



X-ray crystal structure of proliferating cell nuclear antigen 1 from *Aeropyrum pernix*

Takahiro Yamauchi,^{a,b} Makiko Kikuchi,^{c,d,‡} Yasuhito Iizuka^{a,d} and Masaru Tsunoda^{a,c,d,*}

^aGraduate School of Life Science and Technology, Iryo Sosei University, Iwaki, Fukushima, Japan, ^bDepartment of Pharmacy, Fukushima Rosai Hospital, Iwaki, Fukushima, Japan, ^cGraduate School of Science and Engineering, Iryo Sosei University, Iwaki, Fukushima, Japan, and ^dFaculty of Pharmacy, Iryo Sosei University, Iwaki, Fukushima, Japan.

*Correspondence e-mail: masaru@isu.ac.jp

Proliferating cell nuclear antigen (PCNA) plays a critical role in DNA replication by enhancing the activity of various proteins involved in replication. In this study, the crystal structure of ApePCNA1, one of three PCNAs from the thermophilic archaeon *Aeropyrum pernix*, was elucidated. ApePCNA1 was cloned and expressed in *Escherichia coli* and the protein was purified and crystallized. The resulting crystal structure determined at 2.00 Å resolution revealed that ApePCNA1 does not form a trimeric ring, unlike PCNAs from other domains of life. It has unique structural features, including a long inter-domain-connecting loop and a PIP-box-like sequence at the N-terminus, indicating potential interactions with other proteins. These findings provide insights into the functional mechanisms of PCNAs in archaea and their evolutionary conservation across different domains of life. A modified medium and protocol were used to express recombinant protein containing the *lac* operon. The expression of the target protein increased and the total incubation time decreased when using this system compared with those of previous expression protocols.

1. Introduction

DNA replication is a key event in the cell cycle that precedes cell division. It involves various proteins, including DNA-dependent DNA polymerases (Yao & O'Donnell, 2016). Among these factors, proliferating cell nuclear antigen (PCNA) forms a trimeric ring (sliding clamp) around DNA and directly enhances the activity of proteins involved in replication (Maga & Hübscher, 2003). Several enzymes bind to PCNA via the PCNA-interacting peptide-box (PIP-box) motif, including DNA-dependent DNA polymerases, flap endonucleases and DNA ligases involved in DNA replication (MacNeill, 2016; Moldovan *et al.*, 2007; Winter & Bunting, 2012). PCNA has a PIP-box binding site near the C-terminus that is oriented in the direction of DNA replication (Horsfall *et al.*, 2021). Sliding clamps can be considered to be DNA scaffolds that allow specific enzymes (such as DNA polymerase) to bind and catalyze DNA replication.

Sliding clamps are involved in cell proliferation and are frequently expressed in cancer cells (Chen *et al.*, 2021). They have attracted attention as a potential target for disease diagnosis and treatment (Wendel *et al.*, 2023). Furthermore, bacterial β -subunit (a monomer of the bacterial sliding-clamp complex) inhibitors have been explored to develop new antimicrobial agents (Liu *et al.*, 2021; Nedal *et al.*, 2020; Pandey *et al.*, 2017). The sliding clamp is one of the earliest reported proteins to have its three-dimensional structure

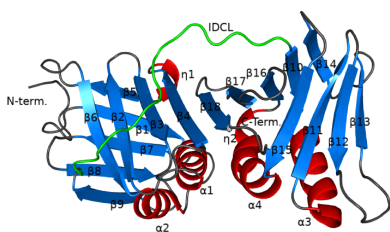


Table 1
Macromolecule-production information.

Source organism	<i>A. pernix</i> K1 ^T
Expression vector	pET-11a
Expression host	<i>E. coli</i> Rosetta-gami (DE3)
Complete amino-acid sequence of the construct produced	MSSEATLDSEFTDYKAMFRYEAKVFKEL VDSVSKILDEGLFIITGEGRLRLGMDF ARVALVDIEIPSSSFDFYMGADVERV ELGVNMETLKGVARAKKGDQLEVRVR EDKVLFFIVESVVLRRYLLPNLEIVDV PEDISLEFDATATVIADVKKTLRDVE LVGDIVEFDAGEDYLSIRSVGPERRRV ETRLTRESPALIDLEVKEPATSRVDVG YLKRMGLGVAKIAESIELSFSTDKPLKM VFKSPDGSRVTYLLAPSTG

determined, and has been described in many species, including humans, eukaryotes and bacteria (Georgescu *et al.*, 2008; Gulbis *et al.*, 1996; Williams *et al.*, 2006). The sliding clamp is widely present in the three major domains of living organisms, forming a β -clamp dimer in bacteria and a PCNA trimer in eukaryotes. In these two domains, a single PCNA or β -clamp gene is present in the genome, and the sliding clamp refers to a homo-complex. In the archaeal domain, Euryarchaeota form homotrimers encoded by a single gene, whereas some species belonging to the Crenarchaeota family have multiple PCNAs that form heterotrimers and homotrimers (MacNeill, 2016).

Aeropyrum pernix is a crenarchaeon that inhabits high-temperature and aerobic conditions and was discovered on Kodakara-Jima Island, Japan (Sako *et al.*, 1996). Analysis of the *A. pernix* genome showed that the sequence contains three PCNA genes (Kawarabayasi *et al.*, 1999): ApePCNA1 (*APE_0162*), ApePCNA2 (*APE_0441.1*) and ApePCNA3 (*APE_2182*) in the KEGG database (Kanehisa *et al.*, 2017). Of the *A. pernix* genes, approximately 30% have an ATG start codon, approximately 50% have a TTG start codon and 20% have a GTG start codon (Yamazaki *et al.*, 2006). The start codon of the *APE_0162* gene is GTG, and the molecular weights of the recombinant and native (cell-extract) proteins are identical on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE; Daimon *et al.*, 2002). The sliding clamp of *A. pernix* forms heterotrimers and enhances polymerase activity in the ApePCNA1–ApePCNA2–ApePCNA3 and ApePCNA2–ApePCNA3 complexes (Imamura *et al.*, 2007). ApePCNA1 is expressed in *A. pernix* cells but is not essential for its *in vitro* activity. However, the specific activity of ApePCNA1 has not been determined, and its three-dimensional structure has not yet been analyzed. In this study, we determined the crystal structure of ApePCNA1 and described its features.

2. Materials and methods

2.1. Macromolecule production

The *APE_0162* gene encoding ApePCNA1 from *A. pernix* K1^T was amplified from the genomic DNA of strain NBRC 100138^T (obtained from NBRC, NITE, Kisarazu, Japan) using polymerase chain reaction (PCR). The PCR product was

Table 2
Crystallization.

Method	Vapor diffusion, hanging drop
Plate type	VDX plate
Temperature (K)	298
Protein concentration (mg ml ⁻¹)	10
Buffer composition of protein solution	20 mM Tris–HCl pH 8.0
Composition of reservoir solution	McIlvaine buffer, lithium sulfate, propan-2-ol
Volume and ratio of drop	4 μ l, 1:1
Volume of reservoir (μ l)	1000

treated with the NdeI and BglII restriction enzymes and was subcloned into pET-11a vector (Merck KGaA, Darmstadt, Germany). The plasmid was introduced into *Escherichia coli* Rosetta-gami (DE3) cells by electroporation on ice for 4 ms at 1.80 kV and 25 μ F; the cells were incubated in Super Optimal broth with catabolite repression at 310 K for 30 min and spread onto Luria–Bertani (LB) agar containing ampicillin and chloramphenicol. A colony was selected and grown as a primary culture in 15 ml LB broth (Lennox) containing 50 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol for 9 h at 310 K. Subsequently, 7.5 ml bacterial suspension from the primary culture was grown in 1 l 1% Bacto tryptone (Becton, Dickinson & Co., Franklin Lakes, New Jersey, USA), 0.5% peptone (Kyokuto Pharmaceutical Industrial, Tokyo, Japan), 0.5% sodium chloride, 50 μ g ml⁻¹ ampicillin and 34 μ g ml⁻¹ chloramphenicol for 12 h at 310 K. The cells were cultured to an *A*₆₀₀ of 0.6, and expression of the *APE_0162* gene was induced by adding 10% lactose to a final concentration of 0.4%. Cultivation continued for a further 4 h at 310 K.

The cells were harvested, resuspended in 20 mM Tris–HCl pH 8.0, 50 mM NaCl, disrupted by sonication and heated for 30 min at 353 K. Heat-resistant fractions were obtained by centrifugation (8000g for 15 min at 277 K). Soluble proteins were precipitated using 40% saturated ammonium sulfate. The precipitate was resuspended in 10 ml 20 mM Tris–HCl pH 8.0 and dialyzed three times against 1 l 20 mM Tris–HCl pH 8.0 containing 200 mM NaCl (the three dialysis times were 30 min, 1 h and overnight). The dialysate was loaded onto a HiTrap Q HP column (5 ml; Cytiva, Marlborough, Massachusetts, USA) and ApePCNA1 protein was eluted with 200–500 mM sodium chloride.

The eluted protein was concentrated to 100 μ l by ultrafiltration using an Amicon Ultra-15 10K centrifugal filter (Merck) and applied onto a HiLoad 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated with 20 mM Tris–HCl pH 8.0, 200 mM sodium chloride. The macromolecule-production process is summarized in Table 1.

2.2. Crystallization

Purified ApePCNA1 was concentrated to 10 mg ml⁻¹ via ultrafiltration using an Amicon Ultra 10K centrifugal filter. The buffer solution was changed to 20 mM Tris–HCl pH 8.0 at the same time. Crystallization was performed by hanging-drop vapor diffusion in a VDX Plate with sealant (Hampton Research, Aliso Viejo, California, USA) at 298 K. The initial

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Beamline BL-1A, Photon Factory
Wavelength (Å)	1.1000
Temperature (K)	100
Detector	Dectris EIGER X 4M
Crystal-to-detector distance (mm)	92.260
Rotation range per image (°)	1
Total rotation range (°)	180
Exposure time per image (s)	10
Space group	<i>P</i> 4 ₃ 2 ₁ 2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	69.03, 69.03, 120.86
α , β , γ (°)	90, 90, 90
Mosaicity (°)	0.138
Resolution range (Å)	48.86–2.00 (2.05–2.00)
Total No. of reflections	258157 (12621)
No. of unique reflections	20481 (1056)
Completeness (%)	100.0 (100.0)
Multiplicity	12.5 (13.4)
$\langle I/\sigma(I) \rangle$	34.4 (15.0)
$R_{\text{merge}}^{\dagger}$	0.048 (0.164)
$R_{\text{r.i.m.}}^{\ddagger}$	0.051 (0.177)
$R_{\text{p.i.m.}}^{\S}$	0.019 (0.067)
Overall <i>B</i> factor from Wilson plot (Å ²)	19.354

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl . $\ddagger R_{\text{r.i.m.}} = \sum_{hkl} \{ [N(hkl) / (N(hkl) - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl) \}$, where $N(hkl)$ is the data multiplicity. $\S R_{\text{p.i.m.}} = \sum_{hkl} \{ [1 / (N(hkl) - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl) \}$.

crystallization conditions were screened using Wizard Classics 1 and 2 (Rigaku Reagents, The Woodlands, Texas, USA) and Index (Hampton Research). Finally, colorless, origami shuriken-shaped (paper ninja star-like) crystals grew after two days in 100 mM McIlvaine buffer pH 4.2–4.6, 100–250 mM lithium sulfate, 7% propan-2-ol. The crystallization conditions are summarized in Table 2.

2.3. Data collection and processing

The ApePCNA1 crystals were collected and packed into a Universal V1-Puck at 77 K in liquid nitrogen. Diffraction data for ApePCNA1 were collected on BL-1A at the Photon Factory, Tsukuba, Japan. The diffraction images were processed (indexed, integrated, scaled and merged) using *XDSGUI* (Brehm *et al.*, 2023). The data-collection statistics are summarized in Table 3.

2.4. Structure solution and refinement

Phases were determined by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) and the PCNA model from *Pyrococcus furiosus* (PDB entry 1iz4; Matsumiya *et al.*, 2001). The model was rebuilt using *ARP/wARP Classic* (Langer *et al.*, 2008) and 92 water molecules were added. The structure was modified using *Coot* (Emsley *et al.*, 2010) and refined using *REFMAC5* (Murshudov *et al.*, 2011). These processes were performed within the *CCP4i* interface (Potterton *et al.*, 2003) using the *CCP4* suite (Agirre *et al.*, 2023). The ApePCNA1 structure was validated using *Coot* and the wwPDB validation services (Young *et al.*, 2017) during deposition. Figures were constructed using *PyMOL* ([https://](https://www.pymol.org/)

Table 4

Structure refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	48.86–2.00 (2.052–2.000)
Completeness (%)	99.9
σ Cutoff	2.00
No. of reflections, working set	19446 (1403)
No. of reflections, test set	991 (69)
Final $R_{\text{work}}^{\dagger}$	0.185 (0.196)
Final $R_{\text{free}}^{\ddagger}$	0.226 (0.251)
Cruickshank DPI	0.1220
No. of non-H atoms	
Protein	1959
Ligand	0
Water	63
Total	2022
R.m.s. deviations	
Bond lengths (Å)	0.014
Angles (°)	1.583
Average <i>B</i> factors (Å ²)	
Protein	29.6
Water	32.6
Ramachandran plot	
Most favored (%)	98.8
Allowed (%)	1.2

$\dagger R_{\text{work}} = \sum_{hkl} | |F_{\text{obs}}| - |F_{\text{calc}}| | / \sum_{hkl} |F_{\text{obs}}|$. $\ddagger R_{\text{free}}$ is calculated the same way as R_{work} for data omitted from refinement (5% of the reflections in the data set).

www.pymol.org/). The refinement statistics are summarized in Table 4.

3. Results and discussion

3.1. Sequence comparison of PCNAs

The amino-acid sequence identities between ApePCNA1 and ApePCNA2, between ApePCNA2 and ApePCNA3 and between ApePCNA3 and ApePCNA1 were 28.16%, 21.81% and 28.57%, respectively. The average identity among the three PCNAs was 26.18% (Fig. 1). The ApePCNA1 sequence was approximately ten residues longer at the N-terminus than the other two sequences. Similarly, PCNAs from *Saccharolobus solfataricus* contained a long N-terminal region, as in SsoPCNA3 (PDB entry 2ix2, chain C; protein ID AAK40734.1; Williams *et al.*, 2006). However, the homology between ApePCNA1 and SsoPCNA3 (24.9%) was lower than those between the ApePCNAs, and only Met1 matched the N-terminal region before the second methionine residue of each PCNA (Met17 in ApePCNA1 and Met16 in SsoPCNA3). These long N-terminal amino-acid sequences (ApePCNA1, Q9YFT8; SsoPCNA3, P57765) are considered to be errors in the initial positions in the UniProt database (The UniProt Consortium, 2023). However, as discussed later, they can be used as anchor sequences.

3.2. Evaluating the enhanced expression system

ApePCNA1 was effectively expressed in an amino-acid-rich medium that differs from the Lennox LB medium (Lennox, 1955) that is widely used in protein-expression systems. SDS-PAGE showed that the expression in 1% tryptone, 0.5% peptone, 0.5% NaCl broth (TP broth) was higher than that in Lennox LB medium (Fig. 2). Furthermore, the expression of

E. coli-derived proteins (other than the target proteins) was significantly lower in this medium than in conventional LB medium. This medium is mostly composed of amino acids, which limit energy and nucleic acid metabolism. For example, the expression of these metabolism-related proteins is reduced owing to the extremely low amount of nucleic acids in the medium, since the *de novo* pathway is dominant in nucleic acid metabolism and the salvage pathway is of little significance. Therefore, the specificity of the expression levels of the target proteins may be attributed to the reduced expression of proteins required for their metabolism. We examined the timing of lactose addition by inducing protein expression using the *lac* operon. Considering the simplicity of this technique and the risk of contamination during incubation, it is preferable to mix lactose with the culture medium and sterilize it at the time of use, rather than separately preparing and sterilizing the culture medium and lactose solution before mixing them. However, the growth rate of the host *E. coli* was suppressed when ApePCNA1 was overexpressed in a medium containing lactose, and the incubation time required for the solution to reach visually observable turbidity doubled to 26 h.

Thus, the premix protocol was unsuitable for *E. coli* growth. To resolve this issue, lactose was added after *E. coli* reached an A_{600} of 0.6, which facilitated the successful growth of *E. coli* cells overexpressing ApePCNA1.

3.3. Crystal structure of ApePCNA1

ApePCNA1 crystals diffracted to a maximum of 1.59 Å resolution and the final structure was determined at a resolution of 2.00 Å (Fig. 3). The final resolution cutoff was 2.0 Å, because the R_{merge} and $I/\sigma(I)$ showed abnormal values in the resolution range 1.96–1.66 Å, although the statistics for the outer shell (resolution range 1.66–1.59 Å) were normal. Most amino-acid residues of ApePCNA1 (except for the N-terminal Met1-Ser2-Ser3-Glu4 and the C-terminal Gly263) were assigned in the final structure. The crystal structure contains a single ApePCNA1 subunit within the asymmetric unit. The crystals did not contain a threefold axis and the subunits did not form a ring-shaped trimeric structure. The ApePCNA1 subunit is composed of N- and C-terminal domains and a long loop called the inter-domain-connecting loop (IDCL) that

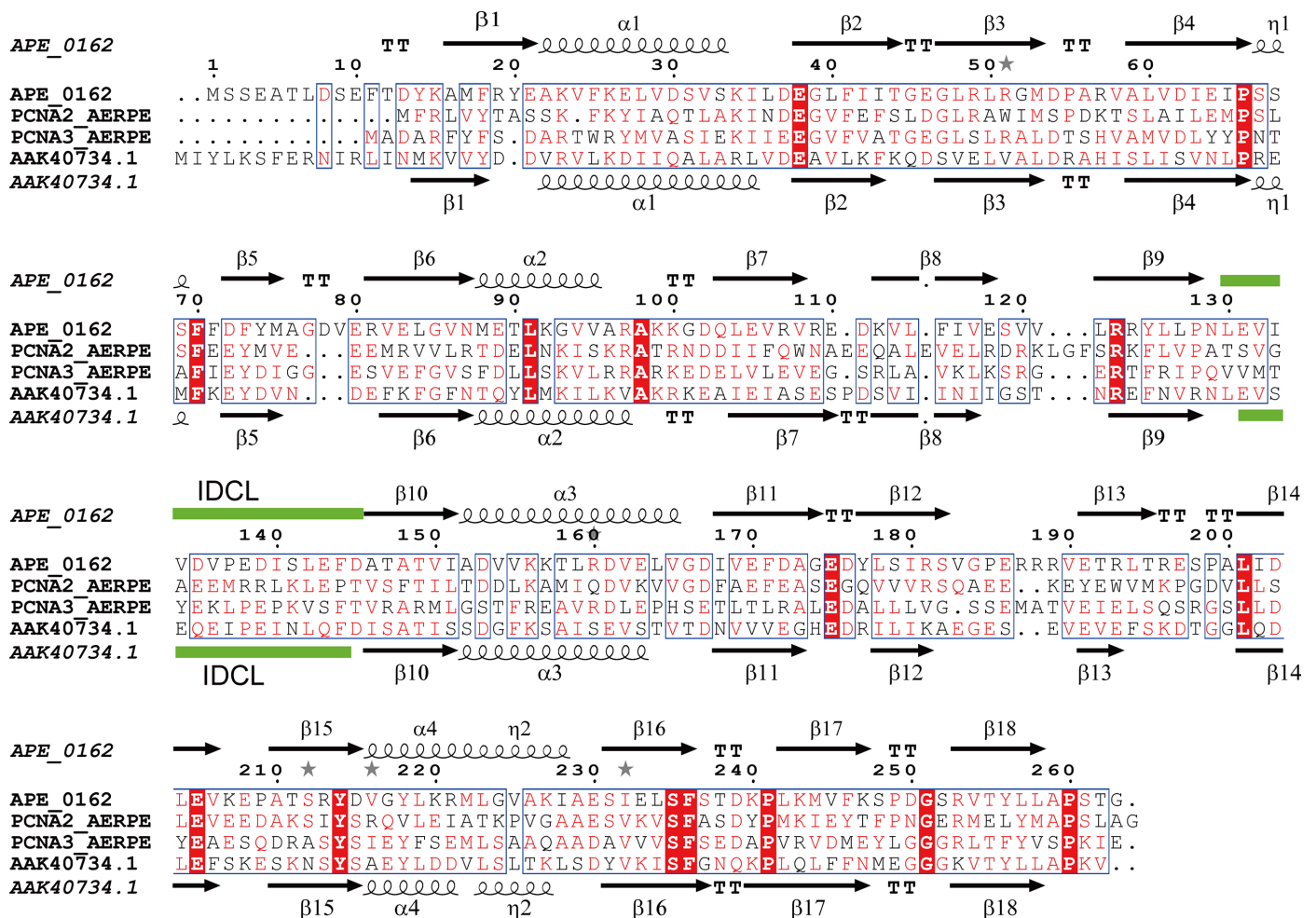


Figure 1
Multiple sequence alignment of ApePCNAs and SsoPCNA3. Helices, β -sheets and the IDCL are represented by coils, arrows and green bars, respectively. Columns with residues that are more than 70% similar according to physicochemical properties (threshold set to 0.7) are framed in blue and amino-acid residues with 100% identity are highlighted by a red background. This figure was created using *ESPrIpt* 3.0 (Robert & Gouet, 2014).

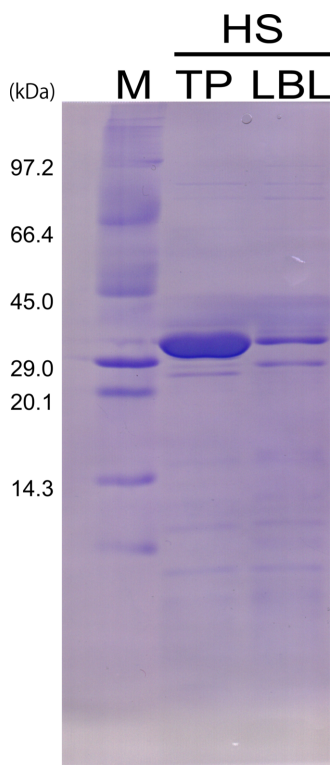


Figure 2
Comparison of ApePCNA1 expression between the two media. In the SDS-PAGE image, the molecular-weight marker (lane M) is on the left, TP medium (lane TP) is in the middle and Lennox LB medium (lane LBL) is on the right. Supernatants (lanes HS) of each sample with the same weight were compared after heat treatment.

connects the two domains formed by Leu130–Gln145. The N- and C-terminal domains consist of two β -sheets and two α -helices. In the N-terminal domains, each β -strand (β 5, β 1, β 7, β 8, β 9 and β 6, β 2, β 3, β 4) forms small β -sheets. In the C-terminal domains, β 15, β 11, β 12, β 13 and β 14, β 10, β 16, β 17, β 18 form β -sheets. One of the two β -sheets in each domain is linked to a β -sheet in the neighboring domain in the same molecule to form a large continuous β -sheet (β 6, β 2, β 3, β 4, β 18, β 17, β 16, β 10, β 14). Therefore, the ApePCNA1 subunit contains one large β -sheet and two small β -sheets. All of these β -sheets are arranged in an antiparallel manner. The four α -helices are arranged in an inverse parallel manner and are positively charged. Of these, η 2 follows α 4 to form a tighter helix. The individual β -strands that make up the large β -sheet

are aligned parallel to the antiparallel α -helices, but although the middle of the sheet is close, the edges of the sheet are further away from the α -helices. The two small β -sheets that do not belong to the large β -sheet are arranged so that one covers one end of the large β -sheet and the other covers one end of the α -helices. The inner side of the ApePCNA1 subunit surrounded by these secondary structures is rich in hydrophobic residues that form a series of hydrophobic interactions (Fig. 4a). Cavity regions are present near the contacts between the two sides and the dorsal β -sheet (Fig. 4b). A cavity is also present between the C-terminal side of the IDCL region and the back β -sheet, which is rich in hydrophobic residues. The overall structure is similar to those of the homotrimeric ring of human PCNA (PDB entry 1axc; Gulbis *et al.*, 1996), a monomer of the PCNA ring from yeast (PDB entry 1plq; Krishna *et al.*, 1994) and a PCNA monomer from the thermophilic archaeon *P. furiosus* (PDB entry 1ge8; Matsumiya *et al.*, 2001). Structural alignment was performed for these PCNAs (human PCNA, yeast PCNA and *P. furiosus* PCNA). *PyMOL* was used for the alignment, which was performed using C^α atoms. The r.m.s.d.s for each were 1.178 Å (the average value for the trimer), 1.061 Å and 0.881 Å, respectively.

3.4. Interaction of ApePCNA1 with other molecules

The PCNA rings from other species have an inward α -helix region (Li *et al.*, 2021). The positive surface charge of the α -helix region suggests that ApePCNA1 has an α -helix region located in the inner part of the ring that interacts with the negative charge of the DNA double helix like that in other species (Fig. 4c). The trimeric ring model based on the structure of ApePCNA1 is connected by two adjacent small β -sheets between each subunit (Fig. 4d). In addition, the long N-terminus may be disordered in aqueous solution, suggesting that it does not form a homotrimer. This insight follows the results of gel-filtration analysis in a previous study (Imamura *et al.*, 2007). The IDCL region of PDB-deposited PCNAs is a movable, partially unfolded loop that often has a disordered structure. In contrast, electron density was observed in the IDCL region of ApePCNA1. Contrary to our expectations, clear electron density was observed at the N-terminus of ApePCNA1, which was expected to be a partially unfolded loop owing to its length. These long loops are the contact points of the adjacent subunits related to the screw axis in the

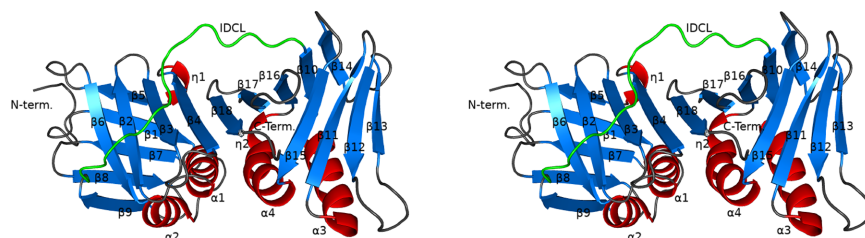


Figure 3
Stereoview of the ApePCNA1 structure. The protein is depicted as a ribbon diagram, with the α -helices, β -sheets, loops and IDCL colored red, marine blue, gray and green, respectively.

space group ($P4_32_12$) of the crystal. In general, the C-terminal side of the sliding clamp is considered to be the direction of DNA replication, and a group of enzymes responsible for DNA replication bind to the binding pocket near the IDCL (González-Magaña & Blanco, 2020). Interestingly, the N-terminus of ApePCNA1 interacted with the cleavage groove of the adjacent ApePCNA1 molecule (Fig. 5*a*). In this interaction, the carboxyl group of the N-terminal Asp8 is connected to Arg57 of the neighboring subunit by a hydrogen bond. Additionally, the NH of the main chain of Leu7 formed a hydrogen bond to the CO of the main chain of Arg57 in the neighboring subunit. The hydrophobic side chain of Leu7 is placed in a hydrophobic pocket formed by Met53, Leu60, Pro137, Pro241, Leu257 and Ala259 in an adjacent subunit. This interaction is like that observed between human PCNA and UHRF2^{PIP} (PDB entry 5yco; Chen *et al.*, 2017; Fig. 5*b*). The consensus sequence of the PIP-box is QXXhXXaa, where *X* is any amino acid, *h* is a hydrophobic amino acid and *a* is an

aromatic amino acid. The 4-EATLDSEF-11 sequence in the N-terminal region of ApePCNA1 is similar to the consensus sequence of the PIP-box (Fig. 5*c*). These results suggest that the PIP-box binding site of ApePCNA1 is in a similar location and binds in an analogous manner to those of PCNAs from other species. Furthermore, the N-terminal region of ApePCNA1 was suggested to be an anchor sequence that induces PIP-box binding. These results suggest that ApePCNA1 has an unknown function mediated by an N-terminal anchor sequence that results in PIP-box-like interactions.

In conclusion, we determined the crystal structure of ApePCNA1. A unique interaction via the N-terminus (which contains a PIP-box-like sequence) may be involved in the uncharacterized DNA-replication machinery in thermophilic archaea. We developed a new approach using a modified medium and protocol to express recombinant proteins containing the *lac* operon. These findings may be useful for protein research.

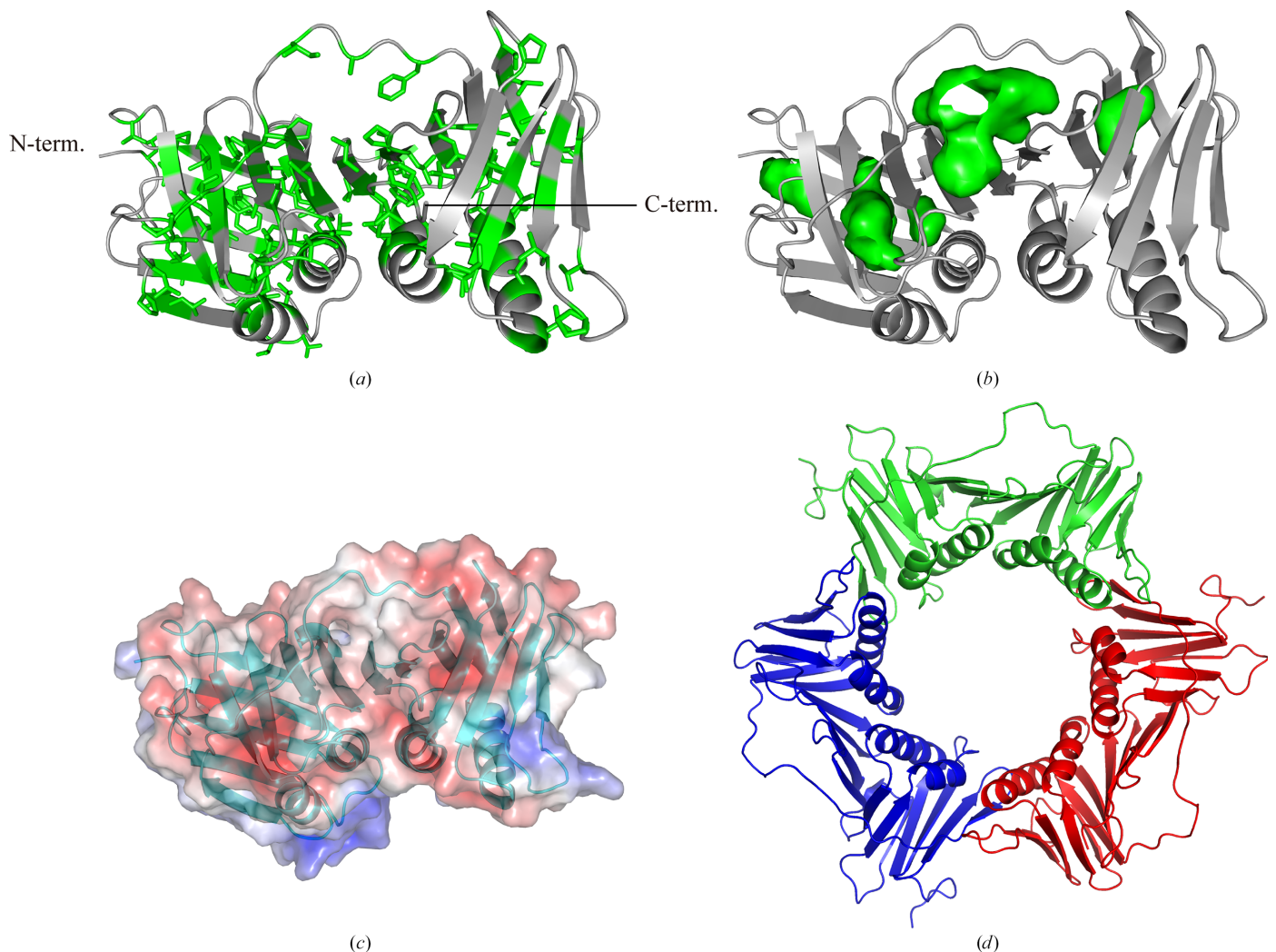


Figure 4

Structural features of ApePCNA1. (*a*) Hydrophobic amino acids are shown as green sticks. (*b*) Cavity regions are shown as green meshes. The cavities were drawn using *PyMOL*. (*c*) Electrostatic potentials on the molecular surface are shown in blue and red for positive and negative charge, respectively. (*a*), (*b*) and (*c*) are viewed from the same orientation as in Fig. 3. (*d*) Model of the ApePCNA1 homotrimeric ring viewed from the threefold axis direction. Each subunit is colored differently.

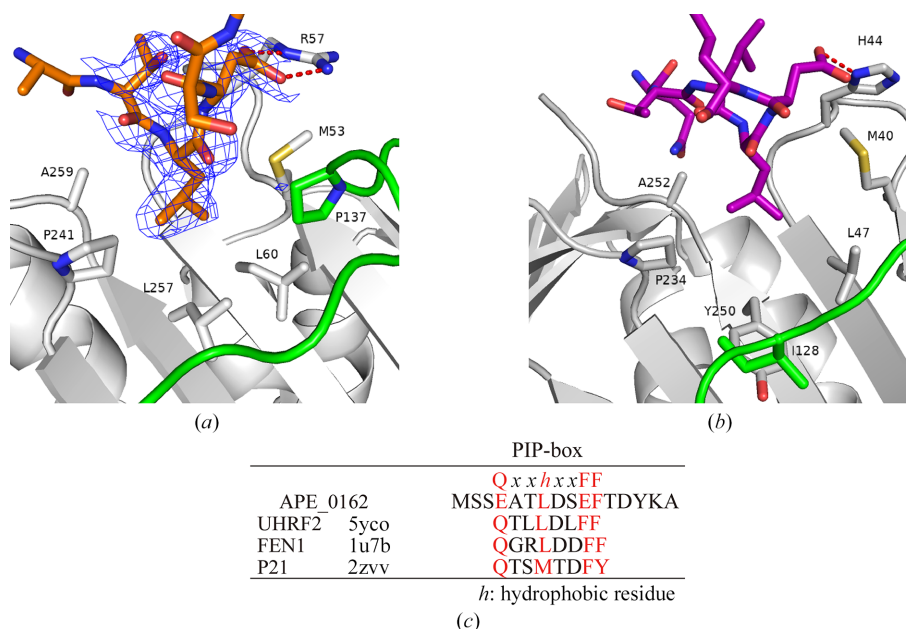


Figure 5
 Comparison of the PCNA structures and PIP-box sequences. (a) Omit $F_o - F_c$ difference map (blue mesh) calculated without the N-terminus of ApePCNA1 (orange) contoured at 2σ . The ICDL is colored green and red dashes represent hydrogen bonds. (b) Human PCNA with UHRF2^{PIP} (PDB entry 5yco) is shown in the same orientation as in (a). (c) The PIP-box consensus sequence was compared with the N-terminal sequence of ApePCNA1 and proteins with the PIP-box sequence [PDB entries 1u7b (Bruning & Shamoo, 2004) and 2zvz (Strzalka *et al.*, 2009)].

Acknowledgements

We thank Professor T. Senda of the Structural Biology Research Center (SBRC), Professor N. Matsugaki, Dr Y. Yamada and Dr M. Hikita for their support with data collection at the Photon Factory of the High Energy Accelerator Research Organization (KEK), Japan. We thank Editage (<https://www.editage.com>) for English language editing.

Funding information

This study was funded by the National Project on Protein Structural and Functional Analysis of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley, P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L., Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K., Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Malý, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray, J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito, M., Sánchez Rodríguez, F., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T.,

Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M. & Yamashita, K. (2023). *Acta Cryst.* **D79**, 449–461.
 Brehm, W., Triviño, J., Krahn, J. M., Usón, I. & Diederichs, K. (2023). *J. Appl. Cryst.* **56**, 1585–1594.
 Bruning, J. B. & Shamoo, Y. (2004). *Structure*, **12**, 2209–2219.
 Chen, W., Wu, M., Hang, T., Wang, C., Zhang, X. & Zang, J. (2017). *Biochem. Biophys. Res. Commun.* **494**, 575–580.
 Chen, X., Sun, J. & Wang, Y. (2021). *J. BUON*, **26**, 72–78.
 Daimon, K., Kawarabayasi, Y., Kikuchi, H., Sako, Y. & Ishino, Y. (2002). *J. Bacteriol.* **184**, 687–694.
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
 Georgescu, R. E., Kim, S. S., Yurieva, O., Kuriyan, J., Kong, X. P. & O'Donnell, M. (2008). *Cell*, **132**, 43–54.
 González-Magaña, A. & Blanco, F. J. (2020). *Biomolecules*, **10**, 570.
 Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. (1996). *Cell*, **87**, 297–306.
 Horsfall, A. J., Vandborg, B. A., Kowalczyk, W., Chav, T., Scanlon, D. B., Abell, A. D. & Bruning, J. B. (2021). *J. Biol. Chem.* **296**, 100773.
 Imamura, K., Fukunaga, K., Kawarabayasi, Y. & Ishino, Y. (2007). *Mol. Microbiol.* **64**, 308–318.
 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. (2017). *Nucleic Acids Res.* **45**, D353–D361.
 Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankaï, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K., Nakamura, Y., Nomura, N., Sako, Y. & Kikuchi, H. (1999). *DNA Res.* **6**, 83–101.
 Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M. & Kuriyan, J. (1994). *Cell*, **79**, 1233–1243.
 Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.
 Lennox, E. S. (1955). *Virology*, **1**, 190–206.
 Li, H., Zheng, F. & O'Donnell, M. (2021). *FEBS J.* **288**, 7256–7262.

- Liu, B., Li, S., Liu, Y., Chen, H., Hu, Z., Wang, Z., Zhao, Y., Zhang, L., Ma, B., Wang, H., Matthews, S., Wang, Y. & Zhang, K. (2021). *Nucleic Acids Res.* **49**, 11367–11378.
- MacNeill, S. A. (2016). *Curr. Genet.* **62**, 527–532.
- Maga, G. & Hübscher, U. (2003). *J. Cell Sci.* **116**, 3051–3060.
- Matsumiya, S., Ishino, Y. & Morikawa, K. (2001). *Protein Sci.* **10**, 17–23.
- Moldovan, G. L., Pfander, B. & Jentsch, S. (2007). *Cell*, **129**, 665–679.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Nedal, A., Raeder, S. B., Dalhus, B., Helgesen, E., Forstrøm, R. J., Lindland, K., Sumabe, B. K., Martinsen, J. H., Kragelund, B. B., Skarstad, K., Bjørås, M. & Otterlei, M. (2020). *Nucleic Acids Res.* **48**, 5540–5554.
- Pandey, P., Verma, V., Gautam, G., Kumari, N., Dhar, S. K. & Gourinath, S. (2017). *FEBS Lett.* **591**, 2311–2322.
- Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* **D59**, 1131–1137.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., Hoaki, T. & Maruyama, T. (1996). *Int. J. Syst. Bacteriol.* **46**, 1070–1077.
- Strzalka, W., Oyama, T., Tori, K. & Morikawa, K. (2009). *Protein Sci.* **18**, 1072–1080.
- The UniProt Consortium (2023). *Nucleic Acids Res.* **51**, 523–531.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Wendel, S. O., Snow, J. A., Gu, L., Banerjee, N. S., Malkas, L. & Wallace, N. A. (2023). *J. Med. Virol.* **95**, e29244.
- 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.* **F62**, 944–948.
- Winter, J. A. & Bunting, K. A. (2012). *Archaea*, **2012**, 951010.
- Yamazaki, S., Yamazaki, J., Nishijima, K., Otsuka, R., Mise, M., Ishikawa, H., Sasaki, K., Tago, S. & Isono, K. (2006). *Mol. Cell. Proteomics*, **5**, 811–823.
- Yao, N. Y. & O'Donnell, M. E. (2016). *Crit. Rev. Biochem. Mol. Biol.* **51**, 135–149.
- Young, J. Y., Westbrook, J. D., Feng, Z., Sala, R., Peisach, E., Oldfield, T. J., Sen, S., Gutmanas, A., Armstrong, D. R., Berrisford, J. M., Chen, L., Chen, M., Di Costanzo, L., Dimitropoulos, D., Gao, G., Ghosh, S., Gore, S., Guranovic, V., Hendrickx, P. M. S., Hudson, B. P., Igarashi, R., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Liang, Y., Mading, S., Mak, L., Mir, M. S., Mukhopadhyay, A., Patwardhan, A., Persikova, I., Rinaldi, L., Sanz-Garcia, E., Sekharan, M. R., Shao, C., Swaminathan, G. J., Tan, L., Ulrich, E. L., van Ginkel, G., Yamashita, R., Yang, H., Zhuravleva, M. A., Quesada, M., Kleywegt, G. J., Berman, H. M., Markley, J. L., Nakamura, H., Velankar, S. & Burley, S. K. (2017). *Structure*, **25**, 536–545.