doi:10.1016/j.jmb.2006.06.050 J. Mol. Biol. (2006) 361, 537–550

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# Directed Evolution of DNA Polymerase, RNA Polymerase and Reverse Transcriptase Activity in a Single Polypeptide

**Jennifer L. Ong, David Loakes, Szymon Jaroslawski Kathleen Too and Philipp Holliger***

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

DNA polymerases enable key technologies in modern biology but for many applications, native polymerases are limited by their stringent substrate recognition. Here we describe short-patch compartmentalized self-replication (spCSR), a novel strategy to expand the substrate spectrum of polymerases in a targeted way. spCSR is based on the previously described CSR, but unlike CSR only a short region (a "patch") of the gene under investigation is diversified and replicated. This allows the selection of polymerases under conditions where catalytic activity and processivity are compromised to the extent that full self-replication is inefficient. We targeted two specific motifs involved in substrate recognition in the active site of DNA polymerase I from *Thermus aquaticus* (Taq) and selected for incorporation of both ribonucleotide- (NTP) and deoxyribonucleotide-triphosphates (dNTPs) using spCSR. This allowed the isolation of multiple variants of Taq with apparent dual substrate specificity. They were able to synthesize RNA, while still retaining essentially wild-type (wt) DNA polymerase activity as judged by PCR. One such mutant (AA40: E602V, A608V, I614M, E615G) was able to incorporate both NTPs and dNTPs with the same catalytic efficiency as the wt enzyme incorporates dNTPs. AA40 allowed the generation of mixed RNA–DNA amplification products in PCR demonstrating DNA polymerase, RNA polymerase as well as reverse transcriptase activity within the same polypeptide. Furthermore, AA40 displayed an expanded substrate spectrum towards other 2′-substituted nucleotides and was able to synthesize nucleic acid polymers in which each base bore a different 2′-substituent. Our results suggest that spCSR will be a powerful strategy for the generation of polymerases with altered substrate specificity for applications in nano- and biotechnology and in the enzymatic synthesis of antisense and RNAi probes.



*Keywords:* DNA polymerases; directed evolution; protein engineering; PCR; enzyme evolution

*Corresponding author

## Introduction

DNA polymerases with altered substrate specificity have a number of important applications in biotechnology but their generation presents formidable challenges. DNA polymerases are unique in recognising four different substrates with exceptionally high specificity. For example, DNA pol I from *Thermus aquaticus* (Taq) misincorporates an incorrect deoxynucleotide triphosphate (dNTP) opposite the template base on average only once every 10,000 bases.[1] This remarkable specificity is due to the fact that polymerase catalytic activity is exquisitely

Present addresses: J.L. Ong, Department of Molecular Biology and Center for Computational and Integrative Biology, Simches Research Center, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA; S. Jaroslawski, Department of Plant Physiology and Biochemistry, Faculty of Biotechnology, Jagiellonian University, ul. Gronostajowa 7, 30-387 Krakow, Poland.

Abbreviations used: spCSR, short-patch compartmentalized self-replication; wt, wild-type; ELISA, enzyme-linked immunosorbent assay.

E-mail address of the corresponding author: ph1@mrc-lmb.cam.ac.uk

sensitive to even slight distortions in the primer–template duplex. Structural studies of Taq[2] and other A-family (Pol I-like) DNA polymerases[3,4] have begun to reveal the molecular basis of polymerase fidelity, revealing how conformational changes upon substrate binding exclude non-cognate base-pairing geometries because of steric conflicts within the closed active site.

In particular, a series of remarkable structures of the DNA pol I from *Bacillus stearothermophilus* (Bst),[5,6] have provided a direct glimpse of how DNA polymerases enhance replication fidelity by a series of interlocking conformational changes. During these, the nascent base-pair and DNA strand proceed through a series of molecular "checkpoints" that stall progression of the catalytic cycle if non-cognate H-bonding[7] or unfavourable steric interactions are detected (reviewed by Kool[8]). These fidelity checkpoints not only prevent the incorporation of mispairs into the nascent DNA but also very effectively exclude the incorporation and/or extension of unnatural non-cognate nucleotide substrates.

Unnatural nucleotide substrates can differ from their natural counterparts in many ways. They may contain bulky substituents on the nucleobase or ribofuranose scaffold that may mimic forms of alkylation damage in DNA and conflict directly with the steric control in the polymerase active site.[9] They may lack minor groove hydrogen-bonding characteristic of cognate base-pairs, absence of which can reduce incorporation and extension by several orders of magnitude.[7] Alternatively, they may cause direct or indirect distortions of the DNA geometry in the active site, for example through intercalation into the template strand base-stack (P. H. & D.L., unpublished results) or by non-cognate conformational preferences.[10] Engineering polymerases therefore requires a reshaping of specific regions of the polymerase active site to accommodate non-cognate chemical modifications, while maintaining cognate interactions or introducing a sufficient number of compensatory interactions to ensure progression through the fidelity checkpoints.

Polymerases have been engineered for the acceptance of unnatural substrates by design, screening and selection. Rational design has allowed the engineering of mutants of Taq polymerase with a much improved ability to incorporate dideoxynucleotide triphosphates (ddNTPs),[11] and mutants of Moloney murine leukemia reverse transcriptase,[12] *Escherichia coli* DNA pol I (Klenow fragment)[13] and *Sulfolobus solfataricus* P1 Dbh[14] with improved ability to utilize ribonucleotides have been engineered. *In vivo* selection and screening has resulted in the identification of a number of polymerases with interesting properties including variants of Taq with up to $10^3$-fold improved incorporation of NTPs.[15] Phage selection has allowed the isolation of mutants of the Stoffel fragment of Taq polymerase with even more strikingly improved incorporation of NTPs (up to $10^4$-fold),[16] 2′-$OCH_3$ substituted NTPs[17] as well as the unnatural deoxyribosyl self-pair PICS.[18]

We have previously described a strategy for the directed evolution of polymerases called compartmentalized self-replication (CSR).[19] CSR is based on a positive feedback loop, whereby a polymerase replicates its own encoding gene. Compartmentalization of the self-replication reactions into individual, non-communicating aqueous compartments of a water-in-oil emulsion[20] ensures a linkage of genotype and phenotype, i.e. it ensures that each polymerase only replicates its own encoding gene. Under these circumstances adaptive gains by the polymerase translate directly (and proportionally) into more efficient self-replication and hence an increased copy number of those genes encoding an active polymerase. Genes encoding inactive polymerases (or polymerases that are poorly active under the selection conditions) will decrease in number and eventually disappear from the gene pool.

CSR has proven to be a powerful method for the directed evolution of polymerase function and has yielded, among others, mutants of Taq polymerase with enhanced thermostability,[19] increased resistance to the potent inhibitor heparin[19] and a generically expanded substrate spectrum by selecting for extension of distorting 3′-mismatches.[21]

Although selection for mismatch extension appears a promising and potentially general strategy for obtaining polymerases with a desired substrate spectrum, we reasoned that it would be desirable to be able to select directly for specific alterations in substrate specificity. Depending on the substrate, such an adaptation may require reshaping of the polymerase active site, which is likely to be accompanied by an at least temporary drop in catalytic efficiency. CSR, however, requires replication of the entire polymerase gene (Figure 1) and thus makes stringent demands on the catalytic efficiency and processivity of selected polymerases. Such onerous requirements, while desirable at later stages of polymerase evolution, were likely to limit our ability to select for incorporation of unfavourable substrates (or for adaptation to other selection conditions that strongly inhibit polymerase activity). In order to reduce the adaptive burden, we devised an alternative strategy called short-patch CSR (spCSR), in which only a short, defined segment of the polymerase gene is self-replicated and evolved (Figure 1). Consequently, the activity and processivity barrier required for a polymerase variant to self-replicate and "survive" a round of selection is lowered. Therefore, spCSR should be a considerably more sensitive and less stringent method, allowing the rescue of a wider spectrum of polymerase mutants.

Here we evaluate the potential of spCSR as a directed evolution strategy by selecting for RNA polymerase activity in a DNA polymerase as a model system. Ribonucleotides differ from deoxyribonucleotides only by the presence of an additional oxygen atom in the 2′-position of the ribose sugar, but are excluded from incorporation into DNA by a factor of $10^5$. Many of the structural mechanisms of ribonucleotide discrimination by