, K.-Y.; Matteucci, M. D. *J. Am. Chem. Soc.* **1998**, *120*, 8531−8532.
(17) Flanagan, W. M.; Wolf, J. J.; Olsen, P.; Grant, D.; Lin, K. Y.; Wagner, R. W.; Matteucci, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3515−3518.
(18) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990−1009.
(19) Kool, E. T. *Chem. Rev.* **1997**, *97*, 1473−1487.
(20) Ahmadian, M.; Zhang, P.; Bergstrom, D. E. *Nucleic Acids Res.* **1998**, *26*, 3127−3135.
(21) He, J.; Seela, F. *Nucleic Acids Res.* **2002**, *30*, 5485−5496.
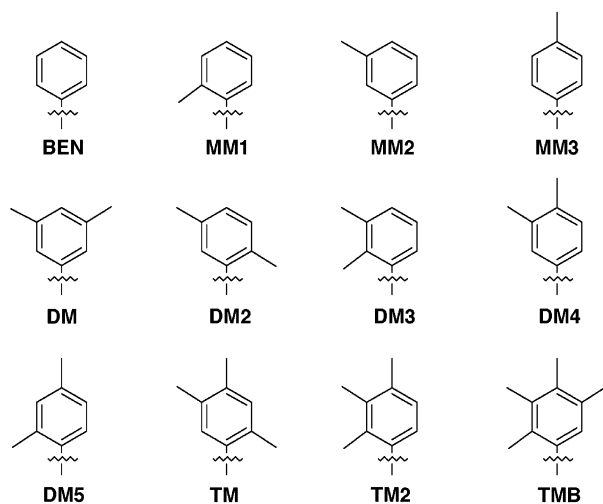(22) Seela, F.; Zulauf, M. *Helv. Chim. Acta* **1999**, *82*, 1878−1898.

Figure 2. Unnatural nucleobases used in this study.

is then a poor substrate for continued extension. One explanation for the poor extension rates may be the absence of heteroatoms in both the major and the minor grooves. These atoms would contribute to the physical properties of the nucleobases, including dipole moment and polarizability.[23−27] Indeed, we have recently demonstrated that heteroatom substitution may be used to improve unnatural base pair extension rates.[3,4,10] It may also be that the large aromatic surface area itself contributes to inefficient extension. Steric repulsion or interstrand intercalation between the unnatural bases may distort the primer terminus and disfavor extension. In fact, preliminary NMR studies suggest that the isocarbostyril-based unnatural nucleobases do interstrand intercalate and appear to adopt a geometry that is favorable for packing but inappropriate for polymerase-catalyzed primer extension (Wemmer, D., unpublished data). We hypothesize that small nucleobases, having an aromatic surface area that is too limited to allow interstrand intercalation and having suitably positioned heteroatoms, will form unnatural base pairs that are thermally stable and selective and also efficiently synthesized and extended by a DNA polymerase.

To test this hypothesis, we first asked whether packing interactions between benzene-based nucleobase analogues that are too small to intercalate might be optimized for stability and selectivity within duplex DNA. Herein, we report the synthesis and thermodynamic characterization of 10 novel nucleotides bearing methyl-derivatized benzene rings which, when combined with two previously reported analogues,[6] provide a systematic analysis of heteropair and self-pair stability (Figure 2). All of the unnatural base pairs have good selectivity, preferring to pair with one another rather than pair with a natural base, and several are also reasonably stable relative to a natural base pair. These studies demonstrate that hydrophobic and packing forces are

sufficient for controlling molecular recognition within duplex DNA without H-bonding or extended aromatic surface area.

2. Results

2.1. Nucleobase Analogues. The synthesis and characterization of the unnatural nucleoside is described in the Supporting Information. Briefly, nucleobase−sugar coupling relied on Heck-type coupling reactions[28−31] between the silyl-protected glycal[32] and each methyl-substituted iodobenzene or addition of the aryllithium reagent to the disiloxane-protected 2-deoxy-D-ribono-1,4-lactone.[33] In all cases, the phosphoramidite used in automated DNA synthesis was obtained by standard dimethoxy-tritylation and phosphitylation. Oligonucleotides were synthesized on an Applied Biosystems 392 DNA synthesizer using standard solid-phase β-cyanoethyl phosphoramidite chemistry. After deprotection under basic conditions, the crude oligonucleotides were purified by polyacrylamide gel electrophoresis.

2.2. Preliminary Structural Characterization of DNA Containing Unnatural Self-Pairs. To determine whether the unnatural self-pairs perturbed the overall structure of a short DNA duplex, each unnatural nucleoside was incorporated, at position **X**, into the complementary oligonucleotides 5′-d(GCG-TAC**X**CATGCG)-3′ and 5′-d(CGCATG**X**GTACGC)-3′. The circular dichroism (CD) spectrum, from 220 to 360 nm, was recorded for each unnatural duplex, as well as for the fully natural duplexes containing dA:dT or dG:dC (Figure 3). The λ_{max} and λ_{min} of all of the spectra are virtually identical, indicating that the unnatural base pairs do not significantly distort the B-form conformation of the DNA duplex. The amplitudes of the negative feature for each duplex containing a self-pair are similar and intermediate between the duplex containing a dA:dT and dG:dC base pair. There is more variation in the amplitude of the positive feature. While the varying ellipticity is difficult to interpret, we note that it at least roughly follows the stability of the self-pairs and thus may reflect differences in packing with neighboring base pairs.

To generate a more detailed structural model, we energy-minimized the same duplex containing the **DM5** self-pair (Figure 4) using the AMBER* force field,[34,35] implemented with the MacroModel program (version 6.0).[36] The resulting structure predicts that the DNA double helix can accommodate the **DM5** self-pair without major backbone or stacking distortions (Figure 4a). The torsion angle about the exocyclic $C_{4'}-C_{5'}$ bond, critical for positioning of the 5′ phosphate group, relative to the sugar and the nucleobase, is ∼63° for both unnatural nucleotides and is between 62 and 65 °C for each flanking natural nucleotide. The intrastrand phosphate distance (6.6−7.0 Å) is typical of B-form DNA.[37]

(23) Bergmann, E. D.; Weiler-Feilchenfeld, H. In *The Dipole Moments of Purines*; Bergmann, E. D., Pullman, B., Eds.; The Israel Academy of Sciences and Humanities: Jerusalem, 1971; pp 21−28.
(24) Lister, J. H. *The Purines. Supplement 1*; John Wiley and Sons: New York, 1996; Vol. 54.
(25) Hurst, D. T. *An Introduction to the Chemistry and Biochemistry of Pyrimidines, Purines and Pteridines*; Page Bros.: Norwich, Great Britain, 1980.
(26) Gilchrist, T. L. *Heterocyclic Chemistry*; Longman Scientific & Technical: Essex, 1985.
(27) Bugg, C. E. In *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.

(28) Farr, R. N.; Outten, R. A.; Cheng, J. C. Y.; Daves, G. D. *Organometallics* **1990**, *9*, 3151−3156.
(29) Zhang, H. C.; Daves, G. D. *J. Org. Chem.* **1992**, *57*, 4690−4696.
(30) Hsieh, H. P.; McLaughlin, L. W. *J. Org. Chem.* **1995**, *60*, 5356−5359.
(31) Zhang, H. C.; Brakta, M.; Daves, G. D. *Nucleosides Nucleotides* **1995**, *14*, 105−116.
(32) Cameron, M. A.; Cush, S. B.; Hammer, R. P. *J. Org. Chem.* **1997**, *62*, 9065−9069.
(33) Wichai, U.; Woski, S. A. *Org. Lett.* **1999**, *1*, 1173−1175.
(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(35) Ferguson, D. M.; Kollman, P. A. *J. Comput. Chem.* **1991**, *12*, 620−626.
(36) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.
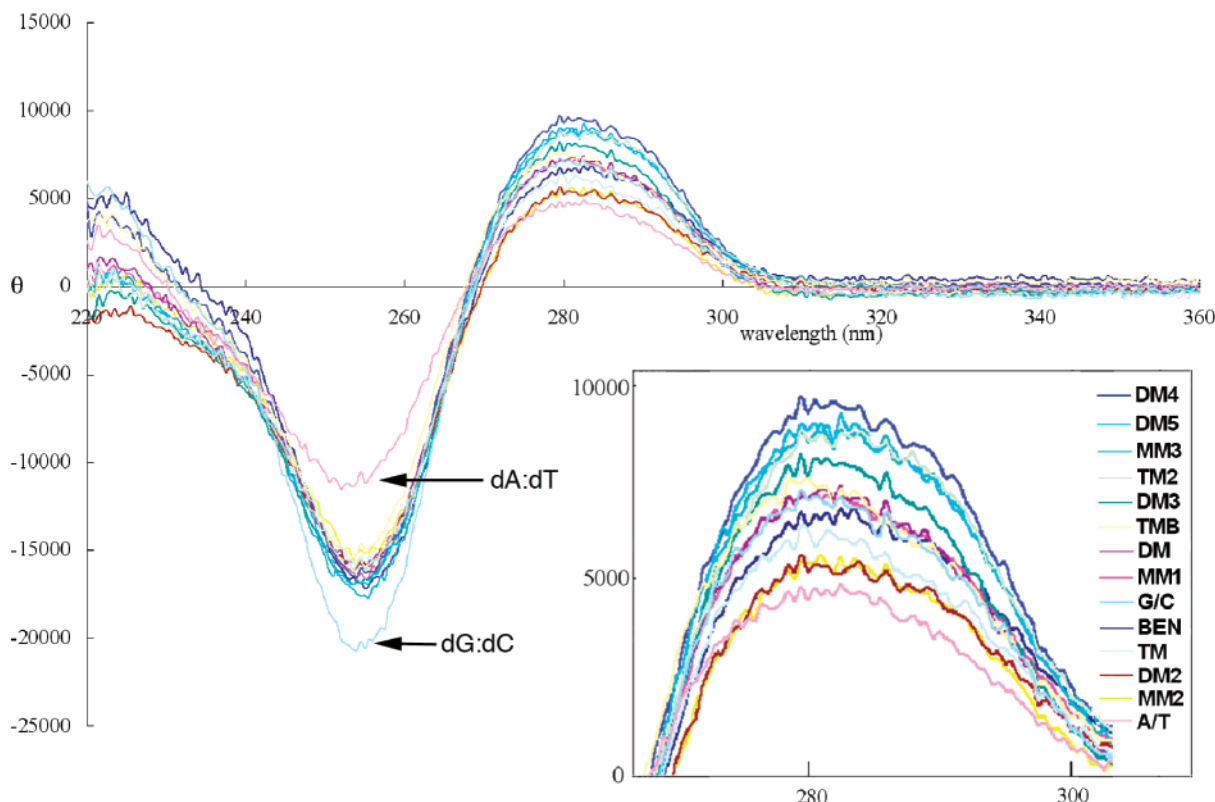(37) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

Figure 3. CD spectra of DNA duplexes 5'-d(GCGTAC**X**CATGCG)-3'/5'-d(CGCATG**X**GTACGC)-3' containing self-pairs or natural base pairs at position **X**. The spectra of the fully natural duplexes are indicated. In the inset are listed the base pairs from top to bottom corresponding to decreasing amplitude of the CD signal at 280 nm.

The O4′−C1′−C1−C2 dihedral angles of the **DM5** nucleotides are −148° (for −G**DM5**G−) and −113° (for −C**DM5**C−). We will refer to this conformation as *anti* by analogy to natural nucleotides. The *anti* conformation positions the *ortho*-methyl group in the minor groove as opposed to over the deoxyribose ring. We speculate that all of the *ortho*-substituted unnatural nucleotides adopt such *anti*-orientations to minimize steric interactions between the methyl group and the sugar ring. This conformation positions the methyl substituents of both **DM5** nucleobases appropriately to pack with one another. As a result, the two **DM5** analogues appear to pack well with each other, with interbase distances approaching the sum of the van der Waals radii (Figure 4b). However, the methyl groups are not equally packed within the duplex (Figure 4b). For the **DM5** nucleobase flanked by two pyrimidines, only the *para*-methyl group is well packed by the flanking bases. In contrast, for the **DM5** nucleobase flanked by two purines, both the *ortho*- and the *para*-methyl groups appear to be well packed, due to the increased aromatic surface area of the neighboring bases. Together, the CD and modeling studies suggest that the unnatural base pairs are accommodated in a B-form DNA duplex with each unnatural nucleotide adopting an orientation that minimizes steric interactions with the sugar and optimizes methyl group packing with the pairing nucleobase. This structural model is used to interpret the thermodynamic data described in the following section.

2.3. Stability of Unnatural Base Pairs. To evaluate the thermodynamic stability and selectivity of the unnatural base pairs in duplex DNA, all unnatural and natural nucleosides were incorporated into the complementary oligonucleotides 5′-d(GCGTACXCATGCG)-3′ and 5′-d(CGCATG**Y**GTACGC)-3′

at the positions labeled **X** and **Y**. This sequence context was chosen to examine the effect of sequence context by comparing the effects of flanking pyrimidines (position **X**) and flanking purines (position **Y**). The melting temperature (T_m) of each duplex was determined by thermal denaturation experiments (Table 1). The term "stability" is used here only to refer to T_m values, unless otherwise indicated. The unnatural nucleotides were examined both as self-pairs and as heteropairs. For comparison, the T_m for a duplex containing a natural base pair (**X**:**Y** = dA:dT) is 59.2 °C.

Generally, the least substituted unnatural base pairs were the least stable. Accordingly, the **BEN** self-pair was the least stable, with a T_m of only 52.8 °C. The heteropairs formed between **BEN** and a monosubstituted derivative (**MM1**, **MM2**, or **MM3**) were only slightly more stable, and independent of sequence context, with a constant T_m of approximately 53 °C. The heteropairs formed between **BEN** and more highly substituted analogues were generally more stable, and in these cases the stability did depend on the sequence context. With **BEN** at position **X** (flanked by pyrimidines) and the more substituted analogue at position **Y** (flanked by purines), the pairs showed stability that increased with substitution, culminating in the **BEN**:**TM2** pair, only 3.7 °C less stable than a dA:dT base pair. Each pair was slightly less stable in the opposite sequence context (**Y** = **BEN**) in which the more substituted analogue was flanked by pyrimidines. The sequence dependence suggests that, with these pairs, intrastrand packing is more important than interstrand packing.

The self-pairs formed among the monosubstituted analogues, **MM1**, **MM2**, and **MM3**, are more stable than the **BEN**:**BEN** self-pair, by 1.4, 0.5, and 1.3 °C, respectively. **MM2** forms the
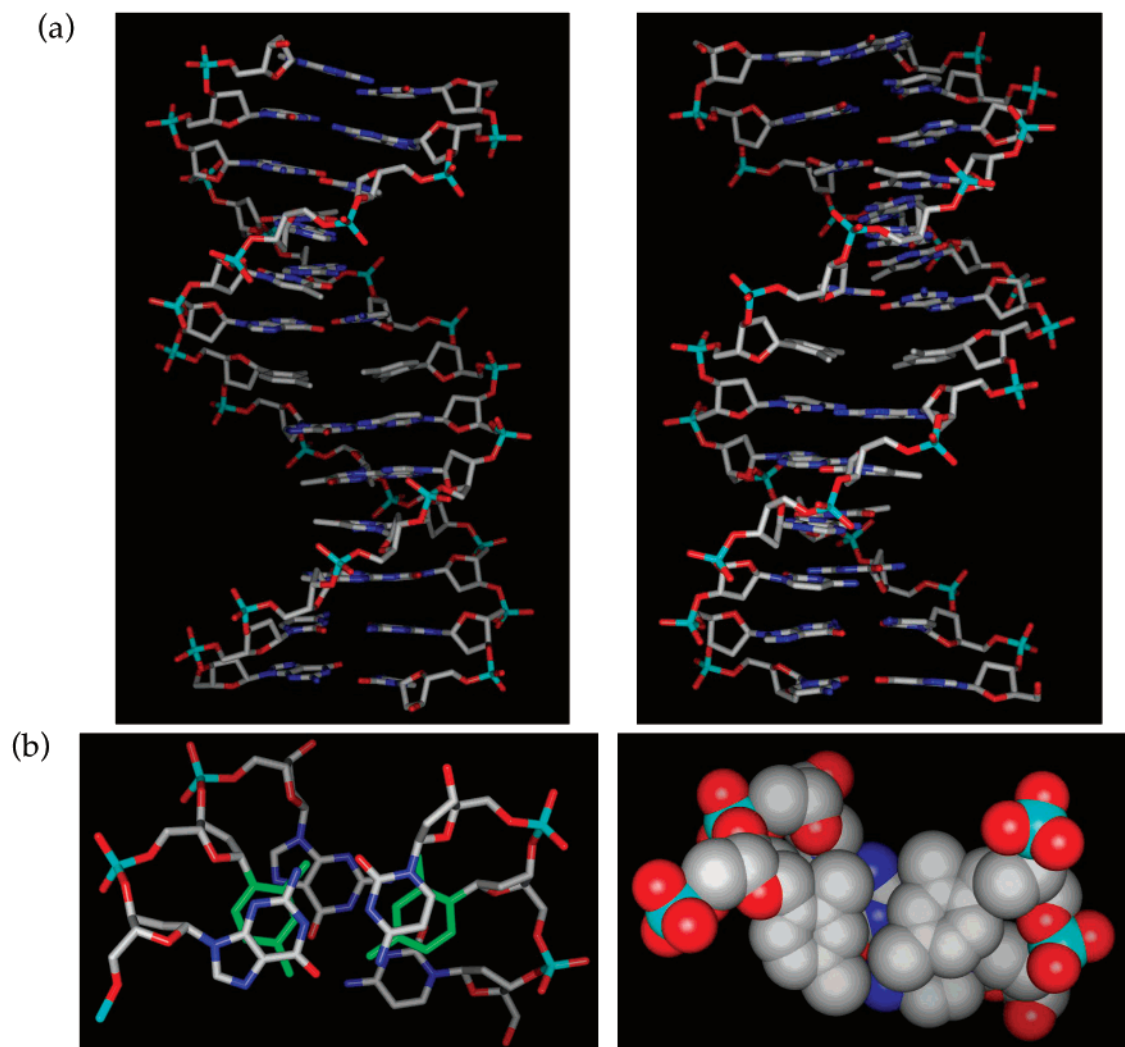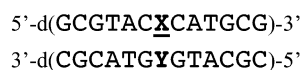
(a)

(b)

Figure 4. Energy-minimized model structure of the duplex containing the **DM5** self-pair. (a) View from major groove (left) and from minor groove (right). (b) Intrastrand stacking of the **DM5** (shown in green) self-pair (left). The **DM5** (gray) self-pair shown as a CPK model (right).

Table 1. T_m Values (°C) for Duplex Containing Unnatural Base Pairs[a]

5'-d(GCGTAC**X**CATGCG)-3'

3'-d(CGCATG**Y**GTACGC)-5'

Y	X											
	BEN	MM1	MM2	MM3	DM	DM2	TM	DM3	DM4	DM5	TM2	TMB
BEN	52.8	53.3	53.0	53.0	53.3	53.1	53.1	54.2	53.1	53.7	54.1	54.0
MM1	53.1	54.2	53.1	54.1	53.2	54.2	54.3	55.0	54.1	55.1	54.2	54.3
MM2	53.0	54.1	53.3	54.0	54.0	54.1	54.2	54.2	54.3	55.0	54.3	54.0
MM3	53.2	54.3	53.2	54.1	54.2	54.0	54.1	55.1	55.1	55.2	56.2	54.0
DM	53.2	54.0	54.0	54.0	53.7	54.1	54.2	55.2	54.0	54.1	55.2	54.2
DM2	53.0	54.1	54.0	54.0	54.3	54.5	55.1	55.1	54.1	55.6	55.1	55.3
TM	54.1	54.1	53.2	54.0	54.2	54.2	55.2	55.1	56.1	56.2	57.1	56.0
DM3	54.7	55.2	54.2	55.1	55.0	55.7	56.1	55.7	56.2	55.7	56.2	55.0
DM4	54.5	55.2	54.2	55.1	55.1	54.6	56.1	56.1	56.1	56.6	57.0	56.1
DM5	54.6	55.3	55.0	55.2	55.2	55.6	56.1	58.1	56.1	57.3	57.0	56.2
TM2	55.5	56.1	55.1	57.2	55.1	56.5	58.0	57.0	57.0	57.5	57.2	57.2
TMB	55.1	55.0	54.2	56.1	54.1	55.2	57.2	56.0	56.0	57.1	57.1	57.2

[a] Determined in 10 mM PIPES, 10 mM MgCl$_2$, 100 mM NaCl, pH 7.0. For reference, T_m = 59.2 °C for **X** = dA, **Y** = dT.

least stable self-pair, and it also forms less stable heteropairs when flanked by pyrimidines. However, the stabilities of all nine of these self-pairs and heteropairs differ by only ~1 °C. When a monosubstituted analogue is paired opposite an analogue with greater methyl substitution, the base pair is stabilized. These pairs also display a sequence dependence that

is similar to that observed with the **BEN** pairs: the pairs tend to be more stable when the more substituted analogue is flanked by purines. This again indicates that stacking between the unnatural analogue and the flanking bases makes an important contribution to the stability of the unnatural base pair. However, these more substituted pairs also begin to show a dependence

on the specific substitution pattern. For example, when paired opposite **DM5**, each of the monosubstituted analogues forms pairs with virtually identical stability (∼55 °C), but the pairs formed with the most highly substituted **TM2** or **TMB** differ by nearly 2 °C. This indicates that specific interstrand interactions between the unnatural analogues also contribute to stability.

The stabilities of the self-pairs formed between disubstituted analogues are more variable. The **DM3**, **DM4**, and **DM5** self-pairs are more stable ($T_m = 55.7-57.3$ °C) than the **DM** and **DM2** self-pairs ($T_m \approx 54$ °C). This is also generally true of the corresponding heteropairs: those formed between **DM3**, **DM4**, and **DM5** are only 2−3 °C less stable than a dA:dT in the same sequence context. The reduced stability of the pairs involving **DM** and **DM2** may be understood in the context of the structural model described above. Assuming that each nucleobase analogue is oriented similarly to **DM5** in the model, it seems unlikely that the two methyl groups of **DM** and **DM2** could simultaneously pack with the pairing base. The increased dependence of stability on substitution pattern suggests that, in addition to intrastrand packing, interstrand packing interactions contribute to molecular recognition within these moderately substituted unnatural nucleobases.

The unnatural pairs involving the more highly substituted analogues are generally quite stable, typically only 3 °C less stable than a dA:dT in the same sequence context. The T_m values for the **TM** and **TM2** self-pairs are 55.2[6] and 57.2 °C, respectively. This again demonstrates the importance of the substitution pattern at the putative interstrand interface (i.e., all three methyl groups can simultaneously pack at the interface only with **TM2**). The **TMB** self-pair shows no increase in stability relative to the **TM2** self-pair, again indicating that methyl groups are most stabilizing when they are positioned to pack the other analogue. Further supporting the importance of interstrand interactions, heteropair stability varied widely depending on the specific substitution pattern of the unnatural nucleobases. The T_m values range between 53.1 and 58.0 °C. The stabilities of the self-pairs and heteropairs formed with **TM**, **TM2**, and **TMB** indicate that these nucleobases generally have sufficient size to contribute to an interstrand interface that is sufficiently well packed to stabilize the unnatural base pairs.

Interestingly, the two most stable base pairs are not the most substituted, again demonstrating that the interstrand interactions become specific with a sufficiently packed interface. Duplex DNA containing the **TM**:**TM2** and **DM3**:**DM5** pairs have T_m values of 58.0 and 58.1 °C, respectively. These unnatural base pairs are only ∼1.2 °C less stable than a dA:dT pair in the same sequence context. Both pairs are formed by nucleobases that have an *ortho*-methyl substituent. These methyl groups may pack with each other in the minor groove, as it is assumed that the nucleobases adopt an orientation in duplex DNA that minimizes repulsion between any *ortho*-methyl group and the sugar (Figure 4). Both pairs also contain a single *meta*-substituted nucleobase, which is likely oriented toward the pairing nucleobase. The stability of **DM3**:**DM5** is especially surprising, considering that these nucleobases have only two methyl substituents. This unnatural base pair may be optimized for interstrand packing.

2.4. Stabilities of Mispairs between Unnatural and Natural Bases. It is important that the unnatural nucleobases are thermally selective against mispairing with the natural bases.

Table 2. T_m Values (°C) for Duplex Containing Self-Pairs or Mispairs[a]

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

X	Y	T_m (°C)	X	Y	T_m (°C)
BEN	**BEN**	52.8	**MM1**	**MM1**	54.2
	C	45.0		C	45.0
	G	49.1		G	49.0
	T	47.6		T	46.1
	A	49.2		A	50.1
MM2	**MM2**	53.3	**MM3**	**MM3**	54.1
	C	45.0		C	46.2
	G	49.0		G	50.0
	T	48.2		T	48.1
	A	49.2		A	50.1
DM	**DM**	53.7	**DM2**	**DM2**	54.5
	C	43.8		C	44.0
	G	48.7		G	49.1
	T	48.2		T	48.1
	A	48.5		A	50.2
DM3	**DM3**	55.7	**DM4**	**DM4**	56.1
	C	46.0		C	47.3
	G	49.0		G	50.0
	T	48.1		T	49.1
	A	49.0		A	50.1
DM5	**DM5**	57.3	**TM**	**TM**	55.2
	C	47.2		C	44.7
	G	50.0		G	50.0
	T	49.0		T	49.2
	A	49.2		A	51.7
TM2	**TM2**	57.2	**TMB**	**TMB**	57.2
	C	47.0		C	45.2
	G	50.1		G	49.2
	T	49.1		T	49.0
	A	50.2		A	50.3

[a] Determined in 10 mM PIPES, 10 mM MgCl₂, 100 mM NaCl, pH 7.0.

The stabilities of all possible mispairs in one sequence context (**Y** = dG, dA, dC, or dT) were determined and are reported in Table 2. All of the mispairs are significantly less stable than even the least stable unnatural pair. This demonstrates the specificity inherent to these hydrophobic nucleobases, which is likely to be caused by forced desolvation of the natural bases. However, desolvation is not sufficient to explain the relative stabilities of the mispairs. For example, in the case of **BEN**, mispairs with dG and dA are the most stable, followed by dT and then dC. The relative stabilities do not reflect hydrophobic-ity, as dA and dT are the most hydrophobic.[38] Instead, it seems likely that intrastrand stacking underlies the stability of the purine mispairs, just as it contributes to the stability of natural base pairs.[37] Thus, similar to the unnatural pairs involving **BEN**, the stabilities of its mispairs appear to depend on intrastrand stacking of the pairing base. Nonetheless, even the most stable mispairs, **BEN**:dA and **BEN**:dG, are almost 4 °C less stable than the **BEN** self-pair.

Similarly, mispairs between the monosubstituted analogues (**MM1**, **MM2**, and **MM3**) and dC are least stable, and those involving dG and dA are most stable. The stabilities were similar to those with **BEN**, implying an absence of specific interactions. The only exceptions were the **MM1**:dA and **MM1**:dT mispairs, which are slightly more and less stable, respectively, than the same mispairs with **MM2** or **MM3**. Assuming that the *ortho*-methyl group is positioned in the minor groove (see above), this may be the result of a stabilizing (dA) or a destabilizing

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

(dT) interaction in the minor groove. Similar interactions have previously been evoked to explain the selective recognition of dA in duplex DNA, and during polymerase-mediated base pair synthesis.[7] Most importantly, all mispairs were significantly less stable than the "correct" unnatural pairs.

Mispairs between the disubstituted analogues (**DM**, **DM2**, **DM3**, **DM4**, or **DM5**) and dC are all less stable than their mispairs with the other natural bases. In addition, the mispair stability varied by 3.5 °C, depending on the substitution pattern of the unnatural base. This greater range in stability results largely from the smaller T_m values of the **DM**:dC and **DM2**:dC mispairs. These were the least stable mispairs of all, including the mispairs with **BEN** or with the monosubstituted analogues. This may again result from **DM** and **DM2** being unable to present both methyl substituents to the pairing base, as was discussed above in relation to the unnatural pairs involving these two analogues. However, the second methyl group not only fails to stabilize the mispair, but actually destabilizes it by ~1 °C (compare **MM1**:dC to **DM2**:dC and **MM2**:dC to **DM**:dC). This may result from the size and hydrophilicity of dC exacerbating the poor packing of **DM** and **DM2**. The stabilities of the mispairs between a disubstituted analogue and dG, dA, or dT were all greater and less variable than the mispairs with dC. Again, all of the mispairs between an unnatural and a natural base are less stable than the "correct" unnatural base pairs.

Mispairs involving the three most highly substituted analogues, **TM**, **TM2**, and **TMB**, are also least stable in the case of dC. However, unlike the less-substituted analogues, the stabilities of the mispairs with dA are generally most stable, although the effect is small with **TM2**. This may result from specific minor groove interactions (**TM**, **TM2**, and **TMB** all have a methyl group that is expected to be oriented into the minor groove) and/or from desolvation. Interestingly, unlike the unnatural base pairs discussed above, the stability of the mispairs does not increase with increasing substitution; mispair stability does not exceed ~50 °C. Preferential stabilization of the unnatural pairs, such as **TM**:**TM2** and **DM3**:**DM5**, results in thermal selectivities in excess of 8 °C, which compare favorably with those seen among the natural base pairs. Even with this remarkably simple framework, hydrophobic and van der Waals forces are sufficient to endow the unnatural base pairs with stability and thermal selectivity

2.5. Further Characterization of the DM5 Self-Pair. As discussed above, several of the unnatural base pairs, including the **DM5** self-pair, are remarkably stable and selective in the above sequence context. We were interested in determining the generality of these characteristics and thus determined the T_m values of DNA containing the **DM5** self-pairs in a variety of different sequence contexts (Table 3). The sequences included multiple self-pairs separated by one to five natural pairs, or runs of up to three contiguous self-pairs. While destabilization varied between 1.9 and 5.1 °C per unnatural pair, relative to a dA:dT pair, the melting curves of each duplex showed a well-defined two-state transition (Supporting Information), and the CD spectrum of each indicates a typical B-form duplex (Figure 5). The duplex containing a triplet run is of particular interest. In this case, the DNA duplex is only destabilized by 3.3 °C per self-pair relative to an all-natural sequence, and even in this case the self-pairs remain strongly selective against mispairing with the natural bases (Table 4); the sequences containing the

Table 3. T_m Values (°C) of Duplex with Multiple **DM5** Self-Pairs[a]

Duplex	Sequence	T_m	ΔT_m/self pair.
1	5'-d(GCGTACACATGCG)-3' 3'-d(CGCATGTGTACGC)-5'	59.2	
2	5'-d(GCGTACAAATGCG)-3' 3'-d(CGCATGTTTACGC)-5'	58.2	
3	5'-d(GCGTAC**X**CATGCG)-3' 3'-d(CGCATG**X**GTACGC)-5'	57.3	-1.9[b]
4	5'-d(GCGTAC**XX**ATGCG)-3' 3'-d(CGCATG**XX**TACGC)-5'	52.2	-3.0[c]
5	5'-d(GCGTAC**X**C**X**TGCG)-3' 3'-d(CGCATG**X**G**X**ACGC)-5'	49.2	-5.0[b]
6	5'-d(GCGTAC**X**CA**X**GCG)-3' 3'-d(CGCATG**X**GT**X**CGC)-5'	49.1	-5.1[b]
7	5'-d(GCG**X**AC**X**CA**X**GCG)-3' 3'-d(CGC**X**TG**X**GT**X**CGC)-5'	44.0	-5.1[b]
8	5'-d(GCG**X**ACACA**X**GCG)-3' 3'-d(CGC**X**TGTGT**X**CGC)-5'	52.0	-3.6[b]
9	5'-d(GCGTAAAAATGCG)-3' 3'-d(CGCATTTTTACGC)-5'	57.1	
10	5'-d(GCGTA**XXX**ATGCG)-3' 3'-d(CGCAT**XXX**TACGC)-5'	47.2	-3.3[d]

[a] Determined in 10 mM PIPES, 10 mM MgCl₂, 100 mM NaCl, pH 7.0. [b] As compared to duplex 1. [c] As compared to duplex 2. [d] As compared to duplex 9.

corresponding dC:dC, **DM5**:dC, or dC:**DM5** mispairs failed to form duplex DNA under all conditions. We conclude that the self-pair will be accommodated in an arbitrary sequence context with reasonable stability and selectivity.

To further analyze **BEN** and **DM5** self-pair formation, the free energy change on duplex formation was calculated and partitioned into enthalpic and entropic components by analysis of the melting curves at different DNA concentrations[39] (Table 5 and Supporting Information). The $\Delta G_{298K}°$ values of duplex formation are consistent with those expected from the T_m values. The data imply that the reduced stability of the duplexes containing the self-pairs does not result from entropic effects, but rather from enthalpic effects. The **BEN** and **DM5** self-pairs both show a less favorable enthalpy change upon duplex formation, relative to a natural dA:dT pair in the same sequence context. This may result from the absence of H-bonds and/or the significantly reduced surface area of the nucleobase analogues. In contrast, self-pair formation is accompanied by a more favorable (less negative) entropy change. Interestingly, the entropy change is also responsible for the increased stability of the **DM5** self-pair relative to the **BEN** self-pair. This implies that the classical hydrophobic effect contributes significantly to the stability of the unnatural base pairs (see below).

3. Discussion

The storage and replication of biological information is based on complementary interactions between the natural nucleobases, dA with dT and dG with dC. Expansion of the genetic alphabet by the introduction of a third base pair would enable a wide variety of biotechnology applications based on improved or

(39) Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601−1620.
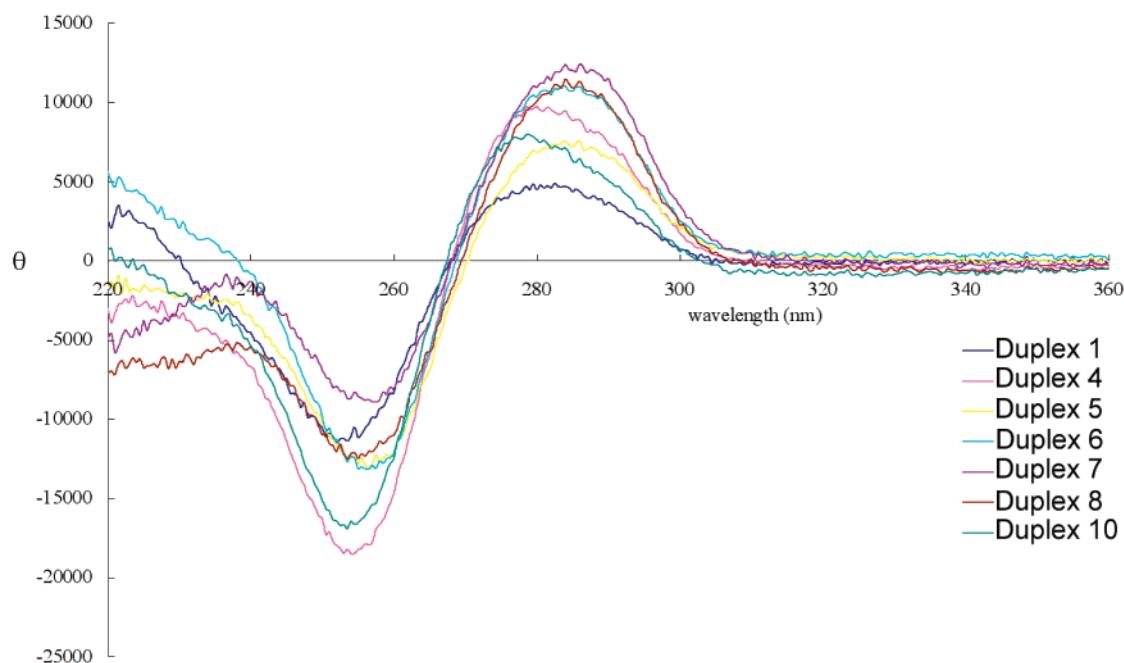
Figure 5. CD spectra of DNA duplexes containing multiple **DM5** self-pairs. See text and Table 3 for details.

Table 4. T_m Values (°C) of Duplex with Multiple DM5 Self-Pairs and Mispairs[a]

Duplex	Sequence	T_m	ΔT_m/self pair.
1	5'-d(GCGTAAAAAATGCG)-3' 3'-d(CGCATTTTTTACGC)-5'	57.1	
2	5'-d(GCGTA**XXX**ATGCG)-3' 3'-d(CGCAT**XXX**TACGC)-5'	47.2	-3.3
3	5'-d(GCGTA**CCC**ATGCG)-3' 3'-d(CGCAT**CCC**TACGC)-5'	nd[b]	
4	5'-d(GCGTA**XXX**ATGCG)-3' 3'-d(CGCAT**CCC**TACGC)-5'	nd[b]	
5	5'-d(GCGTA**CCC**ATGCG)-3' 3'-d(CGCAT**XXX**TACGC)-5'	nd[b]	
6	5'-d(GCGTA**CXX**ATGCG)-3' 3'-d(CGCAT**XXX**TACGC)-5'	45.4	
7	5'-d(GCGTA**CXC**ATGCG)-3' 3'-d(CGCAT**XXX**TACGC)-5'	36.4	
8	5'-d(GCGTA**CXCX**TGCG)-3' 3'-d(CGCAT**XXX**TACGC)-5'	34.7	

[a] Determined in 10 mM PIPES, 10 mM $MgCl_2$, 100 mM NaCl, pH 7.0.
[b] Cooperative melting behavior was not detected.

Table 5. Thermodynamic Parameters for Duplex Formation

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

	T_m (°C)	$\Delta H°$ (kcal mol^{-1})	$\Delta S°$ (cal K^{-1} mol^{-1})	$\Delta G_{298K}°$ (kcal mol^{-1})
X = A, Y = T	59.2	−95.1	−259.7	−17.7
X = Y = BEN	52.8	−91.6	−254.5	−15.7
X = Y = DM5	57.3	−83.1	−225.5	−16.0

A wide variety of stable unnatural base pairs were identified; however, they were generally not good substrates for DNA polymerases. Some were efficiently synthesized (by incorporation of the unnatural triphosphate opposite the unnatural base in the template), but generally the nascent unnatural pair was not a good substrate for continued primer extension. We have found that continued synthesis may be made more efficient by introducing heteroatoms such as nitrogen, oxygen, or sulfur.[4,10] However, preliminary structural studies indicate that pairs formed between two large aromatic nucleobase analogues are prone to interstrand intercalate, and we suspect that a similar structural distortion at the primer terminus might inhibit continued synthesis. It appeared that these first generation unnatural base pairs achieved stability at the cost of replicability. We have thus been interested in examining second generation base pairs that are both modified with heteroatoms and also too small to interstrand intercalate. However, it was unclear whether this last design criterion could be accomplished while simultaneously maintaining reasonable base pair stability and thermal selectivity. To systematically evaluate this issue, we synthesized and characterized a series of unnatural nucleosides bearing simple methyl-substituted benzene rings as nucleobase analogues.

3.1. Stability and Selectivity of the Unnatural Base Pairs. Without substitution, the unnatural base pair formed between two benzene rings is less stable than a typical mispair among the natural bases (i.e., the T_m for a dT:dG mispair in the same sequence context is 53.3 °C). This demonstrates that, as

altered hybridization or replication properties, and also the encoding of additional information for both in vitro and in vivo applications. We have examined a wide variety of candidate unnatural base pairs, including both self-pairs formed between identical unnatural nucleotides and heteropairs formed between two different unnatural nucleotides, whose interstrand interactions are largely based on hydrophobic and van der Waals forces. At the time of our first publication, no predominantly hydrophobic base pairs had ever been found to stabilize duplex formation. Thus, much of our initial efforts focused on the design of unnatural nucleotides having a large aromatic surface area to ensure base pair stability via intrastrand stacking interactions.

expected, two simple benzene rings are not able to efficiently fill the space occupied in the natural case by a purine—pyrimidine pair. However, even the least stable unnatural pairs are thermally selective, by almost 4 °C, against mispairing with the natural bases. This selectivity is comparable to that observed among the natural bases. With increasing substitution, the unnatural base pairs become progressively more stable. Remarkably, the most stable base pairs, **TM**:**TM2** and **DM3**:**DM5**, are only ∼1.2 °C less stable than a dA:dT pair in the same sequence context. In addition, several more of the duplexes are only slightly less stable than natural duplex DNA. While the unnatural pairs generally become more stable with substitution, the mispairs show little increase in stability, resulting in thermal selectivities of greater than 8 °C. This selectivity is actually greater than that typically observed among natural nucleobases. It is thus apparent that even base pairs with both reduced surface area and no ability to H-bond can form stable and selective pairs within duplex DNA.

Efforts to develop an unnatural base pair require that the base pair be accommodated in DNA in a sequence-independent manner. We thus characterized the stability of the **DM5** self-pair in a variety of sequence contexts, including those with multiple self-pairs. While each duplex was synergistically destabilized (the destabilization resulting from multiple substitutions is not simply additive), each sequence still forms a B-form duplex by CD spectroscopy, demonstrates a well-behaved two-state transition during melting, and is significantly more stable than the corresponding duplexes containing mispairs. This implies, at least for **DM5**, that most sequence contexts should selectively accommodate the self-pair.

3.2. Origins of Base Pair Stability. It is interesting to speculate about how nucleobase analogues with such limited surface area, and no ability to form H-bonds, can pair with reasonable stability and thermal selectivity. In general, dipole and dipole—induced dipole are all known to contribute to stable intrastrand stacking in natural DNA.[37] However, because the unnatural analogues examined in this study have neither large permanent dipole moments nor significant polarizability, it seems unlikely that these electrostatic interactions contribute significantly to their pairing upon duplex formation. The hydrophobic effect is also thought to contribute to base stacking interactions in natural DNA.[37] In this case, the driving force would result from the unnatural base having greater solvent exposure in single-stranded DNA, and thus greater solvent ordering (solvent surrounding hydrophobic moieties such as the unnatural nucleobase is thought to become structured in a fashion similar to ice clathrate formation[37]). Duplex formation reduces the exposed hydrophobic surface area and thus decreases the water ordering. This decrease in order increases the entropy of the system, which should be apparent in the thermodynamic data, if the hydrophobic effect contributes significantly to unnatural base pair formation.[40−43] When the entropy change associated with formation of a dA:dT pair is compared to that of **BEN** and **DM5** self-pairs, it is apparent that unnatural base pair formation is associated with a significantly more positive

entropy change. In addition, the larger effect associated with the **DM5** self-pair is consistent with a dominant contribution from the hydrophobic effect, as in this case duplex formation buries more hydrophobic surface area than with **BEN**. We thus suggest that pairing of the unnatural nucleobases is mediated predominantly by the hydrophobic effect.

In addition to the hydrophobic effect, specific intra- and interstrand packing interactions appear to also contribute to unnatural base pair stability. A structural model, based on CD spectroscopy and molecular modeling, suggests that the unnatural base pairs pack within an undistorted B-form duplex and orient any *ortho*-substituent into the minor groove. This structural model allows for the identification of the substituents of one nucleobase analogue that are likely to be well packed by flanking bases, and thus contribute to intrastrand packing, or oriented toward the other nucleobases, and are thus likely to contribute to "edge—edge" or interstrand packing. The contribution of these interactions is expected to be made manifest as a dependence of the T_m values on the sequence (intrastrand packing) or on the specific substituent pattern of the nucleobase analogues (interstrand packing). The stability of the pairs with **BEN** is illustrative. When **BEN** is positioned between purines and the pairing base (packed by pyrimidines) is varied, the T_m values change by less than 1 °C. However, when the pairing base is packed by two purines, substitution results in T_m values that vary by almost 3 °C. An origin of this sequence dependence is suggested by the structural model of the **DM5** self-pair (Figure 4). The model predicts that flanking pyrimidines are of insufficient size to efficiently pack with multiple methyl group substituents of the unnatural nucleobases. Thus, the favorable hydrophobic packing may not be fully realized when the unnatural nucleotide is flanked by a dC or a dT. This sequence dependence and specific substitution pattern independence of the pairs with **BEN** implies that the interface between the unnatural nucleobases is not sufficiently developed to mediate specific interstrand interactions.

When both pairing analogues are methyl-substituted, there remains sequence dependence, especially when there is a large discrepancy in the level of substitution; the pairs are generally more stable when the more substituted base is flanked by purines. However, the stability of the more highly substituted unnatural base pairs also depends on the specific substitution pattern. This is apparent from the variable stabilities of the heteropairs formed between the di-, tri-, and tetra-substituted analogues. The dependence on specific interactions is also demonstrated by the stabilities of the self-pairs, which are by definition sequence independent. For example, in the case of the disubstituted analogues, the stabilities of the five self-pairs vary by 3.6 °C. In addition, specific interactions must underlie the stabilities of the **TM**:**TM2** and **DM3**:**DM5** pairs, which are each significantly more stable than other pairs, including those with more substituted analogues. Thus, at least when the interstrand interface is developed with multiple methyl groups, the stability and thermal selectivity appear to result from a combination of the classical hydrophobic effect, intrastrand packing, and specific interactions between the nucleobase analogues. A well-packed interstrand interface that contributes to unnatural base pair stability is supported by the energy-minimized **DM5** self-pair model structure, where the two **DM5** surfaces approach van der Waals contacts. Despite their reduced

(40) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863−1872.
(41) Guckian, K. M.; Morales, J. C.; Kool, E. T. *J. Org. Chem.* **1998**, *63*, 9652−9656.
(42) Brotschi, C.; Haberli, A.; Leumann, C. J. *Angew. Chem., Int. Ed.* **2001**, *40*, 3012−3014.
(43) Brotschi, C.; Leumann, C. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 1655−1658.

surface area, the interstrand interactions of these unnatural base pairs may be optimized for molecular recognition by the judicious placement of methyl groups.

The stability of the duplex containing three adjacent unnatural base pairs merits additional comment. A significant component of duplex stability in this case must derive from both inter- and intrastrand packing of the unnatural nucleobases with each other. Recently, Shionoya and colleagues elegantly demonstrated that a run of five hydroxypyridone self-pairs may mediate duplex formation via Cu^{2+} ligation.[44] While duplex stability and selectivity were not reported, it seems likely that, in addition to interstrand interactions mediated by bidentate metal ligation, interstrand packing interactions between the unnatural base also contribute to stability in this case. Thus, favorable intrastrand packing among small predominantly hydrophobic moieties in DNA may be general, and, when combined with suitable interstrand interactions, for example, those mediated by metal chelation or by hydrophobic packing, runs of unnatural base pairs may form extended, stable duplexes.

3.3. Conclusions and Implications for the Expansion of the Genetic Alphabet. It is well known that increased aromatic surface area may stabilize duplex formation. However, no studies involving the systematic optimization of interstrand packing interactions had been reported, so surface area and H-bonding requirements remained unclear. We have now shown that unnatural nucleotides bearing simple methyl-substituted benzene rings as nucleobase analogues may form base pairs that are virtually as stable as natural pairs, although they possess neither H-bonds nor large aromatic surface area. In fact, these unnatural base pairs are stable and selective despite having significantly less surface area than the natural base pairs. The molecular recognition of these base analogues may be tuned and optimized by the judicious substitution of methyl groups. These smaller nucleobases are not expected to induce distortions in duplex DNA, including those at the primer terminus that seem to limit replication of the larger unnatural base pairs. It seems likely that these nucleobase analogues will serve as scaffolds for the development of unnatural base pairs with further improved properties. For example, heteroatom substitution patterns should exist that further stabilize the unnatural pairs by dipole and induced-dipole interactions. In fact, fluorine substitution has recently been shown to have a pronounced effect on unnatural base pair replication.[45] We are now focused on identifying combinations of methyl group substitutions and heteroatom derivatizations that will impart these smaller unnatural nucleobases with further improvements in stability and replication.

4. Experimental Section

Oligonucleotide Synthesis. Oligonucleotides were prepared by the β-cyanoethylphosphoramidite method on controlled pore glass supports (1 μmol) by using an Applied Biosystems, Inc. 392 DNA/RNA synthesizer as the standard method. After automated synthesis, the oligomers were cleaved from the support by concentrated aqueous ammonia for 1 h, deprotected by heating at 55 °C for 12 h, and purified by polyacrylamide gel electrophoresis. The concentration of oligodeoxynucleotides was determined by UV.

Circular Dichroism Measurements. CD experiments were performed with an Aviv model 61 DS spectropolarimeter equipped with a Peltier thermoelectric temperature control unit (3 μM strand concentration, 10 mM PIPES buffer, pH 7.0, 100 mM NaCl, and 10 mM $MgCl_2$). The data were collected using a 1 cm path length quartz cuvette with scanning from 360 to 220 nm, a time constant of 3 s, and a wavelength step size of 0.5 nm, at 25 °C. CD data were transformed into molar ellipticity in the units of deg cm^2/dm of monomer subunits.

Duplex Melting Temperature Measurements. UV melting experiments were carried out by means of a Cary 200 Bio UV—visible spectrometer. The absorbance of the sample (3 μM strand concentration, 10 mM PIPES buffer, pH 7.0, 100 mM NaCl, and 10 mM $MgCl_2$) was monitored at 260 nm from 21 to 80 °C at a heating rate of 0.5 °C per min. Melting temperatures were determined from the derivative method using the Cary Win UV thermal application software.

Modeling Studies. Proposed duplex structures containing the **DM5** self base pair were obtained by molecular modeling of the 13-mer duplex 5′-d(GCGTAC**X**CATGCG)-3′/5′-d(CGCATG**Y**GTACGC)-3′ (**X** = **Y** = **DM5**). The initial structure of the duplex was obtained by manually docking **DM5** to the sugar C1′ position. Energy minimization of the duplex was carried out by using an Amber* force field with GB/SA solvation treatment of water. Each self-pair orientation (*anti*:*anti*, *anti*:*syn*, *syn*:*anti*, and *syn*:*syn*) was evaluated.

Supporting Information Available: Details of the synthetic procedures, characterizations, melting curves, and plots used for the determination of thermodynamic parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

JA047291M

(44) Tanaka, K.; Tengeiji, A.; Kato, T.; Toyama, N.; Shionoya, M. *Science* **2003**, *299*, 1212−1213.
(45) Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C.; Geierstanger, B. H.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 6923−6931.