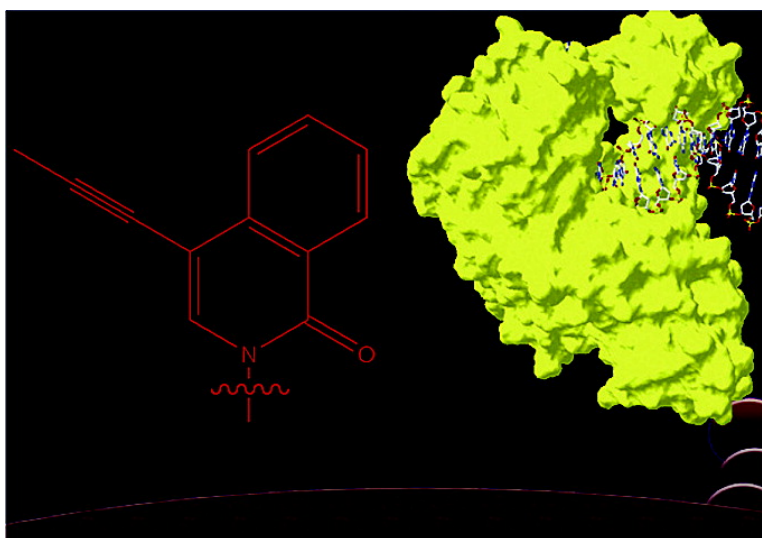Communication

# Polymerase Evolution: Efforts toward Expansion of the Genetic Code

Aaron M. Leconte, Liangjing Chen, and Floyd E. Romesberg

## More About This Article

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

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

View the Full Text HTML

# Polymerase Evolution: Efforts toward Expansion of the Genetic Code

Aaron M. Leconte, Liangjing Chen, and Floyd E. Romesberg*

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

Received May 20, 2005; E-mail: floyd@scripps.edu

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. Expansion of the alphabet to contain a third base pair would allow additional information to be encoded in DNA and would also enable a variety of *in vitro* experiments using unnatural nucleic acids.[1] *A priori*, there is no reason to assume that the requirements for duplex stability and replication must limit the genetic alphabet to only two base pairs, or even to hydrogen-bonded base pairs.[2] The stability and replication of DNA containing nucleobase analogues that are predominantly hydrophobic and that do not bear a H-bonding pattern or shape that is complementary to the natural bases have been previously examined.[3−6] One class of unnatural base pairs that has been extensively examined is that formed between two identical nucleobase analogues. For example, the **PICS** self-pair is stable in duplex DNA and synthesized by the Klenow fragment of *E. coli* (Kf)[3] and by *Thermus aquaticus* (*Taq*) polymerase, but not by the proteolytic fragment of *Taq* (Stoffel fragment, Sf). In addition, none of these polymerases can extend a primer terminating with a **PICS** self-pair.

One strategy for improving the replication of these unnatural base pairs is based on their further derivatization with heteroatoms or alkyl substituents.[6,7] This strategy mimics the optimization of the natural bases in nature. However, in nature, base pair optimization proceeded simultaneously with DNA polymerase evolution. Thus, any expansion of the genetic code is expected to be facilitated by optimizing both the unnatural nucleobase analogues and the polymerases that replicate them. Here, we report our initial efforts toward the directed evolution of polymerases that more efficiently synthesize DNA containing the **PICS** self-pair.

Previously, we reported a phage display selection system designed to evolve DNA or RNA polymerases with novel activities.[8,9] We used the selection system to evolve variants of Sf that efficiently synthesize RNA[8] or DNA containing C2′-O-methyl-modified nucleotides.[9] The selection system is based on the co-display on phage of DNA polymerase libraries and an "acidic peptide" that is used to attach a DNA substrate (Figure 1B). If the displayed polymerase mutant recognizes the attached unnatural substrate, it will synthesize DNA, and only then incorporate a biotin-dUTP to the attached primer. Biotinylated phage particles may be selectively recovered using a streptavidin solid support.

To optimize the replication of DNA containing **PICS**, we decided to select for variants of Sf that more efficiently synthesize and extend a primer-template terminus containing the self-pair. We reasoned that both steps of replication might be optimized by selecting for recognition of the self-pair at a primer terminus. A **PICS**-containing 51-mer oligonucleotide primer was thus conjugated to a "basic peptide" by way of a bismaleimide linker and hybridized to the **PICS**-containing 28-mer oligonucleotide template (Supporting Information). As described previously, the primer-
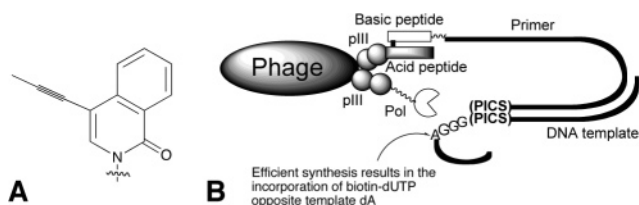


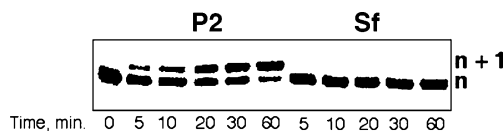**Figure 1.** (A) **PICS** and (B) selection scheme.



**Figure 2.** Pre-steady-state extension of **PICS** self-pair. Primer is labeled *n*, and the extension product is labeled *n* + 1. See Supporting Information for experimental details.

template assembly was then attached to phage particles via coiled-coil formation between the acidic and basic peptides.[8,9] Incorporation of biotin-dUTP requires productive recognition of the primer terminating with a **PICS** self-pair (Figure 1B).

Sf libraries were constructed as described previously.[8,9] Briefly, five focused libraries were constructed by two-step overlapping extension PCR. The doping ratio was set between 8 and 45%, depending on the size of the region to be mutated. Each focused library contained mutations localized to a specific region of the polymerase:[10] two regions of the metal binding site (amino acids 597−615 and 783−786); a portion of the template binding site (728−734); the duplex binding site (568−587); and the O-helix (665−676). The final polymerase library was generated by combining the five focused libraries so that they were equally represented. Based on the number of phage used in each selection ($5 \times 10^{12}$), the size of each library ($5 \times 10^8$), and the efficiency of polymerase display (0.1%), we approximate that each clone was displayed on 10 phage particles.

The phage displayed polymerase library was subjected to four rounds of selection with dCTP (50 $\mu$M) and biotin-dUTP (2 $\mu$M), at 50 °C, for 15 min. Protein from selected phage was prepared and screened for extension of the **PICS** self-pair by insertion of dCTP opposite dG in the template. One clone, P2, was found to insert dCTP with significantly increased efficiency (Figure 2). The unnatural activity was then quantified with pre-steady-state kinetics. While Sf inserts dCTP with a rate that is too low to detect ($k_{pol}$ < 0.01 min⁻¹), P2 inserts the same triphosphate with a $k_{pol}$ of 0.3 min⁻¹, at least 30-fold increased relative to the parental enzyme. Moreover, the rate of self-pair extension by incorporation of an incorrect dNTP was only typical of natural mispair synthesis by wild-type polymerases (<0.04 min⁻¹)[11,12] (see Supporting Information). Neither Sf nor P2 detectably extended the primer under steady-state conditions.

**Table 1.** Steady-State Rate Constants of Synthesis of DNA Containing dPICS

5' - dTAATACGACTCACTATAGGGAGA

3' - dATTATGCTGAGTGATATCCCTCT(**PICS**)GCTAGGTTACGGCAGGATCGC

| enzyme | dNTP | $k_{cat}$ (min$^{-1}$) | $K_M$ ($\mu$M) | $k_{cat}/K_M$ (M$^{-1}$ min$^{-1}$) |
|---|---|---|---|---|
| P2 | PICS | 6.5 (1.0) | 20.5 (1.2) | $3.18 \times 10^5$ |
| P2 | T | 0.23 (0.07) | 148 (73) | $1.55 \times 10^3$ |
| P2 | A, C, G | n.d.$^b$ | n.d.$^b$ | $\leq 1 \times 10^3$ |
| Sf | PICS | n.d.$^b$ | n.d.$^b$ | $\leq 1 \times 10^3$ |

$^a$ See Supporting Information for experimental details. $^b$ Rates too slow for determination of $k_{cat}$ and $K_M$ independently.

To examine how selectively the self-pair was extended relative to a mispair, we examined the rate of extension of d**PICS**:dT and dT:d**PICS** mispairs (the mispairs with dT are the most competitively synthesized; see below). The extension rate of either mispair, using the steady-state assay, is below the detection limit ($k_{cat}/K_M < 1 \times 10^3$). Using more sensitive pre-steady-state conditions, we were able to measure the rate of steady-state incorporation, $k_{ss}$, for both P2 and Sf (Supporting Information). While P2 extends the **PICS** mispairs with dT faster than Sf, it does so only with rates that are comparable to or less than extension of natural mispairs by wild-type enzymes.[13]

We also characterized Sf- and P2-mediated self-pair synthesis. Under steady-state conditions, Sf is unable to synthesize the self-pair by insertion of d**PICS**TP opposite d**PICS** in the template with a detectable rate ($k_{cat}/K_M < 1 \times 10^3$). Remarkably, P2 incorporates d**PICS**TP against d**PICS** under the same conditions with a $k_{cat}/K_M$ = $3.2 \times 10^5$, at least 320-fold increased relative to the parental enzyme. This is at least 200-fold greater than the efficiency with which P2 inserts any natural dNTP opposite d**PICS** (Table 1). The selectivity of this insertion is greater than that for the polymerase-mediated synthesis of any unnatural base pair reported to date.

We then measured the steady-state rate for P2 insertion of each natural dNTP opposite dA. Strikingly, P2 synthesizes the correct dT:dA pair more than 100-fold faster than Sf (Supporting Information), and it does so with fidelity uncompromised relative to the parental enzyme (no mispair was synthesized with a $k_{cat}/K_M$ greater than $1 \times 10^3$). P2 is also able to efficiently synthesize natural DNA under standard PCR conditions, suggesting that P2 retains the ability to synthesize all natural base pairs (unpublished results).

The evolved activity of P2 results from three mutations; F598I, I614F, and Q489H. Gln489 was not included in the designed library, and the His residue was presumably introduced by spontaneous mutation. In the wild-type enzyme, Gln489 forms a salt-bridge with the DNA phosphate backbone at the −8 position.[10] Phe598 is a conserved residue at the interface of the palm and thumb subdomains, which are both important for substrate recognition and catalysis.[10] Thus, while the residue at this position does not directly contact the DNA, it may mediate important long-range interactions. Ile614 is part of the highly conserved motif A of the DNA pol I family and, along with Phe667 and Tyr671, forms part of a hydrophobic pocket that packs on the sugar ring of the incoming

dNTP.[10] Mutation to Phe at this position may improve self-pair recognition by increasing favorable $\pi$-stacking interactions with the large aromatic ring of **PICS**.

P2 both synthesizes and extends the **PICS** self-pair with reasonable efficiency and fidelity, whereas the parental enzyme is unable to catalyze either step at detectable rates. In fact, the evolved enzyme synthesizes the self-pair only ∼10-fold less efficiently than the parental enzyme synthesizes a natural base pair in the same sequence context. The increase in extension rate was 30-fold and only observable under pre-steady-state conditions. This suggests that extension of the **PICS** self-pair may be limited by a step other than bond formation, such as primer-template binding, duplex dissociation, or conformational changes in the polymerase. The evolved properties of P2, as well as the observed mutations, are consistent with an increased affinity for the DNA primer-template containing the self-pair. Increased affinity for DNA, in general, might also underlie the 100-fold increased rate of P2-mediated synthesis of natural DNA. A more thorough examination of this hypothesis is currently in progress.

P2 represents the first polymerase evolved to possess an altered nucleobase substrate repertoire, albeit a not yet optimized one. 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. The ability to evolve polymerases, specifically tailored for an unnatural base pair, will not only facilitate the effort to expand the genetic alphabet but also help develop polymerases with a variety of unnatural activities for novel sequencing methodologies and other biotechnology applications.

**Supporting Information Available:** Experimental details and kinetic data. This material is available free of charge via the Internet at http://pubs.acs.org.

**References**

(1) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37.
(2) Delaney, J. C.; Henderson, P. T.; Helquist, S. A.; Morales, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469−4473.
(3) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(4) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(5) Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125*, 5298−5307.
(6) Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C.; Geierstanger, B. H.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 6923−6931.
(7) Henry, A. A.; Yu, C.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 9638−9646.
(8) Xia, G.; Chen, L.; Sera, T.; Fa, M.; Schultz, P. G.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6597−6602.
(9) Fa, M.; Radeghieri, A.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 1748−1754.
(10) Li, Y.; Korolev, S.; Waksman, G. *EMBO J.* **1998**, *17*, 7514−7525.
(11) Wong, I.; Patel, S. S.; Johnson, K. A. *Biochemistry* **1991**, *30*, 526−537.
(12) Eger, B. T.; Benkovic, S. J. *Biochemistry* **1992**, *31*, 9227−9236.
(13) Eckert, K. A.; Kunkel, T. A. *J. Biol. Chem.* **1993**, *268*, 13462−13471.

JA053322H