J|A|C|S COMMUNICATIONS

Published on Web 09/29/2009

PCR with an Expanded Genetic Alphabet

Denis A. Malyshev,[†] Young Jun Seo,[†] Phillip Ordoukhanian,[‡] and Floyd E. Romesberg*[,†]

Department of Chemistry and Center for Protein and Nucleic Acid Research, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received July 23, 2009; E-mail: floyd@scripps.edu

The genetic alphabet is constrained by the efficient polymerase-mediated replication of DNA and RNA containing the two natural base pairs. In addition to laying the foundation for a semisynthetic organism with an expanded genetic code, an efficiently and selectively replicated and transcribed unnatural base pair would dramatically increase the potential of the already ubiquitous in vitro methodologies based on DNA and RNA and their sequence-specific amplification. PCR amplification of DNA containing unnatural base pairs was first reported by Benner and co-workers,[1] who used pairs with orthogonal H-bonding complementarity, and they were followed by Hirao and co-workers,[2] who used a pair formed between substituted pyrimidine and pyrrole nucleotide analogues. These studies represent landmarks in the effort to expand the genetic alphabet, but the unnatural pairs used were limited by strong sequence dependencies and/or inefficient transcription into RNA. This may not limit their use in some in vitro applications but does reduce their generality and preclude their eventual use in vivo as part of a semisynthetic organism.

We have focused on developing unnatural base pairs formed between predominantly hydrophobic nucleobases that have no structural homology to the natural nucleobases and that pair on the basis of hydrophobic and packing forces. Screening of a library of nucleotides, followed by hit optimization, identified the pair formed between d**5SICS** and d**MMO2** (Figure 1), which is relatively well recognized by different DNA polymerases under steady-state, single-nucleotide incorporation conditions.[3a] Further optimization identified d**NaM**, which pairs with d**5SICS** to form an unnatural pair that is even better replicated under the same steady-state conditions.[3b] Importantly, d**5SICS**−d**MMO2** and especially d**5SICS**−d**NaM** are also efficiently transcribed in both directions,[3c] suggesting that they might have immediate practical applications. Here we examine whether these unnatural base pairs, which are much less natural-like than those previously examined, can also be sufficiently well amplified for practical use.

Nucleosides were synthesized and converted to the corresponding triphosphates or phosphoramidites, and the phosphoramidites were incorporated into DNA duplexes D1−D6 using automated DNA synthesis [Figure 1 and the Supporting Information (SI)]. The duplexes are 134 or 149 nucleotides in length with a single, centrally positioned d**5SICS**−d**MMO2** (D1) or d**5SICS**−d**NaM** (D2−D6) pair. D1 is similar to the duplex used by Hirao et al.,[2] while D2−D5 systematically vary the flanking dG−dC base pairs to probe for sequence dependencies among sequences that are relatively challenging to amplify by PCR.[4] D6 contains randomized nucleotides to further explore sequence-specific effects, and D7 is identical to D1 except that the unnatural base pair is replaced by a natural dA−dT pair. Using a gel-based assay, we first explored PCR amplification of D1 using exonuclease-proficient DeepVent DNA polymerase in the presence and absence of d**5SICS**TP and d**NaM**TP (it should be noted that this results in the replacement of d**MMO2** with d**NaM** during the first round of replication). Promisingly, it was straightforward to identify PCR conditions under which DNA was amplified only when both natural and unnatural triphosphates were present (see the SI).

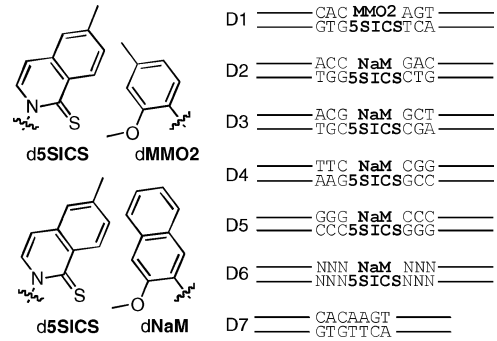

Figure 1. Unnatural base pairs (with sugar and phosphate backbones omitted for clarity) and duplexes employed in this study.

Table 1. Efficiencies and Fidelities of PCR Amplification[a]

template	d**X**TPs incorporated	enzyme	amplification	fidelity[b]
D1	d**NaM**, d**5SICS**	DeepVent	424	99.7
D2	d**NaM**, d**5SICS**	DeepVent	118	99.0
D3	d**NaM**, d**5SICS**	DeepVent	74	98.5
D4	d**NaM**, d**5SICS**	DeepVent	150	99.2
D5	d**NaM**, d**5SICS**	DeepVent	35	98.0
D6	d**NaM**, d**5SICS**	DeepVent	121	99.5
D7	−	DeepVent	556	−
D1[c]	d**NaM**, d**5SICS**	DeepVent	2.7×10^6	99.8
D6[d]	d**NaM**, d**5SICS**	DeepVent	1.9×10^4	98.2
D1	d**MMO2**, d**5SICS**	DeepVent	224	99.4
D5	d**MMO2**, d**5SICS**	DeepVent	25	97.1
D6	d**MMO2**, d**5SICS**	DeepVent	52	92.9
D1	d**NaM**, d**5SICS**	*Taq*	159	98.7
D5	d**NaM**, d**5SICS**	*Taq*	83	99.1
D6	d**NaM**, d**5SICS**	*Taq*	104	92.7
D1	d**NaM**, d**5SICS**	Phusion	257	99.7
D5	d**NaM**, d**5SICS**	Phusion	28	85.7
D6	d**NaM**, d**5SICS**	Phusion	82	95.9

[a] Conditions: 1 ng of DNA template; dNTPs/dXTP = 600/400 μM, 6 mM $MgSO_4$, 0.03 units of enzyme/μL, 8 min extension, 14 cycles. [b] Calculated as average fidelity for unnatural base pair replication in both directions, except with template D5, where it was calculated in one direction (see the text). [c] Using 100 fg of DNA template, 30 cycles. [d] Using 1 pg of DNA template, 30 cycles.

To better characterize amplification, we determined the yields of PCR products after 14 cycles, starting with D1−D6 (1 ng), dNTPs, d**5SICS**TP, and d**NaM**TP (Table 1). The 424-fold amplification of D1 was the highest and compares favorably with the 556-fold amplification of the control D7. The amplification levels of the other duplexes were slightly lower, likely because of their

[†] Department of Chemistry.
[‡] Center for Protein and Nucleic Acid Research.

10.1021/ja906186f CCC: $40.75

GC content, but remained greater than 100-fold, except for D3 and D5, which were amplified 74- and 35-fold, respectively. The lower efficiency observed with D3 and D5 is not surprising, considering that they position the unnatural base pair within a dG:dC run, which is particularly challenging to amplify by PCR.[4] While the amplification levels may be increased with additional rounds of PCR (see below), the data suggest that DNA containing the unnatural base pairs may be amplified with an efficiency that is sufficient for in vitro applications.

To better characterize fidelity, the amplicons were sequenced (Table 1). In most cases, standard sequencing reactions (ABI 3730 DNA Analyzer) lacking unnatural triphosphates terminated at the unnatural nucleotide. In these cases, the fidelities (i.e., the percentage of unnatural base pair retention per doubling) are reported as the average determined by sequencing both amplified strands (see the SI). For D5, significant read-through was observed in one direction. However, because read-through was also observed with chemically synthesized control strands, we conclude that it results from sequencing and not from PCR amplification. In this case, the reported fidelity was determined from one strand context. Remarkably, in four of the six sequence contexts examined, the average fidelity was at least 99%. It was slightly lower in D3 (98.5%) and D5 (98.0%), which again likely resulted from their particularly difficult, GC-rich sequence context. Despite the expected challenges associated with replicating the GC-rich sequences, the 99.5% fidelity observed with the random sequence of D6 suggests that in general, most sequences are compatible with high-fidelity amplification of the unnatural base pair.

Many in vitro applications rely on the efficient amplification of minuscule quantities of a template. To further explore the potential utility of the unnatural base pair, we examined PCR amplification via quantitative real-time PCR with decreasing amounts of D1 and D6 (Table 1 and Figure 2). After 30 cycles of PCR, we found that D1 amplification remained efficient (2.7×10^6) and accurate (99.8% fidelity) even with only 100 fg of template. D6 was examined down to 1 pg, where amplification also remained efficient (1.9×10^4) and accurate (98.2% fidelity). The efficient, high-fidelity amplification of D6 is particularly noteworthy given its randomized sequence. Moreover, sequencing revealed no major differences before and after amplification of D6 (see the SI), further suggesting that amplification is general and not strongly sequence-dependent.

To explore the determinants of efficient unnatural base pair amplification, we compared the amplification of DNA using d**5SICS**TP and d**MMO2**TP with that described above using d**5SICS**TP and d**NaM**TP. We examined amplification with duplexes D1 and D5, which were the best and worst sequence contexts, respectively, for d**5SICS**−d**NaM** amplification, as well as with D6 (Table 1). The two unnatural pairs were replicated with similar efficiencies and fidelities in D1, reinforcing the idea that their incorporation into DNA does not rely solely on hydrophobicity, since d**NaM** is much more hydrophobic than d**MMO2**, and also that the intrabase packing interactions, which are similar in the two base pairs, make an important contribution. However, the data also reveal that d**5SICS**−d**MMO2** was replicated with lower fidelity and efficiency in D5 and D6, suggesting that its recognition is more dependent on sequence-context than that of d**5SICS**−d**NaM**.

We next examined amplification with two other thermostable polymerases commonly used for PCR, *Taq* and Phusion, using templates D1, D5, and D6 and dNTPs as well as d**5SICS**TP and d**NaM**TP (Table 1). Relative to DeepVent, *Taq* recognized the

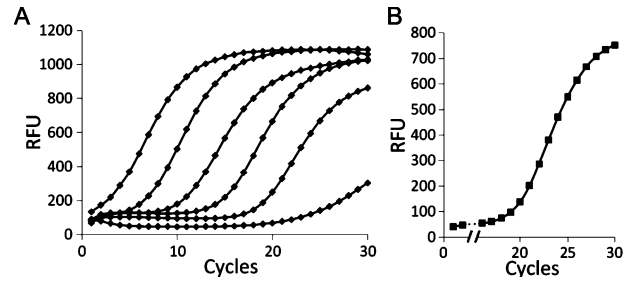

Figure 2. Quantitative real-time PCR analysis. (A) Amplification of (left to right) 1 ng, 100 pg, 10 pg, 1 pg, and 0.1 pg of D1 over 30 cycles. The rightmost curve corresponds to no template (negative control), where the signal results from primer-dimer formation. Each curve represents the average of two independent experiments. (B) Amplification of 1 pg of D6 over 30 cycles. See the SI for details.

unnatural base pair in D1 and D5 with similar efficiency and fidelity, but the fidelity was somewhat reduced with D6. Relative to DeepVent, Phusion polymerase amplified d**5SICS**−d**NaM** in D1 with generally similar efficiency and fidelity, but in this case, the fidelities were somewhat lower with both D5 and D6. The data suggest that the different polymerases behave similarly, except with increasing GC content, where divergent behavior is also observed with fully natural sequences,[4b] further demonstrating that the determinants of replication are contained within the base pairs and are not specific for a given polymerase.

We have now demonstrated not only that d**5SICS**−d**NaM** and d**5SICS**−d**MMO2** are efficiently transcribed in either direction[3c] but also that they may be amplified by PCR and are thus the first pairs to fulfill all of the primary requirements of a fully functional unnatural base pair. From a theoretical perspective, the data reveal that all of the properties required of a functional unnatural base pair may be optimized within predominantly hydrophobic nucleobases that bear no homology to the natural nucleobases. From a practical perspective, the data suggest that d**5SICS**−d**MMO2** and especially d**5SICS**−d**NaM** are sufficiently optimized for use as part of an in vitro expanded genetic alphabet. This should enable a variety of unprecedented applications,[5] which are currently being explored.

Acknowledgment. Funding was provided by NIH GM060005.

Supporting Information Available: Details of template synthesis and PCR analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Sismour, A. M.; Lutz, S.; Park, J. H.; Lutz, M. J.; Boyer, P. L.; Hughes, S. H.; Benner, S. A. *Nucleic Acids Res.* **2004**, *32*, 728. (b) Yang, Z.; Sismour, A. M.; Sheng, P.; Puskar, N. L.; Benner, S. A. *Nucleic Acids Res.* **2007**, *35*, 4238.

(2) (a) Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S. *Nat. Methods* **2006**, *3*, 729. (b) Kimoto, M.; Kawai, R.; Mitsui, T.; Yokoyama, S.; Hirao, I. *Nucleic Acids Res.* **2009**, *37*, e14.

(3) (a) Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E. *J. Am. Chem. Soc.* **2008**, *130*, 2336. (b) Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 3246. (c) Seo, Y. J.; Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 5046.

(4) (a) Hansen, L. L.; Justensen, J. In *PCR Primer: A Laboratory Manual*; Diffenbach, C. W., Dveksler, G. S., Eds.; Cold Spring Harbor Laboratory Press: Woodbury, NY, 2003; pp 43−62. (b) Arezi, B.; Xing, W.; Sorge, J. A.; Hogrefe, H. H. *Anal. Biochem.* **2003**, *321*, 226.

(5) *The Aptamer Handbook: Functional Oligonucleotides and Their Applications*; Klussmann, S., Ed.; Wiley-VCH: Weinheim, Germany, 2006.

JA906186F