Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the β -clamp

Karen A. Bunting, S. Mark Roe and Laurence H. Pearl¹

Section of Structural Biology, The Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

¹Corresponding author
e-mail: Laurence.Pearl@icr.ac.uk

Y-family DNA polymerases can extend primer strands across template strand lesions that stall replicative polymerases. The poor processivity and fidelity of these enzymes, key to their biological role, requires that their access to the primer–template junction is both facilitated and regulated in order to minimize mutations. These features are believed to be provided by interaction with processivity factors, β -clamp or proliferating cell nuclear antigen (PCNA), which are also essential for the function of replicative DNA polymerases. The basis for this interaction is revealed by the crystal structure of the complex between the ‘little finger’ domain of the Y-family DNA polymerase Pol IV and the β -clamp processivity factor, both from *Escherichia coli*. The main interaction involves a C-terminal peptide of Pol IV, and is similar to interactions seen between isolated peptides and other processivity factors. However, this first structure of an entire domain of a binding partner with an assembled clamp reveals a substantial secondary interface, which maintains the polymerase in an inactive orientation, and may regulate the switch between replicative and Y-family DNA polymerases in response to a template strand lesion.

Keywords: β -clamp processivity factor/DNA polymerase IV/Y-family

Introduction

In recent years, the number of characterized DNA polymerases has increased rapidly, with the majority of these novel polymerases belonging to the newly designated Y-family of DNA polymerases (Ohmori *et al.*, 2001). These polymerases have varying functions in bypassing a wide range of DNA lesions that would cause the main replicative polymerase to stall, with subsequent collapse of the replication fork. In order to enable this bypass function, these enzymes all display low fidelity of synthesis, poor processivity and have no 3′–5′ proofreading activity (Friedberg *et al.*, 2001). Depending on both the polymerase involved and the nature of the lesion, bypass synthesis may be error-free, or error-prone with incorporation of nucleotides not directed by the damaged template sequence. Y-family polymerases share little sequence similarity with classical DNA polymerases, although recent structural studies indicate a similar overall

architecture (Friedberg *et al.*, 2001). However, Y-family enzymes possess an additional novel domain involved in DNA contacts, variously termed the ‘little finger’ (LF), ‘wrist’ or polymerase-associated domain (Ling *et al.*, 2001; Silvan *et al.*, 2001; Trincão *et al.*, 2001).

Escherichia coli possesses two Y-family members, Pol IV (DinB) and Pol V (UmuD′₂C) (Tang *et al.*, 1999; Wagner *et al.*, 1999). Pol V can bypass a variety of lesions (Tang *et al.*, 2000) and, prior to the demonstration of its intrinsic polymerase activity, was known to be necessary for damage-induced mutagenesis as part of the SOS response (Tang *et al.*, 1999). Pol IV is less well characterized but it is also induced as part of the SOS response, is necessary for adaptive mutagenesis (McKenzie *et al.*, 2001) and is involved in adduct bypass (Napolitano *et al.*, 2000).

Four Y-family DNA polymerases have been identified in humans. Mutations in Pol η have been shown to result in the autosomal recessive xeroderma pigmentosum variant (XPV) disorder, which renders the sufferer hypersensitive to UV exposure, resulting in a high risk of developing skin cancer (van Steeg and Kraemer, 1999). Pol η carries out the error-free bypass of *cis-syn* cyclobutane dimers, a common lesion caused by UV exposure (Masutani *et al.*, 1999). In addition, it has been shown to bypass a variety of other lesions, with varying degrees of fidelity (Haracska *et al.*, 2000, 2001b). The *in vivo* roles of Pol ι and Pol κ currently are not well defined, but could potentially be involved in cancer predisposition, being located in regions of the chromosome implicated in several cancer types (Gerlach *et al.*, 1999; McDonald *et al.*, 1999). Pol ι has been shown to act in concert with another translesion polymerase, Pol ζ , by inserting nucleotides opposite a highly distorting 6–4 thymine–thymine photoproduct, permitting extension by Pol ζ (Johnson *et al.*, 2000). Pol κ , the human counterpart of Pol IV, performs bypass of a benzo[a]pyrene adduct (Zhang *et al.*, 2000) and may also cause frameshift events in the cell (Ohashi *et al.*, 2000), and has been shown recently to be upregulated in certain lung cancers (O-Wang, 2001). The fourth human Y-family member, Rev1, possesses deoxycytidyl transferase activity, incorporates dCMP opposite abasic sites (Lin *et al.*, 1999) and is necessary for UV-induced mutagenesis (Gibbs *et al.*, 2000).

Translesion DNA polymerases synthesize DNA with a fidelity several orders of magnitude lower than the main replicative polymerases. It has been suggested that the active sites of these enzymes would be more flexible than those of replicative polymerases (Friedberg *et al.*, 2002; Kunkel *et al.*, 2003). Whilst replicative polymerases employ a mechanism whereby adherence to geometric constraints assists in correct base pairing at the active site (Goodman, 1997), a more flexible active site would permit bypass of distorting lesions, with a consequent loss of

fidelity (Ling *et al.*, 2001). By their nature, these Y-family polymerases must be strictly regulated, in order to be available at the replication fork to prevent obstruction of replication while minimizing the introduction of mutations due to their inherently low fidelity.

It has been suggested that processivity factors [the prokaryotic β -clamp and the archaeal/eukaryotic proliferating cell nuclear antigen (PCNA)] are involved in such regulatory processes as they theoretically are able to bind a number of interacting partners within the replisome (Hingorani and O'Donnell, 2000). These toroidal structures are loaded onto DNA, encircling it to act as a sliding clamp, and vastly increase the processivity of the replicative polymerases (Kelman and O'Donnell, 1995). Interactions have been demonstrated between all five characterized *E.coli* DNA polymerases and the β -clamp (Hughes *et al.*, 1991; Naktinis *et al.*, 1996; Wagner *et al.*, 2000; Lopez de Saro and O'Donnell, 2001; Sutton *et al.*, 2001), and at least three human Y-family members have been shown to interact with PCNA (Haracska *et al.*, 2001a, 2002). The bypass and mutagenesis activities of *E.coli* Pol IV in particular have been shown to be dependent upon the clamp both *in vitro* and *in vivo*, with an essential interaction site provided by the five residues at the extreme C-terminus of the polymerase (Lenne-Samuel *et al.*, 2002).

In order to understand the role of clamp processivity factors in recruiting and regulating translesion DNA polymerases, we have determined the structure of the complex between the LF domain of Pol IV and the β -clamp processivity factor. The structure reveals a bipartite interaction in which the polymerase is tethered to the processivity factor via a C-terminal peptide, combined with a substantial protein–protein interface that may play a role in regulating Pol IV activity.

Results

A complex between the LF domain of *E.coli* Pol IV (Pol IV-LF) and the *E.coli* Pol III β -subunit processivity clamp (β -clamp) was assembled in solution from the recombinant proteins, purified and crystallized (see Materials and methods). Phasing was achieved by molecular replacement using the structure of the isolated β -clamp as a search model. The complex crystallized with the β -clamp dimer and two LF domains in the asymmetric unit. The two crystallographically independent copies of the LF domain are bound to opposite edges of the dimeric β -clamp ring and are related by approximate non-crystallographic 2-fold symmetry (Figure 1A). The two LF domains in the complex make different lattice contacts in the crystal. This suggests strongly that their virtually identical orientation with respect to the β -clamp is determined by the stability of that particular interaction with the clamp and is not an artefact of crystal packing. The root mean square deviation in the C α position between the two halves of the β -clamp is 0.33 Å, and 0.52 Å between the two attached LFs.

Structure of the LF domain

Despite the lack of sequence conservation, secondary structure predictions (Ling *et al.*, 2001) suggest that an LF domain is present in all Y-family DNA polymerases. The

structure of the ternary complex of the archaeal Dpo4 enzyme shows the LF domain to be directly involved in DNA binding, and biochemical analysis suggests that it serves to increase the processivity of the polymerase (Ling *et al.*, 2001). In view of the considerable divergence in the amino acid sequences of LF domains from different Y-family polymerases, it has been suggested that the LF domain plays an important role in adaptation of the individual polymerases to bypass specific types of lesions (Ling *et al.*, 2001).

The Pol IV-LF has the $\beta\alpha\beta\beta\alpha\beta$ architecture seen in other LF domains (Figure 1B), and superposition with LF domains from the archaeal Dbh (Silvian *et al.*, 2001) and Dpo4 enzymes (Ling *et al.*, 2001), and from yeast Pol η (Trincao *et al.*, 2001) reveals a considerable degree of structural conservation despite the lack of sequence similarity. Given that the structure of Dpo4-LF is taken from the ternary complex, it seems unlikely that any substantial conformational changes take place within the LF domain upon DNA binding. The region containing the β -sheet, which contacts the DNA (Ling *et al.*, 2001), is most structurally conserved between the four examples. The greatest deviations are seen in yeast Pol η -LF, mainly in the C-terminal end of α 1, where the helix is extended, and in the loop connecting β 3 and α 2. This enzyme belongs to a subfamily different from the other known structures, the Rad30 group (Ohmori *et al.*, 2001), although the bypass mechanism of Dpo4 may be more similar to that of this subfamily (Ling *et al.*, 2001).

Structure of the β -clamp–Pol IV-LF complex

The interface between the Pol IV-LF and the β -clamp can be partitioned into two components: a peptide–protein interaction involving the extended C-terminal tail of Pol IV-LF (residues 340–351) and a hydrophobic channel on the surface of the β -clamp, and a protein–protein interaction between surface loops of the Pol IV-LF core and the edge of the β -clamp ring at the juxtaposition of the two monomers.

Peptide–protein interface

Previous studies had shown that a hydrophobic cluster at the extreme C-terminus of Pol IV-LF (346-QLVLGL-351) is required for *in vivo* function in both spontaneous and induced mutagenesis (Lenne-Samuel *et al.*, 2002). The sequence in this region of Pol IV resembles the consensus β -clamp-binding motif (QL[S/D]LF) identified by bioinformatics analysis of eubacterial β -clamp-binding proteins (Dalrymple *et al.*, 2001), but with some potentially significant differences. In this present structure, the hydrophobic residues comprising the majority of the clamp-binding motif in Pol IV lie in an extended depression on the surface of the clamp, overlying the second and third subdomains of each β -clamp subunit (Figure 2A). At the C-terminus of PolIV-LF, Leu349 and Leu351 are buried together in a pocket lined by the side chains of Leu155, Thr172, Leu177, Pro242, Val247, Val360 and Met362 from the β -clamp (Figure 2B). The intervening Gly350 in the QLVLGL motif adopts a very tight conformation that brings the two leucine side chains into contact, and the LGL tripeptide in Pol IV is probably the equivalent of the LF dipeptide of the consensus motif (Dalrymple *et al.*, 2001). The α -carboxyl of Leu351 is

Fig. 1. β -clamp–Pol IV–LF complex. (A) Secondary structure cartoon of the β -clamp dimer (magenta and yellow) with a Pol IV LF domain bound on each side, superimposed on the solvent-accessible surface of the complex. The two halves of the β -clamp and the two LF domains are related by a non-crystallographic 2-fold axis running perpendicular to the plane of view. [This and all other molecular graphics were generated with PyMOL (DeLano Scientific LLC, www.delanoscientific.com).] (B) Superimposition of the LF domain of Pol IV (red tube) and the equivalent structures from *S.solfaraticus* Dpo4 (cyan), *S.solfataricus* Dbh (yellow) and yeast Pol η (green). (C) As (B) but rotated 90° around the horizontal.

surface exposed and involved in an electrostatic interaction with Arg152 from the clamp. The side chains of the preceding hydrophobic residues in the Pol IV C-terminal motif are more exposed, with Val348 in particular making no interactions, while Leu347 packs against a hydrophobic depression formed by the side chains of Val344, Met362, Pro363 and Arg365 from the clamp. The side chain of Gln346, which is highly conserved in the consensus clamp-binding motif, binds in a pocket between the side chains of His175, Asn320, Tyr323 and Met364 of the clamp, and hydrogen-bonds to the peptide carbonyl of Met362. Upstream of Gln346, the side chains of Glu344 and Arg345 are exposed and make no significant interactions with the clamp, whereas Met343 is again buried in a pocket in the clamp formed by Gly280, Val281, Arg282, Gly318, Met364 and Leu366. Interactions between the Pol IV C-terminal peptide and the clamp become progressively less intimate upstream of Met343. In addition to side chain interactions, a short segment of antiparallel β -sheet is formed between the peptide back-

bones of Arg345–Leu347 from Pol IV and Pro363–Arg365 from the β -clamp. The structure of the Pol IV C-terminus from Pro341 to Leu351 involved in this interface appears to be entirely dictated by its interactions with the β -clamp.

Protein–protein interface

In addition to the interaction involving the C-terminal peptide, the Pol IV–LF makes a substantial protein–protein interaction with the β -clamp, involving exposed residues from the loops connecting consecutive secondary structural elements of the LF domain packing against the surface of the eight-stranded β -sheet formed between the N-terminal domain of one clamp subunit and the C-terminal domain of the other (Figure 3A). The interface centres on the side chain of Leu98 of the clamp which inserts as a ‘peg’ into a hydrophobic depression formed by the side chains of Val303, Trp304 and Pro305 of Pol IV. This core interaction is reinforced by a plethora of ion pairs, and direct and solvent-bridged polar interactions

Fig. 2. Binding of Pol IV C-terminal peptide to the β -clamp. (A) The C-terminal peptide of Pol IV from residues 341 to 351 (stick model) binds into a channel on the surface of the β -clamp, with the side chains of Met343, Leu347, Leu349 and Leu351 inserted into hydrophobic pockets in the base of the channel. The molecular surface of the β -clamp is coloured according to the residue type: acidic, red; basic, blue; polar, orange; hydrophobic, green. (B) View perpendicular to (A) showing residues from the β -clamp involved in interaction with the Pol IV C-terminus.

between the residues from each protein, that surround the hydrophobic core of the interaction (Figure 3B). The interactions in this protein–protein interface with Pol IV-LF primarily involve residues from both β -clamp subunits, so that the combination of the two interfaces would serve to reinforce the dimer–dimer interface in the assembled β -clamp. Overall, the binding of each Pol IV-LF domain to the β -clamp buries $\sim 3160 \text{ \AA}^2$ of molecular surface area, of which the protein–protein interface contributes $\sim 2250 \text{ \AA}^2$.

Comparison with the β -clamp– δ -subunit complex

Although the structure presented here is the first description of a complex between the assembled β -clamp and a partner protein, the structure of the complex between a β -clamp subunit and the δ -subunit of the clamp loader complex has been reported previously (Jeruzalmi *et al.*,

2001). Binding of the δ -subunit to the β -clamp opens one of the monomer–monomer interfaces to permit loading onto DNA. The clamp loader δ -subunit has two main points of contact on the β -clamp, one of which coincides with the surface groove occupied by the C-terminal peptide of Pol IV in the present structure (Figure 4A). This interaction involves a linear peptide motif (69-QAMSLF-74), which makes interactions with the β -clamp extremely similar to the C-terminal peptide of Pol IV. Thus, the side chains of 73-LF-74 from the clamp loader δ -subunit occupy the same pocket as the side chains of 349-LGL-351, while the preceding three residues of the clamp loader, Ala70, Met71 and Ser72, almost precisely overlay the positions of Gln346, Leu347 and Val348 from Pol IV (Figure 4B). The main chain from the clamp loader in this region also forms an antiparallel β -sheet with the C-terminal strand of the β -clamp.

Fig. 3. Protein-protein interface. (A) Stereo pair of the Pol IV LF domain (blue cartoon) binding at the interface between the two monomers of the β -clamp (yellow/magenta surfaces). Pol IV residues involved in direct interaction with the β -clamp are shown as sticks. The C-terminal peptide (see text) runs off to the left in this view. (B) Close-up stereo pair showing details of the protein-protein interface between Pol IV-LF and the β -clamp. The core of the interaction is provided by the insertion of a hydrophobic 'peg' provided by the side chain of Leu98 of the β -clamp into a hydrophobic 'hole' formed by the side chains of Val303, Trp304 and Pro305 of Pol IV. This is surrounded by a ring of polar and electrostatic and solvent-bridged interactions between Pol IV and both monomers of the β -clamp. The view is from the opposite face of the β -clamp ring to that in (A).

Comparison with PCNA-peptide complexes

Although there is very little sequence similarity between them, bacterial, archaeal, phage and eukaryotic clamp processivity factors show remarkable similarity in their domain architecture and superstructural assembly (Krishna *et al.*, 1994; Moarefi *et al.*, 2000; Matsumiya *et al.*, 2001) and in the location of the binding sites they provide for binding peptide motifs of partner and associated proteins. In all cases, the partner protein peptide forms a short antiparallel β -sheet with the C-terminal strand of the clamp subunit, and inserts one or more hydrophobic/aromatic residues into a pocket formed between the outer surface of the eight-stranded β -sheet and the loop segment connecting the C-terminal and preceding domains (Figure 5) (Gulbis *et al.*, 1996; Shamoo and Steitz, 1999; Matsumiya *et al.*, 2002). The most significant difference occurs between the β -clamp complex described here, where the binding motif of Pol IV

maintains a more or less extended conformation, and the archaeal, phage and eukaryotic clamps, where the bound peptide forms a 3_{10} -helical turn as it approaches the interdomain connecting loop of the clamp. With the longer bound peptides, such as p21^{Cip/Waf} bound to PCNA, the peptide continues into an extended β -sheet with the interdomain connector loop (Gulbis *et al.*, 1996). In Pol IV, the β -clamp-binding motif is the C-terminus of the protein, so no extended β -sheet interaction with the inter-domain connector loop is possible, but other β -clamp partners have non-terminal motifs that could in principle make such interactions. However, the β -clamp structure has a unique insertion relative to the other clamps, of ~ 10 residues between the end of the first helix and beginning of the second strand in its middle domain. This structure interacts with the inter-domain connector loop and, unless there is a major change in conformation on binding, would prevent binding of an extended peptide. The observed position of

Fig. 4. Comparison of β -clamp interactions with Pol IV-LF and the clamp loader δ -subunit. **(A)** Superposition of the clamp loader δ -subunit- β -clamp complex (pink, green) with one half of the Pol IV-LF- β -clamp complex (blue, gold) based on a least squares fit of the C α positions in the common β -clamp subunit component. The β -clamp-binding segment of the δ -subunit is highlighted in red. **(B)** Close-up of the superposed β -clamp-binding peptide segments of the clamp loader δ -subunit (pink carbons) and Pol IV-LF (blue carbons). There is a one-to-one correspondence in conformation and composition except for the extreme C-terminus of Pol IV, where the LGL tri-peptide replaces the LF dipeptide of the δ -subunit.

the C-terminus of Pol IV probably therefore represents the C-terminal limit for extended peptide interaction with the β -clamp, so that any part of a partner protein beyond the binding motif would be directed away from the clamp, as it is in the clamp loader δ -subunit complex.

Discussion

Current structural models for the interaction of assembled processivity clamps with the partner proteins whose DNA interactions they facilitate (Hosfield *et al.*, 1998; Shamoo and Steitz, 1999; Pages and Fuchs, 2002) are largely based on complexes between clamps and otherwise unstructured peptide motifs. In the absence of data to the contrary, it has been reasonably assumed that these peptide motifs provide

an anchor, linked flexibly to the rest of the partner protein, and with little or no specific interaction between the globular part of the binding partner and the surface of the clamp (Shamoo and Steitz, 1999). This is clearly not the case in the interaction of Pol IV and the β -clamp described here, where more than two-thirds of the interaction surface between the two proteins is contributed by the surface of the globular LF domain. Analysis of β -clamp and Pol IV-LF sequences from widely divergent bacterial species shows some variation in the identity of the interacting residues, but with a general conservation of the hydrophobic nature of the 'patch' at the core of the protein-protein interface, and of several of the surrounding polar and ion pair interactions, suggesting that this interface is real and probably functional in most bacterial

Fig. 5. Binding site conservation in processivity clamp interactions. (A) Secondary structure cartoon of the *E. coli* β -clamp–Pol IV-LF complex compared with other processivity clamp–peptide complexes. (B) Human PCNA–p21^{cip/waf} peptide. (C) *Pyrococcus furiosus* PCNA–replication factor C peptide. (D) Phage RB69 sliding clamp–DNA polymerase. Despite the very low level of sequence similarity between the bacterial, mammalian, archaeal and viral clamps, the gross topology of the peptide-binding site is conserved in all the systems.

species. Variation is greater in the Pol IV sequences, which show a comparable degree of variation in the residues forming the C-terminal β -clamp-binding peptide and are in general more divergent than the β -clamps. The greater conservation of the interacting residues in the β -clamps probably reflects the evolutionary pressure to conserve common binding sites for a range of different protein partners.

Whether this additional protein–protein interface is unique to Pol IV is unknown. All five identified *E. coli* DNA polymerases have been shown to interact with the β -clamp (Hughes *et al.*, 1991; Naktinis *et al.*, 1995; Kim and McHenry, 1996; Lopez de Saro and O'Donnell, 2001; Sutton *et al.*, 2001; Lenne-Samuel *et al.*, 2002), but the LF domain that mediates the interaction described here is only found in the Y-family enzymes, Pol IV and the UmuC component of Pol V. Interestingly, mutations impairing the lesion bypass activities of UmuD' and UmuC have been identified in residues that are predicted to map in the structure, in and around the loops in the LF domain that in Pol IV interact with the β -clamp (Woodgate *et al.*, 1994; Boudsocq *et al.*, 2002).

While the β -clamp complex structure described here only contains the LF domain of Pol IV, more complete structures of Pol IV/DinB homologues from archaea have been described, including a complex between the Dpo4

enzyme of *Sulfolobus solfataricus*, a DNA primer–template duplex and an incoming nucleotide (Ling *et al.*, 2001). By superimposing the LF domain from SsDpo4 onto the Pol IV-LF in the complex, we can construct a good working model for how the full Pol IV enzyme might interact with the β -clamp via the C-terminal peptide motif and the protein–protein interface (Figure 6A). This reconstruction has several immediate implications. First, there are no steric clashes between the other domains of the docked SsDpo4 and the β -clamp, showing clearly that the LF within the full-length Pol IV could bind readily to the β -clamp in the same way as the isolated LF domain, and that the observed interactions in the crystal structure are therefore not an artefact due to use of an isolated domain. Secondly, it is evident that when bound in this position on the β -clamp, the Pol IV molecule would not have productive access to a linear DNA duplex running perpendicularly through the centre of the β -clamp as is conventionally assumed in the absence of any evidence to the contrary (Kong *et al.*, 1992; Krishna *et al.*, 1994). Even if the DNA passing through the lumen of the clamp were to be severely bent, as it might well be in the vicinity of a primer–template junction, on the basis of the SsDpo4–DNA complex it would still not be possible for Pol IV bound to the clamp in this orientation to access the DNA. A productive orientation for Pol IV, allowing access to

Fig. 6. The Pol IV/DinB 'polymerase switch'. (A) Model of a Pol IV/DinB type Y-class polymerase (red) bound to the β -clamp (gold) in the 'locked-down' inactive position. The position of the polymerase was modelled by superimposing the LF domain of the archaeal Dpo4 enzyme from the DNA complex (PDB code: 1JXL) onto the Pol IV-LF in the complex with the β -clamp described here (blue). Modelled in this position, the polymerase makes no steric clashes with the β -clamp, but cannot access the primer-template junction of the DNA (primer strand, pink, template strand, green). Furthermore, the face of the β -clamp is not obstructed, so that a replicative DNA polymerase bound to the DNA could be accommodated. (B) Model of a Pol IV/DinB type Y-class polymerase bound to the primer-template junction following a polymerase switch. The position of the polymerase was modelled by superimposing the DNA from the Dpo4-DNA complex onto the end of a long DNA molecule running perpendicularly through the lumen of the clamp. Disruption of the substantial protein-protein interface between the clamp and the polymerase in the 'locked-down' conformation would provide a barrier to attainment of this active complex, disfavoured access by the translesion polymerase except when the primer-template junction is vacated for a long period. Contact with the clamp is maintained by the C-terminal clamp-binding peptide (blue), which tethers the enzyme to the replication complex.

duplex DNA running perpendicularly through the lumen of the clamp, as generally envisaged, would require disruption of the protein-protein interface. However, this can be achieved while still maintaining the interaction between the β -clamp and the C-terminal Pol IV peptide, by main chain rotations in the short linker that connects the

C-terminal clamp-binding peptide and the globular core of the LF domain. Pol IV bound to the β -clamp could therefore exist in an equilibrium between two states: a 'locked-down' complex with a maximal protein-protein interface but no access to DNA, and a 'tethered' complex flexibly attached via the C-terminal peptide and capable of productive interaction with DNA (Figure 6B).

The existence of an apparently stable complex between Pol IV and the β -clamp, in which the polymerase cannot access the DNA primer-template junction, suggests a mechanism by which the translesion polymerase can be recruited to the site of DNA replication but with its highly mutagenic activity controlled. Current models for lesion bypass involve a process of DNA polymerase 'switching' (Cordonnier and Fuchs, 1999) in which processivity factors function as 'tool-belts' (Pages and Fuchs, 2002) tethering multiple DNA polymerases (replicative and translesion), which can interchangeably access the primer-template junction depending on its suitability for their active sites. Thus a distorted template containing a lesion that is released by the replicative DNA polymerase can be picked up rapidly by a translesion DNA polymerase attached to the same clamp. One potential problem with this model is that it implies a continuous competition between the replicative and translesion DNA polymerase(s) for access to the primer-template junction, to which they are both tethered at equally high effective concentration via interaction with the processivity clamp. For replication to remain accurate, it would be essential that this competition is weighted greatly in favour of the high-fidelity replicative enzyme and against the error-prone translesion enzyme. While some component of that weighting will come from the slow 'off-rate' implicit in the processivity of the replicative DNA polymerase, a significant contribution could also be provided if the equilibrium between DNA-bound and DNA-unbound states for the translesion DNA polymerase were shifted away from the DNA-bound state by providing a defined and thermodynamically stable DNA-unbound state as in the complex described here. At least for *E.coli* Pol IV, the combination of a flexible peptide-protein 'tether' together with a secondary protein-protein interface provides a solution to the problem of maintaining a translesion DNA polymerase in close proximity to the primer-template junction, while regulating its access to the DNA so as to minimize error-prone DNA synthesis except in circumstances where the replicative enzymes are unable to proceed. Whether this secondary protein-protein interface is a common feature of translesion DNA polymerase interactions with processivity clamps remains to be seen.

Materials and methods

Cloning and expression

A truncated, C-terminal form of Pol IV (Pol IV-LF) encoding the LF domain (residue Val243-Leu351) was amplified from *E.coli* JM109 genomic DNA to incorporate an N-terminal His₆ tag and was cloned into pET11 (Novagen) for expression purposes. The full-length β -clamp was amplified from *E.coli* JM109 genomic DNA and cloned into pET17b (Novagen). Both constructs were expressed in *E.coli* B834 (DE3) cells. Freshly transformed cells were used to inoculate large-scale cultures of LB containing the appropriate antibiotic which were incubated at 37°C to an absorbance at 600 nm of 0.3–0.4. Expression of the β -clamp construct was induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG)

Table I. Crystallographic statistics

Data collection	All data (outer shell)
Resolution range (Å)	62.1–1.90 (2.01–1.90)
R_{merge}	0.060 (0.201)
$I/\sigma(I)$	8.3 (3.5)
Completeness (%)	96.5 (87.4)
Multiplicity	4.9 (2.7)
No. of unique reflections	90 134 (12 565)
Structure refinement	
No. of atoms (protein)	7528
No. of atoms (all)	8485
Resolution range (Å)	62.1–1.90
R_{cryst}	0.180
R_{free}	0.240

and the cells allowed to grow overnight at 20°C, prior to harvesting. The Pol IV-LF construct was grown at 20°C overnight, following the addition of 0.1 mM IPTG.

Purification of Pol IV-LF- β -clamp complex

Cells containing Pol IV-LF were resuspended in buffer A (containing 50 mM HEPES pH 7.0, 2 mM MgCl₂, 0.5 M NaCl and 5 mM imidazole). Following lysis, via sonication and centrifugation, the clarified lysate was applied to Talon affinity resin (Clontech) equilibrated in buffer A. The protein was eluted in a small volume of buffer A containing 300 mM imidazole, for application to a 26/60 S-200 (Pharmacia) size exclusion column equilibrated in buffer B (50 mM HEPES pH 7.0, 2 mM MgCl₂ and 0.2 M NaCl). Purified Pol IV-LF was immobilized on Talon resin equilibrated in buffer C (50 mM HEPES pH 7.0, 2 mM MgCl₂, 0.1 M NaCl, 10 mM imidazole and 0.1% NP-40). Cells containing over-expressed β -clamp were resuspended in 50 mM HEPES pH 7.0, 2 mM MgCl₂ and 0.3 M NaCl. Following sonication, the lysate was clarified via centrifugation prior to application to the Talon resin containing immobilized Pol IV-LF. The resin was incubated with gentle agitation at 4°C for 30 min prior to washing with buffer C. The complex was eluted in a small volume of buffer C containing 300 mM imidazole before being applied to a 26/200 S-200 column equilibrated in buffer B. Fractions containing the complex were concentrated via centrifugation in a Vivaspinn (Vivascience) device (30 000 Da mol. wt cutoff) and were stored at 4°C.

Crystallization and data collection

Crystals were obtained using the hanging drop method at 14°C. Protein at a concentration of 6 mg/ml was mixed in a 1:1 ratio with, and equilibrated against, a solution containing 1.2 M ammonium sulfate and 50 mM sodium acetate pH 4.6. Usable crystals were obtained after 3 days incubation. Crystals were harvested and transferred to a cryoprotectant buffer containing 30% ethylene glycol prior to freezing.

The complex crystallized in space group $P2_12_12$ with cell dimensions $a = 146.45$ Å, $b = 70.12$ Å, $c = 110.93$ Å. Crystal volume calculations suggested the presence of one dimer of the β -clamp and two Pol IV-LF monomers within the asymmetric unit (solvent content 53% by volume). Data were collected to 1.9 Å at ID14-1 at ESRF Grenoble.

Structure determination/refinement

Data were processed using MOSFLM and programmes of the CCP4 suite (CCP4, 1994; Leslie, 1995). Molecular replacement was carried out using AMORE (Navaza, 1994), utilising the dimeric β -clamp structure (PDB code: 2POL) as a search model, giving a clear solution. Difference maps generated using the aligned β -clamp structure showed clear density indicating the position of the Pol IV-LF monomers. An initial model for the LF domain was produced from a polyalanine model based on the equivalent region in the Dpo4 structure (PDB code: 1JXL). Refinement was carried out using CNS (Brünger *et al.*, 1998) and Refmac (Murshudov *et al.*, 1997) interspersed with manual building based on difference maps in O (Jones *et al.*, 1991). Data collection and refinement statistics are given in Table I. Refined coordinates and structure factors for the complex have been deposited in the Protein Databank with accession code 1UNN.

Acknowledgements

We are grateful to the ESRF Grenoble for access to synchrotron radiation, to Cancer Research UK for infrastructural support for structural biology at ICR, and to Robert Fuchs and Dominique Burnouf for useful discussion and sharing unpublished data. This work was supported by a Project Grant from The Wellcome Trust.

References

- Boudsocq,F., Ling,H., Yang,W. and Woodgate,R. (2002) Structure-based interpretation of missense mutations in Y-family DNA polymerases and their implications for polymerase function and lesion bypass. *DNA Repair*, **1**, 343–358.
- Brünger,A.T. *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Cordonnier,A.M. and Fuchs,R.P. (1999) Replication of damaged DNA: molecular defect in xeroderma pigmentosum variant cells. *Mutat. Res.*, **435**, 111–119.
- Dalrymple,B.P., Kongsuwan,K., Wijffels,G., Dixon,N.E. and Jennings,P.A. (2001) A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems. *Proc. Natl Acad. Sci. USA*, **98**, 11627–11632.
- Friedberg,E.C., Fischhaber,P.L. and Kisker,C. (2001) Error-prone DNA polymerases: novel structures and the benefits of infidelity. *Cell*, **107**, 9–12.
- Friedberg,E.C., Wagner,R. and Radman,M. (2002) Specialized DNA polymerases, cellular survival and the genesis of mutations. *Science*, **296**, 1627–1630.
- Gerlach,V.L., Aravind,L., Gotway,G., Schultz,R.A., Koonin,E.V. and Friedberg,E.C. (1999) Human and mouse homologs of *Escherichia coli* DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proc. Natl Acad. Sci. USA*, **96**, 11922–11927.
- Gibbs,P.E., Wang,X.D., Li,Z., McManus,T.P., McGregor,W.G., Lawrence,C.W. and Maher,V.M. (2000) The function of the human homolog of *Saccharomyces cerevisiae* REV1 is required for mutagenesis induced by UV light. *Proc. Natl Acad. Sci. USA*, **97**, 4186–4191.
- Goodman,M.F. (1997) Hydrogen bonding revisited: geometric selection as a principal determinant of DNA replication fidelity. *Proc. Natl Acad. Sci. USA*, **94**, 10493–10495.
- Gulbis,J.M., Kelman,Z., Hurwitz,J., O'Donnell,M. and Kuriyan,J. (1996) Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell*, **87**, 297–306.
- Haracska,L., Prakash,S. and Prakash,L. (2000) Replication past O(6)-methylguanine by yeast and human DNA polymerase η . *Mol. Cell. Biol.*, **20**, 8001–8007.
- Haracska,L., Johnson,R.E., Unk,I., Phillips,B., Hurwitz,J., Prakash,L. and Prakash,S. (2001a) Physical and functional interactions of human DNA polymerase η with PCNA. *Mol. Cell. Biol.*, **21**, 7199–7206.
- Haracska,L., Washington,M.T., Prakash,S. and Prakash,L. (2001b) Inefficient bypass of an abasic site by DNA polymerase η . *J. Biol. Chem.*, **276**, 6861–6866.
- Haracska,L., Unk,I., Johnson,R.E., Phillips,B.B., Hurwitz,J., Prakash,L. and Prakash,S. (2002) Stimulation of DNA synthesis activity of human DNA polymerase κ by PCNA. *Mol. Cell. Biol.*, **22**, 784–791.
- Hingorani,M.M. and O'Donnell,M. (2000) Sliding clamps: a (tail)ored fit. *Curr. Biol.*, **10**, R25–R29.
- Hosfield,D.J., Mol,C.D., Shen,B.H. and Tainer,J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*, **95**, 135–146.
- Hughes,A.J., Jr, Bryan,S.K., Chen,H., Moses,R.E. and McHenry,C.S. (1991) *Escherichia coli* DNA polymerase II is stimulated by DNA polymerase III holoenzyme auxiliary subunits. *J. Biol. Chem.*, **266**, 4568–4573.
- Jeruzalmski,D., Yuriev,O., Zhao,Y., Young,M., Stewart,J., Hingorani,M., O'Donnell,M. and Kuriyan,J. (2001) Mechanism of processivity clamp opening by the δ subunit wrench of the clamp loader complex of *E. coli* DNA polymerase III. *Cell*, **106**, 417–428.
- Johnson,R.E., Washington,M.T., Haracska,L., Prakash,S. and Prakash,L. (2000) Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature*, **406**, 1015–1019.

- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
- Kelman, Z. and O'Donnell, M. (1995) Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res.*, **23**, 3613–3620.
- Kim, D.R. and McHenry, C.S. (1996) Identification of the β -binding domain of the α subunit of *Escherichia coli* polymerase III holoenzyme. *J. Biol. Chem.*, **271**, 20699–20704.
- Kong, X.P., Onrust, R., O'Donnell, M. and Kuriyan, J. (1992) Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, **69**, 425–437.
- Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233–1243.
- Kunkel, T.A., Olovov, I.I. and Bebenek, K. (2003) Functions of human DNA polymerases η , κ and ι suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair*, **2**, 135–149.
- Lenne-Samuel, N., Wagner, J., Etienne, H. and Fuchs, R.P. (2002) The processivity factor β controls DNA polymerase IV traffic during spontaneous mutagenesis and translesion synthesis *in vivo*. *EMBO Rep.*, **3**, 45–49.
- Leslie, A.G.W. (1995) *MOSFLM Users Guide*. MRC Laboratory of Molecular Biology, Cambridge, UK.
- Lin, W., Xin, H., Zhang, Y., Wu, X., Yuan, F. and Wang, Z. (1999) The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Res.*, **27**, 4468–4475.
- Ling, H., Boudsocq, F., Woodgate, R. and Yang, W. (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell*, **107**, 91–102.
- Lopez de Saro, F.J. and O'Donnell, M. (2001) Interaction of the β sliding clamp with MutS, ligase and DNA polymerase I. *Proc. Natl Acad. Sci. USA*, **98**, 8376–8380.
- Masutani, C. *et al.* (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η . *Nature*, **399**, 700–704.
- Matsumiya, S., Ishino, Y. and Morikawa, K. (2001) Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci.*, **10**, 17–23.
- Matsumiya, S., Ishino, S., Ishino, Y. and Morikawa, K. (2002) Physical interaction between proliferating cell nuclear antigen and replication factor C from *Pyrococcus furiosus*. *Genes Cells*, **7**, 911–922.
- McDonald, J.P., Rapic-Otrin, V., Epstein, J.A., Broughton, B.C., Wang, X., Lehmann, A.R., Wolgemuth, D.J. and Woodgate, R. (1999) Novel human and mouse homologs of *Saccharomyces cerevisiae* DNA polymerase η . *Genomics*, **60**, 20–30.
- McKenzie, G.J., Lee, P.L., Lombardo, M.J., Hastings, P.J. and Rosenberg, S.M. (2001) SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Mol. Cell*, **7**, 571–579.
- Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M. and Kuriyan, J. (2000) Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J. Mol. Biol.*, **296**, 1215–1223.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D*, **53**, 240–255.
- Naktinis, V., Onrust, R., Fang, L. and O'Donnell, M. (1995) Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader and sliding clamps in one holoenzyme particle. II. Intermediate complex between the clamp loader and its clamp. *J. Biol. Chem.*, **270**, 13358–13365.
- Naktinis, V., Turner, J. and O'Donnell, M. (1996) A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell*, **84**, 137–145.
- Napolitano, R., Janel-Bintz, R., Wagner, J. and Fuchs, R.P. (2000) All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *EMBO J.*, **19**, 6259–6265.
- Navaza, J. (1994) AMoRE—an automated package for molecular replacement. *Acta Crystallogr. D*, **50**, 157–163.
- Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W.J., Gerlach, V.L., Friedberg, E.C., Ohmori, H. and Kunkel, T.A. (2000) Fidelity and processivity of DNA synthesis by DNA polymerase κ , the product of the human DINB1 gene. *J. Biol. Chem.*, **275**, 39678–39684.
- Ohmori, H. *et al.* (2001) The Y-family of DNA polymerases. *Mol. Cell*, **8**, 7–8.
- O-Wang, Z., Kawamura, K., Tada, Y., Ohmori, H., Kimura, H., Sakiyama, S. and Tagawa, M. (2001) DNA polymerase κ , implicated in spontaneous and DNA damage-induced mutagenesis, is overexpressed in lung cancer. *Cancer Res.*, **61**, 5366–5369.
- Pages, V. and Fuchs, R.P. (2002) How are DNA lesions turned into mutations within cells? *Oncogene*, **21**, 8957–8966.
- Shamoo, Y. and Steitz, T.A. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell*, **99**, 155–166.
- Silvian, L.F., Toth, E.A., Pham, P., Goodman, M.F. and Ellenberger, T. (2001) Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nat. Struct. Biol.*, **8**, 984–989.
- Sutton, M.D., Kim, M. and Walker, G.C. (2001) Genetic and biochemical characterization of a novel umuD mutation: insights into a mechanism for UmuD self-cleavage. *J. Bacteriol.*, **183**, 347–357.
- Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R. and Goodman, M.F. (1999) UmuD'(2)C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl Acad. Sci. USA*, **96**, 8919–8924.
- Tang, M., Pham, P., Shen, X., Taylor, J.S., O'Donnell, M., Woodgate, R. and Goodman, M.F. (2000) Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature*, **404**, 1014–1018.
- Trincao, J., Johnson, R.E., Escalante, C.R., Prakash, S., Prakash, L. and Aggarwal, A.K. (2001) Structure of the catalytic core of *S. cerevisiae* DNA polymerase η : implications for translesion DNA synthesis. *Mol. Cell*, **8**, 417–426.
- van Steeg, H. and Kraemer, K.H. (1999) Xeroderma pigmentosum and the role of UV-induced DNA damage in skin cancer. *Mol. Med. Today*, **5**, 86–94.
- Wagner, J., Gruz, P., Kim, S.R., Yamada, M., Matsui, K., Fuchs, R.P. and Nohmi, T. (1999) The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol. Cell*, **4**, 281–286.
- Wagner, J., Fujii, S., Gruz, P., Nohmi, T. and Fuchs, R.P. (2000) The β clamp targets DNA polymerase IV to DNA and strongly increases its processivity. *EMBO Rep.*, **1**, 484–488.
- Woodgate, R., Singh, M., Kulaeva, O.I., Frank, E.G., Levine, A.S. and Koch, W.H. (1994) Isolation and characterization of novel plasmid-encoded umuC mutants. *J. Bacteriol.*, **176**, 5011–5021.
- Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechko, O., Taylor, J.S., Geacintov, N.E. and Wang, Z. (2000) Error-free and error-prone lesion bypass by human DNA polymerase κ *in vitro*. *Nucleic Acids Res.*, **28**, 4138–4146.

Received July 17, 2003; revised September 15, 2003;
accepted September 16, 2003