EFFORTS TOWARDS EXPANSION OF THE GENETIC ALPHABET

Aaron M. LECONTE and Floyd E. ROMESBERG*

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037-1000, U.S.A.; e-mail: floyd@scripps.edu

In an effort to develop an unnatural base pair with which to expand the genetic code, we have examined a wide variety of large aromatic nucleotides. While these nucleotides are often incorporated by natural polymerases with high efficiency and selectivity, extension beyond the newly formed unnatural terminus is often inefficient. We have undertaken several strategies to overcome this inefficient extension step. We have examined a wide variety of simple phenyl rings derivatized with methyl, fluoro, or nitrogen. The small aromatic surface of these base pairs should prevent inter-strand intercalation, which is thought to inhibit the polymerase-mediated extension of base pairs with larger aromatic surface area. We have also successfully utilized alternative polymerases in a two-polymerase system in which one polymerase incorporates the unnatural nucleoside triphosphate, and a second polymerase extends the unnatural terminus. Also, we have developed an activity based phage selection system which can be used to evolve DNA polymerases that more efficiently recognize the unnatural substrates, and our initial successes are described.

INTRODUCTION

The biological system of information storage, based on the selective Watson–Crick hydrogen bonding (H-bonding) of adenine with thymine (dA:dT base pair) and guanine with cytosine (dG:dC base pair), has been conserved throughout nature. While this system is capable of encoding extremely complex living systems, natural DNA is inherently limited in the number of codons it can encode as well as in its functionality. The development of a third, orthogonal DNA base pair, which can be enzymatically replicated with high efficiency and selectively, has the potential to not only vastly increase the genetic information which can be stored in DNA, but also introduce new functionality into nucleic acid catalysts[1,2]. Initial efforts to develop a third base pair focused on alternative H-bonding schemes, but these base pairs were not replicated with high enough fidelity to be viable, likely because of tautomerization[3]. However, there is no reason, in principle, that a third base pair must utilize H-bonding. In fact, a hydrophobic shape analogue of a natural base has been substituted for its natural counterpart and pairs with selectivity and efficiency similar to the natural pair despite having no heteroatoms[4]. Based on this key observation, we, and others[5–7], have developed base pairs which rely on hydrophobic and van der Waals forces rather than H-bonding. One early success was the **PICS** self

pair, a pair formed between two identical **PICS** nucleosides (Fig. 1)[8]. Self pairs aid in limiting mispair synthesis between natural and unnatural nucleosides. Notably, self pairs are not a limitation for the generation of genetic information storage since the addition of a self pair would create 61 new codons, essentially doubling the coding capacity of DNA. Despite lacking H-bonding and shape complementarity, d**PICS**TP is incorporated by the exonuclease free proteolytic fragment of DNA Polymerase I, Klenow Fragment (Kf exo$^-$), against a templated d**PICS** with moderate efficiency ($k_{cat}/K_M = 2.4 \times 10^5$) and with relatively high fidelity (20-fold more efficient than its fastest mispair, dT). However, as is typical of most large, hydrophobic base pairs, synthesis beyond the unnatural self pair terminus is highly inefficient, likely due to distortion of the nascent primer terminus. We have undertaken several strategies to overcome this barrier. One strategy is the use of smaller aromatic nucleobase scaffolds and their subsequent derivatization to optimize their thermostability and replication efficiency. Another technique is the use of a two-polymerase system, where one polymerase inserts the nucleotide, and a second polymerase extends past the unnatural terminus. A third technique is the use of directed evolution to expand the substrate repertoire of DNA Polymerases to include unnatural DNA.

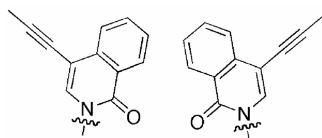

FIG 1.
The **PICS** self pair

RESULTS AND DISCUSSION

The physical properties that may be important for efficient polymerase recognition of unnatural nucleotides include aromatic surface area and heteroatom substitution. For instance, the large aromatic surface area of past unnatural nucleotides likely aided in their recognition and incorporation by DNA polymerases. However, such base pairs, once synthesized, are unable to adopt a primer terminus geometry that is successfully recognized by Kf exo$^-$ because of steric constraints, and synthesis beyond the unnatural terminus is inefficient. We are interested in using small aromatic scaffolds with various methyl, fluoro, and heteroatom substitutions and derivatizations to develop a more efficiently extended unnatural base pair (Fig. 2). One plausible concern about smaller base pairs is that the reduc-

tion in aromatic surface area will result in a large decrease in duplex stability. Remarkably, with careful design, the stability and thermoselectivity may be similar to natural pairs[9]. For instance, despite a significantly reduced aromatic surface area and a complete lack of heteroatoms, the **DM5** self pair is virtually as stable and thermally selective in a DNA duplex as a dA:dT pair.

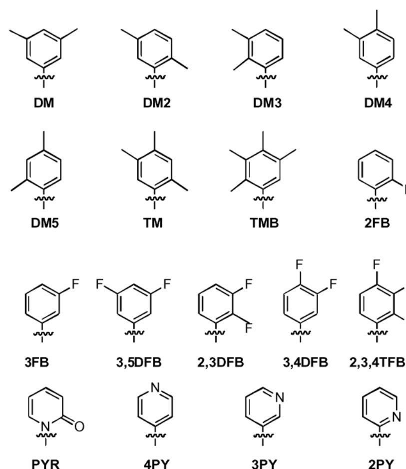

FIG. 2
Small aromatic nucleobases

More importantly, we have found that judicious derivatization can greatly improve the efficiency of extension. The **3FB** self pair is synthesized by Kf exo^{-} with high efficiency, only approximately 10-fold below wild-type synthesis ($k_{cat}/K_M = 2.1 \times 10^6$)[6]. Unlike the larger aromatic nucleobases, the self pair is extended by Kf exo^{-} at a rate only approximately 100-fold below wild-type ($k_{cat}/K_M = 3.3 \times 10^5$). This is a 200-fold increase in extension rate compared to the non-fluorinated benzene nucleobase, **BEN** (Table I, S.M., F.E.R. unpublished results). The synthesis and extension of the **3FB** self pair occurs with moderate selectivity over natural pairs (20-fold selective over dATP). Despite its simplicity, the remarkable efficiency and selectivity of the **3FB** base pair make it the most promising unnatural base pair candidate identified to date. Also noteworthy, extension of the 2-pyridine (**2PY**) and pyridone (**PYR**) self pairs increased 60- and 51-fold, respectively, compared to the rate of extension of the benzene self pair (Table I, Y.K., A.M.L., F.E.R. unpublished results). The impressive extension rates of **3FB**, **2PY**, and **PYR** suggest that careful substitution and derivitazation can lead to marked increases in extension rates.

We are currently characterizing synthetic combinations of these bases (for instance fluoropyridones) to determine whether the effects that lead to increased extension are additive. If they are, we expect to identify an unnatural base pair that approaches the rate of extension as a wild-type base pair.

TABLE I
Steady-state rate constants of Kf exo⁻ mediated extension of unnatural termini[a]

5′-dTAATACGACTCACATAGGGAGA(**X**)
3′-dATTATGCTGAGTGTATCCCTCT(**X**)GCTAGGTTACGGCAGGATCGC

X	k_{cat}, min^{-1}	K_M, μM^{-1}	k_{cat}/K_M, min^{-1} μM^{-1}
3FB[b]	28 (4)	85 (24)	3.3×10^5
PYR[c]	9.6 (1.0)	117 (32)	8.2×10^4
2PY[c]	3.7 (2.8)	46 (33)	9.7×10^4
BEN[c]	0.18 (0.02)	112 (20)	1.6×10^3

[a] Assay conditions were as follows: 40 nM template-primer duplex, 0.12–1.2 nM enzyme, 50 mM Tris buffer (pH 7.5), 10 mM $MgCl_2$, 1 μM DTT, 50 μg/ml BSA. The reactions were initiated by adding the DNA–enzyme mixture to an equal volume (5 μl) of a 2 × dNTP stock solution, incubated at room temperature for 2–12 min, and quenched by the addition of 20 μl of loading buffer (95% formamide, 20 mM EDTA). A 5-μl portion of the reaction mixture was then analyzed by 15% polyacrylamide gel electrophoresis containing 8 M urea. Radioactivity was quantified using a PhosphorImager (Molecular Dynamics), with overnight exposures and the ImageQuant program. The data were fit to the Michaelis–Menten equation using the program Kaleidagraph (Synergy software). The data presented here are averages of triplicates. [b] Ref. 6. [c] Unpublished results.

Another potential solution to the slow extension of unnatural termini is the use of alternative polymerases. Towards this end, we characterized the ability of pol β, a mammalian polymerase that primarily fills gaps during base excision repair, to replicate the 7-azaindole (**7AI**) self-pair[10]. **7AI** is a large aromatic base which was found to be inserted by Kf exo⁻ with relatively high efficiency and selectivity ($k_{cat}/K_M = 2.2 \times 10^5$, 37-fold more efficient than its fastest natural mispair, dA)[11]. However, like most large aromatic base pairs, Kf exo⁻ could not extend the unnatural terminus at a detectable rate ($k_{cat}/K_M < 10^3$). Pol β mediated extension of the **7AI**:**7AI** self-pair proceeded with an efficiency of 3.5×10^5 M^{-1} min^{-1}, which is virtually identical to pol β mediated extension of a natural pair; however, it is unable to synthesize the unnatural pair at a detectable rate. Fortunately, the activities of Kf and pol β are complementary with regard to the **7AI**

self-pair: Kf efficiently synthesizes the self-pair, and pol β efficiently extends it. We characterized extension of the 23-nt primer oligonucleotide in the presence of each polymerase separately, as well as in the presence of both polymerases together (Fig. 3). In the presence of both enzymes, full-length product synthesis was efficient and selective.

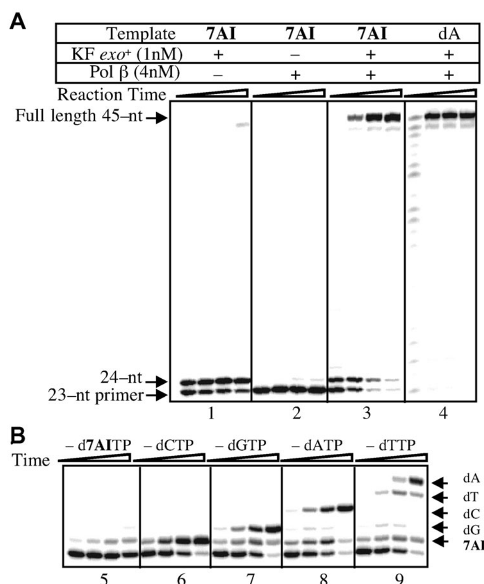

FIG. 3
Extension of 23-nt primer (with 45-nt template). **A**. Reactions 1, 2, and 3 contain either KF exo$^+$, pol β, or both, respectively. Template **X** = **7AI**. The four lanes for each reaction correspond to reaction times of 15, 30, 60, 120 min. Reaction 4 is the same as reaction 3, except template **X** = dA and the reaction times were 3, 10, 30, 60 min. **B**. Reactions 5–9 are identical to reaction 3, except each contains only four of the requisite five nucleoside triphosphates, and the reaction times were 5, 10, 20, 30 min. The omitted triphosphate is indicated along the top of the gel. The template base that resulted in the termination of primer extension is indicated along the right edge of the gel

Although we have had success derivatizing unnatural bases and using alternative polymerases, the rates of synthesis of unnatural DNA are typically less efficient than the rates of natural synthesis. We are thus interested in using the tools of modern directed evolution to develop novel DNA and RNA polymerases which utilize unnatural base pairs. We have developed an activity based phage display selection system based on the co-display on phage of DNA polymerase libraries and an 'acidic peptide' that is used to attach a DNA substrate (Fig. 4). A 'basic peptide', which has been previously

conjugated to a primer-template through a bis-maleimide linker, is attached to the acidic peptide via a leucine zipper and a disulfide bond. In this manner, the primer-template, the encoded polymerase mutant and its gene are all physically associated with a phage particle. Polymerases which can successfully recognize their primer-template substrate will synthesize DNA utilizing modified or unmodified nucleotide triphosphates and only then incorporate a biotin-dUTP. Biotinylated phage particles can be recovered using a streptavidin solid support, enriching the population for active members.

First, as a proof of principle, we used the selection system to evolve variants of the proteolytic fragment of ***Thermus aquaticus*** (***Taq***) polymerase (Stoffel fragment, Sf) that efficiently synthesize RNA [12]. Phage biotinylation and recovery relied on the incorporation of rNTPs. After four rounds of selection, three clones (SFR1, SFR2, and SFR3) could efficiently extend a DNA primer by insertion of rATP opposite dT. Each polymerase exhibited 10^3- to 10^4-fold faster rates for insertion of each rNTP relative to wild-type Sf and required only normal PCR-like conditions. In fact, one mutant (SFR3) synthesized RNA as efficiently as Sf synthesizes DNA.

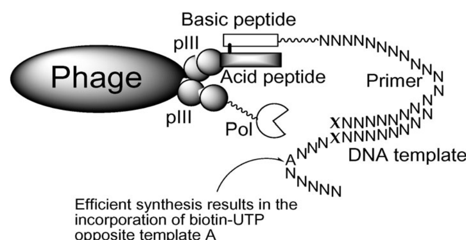

FIG. 4

Phage particle with attached polymerase and substrate. **X** represents a natural or unnatural base. Selections are designed so that biotin incorporation is dependent on DNA synthesis with natural or modified dNTPs

To demonstrate that the activity based phage selection system is capable of evolving a true unnatural activity, we used it to evolve polymerase variants that efficiently synthesize polymers comprised of 2′-*O*-methyl ribonucleosides ($_{OMe}$NTPs)[13]. In this case, biotinylation required the insertion of $_{OMe}$CTP and $_{OMe}$ATP instead of natural substrates. Protein isolated from four clones were identified that could efficiently extend a DNA primer by inserting $_{OMe}$ATP opposite dT in the template. Two clones, SFM01 and SFM18, encode identical amino acid sequences, but have different DNA sequences, showing that a protein sequence was independently isolated twice, which implies that a sufficient number of clones were assayed for ac-

tivity. The kinetic data revealed a remarkable increase in $_{OMe}$NTP incorporation efficiency that was most pronounced for clone SFM19 [13]. The k_{cat}/K_M for correct $_{OMe}$NTP insertion by SFM19 ranged from 1×10^6 to 6×10^6 M^{-1} s^{-1}, even in the case of modified UTP insertion opposite dA, which has proven especially difficult for DNA polymerase mutants to catalyze due to strong discrimination against this base[14]. In contrast to SFM19, the wild-type enzyme showed no detectable activity with any of the modified triphosphates ($k_{cat}/K_M < 10^3$ M^{-1} s^{-1}). Thus, the evolved activity of SFM19 corresponds to an increase in k_{cat}/K_M of at least 10,000-fold. In addition, the fidelity of SFM19 was virtually unchanged relative to Sf.

In our first attempt to evolve DNA polymerases which utilize unnatural nucleobases, we evolved an Sf mutant that synthesizes DNA containing an unnatural **PICS** self-pair. Polymerase mutants were selected based on their ability to recognize and extend a **PICS**:**PICS** terminus. From twenty randomly selected clones, one polymerase, P2, was found to insert dCTP with significantly increased efficiency. The unnatural activity of P2 was quantified with both steady-state and pre-steady-state kinetics (A.L., F.E.R. unpublished results). The evolved enzyme synthesized the **PICS**:**PICS** self pair by insertion of d**PICS**TP against d**PICS** more than 310 fold faster than the wild type polymerase (Table II). In fact, P2 synthesizes the self-pair only ~10-fold less efficiently than the parental enzyme synthesizes a natural base pair in the same sequence context. The significant increase in rate did not compromise selectivity, which is greater than that of any other unnatural nucleobase pair – polymerase system reported to date (200-fold selective over the fastest mispair, dT) (Table II). The pre-steady-state extension rate increased

TABLE II
Steady-state rate constants of polymerase mediated synthesis of d**PICS** self pair and mispairs[a,b]

5′-dTAATACGACTCACATAGGGAGA
3′-dATTATGCTGAGTGTATCCCTCT(**PICS**)GCTAGGTTACGGCAGGATCGC

Enzyme	dNTP	k_{cat}, min^{-1}	K_M, μM^{-1}	k_{cat}/K_M, min^{-1} μM^{-1}
P2	**PICS**	6.5 (1.0)	20.5 (1.2)	3.2×10^5
P2	A, C, G	n.d.[c]	n.d.[c]	$<10^3$
P2	T	0.23 (0.07)	148 (73)	1.6×10^3
Sf	**PICS**	n.d.[c]	n.d.[c]	$<10^3$

[a] See Table I for experimental details. [b] Unpublished results. [c] Rates too slow for determination of k_{cat} and K_M independently.

30-fold relative to the parental enzyme. However, since extension was not observed under steady-state conditions, it must be limited by some step other than bond formation, such as duplex dissociation. Regardless, P2 both synthesizes and extends the **PICS** self-pair with reasonable efficiency whereas the parental enzyme is unable to catalyze either step at detectable rates. The results suggest that, with suitably designed experiments involving more stringent selection criteria or gene shuffled libraries, the selection system should be capable of evolving polymerases with truly expanded repertoires.

In nature, DNA and RNA polymerases coevolved with the two natural base pairs to develop a system in which genetic material is replicated with extremely high fidelity and selectivity. We are interested in introducing a third base pair into DNA and RNA in a manner which mimics nature's evolution of the natural DNA. By modifying nucleobase structure, utilizing the tools of modern organic chemistry, and by evolving DNA polymerases, utilizing the tools of modern directed evolution, we are rapidly approaching a functional third base pair, which will be the first step in expanding the genetic code.

This work was supported by the National Institutes of Health (NIGMS 60005).

REFERENCES

1. Joyce G. F.: *Curr. Opin. Struct. Biol.* **1994**, *4*, 331.
2. Kusser W.: *J. Biotechnol.* **2000**, *74*, 27.
3. Piccirilli J. A., Krauch T., Moroney S. E., Benner S. A.: *Nature* **1990**, *343*, 33.
4. Moran S., Ren R. X.-F., Rumney S. I., Kool E. T.: *J. Am. Chem. Soc.* **1997**, *119*, 2056.
5. Ishikawa M., Hirao I., Yokoyama S.: *Tetrahedron Lett.* **2000**, *41*, 3931.
6. Henry A. A., Olsen A. G., Matsuda S., Yu C., Geierstanger B. H., Romesberg F. E.: *J. Am. Chem. Soc.* **2004**, *126*, 6923.
7. Mitsui T., Kitamura A., Kimoto M., To T., Sato A., Hirao I., Yokoyama S.: *J. Am. Chem. Soc.* **2003**, *125*, 5298.
8. McMinn D. L., Ogawa A. K., Wu Y., Liu J., Schultz P. G., Romesberg F. E.: *J. Am. Chem. Soc.* **1999**, *121*, 11585.
9. Matsuda S., Romesberg F. E.: *J. Am. Chem. Soc.* **2004**, *126*, 14419.
10. Tae E. L., Wu Y. Q., Xia G., Schultz P. G., Romesberg F. E.: *J. Am. Chem. Soc.* **2001**, *123*, 7439.
11. Ogawa A. K., Wu Y., McMinn D. L., Liu J., Schultz P. G., Romesberg F. E.: *J. Am. Chem. Soc.* **2000**, *122*, 3274.
12. Xia G., Chen L., Sera T., Fa M., Schultz P. G., Romesberg F. E.: *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6597.
13. Fa M., Radeghieri A., Henry A. A., Romesberg F. E.: *J. Am. Chem. Soc.* **2004**, *126*, 1748.
14. Patel P., Loeb L.: *J. Biol. Chem.* **2000**, *275*, 40266.