Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and biochemical characterization of proliferating cellular nuclear antigen from a parasitic protozoon

Cesar S. Cardona-Felix, Samuel Lara-Gonzalez‡ and Luis G. Brieba*

Grupo de Bioquímica Estructural, Laboratorio Nacional de Genomica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Km 9.6 Libramiento Norte, Carretera Irapuato-León, 36821 Irapuato, Guanajuato, México

‡ Current address: División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, Camino a la Presa San José 2055, CP 78216 San Luis Potosí, SLP, México.

Correspondence e-mail: lgbrieba@ira.cinvestav.mx

Received 14 February 2011
Accepted 21 March 2011

PDB Reference: proliferating cellular nuclear antigen, 3p91.

Proliferating cellular nuclear antigen (PCNA) is a toroidal-shaped protein that is involved in cell-cycle control, DNA replication and DNA repair. Parasitic protozoa are early-diverged eukaryotes that are responsible for neglected diseases. In this work, a PCNA from a parasitic protozoon was identified, cloned and biochemically characterized and its crystal structure was determined. Structural and biochemical studies demonstrate that PCNA from *Entamoeba histolytica* assembles as a homotrimer that is able to interact with and stimulate the activity of a PCNA-interacting peptide-motif protein from *E. histolytica*, EhDNAligI. The data indicate a conservation of the biochemical mechanisms of PCNA-mediated interactions between metazoa, yeast and parasitic protozoa.

1. Introduction

Proliferating cellular nuclear antigen (PCNA) is an evolutionary conserved protein that is involved in the cell cycle, chromatin remodelling, DNA repair and nuclear DNA replication (Maga & Hubscher, 2003; Ivanov *et al.*, 2006). PCNA is a multifunctional protein; however, it is better known for its role as a sliding clamp that interacts with replicative DNA polymerases in order to increase their processivity. PCNA also interacts with proteins involved in DNA repair such as FEN1, DNA glycosylases and DNA ligase I, cell-cycle control proteins such as p21, CCK2 and cyclin D, and chromatin-remodelling proteins such as DNA methyltransferases (for recent reviews, see Stoimenov & Helleday, 2009; Tsurimoto, 1998). PCNA is composed of three subunits that assemble to adopt a ring-shaped structure. PCNA oligomerization differs between archaea and eukaryotes: PCNAs from eukaryotes assemble as homotrimers, while in crenarchaeota such as *Sulfolobus sulfataricus* PCNA assembles as a functional heterotrimer (Williams *et al.*, 2006; Dionne *et al.*, 2003; Krishna *et al.*, 1994; Chia *et al.*, 2010). Despite this evolutionary divergence, the toroidal structure of PCNA is highly conserved in humans, yeast, plants and archaea (Strzalka *et al.*, 2009; Chia *et al.*, 2010; Krishna *et al.*, 1994; Matsumiya *et al.*, 2001). Each monomer of PCNA is composed of two globular domains connected by an inter-domain connecting loop (IDCL; Krishna *et al.*, 1994). Thus, trimeric PCNA forms a ring with pseudo-sixfold symmetry. The inner surface of the ring is composed of 12 α-helices, whereas the outer surface is composed of 54 β-sheets and three IDCLs (Ivanov *et al.*, 2006). The inner part of the PCNA ring consists of positively charged amino acids oriented to interact with the sugar-phosphate backbone of double-stranded DNA. This structural organization permits PCNA to encircle DNA while allowing lateral movement and provides a scaffold for the assembly of associated proteins.

PCNA interacts with its multiple protein partners mainly through its IDCL *via* a consensus motif that is present in its interacting proteins and named the PCNA-interacting protein box (PIP box). The PIP box has a Q*XX*(M/L/I)*XX*(F/Y)(F/Y) consensus sequence and adopts a 3_{10}-helical structure (Moldovan *et al.*, 2007). Besides this interaction, PCNA can alternatively bind to interacting proteins using additional regions such as its C-terminus, the centre loop and the back-side loop (Mayanagi *et al.*, 2009).

PCNA has been biochemically and structurally characterized in higher eukaryotes, including humans (Almendral *et al.*, 1987), yeast (Bauer & Burgers, 1990) and plants (Strzalka *et al.*, 2009), and archaea (Williams *et al.*, 2006). However, no structural studies of PCNA from protozoan parasites have been carried out to date. Protozoan PCNAs that have been biochemically characterized to date include PCNAs from *Plasmodium falciparum*, *Toxoplasma gondii* and *Leishmania donovani* (Kilbey *et al.*, 1993; Patterson *et al.*, 2002; Guerini *et al.*, 2000; Kumar *et al.*, 2009). *P. falciparum* and *T. gondii* contain two different PCNA genes in their genome, whereas *L. donovani* contains one PCNA gene. Phylogenetic studies suggest that although some PCNAs from archaea assemble as heterotrimers, all PCNAs have diverged from a homotrimeric ancestor followed by gene duplication and specialization (Chia *et al.*, 2010). These findings suggest the possibility that PCNAs from protozoan parasites such as *P. falciparum* and *T. gondii* may assemble as a mixed heterotrimer, as has been demonstrated to occur *in vitro* in *Arabidopsis thaliana* (Strzalka *et al.*, 2009).

Entamoeba histolytica is an early-branching protozoan parasite that is responsible for amoebiasis, a disease that affects close to ten million people worldwide and causes approximately 100 000 deaths annually (Stanley, 2003). The genome of *E. histolytica* indicates that this protozoan parasite may perform DNA-associated processes, such as those involved in DNA replication and DNA repair, in a fashion similar to higher eukaryotes (Clark *et al.*, 2007). However, few DNA-binding proteins from this parasite have been identified and biochemically characterized. As a first step towards understanding the replication process in this parasite, we here report the biochemical and structural characterization of PCNA from *E. histolytica* (EhPCNA).

2. Material and methods

2.1. *In silico* identification of EhPCNA

In silico identification of EhPCNA was carried out by a *BLAST* search of the Pathema database (http://pathema.jcvi.org/cgi-bin/Entamoeba/PathemaHomePage.cgi) using the amino-acid sequence of human PCNA (HsPCNA) as a query. A hidden Markov model consisting of 84 amino-acid sequences of PCNAs from different organisms was constructed using the program *HMMER* v.2.3.2. Hidden Markov analyses of the coding sequences of *E. histolytica* were used to test for the possible presence of multiple genes of PCNA in the parasite's genome.

2.2. Cloning of full-length EhPCNA

The 789 bp coding sequence of EhPCNA was amplified from genomic DNA of *E. histolytica* strain HM1-IMSS. The PCR primers were 5′-GGTCGGAATTC**CATATG**TG-TGCTTTCCACGCCAAATTTAAAG-3′ (forward) and 5′-GGTCTT**GGATCC**TTACTCTTGAGGTTCATCTTCTTC-3′ (reverse). These primers contained restriction sites for *Nde*I and *Bam*HI restriction enzymes, respectively (the recognition sequences are shown in bold). The purified PCR product was digested and ligated into a modified pET19 vector in which the thrombin site was modified to a PreScission Protease site. Positive clones were confirmed by automated DNA sequencing.

2.3. Expression and purification of EhPCNA

Escherichia coli BL21 (DE3) Star was transformed with the pET19-EhPCNA plasmid and plated onto an agar plate supplemented with 100 µg ml^{-1} ampicillin. A single colony was used to grow a 100 ml overnight LB culture supplemented with 100 µg ml^{-1} ampicillin, which was then used to inoculate a 4 l culture. This culture was further grown at 310 K until it reached an OD_{600} of 0.8. Heterologous expression was induced by adding IPTG to a final concentration of 0.5 m*M*. The induced cell culture was centrifuged at 6000 rev min^{-1} for 15 min and the pellet was resuspended in 100 ml lysis buffer (50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, 0.01% Triton X-100). Cells were lysed by sonication and the lysate was centrifuged at 15 000 rev min^{-1} for 30 min at 277 K. The supernatant was filtered through a 0.45 µm pore-size filter and passed through a 5 ml Ni Sepharose High Performance column (GE Healthcare) previously equilibrated with lysis buffer. The first column wash consisted of 75 ml lysis buffer, the second wash consisted of 75 ml lysis buffer supplemented with 40 m*M* imidazole and the third wash consisted of 75 ml lysis buffer supplemented with 80 m*M* imidazole. EhPCNA was eluted with 20 ml 50 m*M* Tris–HCl, 100 m*M* NaCl, 500 m*M* imidazole. The eluate was immediately dialyzed in 2 l of a solution consisting of 50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, 10 m*M* DTT, 2 m*M* EDTA for 12 h. A second purification step was performed using a DEAE anion-exchange column equilibrated with 50 m*M* Tris–HCl, 30 m*M* NaCl and subjected to a NaCl linear gradient from 100 to 800 m*M*. EhPCNA eluted as a single peak between 400 and 500 m*M* NaCl. The histidine tag of EhPCNA was removed by the addition of PreScission Protease (GE Healthcare) using 0.1 mg protease per 10 mg purified protein. Histidine-tag removal was carried out by gel filtration on a Superdex 75 column (GE Healthcare) equilibrated with 50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl. Determination of the multimeric state of EhPCNA was carried out using a Superdex 200 size-exclusion column (GE Healthcare) previously calibrated with molecular-weight standards in a buffer consisting of 50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl. The molecular-weight markers were thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa) and myoglobin (17 kDa).

2.4. Electrophoretic mobility-shift and nick-sealing activity assays

DNA binding and enzymatic activity stimulation of EhDNAligI by EhPCNA were investigated using an electrophoretic mobility-shift assay (EMSA) and an *in vitro* DNA-ligation reaction as previously reported (Cardona-Felix *et al.*, 2010). Briefly, a double-stranded nicked DNA substrate was formed by annealing a downstream 24-mer (5′-CGCAGCCCACCTGCCCACCTACT-3′) and an upstream 21-mer (5′-GGCCCTGCGCTAGTGCCAAGG-3′) to a complementary 45-mer template strand. The upstream 21-mer was 5′-labelled with 100 µCi [γ-^{32}P]-ATP (3000 Ci mmol^{-1}). Binding reactions were assayed in a buffer consisting of 50 m*M* Tris–HCl pH 7.5, 10 m*M* DTT, 5% glycerol, 30 n*M* ^{32}P-labelled nicked double-stranded DNA, 10 n*M* EhDNAligI and increasing concentrations of EhPCNA from 20 to 160 n*M* and were resolved by native polyacrylamide gel electrophoresis. Nick-sealing activity assays were performed in a buffer consisting of 50 m*M* Tris–HCl pH 7.5, 10 m*M* DTT, 10 m*M* $MgCl_2$, 0.5 m*M* ATP, 30 n*M* ^{32}P-labelled nicked double-stranded DNA substrate, 10 n*M* EhDNAligI and 500 n*M* EhPCNA. The reactions were incubated at 310 K and aliquots were taken at 0.5, 1, 2, 4, 8, 16 and 32 min and stopped by adding equal volumes of stop buffer (90% formamide, 50 m*M* EDTA). Reactions were electrophoresed in a 16% polyacrylamide/8 *M* urea gel and the reaction products were detected by phosphorimagery.

2.5. Protein crystallography

The protein concentration was measured by Bradford assay using bovine serum albumin dilutions as standards (Bradford, 1976). Protein crystallization screening was performed with EhPCNA at a concentration of 5 mg ml^{-1} in 50 m*M* Tris pH 7.5, 100 m*M* NaCl, 1 m*M* EDTA and 10 m*M* DTT using the hanging-drop vapour-diffusion technique by mixing 1 µl EhPCNA with 1 µl reservoir solution from commercial screens. Thin crystal plates appeared overnight in a condition consisting of 1.6 *M* sodium citrate tribasic dihydrate pH 6.5 and were used as seeds to grow single cube-shaped crystals that were suitable for X-ray diffraction. The crystals were cryoprotected by quick soaking in 1.6 *M* sodium citrate tribasic dihydrate pH 6.5 and flash-frozen in liquid nitrogen.

2.6. Data collection and refinement

A single protein crystal was used to collect diffraction data at 100 K using a MAR 300 CCD detector on the APS beamline LS-CAT 21-ID-D with 1.0° rotation per image. Diffraction intensities were integrated and scaled in space group *H*3 using *MOSFLM* and *SCALA* (Leslie, 1992). The search model used for molecular replacement was generated with the program *Sculptor* (Bunkóczi & Read, 2011) using human PCNA (PDB entry 1vym, chain *A*; Kontopidis *et al.*, 2005) as a template. A unique solution was readily found with *Phaser* and structure-factor data were analyzed with *phenix.xtriage* to detect outliers and twinning (McCoy *et al.*, 2007; Adams *et al.*, 2010). Rigid-body refinement, simulated annealing, individual atomic coordinate and individual atom isotropic displacement parameter refinement strategies were performed with *phenix.refine* (Adams *et al.*, 2010). Manual model adjustment to improve the fit to likelihood-weighted electron-density maps was carried out using *Coot* (Emsley & Cowtan, 2004). Water molecules were added into difference electron-density maps. The quality of the model was checked using the *MolProbity* server (Chen *et al.*, 2010; Table 1). Structural models were rendered with *PyMOL* and *UCSF Chimera* (Pettersen *et al.*, 2004; DeLano, 2002).

Table 1
PCNA crystal parameters and data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
X-ray source	APS LS-CAT 21-ID-D
Wavelength (Å)	0.9787
Data-collection temperature (K)	100
Resolution range (Å)	27.91–2.40 (2.53–2.40)
Space group	*H*3
Unit-cell parameters (Å, °)	$a = b = 141.1$, $c = 49.2$, $\alpha = \beta = 90.0$, $\gamma = 120.0$
Matthews coefficient (Å^3 Da^{-1})	3.3
Solvent content (%)	60
No. of measured reflections	55605 (8075)
No. of unique reflections	14275 (2087)
Completeness (%)	99.9 (100.0)
Multiplicity	3.9 (3.9)
Mean $I/\sigma(I)$	10.7 (5.8)
$R_{\rm merge}$ (%)	8.9 (15.7)
Refinement	
Resolution range (Å)	26.66–2.40 (2.59–2.40)
$R_{\rm work}$†	0.148 (0.208)
$R_{\rm free}$†	0.193 (0.236)
Reflections, working	13521 (2708)
Reflections, free	713 (143)
Twinning fraction; operator	0.266; $h, -h-k, -l$
Molecules per asymmetric unit	1
Non-H atoms	1839
Water molecules	85
Average *B* factor (Å^2)	28.1
R.m.s.d. bond lengths‡ (Å)	0.007
R.m.s.d. bond angles‡ (°)	1.027
Ramachandran favoured (%)	96.2
Ramachandran outliers (%)	0.0

† $R = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively. $R_{\rm work}$ is for reflections from the working set; $R_{\rm free}$ is calculated with 5% of the reflections chosen at random and omitted from refinement. ‡ Root-mean-square deviation of bond lengths and bond angles from ideal geometry.

3. Results and discussion

3.1. Cloning, expression and purification of recombinant EhPCNA

A hidden Markov analysis using PCNA-related sequences revealed only one positive score in the Pathema database annotated as 'putative proliferating cellular nuclear antigen' in locus EHI_128450 (data not shown). A single hit was obtained during hidden Markov analysis, strongly suggesting that *E. histolytica* only contains one PCNA gene in its genome, which is named EhPCNA in this work. EhPCNA contains 262 amino acids and is annotated in the NCBI GenBank with identification number XP_651510.1. Hidden Markov models of PCNA were initially used to identify duplicated PCNA

genes in *Thermotoga maritima* and an orthologuous PCNA gene in vaccinia virus, therefore validating the use of this approach to identify duplicated PCNA genes (Neuwald & Poleksic, 2000; Da Silva & Upton, 2009). Amino-acid sequence alignments with PCNAs from *Saccharomyces cerevisiae* and *Homo sapiens* revealed that EhPCNA has 33 and 34% amino-acid identity to these PCNAs, respectively. This percentage of amino-acid sequence resembles the amino-acid identity observed between other PCNAs (Naryzhny, 2008).

In order to test the biochemical properties of EhPCNA, its open reading frame was PCR-amplified from genomic DNA and cloned into a modified pET19 vector. This modified vector provides the technical advantage that the histidine tag can be removed with a specific protease, leaving only three extra amino acids (Gly, Pro and His) before the original N-terminal methionine. Several *E. coli* strains were tested for the overexpression of EhPCNA and the best results were obtained with strain BL21 (DE3) Star (Cardona-Felix, data not shown; Fig. 1*b*, lanes 1 and 2). Approximately 50% of the heterologous induced EhPCNA was soluble in a lysis buffer with a near-physiological salt concentration (Fig. 1*b*, lane 3). Histidine-tagged EhPCNA remained bound to the nickel column

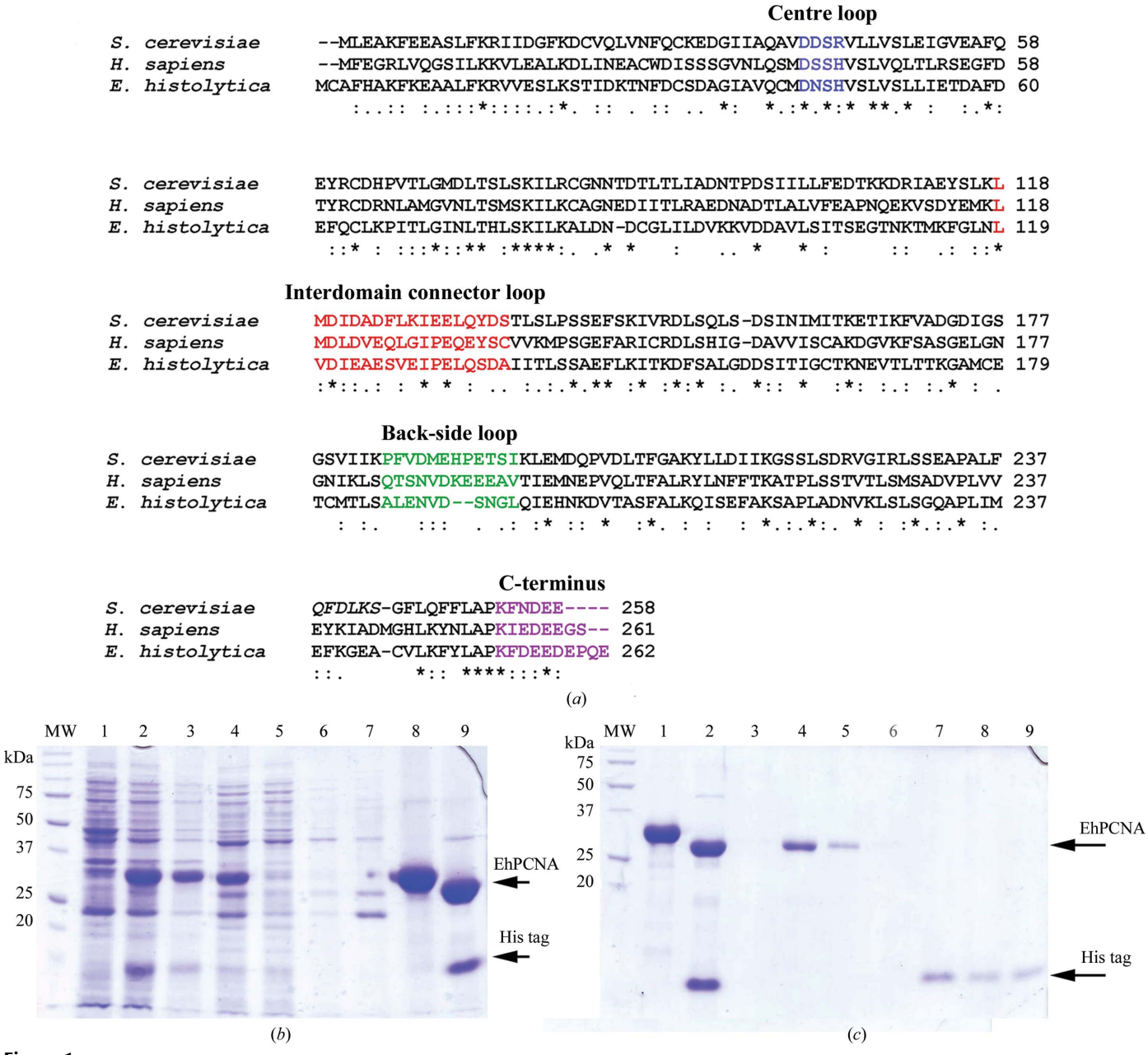

Figure 1
Identification, cloning, expression and purification of EhPCNA. (*a*) Amino-acid sequence alignment of EhPCNA with PCNAs from *H. sapiens* and *S. cerevisiae*. The structural elements present in PCNA are coloured as follows: the centre loop is coloured blue, the inter-domain connecting loop is coloured red, the back-side loop is coloured violet and the C-terminus is coloured green. (*b*) Expression and purification of EhPCNA: 15% SDS–PAGE showing samples taken during the expression and purification protocols. Arrows indicate the positions of EhPCNA and the histidine tag. (*c*) Purification of EhPCNA after removal of the histidine tag: 15% SDS–PAGE analysis of EhPCNA purification by size-exclusion chromatography.

Table 2
PIP-box-containing proteins from *E. histolytica* in comparison with related proteins from *H. sapiens*.

The consensus PIP-box sequence is **Q***XX*(**M/L/I**)*XX*(**F/Y**)(**F/Y**).

Protein	*H. sapiens*	GenBank ID	*E. histolytica*	Pathema locus	GenBank ID
DNA ligase I	1 M**Q**RS**I**MS**FF**HPKK	AAA59518.1	1 MSKR**Q**S**L**SR**FF**KPA	EHI_111060	XP_657595.1
FEN1	332 GST**Q**GR**L**DD**FF**KVT	NP_004102.1	328 KKA**Q**GR**L**DS**FF**NVK	EHI_099740	XP_651270.1
hMYH	509 MG**Q**QV**L**DN**FF**RSH	U6339	294 QI**V**HK**E**GD**FI**VLN	EHI_010700	XP_653060.1
hMSH3	18 PAR**Q**AV**L**SR**FF**QS	NP_002430.3	1 MSI**Q**SK**L**KF**KD**GMF	EHI_193340	XP_651951.1

even after extensive washing (Fig. 1*b*, lanes 5–7). After one purification step EhPCNA was approximately 95% pure (Fig. 1*b*, lane 8); however, a second purification step with a DEAE column was included in order to further increase the purity of the EhPCNA before its crystallization. The histidine tag of recombinant EhPCNA was efficiently cleaved using PreScission protease (Fig. 1*b*, lane 9). To remove the histidine tag and to eliminate putative DNA ligase activity arising from bacterial contamination, a third chromatographic step using a Superdex 75 column was performed. This gel-filtration column efficiently separates recombinant EhPCNA from its histidine tag (Fig. 1*c*). The recombinantly expressed EhPCNA migrates as a protein of approximately 28 kDa on an SDS–PAGE gel after removal of the histidine tag (Fig. 1*c*, lanes 4 and 5).

3.2. EhPCNA assembles as a functional homotrimer

To determine the functional assembly of EhPCNA, we compared its elution profile using molecular-mass standards in a Superdex 200 size-exclusion chromatography column. EhPCNA eluted in two fractions corresponding to molecular masses of 44 and 108 kDa. In relation to the elution profiles of the standards, the molecular masses are close to the theoretically expected molecular masses of 28.5 kDa for monomeric EhPCNA and 85.5 kDa for trimeric EhPCNA (Fig. 2*a*). The anomalous migration of monomeric EhPCNA may be explained by the fact that EhPCNA is a rod-shaped protein. The more abundant peak that corresponds to 108 kDa may correspond to trimeric EhPCNA (Fig. 2*a*). The ratio of trimeric and monomeric PCNA species has been found to be dependent on protein concentration (Zhang *et al.*, 1995). However, in our hands the relationship between the two species was not altered at different protein concentrations, indicating the possibility that monomeric EhPCNA may correspond to a population that is unable to form trimers. Thus, according to the proportions of trimeric and monomeric species in a Superdex 200 gel-filtration column EhPCNA assembles mainly as a homotrimer in solution, although approximately 8% of native EhPCNA exists as a monomer. This result is in agreement with previous reports that indicate that PCNAs from other species exist as a mixture of oligomers in solution (Naryzhny *et al.*, 2005, 2006).

3.3. EhPCNA enhances the binding of EhDNAligI to nicked DNA and stimulates its nick-sealing reaction

PCNA stimulates the interaction between its protein partners and DNA mainly through its PIP box. DNA-binding

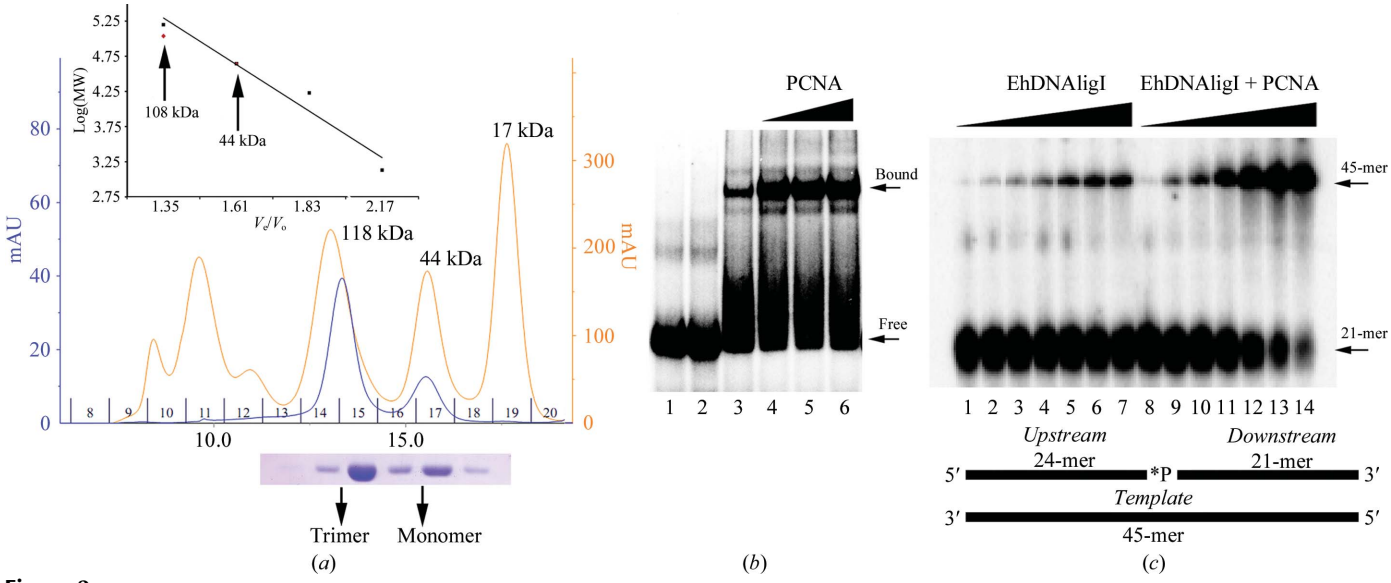

Figure 2
Biochemical characterization of EhPCNA. (*a*) EhPCNA is a homotrimer according to gel filtration. Purified EhPCNA was applied onto a Superdex 200 column depicting its oligomeric state by gel filtration. (*b*) EhPCNA stimulates the binding of EhDNAligI to nicked DNA. The migration of the complex and free probe are indicated by arrows. (*c*) EhPCNA stimulates the nick-sealing activity of EhDNAligI. The reaction time course shows a comparison of the nick-sealing activity of EhDNAligI in the absence (lanes 1–7) and in the presence of a molar excess (lanes 8–15) of EhPCNA. Nicked double-stranded DNA substrate is depicted at the bottom of the panel.

proteins from protozoan parasites often diverge from those of other metazoans. For instance, DNA ligase I from Apicomplexa does not contain a recognizable PIP box (Buguliskis *et al.*, 2007), whereas DNA ligase I from *E. histolytica* contains a *bona fide* PIP box (EhDNAligI; Cardona-Felix *et al.*, 2010). Several other proteins from this parasite contain this interacting motif but with deviations from the consensus sequence (Table 2). Thus, as in other biological systems, EhDNAligI may form a physical interaction with EhPCNA during the sealing of Okazaki fragments (Tom *et al.*, 2001; Mayanagi *et al.*, 2009; Pascal *et al.*, 2006). In order to detect a putative physical interaction between EhPCNA and EhDNAligI, we performed an EMSA analysis in which we tested whether EhPCNA is able to stimulate the binding of EhDNAligI to nicked DNA (Fig. 2*b*). It has been observed that HsPCNA increases the formation of a complex with nicked DNA substrate (Levin *et al.*, 2000; Tom *et al.*, 2001). As shown for other PCNAs, EhPCNA is unable to bind nicked DNA because it slides from linear DNA (Yao *et al.*, 1996), whereas EhDNAligI binds tightly to phosphorylated nicked DNA as observed from the band shift of the labelled probe (Cardona-Felix *et al.*, 2010; Fig. 2*b*, lanes 2 and 3). The presence of twofold, fourfold and eightfold molar excesses of EhPCNA increased the population of EhDNAligI stably bound to nicked DNA (Fig. 2*b*, lanes 4–6). However, increasing concentrations of EhPCNA do not produce a supershift of the EhDNAligI–DNA complex. The same phenomenon has been observed for the interaction of HsPCNA with HsFEN1 and HsDNAlig I (Tom *et al.*, 2000, 2001), indicating that in all cases PCNA dissociates from the ligase or nuclease complex during electrophoresis and suggesting that the role of PCNA is to facilitate the initial binding of DNA ligase I and FEN1 endonuclease to nicked DNA. The complex between human DNA ligase I and HsPCNA stimulates the nick-sealing activity of DNA ligase I by threefold to fivefold (Tom *et al.*, 2001). In order to test whether EhPCNA also stimulates EhDNAligI nick-sealing activity, we performed a time-course experiment with a 50-fold molar excess of EhPCNA with respect to EhDNAligI. As shown in Fig. 2(*c*), the nick-sealing reaction is stimulated by approximately fourfold in the experiment with added EhPCNA. EhPCNA is not able to perform nick sealing, nor is it able to stimulate a non-PIP box contained in T4 DNA ligase (Cardona-Felix, data not shown). Thus, EhPCNA causes specific stimulation of EhDNAligI nick-sealing activity throughout the reaction course. The similarity in ring size between HsPCNA and HsDNAligI suggests that PCNA may facilitate the transition of HsDNAligI from an extended conformation to a closed conformation (Pascal *et al.*, 2004). However, from these kinetic experiments is not possible to conclude whether the observed stimulation of the nick-sealing activity of EhDNAlig I arises from an increase in substrate binding or from an influence on the catalytic step, as has been suggested for DNA nucleases and PCNA (Hutton *et al.*, 2008, 2010).

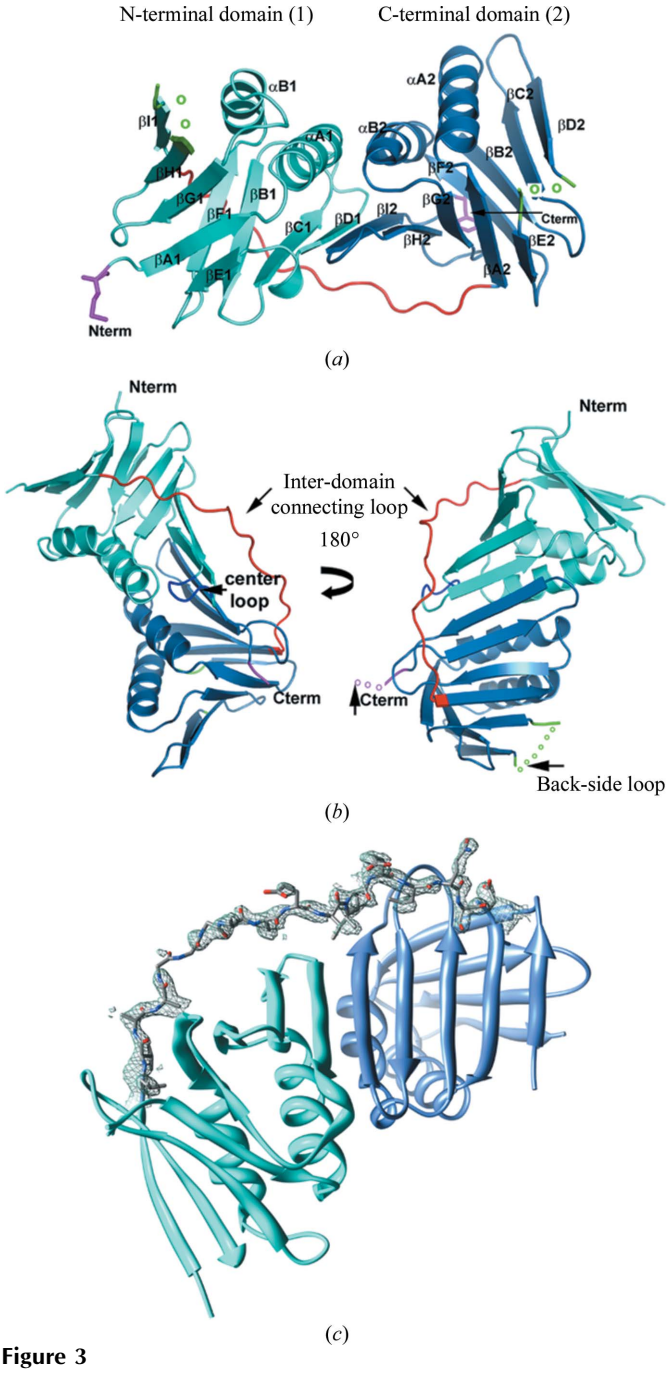

Figure 3
Monomeric EhPCNA is composed of two *β*-*α*-*β*-*β*-*β*-*β*-*α*-*β*-*β*-*β* domains (cyan and blue) linked by an inter-domain connecting loop (IDCL; red). (*b*) Crystal structure of monomeric EhPCNA. The inter-domain connecting loop is coloured red, the centre loop is coloured blue, the back-side loop is coloured green and the C-terminus is coloured magenta. (*c*) Difference electron σ_A-weighted $F_o - F_c$ annealed electron-density map of the IDCL region contoured at 1.5σ. The electron-density map is superimposed on the final secondary structure of the IDCL from residues Leu119 to Ser134. The N-terminal and C-terminal subdomains of EhPCNA are depicted as ribbons and are coloured cyan and blue, respectively.

3.4. Crystal structure of EhPCNA

EhPCNA crystallized in space group *H*3, contained one molecule per asymmetric unit and diffracted to a resolution of 2.4 Å (Table 1). The crystal structure of EhPCNA was solved by molecular replacement using the structure of human PCNA

(Gulbis *et al.*, 1996) as a search model. The structure of PCNA consists of two structural domains with β-α-β-β-β-β-α-β-β-β topology joined by an inter-domain connecting loop (IDCL; Fig. 3*a*). PCNA interacts with several proteins *via* specific structural elements. Although the majority of the interactions are mediated by the IDCL, the centre loop is important for binding to cyclin D and the C-terminal extension is important for binding to DNA polymerase ε and replication factor C (reviewed in Maga & Hubscher, 2003; Fig. 3*b*). Continuous electron density was present during amino-acid tracing of EhPCNA; however, regions in which no detectable electron density was observed included residues 107–110, 186–193 and the C-terminal residues 255–262. Residues Asp123, Ala124 and Asp125 were modelled as glycines because their side chains could not be reliably modelled (Fig. 3*b*). The IDCL is well defined in the EhPCNA crystal structure as observed in a simulated-annealing $F_o - F_c$ electron-density map, although this loop could not be modelled in other PCNAs (Matsumiya *et al.*, 2001; Figs. 3*b* and 3*c*). The architecture of the IDCL is similar to that observed in other PCNAs, indicating that protein partners may interact with EhPCNA *via* a consensus PIP box. Thus, EhPCNA may mediate similar protein interactions to PCNAs from metazoans and yeast.

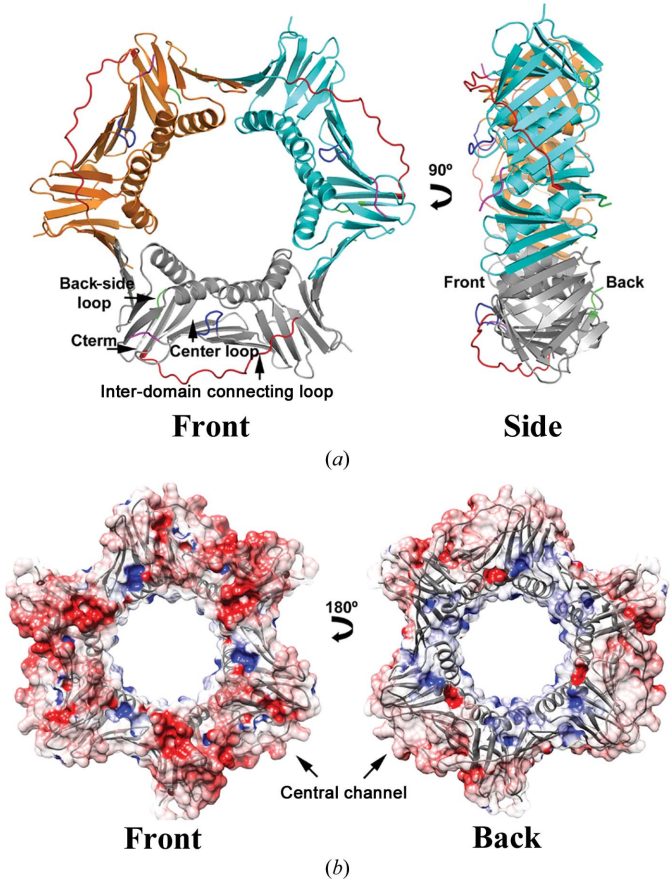

Figure 4
Crystal structure of trimeric EhPCNA. (*a*) Ribbon representation of EhPCNA as a crystallographic trimer. Subunit *A* is coloured grey, subunit *B* is coloured blue and subunit *C* is coloured orange. The IDCL is coloured red. (*b*) Electrostatic surface representation of EhPCNA. Surface representation of EhPCNA showing its calculated electrostatic potential. Positively charged segments are coloured blue and negatively charged segments are coloured red. The highly positively charged central channel corresponds to the double-stranded DNA-binding region.

3.5. EhPCNA is a positively charged homotrimeric ring

EhPCNA forms a trimer around a crystallographic three-fold axis. The trimer is composed of 12 α-helices in the inner surface and six β-sheets in the outer surface. The assembly of the trimer produces a central channel of approximately 34 Å in diameter; within this assembly the inner 12 α-helices are oriented to form a central hole that can encircle double-stranded DNA. As in other PCNAs, EhPCNA has distinct front and back sides, although PCNA interacts with protein partners using only its front side (Fig. 4*a*). EhPCNA is an acidic protein with an isoelectric point of 4.63; however, the central PCNA channel contains 27 positively charged amino acids (arginines and lysines) located in αA1, αB1, αA2 and αB2 as depicted in a molecular-surface representation of its electrostatic potential (Fig. 4*b*). These positively charged amino acids are important for nonspecific interactions between PCNA and the sugar–phosphate backbone of DNA, as demonstrated by electron microscopy and X-ray crystallography (McNally *et al.*, 2010; Mayanagi *et al.*, 2009).

3.6. Structural comparison of EhPCNA

Monomeric EhPCNA superimposes with PCNA monomers from *H. sapiens*, *S. cerevisiae* and *A. thaliana* with r.m.s.d.s of 1.3, 1.7 and 1.8 Å, respectively, using the secondary-structure matching (*SSM*) algorithm (Krissinel & Henrick, 2004) in *Coot* (Emsley & Cowtan, 2004). The main structural deviations are located at the IDCL (Fig. 5*a*). The structural differences at the IDCL may be related to its inherent flexibility compared with the relative rigidity of the N- and C-terminal domains. The PCNAs from *S. cerevisiae*, *H. sapiens* and *A. thaliana* maintain the interface between subunits by an interaction between two antiparallel β-strands. Specifically, this interaction is formed by hydrogen-bond backbone interactions between βD2 of one monomer and βI of the other monomer. The conserved subunit–subunit interaction in HsPCNA involves the formation of eight hydrogen bonds (Krishna *et al.*, 1994; Gulbis *et al.*, 1996; Fig. 5*b*). In EhPCNA we can identify seven hydrogen bonds between adjacent subunits, although the N-terminus of EhPCNA βD2 is not reliable in the electron-density maps, suggesting the possibility that up to eight hydrogen bonds maintain the subunit interface of EhPCNA (Fig. 5*b*). This in contrast to the subunit interface of *Pyrococcus furiosus*, which only contains four hydrogen bonds (Matsumiya *et al.*, 2001). A global structural comparison of PCNAs indicates that the overall architecture of EhPCNA is similar to the structures of PCNAs from human, yeast and plants, demonstrating the conservation of this architecture in different eukaryotes (Fig. 5*c*). This is in agreement with previous work, which indicated that the processivity mechanisms are conserved between archaea and eukaryotes (Matsumiya *et al.*, 2001).

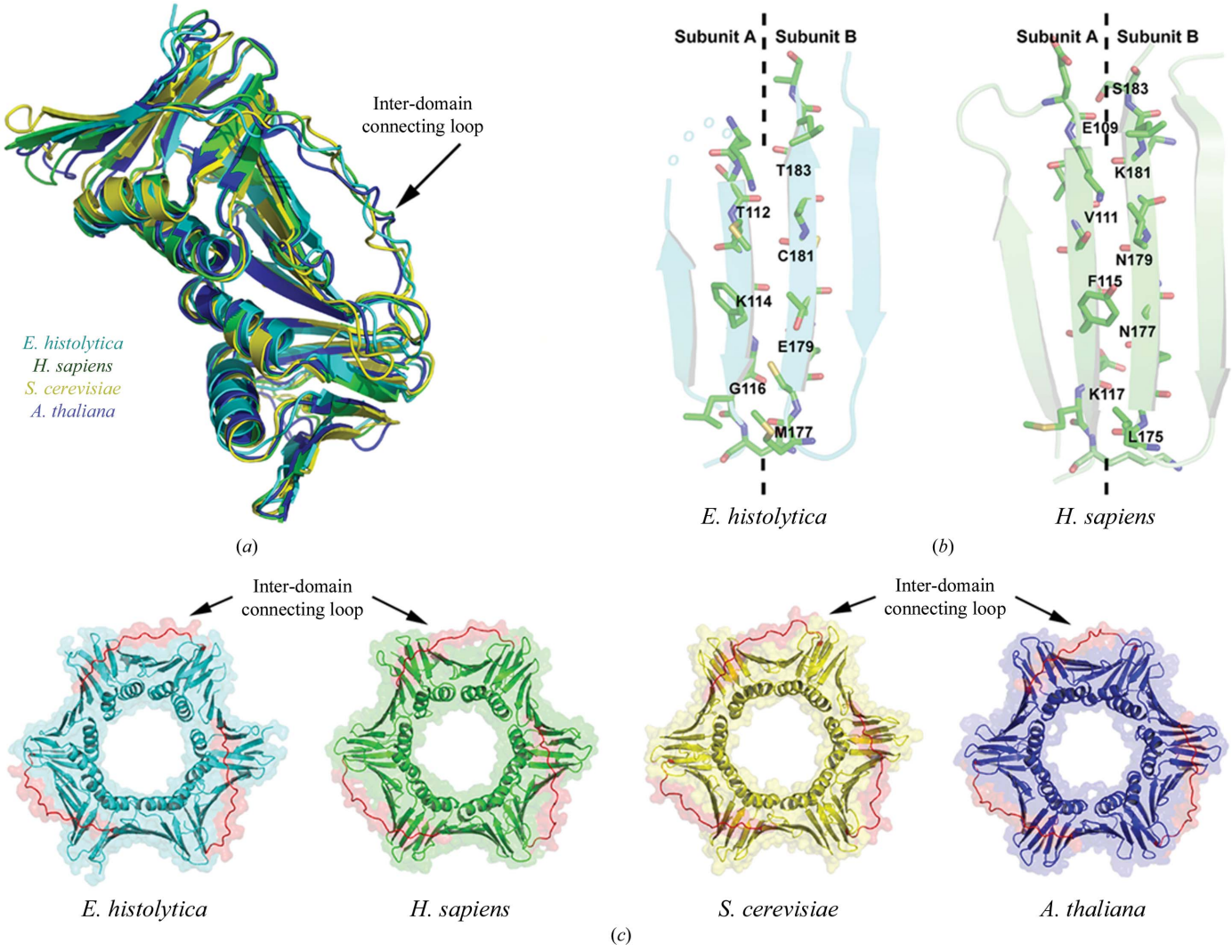

Figure 5
(*a*) Superimposition of monomeric EhPCNA (blue) with monomeric PCNAs from *S. cerevisiae* (gold), *H. sapiens* (green) and *A. thaliana* (blue). (*b*) Intermolecular β-sheet interface of EhPCNA in comparison to that of PCNA from *H. sapiens*. Side chains involved in hydrogen-bond formation between β-strands of adjacent subunits are shown as sticks and hydrogen bonds are presented as dotted lines. (*c*) Global structural comparison of PCNA from *E. histolytica* with PCNAs from *H. sapiens*, *S. cerevisiae* and *A. thaliana*.

4. Conclusions

We have determined the crystal structure and biochemically characterized the processivity factor PCNA from *E. histolytica*. This protein assembles as a homotrimer in solution and is able to bind and stimulate the activity of EhDNAligI, a *bona fide* PIP-box-containing protein. The crystal structure of EhPCNA illustrates a conservation in fold, subunit–subunit interactions and electrostratic distribution between the structure of PCNA from *E. histolytica* and PCNA structures from higher eukaryotes. Our data indicate that PCNA is a structural platform that is highly conserved between parasitic protozoans, yeast, plants and metazoans.

This work was supported by grants from CONACYT-México (No. 128647) and the International Society for Infectious Diseases to LGB. CSCF and SLG are supported by a doctoral and a postdoctoral fellowship from CONACYT-México, respectively. We thank Dr Elisa Azuara Liceaga for kindly providing genomic DNA and Varinia Lopez for her assistance with the hidden Markov analysis. We are indebted to Dr Alfredo Torres-Larios for data collection and the staff members of Argonne National Laboratory for use of beamline 21-ID-D.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Almendral, J. M., Huebsch, D., Blundell, P. A., Macdonald-Bravo, H. & Bravo, R. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 1575–1579.
Bauer, G. A. & Burgers, P. M. (1990). *Nucleic Acids Res.* **18**, 261–265.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Buguliskis, J. S., Casta, L. J., Butz, C. E., Matsumoto, Y. & Taraschi, T. F. (2007). *Mol. Biochem. Parasitol.* **155**, 128–137.
Bunkóczi, G. & Read, R. J. (2011). *Acta Cryst.* D**67**, 303–312.
Cardona-Felix, C. S., Pastor-Palacios, G., Cardenas, H., Azuara-Liceaga, E. & Brieba, L. G. (2010). *Mol. Biochem. Parasitol.* **174**, 26–35.

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.
Chia, N., Cann, I. & Olsen, G. J. (2010). *PLoS One*, **5**, e10866.
Clark, C. G. *et al.* (2007). *Adv. Parasitol.* **65**, 51–190.
Da Silva, M. & Upton, C. (2009). *PLoS One*, **4**, e5479.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
Dionne, I., Nookala, R. K., Jackson, S. P., Doherty, A. J. & Bell, S. D. (2003). *Mol. Cell*, **11**, 275–282.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Guerini, M. N., Que, X., Reed, S. L. & White, M. W. (2000). *Mol. Biochem. Parasitol.* **109**, 121–131.
Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. (1996). *Cell*, **87**, 297–306.
Hutton, R. D., Craggs, T. D., White, M. F. & Penedo, J. C. (2010). *Nucleic Acids Res.* **38**, 1664–1675.
Hutton, R. D., Roberts, J. A., Penedo, J. C. & White, M. F. (2008). *Nucleic Acids Res.* **36**, 6720–6727.
Ivanov, I., Chapados, B. R., McCammon, J. A. & Tainer, J. A. (2006). *Nucleic Acids Res.* **34**, 6023–6033.
Kilbey, B. J., Fraser, I., McAleese, S., Goman, M. & Ridley, R. G. (1993). *Nucleic Acids Res.* **21**, 239–243.
Kontopidis, G., Wu, S., Zheleva, D., Taylor, P., McInnes, C., Lane, D., Fischer, P. & Walkinshaw, M. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 1871–1876.
Krishna, T. S., Kong, X.-P., Gary, S., Burgers, P. M. & Kuriyan, J. (1994). *Cell*, **79**, 1233–1243.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Kumar, D., Minocha, N., Rajanala, K. & Saha, S. (2009). *Microbiology*, **155**, 3748–3757.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Levin, D. S., McKenna, A. E., Motycka, T. A., Matsumoto, Y. & Tomkinson, A. E. (2000). *Curr. Biol.* **10**, 919–922.
Maga, G. & Hubscher, U. (2003). *J. Cell Sci.* **116**, 3051–3060.
Matsumiya, S., Ishino, Y. & Morikawa, K. (2001). *Protein Sci.* **10**, 17–23.
Mayanagi, K., Kiyonari, S., Saito, M., Shirai, T., Ishino, Y. & Morikawa, K. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 4647–4652.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McNally, R., Bowman, G. D., Goedken, E. R., O'Donnell, M. & Kuriyan, J. (2010). *BMC Struct. Biol.* **10**, 3.
Moldovan, G. L., Pfander, B. & Jentsch, S. (2007). *Cell*, **129**, 665–679.
Naryzhny, S. N. (2008). *Cell. Mol. Life Sci.* **65**, 3789–3808.
Naryzhny, S. N., Desouza, L. V., Siu, K. W. & Lee, H. (2006). *Biochem. Cell Biol.* **84**, 669–676.
Naryzhny, S. N., Zhao, H. & Lee, H. (2005). *J. Biol. Chem.* **280**, 13888–13894.
Neuwald, A. F. & Poleksic, A. (2000). *Nucleic Acids Res.* **28**, 3570–3580.
Pascal, J. M., O'Brien, P. J., Tomkinson, A. E. & Ellenberger, T. (2004). *Nature (London)*, **432**, 473–478.
Pascal, J. M., Tsodikov, O. V., Hura, G. L., Song, W., Cotner, E. A., Classen, S., Tomkinson, A. E., Tainer, J. A. & Ellenberger, T. (2006). *Mol. Cell*, **24**, 279–291.
Patterson, S., Whittle, C., Robert, C. & Chakrabarti, D. (2002). *Biochem. Biophys. Res. Commun.* **298**, 371–376.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Stanley, S. L. (2003). *Lancet*, **361**, 1025–1034.
Stoimenov, I. & Helleday, T. (2009). *Biochem. Soc. Trans.* **37**, 605–613.
Strzalka, W., Oyama, T., Tori, K. & Morikawa, K. (2009). *Protein Sci.* **18**, 1072–1080.
Tom, S., Henricksen, L. A. & Bambara, R. A. (2000). *J. Biol. Chem.* **275**, 10498–10505.
Tom, S., Henricksen, L. A., Park, M. S. & Bambara, R. A. (2001). *J. Biol. Chem.* **276**, 24817–24825.
Tsurimoto, T. (1998). *Biochim. Biophys. Acta*, **1443**, 23–39.
Williams, G. J., Johnson, K., Rudolf, J., McMahon, S. A., Carter, L., Oke, M., Liu, H., Taylor, G. L., White, M. F. & Naismith, J. H. (2006). *Acta Cryst.* F**62**, 944–948.
Yao, N., Turner, J., Kelman, Z., Stukenberg, P. T., Dean, F., Shechter, D., Pan, Z.-Q., Hurwitz, J. & O'Donnell, M. (1996). *Genes Cells*, **1**, 101–113.
Zhang, P., Zhang, S.-J., Zhang, Z., Woessner, J. F. & Lee, M. Y. (1995). *Biochemistry*, **34**, 10703–10712.