Published online 10 June 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 14* ***4757–4763***
doi:10.1093/nar/gkp467

The effects of unnatural base pairs and mispairs on DNA duplex stability and solvation

Gil Tae Hwang[1], Yoshiyuki Hari[2] and Floyd E. Romesberg[2,*]

[1]Department of Chemistry, Kyungpook National University, Daegu 702-701, Korea and [2]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Received March 18, 2009; Revised May 6, 2009; Accepted May 17, 2009

ABSTRACT

In an effort to develop unnatural DNA base pairs we examined six pyridine-based nucleotides, d3MPy, d4MPy, d5MPy, d34DMPy, d35DMPy and d45DMPy. Each bears a pyridyl nucleobase scaffold but they are differentiated by methyl substitution, and were designed to vary both inter- and intra-strand packing within duplex DNA. The effects of the unnatural base pairs on duplex stability demonstrate that the pyridine scaffold may be optimized for stable and selective pairing, and identify one self pair, the pair formed between two d34DMPy nucleotides, which is virtually as stable as a dA:dT base pair in the same sequence context. In addition, we found that the incorporation of either the d34DMPy self pair or a single d34DMPy paired opposite a natural dA significantly increases oligonucleotide hybridization fidelity at other positions within the duplex. Hypersensitization of the duplex to mispairing appears to result from global and interdependent solvation effects mediated by the unnatural nucleotide(s) and the mispair. The results have important implications for our efforts to develop unnatural base pairs and suggest that the unnatural nucleotides might be developed as novel biotechnological tools, diagnostics, or therapeutics for applications where hybridization stringency is important.

INTRODUCTION

Significant effort has been directed towards developing unnatural nucleotides that selectively pair within duplex DNA and during DNA replication in order to expand the genetic alphabet (1–9). Our approach has been based on the use of predominantly hydrophobic nucleotide analogs that interact within duplex DNA via hydrophobic and packing forces (4–6,10). Much effort has been focused on characterizing how these and other predominantly hydrophobic analogs are recognized by DNA polymerases, and this has led to many insights into DNA replication and helped to develop unnatural base pairs that are replicated and transcribed with efficiencies and fidelities that are beginning to approach those of a natural base pair (6,10–12). However, much less effort has been directed towards characterizing how such unnatural nucleotides impact other properties of a DNA duplex, such as structure, solvation, and/or hybridization fidelity (13,14).

Our studies of unnatural base pairs formed between nucleotides bearing predominantly hydrophobic nucleobase analogs have included the characterization of both self pairs (formed between two identical nucleotide analogs) and heteropairs (formed between different analogs) of isocarbostiryl-, napthyl- and azaindole-nucleotides (15–17). The unnatural pairs are typically stable, presumably due to their extended aromatic surface area and an intercalative mode of pairing (18). They are also typically formed with high selectivity against mispairing with the natural nucleotides, presumably due to forced desolvation of the natural hydrogen-bond (H-bond) donors and acceptors. More recently we explored the use of smaller phenyl- and pyridone-based nucleoside analogs, whose reduced size is expected to preclude intercalative base pairing (18–21). With suitable derivatization, we found that the phenyl nucleobase scaffold may be optimized for reasonably selective and stable pairing within duplex DNA (19), while the pyridone analogs are generally less stable and less selective, especially relative to mispairing with dG, presumably due to the formation of a minor groove H-bond (21). To identify more stable and orthogonal analogs, we have also examined the effect of aza substitution within the phenyl nucleobase scaffold (20). Preliminary analysis revealed that of the unsubstituted pyridyl-nucleotides, d**2Py** (Figure 1) forms unnatural pairs with higher stability and selectivity, including higher selectivity against mispairing with dG.

Here we report the thermodynamic analysis of oligonucleotides containing six d**2Py** analogs, d**3MPy**, d**4MPy**, d**5MPy**, d**34DMPy**, d**35DMPy** and d**45DMPy** (Figure 1). These analogs were designed to systematically examine the effect of mono- or di-methyl substitution at

*To whom correspondence should be addressed. Tel: +1 858 784 7290; Fax: +1 858 784 7472; Email: floyd@scripps.edu

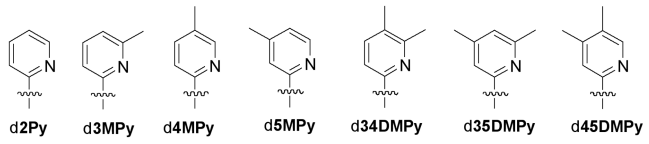

Figure 1. 2-Pyridine analogs. Sugar and phosphate backbone omitted for clarity.

each unique position of the nucleobase scaffold, excluding the 6-position due to the expected eclipsing interactions with the ribosyl oxygen (we assume that like a natural base pair, the unnatural nucleotides adopt an anti-geometry as defined by the 2-pyridyl nitrogen). The results reveal that the pyridine scaffold may be optimized for stable and selective pairing within DNA. Surprisingly, we also found that the optimized nucleotide, d**34DMPy**, hypersensitizes the duplex to mispairing at positions throughout its entire length when it is incorporated in both oligonucleotide strands, forming the self pair, as well as when it is incorporated in only one of the strands, paired opposite a natural nucleotide. This hypersensitization to mispairing appears to result from global and interdependent solvation effects that result from the presence of the unnatural nucleotide(s) and a mispair in the same duplex. The results have important implications for our understanding of the unnatural nucleotides and the base pairs they form, and they also suggest that unnatural nucleotides might have unexpected utility in different technologies where hybridization stringency is important.

MATERIALS AND METHODS

Synthesis of unnatural nucleotides and oligonucleotides

Unnatural nucleosides and nucleotides used in this study were synthesized as previously reported (22). 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. Natural oligonucleotides were either synthesized analogously or purchased from Integrated DNA Technologies (San Diego, CA, USA).

Thermal stability

UV melting experiments were carried out by means of a Cary 300 Bio UV-visible spectrophotometer. 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°C to 80°C at a heating rate of 0.5°C per min. Melting temperatures were determined via the derivative method using the Cary Win UV thermal application software.

Calculation of thermodynamic parameters and Δn_w

Thermodynamic parameters were determined by van't Hoff analysis (23): $T_m^{-1} = R[\ln([C_T]/4)]/\Delta H + \Delta S^\circ/\Delta H^\circ$, where ΔH° and ΔS° are the standard enthalpy and entropy changes determined from UV experiments, respectively, R is the universal gas constant and $[C_T]$ is the total strand concentration. The changes in the number of water molecules associated with the melting process, Δn_w, were obtained from the dependence of T_m on water activity (a_w) according to the equation $\Delta n_w = (-\Delta H/R)[\delta(T_m^{-1})/\delta(\ln a_w)]$ (24). The slope of the plot of reciprocal temperature (K^{-1}) of melting versus the logarithm of water activity at different concentrations (0, 2, 5, 7, 10, 12 and 15 wt%) of ethylene glycol was taken as the value of $\delta(T_m^{-1})/\delta(\ln a_w)$ (see Supplementary Data).

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.

RESULTS AND DISCUSSION

Each deoxynucleoside was synthesized as reported previously (22), and then converted to its phosphoramidite and incorporated into DNA at the positions labeled **X** and **Y** (Table 1) using standard procedures. The melting temperatures (T_m) of the duplexes were determined and are used to evaluate the stability of the unnatural self pairs and mispairs with each natural nucleotide (Table 1). For reference, the T_m of the duplex containing **X**:**Y** = dA:dT and dA:dG are 59.2 and 55.4°C, respectively, and the T_m of the analogous duplex containing the parent d**2Py** self pair (**X**:**Y** = d**2Py**:d**2Py**) is 52.2°C (20).

Methyl derivatization at the 3-, 4- or 5-position stabilizes the unnatural self pairs (Table 1). Substitution at the 3- or 4-position (d**3MPy** and d**4MPy**, respectively) has the largest effect, stabilizing the self pairs by 3 to 4°C, relative to the d**2Py** self pair, while substitution at the 5-position (d**5MPy**) is also stabilizing, but less so. The effects of substitution at the 3- and 4-positions are approximately additive, resulting in a duplex T_m for the d**34DMPy** self pair of 58.2°C. The d**34DMPy** self pair is more stable than any phenyl-, pyridone- or pyridyl-based self pair or heteropair identified to date, and is actually more stable than most pairs formed between analogs with much larger nucleobases (16,17,19–21). In fact, the d**34DMPy** self pair is nearly as stable as a dA:dT pair in the same sequence context (T_m = 59.2°C). The stability of the d**34DMPy** self pair is remarkable considering the limited stacking and H-bonding potential of its nucleobase, and suggests that the nucleobases are optimized to pack within the self pair in an edge-to-edge manner.

Methyl substitution also stabilizes the mispairs with natural bases, but generally less so than it stabilizes self pairing (Table 1). Mispair stabilization is maximal with

the two methyl groups of d**34DMPy** which, relative to the mispairs with d**2Py**, stabilize the mispairs with dG and dT by ~2.5°C, the mispair with dA by 1.4°C, and surprisingly, the mispair with dC by almost 5°C. The observed trend in mispair stability is opposite that of natural nucleobase hydrophobicity, which suggests that it is mediated, at least in part, by specific inter-strand interactions or solvation effects. Overall, the data reveal that the nucleotides are selective for self pairing, with the stability of each self pair at least 2.3–6.5°C higher than that of the most stable mispair. The d**34DMPy** self pair shows the greatest thermal selectively of 6.4–9.1°C, which compares favorably with that of the natural base pairs. For example, thermal selectivity with dA in the same sequence context is 2.7–10.8°C (25).

To better understand the origins of the stability and thermal selectivity of the d**34DMPy** self pair, we examined the free energy of duplex formation and deconvoluted it into enthalpic and entropic contributions (26,27) (Table 2). Relative to a dA:dT base pair, formation of the d**34DMPy** self pair is slightly less favorable enthalpically, but slightly more favorable entropically. Considering the lack of H-bonds within the self pair and its reduced aromatic surface area, it is remarkable that its formation is so similar enthalpically to a natural base pair.

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)
d**3MPy**	d**3MPy**	56.0	d**34DMPy**	d**34DMPy**	58.2
	dA	51.0		dA	51.6
	dC	47.6		dC	49.1
	dG	51.0		dG	51.8
	dT	49.6		dT	50.1
d**4MPy**	d**4MPy**	55.0	d**35DMPy**	d**35DMPy**	56.4
	dA	50.5		dA	50.6
	dC	47.3		dC	46.9
	dG	50.5		dG	50.9
	dT	49.3		dT	49.6
d**5MPy**	d**5MPy**	53.1	d**45DMPy**	d**45DMPy**	54.7
	dA	50.8		dA	50.7
	dC	46.1		dC	46.9
	dG	50.1		dG	50.5
	dT	48.8		dT	49.2

[a]Uncertainty in values is <0.1°C; see 'Materials and Methods' section for details.

Regardless of the specific forces underlying this stability, including possibly the hydrophobic effect and electrostatic interactions mediated by the aza substituent, the data again reinforce the suggestion that both inter- and intrastrand packing interactions are relatively well optimized for pairing. While increased dynamics of the self pair within the duplex, relative to a natural base pair (which is rigidified by inter-strand H-bonding) may contribute to its favorable entropy of pairing, considering its hydrophobicity, solvation is also likely to make an important contribution.

To further explore how solvation contributes to self pair stability, we used a method based on osmotic stressing (24,28) to determine the number of water molecules that are liberated from the duplex upon melting (Δn_w), and whether the number depends on the presence of the self pair. Using ethylene glycol as a cosolute, we found that 4.46 ± 0.26 water molecules per base pair are liberated from the DNA upon melting of the fully natural duplex (Table 2). This likely reflects the liberation of waters that are at least weakly ordered within the minor and major grooves and is in good agreement with values reported previously (28). Upon melting of the duplex containing the d**34DMPy** self pair, we found that 3.96 ± 0.21 water molecules per base pair are liberated. Although small, this difference suggests that less water is liberated during melting when the duplex contains a self pair. In all, the data suggest that the stability of the d**34DMPy** self pair results from optimized edge-to-edge inter-strand packing and from reduced ordering of water molecules.

To further understand the effects of the unnatural base pair on the properties of the DNA duplex, we next examined the effect of the d**34DMPy** self pair on hybridization fidelity. Clearly the presence of an unnatural nucleotide will impart an oligonucleotide with fidelity against hybridization with an all natural oligonucleotide. However, it is less clear how the presence of the unnatural base pair will impact the stability of the duplex with mispairs between natural bases at other positions (13,14). Thus, we introduced single mispairs into the duplex 5′-d($G_1C_2G_3T_4A_5C_6\mathbf{X}_7C_8A_9T_{10}G_{11}C_{12}G_{13}$):5′-d($C_{14}G_{15}C_{16}A_{17}T_{18}C_{19}\mathbf{Y}_{20}G_{21}T_{22}A_{23}C_{24}G_{25}C_{26}$) with $\mathbf{X}_7$:$\mathbf{Y}_{20}$ = dA:dT or d**34DMPy**:d**34DMPy**. The nucleotide dC_6 remained constant, and we determined the stability of the duplex containing dG_{21}, dC_{21}, dA_{21} or dT_{21} (Table 3). Mispairing between fully natural DNA ($\mathbf{X}_7$:$\mathbf{Y}_{20}$ = dA:dT) destabilizes duplex formation by 13.0 to 16.7°C. The most destabilizing mispair is dC_6:dC_{21}, followed by dC_6:dA_{21} and

Table 2. Thermodynamic parameters and Δn_w values for correctly paired DNA duplexes[a]

DNA duplexes	T_m (°C)	$-\Delta H^\circ$ kcal mol^{-1}	$-\Delta S^\circ$ cal·K^{-1} mol^{-1}	Δn_w
5′-d(GCGTAC**A**CATGCG) 3′-d(CGCATG**T**GTACGC)	59.2 ± 0.02	99.2 ± 5.2	271.8 ± 14.3	4.46 ± 0.26
5′-d(GCGTAC(**34DMPy**)CATGCG) 3′-d(CGCATG(**34DMPy**)GTACGC)	58.2 ± 0.11	96.2 ± 6.7	264.4 ± 18.6	3.96 ± 0.21

[a]See text and 'Materials and methods' section for details.

Table 3. T_m values for duplexes containing single mispairs[a]

5′-d(G C G T A C (**34DMPy**)CATGCG)
3′-d($N_{26}N_{25}N_{24}N_{23}N_{22}N_{21}$(**34DMPy**)GTACGC)

N_{21}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]	N_{22}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]	N_{23}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]
dA	43.7	1.3	dA	49.7	2.6	dA	58.2	
dC	41.9	−0.4	dC	44.7	5.5	dC	45.9	3.4
dG	58.2		dG	54.7	2.7	dG	50.4	2.7
dT	43.6	1.6	dT	58.2		dT	47.9	2.8
N_{24}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]	N_{25}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]	N_{26}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]
dA	46.5	2.2	dA	49.3	2.4	dA	53.7	2.1
dC	58.2		dC	49.1	3.4	dC	58.2	
dG	46.4	3.0	dG	58.2		dG	53.7	3.2
dT	47.6	2.5	dT	48.6	3.8	dT	53.7	2.8

[a]Uncertainty in values is less than 0.1°C; see 'Materials and methods' section for details.
[b]$\Delta\Delta T_m$ corresponds to the difference in destabilization resulting from the mispair in the shown sequence versus the destabilization resulting from the same mispair in the sequence with dA:dT replacing the unnatural base pair. Positive values correspond to a greater destabilization in the duplex with the unnatural base pair.

dC_6:dT_{21}. Interestingly, two of the three mispairs with dC_6 are more destabilizing in the duplex containing a self pair ($\mathbf{X}_7$:$\mathbf{Y}_{20}$ = d**34DMPy**:d**34DMPy**), by approximately 1.5°C ($\Delta\Delta T_m$ in Table 3), while the dC_6:dC_{21} mispair results in a similar destabilization.

The increased sensitivity to mispairing of the duplexes containing the self pair could result from either localized interactions between the unnatural and flanking nucleotides, or more delocalized changes within the duplex. To differentiate between these possibilities, we determined the effect of mispairing at other positions, including dG_1:dN_{26}, dC_2:dN_{25}, dG_3:dN_{24}, dT_4:dN_{23} and dA_5:dN_{22} (Table 3). Surprisingly, despite being insulated from the self pair by a natural base pair, the dA_5:dN_{22} mispairs were again more destabilizing when the duplex also contained the self pair than when it did not (by 2.6 to 5.5°C). In fact, the increased sensitivity to mispairing was also observed when the position of the mispair was even more insulated from the self pair (dG_1:dN_{26} to dT_4:dN_{23}). This is especially remarkable for the dG_1:dN_{26} mispairs, since they are located at the end of the duplex, a position typically thought to be less ordered (29,30), and thus less sensitive to mispairing. The data strongly suggest that the differences in sensitivity to mispairing are not the result of specific interactions between the mispairing and unnatural nucleotides, but rather the result of more global changes induced within the duplex by the self pair.

One global property of the duplex that could be affected by the introduction of the self pair is its conformation which might be altered by the self pair such that the duplex is more sensitive to mispairing. To examine this possibility, we used circular dichroism to characterize the correctly paired duplexes (with and without the self pair), as well as the analogous duplexes containing a dT_4:dC_{23} mispair (Figure 2). While the self pair induces slightly stronger and red-shifted absorptions, the spectra

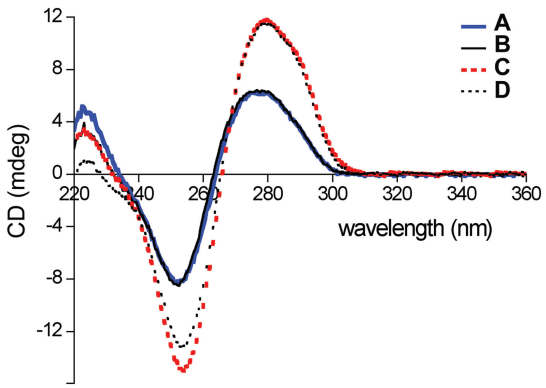

Figure 2. CD spectra of DNA duplexes. (Curve A) Fully natural and correctly paired duplex (N_4:N_{23} = dT:dA, X_7:Y_{20} = dA:dT); (curve B) Fully natural duplex containing a single mispair (N_4:N_{23} = dT:dC, X_7:Y_{20} = dA:dT); (curve C) Duplex containing a d**34DMPy** self pair and no mispairs (N_4:N_{23} = dT:dA, X_7:Y_{20} = d**34DMPy**:d**34DMPy**) and (curve D) Duplex containing a d**34DMPy** self pair and a single mispair (N_4:N_{23} = dT:dC, X_7:Y_{20} = d**34DMPy**:d**34DMPy**).

suggest that it does not significantly perturb duplex structure. Most importantly, mispairing has little effect on the spectrum of either duplex. Thus, we conclude that the self pair-mediated hypersensitivity to mispairing does not result from a global conformational change. This conclusion is consistent with previous studies which showed that different hydrophobic unnatural base pairs (18) or even a hydrophobic shape mimic of dT paired opposite dA (31) do not significantly perturb the structure of the duplex.

Another global property of the duplex that could be affected by the introduction of the d**34DMPy** self pair is its solvation. For example, the predominantly hydrophobic unnatural base pair may disrupt the network of waters of solvation within the minor and/or major grooves, and this disruption may then destabilize the entire network,

Table 4. Thermodynamic parameters and Δn_w values for mismatched DNA duplexes[a]

DNA duplexes	T_m (°C)	$-\Delta H^\circ$ kcal mol^{-1}	$-\Delta S^\circ$ cal K^{-1} mol^{-1}	Δn_w
5′-d(GCG**T**AC**A**CATGCG) 3′-d(CGC**C**TG**T**GTACGC)	50.3 ± 0.04	89.7 ± 3.7	250.8 ± 10.4	4.60 ± 0.36
5′-d(GCG**T**AC(**34DMPy**)CATGCG) 3′-d(CGC**C**TG(**34DMPy**)GTACGC)	45.9 ± 0.04	80.3 ± 6.3	225.1 ± 17.9	3.17 ± 0.26

[a]See text and 'Materials and Methods' section for details. The dT:dC mispair as well as the central dA:dT or d**34DMPy**:d**34DMPy** pairs are shown in bold.

Table 5. T_m values for duplexes containing single mispairs[a]

5′-d(GC G TA C A CATGCG)
3′-d($CN_{25}N_{24}AN_{22}G$(**34DMPy**)GTACGC)

N_{22}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]	N_{24}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]	N_{25}	T_m (°C)	$\Delta\Delta T_m$ (°C)[b]
dA	44.4	2.0	dA	52.3		dA	42.9	2.9
dC	38.8	5.5	dC	39.9	3.5	dC	43.1	3.5
dG	49.5	2.0	dG	43.8	3.4	dG	52.3	
dT	52.3		dT	41.8	3.0	dT	42.3	4.2

[a]Uncertainty in values is <0.1°C.
[b]$\Delta\Delta T_m$ corresponds to the difference in destabilization resulting from the mispair in the shown sequence versus the destabilization resulting from the same mispair in the sequence with dA:dT replacing the dA:d**34DMPy** pair. Positive values correspond to a greater destabilization in the duplex with the dA:d**34DMPy** pair. See 'Materials and methods' section for details.

which depends, at least in part, on H-bonding between waters (32–35). Indeed, such effects have been predicted computationally for similar hydrophobic unnatural base pairs (36). To examine this possibility, we again employed ethylene glycol-mediated osmotic stressing experiments (Table 4 and Supplementary Data). The data suggest that the introduction of a single natural mispair within the duplex has little effect on the number of water molecules liberated during the double to single-stranded transition ($\Delta n_w = 4.60 \pm 0.36$ versus 4.46 ± 0.26). In contrast, the introduction of the same mispair results in a significant decrease in the number of water molecules liberated by melting when the duplex also contains the self pair ($\Delta n_w = 3.17 \pm 0.26$). Overall, the data reveal that individually the self pair and the mispair each induce small or negligible decreases in duplex solvation, but when combined they cause a synergistic decrease in duplex solvation.

To further test this hypothesis, we determined the enthalpy and entropy of melting for the mispaired duplexes (Table 4). Mispairing within the fully natural duplex results in similar, although somewhat larger changes than the introduction of the self pair, decreasing the enthalpy but increasing the entropy of duplex formation. Consistent with the above conclusion, the effects were significantly larger when the mispair was introduced within a duplex containing the self pair. The decrease in enthalpy and the increase in entropy of duplex formation are consistent with a reduction in duplex solvation when both the self pair and the mispair are present. In all, the data suggest that the self pair hypersensitizes the duplex to mispairing by rendering it susceptible to a mispair-mediated collapse of the network of water molecules within its major and/or minor grooves. Such cooperative collapse of water networks, or 'clusters' is not without precedent and has already been observed during duplex dehydration (34).

The use of the d**34DMPy** self pair for increased hybridization fidelity might be useful for a variety of biotechnology applications, however it is limited to applications where the unnatural nucleotide is present in both hybridizing strands. The increased fidelity would be more generally useful for different applications if it were manifest with the unnatural nucleotide present in only one of the hybridizing strands, with the other strand being unmodified. Indeed, such effects have been observed with nucleotides bearing the hydrophobic 3-nitropyrrole nucleobase analog (13,14). To examine whether a single d**34DMPy** is also sufficient to mediate hypersensitization to mispairing, we characterized duplex DNA containing a single $\mathbf{X}_7$:$\mathbf{Y}_{20}$ = dA:d**34DMPy** mispair and a dC_2:dN_{25}, dG_3:dN_{24}, or dA_5:dN_{22} mispair (Table 5). Consistent with the proposed model, the central dA:d**34DMPy** significantly sensitized the duplex to mispairing, with each mispair being destabilized by a ~3°C relative to the fully matched sequence. This includes the dA_5:dG_{22} mispair, which destabilizes the fully natural duplex by only 0.8°C, but destabilizes the duplex containing the central dA:d**34DMPy** by 2.8°C. Thus, the presence of the unnatural nucleotide in one strand significantly increases

hybridization fidelity, including against mispairs that are relatively stable and otherwise difficult to discriminate against.

CONCLUSION

Despite its reduced size and limited H-bonding capacity, a single d**34DMPy** self pair is nearly as stable as a natural dA:dT pair in the same sequence context and the self pair does not appear to significantly perturb duplex structure. Thus, at least from a thermodynamic perspective, the self pair and heteropairs that it forms are promising as part of an expanded genetic alphabet. Previously, we noted that different predominantly hydrophobic unnatural base pairs are destabilized when included at multiple positions within a duplex. It seems likely that this results from the same solvation effects that appear to underlie the self pair mediated hypersensitization to mispairing. While this has important implications for the future design of unnatural base pairs, it also suggests that the unnatural nucleotides might find uses in different oligonucleotide-based technologies where hybridization fidelity is important, such as for 'zip-coding' PCR products for hybridization to oligonucleotide arrays (37). The hypersensitization associated with the presence of a single d**34DMPy** nucleotide may be particularly useful since the hybridizing strand is unmodified. This suggests that the unnatural nucleotide could be used to increase the fidelity of oligonucleotides used in oligonucleotide arrays or for therapeutic oligonucleotides that target natural DNA or RNA sequences (38,39). The continued exploration of these nucleotides promises not only to help identify promising unnatural base pairs, but also to help elucidate the forces underlying duplex stability and to identify analogs that may be used to optimize oligonucleotides for biotechnological or therapeutic applications where high fidelity hybridization and discrimination is critical.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Professor Charles Spink (SUNY Cortland) for kindly providing water activity values at various ethylene glycol cosolute concentrations.

FUNDING

National Institutes of Health [GM060005 to F.E.R.]; and The Uehara Memorial Foundation [to Y.H.]. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Leconte,A.M. and Romesberg,F.E. (2009) Engineering nucleobases and polymerases for an expanded genetic alphabet. In Köhrer,C. and RajBhandary,U.L. (eds), *Protein Engineering*, Springer-Verlag, Berlin, pp. 291–314.
2. Piccirilli,J.A., Krauch,T., Moroney,S.E. and Benner,S.A. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature*, **343**, 33–37.
3. Yang,Z., Sismour,A.M., Sheng,P., Puskar,N.L. and Benner,S.A. (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res.*, **35**, 4238–4249.
4. Leconte,A.M., Hwang,G.T., Matsuda,S., Capek,P., Hari,Y. and Romesberg,F.E. (2008) Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.*, **130**, 2336–2343.
5. Hwang,G.T. and Romesberg,F.E. (2008) The unnatural substrate repertoire of A, B, and X family DNA polymerases. *J. Am. Chem. Soc.*, **130**, 14872–14882.
6. Seo,Y.J., Hwang,G.T., Ordoukhanian,P. and Romesberg,F.E. (2009) Optimization of an unnatural base pair towards natural-like replication. *J. Am. Chem. Soc.*, **131**, 3246–3252.
7. Mitsui,T., Kimoto,M., Harada,Y., Yokoyama,S. and Hirao,I. (2005) An efficient unnatural base pair for a base-pair-expanded transcription system. *J. Am. Chem. Soc.*, **127**, 8652–8658.
8. Hirao,I., Harada,Y., Kimoto,M., Mitsui,T., Fujiwara,T. and Yokoyama,S. (2004) A two-unnatural-base-pair system toward the expansion of the genetic code. *J. Am. Chem. Soc.*, **126**, 13298–13305.
9. Hirao,I., Mitsui,T., Kimoto,M. and Yokoyama,S. (2007) An efficient unnatural base pair for PCR amplification. *J. Am. Chem. Soc.*, **129**, 15549–15555.
10. Seo,Y.J., Matsuda,S. and Romesberg,F.E. (2009) Transcription of an expanded genetic alphabet. *J. Am. Chem. Soc.*, **131**, 5046–5047.
11. Hirao,I., Ohtsuki,T., Fujiwara,T., Mitsui,T., Yokogawa,T., Okuni,T., Nakayama,H., Takio,K., Yabuki,T., Kigawa,T. *et al.* (2006) An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Methods*, **3**, 729–735.
12. Kimoto,M., Kawai,R., Mitsui,T., Yokoyama,S. and Hirao,I. (2009) An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.*, **37**, e14.
13. Guo,Z., Liu,Q. and Smith,L.M. (1997) Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nat. Biotechnol.*, **15**, 331–335.
14. Burgner,D., D'Amato,M., Kwiatkowski,D.P. and Loakes,D. (2004) Improved allelic differentiation using sequence-specific oligonucleotide hybridization incorporating an additional base-analogue mismatch. *Nucleosides, Nucleotides Nucleic Acids*, **23**, 755–765.
15. 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.
16. 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.
17. 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.
18. Matsuda,S., Fillo,J.D., Henry,A.A., Rai,P., Wilkens,S.J., Dwyer,T.J., Geierstanger,B.H., Wemmer,D.E., Schultz,P.G., Spraggon,G. *et al.* (2007) Efforts toward expansion of the genetic alphabet: Structure and replication of unnatural base pairs. *J. Am. Chem. Soc.*, **129**, 10466–10473.
19. 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.
20. Kim,Y., Leconte,A.M., Hari,Y. and Romesberg,F.E. (2006) Stability and polymerase recognition of pyridine nucleobase

analogues: role of minor groove H-bond acceptors. *Angew. Chem. Int. Ed.*, **45**, 7809–7812.
21. Leconte,A.M., Matsuda,S. and Romesberg,F.E. (2006) An efficiently extended class of unnatural base pairs. *J. Am. Chem. Soc.*, **128**, 6780–6781.
22. Hari,Y., Hwang,G.T., Leconte,A.M., Joubert,N., Hocek,M. and Romesberg,F.E. (2008) Optimization of the pyridyl nucleobase scaffold for polymerase recognition and unnatural base pair replication. *ChemBioChem*, **9**, 2796–2799.
23. Marky,L.A. and Breslauer,K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, **26**, 1601–1620.
24. Spink,C.H. and Chaires,J.B. (1999) Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA. *Biochemistry*, **38**, 496–508.
25. 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.
26. Aboul-ela,F., Koh,D., Tinoco,I. Jr. and Martin,F.H. (1985) Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A,C,G,D. *Nucleic Acids Res.*, **13**, 4811–4824.
27. Breslauer,K.L. (1995) Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods Enzymol.*, **259**, 221–242.
28. Rozners,E. and Moulder,J. (2004) Hydration of short DNA, RNA and 2′-OMe oligonucleotides determined by osmotic stressing. *Nucleic Acids Res.*, **32**, 248–254.
29. Leroy,J.L., Kochoyan,M., Huynh-Dinh,T. and Guéron,M. (1988) Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.*, **200**, 223–238.
30. Nonin,S., Leroy,J.L. and Guéron,M. (1995) Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. *Biochemistry*, **34**, 10652–10659.
31. Guckian,K.M., Krugh,T.R. and Kool,E.T. (1998) Solution structure of a DNA duplex containing a replicable difluorotoluene-adenine pair. *Nat. Struct. Biol.*, **5**, 954–959.
32. Saenger,W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
33. Hunter,W.N., Brown,T., Kneale,G., Anand,N.N., Rabinovich,D. and Kennard,O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-A resolution. *J. Biol. Chem.*, **262**, 9962–9970.
34. Brovchenko,I., Krukau,A., Oleinikova,A. and Mazur,A.K. (2007) Water clustering and percolation in low hydration DNA shells. *J. Phys. Chem. B*, **111**, 3258–3266.
35. Arai,S., Chatake,T., Ohhara,T., Kurihara,K., Tanaka,I., Suzuki,N., Fujimoto,Z., Mizuno,H. and Niimura,S. (2005) Complicated water orientations in the minor groove of the B-DNA decamer d(CCATT AATGG)2 observed by neutron diffraction measurements. *Nucleic Acids Res.*, **33**, 3017–3024.
36. Sherer,E.C. and Cramer,C.J. (2004) Structural and dynamic variations in DNA hexamers containing T–T and F–F single and tandem internal mispairs. *Theor. Chem. Acc.*, **111**, 311–327.
37. Hoheisel,J.D. (2006) Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.*, **7**, 200–210.
38. Krützfeldt,J., Rajewsky,N., Braich,R., Rejeev,K.G., Tuschl,T., Manoharan,M. and Stoffel,M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685–689.
39. Schwarz,D.S., Ding,H., Kennington,L., Moore,J.T., Schelter,J., Burchard,J., Linsley,P.S., Aronin,N., Xu,Z. and Zamore,P.D. (2006) Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet.*, **2**, e140.