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

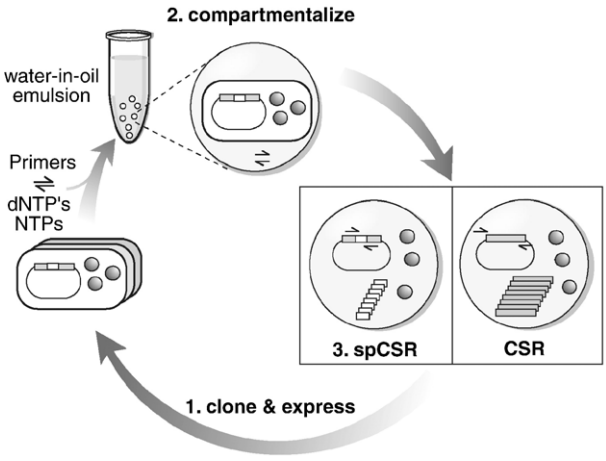


Figure 1. General scheme of short-patch CSR. Compartmentalized self-replication (CSR)[19] is based on a positive feedback loop whereby a polymerase replicates its own encoding gene within the aqueous compartments of a water-in-oil emulsion.[20] The CSR cycle begins with cloning and expression of a repertoire of diversified polymerase genes into *E.coli* (1). In this study, regions of the *T. aquaticus* pol I gene, which are structurally implicated in nucleotide discrimination, were targeted for random mutagenesis. In the next step (2), the bacterial cells expressing Taq polymerase variants are combined with suitable reagents for replication of the diversified region (e.g. flanking primers, NTPs, dNTPs, etc.) and segregated into the aqueous compartments of a heat-stable water-in-oil emulsion[19] in such a way that each compartment on average contains just a single cell (2). Thermocycling ruptures the bacterial cell (and inactivates background cellular enzymatic activities) releasing polymerase and encoding genes into the compartments allowing self-replication to proceed (3). In CSR (right panel) self-replication targets the whole gene, therefore only polymerases capable of replicating their whole encoding gene are selected, while those that are inactive or incapable of full self-replication are lost from the gene pool. In short-patch CSR (left-panel) only a short, defined segment is replicated and propagated to the next generation, therefore polymerases need only be capable of replicating a short segment to be selected. After self-replication, the emulsion is broken, self-replicated polymerase genes (CSR) or polymerase gene fragments (short-patch CSR) are harvested, pooled, reverse transcribed, reamplified and recloned for another round of selection.

DNA polymerases have been studied in some detail,[13,15,16] and selected mutations can be rationalized within this context. Finally, there is a precedent for the successful selection of mutants of the Stoffel fragment of Taq DNA polymerase with non-processive RNA polymerase activity from a phage display library.[16]

We diversified two regions of the active site of Taq DNA polymerase and performed selections for the incorporation of both NTPs and dNTPs using spCSR. We expressed the selected polymerases recombinantly and assayed their ability to incorporate NTPs and dNTPs by steady-state kinetic analysis, primer extension, PCR and a novel enzyme-linked immunosorbent assay (ELISA)-based endpoint assay.

Results

Selection strategy and model selections

spCSR presupposes some knowledge as to where useful diversity may be introduced since only a segment of the polymerase gene is replicated and propagated to the next round of selection (Figure 1). Molecular diversity outside this segment is not passed on to the next generation of polymerases. We prepared two Taq polymerase libraries (A1 and A2) in which diversity was targeted to two short adjacent segments comprising part of the polymerase active site and proximal region. In library A1, residues 611–617, were diversified comprising the floor of the nucleotide binding pocket and the steric gate residue E615. In library A2, residues 597-609, were diversified comprising the third β-sheet of the polymerase palm domain. The two patches of diversity flank the invariant D610, one of two aspartates that chelate the catalytic Mg ions. Diversity was introduced by cloning of degenerate oligonucleotides with a doping ratio of 15% (library A1) or 7.5% (library A2), resulting in an average of three mutations per clone for both libraries.

To test and optimize spCSR, we performed parallel activity selections using CSR and spCSR. *E. coli* cells expressing Taq library A1 were emulsified and selected for their ability to replicate either a short segment covering the region diversified in library A1 (109 bp, spCSR) or the entire Taq gene (2.5 kb, CSR) using their cognate dNTP substrates. After one round of selection, individual mutants were cloned and screened for their ability to amplify a short target sequence (100 bp) in PCR. Ten active mutants from each selection were sequenced. We observed few striking differences between spCSR and CSR selections in either the percentage of PCR-active mutants (ca 30%) or their sequences. In both spCSR and CSR selections, the majority of active mutants corresponded to wtTaq (6/10) but contained a number of silent mutations, indicating they derived from library A1, rather than from contamination. Non-wt mutant sequences contained divergent but conservative single mutations for spCSR (I614M (2×), Y611F, L616V) and for CSR, a single non-conservative mutation (S612R) (3×).

Early on we noticed that there was a significantly greater degree of background signal for spCSR than for CSR selections (i.e. for negative controls in which one of the dNTPs was left out). This is due to the absence of cutting sites for DpnI restriction endonuclease in the short-patch region, preventing the efficient removal of the parent sequences by DpnI digestion (which specifically targets *dam* methylated DNA). We transformed spCSR selections into the *dut*$^-$ *ung*$^-$ *E. coli* strain CJ236[22] as the cloning and expression host. In CJ236, deoxythymidine (dT) in DNA is largely replaced by deoxyuridine (dU), which can be excised by uracil-DNA glycosylase (UDG) to yield abasic sites and render the DNA

"unamplifiable". This strategy indeed greatly reduced spCSR background.

Selection for NTP incorporation

We proceeded to apply spCSR to the selection of polymerases with an enhanced ability to incorporate NTPs. We performed spCSR selections on both libraries A1 and A2 using reaction mixtures comprising three dNTPs and one NTP (e.g. dTTP, dGTP, dCTP + ATP). The selection products were reverse-transcribed; PCR amplified and cloned back into the Taq gene using appropriate restriction enzymes. Library A1 (109 bp) and B (152 bp) both yielded successful selections for the incorporation of ATP, while library A2 was also successfully selected for the incorporation of UTP (Figure 2(a)). Selections for CTP or GTP incorporation were unsuccessful from both presumably due to the relatively high GC content in the Taq gene (63% GC for the region 597–617). Selected clones from the first round were combined and recloned (Figure 2(b)) for a further round of selection for ATP incorporation. The segment amplified in round two now spanned both libraries A1 and A2 (597–617) (152 bp, requiring incorporation and replication of 60 A on both strands) and was selected for ATP incorporation (Figure 2(c)) and recloned for screening.

Polymerase ELISA end-point assay

For the effective screening of large number of individual clones, we developed a rapid and sensitive ELISA-based assay for polymerase activity. It is based on a hairpin DNA oligonucleotide, which serves as both primer and template in one (Figure 3(a)). The hairpin includes a biotin tag for capture on streptavidin-coated microtiter plates. Solid-phase capture allows detection of primer extension by incorporation of tagged nucleotides, removal of unincorporated nucleotides by simple wash steps and high-throughput analysis on crude lysates in a 96 or 384 well format. To detect NTP incorporation, we performed primer extensions under selection conditions, i.e. in the presence of three dNTPs (dCTP, dGTP, dATP) and one NTP (UTP), or two dNTPs (e.g. dCTP, dGTP) and two NTPs (ATP, UTP) with trace amounts of digoxigenin (Dig) labelled UTP (Dig-12-UTP)) (Figure 3(b)). This allows colorimetric detection of Dig-12-UTP incorporation using an anti-Dig antibody horseradish peroxidase (HRP) conjugate. The assay provides a reliable, semi-quantitative measurement of ribonucleotide incorporation and proved much more sensitive than a PCR-based screening assay[19] (not shown).

Of the 600 variants tested for incorporation of UTP using the primer extension ELISA, the best 36 were chosen for further study and assayed for ATP (and UTP) (Figure 3(b)), for CTP and GTP incorporation (not shown) as well as for incorporation of all four NTPs (Figure 3(c)). The best seven clones (as ranked by their ability to incorporate all four NTPs in the ELISA assay) were sequenced. Sequences revealed a

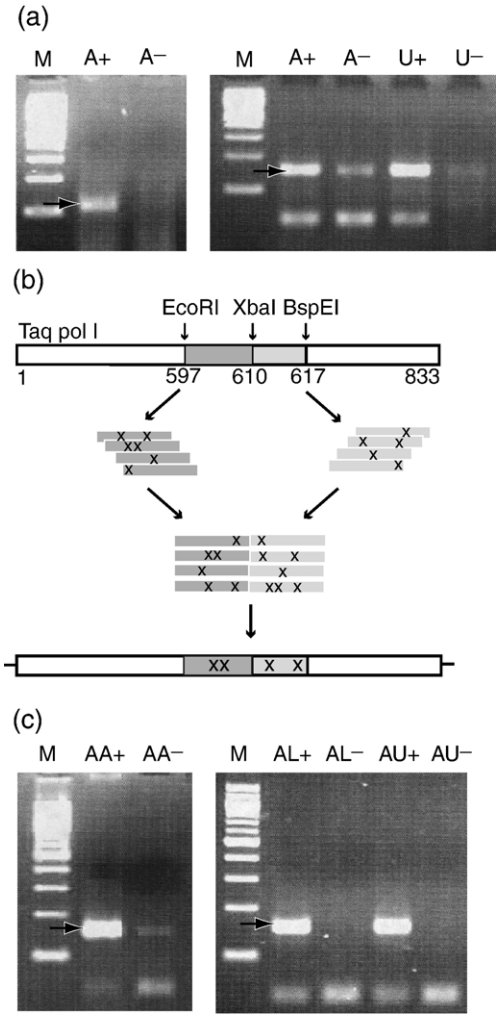


Figure 2. Short-patch CSR selection with ATP and UTP. (a) First round selection of library A1 (611-617) (left panel) and library A2 (597-609) (right panel) with complete replacement of dATP by ATP or dTTP by UTP. RT-PCR bands (arrow) denote successful self-replication in the presence (A+, U+) or absence (A-, U-) of ATP or UTP. (b) First round selections from library A1 and A2 were recombined through an internal XbaI site and recloned for round 2 selection. (c) Round 2 selections (597–617) for ATP incorporation. RT-PCR bands (arrow) denote successful self-replication from combined libraries: AA+/− from library A1 (ATP) × library A2 (ATP), AA+/− from library A1 (ATP) × library A2 (ATP); AU+/− from library A1 (ATP) × library A2 (UTP); AL+/− from library A1 (ATP × unselected library A2). M shows a 100 bp DNA ladder.

strict requirement for the steric gate residue (E615) to be mutated to a glycine. Other mutations were less conserved, although several clones displayed mutations at the active site residues I614, S612 as well as the more distal residue A608 (Figure 4).

Primer extension

Two Taq mutants, I614K[15] and SFR3 (A597T, W604R, L605Q, I614T, E615G)[16] previously described to display an enhanced ability to incorporate

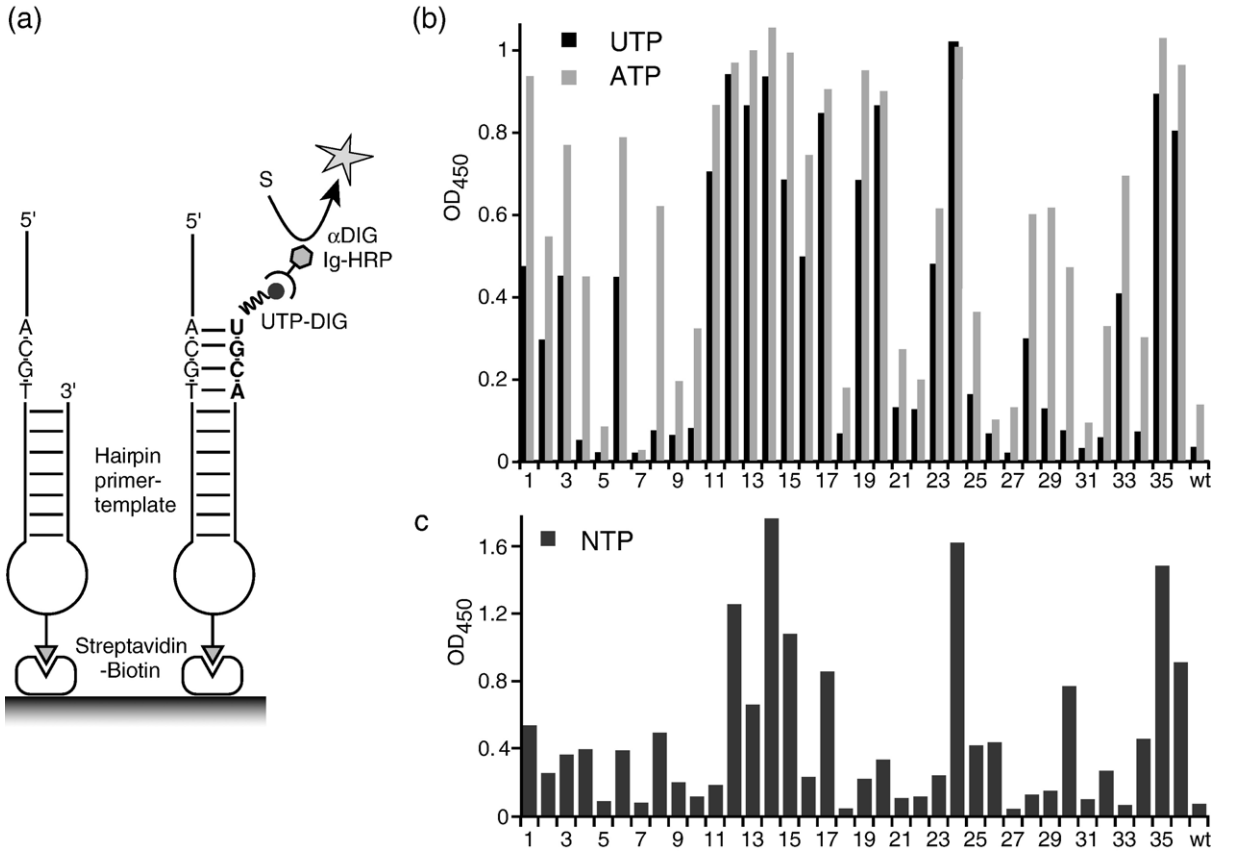


Figure 3. Polymerase ELISA. (a) Schematic of polymerase ELISA endpoint assay. A hairpin primer comprising dU-biotin in the loop is extended by a polymerase using NTPs (or mixtures of dNTPs and NTPs) as well as digoxigenin-12-UTP (UTP-DIG). Reactions are captured on streptavidin-coated microtiter plates and levels of end-point incorporation are analysed using an anti-DIG antibody-horseradish peroxidase (HRP) conjugate. (b) and (c) ELISA data for (b) ATP (black bars), UTP (grey bars) or (c) NTP incorporation for selected 35 clones. For comparison wtTaq is on the far right.

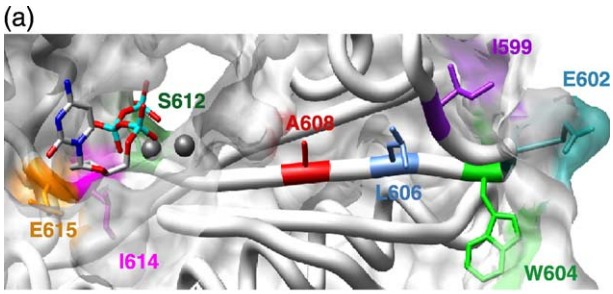
ribonucleotides were constructed for comparison. I614K, SFR3 and spCSR selected clones were expressed and purified using an N-terminal hexahistidine tag[19] and the ability of the purified enzymes to incorporate NTPs and dNTPs in primer extension, PCR and single nucleotide incorporation kinetic assays was determined.

The ability of the spCSR selected mutants as well as I614K and SFR3 to extend a short DNA primer using NTPs under standard conditions (1.5 mM $MgCl_2$ and 200 mM NTPs, 60 °C) was analysed first. There was no detectable primer extension by wtTaq under these conditions, while all but one of the selected Taq mutants were able to extend the primer by six NTPs (Figure 5(a)), although one clone (AL42), showed substantial pausing. I614K[15] maximally incorporated two NTPs, while the SFR3 mutant[16] incorporated six NTPs, with similar efficiency to the best mutant polymerases selected by spCSR. Under forcing conditions (by increasing the Mg^{2+} (3 mM) and NTP (0.5 mM) and adding Mn^{2+} (0.5 mM)), some mutants were able to synthesize longer stretches of RNA (up to 14 bases) (Figure 5(b)). However, under these conditions, even wtTaq polymerase could (inefficiently) incorporate up to three NTPs.

DNA and RNA–DNA PCR

We also investigated the activity of Taq mutants in PCR. All seven spCSR-selected mutants were able to PCR amplify both a 0.4 kb and a 1 kb target with variable efficiency (Figure 6(a)). Comparison of product band intensity provides a semi-quantitative measure of PCR efficiency. By this measure, the two best mutants, AA40 and AL42, show near wild-type activity in PCR, while the PCR efficiency of the remaining mutants appears to be compromised to varying degrees, with SFR3 unable to produce any amplification products in PCR (not shown).

Hybrid RNA–DNA PCR activity has not previously been described. However, as our selection strategy involved incorporation of both dNTPs and NTPs we tested the ability of the selected enzymes to generate hybrid RNA–DNA PCR products utilizing both dNTPs and NTPs. This is a stringent demand, as exponential amplification in PCR requires not only the synthesis of mixed RNA–DNA products but also their utilization as templates in subsequent rounds of amplification. In other words, the enzyme must display both DNA and RNA polymerase activity as well as substantial reverse transcriptase activity. Nevertheless, we



(b)

Taq	A F I A E E G W L L V A L D Y S Q I E L R
AA22	A F I A E E G W L L V D L D Y S Q M G L R
AU14	A F I A E E G W L L V A L D Y S Q I G L R
AA40	A F I A E V G W L L V V L D Y S Q M G L R
AA14	A F I A E E G W L L V A L D Y G Q I G L R
AA27	A F V A E E G C L L V A L D Y S Q I G L R
AL42	A F I A E E G W L F V A L D Y S Q I G L R
AA24	A F I A E E E W L L V A L D Y R Q I G L R
I614K	A F I A E E G W L L V A L D Y S Q K E L R
SFR3	T F I A E E G R Q L V A L D Y S Q T G L R

Figure 4. Mutations conferring RNA polymerase activity. (a) Residues implicated in conferring RNA polymerase activity are highlighted and side-chains are shown as stick models in the Taq structure 3ktq.pdb.[2] Also shown is bound ddCTP substrate. (b) Sequences of selected clones are shown and mutated residues highlighted in colour. Also shown are the sequences of mutants I614K[15] and SFR3.[16]

found that under optimized conditions two mutants AL42 (L606F, E615G) and AA40 (E602V, A608V, I614M, E615G) were able to synthesize hybrid DNA–RNA products in PCR (Figure 6(b)) with robust efficiency utilizing dGTP, dTTP, dCTP and ATP. In order to exclude the possibility that the observed PCR products were the product of dATP contamination of the ATP stock and confirm that the PCR products were indeed DNA–RNA hybrid sequences, we tested their susceptibility to base-catalyzed hydrolysis. Indeed, we found that the PCR products were completely degraded by incubation in sodium hydroxide, while an equivalent all dNTP PCR product remained intact (Figure 6(c)).

Single nucleotide incorporation kinetics

We determined single nucleotide incorporation kinetics for wtTaq and mutant AA40 using a gel-based assay[23] (Table 1). AA40 was chosen because of its apparent dual substrate specificity: it was capable of both chimeric RNA–DNA synthesis in PCR (Figure 6(b)) as well as RNA synthesis with only minor pausing (Figure 5(a)). Single nucleotide incorporation kinetics revealed a 10^3–10^5-fold improvement in the efficiency of incorporation of the first ribonucleotide by AA40 (Table 1). Remarkably, while dNTPs are incorporated by AA40 with comparable efficiency to wtTaq, NTPs are incorporated only slightly less efficiently. In other words, although AA40 was selected for its ability to recognize and incorporate ATP, it has acquired a generalized ability to recognize and incorporate all four ribonucleotides with the catalytic efficiency with which the wild-type DNA polymerase incorporates its cognate dNTP substrates.

We compared the relative efficiencies (f) of dNTP *versus* NTP incorporation in AA40 and a number of other mutants. wtTaq was found to favour dNTPs over NTPs by a factor of 10^3 (dATP/ATP; dGTP/GTP; dCTP/CTP) to 10^6 (dTTP/UTP), while AA40 has a f value (dNTP/NTP) close to 1 for all four bases, indicating that it has completely lost any discrimination between deoxy- and ribo-nucleotide triphosphates, while retaining near wild-type catalytic activity (Table 2).

Expanded substrate spectrum

We found that the activity of AA40 is enhanced not only towards NTPs, but also towards nucleotide triphosphates with other 2′-substituents, in particular 2′-fluoro- (2′-F) and 2′-azido- (2′-N_3) derivatives (Figure 7(a)). In the case of 2′-fluoro substitution, AA40 can even perform PCR amplifications in which dTTP is entirely replaced by 2′-fluoro-2′-deoxyuridine-5′-triphosphate (2′-F-dUTP) (Figure 7(b)). Larger 2′-substituents such as NH_2 or OCH_3, while incorporated with good efficiency, stall further extension after just two incorporation steps. An even poorer substrate is the 2′-arabino–triphosphate, in

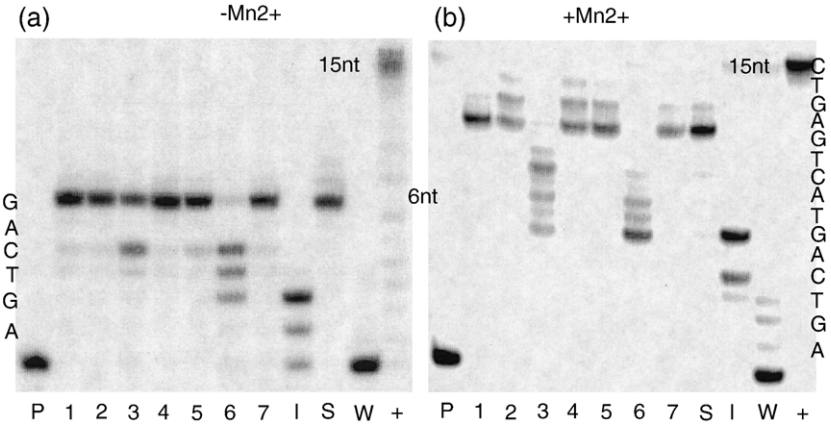


Figure 5. Ribonucleotide incorporation. Primer extension of a 15 nucleotide template under standard ($-Mn^{2+}$) and forcing ($+Mn^{2+}$) conditions is shown. Lane P is non-extended primer. Lanes 1–7 contain selected mutants, in order: AA22, AU14, AA40, AA14, AA27, AL42, AA24. Lane I contains mutant I614K,[15] lane S mutant SFR3[16] and lanes W and + contain wtTaq in the presence of NTPs (W) and dNTPs (+).

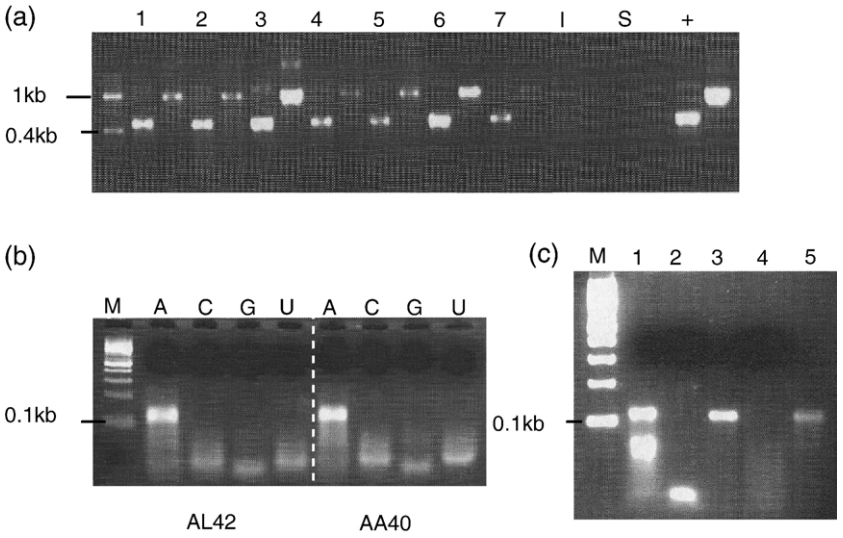


Figure 6. PCR activity of selected clones. (a) PCR amplification of 0.4 kb and 1 kb targets are shown. Lanes 1–7 contain selected mutants, in order: AA22, AU14, AA40, AA14, AA27, AL42, AA24. Lane I contains mutant I614K,[15] lane S mutant SFR3[16] and lane + wtTaq. (b) Hybrid DNA–RNA PCR by mutants AL42 and AA40 of a 0.15 kb target using a mixture of 3dNTPs and 1NTP (for example lane A denotes ATP, dGTP, dCTP, dTTP). (c) NaOH hydrolysis of DNA–RNA PCR products. Lanes 1 and 2 are DNA–RNA PCR amplifications (with ATP, dGTP, dCTP, dTTP) by AL42 (1) or wtTaq (2). wtTaq fails to yield any product. Lane 3 is the same fragment generated by standard PCR (all dNTP). Lanes 4 and 5 show NaOH hydrolysis of fragments from lanes 1(4) and 3(5). The hybrid RNA–DNA PCR product (1) is completely degraded by NaOH (4), whereas the all DNA PCR product (3) is not (5). M shows a 100 bp DNA ladder.

which the 2′-OH group is *trans* to the 3′-OH (as opposed to *cis* for NTPs) acting as a terminator after just one incorporation step. Nevertheless, the expanded substrate spectrum of AA40 allows the synthesis of short nucleic acid polymers comprising a different 2′-substituent on every base such as dGTP (2′-H), CTP (2′-OH), 2′-azido-dATP (2′-N_3) and 2′-fluoro-dUTP (2′-F) (Figure 7(c)).

Discussion

Selected mutations

Comparison of selected sequences revealed a distinct pattern of substitutions with 11 of the 21 randomized amino acid residues mutated at least once (Figure 4). Most frequently mutated are the steric gate residue E615 as well as the adjacent I614. E615 and I614 are in direct proximity to the incoming nucleotide triphosphate. E615 was previously identified as the so-called steric gate residue.[13] The side-chain of E615 packs directly on the 2′ position of the ribofuranose sugar and therefore sterically excludes the incorporation of 2′ substituted nucleotides. Indeed, it was found that mutation to a smaller side-chain (Ala in *E.coli* DNA polymerase I Klenow fragment)[13] allowed efficient incorporation of ribonucleotides. Comparison to sequences obtained in previous screening and selection experiments using *in vivo* complementation[15] or phage display[16] identifies both convergent as well as divergent features (Figure 4(b)). While mutation hot-spots such as E615 and I614 are largely shared the detailed pattern of substitutions is not.

For example, in bacterial complementation and screening,[15] mutations at E615 are less frequent than in either phage display or spCSR selection and more conservative (E615D). In phage display and spCSR selections, substitutions at E615 occur more frequently and are invariably E615G. Indeed, the E615G mutation is present in all the best mutants selected by spCSR. One of the mutants (AU14) contains only the single mutation E615G and is among the best mutants at synthesizing RNA, as judged by primer extension assay (Figure 5(a) and (b)), suggesting that the E615G mutation is probably responsible for most of the improved activity on NTPs. However, additional compensating mutations are needed to maintain processive DNA polymerase and reverse transcriptase activity, since this mutant is much less active in PCR. Mutation of a conserved residue at the centre of an enzyme active site is likely to have diverse functional consequences as they participate in manifold interactions.[13,14] Structural studies of Taq and other polA enzymes indeed reveal potential functions of E615 other than the steric gate. For example, the carboxy group of E615 coordinates a water molecule (with N750 and Q754) that defines the minor groove side of the nucleotide-binding pocket in Taq. Furthermore, there is a hydrogen bond to Y671, stabilizing the conserved tyrosine after it is released from the stacking arrangement of the duplex DNA.[2] E615 may thus participate not only in substrate binding but also in translocation and mutations may disrupt these functions.[24]

At the adjacent I614, the three different methods yield different mutations. While the majority of mutations in bacterial complementation are I614N or I614K, phage display preferentially selects I614T, whereas in spCSR mostly a conservative I614M is found. Different substitutions may reflect subtle adaptations to different selection conditions (for example I614K displays significantly reduced efficiency in PCR (Figure 5(c))) or a compromise between conflicting evolutionary demands.

This may also be the case in mutations more distal to the active site such as substitutions at A608 or E602. Two mutants from the spCSR selections, AA22 (A608D, I614M, E615G) and AA40 (E602V, A608V, I614M, E615G) illustrate how these distal mutations may fine-tune disparate activities. AA22 and AA40

Table 1. Single nucleotide incorporation kinetics

	dNTP			NTP		
Nucleotide	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} μM^{-1})	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (min^{-1} μM^{-1})
AA40						
A	6.8	1.8	3.7	6.2	2.3	2.7
C	8.7	1.3	6.5	5.7	0.70	8.2
G	8.5	1.4	6.2	6.5	0.76	8.5
dT/U	7.6	1.9	4.0	4.8	2.7	1.8
wtTaq						
A	4.8	2.0	2.4	n.d.	n.d.	8.0×10^{-6} [a]
C	4.4	0.44	10	n.d.	n.d.	1.5×10^{-3}
G	4.0	0.92	4.3	n.d.	n.d.	1.2×10^{-3} [a]
dT/U	3.7	1.5	2.5	n.d.	n.d.	8.3×10^{-6} [a]

n.d., not determined.
[a] From Xia *et al.*,[16] measured for the wtTaq Stoffel fragment.

differ only by a mutation E602V and the type of substitution at position 608 but display strikingly different activities in PCR. A conservative mutation at 608 (A608V, (AA40)) in combination with E602V retains full activity in PCR and even allows the generation of mixed RNA–DNA amplification products but leads to some pausing during RNA synthesis as judged by primer extension (Figure 5(a) and (b)), whereas a non-conservative substitution A608D (AA22) increases processive RNA synthesis in primer extension (Figure 5(a) and (b)),) but significantly decreases activity in PCR with both dNTPs and dNTP/NTP mixtures.

Comparison of mutants

Previous efforts towards engineering an RNA polymerase from Taq DNA polymerase yielded mutants, which can incorporate ribonucleotides to a greater extent than wtTaq. However, RNA polymerase activity was obtained at the expense of DNA polymerase activity and processivity.

A comparison of the Taq mutants with enhanced RNA polymerase activity isolated by three different methods reveals how their properties reflect the method by which they were obtained. Screening of a complementation library yielded mutants such as I614K, which displays weak RNA polymerase but robust DNA polymerase activity reflecting its ability to complement for DNA polymerase I *in vivo*.[15] Selection for the ability to synthesize a short (4nt) stretch of RNA by phage display yielded several mutants such as SFR3[16] with good RNA polymerase activity but reduced DNA polymerase activity and processivity as judged by PCR (Figure 5(c)). The mutants described here were selected to replicate a

Table 2. Comparison of single nucleotide incorporation kinetics of wtTaq and different mutant polymerases

	Nucleotide	*f*(dNTP)[a]	*f*(NTP)[a]	*f* (dNTP/NTP)[b]	Reference/*T*(°C)
wtTaq	A	10	0.00041	24,000	From Patel and Loeb[15] 55 °C
	C	19	0.00084	23,000	
	G	57	0.0021	27,000	
	dT/U	160	0.00011	1,500,000	
wtTaq	A	0.85	8.0×10^{-6}	110,000	From Xia *et al.*[16] 50 °C
Stoffel	C	3.0	0.0012	2,=500	
fragment	G	4.4	0.0015	2,=900	
	dT/U	0.28	8.3×10^{-6}	33,000	
AA40	A	3.7	2.7	1.4	This work 60 °C
	C	6.5	8.2	0.78	(isolated by spCSR)
	G	6.2	8.5	0.72	
	dT/U	4.0	1.8	2.2	
SFR3	A	0.16	0.097	1.6	From Xia *et al.*[16] 50 °C
	C	0.67	1.3	0.51	(isolated by phage display)
	G	0.50	1.3	0.38	
	dT/U	0.046	0.14	0.32	
I614K	A	19	0.18	110	From Patel and Loeb[15] 55 °C
	C	57	0.40	140	(isolated by *in vivo* complementation)
	G	13	0.64	19	
	dT/U	89	0.021	4,=200	

[a] The efficiency of incorporation, *f*, is the ratio of k_{cat}/K_M (min^{-1} μM^{-1}).
[b] The selectivity for discrimination between dNTPs and NTPs, *f*(dNTP/NTP), is the ratio of *f*(dNTP)/*f*(NTP).

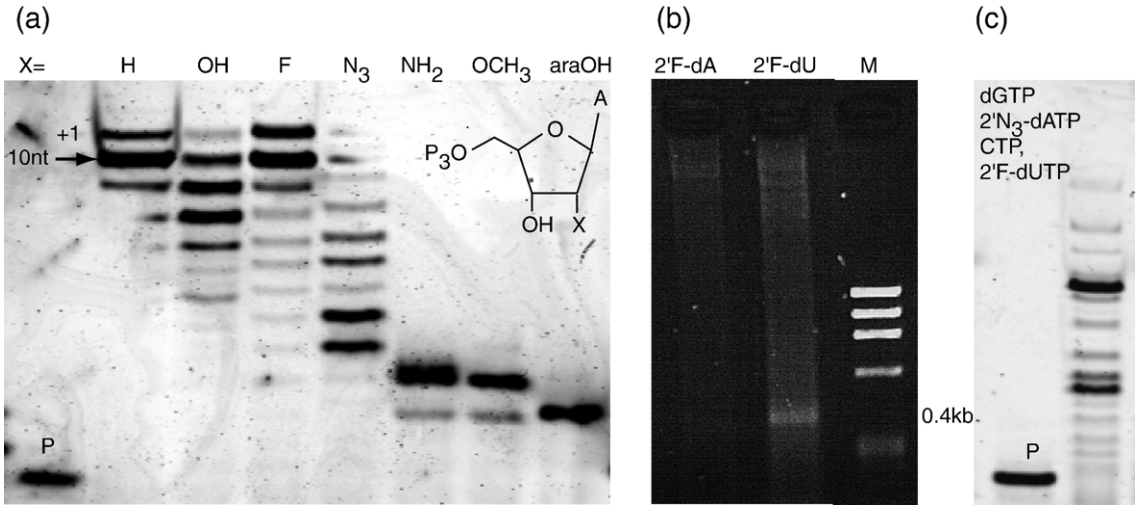


Figure 7. Ability of AA40 to use 2′ substituted nucleotides. (a) Primer extension by AA40 of a ten nucleotide template using 2′-X-2′-deoxyadenosine-5′-triphosphate, with (from left to right) X = H, OH, F, N_3, OCH_3, NH_2 and araOH (2′-arabino-adenosine-5′-triphosphate). (b) PCR amplification by AA40 of a 0.4 kb fragment using either (2′-F-ATP, dCTP, dGTP and dTTP) (left) or (dATP, dCTP, dGTP and 2′-F-dUTP) (right). Only replacement of dTTP with 2′F-dUTP yields a band. M is a ϕX174 DNA-HaeIII digest. (c) Primer extension by AA40 using four different 2′ substituted substrates: dGTP, 2′-N_3-dATP, CTP and 2′-F-dUTP.

short stretch of DNA using a mixture of dNTPs and NTPs and as a result most mutants retain excellent DNA polymerase activity (as judged by PCR) in combination with good RNA polymerase activity. In fact two mutants (AL42 and AA40) were even found to be capable of mixed RNA–DNA PCR amplifications thus displaying DNA and RNA polymerase as well as reverse transcriptase activity.

The respective methods of isolation are also evident in the incorporation kinetics (Table 1), in particular the relative efficiency, *f*, of dNTP *versus* NTP incorporation (Table 2). While I614K preferentially incorporates dNTPs over NTPs by 19 to 4200-fold depending on the nucleotide (Table 2), SFR3 favours NTPs over dNTPs by a factor of two to three with the exception of ATP, which is incorporated less efficiently than dATP (by a factor of 1.6) (Table 2). AA40, one of the mutants selected by spCSR, displays remarkable incorporation kinetics: not only has it retained essentially wild-type catalytic efficiency for dNTP incorporation (as reflected by its proficiency in PCR), but now displays essentially the same k_{cat}/K_M values for all four NTPs and dNTPs as the wild-type enzyme does for the incorporation of dNTPs. Therefore, AA40 does not discriminate at all between NTPs and its natural substrate dNTPs (displaying *f* values close to 1) (Table 2), while retaining essentially wild-type catalytic activity and processivity.

AA40 thus appears to have acquired wild-type activity towards a non-native substrate, without any loss of activity towards its native substrate. It has generally been assumed that adaptation to new substrates involves evolutionary intermediates of lower catalytic efficiency and/or reduced substrate specificity[25,26] and that adaptation to a new substrate invariably entails a negative trade-off in native activity. However, taking advantage of the promiscuous activities of enzymes towards non-native substrates it was found recently that it is possible to enhance the activity of some enzymes for such substrates by >100-fold without compromising native activity.[27] AA40 might be viewed as an extreme example of this, whereby the very weak inherent activity of the wild-type enzyme towards NTPs has been enhanced by up to 10^5-fold.

AA40 displays an expanded substrate spectrum towards a range of non-native 2′ substituted nucleotide substrates (Figure 7), and is even able to synthesize short nucleic acid polymers in which every nucleotide bears a different 2′ substituent approaching the activities of engineered RNA polymerases such as phage T7 RNA polymerase.[28,29] While larger substituents such as NH_2 or OCH_3 are only poorly extended, it is clear that AA40 represents an ideal starting point for the further engineering of a polymerase ultimately capable of utilizing 2′-modified nucleotides in PCR. AA40 must also display substantial reverse transcriptase activity as otherwise exponential PCR amplification of chimeric DNA–RNA fragments would be impossible. Reverse transcriptase activity has previously been observed in the related DNA polymerase I from *T. thermophilus* (Tth) in the presence of Mn^{2+}.[30] Comparison of the Taq and Tth sequence in the 597-617 region reveals divergence at positions also found in selected polymerases such as I559V (Tth), (I599V, AA27); E602A (Tth), (E602V, AA40) and L606A (Tth), (L606Q, SFR3,[16]), strongly suggesting that some of these conservative mutations may contribute to reverse transcriptase activity.

Despite its highly efficient incorporation kinetics for ribonucleoside triphosphates (NTPs), AA40 is still only able to synthesize short RNA polymers (hexaribonucleotide (Np_6), under standard conditions (Figure 5(a))). Indeed, DNA polymerase variants derived by design,[12] screening,[15] or selection[16] for

NTP incorporation display a synthesis arrest at around the same length. Why this sudden termination of RNA polymerization after six nucleotides from such structurally and functionally diverse polymerases? We would like to propose the hypothesis that the termination of synthesis after six nucleotides reflects the conformational preferences of the primer template duplex, which are altered upon synthesis of an increasing stretch of RNA–DNA duplex.

RNA prefers a C3′-*endo* conformation and as a result double-stranded RNA helices adopt *A*-form. While DNA preferentially adopt a C2′-*endo* sugar pucker resulting in *B*-form helix, its conformational preferences can be shifted by interactions with proteins or other nucleic acids. DNA–RNA duplexes appear to either adopt an overall *A*-form[31] or an intermediate between an *A*- and *B*-form.[32,33] In the active site of Taq polymerase the DNA in the primer-template duplex is unwound up to *n*–2 (with n the 3′-nucleotide).[2] The ribofuranose ring of base pairs (*n*, *n*-1, *n*-2) thus display the C3′-*endo* sugar pucker characteristic of the *A*–form. Therefore, the first few nucleotides of RNA synthesized on the DNA template can adopt their cognate, preferred *A*-conformation. However, as the primer is extended further, longer DNA–RNA duplex products increasingly deviate from the preferred conformation as base pairs (*n*-3, *n*-4, *n*-5 etc.) preferentially adopt C2′-*endo* sugar pucker and *B*-form DNA.[2]

Superposing duplex DNA (as bound to the active site of Taq pol I[2]) over an RNA–DNA hybrid[31] illustrates convergence and divergence of the duplexes (Figure 8). While there is a close match for the first five base-pairs, after *n*-5 DNA and RNA-DNA diverge sharply and the RNA–DNA duplex clashes with the thumb domain of the polymerase (not shown). Based on this model, we speculate that the synthesis of a RNA–DNA duplex of a certain length stabilizes *A*-form conformation and disrupts the transition to *B*-form DNA. Consequently, extension arrests around *n*-5, after synthesis of Np_6, due to steric conflict with the polymerase structure.

Detecting aberrant conformational preferences (such as altering the transition equilibrium between the apomorphic *A* and *B* forms) may be a sophisticated molecular mechanism to minimize the permanent incorporation of non-cognate nucleic acids into the nascent strand, just as it appears to prevent permanent incorporation of mispairs. Structural snapshots of the extension of a G·T mispair in the catalytically active crystals of Bst pol I revealed a disruption of the *A* to *B*-form transition by the mispair, which was "transmitted" back to the active site and promoted stalling of the polymerase,[5] allowing for enhanced error detection and editing.

spCSR

Here, we have expanded on the previously described strategy of CSR (compartmentalized self-replication) for the directed evolution of polymerases[19,21] and developed short-patch CSR (spCSR). As we have shown here spCSR allows a direct selection for incorporation, extension and

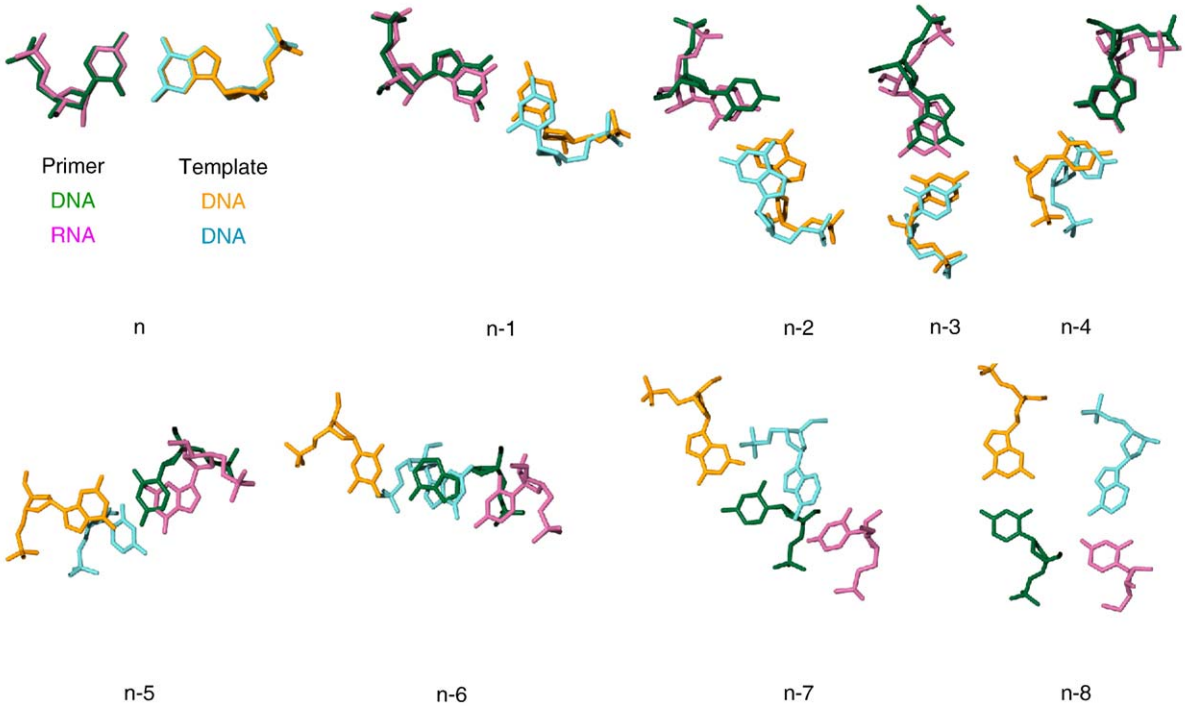


Figure 8. Superposition of DNA and RNA–DNA duplex. DNA primer (green) and template (blue) duplex bound to the active site of Taq polymerase from 3ktq.pdb is superimposed at the 3′ base-pair *n* with RNA (magenta)–DNA (orange) duplex up to base-pair *n*-8. Superposition is good for base-pairs *n* to *n*-4, as interaction with the polymerase induces *A*-form DNA in base-pairs *n* to *n*-3. From base-pair *n*-5 there is increasingly poor overlap and steric clash with the polymerase thumb domain (not shown).

replication of modified nucleotide substrates. The spCSR strategy is "tuneable", such that with increasing polymerase activity and processivity, the size of the region mutated and replicated in later rounds can also be increased. In this way, larger regions of the polymerase gene can be scanned for mutations. This may include either the extension of the original region of diversity to either or both sides or the link-up of separate and separated patches, ultimately leading to complete or near-complete self-replication as described in CSR.[19,21]

Unlike phage display or bacterial complementation methods, the selection pressure in spCSR is towards a polymerase active in PCR (Figures 6 and 7). Thus, multiple properties are optimized simultaneously including nucleotide triphosphate binding, incorporation and extension efficiency, as well as fidelity and to a lesser extent processivity. In addition, other properties important for polymerase function are also under selective pressure including thermostability, folding and solubility.

spCSR requires an educated guess about which regions of the polymerase will yield increases in the incorporation of the nucleotide of choice when mutated. This remains challenging despite a great amount of biochemical and structural data available to guide selection. spCSR also still requires the replication of (at least) 50–100 bases, which may exclude a number of exceedingly poor substrates. Finally, spCSR may also be useful for probing the active site of DNA polymerases. Selections for active polymerases in which different regions have been subject to diversification may yield insights into polymerase function and pinpoint critical residues. For example, when we applied spCSR to the selection of polymerases capable of substituting dATP with the fluorescent dye-labelled nucleotide FITC-12-dATP, we identified an universally mutated residue S612 (N, R) (J.L.O., N. Ramsay, P.H., unpublished results), which may act as a "Fluorogate" in analogy to the steric gate E615G for ribonucleotides.

In conclusion, spCSR allowed the isolation of polymerase mutants, which are able to incorporate NTPs and dNTPs with equal kinetic efficiency than the wt enzyme incorporates dNTPs, while retaining near wt activity in PCR. This is reflected in the astonishing ability of some of the selected polymerases to generate hybrid DNA–RNA products in PCR thus uniting DNA polymerase, RNA polymerase activity and reverse transcriptase activity in a single polypeptide. This combination of activities has not been previously observed, although sections of the mitochondrial genome have been reported to comprise mixed DNA-RNA oligomers[34] indicating that DNA polγ may possess such activities. Finally, the ability of the selected polymerases to generate hybrid DNA–RNA PCR products and 2′ substituted polymers may have potential applications in biotechnology and molecular genetics, for example in enabling sequencing by partial ribosubstitution[35] or the *in situ* synthesis of substituted nucleic acid polymers for applications in antisense, antigene or RNAi strategies.

Materials and Methods

DNA manipulation

The XbaI site in the multiple cloning site of the Taq polymerase expression vector pASK75-TaqHis6[19] was removed by digesting the plasmid with XbaI, filling in with Klenow (Stratagene) and re-circularizing the plasmid with T4 DNA ligase (New England Biolabs) to give pASK75-TaqHis6(-XbaI). Silent mutations were introduced into the Taq gene to include an unique EcoRI (at 1778), XbaI (1824), and BspEI (1872) restriction sites using primers 1: (5′-CAC CCC GCT TGG GCA GAG AAT TCG CCG GGC CTT CAT CG-3′) and 2: (5′- CGA TGA AGG CCC GGC GAA TTC TCT GCC CAA GCG GGG TG-3′) (EcoRI); 3: 5′- GGT GGC TAT TGG TGG CTC TAG ACT ATA GCC AGA TAG AG-3′) and 4: (5′- CTC TAT CTG GCT ATA GTC TAG AGC CAC CAA TAG CCA CC-3′) (XbaI); 5: (5′- GTG CTG GCC CAC CTC TCC GGA GAC GAG AAC CTG ATC-3′) and 6: (5′- GAT CGA TCA GGT TCT CGT CTC CGG AGA GGT GGG CCA GCA C-3′) (BspE2) and Quikchange mutagenesis (Stratagene) to give pASK75-TaqHis6-X(+Xba1824) (+BspE1872). A stuffer fragment introducing a BglII site, a +1 frameshift and a stop codon was inserted using primers 6: (5′- CTA TTG GTG GCT CTA GAC TAA AGC CAG ATC TAG CTC AGG GTG CTG GCC-3′) and 7 (5′- GGC CAG CAC CCT GAG CTA GAT CTG GCT TTA GTC TAG AGC CAC CAA TAG-3′) and Quikchange mutagenesis as above to give pASK75-TaqΔ611stop. pASK75-TaqΔ611stop does not yield any Taq polymerase activity upon expression but Taq activity can be reconstituted by insertion of the XbaI/BspEI insert from pASK75-Taq-X(+Xba1824)(+BspE1872). Polymerase mutants I614K[15]) and SFR3[16] were constructed using QuikChange mutagenesis as above with primers (5′- TGG ACT ATA GCC AGA AAG AAC TCA GGG TGC TGG CC-3′) and (5′- GGC CAG CAC CCT GAG TTC TTT CTG GCT ATA GTC CA-3′) (I614K) and primers (5′- GCC CTG GAC TAT AGC CAG ACA GGG CTC AGG GTG CTG GCC CAC-3′) and (5′- GTG GGC CAG CAC CCT GAG CCC TGT CTG GCT ATA GTC CAG GGC-3′) (SFR-3).

Library construction

For library construction pASK75-TaqΔ611stop was purified by caesium chloride gradient and digested with XbaI/BspEI (Library A1) and EcoRI/BspEI (Library A2). Library inserts were prepared using a degenerate oligonucleotide template 8 (Library A1): (5′-TGG CTA TTG GTG GCT CTA GAC TAT AGC CAG ATA GAG CTC AGG GTG CTG GCC CAC CTC TCC GGA GAC GAG AAC CTG ATC-3′, the underlined sequence was 85% wild-type, 15% mutant nucleotides, resulting in an average of 3.15 mutations per clone) and template 9 (Library A2): (5′- GCA GAG AAT TCG CCG GGC CTT CAT CGC CGA GGA GGG GTG GCT ATT GGT GGC CCT GGA CTA TAG CC AGA TAG AGC TC-3′; the underlined sequence was 92.5% wild-type, 7.5% mutant nucleotides, resulting in an average of 2.9 mutations per clone), which were amplified using primers 10: (5′- CAG GAA ACA GCT ATG ACC ACG TGG CTA TTG GTG GCC CTG-3′) and 11: (5′- GTA AAA CGA CGG CCA GTG ACT CTC GTC GCC GGA GAG GTG-3′) (Library A1) and 12: (5′-CAG GAA ACA GCT ATG ACC GCT TGG GCA GAG AAT TCG C-3′) and 11, digested with XbaI/BspEI (Library A1) and EcoRI/BspEI (Library A2) cloned into pASK75-TaqΔ611stop. After ligation, non-library vector was destroyed by

digestion with BglII. The library plasmid constructs were transformed into electrocompetent *E. coli* Ace6 (MC1061, *endA*) (Library A1, 2: 8×10^8 cfu (97% insert). For spCSR selections, plasmid DNA was prepared and used to transform *E. coli* CJ236 (*dut*$^-$ *ung*$^-$) (Library A1 (1.6×10^7 cfu); Library A2 (1.0×10^7 cfu)).

Selection

For spCSR test selections, Ace6 (MC1061, *endA*$^-$) cells expressing library A1 were induced for protein expression and induced cells emulsified as described[19] in 1× Taq buffer with primers 10, 11 (for short patch CSR activity selection) and with previously described[19] (primers 4 and 5) primers (4) 13, (5) 14 (for CSR). Emulsions were broken by diethyl ether extraction as described[19] and purified using Nucleotide Removal kit (Qiagen). Remaining primer DNA was degraded with ExoSAP-IT(US Biologicals) and selections were reamplified with primers 15: primer 6[19] and 16: S1211S (New England Biolabs) and cloned XbaI/BspEI into pASK75-TaqΔ611stop or XbaI/SalI into pASK75.

For spCSR selection for NTP incorporation, CJ236 cells expressing libraries A1 and A2 were induced for protein expression, emulsified (as above) in 1× Taq buffer with primers 10 and 11 (0.5 mM each, Library A1) or primers 11 and 12 (0.5 mM each, Library A2) and ATP, dTTP, dCTP, and dGTP (100 mM each) (or UTP, dATP, dCTP and dGTP for UTP selections on library A2) and thermocycled 94 °C 5 min, 20× (94 °C 30 s, 50 °C 1 min, 72 °C 5 min). Emulsions were broken as described above and DNA purified using buffer PN and a PCR purification kit (Qiagen). The cleaned products were digested with uracil-DNA glycosylase (UDG) (New England Biolabs) (to remove carry-over plasmid DNA from CJ236), phenol/chloroform extracted and ethanol precipitated. DNA was further treated with ExoSAP-IT (USB) in 1× SuperScript buffer, then reamplified using SuperScript One-Step RT-PCR system (both Invitrogen) and primers 15 and 16 and the protocol: 50 °C, 30 min; 94 °C, 2 min; 20× (94 °C, 30 s; 50 °C, 1 min; 72 °C, 2 min). Selections were resolved on 3% (w/v) NuSieve agarose gel, extracted using QIAExII gel extraction kit (Qiagen), reamplified using primers 15 and 16, worked up with a PCR purification kit (Qiagen) and digested with XbaI/BspEI (library A1) or EcoRI and BspEI (library A2). The digest products were resolved on a 3% NuSieve agarose gel (Cambrex), and extracted using a QIAExII gel extraction kit (Qiagen) and ligated into pASK75-TaqΔ611-stop as described above. The ligation was digested with BglII and transformed into Ace6 cells (5×10^6 cfu, ATP selection library A1: 2.3×10^7 cfu (ATP) and library A2: 6.7×10^6 cfu (UTP) library A2). For round 2 spCSR selections, round 1 selections were amplified using primers 10, 11 (library A1) and 12, 17: (5′- GCT CTA TCT GGC TAT AGT CTA GA-3′), digested with XbaI and ligated together with T4 DNA ligase (New England Biolabs). The ligation was reamplified in PCR using primers 15 and 16, and cloned into the pASK75-TaqΔ611stop as described above. spCSR selection for ATP incorporation were carried out as described above for round 1.

Screening for polymerase activity

Polymerase variants were screened (as lysates or directly as induced cells in 1× Taq buffer) by using an end-point assay based on enzyme linked immunosorbent assay (ELISA) with a biotinylated hairpin primer 17: 5′-AGC TAC CAT GCC TGC ACG ACG TCG GCA TCC GTC GCG ACC ACG TTT TTC GTG GTC GCG ACG GAT GCC G-3′, *T* = Biotin-dUTP), 200 μM each dATP, dCTP, dGTP and UTP, 6 μM digoxigenin-11-UTP (or 200 μM each ATP, dCTP, dGTP and dTTP, 6 μM Dig-11-dUTP (Roche) or 200 μM each dNTP, 6 μM Dig-11-UTP) in 1× Taq buffer and incubated 94 °C 5 min, 50 °C 5 min, 72 °C 5 min. Extension products were captured with Streptawell Hi-bind, microtiter plates (Roche) and assayed using anti-DIG antibody-peroxidase conjugate (Roche) according to standard ELISA protocols.

Protein expression

Hexahistidine-tagged pol variants were purified using Ni-NTA agarose (Qiagen), desalted and concentrated using Centricon YM-50 concentrator (Amicon) and stored in 50 mM Tris–HCl (pH 8), 1 mM EDTA, 1 mM DTT and 50% (v/v) glycerol at −20 °C. Protein concentration was determined using BCA Assay (Pierce).

Primer extensions

^{32}P-labeled primer 18: (5′-ATT ATG CTG AGT GAT ATC CCT CT-3′) (50pmol) was annealed to template 19: (5′-GAC TCA GTA CTG ACT AGA GGG ATA TCA CTC AGC ATA AT-3′) or Cy3 5′-labelled primer 20: (5′-Cy3- C*T*C *A*CT ATA GGG A-3′ (*T*, *A* denote LNA-T, A) with template 21: (5′- TTT TTT TTT TCT CCC TAT AGT GAG TCG TAT TA-3′) in 1× Taq buffer (or 1×Taq buffer + 1.5 mM $MgCl_2$, 0.5 mM $MnCl_2$), extended at 60 °C for 30 min using 7.5 nM of enzyme and NTP (0.5 mM each), dNTP or 2′-X-modified NTP (X = F, N_3, NH_2, OCH_3, or arabino-OH (TriLink Biotechnologies)) (0.1 mM), resolved by 20% polyacrylamide/7M urea gel electrophoresis. Band intensities were analysed on a Typhoon 8600 Phosphoimager (Molecular Dynamics) either directly (Cy3) or after 1 h exposure of a phosphoimager storage screen (Molecular Dynamics).

PCR analysis

PCR reactions contained 10 ng of template (pASK75-Taq), 0.2 mM dNTPs, 10 pmol primers 22:(5′-GCG GCC AAG ACC ATC AAC TT-3′) and 23:(5′-ACC ACC GAA CTG CGG GTG ACG CCA AGC G-3′) (0.4 kb product) and 24: (5′-GTG GAG AAG ATC CTG CAG TAC C-3′) and 23 (1 kb product)), 1× Taq buffer, and enzyme (activity normalized to 0.5unit). Reactions were hot-started and cycled 30× (94 °C, 30s; 50 °C, 1 min; 72 °C, 5 min) and resolved on a 1% agarose gel. For hybrid RNA–DNA PCR reactions (0.15 kb product) contained 100 ng template, 1× Taq buffer + 1.5 mM $MgCl_2$, 0.5 mM each (ATP, dTTP, dCTP, dGTP), and primers 22 and 25: (5′-GAC GGG CAT GTT GAA GGC CAT GC-3′) and were cycled: 40× (94 °C, 5 s; 50 °C, 30 s; 72 °C, 5 min).

Single nucleotide incorporation kinetics

Kinetic parameters were determined using a polyacrylamide gel assay essentially as described.[36] Oligonucleotide 26: (5′-TAA TAC GAC TCA CTA TAG GGA GA-3′) was ^{32}P-labeled, purified on a Microspin G-25 column (Amersham Pharmacia) and annealed to templates 27: (5′-ACT G**A**T CTC CCT ATA GTG AGT CGT ATT A-3′; to measure UTP/dTTP incorporation); 28: (5′-ACT G**TT** CTC CCT ATA GTG AGT CGT ATT A-3′; to measure ATP/dATP incorporation); 29: (5′-ACT G**C**T CTC CCT ATA

GTG AGT CGT ATT A-3′; to measure GTP/dGTP incorporation); 30 and 31:(5′-ACT G**G**T CTC CCT ATA GTG AGT CGT ATT A-3′ and 5′-GAC T**G**T CTC CCT ATA GTG AGT CGT ATT A-3′; to measure CTP/dCTP incorporation);. Duplex substrates were used at 50 nM final concentration in 1×Taq buffer and 2.5 nM enzyme and various concentrations dNTPs or NTPs (156 nM–80 mM). Reactions were carried out at 60 °C for variable times, resolved on 20% (w/v) PAGE/7M urea gels. Band intensities were analyzed on a Typhoon 8600 Phosphoimager (Molecular Dynamics) after 1 h exposure of a phosphoimager storage screen (Molecular Dynamics), quantified using ImageQuant software, and data fit to the Michaelis-Menten equation using GraphPad Prism 4.

Acknowledgements

We thank our colleagues Leo James (MRC-LMB) and Nicola Ramsay (Domantis Ltd) for helpful comments on the manuscript and John Schwabe (MRC-LMB) for help in preparing Figure 8. J.L.O. was supported by stipends from the Cambridge Overseas Trust, Overseas Research Award and Gonville & Caius College.

References

1. Kunkel, T. A. (1992). DNA replication fidelity. *J. Biol. Chem.* **267**, 18251–18254.
2. Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525.
3. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution (see comments). *Nature*, **391**, 251–258.
4. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304–307.
5. Johnson, S. J. & Beese, L. S. (2004). Structures of mismatch replication errors observed in a DNA polymerase. *Cell*, **116**, 803–816.
6. Johnson, S. J., Taylor, J. S. & Beese, L. S. (2003). Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc. Natl Acad. Sci. USA*, **100**, 38895–38900.
7. Morales, J. C. & Kool, E. T. (2000). Functional hydrogen-bonding map of the minor groove binding tracks of six DNA polymerases. *Biochemistry*, **39**, 12979–12988.
8. Kool, E. T. (2002). Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191–219.
9. Hsu, G. W., Huang, X., Luneva, N. P., Geacintov, N. E. & Beese, L. S. (2005). Structure of a high fidelity DNA polymerase bound to a benzo[a]pyrene adduct that blocks replication. *J. Biol. Chem.* **280**, 3764–3770.
10. Hsu, G. W., Ober, M., Carell, T. & Beese, L. S. (2004). Error-prone replication of oxidatively damaged DNA by a high-fidelity DNA polymerase. *Nature*, **431**, 2172–2221.
11. Li, Y., Mitaxov, V. & Waksman, G. (1999). Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl Acad. Sci. USA*, **96**, 9491–9496.
12. Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A. & Goff, S. P. (1997). Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl Acad. Sci. USA*, **94**, 407–411.
13. Astatke, M., Grindley, N. D. & Joyce, C. M. (1998). How *E. coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. *J. Mol. Biol.* **278**, 147–165.
14. DeLucia, A. M., Grindley, N. D. & Joyce, C. M. (2003). An error-prone family Y DNA polymerase (DinB homolog from *Sulfolobus solfataricus*) uses a 'steric gate' residue for discrimination against ribonucleotides. *Nucl. Acids Res.* **31**, 4129–4137.
15. Patel, P. H. & Loeb, L. A. (2000). Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J. Biol. Chem.* **275**, 40266–40272.
16. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G. & Romesberg, F. E. (2002). Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl Acad. Sci. USA*, **99**, 6597–6602.
17. Fa, M., Radeghieri, A., Henry, A. A. & Romesberg, F. E. (2004). Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* **126**, 1748–1754.
18. Leconte, A. M., Chen, L. & Romesberg, F. E. (2005). Polymerase evolution: efforts toward expansion of the genetic code. *J. Am. Chem. Soc.* **127**, 12470–12471.
19. Ghadessy, F. J., Ong, J. L. & Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA*, **98**, 4552–4557.
20. Tawfik, D. S. & Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nature Biotechnol.* **16**, 652–656.
21. Ghadessy, F. J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S. *et al.* (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nature Biotechnol.* **22**, 755–759.
22. Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.
23. Creighton, S., Bloom, L. B. & Goodman, M. F. (1995). Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.* **262**, 232–256.
24. Minnick, D. T., Bebenek, K., Osheroff, W. P., Turner, R. M. J., Astatke, M., Liu, L. *et al.* (1999). Side chains that influence fidelity at the polymerase active site of *Escherichia coli* DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **274**, 3067–3075.
25. Matsumura, I. & Ellington, A. D. (2001). In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* **305**, 331–339.
26. Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425.
27. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C. & Tawfik, D. S. (2005). The 'evolvability' of promiscuous protein functions. *Nature Genet.* **37**, 73–76.
28. Sousa, R. & Padilla, R. (1995). A mutant T7 RNA

polymerase as a DNA polymerase. *EMBO J.* **14**, 4609–4621.
29. Padilla, R. & Sousa, R. (1999). Efficient synthesis of nucleic acids heavily modified with non-canonical ribose 2′-groups using a mutantT7 RNA polymerase (RNAP). *Nucl. Acids Res.* **27**, 1561–1563.
30. Myers, T. W. & Gelfand, D. H. (1991). Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry*, **30**, 7661–7666.
31. Xiong, Y. & Sundaralingam, M. (2000). Crystal structure of a DNA.RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucl. Acids Res.* **28**, 2171–2176.
32. Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S. & Reid, B. R. (1993). The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry*, **32**, 4207–4215.
33. Fedoroff, O. Y., Ge, Y. & Reid, B. R. (1997). Solution structure of r(gaggacug):d(CAGTCCTC) hybrid: implications for the initiation of HIV-1 (+)-strand synthesis. *J. Mol. Biol.* **269**, 225–239.
34. Yang, M. Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E. *et al.* (2002). Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell*, **111**, 495–505.
35. Barnes, W. M. (1978). DNA sequencing by partial ribosubstitution. *J. Mol. Biol.* **119**, 83–99.
36. Creighton, S., Bloom, L. B. & Goodman, M. F. (1995). Gel fidelity assay measuring nucleotide misinsertion, exonucleolytic proofreading, and lesion bypass efficiencies. *Methods Enzymol.* **262**, 232–256.

Edited by F. Schmid

(Received 29 March 2006; received in revised form 13 June 2006; accepted 21 June 2006)
Available online 5 July 2006