Published online 9 November 2010 *Nucleic Acids Research, 2011, Vol. 39, No. 5* ***1801–1810***
doi:10.1093/nar/gkq1053

Engineered split in *Pfu* DNA polymerase fingers domain improves incorporation of nucleotide γ-phosphate derivative

Connie J. Hansen[1], Lydia Wu[2], Jeffrey D. Fox[1], Bahram Arezi[1] and Holly H. Hogrefe[1,*]

[1]Agilent Technologies Inc., Stratagene Products Division, 11011 N. Torrey Pines Road, La Jolla, CA 92037 and [2]Thermo Fisher Scientific, 9389 Waples Street, San Diego, CA 92121, USA

Received June 28, 2010; Revised October 6, 2010; Accepted October 13, 2010

ABSTRACT

Using compartmentalized self-replication (CSR), we evolved a version of *Pyrococcus furiosus (Pfu)* DNA polymerase that tolerates modification of the γ-phosphate of an incoming nucleotide. A Q484R mutation in α-helix P of the fingers domain, coupled with an unintended translational termination-reinitiation (split) near the finger tip, dramatically improve incorporation of a bulky γ-phosphate-*O*-linker-dabcyl substituent. Whether synthesized by coupled translation from a bicistronic (−1 frameshift) clone, or reconstituted from separately expressed and purified fragments, split *Pfu* mutant behaves identically to wild-type DNA polymerase with respect to chromatographic behavior, steady-state kinetic parameters (for dCTP), and PCR performance. Although naturally-occurring splits have been identified previously in the finger tip region of T4 gp43 variants, this is the first time a split (in combination with a point mutation) has been shown to broaden substrate utilization. Moreover, this latest example of a split hyperthermophilic archaeal DNA polymerase further illustrates the modular nature of the Family B DNA polymerase structure.

INTRODUCTION

Terminal phosphate-labeled nucleotides have been used in a variety of applications including DNA sequencing, single-molecule DNA sequencing, SNP detection, quantitative PCR, and enzymatic assays to monitor polymerase activity (1–8). Homogeneous polymerase assays employ phosphate-labeled nucleotides that change spectral properties when tagged polyphosphates are released and nucleoside monophosphates are added to 3′-OH primer termini. Analogs used for qPCR, SNP typing, and enzymatic assays include internally-quenched nucleotides (IQNs) that are doubly modified with a fluorophore on the base and a quencher on the pyrophosphate (1,2) or *vice versa* (3), and terminal phosphate-labeled nucleotides that when incorporated, liberate dye-labeled polyphosphate esters that are immediately hydrolyzed with alkaline phosphatase to generate readily-detectable free dye [anionic form (4–5)].

By far, the greatest utility of terminal phosphate-labeled nucleotides lies in single-molecule DNA sequencing. Emerging third-generation DNA sequencing technologies achieve single-molecule detection and longer read length through the use of nucleoside tri- or penta-phosphates containing fluorescent dyes on the terminal phosphate (6–8). Single-molecule DNA sequencing requires sophisticated imaging platforms to detect the temporal order of addition of four different phosphate-labeled nucleotides by an immobilized DNA polymerase.

Native polymerases incorporate terminal phosphate-labeled nucleotides to varying extents, depending on number of phosphates and ligand-attachment site, choice of fluorophore, and the chemical structure of the phosphate-dye linkage (9). In general, RNA polymerases and reverse transcriptases (RNA-dependant polymerases) incorporate γ-phosphate-modified nucleotides more efficiently than DNA-dependant DNA polymerases (5,10–12), and at least in one case (HIV-1 RT; dNppp-1-aminonaphthalene-5-sulfonate), with greater fidelity than natural nucleotides (10). In a study employing numerous DNA polymerases and polyphosphate derivatives, incorporation of γ-phosphate-labeled nucleotides was modest (e.g. 0.2–6% efficiency with dTppp-C_7-TAMRA; C_7, optimized heptyl linker), and varied up to 30-fold depending on the DNA-dependant DNA polymerase employed (9). In contrast, terminal phosphate-labeled tetra- and penta-phosphates were incorporated up to

*To whom correspondence should be addressed. Tel/Fax: 858 373 6520; Email: holly.hogrefe@agilent.com

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

50-fold more efficiently compared to the corresponding triphosphate, prompting Kumar *et al.* (9) to conclude that triphosphates labeled on the terminal (γ) phosphate are generally poor substrates for DNA and RNA polymerases. These authors attribute improved incorporation of terminal phosphate-labeled tetra- and penta-phosphates to increased distance between the dye and polymerase active site, or alternatively, to compensation of loss of charge when dyes are attached to γ-phosphates (9). Increased tolerance for modified pentaphosphates is exemplified by Korlach *et al.* (13) who demonstrate processive synthesis with φ29 DNA polymerase and terminal phosphate-labeled nucleoside pentaphosphates (dNppppp-Alexa Fluor 488) at 100% substitution.

In this report, we examine structural modifications required to improve DNA polymerase incorporation of a terminal phosphate-modified nucleotide. We use the compartmentalized self-replication (CSR) technique (14) to evolve a mutant of *Pfu* DNA polymerase (*Pfu* PolB) that incorporates γ-phosphate-modified nucleotide with greater efficiency. In these studies, we employ dCppp-Dabcyl, a precursor to IQNs containing a quencher on the pyrophosphate (1,2) and a poor substrate for wild-type *Pfu* DNA polymerase. As we will show, the combination of a split (translational termination-reinitiation) and an amino acid replacement in the fingers domain of *Pfu* PolB increases tolerance for modification of the γ-phosphate without compromising affinity for natural nucleotides.

MATERIALS AND METHODS

Reagents

All molecular biology reagents were from Agilent Technologies-Stratagene Products, unless otherwise noted. dCppp-Dabcyl (Figure 1A) was synthesized by TriLink Biotechnologies (San Diego, CA, USA).

Random mutagenesis and CSR selection

Mutant *Pfu* PolB libraries were constructed by error-prone PCR amplification with Mutazyme I, followed by restriction-free cloning of randomized fragments with the GeneMorph II EZClone Domain Mutagenesis Kit. pET11a-*Pfu*(V93R)S_m expresses a derivative of *Pfu* (exo^+) that contains a V93R mutation to eliminate sensitivity to deoxyuracil in template DNA (15), a C-terminal processivity tag [Sso7d_m, >80% identity to Sso7d, (16)], and a C-terminal polyhistidine tag. A 2.3 Kb *Pfu* gene fragment was amplified (PfuF: 5′-TTTGTTTAACTTTA AGAAGGAGATATAC; Sso7d_mR: 5′-CCGCCACCGC CGGTACC) from 100 ng pET11a-*Pfu* (V93R)-S_m plasmid DNA using 1–2 U Mutazyme I enzyme ('low-mutation rate' conditions of 1–3/Kb, according to manual) to generate a mutant mega-primer that replaced the parental *Pfu* (V93R) sequence in the EZClone reaction. After *Dpn*I digestion and transformation, plasmid DNA was isolated from a pool of approximately 10^5 clones and used to transform BL21-CodonPlus (DE3)-RIL competent cells. Transformants were pooled (approximately 10^5 clones), grown up in 20 ml LB/AMP/CAM to an OD_{600} of 0.4–0.6, induced with 1 mM IPTG, and then incubated 4 h at 30°C. Cells were collected, washed, and resuspended in 1× cPfu buffer.

CSR emulsions were prepared essentially as described (14), except that *cPfu* buffer, *Pfu* expresser cells, and *Pfu*-specific primers were substituted for *Taq* components. CSR conditions were initially verified by enriching *Pfu* V93R clones from artificial libraries consisting of varying proportions (100–1000:1) of polymerase-deficient (*Pfu* V93R/G387P): polymerase-proficient (*Pfu* V93R) expressor cells. To enrich for *Pfu* mutants that utilize dCppp-Dabcyl, dCTP was replaced with dCppp-Dabcyl:dCTP mixtures (250 μM total) consisting of 90%:10% (round 1), 95%:5% (round 2) or 100%:0% (rounds 3–4). Emulsions (100 μl) were amplified in a RoboCycler Temperature Cycler using the following cycling regimen: 94°C 2 min (one cycle); 94°C 1 min, 52°C 1 min, 72°C 4 min (20 cycles). The aqueous phase was extracted (14), and CSR selection products were gel purified and re-amplified using *PfuTurbo* DNA polymerase and PfuF/Sso7d_mR primers. Re-amplification products were gel purified and swapped back into the pET11a-*Pfu* (V93R)-S_m vector using the EZClone method described earlier.

Mutant screening and characterization

Clones were screened for dCppp-Dabcyl uptake by PCR. Colonies were grown up in 96-(deep) well plates. Cells were collected, lysed with B-PER Protein Extraction reagent (Pierce), and heated at 85°C for 15 mins to inactivate *Escherichia coli* protein. After centrifugation, supernatants (3 μl) were added to amplifications (25 μl) of a 0.5 Kb fragment, carried out with 200 μM each dATP/dGTP/TTP, 40 μM dCppp-Dabcyl, primers, and plasmid DNA (wild-type *Pfu*-Sso7d_m fails to amplify in the presence of 100% dCppp-Dabcyl). Supernatants were also analyzed for *Pfu* PolB synthesis by SDS-PAGE (full-length *Pfu*, 90 kDa) and western blotting. Positive clones were subject to DNA sequencing.

Mutant expression and purification

Directed mutations were incorporated using the QuikChange II XL site-directed mutagenesis kit. Primers were designed to introduce Q484R and split mutations into exo^+ or exo^- (D141A/E143A) *Pfu* constructs (pET21b-*Pfu*). *Pfu* mutants were cultured, and induced cells were collected, resuspended (40 mM Tris–HCl pH 7.5, 1 mM EDTA, protease inhibitors, 0.25 mg/ml lysozyme and 10 mM β-ME), and lysed by sonication. Extracts were heated at 85°C for 15 min and centrifuged to remove denatured *Escherichia coli* protein, and contaminating nucleic acids were removed by polyethyleneimine precipitation (1 M NaCl, 0.175% PEI). Clarified supernatant was brought to 65% saturation with ammonium sulfate, and precipitated protein was pelleted, washed, and dissolved in buffer A (40 mM Tris-HCl pH 7.5, 1 mM EDTA, 10 mM β-ME). After dialysis overnight, each protein sample was loaded on a Q-Sepharose Fast Flow column that had been equilibrated with buffer A. Flow-thru fractions were collected and loaded directly on a SP-Sepharose

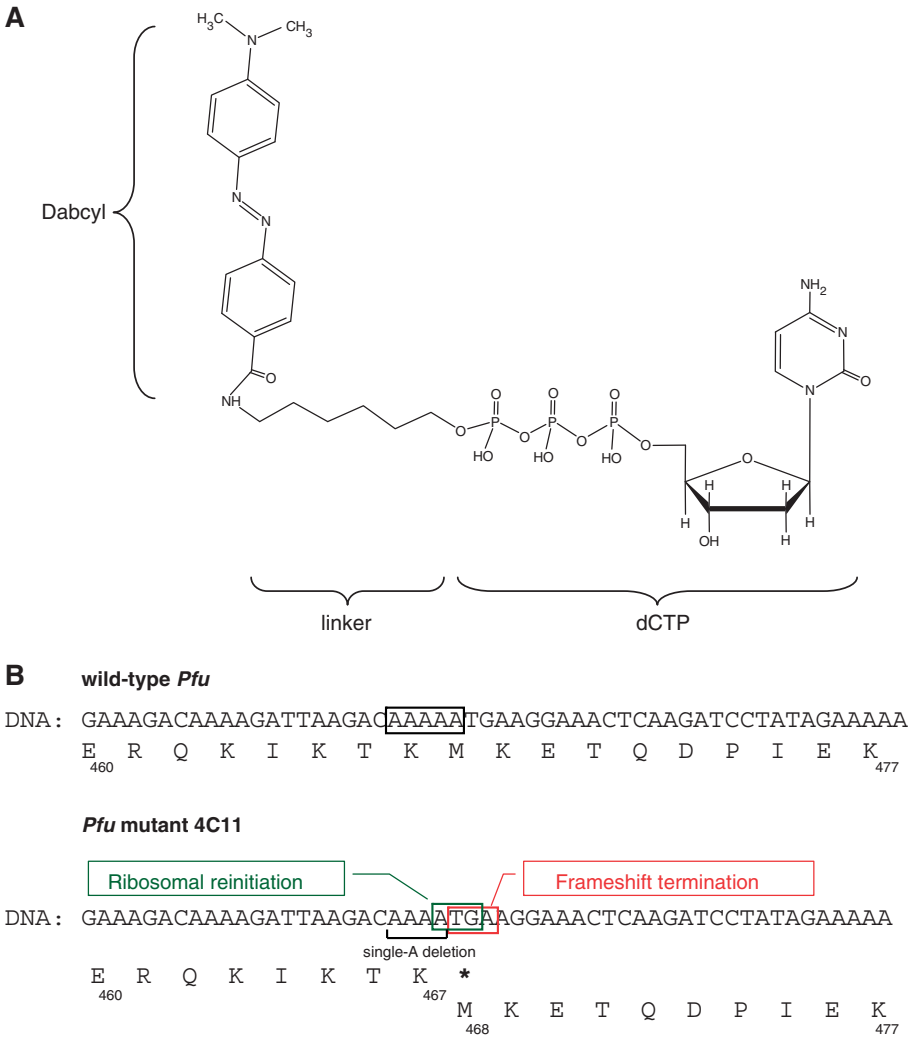

Figure 1. Key structures. (**A**)The structure of dCppp-Dabcyl used for CSR enrichment. (**B**) The DNA (1378–1431 nt) and amino acid (460–477) sequences of the split region in *Pfu* 4C11. A single dA deletion produces a frameshift termination (TGA) immediately downstream of a ribosome initiation site (ATG) that halts *Pfu* synthesis at K467.

HP column, pre-equilibrated with buffer A. After washing, *Pfu* proteins were eluted with a 30 column-volume gradient to 500 mM KCl. Peak fractions were pooled, dialyzed against buffer A, and then loaded on a Heparin Sepharose HP column equilibrated in buffer A. After washing with buffer A, *Pfu* mutants were eluted with a 20 column-volume gradient to 750 mM KCl. Peak fractions were pooled, concentrated and dialyzed into final dialysis buffer (50 mM Tris–HCl, 0.1 mM EDTA, 100 mM KC1, 50% glycerol, pH 8.2), and stored at −20°C.

Refolding

Refolding experiments were carried out using fragments derived from an exo$^-$ (D141A/E143A) derivative of *Pfu* 4C11. N-fragment (*Pfu* 1–467 with V93R/D141A/E143A/A318T mutations) and C-fragment (*Pfu* 468–775-Sso7d$_m$ with Q484R/V604L/A662V mutations) were cloned separately into pET 21b using conventional methods. Clones were transformed into BL21-DE3-CodonPlus RIL cells and grown up separately in 21 cultures as described above. For refolding experiments, inclusion bodies were isolated and washed in turn with buffer containing 0.15% sodium deoxycholate, 1% Triton X-100, and then with buffer alone (40 mM Tris–HCl pH 7.5 and 10 mM β-ME). N- and C-fragment pellets were solubilized with 4 and 5 M guanidine HCl, respectively, and relative recoveries determined by SDS–PAGE analysis. Refolding was carried out by combining equivalent amounts of the two fragments, and dialyzing the mixture against 2 × 200 volumes of buffer (40 mM Tris–HCl pH 7.5, 1 mM EDTA, 100 mM KCl, 10% glycerol). Insoluble protein was removed by centrifugation, and refolded samples were analyzed by Bradford assay, SDS–PAGE and PCR. Refolded protein was further purified by column chromatography as described above.

Endpoint/real-time PCR assays

Amplifications were carried out with *Pfu* mutants (amounts indicated in figure legends) in *cPfu* (non-fusion) or *PfuUltra* II (Sso7d$_m$-fusion) PCR reaction buffer.

Nucleotide mixes consisted of 200 μM each dATP, dGTP, TTP and either 200 μM dCTP (for polymerase activity assays) or 50 μM [dCTP+ dCppp-Dabcyl] (for analog incorporation assays). A 900 bp α-1-antitrypsin gene fragment was amplified (50 μl reactions) from 100 ng human genomic DNA using 100 ng each primer (forward: 5′-GAGGAGAGCAGGAAAGGTGGAAC; reverse: 5′-GAGGTACAGGGTTGAGGCTAGTG). Products were analyzed on 1% agarose/1× TBE gels stained with ethidium bromide and digitally imaged. The 69 bp β-actin fragment was amplified in 25 μl reactions containing cPfu PCR buffer, 5 ng human cDNA, 200 nM each primer (forward: 5′-CTGGCACCCAGCACAATG; reverse: 5′-GCCGATCCACACGGAGTACT), 3% DMSO and 0.24× SYBR Green dye. Amplification was monitored in real-time on the Mx3005P QPCR System.

Polymerase kinetic assays

DNA polymerase activity was measured at 72°C using activated calf thymus DNA as described (17). K_m and K_{cat} were measured at 72°C using primed M13 DNA substrate, 200 μM each dATP, dGTP, TTP (with ^{3}H-TTP as tracer), and varying concentrations of dCTP (range: 1–200 μM) or dCppp-Dabcyl (range: 25–600 μM).

RESULTS

Split *Pfu* PolB isolated by CSR selection

To engineer DNA polymerases to incorporate IQNs, initial efforts were focused on increasing tolerance for modifications at the γ-phosphate of an incoming nucleotide. We employed dCppp-Dabcyl (Figure 1A), a precursor to IQNs that employ a fluorophore on the base and dabcyl quencher on the γ-phosphate (1,2). DNA polymerase mutants were enriched from randomized *Pfu* libraries using the CSR PCR-emulsion technique as described (14) except that *cPfu* buffer, *Pfu* expresser cells, and *Pfu*-specific primers were substituted for *Taq* components, and dCTP was replaced with dCppp-Dabcyl:dCTP mixtures consisting of 90%:10% (round 1), 95%:5% (round 2) or 100%:0% (rounds 3 and 4). To improve amplification efficiency, mutant libraries were prepared from a derivative of *Pfu* (exo$^+$) PolB that contains a V93R mutation to eliminate sensitivity to deoxyuracil in template DNA (15) and a C-terminal processivity tag ('Materials and Methods' section). Earlier attempts to select from *Pfu* (exo$^-$) libraries (derived from pET11a-*Pfu* V93R/D141/E143A-S_m) were unsuccessful; however, higher-than-expected mutant frequencies indicated that the error rate of proofreading-deficient *Pfu* mutants [5- or 40-fold greater than *Taq* or *Pfu*, respectively (18)] may be too high to provide efficient self-replication. Although not formally tested, we suspect that proofreading and/or higher processivity (incorporation of $-S_m$ tag) enhance selectivity of CSR and obviate the need for specialized conditions for *Pfu* (19).

After four rounds of selection, *Pfu* mutants were screened for dCppp-Dabcyl uptake by PCR. Of the several hundred clones tested, only one (4C11) successfully amplified a 0.5 Kb plasmid target in the presence of 100% dCppp-Dabcyl. However, when analyzed by SDS–PAGE, clone 4C11 failed to show detectable synthesis of full-length protein (*Pfu*, 90.1 kDa). Western blotting with polyclonal anti-*Pfu* antibody revealed synthesis of truncated fragments, including a predominant species with an apparent molecular weight around 50 kDa. DNA sequence analysis confirmed that a termination codon had been introduced at amino acid 468 (*Pfu* fragment MW_{calc} = 54.5 kDa), along with four other amino acid changes (A318T, Q484R, V604L, A662V). Upon further inspection, we recognized that a single-base deletion in a run of dAs between 1398 and 1402 (AAAA**ATG**A → AAA**ATG**A) produced a frame-shift termination (TGA) immediately downstream of a ribosome initiation site (ATG) (Figure 1B). Together, these results suggested that clone 4C11 expresses a bicistronic message that directs synthesis of two *Pfu* fragments which assemble to create a PCR-stable polymerase.

Pfu 4C11 was purified using the same procedures employed with wild-type *Pfu*. During sequential chromatography on Q, SP and Heparin Sepharose columns, *Pfu* 4C11 elutes as a single peak with similar migration to wild-type *Pfu*. By SDS-PAGE analysis, the purified protein consisted of two fragments of ~54 and 36 kDa (Figure 2A). In addition to showing identical chromatographic behavior, *Pfu* 4C11 exhibits comparable specific activity to wild-type *Pfu* in incorporation assays carried out at 72°C with activated calf-thymus DNA (data not shown). Apparent size of the C-fragment is consistent with translation initiating at the methionine corresponding to M468 in full-length *Pfu*.

Previous studies in *E. coli* have shown that after a premature termination event, the same ribosome can reinitiate without dissociation in the absence of ribosome recycling factor or a Shine–Dalgarno consensus sequence when a methionine (AUG) codon is located in close proximity (20). As discussed below, we expect that translation and proper folding of *Pfu* C-fragment is dependant on synthesis of the proximal N-fragment. This process, referred to as translational coupling (21), is particularly efficient when initiation and termination codons overlap (e.g. AUGA in 4C11) and is a common feature of bacterial operons for regulating gene expression and stoichiometry of gene products (22,23). In 4C11, termination of N-fragment (*Pfu* 1–467) and initiation of C-fragment (*Pfu* 468–775) synthesis appears to occur seamlessly, with no change to amino acid sequence (Figure 1B). As modeled in Figure 2B, the translational stop–start site introduces a split in the fingers domain of *Pfu* 4C11. One of the four amino acid replacements in 4C11 (Q484R) lies in close proximity to the split, on an opposing α-helix near the tip of the finger (Figure 2B).

DNA polymerase activity reconstituted from separate fragments

To rule out the possibility that polymerase activity is simply the result of translational read-through at stop codon 468 (low-level contamination with full-length *Pfu*), we cloned *Pfu* N- and C-fragments into separate expression vectors. When expressed independently, N-fragment is synthesized

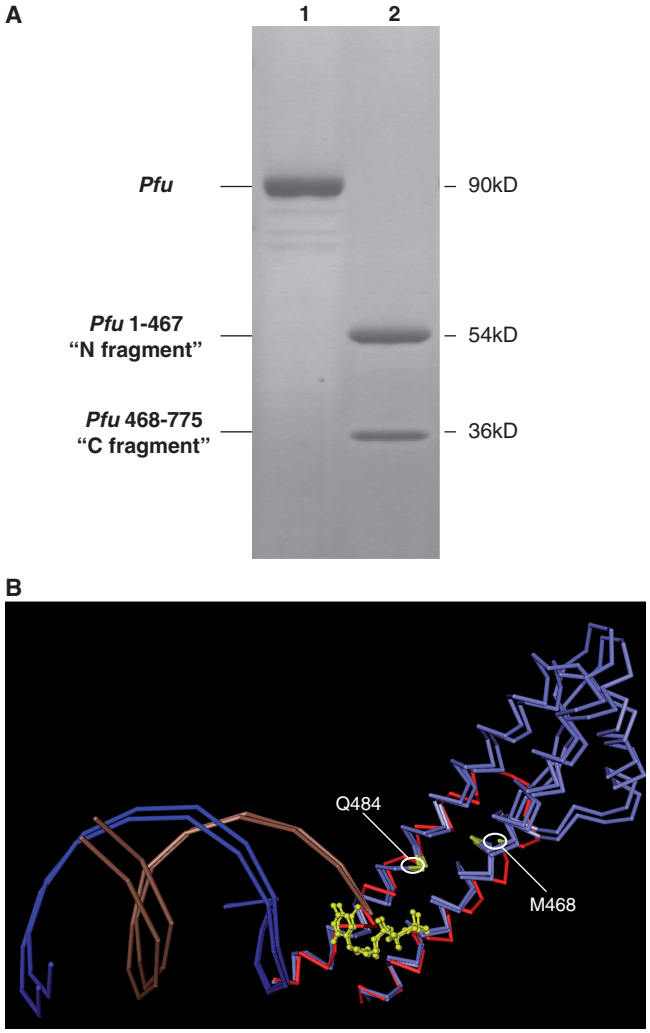

Figure 2. Split *Pfu* 4C11. (**A**) SDS–PAGE analysis of *Pfu* (lane 1) and *Pfu* 4C11 (lane 2; subcloned to remove processivity tag). Calculated pI and molecular weight values are 5.9 and 54.5 kDa, respectively, for N-fragment (*Pfu* 1–467 V93R/A318T) and 9.4 and 35.8 kDa, respectively, for C-fragment (*Pfu* 468–775 Q484R/V604L/A662V). In (**B**), the location of the split (M468, *Pfu*; M489, RB69) and Q484 (Q556, RB69) mutation are indicated in yellow on the C_α backbone structure of the fingers domain. Amino acids 449–499 of *Pfu* (2JGU_A) are shown in red (38). VAST aligned (39) RB69 segments are shown in light blue as follows: amino acids 473–574 from RB69 1Q9Y_A (40) and amino acids 470–571 from RB69 3CQ8_A (41). DNA strands are shown in brown and dark blue, and nucleotides are shown in yellow.

in high yield (~50% soluble), while C-fragment is much less abundant and completely insoluble (Figure 3A). To reconstitute split DNA polymerase activity, guanidine-solubilized protein fractions (inclusion body preps) were quantified, mixed in equimolar ratios (1:1 N:C), and refolded from denaturant by dialysis against buffer. Bradford and SDS–PAGE analyses revealed that N-fragment assumes a soluble conformation when refolded on its own, but C-fragment requires the presence of N-fragment to fold into soluble protein.

Polymerase activity of refolded samples was assayed by PCR. As shown in Figure 3B, PCR activity of split *Pfu* 4C11 can be reconstituted from separately expressed and purified N- and C-fragments ('refolding reaction N + C'). As expected, equivalent amounts of refolded N- or C-fragment (refolded in the absence of the complementary fragment) failed to amplify the 0.9 Kb genomic DNA target. To assess integrity, the refolded 'N + C' preparation was further purified using standard methods for wild-type *Pfu*, which include Q- (in flow-through fractions) and SP- (binds and elutes with KCl gradient) Sepharose chromatography. As shown in Figure 3C, the refolded 'N + C' sample exhibits identical chromatographic behavior to wild-type *Pfu* and to split *Pfu* synthesized from a bicistronic expression vector. Whether N- and C-fragments are expressed together through apparent coupled translation (Figure 2A; from 4C11 clone) or reconstituted by refolding from separate preparations (Figure 3C; fractions 41 and 44), the two fragments co-purify in roughly equivalent amounts, suggesting that they form a stable 1:1 complex.

Split plus Q484R increases tolerance to γ-phosphate substituent

We next shifted our attention to characterize the relative contributions of the split and each of the four mutations to improved dCppp-Dabcyl uptake. We introduced mutations (A318T, Q484R, V604L, A662V or split) separately and in pair-wise combinations into exo^+ or exo^- (D141A/E143A) *Pfu* constructs. Although 4C11 originated from *Pfu* (exo^+) libraries, proofreading-deficient mutants are ultimately required for IQN studies, and therefore were used here to evaluate polymerase activity and dCppp-Dabcyl incorporation. Of the mutants tested, only *Pfu* SQ (*Pfu* split + Q484R) successfully amplified a 69 bp β-actin target in the presence of 100% dCppp-Dabcyl (Figure 4). Incorporation of dCppp-Dabcyl was confirmed by failure of *Pfu* SQ to amplify in the absence of dCTP (dATP/TTP/dGTP-only reactions), and subsequently by synthesis of fluorescent amplicon with internally-quenched dCTP (FAM-dCppp-Dabcyl; Arezi,B. and Hogrefe,H.H., manuscript in preparation). When examined using real-time *PCR*, C_t values for *Pfu* S (*Pfu* split) and *Pfu* Q (*Pfu* Q484R) were slightly ahead of those for *Pfu* in 2.5/47.5 reactions (mirroring variations in band intensity in Figure 4A). These results suggest that the split or Q484R mutation (alone) can broaden specificity, but the combination of mutations is significantly more effective in reducing discrimination against the dabcyl substituent (Figure 4B). Other mutations in 4C11 (V93R, A318T, V604L, A662V, $Sso7d_m$ fusion) had little-to-no impact on dCppp-Dabcyl uptake or stability of split *Pfu* PolB (data not shown).

Broader specificity occurs with no detectable impact on polymerization of natural nucleotides. For example, equivalent amounts of *Pfu*, *Pfu* S, *Pfu* Q and *Pfu* SQ produce identical C_t values in real-time PCR assays (50/0 reactions; Figure 4B), while specific activities of *Pfu* and *Pfu* SQ vary by <6% (Table 1). Moreover, in steady-state kinetic studies employing dCTP or dCppp-Dabcyl (and fixed excess concentrations of dATP, TTP and dGTP), *Pfu* and *Pfu* SQ exhibit comparable K_m and K_{cat} values for dCTP. In similar studies with analog, K_m [dCppp-Dabcyl] for *Pfu* SQ was ~10-fold higher than K_m

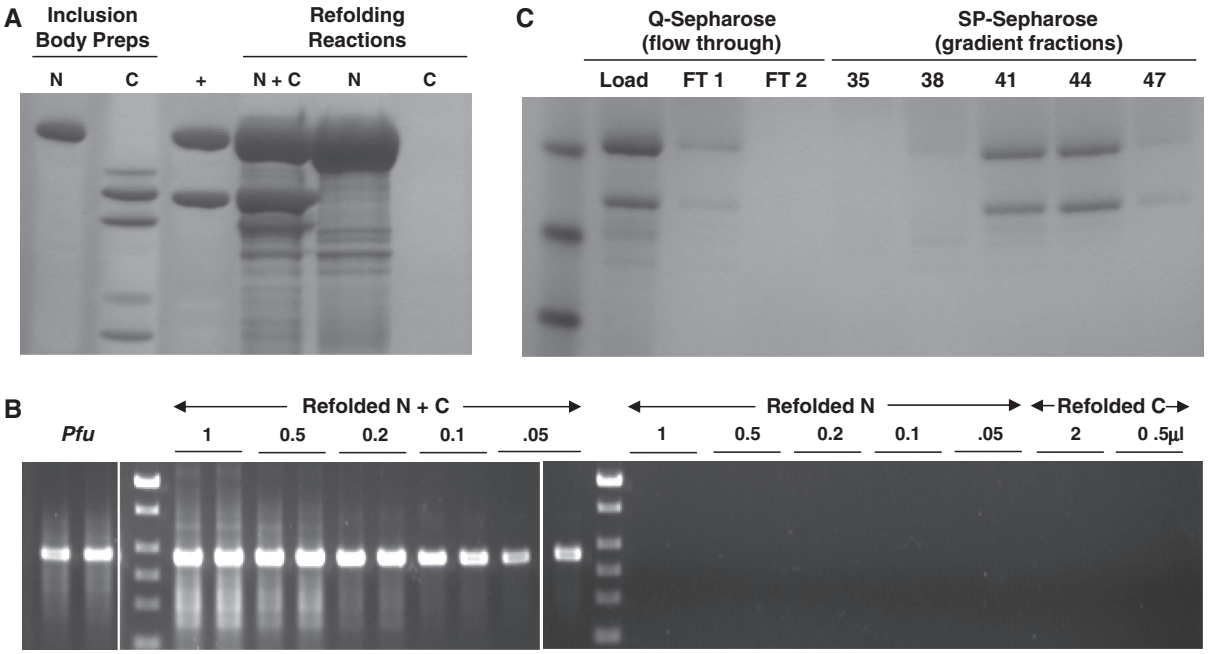

Figure 3. Reconstitution of PCR activity from individual fragments. (**A**) SDS–PAGE analysis of *Pfu* N- and C-fragments that were expressed from separate constructs, purified from inclusion bodies ('inclusion body preps'), and subject to refolding by dialysis from denaturant ('refolding reactions'). Refolding reactions were carried out with guanidine-solubilized preparations of N-fragment ('N'), C-fragment ('C'), or an equimolar mixture of N- and C-fragments ('N + C'). Recoveries of soluble folded protein (in similar volumes) were: 1.1 mg/ml 'N'; ≤0.1 mg/ml 'C'; and 1 mg/ml 'N + C'. Split *Pfu* expressed and purified from clone 4C11 (synthesized from bicistronic mRNA) was run in lane '+'. SDS–PAGE samples were prepared from guanidine-containing samples ('inclusion body preps') by TCA precipitation. In (**B**), a 0.9 Kb fragment was amplified in duplicate with the indicated amounts of each refolded protein sample. PCRs (50 µl) were cycled (one cycle of 95°C 2 min; 30 cycles of 95°C 30 s, 58°C 30 s and 72°C 15 s) and then analyzed on a 1% agarose gel. In (**C**), the refolded N + C sample ('load') was further purified by Q- (FT, in flow-through fractions) and SP- (KCl gradient fractions 35, 38, 41, 44, 47) Sepharose chromatography. Under these conditions, N-fragment in the assembled complex passes through the Q column (FT1), while excess (unassociated) N-fragment in the refolded 'N + C' sample binds to the anion exchange column. Column fractions were analyzed by SDS–PAGE.

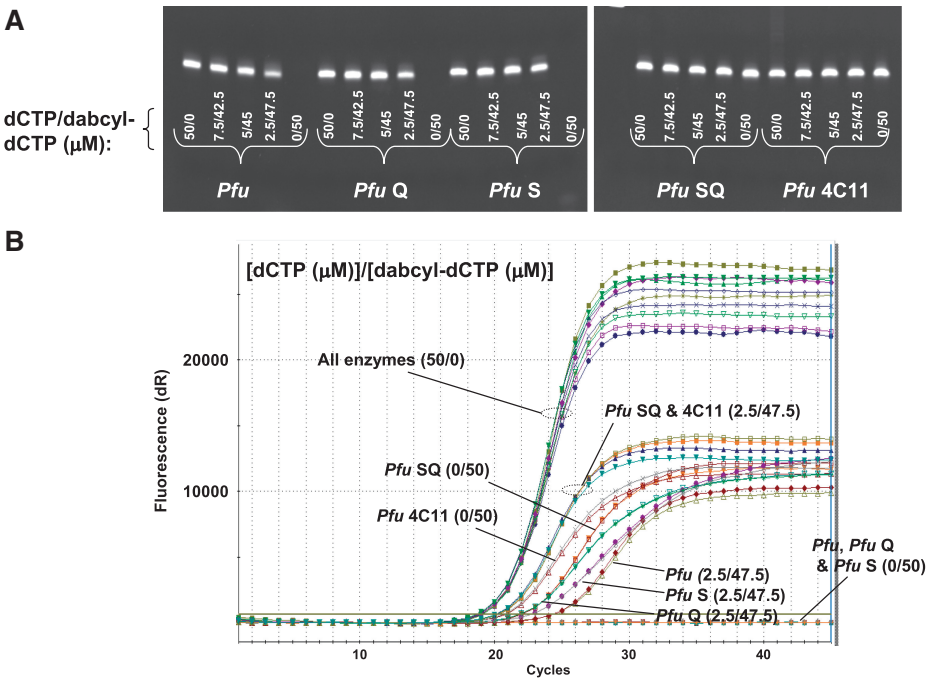

Figure 4. Split plus Q484R mutation required for dCppp-Dabcyl incorporation. A 69 bp β-actin target was amplified as described ('Materials and Methods' section) using the following His-tag purified exo⁻ *Pfu* (non-fusion) enzymes: 1.25 U *Pfu* or 25 ng *Pfu* mutant (S, split; Q, Q484R; SQ, split + Q484R; 4C11, V93R/A318T/split/Q484R/V604L/A662V). PCRs contained 50 µM total deoxycytosine nucleotide (dCTP + dCppp-Dabcyl), and percent dCppp-Dabcyl was varied from 0% to 100%. Reactions were monitored in real-time using SYBR Green (**B**; 0, 95%, and 100% dCppp-Dabcyl reactions), and products generated at cycle 45 were visualized on gels (**A**).

[dCTP] (240 versus 25 nM, respectively), while K_{cat} values were comparable (8.3 s^{-1} for dCppp-Dabcyl versus 6.7 s^{-1} for dCTP). These results indicate that the dabcyl substituent reduces nucleotide-binding affinity, but does not interfere with catalytic activity of *Pfu* SQ. Finally, fidelity studies indicate that neither the split nor Q484R mutation compromise fidelity of incorporating natural nucleotides [*lac* I assay (18)]. In fact, preliminary results suggest that the fidelity of *Pfu* (exo$^+$) SQ is somewhat higher than that of wild-type *Pfu*, which perhaps is not too surprising given the stringent requirement for accurate self-replication in CSR (19), that in our case calls for replicating 2.5 Kb coding regions in the presence of 100% dCppp-Dabcyl. Further studies will address the fidelity of dCppp-Dabcyl incorporation, as well as the molecular basis of any accuracy improvements by the split or Q484R mutation (Hansen,C.J. and Hogrefe,H.H., manuscript in preparation).

Alignment of *Pfu* SQ with native split DNA polymerases

The general architecture of the DNA polymerase core consists of three domains, termed the fingers, palm and thumb (by analogy to a right hand), and proofreading DNA polymerases (e.g. PolB from archaea and phage) additionally contain an N-terminal 3′–5′ exonuclease domain. When substrate is bound, the fingers and thumb domains make numerous contacts with incoming nucleotide and duplex DNA, respectively, and move towards the palm domain (closed conformation) where critical residues for substrate discrimination and catalysis reside. In Figure 5, we compare the location of our engineered split, which lies in the fingers domain of *Pfu* between 467 and 468, to the locations of naturally-occurring splits identified in 7 PolB DNA polymerases (24–26). Like our engineered split, splits identified in five T4-like phage DNA polymerases (gp43) occur within the fingers domain (24,25). In contrast, the split in *Methanobacterium thermoautotrophicum* (*Mth;* thermophilic archaea) PolB occurs in the palm domain, approximately 83 amino acids carboxyl to the *Pfu* split (26), and the *polB* gene of *Rhodothermus marinus* phage RM378 is predicted to encode a split between the N and palm domains (25). While *Pfu* 4C11 (SQ) is synthesized

Table 1. Polymerase activity measurements[a]

Polymerase	Specific activity	Steady-state kinetic parameters					
	dCTP	dCTP			dCppp-Dabcyl		
	(U/mg)	K_m (μM)	K_{cat} (s^{-1})	K_{cat}/K_m (s^{-1} μM^{-1})	K_m (μM)	K_{cat} (s^{-1})	K_{cat}/Km (s^{-1} μM^{-1})
Pfu	166 000	23[b]	9.2	0.4	nd	nd	nd
Pfu SQ[c]	176 000	25	6.7	0.3	240	8.3	0.03

[a]Parameters determined using exo$^-$ *Pfu* or *Pfu* SQ, and activated calf thymus (specific activity) or primed M13 (kinetic parameters) DNA; kinetic parameters were determined in the presence of fixed (excess) concentrations of dATP, dTTP and dGTP.
[b]K_m [dNTP] value reported for exo$^-$ *Pfu* is 12 μM each dNTP in titrations with equimolar dNTPs (17).
[c]Results are expressed as the average of duplicate measurements (ranges are <20%).
nd, no detectable incorporation.

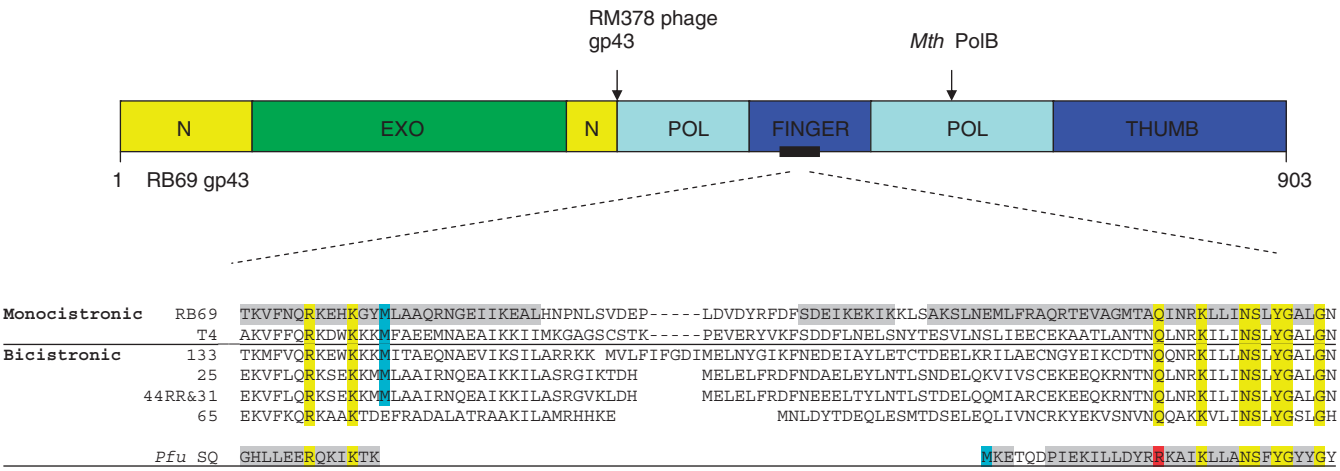

Figure 5. Comparison of engineered versus native splits. Locations of naturally-occurring splits in PolB/gp43 from *Mth* (26), RM378 phage and T4-like phages are indicated on the RB69 gp43 map [adapted from Petrov *et al.* (25)]. Finger tip structures of T4 phage gp43 have been classified according to number of ORFs as: (i) monocistronic, an uninterrupted fingertip exemplified by T4 and RB69 gp43; or (ii) bicistronic, a split fingertip characteristic of the T4-like phage 133/44R/31/25/65 sub-family (24). Conserved residues (yellow) were used to align the corresponding amino acid sequence of *Pfu* SQ. Other features are denoted as follows: turquoise, conserved methionine (M) corresponding to translation reinitiation site in *Pfu* SQ (*Pfu* M468); red, Q484R mutation in *Pfu* SQ; grey, α-helices comprising the fingers domain of *Pfu* [αO extends to E470 in uninterrupted finger; (38)] and RB69 (42).

from one bi-cistronric gene, naturally-occurring split polymerases are encoded by two genes that are separated by anywhere from 2 bp to 3 Kb (T4-like phage) to 85 Kbp (*Mth*) in the genome.

DISCUSSION

In this report, we describe the serendipitous creation of a split in the fingers domain of *Pfu* PolB that facilitates incorporation of a nucleotide γ-phosphate analog. The fingers domain of *Pfu* (aa 451–500) encompasses two anti-parallel helices, αO and αP (notation as per 27), which make numerous contacts with an incoming nucleotide (27–29). As discussed above, dCppp-Dabcyl incorporation requires the combined effects of a split in αO between K467 and M468 and a Q→R substitution at position 484, which resides in αP opposing the split (Figure 2B). In RB69 gp43, amino acid side chains of K560, R482 and K486 (equivalent to *Pfu* K488, R461 and K465, respectively) are involved in H-bonding interactions with oxygen atoms of the γ-phosphate, while K560 and N564 (equivalent to *Pfu* K488 and N492) make additional contacts with α- and β-phosphate oxygens, respectively (30,31). In *Pfu*, the engineered split (467/468) is located just C-terminal to K465, and by analogy to RB69 gp43, should reside sufficiently close to the γ-phosphate of a bound nucleotide to accommodate a bulky O-linker-dabcyl substituent (Figures 1A and 2B).

Q484 resides on the same face of αP as K488, which by analogy to RB69 gp43, forms H-bonds with both α- and γ-phosphate oxygens. Although highly conserved (27), Q484 can be substituted with R with no deleterious effects on *Pfu* DNA polymerase activity (Figure 4A). In earlier studies, mutations at Q484 (or the analogous position) produced varying results, ranging from loss of polymerase activity [<1%; *Pfu* Q484A (32)] to comparable (Vent Q486N/L) or higher (2.5-fold; Vent Q486E) specific activity compared to wild-type (33). In *Pfu* SQ, an arginine at 484 may facilitate dCppp-Dabcyl incorporation by forming alternative H-bonds (e.g. alternative to K488 or N492) with triphosphate oxygens. Changes to the H-bonding network may be required to reposition the γ-phosphate closer to the finger tip to allow the *O*-linker-dabcyl substituent to pass through the split. The results of kinetic studies indicate that the quencher moiety reduces nucleotide-binding affinity by ~10-fold. Once bound, however, the bulky substituent appears to be oriented away from the active site, where it has no impact on phosphodiester bond formation, primer-template translocation, or dissociation of pp-dabcyl (similar K_{cat} values for dCTP and dCppp-Dabcyl; Table 1).

The fingers domain is a common target for mutagenesis studies aimed at reducing fidelity or increasing the incorporation of nucleotide analogs. For example, in PolB-type DNA polymerases, amino acid side chain replacements in the finger tip (*Pfu* T471–Q472–D473) reduce nucleotide insertion fidelity (34). In other studies, amino acid changes on one face of αP (corresponding to Q484–K488–N492 of *Pfu*) reduce affinity for natural nucleotides while improving dNDP incorporation (30,31), and substitutions on the opposite side of αP (*Pfu* A486, L490, A491, Y497) improve incorporation of dideoxynucleotides (32,35). A common dogma of mutagenesis studies like these is that termination codons are deleterious, an undesired consequence of a frameshift or point mutation that leads to premature truncation and synthesis of non-functional protein. With the fortuitous introduction of a split that facilitates nucleotide analog uptake, we now recognize that translational stop-start sites may be used like amino acid replacements, to modify or broaden substrate utilization through rational design (e.g. increasing tolerance for bulky substituents by increasing conformational flexibility or reducing steric hindrance).

The apparent seamless nature of the 467–468 split is remarkable for a hyperthermophilic enzyme. As discussed above, *Pfu* SQ is comparable to wild-type with respect to chromatographic behavior, steady-state kinetic parameters (for dCTP), and PCR performance, which implies that the contact interfaces between N- and C-fragments are stabilized by the same (or nearly the same) hydrophobic, electrostatic and H-bonding interactions employed in wild-type *Pfu*. The only difference observed to date is that unlike *Pfu*, *Pfu* SQ is sensitive to reducing agents and loses polymerase activity in the presence of ≥1 mM DTT or 2-mercaptoethanol (β-ME) (data not shown). In the crystal structures of archaeal PolB DNA polymerases from *Thermococcus gorgonarius*, *Desulfurococcus* strain Tok, *Thermococcus* species 9°N, and *T. kodakarensis* (KOD), there are two cysteine pairs (428C–442C; 506C–509C) that form, or are positioned to form, disulfide bonds (27–29,36). These disulfide bonds are thought to play a role in stabilizing hyperthermophilic DNA polymerases at high temperature (27–29). Increased sensitivity suggests that one or both disulfide bonds are more exposed (less buried) to reducing agent in the split polymerase structure. Alternatively, disulfide bonds may play a greater role in *Pfu* SQ in maintaining the structure of fingers and palm subdomains in the absence of other stabilizing contacts.

Correct assembly of *Pfu* SQ is dependant on folding the otherwise-insoluble C-fragment in the presence of N-fragment. As shown here, this can be accomplished by translation of *Pfu* N- and C-fragments from a bicistronic expression vector or by refolding an equimolar mixture of N- and C-fragments purified from separate expression constructs (Figures 2 and 3). Like *Pfu* SQ, *Mth* PolB requires co-expression of PolB1 (marginally soluble N-fragment) and PolB2 (insoluble C-fragment) in *E. coli* to produce a soluble 1:1 dimer with both polymerase and 3′–5′ exonuclease activities [Figure 5; (26)]. Fully-assembled PolB dimer presumably exists within the *Mth* cell, although the distance between ORFs (85 Kb) precludes coupled translation of subunits. In *Pfu* SQ, deletion of a single A occurs within a short poly(A) track (1398–1402 nt), presumably the result of primer slippage during library construction by error-prone PCR (37), creating an ideal sequence [AUGA; (21)] for coupled translation of N- and C-fragments with no change to amino acid sequence (Figure 1B).

Bicistronic DNA polymerases have also been identified in phylogenetic variants of T4 phage (Figure 5). Genome sequences reveal insertions in gene 43 that split PolB within the finger tip, which is highly divergent among T4 gp43 relatives and significantly longer than the finger tip in *Pfu* (24). Petrov *et al*. classified the various finger tip structures of gp43 according to number of ORFs (monocistronic, uninterrupted fingertip; bicistronic, split fingertip) and length of the intervening sequence between the two helices of the fingers domain (group I, large as exemplified by T4 and RB69; group II, short like archaeal DNA PolB). Compared to bicistronic groups I and II DNA polymerases which are split at various locations within a 21–38 residue intervening sequence, *Pfu* SQ is split within αO on the N-terminal side of an abbreviated finger tip (Figure 2B) that is thought to be an evolutionary adaptation for increased thermostability (27–29,36). Remarkably, all configurations of a split finger tip accommodate nucleotide binding and incorporation. Like *Mth* PolB, subunits of T4 gp43 are synthesized independently from ORFs separated by anywhere from 2bp (*Aeromonas* phage) to 3 Kb (*Acinetobacter* phage) (24). The diversity of intercistronic sequences among split forms of phage gp43 suggests that the fingers domain is a hotspot for mutations, including invasion by DNA insertion elements, and prompted Petrov *et al*. (24,25) to conclude that the modular nature of the T4 genome and its replication proteins provides resilience to mutations and facilitates adaptation of T4-like phage to different bacterial hosts in nature.

In this study, we evolved a mutant of *Pfu* PolB that appears to mimic an evolutionary variant of T4 phage. When split near the finger tip, *Pfu* retains full DNA polymerase activity at PCR temperatures, further illustrating the modular nature of the Family B DNA polymerase structure (24,25). To date, *Pfu* SQ is the only split DNA polymerase shown to incorporate nucleotide γ-phosphate analogs. However, one can speculate that naturally-occurring split DNA polymerases (e.g. evolutionary intermediates) may also show increased tolerance for certain analogs or altered nucleotide fidelity.

ACKNOWLEDGEMENTS

We thank Vanessa Gurtu for initial optimization of CSR conditions, and Jeff Braman (Agilent Technologies Inc.) and the R&D team at Trilink Biotechnologies for design and synthesis of γ-phosphate modified nucleotides.

FUNDING

Funding for open access charge: Agilent Technologies.

Conflict of interest statement. None declared.

REFERENCES

1. Anderson,J.D. and Braman,J.C. (2003) Dual-labeled nucleotides, WO/2003/087302.
2. Lawler,J.L. (2006) Reagents for monitoring nucleic acid amplification and methods of using same, US Patent 7118871.
3. Williams,J.G.K. (2001) System and methods for nucleic acid sequencing of single molecules by polymerase synthesis, US Patent 6255083.
4. Sood,A., Kumar,S., Nampalli,S., Nelson,J.R., Macklin,J. and Fuller,C.W. (2005) Terminal phosphate-labeled nucleotides with improved substrate properties for homogeneous nucleic acid assays. *J. Am. Chem. Soc*., **127**, 2394–2395.
5. Vassiliou,W., Epp,J.B., Wang,B.-B., Del Vecchio,A.M., Widlanski,T. and Kao,C.C. (2000) Exploiting polymerase promiscuity: a simple colorimetric RNA polymerase assay. *Virology*, **274**, 429–437.
6. Hardin,S., Gao,X., Briggs,J., Willson,R. and Tu,S.-C. (2008) Methods for real-time single molecule sequence determination, US Patent 7329492.
7. Eid,J., Fehr,A., Gray,J., Luong,K., Lyle,J., Otto,G., Peluso,P., Rank,D., Baybayan,P., Bettman,B. *et al*. (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133–138.
8. Metzker,M.L. (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet*., **11**, 31–46.
9. Kumar,S., Sood,A., Wegener,J., Finn,P.J., Nampalli,S., Nelson,J.R., Sekher,A., Mitsis,P., Macklin,J. and Fuller,C.W. (2005) Terminal phosphate labeled nucleotides: synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides, Nucleotides Nucleic Acids*, **24**, 401–408.
10. Mulder,B.A., Anaya,S., Yu,P., Lee,K.W., Nguyen,A., Murphy,J., Willson,R., Briggs,J.M., Gao,X. and Hardin,S.H. (2005) Nucleotide modification at the {gamma}-phosphate leads to the improved fidelity of HIV-1 reverse transcriptase. *Nucleic Acids Res*., **33**, 4865–4873.
11. Alexandrova,L.A., Skoblov,A.Y., Jasko,M.V., Victorova,L.S. and Krayevsky,A.A. (1998) 2′-Deoxynucleoside 5′-triphosphates modified at alpha-, beta- and gamma- phosphates as substrates for DNA polymerases. *Nucleic Acids Res*., **26**, 778–786.
12. Arzumanov,A.A., Semizarov,D.G., Victorova,L.S., Dyatkina,N.B. and Krayevsky,A.A. (1996) γ-Phosphate-substituted 2′-deoxynucleoside 5′-triphosphates as substrates for DNA polymerases. *J. Biol. Chem*., **271**, 24389–24394.
13. Korlach,J., Bibillo,A., Wegener,J., Peluso,P., Pham,T.T., Park,I., Clark,S., Otto,G.A. and Turner,S.W. (2008) Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids*, **27**, 1072–1083.
14. Ghadessy,F.J., Ong,J.L. and Holliger,P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci*., **98**, 4552–4557.
15. Fogg,M.J., Pearl,L.H. and Connolly,B.A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struct. Biol*., **9**, 922–927.
16. Wang,Y., Prosen,D.E., Mei,L., Sullivan,J.C., Finney,M. and Vander Horn,P.B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res*., **32**, 1197–1207.
17. Hogrefe,H.H., Cline,J., Lovejoy,A. and Nielson,K.B. (2001) DNA polymerases from hyperthermophiles. *Methods Enzymol*., **343**, 91–116.
18. Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res*., **24**, 3546–3551.
19. Ramsay,N., Jemth,A.-S., Brown,A., Crampton,N., Dear,P. and Holliger,P. (2010) CyDNA: synthesis and replication of highly Cy-dye substituted DNA by an evolved polymerase. *J. Am. Chem. Soc*., **132**, 5096–5104.
20. Karamyshev,A.L., Karamysheva,Z.N., Yamami,T., Ito,K. and Nakamura,Y. (2004) Transient idling of posttermination ribosomes ready to reinitiate protein synthesis. *Biochimie*, **86**, 933–938.
21. Oppenheim,D.S. and Yanofsky,C. (1980) Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics*, **95**, 785–795.
22. Normark,S., Bergstrom,S., Edlund,T., Grundstrom,T., Jaurin,B., Lindberg,F.P. and Olsson,O. (1983) Overlapping genes. *Ann. Rev. Genet*., **17**, 499–525.

23. Salgado,H., Moreno-Hagelsieb,G., Smith,T. and Collado-Vides,J. (2000) Operons in *Escherichia coli*: genomic analyses and predictions. *Proc. Natl Acad. Sci. USA*, **97**, 6652–6657.
24. Petrov,V.M., Nolan,J.M., Bertrand,C., Levy,D., Desplats,C., Krisch,H.M. and Karam,J.D. (2006) Plasticity of the gene functions for DNA replication in the T4-like phages. *J. Mol. Biol.*, **361**, 46–68.
25. Petrov,V.M., Ratnayaka,S. and Karam,J.D. (2010) Genetic insertions and diversification of the polB-type DNA polymerase (gp43) of T4-related phages. *J. Mol. Biol.*, **395**, 457–474.
26. Kelman,Z., Pietrokovski,S. and Hurwitz,J. (1999) Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* ΔH. *J. Biol. Chem.*, **274**, 28751–28761.
27. Hopfner,K.-P., Eichinger,A., Engh,R.A., Laue,F., Ankenbauer,W., Huber,R. and Angerer,B. (1999) Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc. Natl Acad. Sci.*, **96**, 3600–3605.
28. Hashimoto,H., Nishioka,M., Fujiwara,S., Takagi,M., Imanaka,T., Inoue,T. and Kai,Y. (2001) Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J. Mol. Biol.*, **306**, 469–477.
29. Rodriguez,A.C., Park,H.-W., Mao,C. and Beese,L.S. (2000) Crystal structure of a polB family DNA polymerase from the hyperthermophilc archaeon *Thermococcus sp. 9°N-7*. *J. Mol. Biol.*, **299**, 447–462.
30. Yang,G., Franklin,M., Li,J., Lin,T.-C. and Konigsberg,W. (2002) Correlation of the Kinetics of Finger Domain Mutants in RB69 DNA Polymerase with Its Structure. *Biochemistry*, **41**, 2526–2534.
31. Yang,G., Lin,T.-C., Karam,J. and Konigsberg,W.H. (1999) Steady-state kinetic characterization of Rb69 DNA polymerase mutants that affect dNTP incorporation. *Biochemistry*, **38**, 8094–8101.
32. Evans,S.J., Fogg,M.J., Mamone,A., Davis,M., Pearl,L.H. and Connelly,B.A. (2000) Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.*, **28**, 1059–1066.
33. Gardner,A.F. and Jack,W.E. (1999) Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.*, **27**, 2545–2553.
34. Biles,B.D. and Connolly,B.A. (2004) Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res.*, **32**, e176.
35. Mamone,J.A. (1998) Modified pol-II type DNA polymerases, US Patent 5827716.
36. Zhao,Y., Jeruzalmi,D., Moarefi,I., Leighton,L., Lasken,R. and Kuriyan,J. (1999) Crystal structure of an archaebacterial DNA polymerase. *Structure*, **7**, 1189–1199.
37. Kroutil,L.C., Register,K., Bebenek,K. and Kunkel,T.A. (1996) Exonucleolytic proofreading during replication of repetitive DNA. *Biochemistry*, **35**, 1046–1053.
38. Kim,S.W., Kim,D.-U., Kim,J.K., Kang,L.-W. and Cho,H.-S. (2008) Crystal structure of *Pfu*, the high fidelity DNA polymerase from *Pyrococcus furiosus*. *Int. J. Biol. Macromol.*, **42**, 356–361.
39. Gibrat,J.F., Madej,T. and Bryant,S.H. (1996) Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.*, **6**, 377–385.
40. Freisinger,E., Grollman,A.P., Miller,H. and Kisker,C. (2004) Lesion (in)tolerance reveals insights into DNA replication fidelity. *EMBO J.*, **23**, 1494–1505.
41. Zhong,X., Pedersen,L.C. and Kunkel,T.A. (2008) Characterization of a replicative DNA polymerase mutant with reduced fidelity and increased translesion synthesis capacity. *Nucleic Acids Res.*, **36**, 3892–3904.
42. Shamoo,Y. and Steitz,T.A. (1999) Building a replisome from interacting pieces. *Cell*, **99**, 155–166.