Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The PAD region in the mycobacterial DinB homologue MsPolIV exhibits positional heterogeneity

Amit Sharma, Vidya Subramanian and Deepak T. Nair*

National Centre for Biological Sciences (NCBS-TIFR), UAS-GKVK Campus, Bellary Road, Bangalore 560 065, India

Correspondence e-mail: deepaknair@ncbs.res.in

Received 4 March 2012
Accepted 20 April 2012

PDB Reference: MsPolIV, 4dez.

Y-family DNA polymerases (dPols) have evolved to carry out translesion bypass to rescue stalled replication; prokaryotic members of this family also participate in the phenomenon of adaptive mutagenesis to relieve selection pressure imposed by a maladapted environment. In this study, the first structure of a member of this family from a prokaryote has been determined. The structure of MsPolIV, a Y-family dPol from *Mycobacterium smegmatis*, shows the presence of the characteristic finger, palm and thumb domains. Surprisingly, the electron-density map of the intact protein does not show density for the PAD region that is unique to members of this family. Analysis of the packing of the molecules in the crystals showed the existence of large solvent-filled voids in which the PAD region could be located in multiple conformations. In line with this observation, analytical gel-filtration and dynamic light-scattering studies showed that MsPolIV undergoes significant compaction upon DNA binding. The PAD region is known to insert into the major groove of the substrate DNA and to play a major role in shaping the active site. Comparison with structures of other Y-family dPols shows that in the absence of tertiary contacts between the PAD domain and the other domains this region has the freedom to adopt multiple orientations. This structural attribute of the PAD will allow these enzymes to accommodate the alterations in the width of the DNA double helix that are necessary to achieve translesion bypass and adaptive mutagenesis and will also allow regulation of their activity to prevent adventitious error-prone DNA synthesis.

1. Introduction

DNA polymerases (dPols) can carry out template-dependent nucleotide incorporation and are therefore centrally involved in the replication of genomic and episomal DNA. In order to carry out this crucial function, these replicative dPols exhibit extremely high accuracy and processivity. The presence of damaged nucleotides, termed DNA lesions, in the genome is highly inhibitory to the activity of these dPols. DNA can be damaged by a variety of external agents (radiation and chemicals) and internal agents (free radicals and reactive intermediates). The high fidelity exhibited by replicative polymerases arises from the ability of their active sites to stringently select the correct Watson–Crick base-pairing partner for a given template nucleotide. These enzymes are unable to stabilize damaged template nucleotides and the correct incoming nucleotides in their active sites, leading to stalling of the replication fork. To overcome this predicament, all organisms have been found to have specialized dPols that usually belong to the Y-family.

The members of the Y-family usually exhibit low fidelity and low processivity and are able to carry out DNA synthesis past various kinds of damaged nucleotides, thus achieving translesion bypass to rescue replication (Goodman, 2002). The presence of specialized active sites allows many members of this family to stabilize damaged nucleotides and incorporate nucleotides opposite them. Each of these enzymes is probably capable of productive incorporation opposite a distinct set of lesions. Sometimes, after this initial incorporation certain other members of the family may be recruited to act as extender polymerases and add several further nucleotides to achieve complete translesion bypass. It is believed that once these enzymes have helped the replication fork to cross the damage in this manner, they are replaced by normal replicative dPols (Sutton, 2010; Indiani *et al.*, 2005). In addition, it has also been shown that the low fidelity of these molecules is exploited to generate adaptive mutations that can relieve selection pressure arising from adverse environmental conditions (Galhardo *et al.*, 2007; Foster, 2007). Owing to this property, prokaryotic Y-family dPols have been implicated in the appearance of drug-resistant strains of pathogenic bacteria (Cirz *et al.*, 2005; Karpinets *et al.*, 2006).

The structures of archaeal and eukaryotic members of the Y-family are available and the overall structure of these polymerases exhibits coarse overlap and shows the presence of four domains arranged to form the classical right-handed topology (Yang & Woodgate, 2007; Prakash *et al.*, 2005; Washington *et al.*, 2010). In addition to the palm (which houses the catalytic residues), fingers and thumb domains that are observed in other dPols, the members of this family possess an additional large domain called the PAD (polymerase-associated domain) or little finger domain. Although the domain organization shows overlap, the sequences of different members of this family exhibit a high level of divergence and different members of the family are observed to possess unique structural and sequence elements that shape the active site to provide the enzyme with distinctive biochemical properties.

Escherichia coli possesses two members of the Y-family: dPolIV (encoded by the *dinb* gene) and dPolV (encoded by the *umuc* and *umud* genes). These enzymes have been implicated in translesion synthesis as well as in adaptive mutagenesis (Galhardo *et al.*, 2007; Foster, 2007; Walsh *et al.*, 2012; Cirz *et al.*, 2005). dPolIV has been shown to accurately bypass N2 adducts of guanine and has also been shown to be responsible for the appearance of frameshift mutations through single-base deletions (McKenzie *et al.*, 2001; Jarosz *et al.*, 2006). Orthologues of these two enzymes have also been implicated in adaptive mutagenesis in other prokaryotes such as *Bacillus subtilis*, *Pseudomonas aeruginosa* and *P. putida* (Tegova *et al.*, 2004; Sung *et al.*, 2003; Sanders *et al.*, 2006).

In a previous study, we showed that a DinB homologue from *Mycobacterium smegmatis* named MsPolIV is capable of promoting G·T and T·G mismatches and is therefore capable of participating in the initial steps of adaptive mutagenesis (Sharma & Nair, 2012). Here, we describe the structure of MsPolIV in its apo state. The structure shows the presence of thumb, palm and fingers domains, but there is no electron density for the PAD region. SDS–PAGE analysis showed that intact protein had been crystallized. Analysis of the packing of the molecules in the crystals showed the existence of large solvent-filled voids within this crystal form. The PAD region probably exists in a multitude of conformations relative to the rest of the protein. Analytical gel-filtration and DLS experiments showed that MsPolIV undergoes a significant compaction on transition from the apo state to the DNA-bound state. Overall, these observations and comparison with available structures of other members of the Y-family suggest that in the case of these dPols a lack of stabilizing interactions between the PAD/little finger and other domains in the apo state allows this region to exhibit positional heterogeneity. The implications of this structural attribute in facilitating adaptive mutagenesis and translesion bypass and the regulation of these activities are discussed.

Table 1
Data-collection and refinement statistics for MsPolIV.

Values in parentheses are for the highest resolution shell.

	SeMet	Native
Data collection		
Wavelength (Å)	0.97864	1.0
Space group	$P6_522$	$P6_522$
Unit-cell parameters (Å, °)	$a = b = 124.6$, $c = 141.9$, $\alpha = \beta = 90.00$, $\gamma = 120.00$	$a = b = 124.1$, $c = 141.0$, $\alpha = \beta = 90.00$, $\gamma = 120.00$
Resolution (Å)	3.4 (3.52–3.40)	2.6 (2.69–2.60)
R_{merge}†	16.8 (49.5)	9.1 (47.4)
$\langle I/\sigma(I)\rangle$	12.35 (4.1)	23.49 (4.5)
Completeness (%)	100 (100)	100 (100)
Multiplicity	6.8 (7.0)	7.2 (7.3)
Refinement		
Resolution (Å)	50–2.60	
No. of reflections	20354	
R_{work}/R_{free}‡ (%)	21.1/23.8	
No. of atoms		
Protein	1739	
Water	142	
Average B factors (Å^2)		
Protein		
Chain A	36.86	
Water	39.1	
R.m.s. deviations		
Bond lengths (Å)	0.0082	
Bond angles (°)	1.49	

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the integrated intensity of reflection hkl. ‡ $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl} |F_{obs}|$. R_{free} was calculated using 5% of data that were excluded from refinement.

2. Materials and methods

2.1. Structure determination

The purification and crystallization of MsPolIV have been described in detail elsewhere (Sharma & Nair, 2011). Briefly, crystals were obtained at 278 K in 2.4 *M* NaCl, 0.1 *M* sodium/potassium phosphate pH 6.2. Single-wavelength anomalous diffraction (SAD) data were collected to 3.4 Å resolution from an SeMet-derivative crystal on beamline BM14 at the European Synchrotron Radiation Facility (ESRF) along with

a native data set to 2.6 Å resolution. All data were indexed and integrated using *DENZO* and reduced using *SCALEPACK* (Table 1; Otwinowski & Minor, 1997). The positions of five Se atoms of a possible seven were found from the SeMet-derivative data using *SHELXC* (Sheldrick, 2008). *CNS* was used to apply the initial experimental phases (3.6 Å) to the native data and then to extend to 2.6 Å resolution with solvent flattening (Brünger *et al.*, 1998). This yielded a readily interpretable electron-density map, which was used to build an initial model of the protein with *O* (Jones *et al.*, 1991). This was followed by iterative rounds of refinement and model building plus water picking; $R_{\rm free}$ and $R_{\rm cryst}$ converged to final values of 23.8% and 21.5%, respectively. The quality of the chain tracing and side-chain conformations in the final model was confirmed using a composite OMIT map. The geometric and stereochemical parameters of the structure were assessed and validated using *PROCHECK* (Laskowski *et al.*, 1993). The structure was analyzed using *O* and various programs from the *CCP*4 suite (Winn *et al.*, 2011); figures depicting structures were prepared using *PyMOL* (Schrödinger LLC). The structure factors and final refined coordinates were deposited in the PDB as entry 4dez.

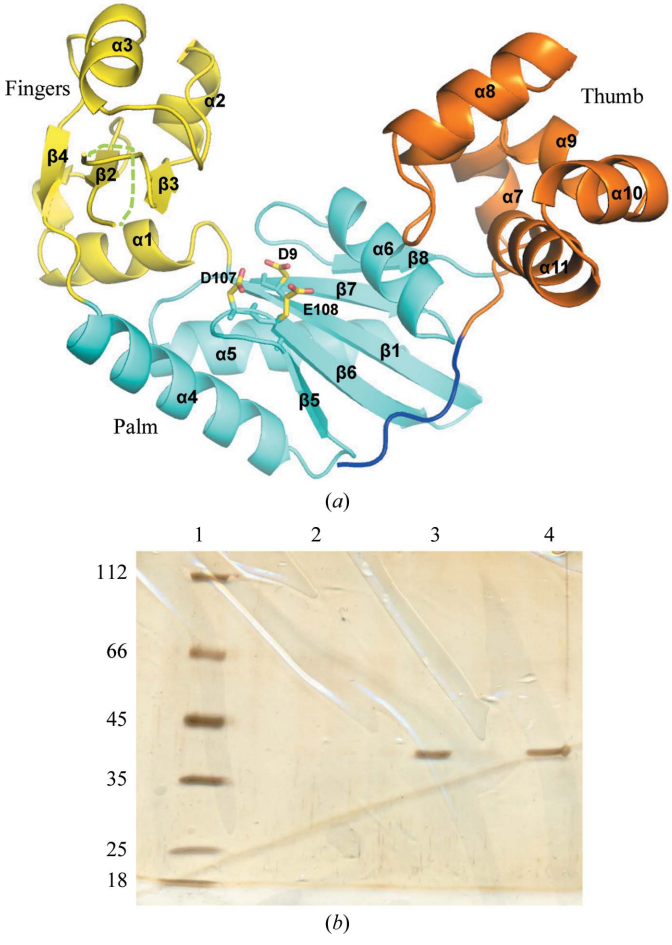

Figure 1
Structure of apo MsPolIV and SDS–PAGE analysis of the crystals. (*a*) The structure of apo MsPolIV is displayed in ribbon representation, with the palm domain coloured cyan, the fingers domain coloured yellow, the thumb domain coloured orange and the linker region coloured dark blue. The secondary-structure elements are labelled. (*b*) Results of SDS–PAGE analysis. Lane 1, molecular-mass marker (labelled in kDa). Lane 2, fourth wash. Lane 3, dissolved crystals. Lane 4, freshly purified MsPolIV. The dissolved crystals and freshly purified protein migrated to a distance corresponding to a molecular mass of 40 kDa, indicating that the intact protein had crystallized.

2.2. SDS–PAGE and packing analysis of crystals

The crystals were picked up using mounted loops (Hampton Research, USA) and washed by serial transfer into four different drops (10 µl) of reservoir solution. Subsequently, the washed crystals were dissolved in 5 µl purified water. The dissolved crystals, the drop corresponding to the fourth wash, freshly purified MsPolIV protein (100 ng) and molecular-mass marker (Fermentas, Canada) were loaded onto a discontinuous SDS–PAGE gel. The gel was subsequently subjected to a silver-staining protocol to detect the presence and positions of protein bands (Merril *et al.*, 1984).

A packing analysis of the apo MsPolIV crystals was carried out using *Coot* (Emsley *et al.*, 2010). The structure of MsPolIV was displayed together with symmetry-related molecules (C^{α} trace) over a radius of 130 Å to view the manner in which the molecules were packed against each other. The coordinates of a number of symmetry-related molecules were written out and were subsequently displayed in *PyMOL* to generate a partial picture of the arrangement of the molecules. The ssDpo4 molecule from PDB entry 3fds (Xing *et al.*, 2009) was aligned with the primary MsPolIV molecule using *PyMOL* to ascertain where the PAD region would be located in the MsPolIV crystal.

2.3. Molecular size-exclusion chromatography

Gel-filtration experiments were performed on a Superdex 200 HR 10/30 column (GE Healthcare). Chromatography was carried out in 25 m*M* Tris buffer pH 8.0 containing 50 m*M* NaCl, 2% glycerol and 2 m*M* DTT. The flow rate was maintained at 0.3 ml min^{-1} and the elution profile was monitored by the absorbance at 280 nm. The void volume ($V_{\rm o}$) was determined using blue dextran and the column was calibrated using the following standard molecular-mass markers (Sigma–Aldrich): thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B_{12} (1.35 kDa). A standard plot of log(molecular mass) *versus* $V_{\rm e}/V_{\rm o}$ was prepared and showed the characteristic linear correlation. The molecular masses of the proteins were calculated from the standard plot of log(molecular mass) against $V_{\rm e}/V_{\rm o}$. The DNA duplex was prepared by annealing an 18-mer template with a 13-mer primer; a 1.2-fold molar excess of the duplex was added to the protein and incubated on ice for 30 min prior to injection. 500 µg MsPolIV was injected into the column in the presence and absence of DNA and the resulting elution profiles were overlaid onto each other.

2.4. Dynamic light scattering

DLS measurements were carried out using a DynaPro 99 unit (Protein Solutions Ltd). All measurements were carried out using a 45 µl cuvette at 298 K. The protein was diluted from the stock solution to a concentration of 1 mg ml^{-1} in buffer consisting of 25 m*M* Tris pH 8.0, 50 m*M* NaCl, 2 m*M* DTT. The buffer and protein solution were filtered using a 0.02 µm filter and the protein solution was centrifuged at 14 000 rev min^{-1} for 45 min immediately before measurements. The data acquired were processed and analyzed using the *DynaLS* software associated with the DynaPro 99 instrument. The duplex DNA utilized for gel filtration was also used for DLS, and the protein and DNA were again maintained in a 1:1.2 molar ratio. The distribution of R_h values for MsPolIV in the presence and absence of DNA were overlaid onto each other.

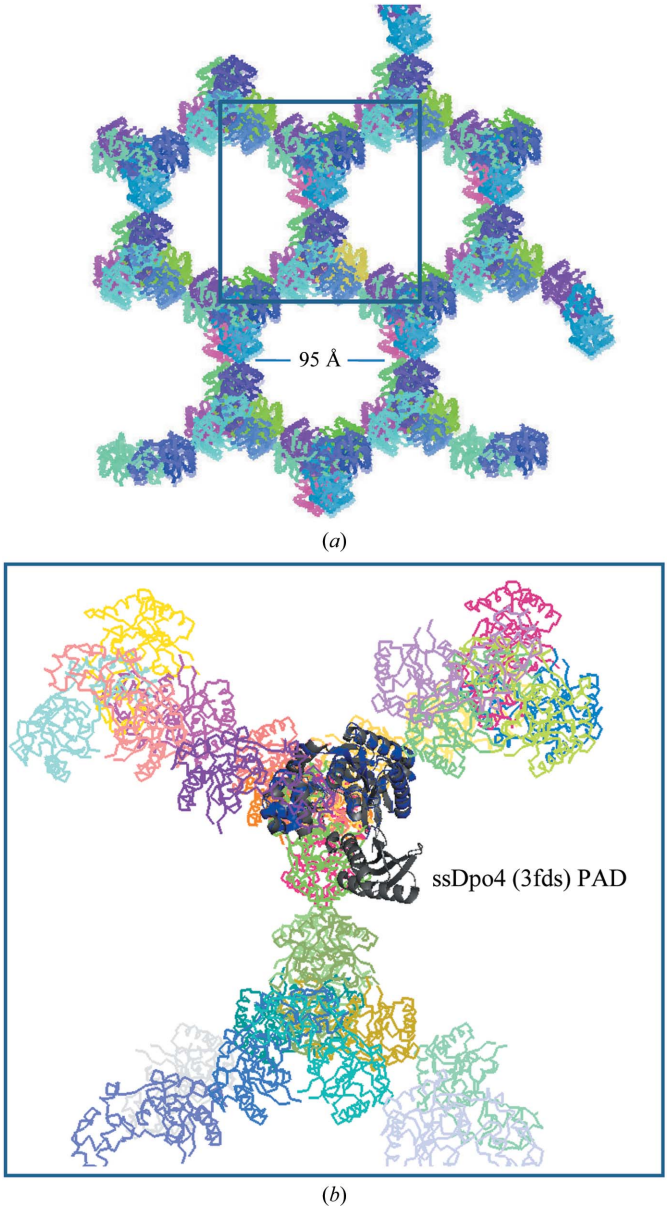

Figure 2
Packing analysis of apo MsPolIV crystals. (*a*) The figure shows the arrangement of MsPolIV molecules within the crystal. The molecules are arranged in such a way that large voids (with a diameter of 95 Å) are present within the crystal. A closer view of the region demarcated by the box is shown in (*b*). (*b*) The ssDpo4 molecule (from PDB entry 3fds) was overlaid over the primary MsPolIV molecule followed by display of symmetry-related molecules. MsPolIV (corresponding to one asymmetric unit) and ssDpo4 are shown in ribbon representation and coloured dark blue and grey, respectively. The figure clearly shows that the PAD region of MsPolIV would be present in the cavities formed within the crystal.

3. Results

3.1. Structure of MsPolIV in the apo state

The asymmetric unit shows the presence of one molecule of MsPolIV. The domain arrangement of MsPolIV resembles that of other Y-family dPols (Fig. 1*a*). The N-terminus forms a β-strand (β1) that is part of the five-stranded β-sheet present in the palm domain and also houses one residue of the catalytic triad (Asp9). This is followed by an α-helix (α1) leading to β-strand β2. The linker between β2 and β3 is disordered, and β3 is followed by two α-helices α2 and α3 leading to β4. β2, β3 and β4 form β-sheet B1 and together with α1, α2 and α3 form the fingers domain (residues 12–79). Next is α-helix α4, which is part of the palm domain. α4 is followed by β-strands β5 and β6, and two of the catalytic residues are located on the linker between the two strands (Asp107 and Glu108). It has previously been shown that mutation of Asp107 to alanine resulted in a loss of DNA polymerase activity (Sharma & Nair, 2012). Asp9, Asp107 and Glu108 are proximal, as are the catalytic residues in other Y-family dPols, and will form the triad that catalyses DNA synthesis. β6 leads to α5, followed by β7, α6 and β8. The N-terminal β-strand β1, along with β5, β6, β7 and β8, forms the five-stranded β-sheet characteristic of the palm region of dPols (80–170). β8 leads to the thumb domain, which is completely α-helical and is made up of α-helices α7, α8, α9, α10 and α11 (171–233). This is followed by the linker that connects the thumb domain to the region that is unique to Y-family dPols and that is known as the polymerase-associated domain (PAD; Trincao *et al.*, 2001). The PAD region exhibits low sequence homology in members of this family, but adopts a similar structure in the form of a four-stranded β-sheet apposed against two α-helices towards the outside of the protein. In the case of MsPolIV, however, no electron density was visible for the PAD region and the structure of the enzyme could only be built up to part of the linker.

3.2. The crystals of MsPolIV consist of intact protein

The lack of electron density for the PAD region suggested that MsPolIV had fragmented during crystallization and that a truncated polypeptide had been crystallized. To ascertain whether this were true, crystals of MsPolIV were washed, dissolved in water and subjected to SDS–PAGE followed by detection of the protein by silver staining (Fig. 1*b*). The stained gel showed the presence of a band corresponding to ~40 kDa for the sample prepared from dissolved crystals. Also, this band migrated to the same extent as a band corresponding to freshly purified MsPolIV protein. These results

clearly showed that the intact protein had been crystallized and that MsPolIV had not fragmented during crystallization.

3.3. Packing analysis of MsPolIV crystals

The fact that the intact protein had crystallized led to the hypothesis that the region corresponding to the PAD was either disordered or existed in a multitude of conformations. Biochemical analysis of the MsPolIV protein had previously shown that it was capable of template-dependent deoxyribonucleotide incorporation and therefore was a dPol (Sharma & Nair, 2012). In the Y-family dPols, the palm domain, which houses the catalytic residues, forms the floor of the DNA-binding groove. The PAD, fingers domain and thumb domain, arranged in a clockwise manner, form the walls surrounding this groove from three sides. In the structures of DNA-bound complexes of Y-family dPols, it is observed that the PAD region inserts into the major groove and forms the majority of stabilizing contacts to orient the substrate DNA in the active site. Since MsPolIV shows functional dPol activity, the PAD region is probably not disordered and adopts the structure that is required to hold the substrate DNA in place for enzymatic activity (Sharma & Nair, 2012). Hence, it is possible that owing to the flexible linker that extends from the thumb domain to the PAD region, the PAD exhibits positional heterogeneity with respect to the rest of the protein. Packing analysis of the MsPolIV unit cell reveals that the molecules are arranged in such a way that there are large solvent-filled cavities within the crystal (Fig. 2*a*). The average diameter of these voids is about 95 Å, while the longest dimension of the PAD region in the closest homologue ssDpo4 is 39 Å. When the structure of ssDpo4 (PDB entry 3fds) was overlaid onto that of one of the monomers in the packing diagram, the PAD region was found to be located within one such cavity (Fig. 2*b*). The linker extends into these cavities and it is expected that the PAD region would be present within these cavities. Since there are no interactions between the PAD region and other domains of the same as well as of symmetry-related molecules to stabilize this region in one orientation, the flexible linker allows this region to occupy multiple positions within the solvent-filled cavity. This accounts for the lack of electron density for the PAD region in the crystal structure.

3.4. MsPolIV exhibits a decrease in hydrodynamic volume upon DNA binding

The structure of apo MsPolIV suggests that the PAD region exhibits positional heterogeneity in the absence of substrate DNA. Since it has been well established that the PAD region of Y-family DNA polymerases docks into the major groove of substrate DNA, it is possible that the PAD region undergoes a transition from positional heterogeneity to a single stable conformation upon DNA binding. Consequently, the PAD will transition from a situation in which it is extended from the rest of the protein in a multitude of positions to a unitary DNA-bound situation in which it is present docked into the major groove. As a result, the protein should exhibit a reduction in hydrodynamic volume on DNA binding, and this change can be monitored by gel filtration and dynamic light scattering (DLS).

The gel-filtration experiments showed that the retention volume for apo MsPolIV was 14.6 ml and that the major peak in the presence of DNA had a retention volume of 15.7 ml (Fig. 1*a*). Based on the calibration plot, the apparent molecular mass of apo MsPolIV would be 50 kDa and that for MsPolIV–DNA would be 40 kDa. The latter value is closer to the molecular mass of the protein (38 kDa) than the former, implying that MsPolIV undergoes a transition from a situation in which the PAD is in an extended conformation to one in which the protein has acquired a globular shape upon binding

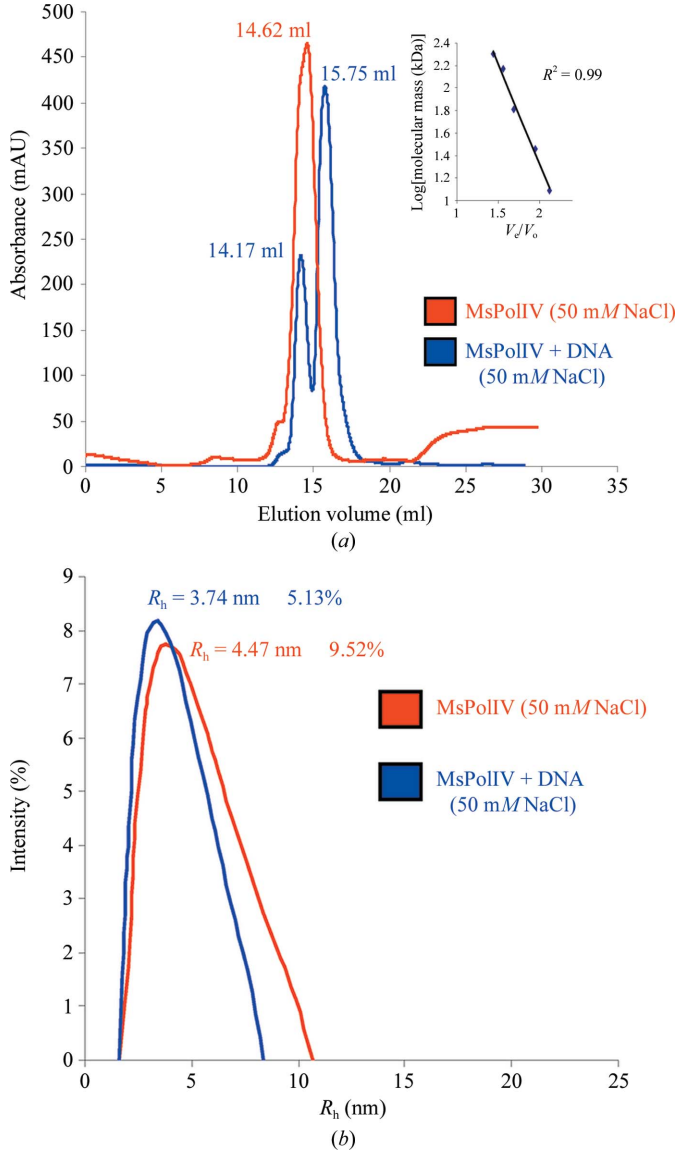

Figure 3
Analytical gel-filtration and dynamic light-scattering (DLS) experiments. (*a*) Analytical gel filtration of MsPolIV in apo and DNA-bound states. The elution profile corresponding to apo MsPolIV is coloured red and that with DNA is coloured blue. The inset shows the calibration curve for the gel-filtration column with standard molecular-mass markers. (*b*) The distribution of R_h values for MsPolIV obtained by DLS in the absence (red) and the presence (blue) of DNA are displayed together with the corresponding statistics.

DNA. Overall, the gel-filtration experiments clearly show that MsPolIV exhibits a higher retention volume in the presence of DNA, which can only be a consequence of a reduction in its hydrodynamic volume. In addition, the DLS experiments show that the hydrodynamic radii in the absence and presence of DNA are 4.47 and 3.74 n*M*, respectively (Fig. 3*b*). Thus, the results of both the gel-filtration and the DLS experiments suggest that MsPolIV undergoes compaction in the presence of DNA, resulting in a decrease in its hydrodynamic volume. This change in the hydrodynamic volume on DNA binding is consistent with the hypothesis that the PAD region transits from positional plurality to a singular orientation relative to the other domains of the protein.

4. Discussion

To our knowledge, the structure of MsPolIV is the first structure of a prokaryotic member of the Y-family of dPols. The structures of Polη (*Saccharomyces cerevisiae*), Dbh (*Sulfolobus acidocaldarius*), ssDpo4 (*S. solfataricus*) and Polκ (*Homo sapiens*) have been reported in the apo state (Silvian *et al.*, 2001; Trincao *et al.*, 2001; Uljon *et al.*, 2004). The structures of these four enzymes have also been reported in their functional states (in complex with DNA and incoming nucleotide; Wilson & Pata, 2008; Ling *et al.*, 2001; Silverstein *et al.*, 2010; Lone *et al.*, 2007). In the case of Dbh, structures of the ternary complex have been solved with DNA containing a looped-out nucleotide in the template strand and these structures provide a basis for the appearance of frameshift mutations (Wilson & Pata, 2008). A similar study has also recently been carried out in the case of ssDpo4 (Wu *et al.*, 2011). All four apo structures show the presence of the PAD region, but in orientations that differ to varying degrees compared with that observed in the structures of the ternary complexes (with DNA and nucleotide; Fig. 4). In the case of Polκ, the PAD region in the apo structure is located in two different positions in two molecules in the asymmetric unit (Uljon *et al.*, 2004). In the more extreme case the PAD region is present dorsal to the palm and moves through 50 Å to insert into the major groove of DNA. In the case of ssDpo4 the PAD region is present at an angle across the DNA-binding groove and shows a substantial movement to allow DNA binding. Additionally, in the structure of ssDpo4 bound to the processivity clamp the PAD is oriented in an extended conformation relative to the rest of the molecule (Xing *et al.*, 2009). However, in the case of yeast Polη only a smaller movement of the PAD region is required to facilitate DNA

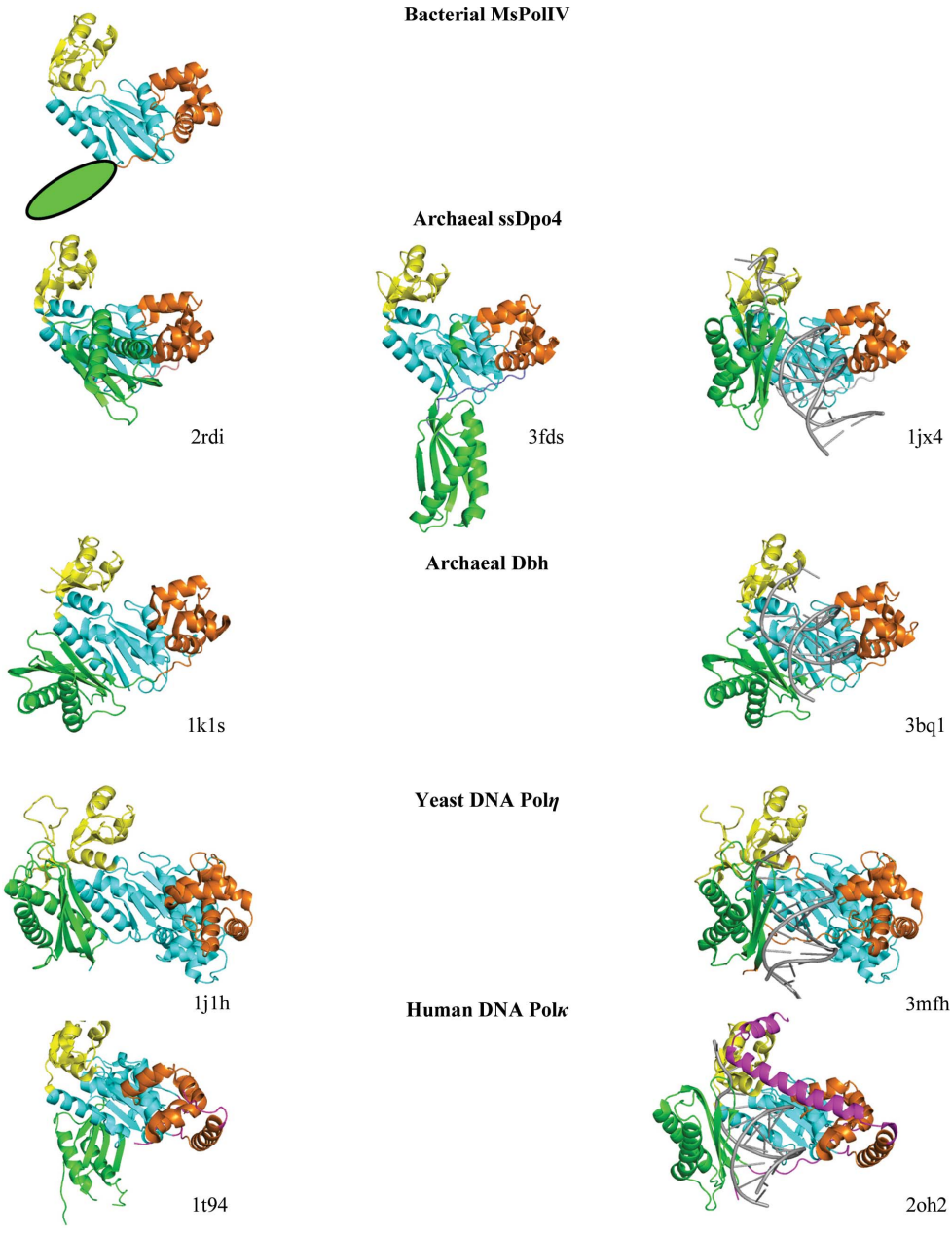

Figure 4
The PAD region in Y-family dPols can adopt different orientations. The structures of apo MsPolIV, ssDpo4 in its apo (PDB entry 2rdi), clamp-bound (PDB entry 3fds) and DNA-bound (PDB entry 1jx4) states, Dbh in its apo (PDB entry 1k1s) and DNA-bound (PDB entry 3bq1) states, yeast Polη in its apo (1jih) and DNA-bound (PDB entry 3mfh) states and Polκ in its apo (PDB entry 1t94) and DNA-bound (2oh2) states are displayed. The structures show that the PAD region can adopt a variety of orientations relative to the rest of the protein.

binding, involving a rotation of 13° around the helical axis of DNA (Silverstein *et al.*, 2010). The PAD in this case adopts an orientation that is very close to that in the ternary complex, possibly because of a close interaction between the fingers and palm domains through a loop region (between β-strands 5 and 6) in the fingers that is unique to yeast Polη (Trincao *et al.*, 2001). In the case of Dbh, the PAD also shows a slight inwards movement to bind DNA. These structures were determined with substrate DNA containing a looped-out nucleotide in the template strand in the double-stranded region; to accommodate this nucleotide, the PAD region does not insert into the major groove to the extent that is observed in the case of the close homologue ssDpo4. For ssDpo4, structures are available in complex with normal DNA and with DNA containing a looped-out nucleotide in the template strand of the double-stranded region (Wu *et al.*, 2011). There is no difference in the orientation of the ssDpo4 PAD when bound to these different substrate duplexes. As opposed to the two states adopted by the PAD region in the case of the other proteins, the PAD region in MsPolIV exhibits conformational heterogeneity in its apo state and provides further information regarding the range of possible states that it can occupy. Comparison of these structures implies that the lack of tertiary interactions between the fingers domain and the PAD region can allow the PAD to adopt a multitude of orientations relative to the rest of the protein. It is therefore possible that the single nonstandard orientation adopted by the PAD region in the case of ssDpo4 and Polκ is a consequence of the formation of stabilizing interactions with neighbouring molecules within the crystal lattice. The structures of DNA polymerases iota (hPolι) and eta (hPolη) from *H. sapiens* are available in their functional states (Nair *et al.*, 2004; Biertümpfel *et al.*, 2010). In the case of hPolι there are no interactions between the fingers and PAD when the enzyme is bound to DNA. Therefore, it is expected that the PAD would also have the freedom to exhibit positional heterogeneity in the apo state in this enzyme. In the case of hPolη a number of interactions are formed between the residues of a substructure present between the fingers and the palm domains with those from an α-helix in the PAD region. Hence, it is possible that, like the yeast homologue, hPolη will only exhibit a small reorientation of the PAD region on DNA binding.

Normally, DNA adopts structural characteristics that are collectively classified as that of B-DNA. The width of the DNA double helix is constant for both G·C and A·T pairs, with a mean C1′–C1′ distance of 10.8 Å. This property is utilized by replicative dPols to select and incorporate the correct incoming nucleotide opposite a particular template. These polymerases are known to utilize a shape filter and it has been observed that perturbing hydrogen bonding between the bases does not compromise the ability of the polymerase to select the correct nucleotide (Sintim & Kool, 2006; Delaney *et al.*, 2003). However, variations in the shape of the template nucleotide occur when the nucleotides are damaged and compromise their ability to carry out template-dependent nucleotide incorporation. Members of the Y-family are known to achieve translesion bypass by accommodating nucleotides opposite damaged templates as they possess specialized active sites that can accommodate nonstandard widths of the nascent base pair. During this process these enzymes should also possess the ability to stabilize deviations in the width of the newly synthesized double-helical DNA as it exits the molecule. The structure and allied data clearly show that in the case of MsPolIV the PAD region exhibits conformational heterogeneity in the absence of DNA. The PAD has been shown to play a critical role in determining the biochemical properties of these enzymes (Boudsocq *et al.*, 2004). It is possible that the flexibility associated with the PAD region is a consequence of the need to reshape the active site in order to accommodate damaged nucleotides both at the templating position and within newly synthesized double-stranded DNA extruding out of the molecule. In addition, this property would also enhance the ability of the enzymes to accommodate mismatches and base deletions and therefore promote adaptive mutagenesis. Xing and coworkers observed that the orientation of the PAD of ssDpo4 when bound to the processivity clamp is very different from that observed in the case of apo ssDpo4 and the ternary complex of the enzyme (Xing *et al.*, 2009). This led them to propose a scheme in which the PAD occupies a nonfunctional orientation on the clamp until it is recruited to the replication fork to carry out lesion bypass.

The comparison of apo and DNA-bound structures in the cases of Polκ, Polη and Dbh and of apo, DNA-bound and clamp-bound structures of ssDpo4 shows that the PAD exists in different orientations in these different states. The present structure of MsPolIV implies that the PAD adopts a multitude of orientations in the absence of DNA. Together, all these structures suggest that the PAD region is endowed with positional flexibility to allow it to accommodate non-Watson–Crick base pairs in the enzyme active site and in dsDNA extruding out of the DNA exit tunnel. This structural attribute allows these enzymes to achieve translesion bypass and to participate in adaptive mutagenesis through the promotion of mismatches and base deletions. Additionally, the positional flexibility could also allow regulation of enzyme activity until necessary and this aspect would be important to prevent adventitious error-prone DNA synthesis. Overall, the unique design of Y-family dPols involving a large functionally critical domain connected by a flexible linker to the rest of the protein allows these class of enzymes to carry out their multiple functions.

5. Conclusion

The structure of the prokaryotic MsPolIV and allied experiments suggest that the PAD region is endowed with the ability to adopt multiple orientations. This positional flexibility with respect to the rest of the protein might be a general feature of this class of enzymes. This attribute probably allows these enzymes to accommodate alterations in the width of the DNA double helix during and post DNA synthesis to facilitate their twin functions of translesion bypass and adaptive mutagenesis.

DTN thanks the X-ray diffraction facility located in the Molecular Biophysics Unit of the IISc [funded by the Department of Biotechnology (DBT) and the Department of Science and Technology (DST), Government of India] for facilitating screening and data collection. DTN acknowledges the help rendered by Dr Hassan Belrhali (ESRF) during data collection on the BM14 beamline at the ESRF. The authors thank Professor Jayant B. Udgaonkar for access to the Dynapro 99 unit for DLS studies. This research was funded by NCBS-TIFR. DTN is a recipient of a Ramanujan fellowship from DST, and AS and VS are recipients of senior and junior research fellowships from Council of Scientific and Industrial Research (Government of India), respectively. Data collection on the BM14 beamline at the ESRF, Grenoble, France was funded by the BM14 project, which is a collaboration between DBT (Government of India), EMBL and the ESRF.

References

Biertümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maiques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F. & Yang, W. (2010). *Nature (London)*, **465**, 1044–1048.

Boudsocq, F., Kokoska, R. J., Plosky, B. S., Vaisman, A., Ling, H., Kunkel, T. A., Yang, W. & Woodgate, R. (2004). *J. Biol. Chem.* **279**, 32932–32940.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Cirz, R. T., Chin, J. K., Andes, D. R., de Crécy-Lagard, V., Craig, W. A. & Romesberg, F. E. (2005). *PLoS Biol.* **3**, e176.

Delaney, J. C., Henderson, P. T., Helquist, S. A., Morales, J. C., Essigmann, J. M. & Kool, E. T. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 4469–4473.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Foster, P. L. (2007). *Crit. Rev. Biochem. Mol. Biol.* **42**, 373–397.

Galhardo, R. S., Hastings, P. J. & Rosenberg, S. M. (2007). *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435.

Goodman, M. F. (2002). *Annu. Rev. Biochem.* **71**, 17–50.

Indiani, C., McInerney, P., Georgescu, R., Goodman, M. F. & O'Donnell, M. (2005). *Mol. Cell*, **19**, 805–815.

Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M. & Walker, G. C. (2006). *Nature (London)*, **439**, 225–228.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Karpinets, T. V., Greenwood, D. J., Pogribny, I. P. & Samatova, N. F. (2006). *Curr. Genomics*, **7**, 481–496.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. (2001). *Cell*, **107**, 91–102.

Lone, S., Townson, S. A., Uljon, S. N., Johnson, R. E., Brahma, A., Nair, D. T., Prakash, S., Prakash, L. & Aggarwal, A. K. (2007). *Mol. Cell*, **25**, 601–614.

McKenzie, G. J., Lee, P. L., Lombardo, M. J., Hastings, P. J. & Rosenberg, S. M. (2001). *Mol. Cell*, **7**, 571–579.

Merril, C. R., Goldman, D. & Van Keuren, M. L. (1984). *Methods Enzymol.* **104**, 441–447.

Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L. & Aggarwal, A. K. (2004). *Nature (London)*, **430**, 377–380.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Prakash, S., Johnson, R. E. & Prakash, L. (2005). *Annu. Rev. Biochem.* **74**, 317–353.

Sanders, L. H., Rockel, A., Lu, H., Wozniak, D. J. & Sutton, M. D. (2006). *J. Bacteriol.* **188**, 8573–8585.

Sharma, A. & Nair, D. T. (2011). *Acta Cryst.* F**67**, 812–816.

Sharma, A. & Nair, D. T. (2012). *J. Nucleic Acids*, **2012**, 285481.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Silverstein, T. D., Johnson, R. E., Jain, R., Prakash, L., Prakash, S. & Aggarwal, A. K. (2010). *Nature (London)*, **465**, 1039–1043.

Silvian, L. F., Toth, E. A., Pham, P., Goodman, M. F. & Ellenberger, T. (2001). *Nature Struct. Biol.* **8**, 984–989.

Sintim, H. O. & Kool, E. T. (2006). *Angew. Chem. Int. Ed. Engl.* **45**, 1974–1979.

Sung, H.-M., Yeamans, G., Ross, C. A. & Yasbin, R. E. (2003). *J. Bacteriol.* **185**, 2153–2160.

Sutton, M. D. (2010). *Biochim. Biophys. Acta*, **1804**, 1167–1179.

Tegova, R., Tover, A., Tarassova, K., Tark, M. & Kivisaar, M. (2004). *J. Bacteriol.* **186**, 2735–2744.

Trincao, J., Johnson, R. E., Escalante, C. R., Prakash, S., Prakash, L. & Aggarwal, A. K. (2001). *Mol. Cell*, **8**, 417–426.

Uljon, S. N., Johnson, R. E., Edwards, T. A., Prakash, S., Prakash, L. & Aggarwal, A. K. (2004). *Structure*, **12**, 1395–1404.

Walsh, J. M., Hawver, L. A. & Beuning, P. J. (2012). *Front. Biosci.* **17**, 3164–3182.

Washington, M. T., Carlson, K. D., Freudenthal, B. D. & Pryor, J. M. (2010). *Biochim. Biophys. Acta*, **1804**, 1113–1123.

Wilson, R. C. & Pata, J. D. (2008). *Mol. Cell*, **29**, 767–779.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Wu, Y., Wilson, R. C. & Pata, J. D. (2011). *J. Bacteriol.* **193**, 2630–2636.

Xing, G., Kirouac, K., Shin, Y. J., Bell, S. D. & Ling, H. (2009). *Mol. Microbiol.* **71**, 678–691.

Yang, W. & Woodgate, R. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 15591–15598.