STRUCTURAL BIOLOGY
ISSN 2059-7983

# Structure of the family B DNA polymerase from the hyperthermophilic archaeon *Pyrobaculum calidifontis*

Jingxu Guo,[a] Wenling Zhang,[b] Alun R. Coker,[a] Steve P. Wood,[a] Jonathan B. Cooper,[a,c]* Shazeel Ahmad,[d] Syed Ali,[d,e] Naeem Rashid[d] and Muhummad Akhtar[d]

[a]Wolfson Institute for Biomedical Research, Division of Medicine, UCL, Gower Street, London WC1E 6BT, England, [b]School of Pharmacy, UCL, 29–39 Brunswick Square, London WC1N 1AX, England, [c]Department of Biological Sciences, Birkbeck, University of London, Malet Street, Bloomsbury, London WC1E 7HX, England, [d]School of Biological Sciences, University of the Punjab, Quaid-e-Azam Campus, Lahore 54590, Pakistan, and [e]Department of Biological Sciences, Forman Christian College, Lahore 54600, Pakistan. *Correspondence e-mail: jon.cooper@ucl.ac.uk

Received 16 November 2016
Accepted 13 March 2017

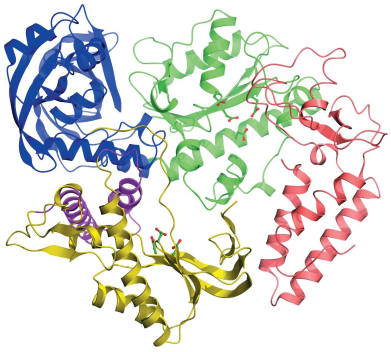
Edited by P. Langan, Oak Ridge National Laboratory, USA

**Keywords:** DNA polymerase; *Pc*-polymerase; protein crystallography; protein structure; molecular replacement; refinement; *Pyrobaculum calidifontis*.

**PDB reference**: DNA polymerase, 5mdn

**Supporting information**: this article has supporting information at journals.iucr.org/d

The family B DNA polymerase from *Pyrobaculum calidifontis* (*Pc*-polymerase) consists of 783 amino acids and is magnesium-ion dependent. It has an optimal pH of 8.5, an optimal temperature of 75°C and a half-life of 4.5 h at 95°C, giving it greater thermostability than the widely used *Taq* DNA polymerase. The enzyme is also capable of PCR-amplifying larger DNA fragments of up to 7.5 kb in length. It was shown to have functional, error-correcting 3′–5′ exonuclease activity, as do the related high-fidelity DNA polymerases from *Pyrococcus furiosus*, *Thermococcus kodakarensis* KOD1 and *Thermococcus gorgonarius*, which have extensive commercial applications. *Pc*-polymerase has a quite low sequence identity of approximately 37% to these enzymes, which, in contrast, have very high sequence identity to each other, suggesting that the *P. calidifontis* enzyme is distinct. Here, the structure determination of *Pc*-polymerase is reported, which has been refined to an *R* factor of 24.47% and an $R_{free}$ of 28.81% at 2.80 Å resolution. The domains of the enzyme are arranged in a circular fashion to form a disc with a narrow central channel. One face of the disc has a number of connected crevices in it, which allow the protein to bind duplex and single-stranded DNA. The central channel is thought to allow incoming nucleoside triphosphates to access the active site. The enzyme has a number of unique structural features which distinguish it from other archaeal DNA polymerases and may account for its high processivity. A model of the complex with the primer-template duplex of DNA indicates that the largest conformational change that occurs upon DNA binding is the movement of the thumb domain, which rotates by 7.6° and moves by 10.0 Å. The surface potential of the enzyme is dominated by acidic groups in the central region of the molecule, where catalytic magnesium ions bind at the polymerase and exonuclease active sites. The outer regions are richer in basic amino acids that presumably interact with the sugar-phosphate backbone of DNA. The large number of salt bridges may contribute to the high thermal stability of this enzyme.





## 1. Introduction

Hyperthermophilic archaea of the genus *Pyrobaculum* (Latin for ‘fire stick’) are able to thrive in the boiling water conditions found in hydrothermal springs and underwater volcanic vents. The organism *P. calidifontis* is a facultative aerobic hyperthermophile that was first isolated from a hot spring in the Philippines (Amo *et al.*, 2002).

Archaea possess replication, transcription and translation machineries that have many similarities to the systems present

in eukaryotes and are certainly more complex than those present in bacteria (Gribaldo *et al.*, 2010). DNA-dependent DNA polymerases are key enzymes in DNA replication and are classified into six main families named A, B, C, D, X and Y (Garcia-Diaz & Bebenek, 2007). DNA polymerases (EC 2.7.7.7) play a pivotal role in DNA replication and repair, and those from thermophilic organisms have extensive biotechnological applications in polymerase chain reaction (PCR) methods (Wu *et al.*, 2014). These include the amplification of specific genes or regions of interest in the genome, DNA sequencing and site-directed mutagenesis, as well as diagnostic and forensic work. The well-known DNA polymerase from *Thermus aquaticus* (*Taq* DNA polymerase) was the first to be characterized (Chien *et al.*, 1976) and used biotechnologically (Bartlett & Stirling, 2003). However, it has no 3′–5′ exonuclease activity, which gives other DNA polymerases their error-correcting or proofreading ability. Archaeal DNA polymerases are known to possess 3′–5′ exonuclease activity and hence have better fidelity when compared with *Taq* DNA polymerase. Curiously, the archaeal DNA polymerases belong to the same family as many of the eukaryotic DNA polymerases and are thus thought to be representative of the replication machinery in higher organisms. Archaeal DNA polymerases are also of interest since many of them are produced by the unusual and complex mechanism of protein splicing, which involves the excision of an intein region from between two polymerase-encoding extein regions (Perler, 1999; Hashimoto *et al.*, 2001). The intervening intein is a homing endonuclease which is thought to facilitate entry of the DNA encoding the intein itself into the chromosome, as well as maintaining its presence and transferability thereafter. However, the gene for *Pc*-polymerase does not contain an intein.

The cloning of the gene encoding the family B DNA polymerase from *P. calidifontis* into the *Escherichia coli* expression vector pET-21a has been reported previously (Ali *et al.*, 2011). Studies of the purified enzyme (*Pc*-polymerase), which consists of 783 amino acids, established that its activity is magnesium-ion dependent, with an optimal pH of 8.5, an optimal temperature of 75°C and a half-life of 4.5 h at 95°C. Thus, the enzyme has greater thermostability than the widely used *Taq* DNA polymerase. Whilst *Taq* DNA polymerase requires monovalent cations for activity, *Pc*-polymerase was found to be inhibited by potassium and ammonium ions and was capable of PCR-amplifying larger DNA fragments of up to 7.5 kb in length. It was shown to have functional, error-correcting 3′–5′ exonuclease activity, which is in accord with its relatively close relationship to the high-fidelity DNA polymerases from *Pyrococcus furiosus* (*Pfu* DNA polymerase) and *Thermococcus kodakarensis* (previously known as *Pyrococcus kodakaraensis*) KOD1 (KOD DNA polymerase), which have extensive commercial applications.

Structures are available for the DNA polymerases from *P. furiosus* (Kim *et al.*, 2008), *T. kodakarensis* KOD1 (Kuroita *et al.*, 2005) and *T. gorgonarius* (Hopfner *et al.*, 1999; Firbank *et al.*, 2008). The structures consist of five domains starting with the N-terminal domain followed by the 3′–5′ exonuclease domain and then three domains associated with the 5′–3′ DNA polymerase activity, which are known as the palm, fingers and thumb domains (Brautigam & Steitz, 1998). The exonuclease domain contains a region of antiparallel $\beta$-sheet with a number of catalytic carboxylate residues. The palm domains, which possess three catalytic carboxylate groups binding two magnesium ions, tend to have a more conserved structure than the fingers and thumb domains in the different DNA polymerase families. Typically, the palm domain consists of a four- to six-stranded $\beta$-sheet that is flanked on one side by two $\alpha$-helices. The thumb and fingers domains arise as 'insertions' in the loops of the palm domain, which vary in position and structure from one family of polymerase to another. However, the thumb domains all tend to be predominantly helical and use at least one helix to make extensive contacts with the minor groove of the bound duplex. The thumb domain is also involved in binding other proteins known as processivity factors, which stabilize the complex with DNA over many thousands of nucleotide additions. For example, phage T7 DNA polymerase recruits the small protein thioredoxin from the *E. coli* host cell, which increases processivity at least 100 times (Brautigam & Steitz, 1998). The high processivity and elongation rate of KOD DNA polymerase are thought to arise from a group of seven arginine residues which interact with the DNA molecule where the two oligonucleotide strands separate. These residues are likely to stabilize the molten form of the DNA for replication and editing (Hashimoto *et al.*, 2001). In addition, there is another extended region which is rich in basic residues in the fingers domain and this is thought to assist in the delivery of incoming nucleotides to the active site of the palm domain.

Much effort has been directed towards engineering DNA polymerase mutants with greater processivity and fidelity, and information gleaned from structural studies has been of great importance. For example, there is evidence that amino acids at the interface between the thumb and exonuclease domains are important in the fidelity of the polymerase (Kim *et al.*, 2008). The control of the exonuclease *versus* polymerase activity is also mediated by a flexible loop containing a Y-GG/A motif which resides between the 3′–5′ exonuclease domain and the palm domain (Böhlke *et al.*, 2000).

Despite the structural differences between the polymerase enzymes from different classes, the arrangement of the catalytic carboxyl groups in the palm domain and the bound metal ions are strongly conserved with respect to the bound DNA molecule (Wu *et al.*, 2014). Of the two bound metal ions, which are denoted A and B, metal ion A is thought to activate the 3′-hydroxyl group of the primer strand by lowering its p$K_a$. The hydroxyl can then act as a nucleophile to attack the $\alpha$-phosphate group of the incoming nucleotide to generate a pentacoordinated intermediate which is stabilized by both of the metal ions. Indeed, both of the metal ions are thought to leave the active site with the pyrophosphate group product as it dissociates from the enzyme after completion of the catalytic cycle. Between cycles, the bound DNA and/or polymerase must translocate so that the new primer terminus is positioned correctly for the next catalytic cycle.

**Table 1**
X-ray data statistics for the structure.

Values in parentheses are for the outer resolution shell.

| | |
|---|---|
| Beamline | I02, DLS |
| Wavelength (Å) | 0.9795 |
| Space group | $P2_1$ |
| Unit-cell parameters | |
| $a$ (Å) | 74.19 |
| $b$ (Å) | 100.74 |
| $c$ (Å) | 119.34 |
| $\beta$ (°) | 94.72 |
| Resolution (Å) | 118.93–2.80 (2.90–2.80) |
| $R_{merge}$† (%) | 10.3 (77.0) |
| $R_{meas}$‡ (%) | 12.1 (91.2) |
| $CC_{1/2}$§ (%) | 99.1 (58.8) |
| Completeness (%) | 97.5 (96.9) |
| Average $I/\sigma(I)$ | 6.8 (1.8) |
| Multiplicity | 3.6 (3.7) |
| No. of observed reflections | 154369 (16405) |
| No. of unique reflections | 42313 (4403) |
| Wilson plot $B$ factor (Å$^2$) | 66.84 |
| Solvent content (%) | 50.68 |
| $R$ factor (%) | 24.47 |
| Free $R$ factor (%) | 28.81 |
| R.m.s.d., bond lengths (Å) | 0.004 |
| R.m.s.d., bond angles (°) | 0.818 |
| No. of reflections in working set | 42206 (4123) |
| No. of reflections in test set | 2132 (236) |
| Mean protein $B$ factor (Å$^2$) | 73.47 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of the $N(hkl)$ observations $I_i(hkl)$ of each unique reflection $hkl$ after scaling. § $CC_{1/2}$ values are the half-set correlation coefficients (Karplus & Diederichs, 2012).

Here, we report the structure determination of the DNA polymerase from *P. calidifontis* (*Pc*-polymerase), which has been refined to an $R$ factor of 24.47% and an $R_{free}$ of 28.81% at 2.80 Å resolution. The enzyme has a number of unique structural features which distinguish it from other archaeal DNA polymerases and may account for its high processivity.

## 2. Methods

### 2.1. Crystallization

Screening for crystallization conditions was conducted by the hanging-drop vapour-diffusion method using *Pc*-polymerase which had been expressed and purified as described by Ali *et al.* (2011). The enzyme, at a concentration of 15 mg ml$^{-1}$ in 20 m*M* Tris pH 8.2, was dispensed using a Mosquito crystal screening robot (TTP Labtech, Hertfordshire, England) into 96-well SBS flat-bottom plates (Molecular Dimensions, Suffolk, England). Three screening kits, Structure Screen 1+2, PACT *premier* and JCSG-*plus* (Molecular Dimensions, Suffolk, England), were used and the plates were stored at both 4 and 21°C. Two crystals, with lengths of approximately 200 µm in two dimensions and of approximately 50 µm in thickness, were obtained after four weeks in Structure Screen 1+2 condition F4 (0.2 *M* potassium thiocyanate, 0.1 *M* bis-tris propane pH 6.5, 20% PEG 3350) at 21°C, but attempts to reproduce these crystals for optimization did not succeed. However, the crystals from the original screen were transferred to mother-liquor solution to which the cryoprotectant glycerol was added to 30%(*v*/*v*), and these were then mounted in loops before flash-cooling and storage in liquid nitrogen.

### 2.2. Data collection and processing

X-ray data collection was undertaken at station I02 at Diamond Light Source (DLS), Didcot, England and integration of the data was performed using *DIALS* (Waterman *et al.*, 2013; Gildea *et al.*, 2014). This revealed that the crystal belonged to space group $P2_1$, which was confirmed using *POINTLESS* (Evans, 2006; Evans, 2011). Scaling and data reduction using *AIMLESS* (Evans & Murshudov, 2013) showed that the best crystal diffracted to 2.80 Å resolution and *phenix.xtriage* (Zwart *et al.*, 2005) indicated that neither twinning nor translational NCS were present. The solvent content was estimated to be 51%, with two molecules per asymmetric unit, using *MATTHEWS_COEF* (Matthews, 1968; Kantardjieff & Rupp, 2003).

### 2.3. Structure determination

Molecular replacement, which was carried out with the *BALBES* server (Long *et al.*, 2008), gave a top solution with a *Q*-factor of 0.67 and suggested that the probability of this being the solution was 99%. The corresponding $R$ factor and $R_{free}$ values were 37.6 and 44.1%, respectively. In this analysis, four different homologous DNA polymerase structures were used as search models to locate different parts of the *Pc*-polymerase structure. These were from *T. gorgonarius* (PDB entry 2xhb; 35.2% overall sequence identity; Killelea *et al.*, 2010) for the first segment, *P. furiosus* (PDB entry 3a2f; 36.0% overall sequence identity; H. Nishida, K. Mayanagi, S. Kiyonari, Y. Sato, Y. Ishino & K. Morikawa, unpublished work) for the second, *T. kodakarensis* KOD1 (PDB entry 1wn7; 35.2% overall sequence identity; Kuroita *et al.*, 2005) for the third part and *Sulfolobus solfataricus* (PDB entry 1s5j; 35.3% overall sequence identity; Savino *et al.*, 2004) for the last region of the model. Since there were many residues which fitted the electron density poorly at this stage, multiple rounds of manual rebuilding were undertaken using *Coot* (Emsley *et al.*, 2010). This included the relocation of large stretches of the protein, which was guided by the electron density for large aromatic side chains. Following this, refinement using *phenix.refine* (Adams *et al.*, 2010; Chen *et al.*, 2010; Afonine *et al.*, 2012; Echols *et al.*, 2012) decreased the $R_{free}$ value to 35.8%. This was followed by NCS refinement with torsion-angle restraints using *phenix.refine*, which gave an $R_{free}$ value of 32.7% and revealed more electron density in the last domain of chain *A*. Several rounds of manual rebuilding and NCS refinement along with the introduction of metal ions and water molecules, followed by further stereochemically restrained refinement, gave $R$ factor and $R_{free}$ values of 24.47 and 28.81%, respectively. All of the statistics for data collection, data processing and refinement of the final model are given in Table 1. Analysis of hydrogen bonds and salt bridges was conducted using the online *VADAR* (Willard *et al.*, 2003) and *ESBRI* (Costantini *et al.*, 2008) servers. Sequence-alignment and molecular-graphics figures were prepared

with *ALSCRIPT* (Barton, 1993) and *CueMol* (http://www.cuemol.org/en), respectively.

## 3. Results and discussion

When compared with other DNA polymerases of known structure, *Pc*-polymerase appears to have a number of distinct features. An alignment of *Pc*-polymerase with the *T. gorgonarius*, *Pfu* and KOD DNA polymerases showing the secondary-structure elements observed in our analysis is presented in Fig. 1. *Pc*-polymerase has a quite low sequence identity of approximately 37% to the other three enzymes, although the secondary structure is substantially conserved. In contrast, the other three DNA polymerases have 80–90% sequence identity to each other, suggesting that they belong to a phylogenetically distinct group. Indeed, *P. calidifontis* is classified as a member of the order Thermoproteales, whereas the others belong to the order Thermococcales.

### 3.1. Tertiary structure

The N-terminal domain of *Pc*-polymerase has a bilobal or dumbbell appearance, which is somewhat more elaborate than the *T. gorgonarius*, *Pfu* and KOD DNA polymerases owing to the insertion of a large 28-residue $\beta$-hairpin in the region just before the polypeptide enters the exonuclease domain (see Fig. 2). This additional $\beta$-hairpin gives the first lobe of this domain an elegant seven-stranded $\beta$-barrel structure.

The exonuclease domain extends from residue 164 to approximately residue 350 and consists of an eight-stranded antiparallel $\beta$-sheet flanked on both sides by helical segments. In the other thermophilic DNA polymerases the sheet is slightly larger, being formed of nine strands. The loss of one strand in *Pc*-polymerase arises from the deletion of approximately eight residues, which removes a short $\beta$-strand at the exposed outer edge of the protein. There are other large differences in the loop regions, in particular the loop following the first $\beta$-strand of this domain. Other loops adopt very

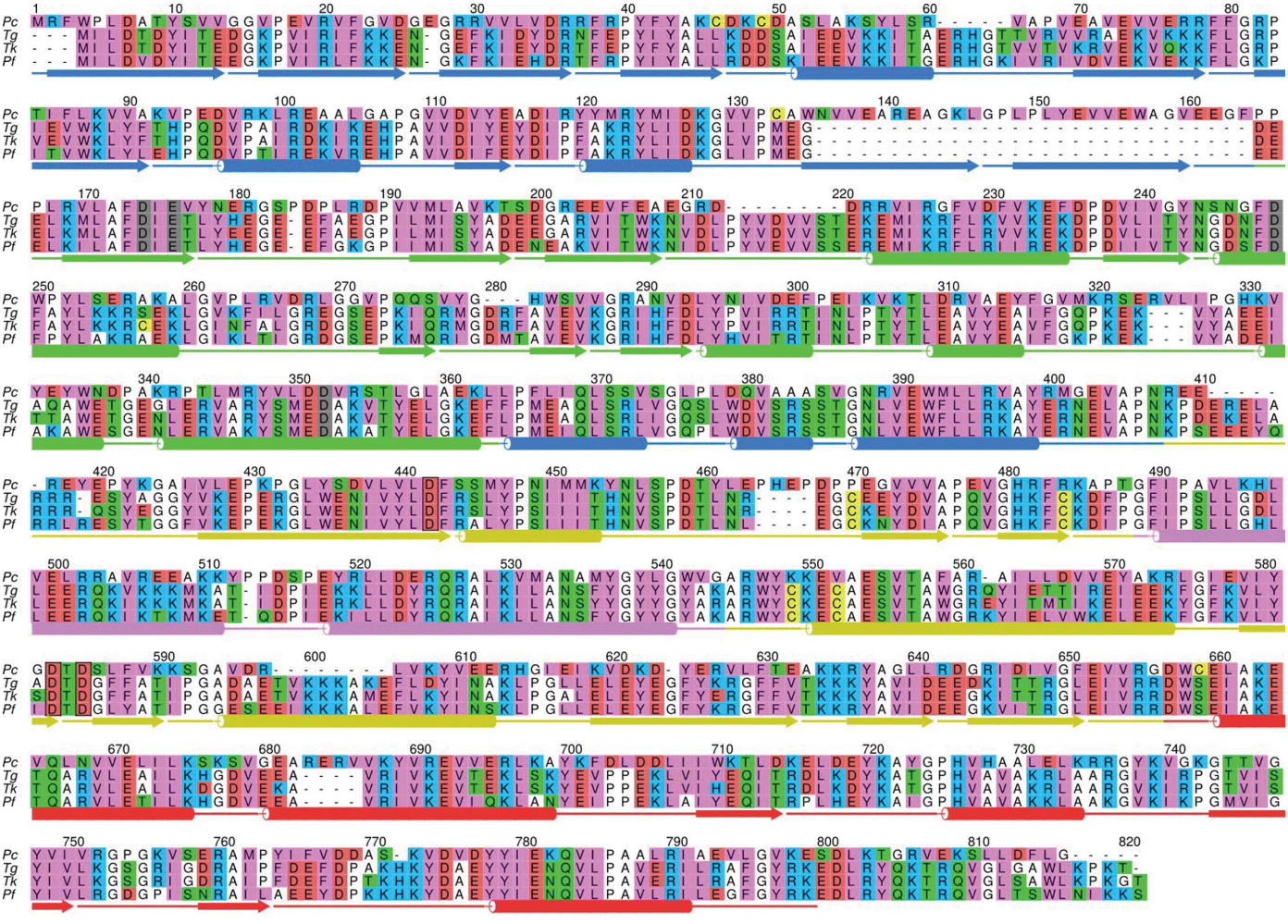



**Figure 1**
A structure-based sequence alignment of *Pc*-polymerase with the enzymes from *T. gorgonarius* (*Tg*), *T. kodakarensis* (*Tk*) and *P. furiosus* (*Pf*). The amino acids are coloured according to the following scheme: acidic, red; basic, pale blue; neutral polar, green; hydrophobic, purple; cysteine, yellow; the structurally important residues Gly, Ala and Pro are in white. The catalytic metal-binding carboxylate residues are shown boxed in the polymerase domain and shaded grey in the exonuclease domain. The secondary-structure elements are shown in the bottom row and are coloured blue to indicate the N-terminal domain, green for the exonuclease domain and yellow, purple and red for the palm, fingers and thumb domains, respectively.

different conformations in the homologous structures, and in superpositions these regions appear to splay apart by as much as 6 Å.

The catalytic residues in the exonuclease domain consist of Asp169, Glu171, Asp236 and Asp336, which are involved in binding two magnesium ions. The first two of these residues are located within the first $\beta$-strand of this domain and are most closely involved in binding the catalytic metal ions. The remaining two aspartates are located within $\alpha$-helical segments which flank the central $\beta$-sheet.

Immediately following this domain is a stretch of 40 residues (350–390) which forms a small $\alpha$-helical region that is considered to be part of the N-terminal domain. Indeed, the helical segments in this region partially plug the gap between the lobes of the N-terminal domain and appear to spatially link this domain to the remainder of this disc-shaped protein. Following this region, there is an extended stretch of polypeptide (residues 390–403) which is on the surface of the protein and joins this region to the palm domain. The linker region (shown by a # symbol in Fig. 2) exhibits a significant structural difference from the other archaeal DNA polymerases, which adopt a helical conformation in this region, which is at a distance of some 10 Å away from the equivalent part of *Pc*-polymerase. However, this difference is likely to be a consequence of the exposed nature and flexibility of this region, which would form unfavourable contacts with symmetry-related molecules if it adopted the conformation found in homologous structures.

The palm domain is formed by residues 405–465 and by residues 525–620, with the intervening region (approximately residues 466–524) forming the fingers domain. The first few dozen residues of the palm domain adopt a $\beta$-strand conformation which straddles both $\beta$-sheets of this domain before forming a short helix and then a relatively irregular region prior to entering the fingers domain. On re-entering the palm domain, the polypeptide forms a large helix of approximately seven turns which leads into the main $\beta$-barrel of this domain. In essence, the barrel is cup-shaped and consists of some ten $\beta$-strand regions, with one of the intervening loops containing a region of $\alpha$-helix which is around four turns in length. The active site consists of Asp420, Asp560 and Asp562 and is well defined by the electron-density map, although there is no electron density for the catalytically essential magnesium ions. The aspartates are in adjacent strands of the $\beta$-barrel, which has the appearance of a cup with its mouth facing towards the centre of the molecule on its DNA-binding face. The fingers domain consists of two antiparallel $\alpha$-helices, each of six to seven turns, that form a helical hairpin which appears to lean over the active site of the palm domain.

The final domain of the protein is the thumb domain, which spatially lies between the palm and exonuclease domains and is formed by residues 625–766. The thumb domain consists of three $\alpha$-helices, two of which occur at the N-terminal end of the polypeptide forming this domain and the last of which occurs at the C-terminus of the protein. The intervening region of the polypeptide forms an irregular three-stranded

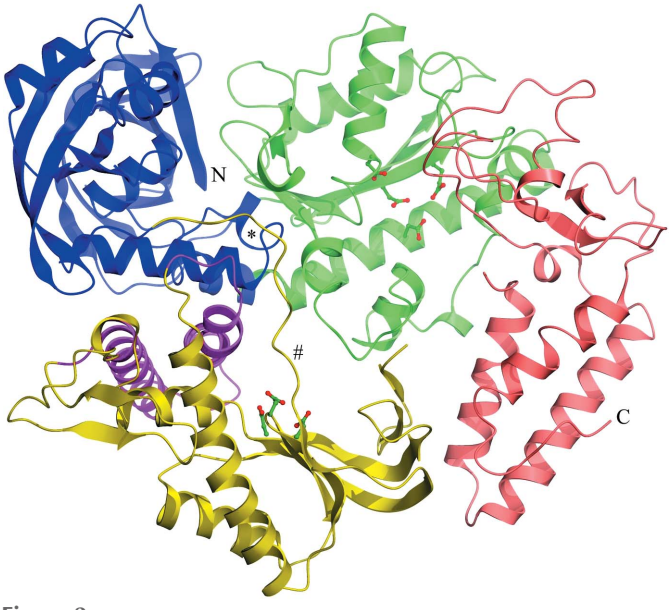



**Figure 2**
The overall structure of *P. calidifontis* DNA polymerase. Blue indicates the N-terminal domain and green indicates the exonuclease domain, followed by yellow, purple and red for the palm, fingers and thumb domains, respectively. The catalytic metal-binding carboxylate side chains forming the active sites of the exonuclease and palm (polymerase) domains are shown in ball-and-stick representation. The small helix which plugs the gap between the lobes of the N-terminal domain is indicated by an asterisk (*) and the linker region between the N-terminal domain and the palm domain is indicated by a hash (#). Overall, the molecule is disc-shaped, with pronounced grooves on one face of the disc which are associated with DNA binding.

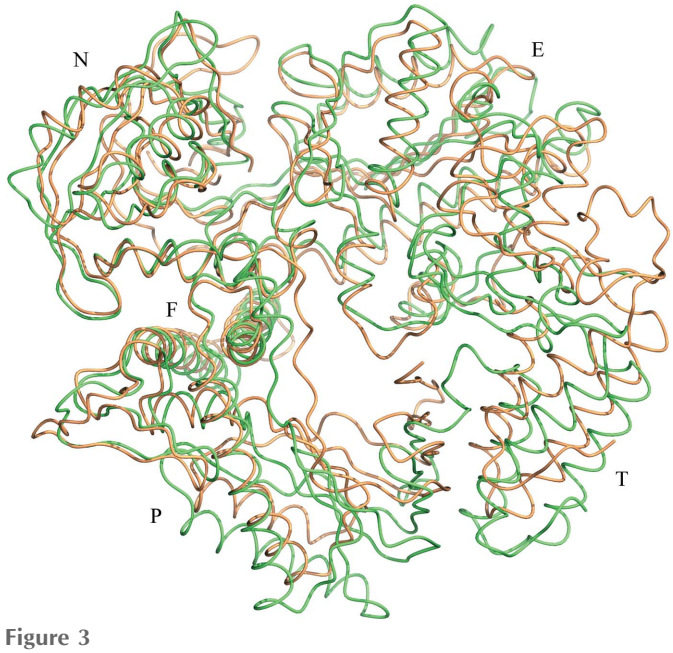



**Figure 3**
A superposition of *Pc*-polymerase with *Pfu* DNA polymerase. The structures are coloured cream for *Pc*-polymerase and green for *Pfu* DNA polymerase, with the N-terminal, exonuclease, palm, fingers and thumb domains labelled N, E, P, F and T, respectively. The thumb domain undergoes the greatest movement upon binding to the DNA substrate, moving away from the catalytic centre.

$\beta$-sheet with one more short $\alpha$-helical region flanking the outer face of the sheet.

### 3.2. Comparison with *Pfu* polymerase

The superposition shown in Fig. 3 demonstrates a number of quite significant structural differences between *Pc*-polymerase and *Pfu* DNA polymerase (hereafter referred to as *Pfu*). The first of these is the insertion of a $\beta$-hairpin at the junction between the N-terminal and exonuclease domains, although this region is not close to the active site. In the latter domain the largest difference between the two structures appears to be the additional $\beta$-hairpin formed by residues 173–186 of *Pc*-polymerase which, again, is not close to the active-site region. It is interesting that in *Pfu* there is a short insertion in a $\beta$-hairpin formed by residues 244–247 which points towards the active site and could therefore affect its catalytic properties. Other regions with differences of potential significance include the loop region from residue 284 to 289 in *Pc*-polymerase, which is closer to the catalytic centre of the enzyme than in *Pfu*. There are additional differences in the loop region involving residues 302–311 of *Pc*-polymerase. As mentioned above, a large difference between *Pc*-polymerase and *Pfu* occurs in the region involving residues 391–404 of *Pc*-polymerase. In the palm domain the loop between residues 440 and 450 of *Pc*-polymerase is shifted relative to the equivalent loop of the *Pfu* enzyme, and in the fingers domain the helix hairpin loop involving residues 490–494 adopts a different conformation, although this difference is at some distance from the active site. In the palm domain at the C-terminal end of the protein there are a number of significant conformational changes, the largest being within the hairpin loop between residues 644 and 650 of *Pc*-polymerase.

### 3.3. Model of the complex with DNA

In three dimensions, the domains of DNA polymerase are arranged in a circular fashion to form a disc with a narrow central channel, reminiscent of a ring doughnut. One face of the disc or ring has a number of connected crevices in it which will allow the protein to bind duplex and single-stranded DNA. One of these crevices, known as the T cleft, binds single-stranded template DNA, whereas the other (the D cleft) binds the duplex that is formed by the action of the polymerase. Another cleft, which is referred to as the editing channel, leads to the 3′–5′ exonuclease active site. This channel directs the primer strand towards the exonuclease active site when a mismatched nucleotide is incorporated. This involves the unwinding of the newly synthesized DNA duplex so that the enzyme can then operate in its editing mode. The fingers and thumb domains tend to move relative to the palm domain when DNA binds and hold it in the manner of a clamp.

Construction of a simple model of the complex that *Pc*-polymerase forms with DNA was undertaken by fitting our structure of the enzyme to the DNA-bound structure of *P. furiosus* DNA polymerase (PDB entry 4ail; Wynne *et al.*, 2013). This structure was derived from a mutant form of *Pfu* which was co-crystallized with a primer-template duplex of DNA with a 2′–3′ dideoxy terminal cytosine at the 3′ end of the primer strand to stall the polymerase. Visual inspection of the superposed structures suggested that the domains of *Pc*-polymerase are likely to move appreciably on binding to DNA, although the N-terminal and exonuclease domains appeared to fit well to the *Pfu* structure without further manipulation. To attempt a correction for the domain movements, the structure of *Pc*-polymerase was first fitted to the *Pfu* structure using the N-terminal and exonuclease domains. The remaining domains were then fitted separately to the equivalent domains of the *Pfu* protein–DNA complex, and the resulting composite model is shown in Fig. 4. In this model of the conformational changes that occur upon DNA binding, the thumb domain rotates by 7.6° and its centroid moves by 10.0 Å. These movements cause the small $\beta$-sheet region of this domain to shift towards the DNA, whereas the helical hairpin region moves further away. As was found when the *Pfu*–DNA complex was compared with the native protein, the other domains move to a lesser extent. For example, the palm domain of *Pc*-polymerase moves by some 4.0 Å and rotates by 15.5° away from the central region of the molecule to accommodate both strands of the bound DNA. The fingers domain rotates by 14.1° and moves by only 1.4 Å. In this model of the complex, the 3′ end of the primer strand is positioned close to the active-site residues of the palm domain,

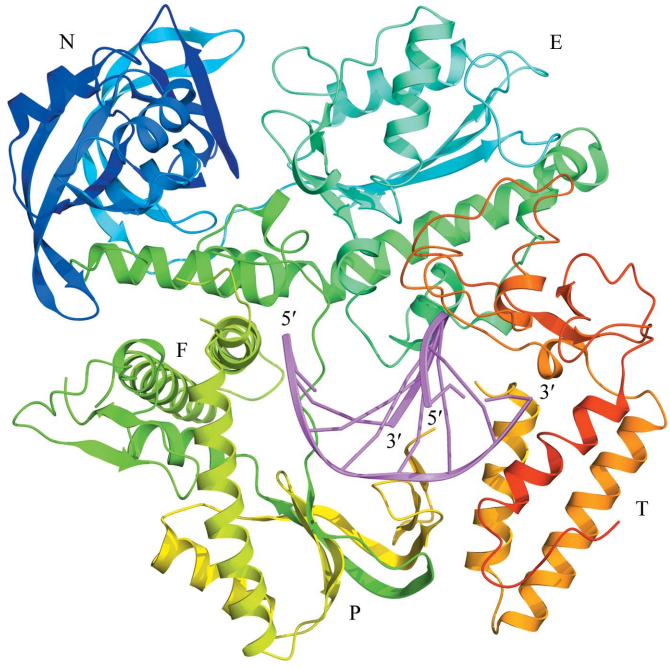



**Figure 4**
A model of the complex with DNA. This composite model was constructed by fitting each of the *Pc*-polymerase domains to the *Pfu*–DNA complex (PDB entry 4ail) one at a time and is shown in the same view as Fig. 2. The N-terminal, exonuclease, palm, fingers and thumb domains are labelled N, E, P, F and T, respectively, while the two DNA strands from the *Pfu* complex and their associated base-pairing are indicated in purple. The shorter of the two DNA strands is the primer strand and has its 3′ end close to the catalytic site of the palm domain. The N-terminal end of the polypeptide is coloured blue and the C-terminal end is coloured red.

being slightly closer to Asp560 and Asp562 than to Asp420. Curiously, the large linker region between the exonuclease and palm domains appears to partially block the binding site for the template strand of the DNA bound in the *Pfu* polymerase structure, suggesting that this region of *Pc*-polymerase may adopt a different local conformation from that seen in the crystal in order to bind its substrate. As described above, the conformation of this region of *Pc*-polymerase is most likely to be affected by crystal contacts.

### 3.4. Electrostatic surface

The solvent-accessible surface of the enzyme demonstrates the presence of two marked clefts for binding the DNA substrate (see Fig. 5). These are referred to as the D and T clefts, which bind the duplex and template DNA, respectively. The electrostatic surface potential of the enzyme is dominated by acidic groups in the central region of the molecule, which has a high propensity for binding the catalytic magnesium ions at the polymerase and exonuclease active sites. The outer regions have a more positive potential, which presumably allows interactions with the sugar-phosphate backbone of DNA.

### 3.5. Thermostability

Recent work has indicated that the thermostability of proteins from extremophilic organisms may stem from increased rigidity of the structure in the room-temperature range owing to greater hydrogen-bonding constraints (Wells *et al.*, 2014). These form a network of polar interactions stabilizing the extremophile structure at high temperature. However, a simple analysis suggests that the number of hydrogen bonds present in *Pc*-polymerase is very similar to that in the native structure of the mesophilic family B *E. coli* DNA polymerase II (PDB entry 1q8i; Midwest Center for Structural Genomics, unpublished work). Indeed, the 2 Å resolution *E. coli* DNA polymerase structure has a slightly higher hydrogen-bond content, with 77% of residues forming hydrogen bonds compared with only 75% in *Pc*-polymerase. However, analysis of the salt-bridge content reveals a more marked difference. For instance, using a donor–acceptor atom cutoff distance of 4 Å indicates that *Pc*-polymerase contains 60 ionic side-chain interactions, whereas *E. coli* DNA polymerase II contains only 40. Since buried electrostatic interactions are likely to have a slightly longer range than this owing to the lower dielectric constant of the protein interior, the cutoff distance was raised to 6 Å. This increased the estimated number of ionic interactions in *Pc*-polymerase and the *E. coli* enzyme to 212 and 124, respectively: a ratio of almost 2:1, a finding which is repeated with other archaeal DNA polymerases. Thus, the difference between the thermophilic and mesophilic enzyme in this analysis becomes more marked if we relax the conditions for electrostatic interactions, perhaps suggesting that these longer range electrostatic effects may be important in the extreme thermostability of *Pc*-polymerase.

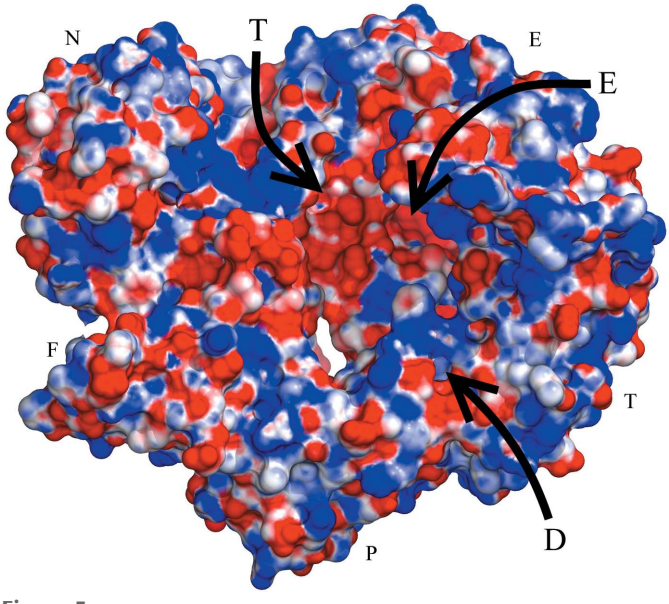



**Figure 5**
The solvent-accessible surface of *Pc*-polymerase coloured by electrostatic potential. The acidity of the central region of the active site stems from the abundance of metal-binding carboxylate groups, which are essential for activity, whereas the more peripheral basic regions allow the protein to bind the phosphate backbone of the DNA substrate ionically. The hole which appears to pass right through the centre of the molecule is thought to be the entry passage for incoming dNTPs, which are added to the primer strand. The arrows indicate the general directions of the duplex cleft (D), the template cleft (T) and the exonuclease or editing channel (E). The molecule is shown in the same orientation as Figs. 2 and 3, with the domains labelled in small text.

## Acknowledgements

We gratefully acknowledge Diamond Light Source for X-ray beam time and travel support for data collection (award MX12342).

## References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Ali, S. F., Rashid, N., Imanaka, T. & Akhtar, M. (2011). *J. Biosci. Bioeng.* **112**, 118–123.
Amo, T., Paje, M. L., Inagaki, A., Ezaki, S., Atomi, H. & Imanaka, T. (2002). *Archaea*, **1**, 113–121.
Bartlett, J. M. & Stirling, D. (2003). *Methods Mol. Biol.* **226**, 3–6.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.
Böhlke, K., Pisani, F. M., Vorgias, C. E., Frey, B., Sobek, H., Rossi, M. & Antranikian, G. (2000). *Nucleic Acids Res.* **28**, 3910–3917.
Brautigam, C. A. & Steitz, T. A. (1998). *Curr. Opin. Struct. Biol.* **8**, 54–63.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Chien, A., Edgar, D. B. & Trela, J. M. (1976). *J. Bacteriol.* **127**, 1550–1557.
Costantini, S., Colonna, G. & Facchiano, A. M. (2008). *Bioinformation*, **3**, 137–138.
Echols, N., Grosse-Kunstleve, R. W., Afonine, P. V., Bunkóczi, G., Chen, V. B., Headd, J. J., McCoy, A. J., Moriarty, N. W., Read, R. J.,

Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Adams, P. D. (2012). *J. Appl. Cryst.* **45**, 581–586.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J. & Connolly, B. A. (2008). *J. Mol. Biol.* **381**, 529–539.
Garcia-Diaz, M. & Bebenek, K. (2007). *Crit. Rev. Plant Sci.* **26**, 105–122.
Gildea, R. J., Waterman, D. G., Parkhurst, J. M., Axford, D., Sutton, G., Stuart, D. I., Sauter, N. K., Evans, G. & Winter, G. (2014). *Acta Cryst.* D**70**, 2652–2666.
Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P. & Brochier-Armanet, C. (2010). *Nature Rev. Microbiol.* **8**, 743–752.
Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T. & Kai, Y. (2001). *J. Mol. Biol.* **306**, 469–477.
Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. & Angerer, B. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 3600–3605.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Killelea, T., Ghosh, S., Tan, S. S., Heslop, P., Firbank, S. J., Kool, E. T. & Connolly, B. A. (2010). *Biochemistry*, **49**, 5772–5781.
Kim, S. W., Kim, D.-U., Kim, J. K., Kang, L.-W. & Cho, H.-S. (2008). *Int. J. Biol. Macromol.* **42**, 356–361.
Kuroita, T., Matsumura, H., Yokota, N., Kitabayashi, M., Hashimoto, H., Inoue, T., Imanaka, T. & Kai, Y. (2005). *J. Mol. Biol.* **351**, 291–298.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Perler, F. B. (1999). *Trends Biochem. Sci.* **24**, 209–211.
Savino, C., Federici, L., Johnson, K. A., Vallone, B., Nastopoulos, V., Rossi, M., Pisani, F. M. & Tsernoglou, D. (2004). *Structure*, **12**, 2001–2008.
Waterman, D., Winter, G., Parkhurst, J., Fuentes-Montero, L., Hattne, J., Brewster, A., Sauter, N. & Evans, G. (2013). *CCP4 Newsl. Protein Crystallogr.* **49**, 13–15.
Wells, S. A., Crennell, S. J. & Danson, M. J. (2014). *Proteins*, **82**, 2657–2670.
Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). *Nucleic Acids Res.* **31**, 3316–3319.
Wu, S., Beard, W. A., Pedersen, L. G. & Wilson, S. H. (2014). *Chem. Rev.* **114**, 2759–2774.
Wynne, S. A., Pinheiro, V. B., Holliger, P. & Leslie, A. G. W. (2013). *PLoS One*, **8**, e70892.
Zwart, P., Grosse-Kunstleve, R. & Adams, P. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, 27–35.