Nucleic Acids Research, 2025, **53**, gkaf132
https://doi.org/10.1093/nar/gkaf132
Molecular and Structural Biology

Structural basis for cooperative ssDNA binding by bacteriophage protein filament P12

Lena K. Träger[1,†], **Morris Degen**[2,3,†], **Joana Pereira**[2,4], **Janani Durairaj**[2,4], **Raphael Dias Teixeira**[2], **Sebastian Hiller**[2,*], **Nicolas Huguenin-Dezot**[1,*]

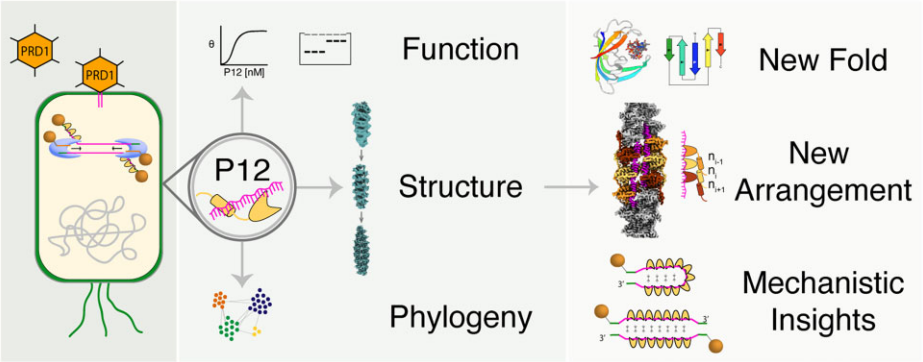
[1]Department of Biosystems Science and Engineering, ETH Zurich, Schanzenstrasse 44, 4056 Basel, Switzerland
[2]Biozentrum, University of Basel, Spitalstrasse 41, 4056 Basel, Switzerland
[3]Swiss Nanoscience Institute, University of Basel, Klingelbergstrasse 82, 4056 Basel, Switzerland
[4]SIB Swiss Institute of Bioinformatics, University of Basel, Elisabethenstrasse 43, 4051 Basel, Switzerland

[*]To whom correspondence should be addressed. Email: nicolas.huguenin-dezot@bsse.ethz.ch
Correspondence may also be addressed to Sebastian Hiller. Email: sebastian.hiller@unibas.ch
[†]The first two authors should be regarded as Joint First Authors.

Abstract

Protein-primed DNA replication is a unique mechanism, bioorthogonal to other known DNA replication modes. It relies on specialised single-stranded DNA (ssDNA)-binding proteins (SSBs) to stabilise ssDNA intermediates by unknown mechanisms. Here, we present the structural and biochemical characterisation of P12, an SSB from bacteriophage PRD1. High-resolution cryo-electron microscopy reveals that P12 forms a unique, cooperative filament along ssDNA. Each protomer binds the phosphate backbone of 6 nucleotides in a sequence-independent manner, protecting ssDNA from nuclease degradation. Filament formation is driven by an intrinsically disordered C-terminal tail, facilitating cooperative binding. We identify residues essential for ssDNA interaction and link the ssDNA-binding ability of P12 to toxicity in host cells. Bioinformatic analyses place the P12 fold as a distinct branch within the OB-like fold family. This work offers new insights into protein-primed DNA replication and lays a foundation for biotechnological applications.

Graphical abstract



Introduction

Protein-primed DNA replication is employed by various bacteriophages as a means to facilitate the specific replication of their linear genome [1]. In this mechanism, a dedicated terminal protein (TP) acts as a primer by offering a free hydroxyl group required to incorporate the first nucleotide. This initial step, as well as the subsequent elongation, is performed by a cognate DNA polymerase that covalently binds the TP to the 5′ end of the replicating DNA strand. Because of its singularity, this replication mechanism behaves in a bioorthogonal manner, being able to form synthetic replicons in *Saccharomyces cerevisiae* [2], *Bacillus thuringiensis* [3], and more recently in *Escherichia coli* [4]. Protein-primed DNA replication has thus been leveraged to perform continuous directed evolution in these organisms by using error-prone DNA polymerase mutants [2–4]. Further conceivable biotechnological applications of protein-primed DNA replication include recombinant protein production and metabolic pathway engineering [5]. Knowledge of the underlying molecular mechanisms has so far mostly been limited to bacteriophage Φ29 [6] and

Received: December 12, 2024. **Revised:** February 5, 2025. **Editorial Decision:** February 7, 2025. **Accepted:** February 11, 2025

adenoviruses [7], and are thus not available for *E. coli* as a host. *E. coli*, however, remains the workhorse of modern molecular biology, due to our extensive knowledge of its genetics, transcriptome, proteome, and metabolome, and the multitude of molecular tools available for its modification [8]. Expanding knowledge on protein-primed DNA replication systems in *E. coli* is thus urgently required to fully unlock the broadest possible spectrum of applications. The only bacteriophages known to replicate their genetic material via protein-primed DNA replication in *E. coli* are members of the *Tectiviridae* family, such as the *Enterobacteria* phage PRD1 [9], making it a prime target for structural and functional studies.

Particularly crucial in the context of protein-primed DNA replication is the recognition and stabilisation of the single-stranded DNA (ssDNA) strands that are generated during the replication process. Initial studies on PRD1 have identified four proteins involved in the replication process (Supplementary Fig. S1), two of which, P12 and P19, have been identified as single-stranded DNA-binding proteins (SSBs) [10, 11]. Whereas P19 appears to be dispensable, P12 is essential for *in vivo* replication of PRD1 [12] and has also been shown to enhance DNA replication *in vitro* [10]. P12 overexpression has a detrimental effect on host cell viability [12–14], overall suggesting a crucial role for P12 in the replication process of PRD1.

In this study, we present a comprehensive structural and biochemical examination of P12. We determine a high-resolution cryo-electron microscopy (cryo-EM) structure of P12 bound to ssDNA. The structure reveals the formation of unique proteofilaments along the ssDNA strand. We characterise functional aspects of the P12 filament architecture through biochemical assays, including truncation, targeted mutagenesis or a nuclease protection assay, and establish a connection between ssDNA binding and host toxicity. Finally, bioinformatic analysis corroborates the singularity of the observed filament arrangement, suggesting an adaptation specifically tailored for protein-primed DNA replication systems.

Material and methods

Cloning, plasmids and strains

Synthetic oligonucleotides, plasmids, and *E. coli* strains are listed in Supplementary Table S1. To generate an expression vector for P12, the P12 coding sequence was PCR-amplified from PRD1 genomic DNA (gift from D. Bamford, Department of Biosciences, University of Helsinki, Finland) with primers P12_pBAD_SapI_fw and P12_pBAD_XhoI_rev. Similarly, a Small Ubiquitin-like Modifier (SUMO) tag preceded by an *N*-terminal hexahistidine was amplified using primers His6_SUMO_pBAD__NcoI_fw and SUMO_pBAD_SapI_rev. The pBAD backbone was digested using NcoI-HF and XhoI-HF restriction enzymes (New England Biolabs [NEB]) and dephosphorylated using alkaline phosphatase (Roche). The P12 and SUMO inserts were digested using SapI (NEB) and XhoI-HF or NcoI-HF respectively. The digested backbone and inserts were ligated using the Rapid DNA Ligation Kit (Roche), yielding plasmid pBAD_His_6_SUMO_P12. Primer pairs P12_K10A_fw/P12_K10A_rev and P12_F44A_fw/P12_F44A_rev were used to introduce specific mutations into P12. The amplified products were digested with BsaI-HFv2 (NEB) and DpnI (NEB), followed by ligation. Similarly, truncated versions of P12 were generated using primers P12_trunc_fw, P12_trunc1-148_rev, P12_trunc1-140_rev and P12_trunc1-122_rev. For toxicity assays, the His6_SUMO-tag was removed with primers P12_removetags_fw/P12_removetags_rev.

Protein expression and purification

The pBAD_His_6_SUMO_P12 plasmid was transformed into *E. coli* BL21 (DE3) cells and cultivated in TB medium with 12.5 μg/mL tetracycline. The cultures were incubated at 37°C with agitation at 220 r.p.m, until reaching an OD_{600} of 0.6. Protein expression was induced using 0.2% L-arabinose. Cells were incubated at 37°C for an additional 4.5 h before being centrifuged at 6000 × *g* for 15 min. For protein purification, the cell pellets were resuspended in buffer A (50 mM Tris-HCl pH 8.5, 50 mM NaCl, 1 mM $MgCl_2$) supplemented with 0.5 mg/mL lysozyme (Sigma), 50 μg/mL DNAse (Sigma) and EDTA-free Pierce™ (ThermoFisher) protease inhibitor tablets. The resuspended cells were lysed by sonication at 50% amplitude (Bandelin WH40 Sonicator) for 15 min, employing an ON/OFF cycle of 1 s/1 s. The lysate was clarified at 39 000 × *g* for 30 min at 4°C. The clarified lysate was loaded onto a 5 mL HisTrap™ HP column (Cytiva) pre-equilibrated with buffer A. The protein was eluted with a linear gradient of imidazole (0–500 mM). Eluted protein was dialysed overnight against low salt buffer (50 mM Tris-HCl pH 8.5, 20 mM NaCl, and 1 mM $MgCl_2$). The His_6-SUMO-tag was cleaved concomitantly by using 40 ng/mg His_6-SENP1 protease. The protease, uncleaved protein and cleaved His_6-SUMO were removed by reverse-affinity chromatography using a 5 mL HisTrap™ HP column. The flow-through was applied to a HiTrap™ Q HP column (Cytiva) pre-equilibrated in low salt buffer. Fractions containing P12 were concentrated with 3 kDa molecular weight cutoff Amicon Ultra Centrifugal Filters (Millipore) and injected on a HiLoad™ 16/600 Superdex 75 pg column (Cytiva) pre-equilibrated with buffer A. Fractions containing purified P12 were pooled, flash frozen and stored at -80°C. All chromatography procedures were conducted at 4°C using an ÄKTA Go system (Cytiva). Protein analysis was performed using 4–12% BOLT Bis-Tris gel (Invitrogen) with MES buffer and stained using QuickBlue protein stain (LubioScience).

Cryo-EM sample preparation

A 3.5 μL aliquot of the SSB–ssDNA complex, containing 750 μM P12 and 50 μM of either $(non\text{-}repetitive)_{80}$, $(ATGCT)_{16}$, or $(dT)_{80}$ ssDNA, was absorbed onto a glow-discharged holey carbon-coated grid (Quantifoil R 1.2/1.3). The sample was blotted with Whatman 1 filter paper and vitrified into liquid ethane at -178°C using a Leica GP2 plunger.

Data collection

A FEI Titan Krios G1 TEM electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct electron detector and a Gatan Quantum-LS Energy Filter (GIF) energy filter was used for data collection. For the P12-$(non\text{-}repetitive)_{80}$ complex we collected 3′492 movies with an accumulated dose of 49 e/Å^2, for the $(ATGCT)_{16}$ ssDNA complex we collected 6′387 movies with an accumulated dose of 49,41 e/Å^2, and for the $(dT)_{80}$ we collected 7′484 movies with an accumulated dose of 48.7 e/Å^2. All collections were done at a pixel size of 0.82Å (nominal magnification of 165k).

Cryo-EM data processing

For all datasets, movies were imported into cryoSPARC [15] v3.4 + for initial patch motion correction, followed by contrast transfer function (CTF) estimation using patch CTF determination. Manually picked filamentous particles were extracted with a box size of 360 pixels and subjected to 2D classification. Low-resolution 2D templates from these initial particles guided filament tracer and template-based particle picking across the datasets. Subsequent 2D classification produced high-resolution class averages, which were used for *ab initio* reconstructions. Initial non-uniform refinements of the *ab initio* models were performed without imposing symmetry constraints. Helical parameters, including rise and twist, were estimated from the initial reconstructions and refined using Fourier–Bessel indexing of averaged power spectra from 2D classes. Final 3D reconstructions were obtained through an iterative process of stringent particle stack cleaning, excluding heavily bent filaments based on 2D class averages. To further improve particle quality, heterogeneous refinements were performed, incorporating low-pass filtered reconstructions alongside 'junk' classes to remove low-quality particles. Cleaned particle stacks were then used for particle picking with Topaz, and duplicate particles were removed through either duplicate removal jobs or alignment-based exclusion during 2D/3D classifications. The most refined particle stack from each dataset underwent final helical refinement, including correction of higher-order aberrations, per-exposure group CTF corrections, and reference-based motion correction with empirical dose weighting. This process was applied consistently across all datasets, yielding final reconstructions with resolutions of 3.71 Å for the $(\text{non-repetitive})_{80}$ substrate, 3.07 Å for the $(\text{ATGCT})_{16}$ substrate, and 2.75 Å for the $(\text{dT})_{80}$ substrate at a GSFSC cutoff of 0.143. Helical parameters were overall consistent across all reconstructions, with a helical rise between 27.5 and 29.2 Å and a helical twist between 17.8° and 19.2° (Supplementary Table S2).

Model building of P12–ssDNA complexes

A map corresponding to a single subunit repeat was extracted using UCSF ChimeraX [16], and initial *de novo* model building was performed in Coot [17]. This was guided by an AlphaFold2 [18] structure prediction of monomeric P12, which was compared and refined based on the solved crystal structure. The model was then fitted into the densities of three consecutive P12 protomers from one filament and three consecutive protomers from the opposing filament. In a next step, 20 nt ssDNA strands were generated in ChimeraX, roughly modelled into the corresponding densities using ISOLDE [19], and later manually refined using Coot. The resulting six-protomer map, including the two ssDNA strands, was refined using phenix.real_space_refine with default settings, incorporating non-crystallographic symmetry (NCS) to restrain the subunit repeats. After each round of refinement, model geometry was evaluated using MolProbity [20] v4.5.2, and poorly fitting regions were manually adjusted in Coot. This iterative process was repeated until satisfactory model:map agreement and stereochemical quality were achieved (Supplementary Table S2).

Crystallographic studies of P12

Crystallisation conditions for $\text{P12}_{(1–122)}$ were found by using commercially available screens (Morpheus[1], Molecular Dimensions) and a protein concentration of 11 mg/mL. Crystals formed within 5 days at 20°C using the sitting drop vapor diffusion method. A 0.2 µL aliquot of protein in buffer A (50 mM Tris-HCl pH 8.5, 50 mM NaCl, 1 mM $MgCl_2$) was mixed with an equal volume of reservoir solution containing 0.12 M monosaccharides (0.2M D-glucose; 0.2M D-mannose; 0.2M D-galactose; 0.2M L-fucose; 0.2M D-xylose; 0.2M N-acetyl-D-glucosamine), 0.1 M buffer system 3 (Tris [base]; BICINE) at pH 8.5, and 37.5% v/v precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). The crystals were flash frozen in liquid nitrogen without requiring additional cryoprotectant. X-ray data were collected using the PXI beamline at the Swiss Light Source located at the Paul Scherrer Institut in Switzerland. Data were processed with XDS and CCP4i [21, 22]. The structure was solved by molecular replacement using Phaser [23], and a P12 model generated by AlphaFold2 as searching model [18]. Subsequent structure building and refinement were performed using Coot [17], phenix.refine [24], and Refmac5 [25]. Structural analysis and figure preparation were conducted using PyMOL and UCSF ChimeraX [16]. The coordinates and structure factors were deposited in the Protein Data Bank [26] under entry code 9GFQ (Supplementary Table S3).

Hill model

EMSA bands were quantified using ImageJ [27]. High-resolution scans were converted to grayscale and analysed. Regions of interest were defined for each band and applied consistently across all lanes. Pixel density measurements were taken for both the bands and their corresponding backgrounds. Background-corrected values were calculated and ratios of the bound band intensity to the total intensity. Hill modelling was used to quantify the binding affinity and cooperativity of P12 across various ssDNA substrates. The fraction of binding sites occupied (θ) was plotted against protein concentration while ssDNA was constant (7.5 nM) and data were fitted to the Hill equation:

$$\theta = \frac{\theta_{max}[L]^n}{K_D^n[L]^n}$$

Curve fitting was conducted using non-linear least squares optimisation via the SciPy [28] curve_fit tool. Two models were applied: An unconstrained model where $\theta_{\max}$ allowed to vary for conditions with incomplete binding and a constrained model where $\theta_{\max}$ fixed at 1 for conditions indicating full binding. Initial guesses for K_D, n and $\theta_{\max}$ were based on the binding saturation and shift behaviour observed in the EMSA raw data. Standard errors were calculated based on residuals between the experimental and fitted data. Error bars were plotted for each data point and coefficient determination (R^2) was used to evaluate the fit quality.

Analysis of nucleic acid binding capacity

EMSAs were employed to investigate the interactions between DNA/RNA substrates and proteins. ssDNA oligomers of variable length ($[\text{non-repetitive}]_{20}$, $[\text{non-repetitive}]_{30}$, $[\text{non-repetitive}]_{34}$, $[\text{non-repetitive}]_{35}$, $[\text{non-repetitive}]_{36}$, $[\text{non-repetitive}]_{37}$, $[\text{non-repetitive}]_{40}$, $[\text{non-repetitive}]_{50}$, $[\text{non-repetitive}]_{c\_50}$, $[\text{non-repetitive}]_{80}$, $[\text{dT}]_{80}$, and $[\text{dA}]_{80}$) were purchased from Merck, while the single-stranded RNA $(\text{non-repetitive})_{50}$ was obtained from Microsynth. Oligomers were tagged with Cyanine-5 (Cy5) fluorophore at their 5′ end.

The dsDNA substrate was prepared by hybridising labelled 50mer with a 1.2 fold excess of its complementary unlabelled counterpart ($[\text{non-repetitive}]_{c\_50}$) in a hybridisation buffer containing 0.2 M NaCl and 60 mM Tris-HCl pH 7.5. The mixture was heated for 10 min at 95°C and left to cool to room temperature. Binding reactions were performed using 7.5 nM of nucleic acid substrates and variable amounts of protein in a reaction buffer containing 50 mM NaCl, 1 mM DTT, 4% glycerol, 0.1 mg/mL BSA, 0.05% Tween20, 1 mM EDTA, 50 mM Tris pH 8.5, and 1 mM $MgCl_2$. The final reaction volume was 20 µL. Reactions were incubated at 4°C for 30 min before being analysed on an 0.7% agarose gel running for 150 min at 20 V and 4°C in 0.5 × TBE buffer. Cy5-labelled substrates were detected using a Typhoon™ FLA 7000 biomolecular imager (GE Healthcare).

Nuclease protection assay

DNA analysis in the nuclease protection assay was performed using a constant ssDNA 80mer substrate concentration of 20 nM. The 20 µL reaction mixture contained 10 units of Exonuclease I (*E. coli*) (NEB), 1 × exonuclease buffer (NEB), 50 mM NaCl, 4% glycerol, and 50 µM of P12. The reaction mixtures were subjected to varying incubation times at 37°C, ranging up to 20 min. To stop the reactions, the samples were boiled for 5 min at 95°C. From the final 20 µL reaction mixture, 10 µL were combined with 2 × Novex™ TBE-Urea sample buffer, and 10 µL of this mixture was loaded onto a 10% Novex™ TBE-Urea polyacrylamide gel. The gel was run at 180 V for one hour and then stained using SYBR™ Green II. DNA band intensities were quantified using Fiji ImageJ software [27].

Toxicity tests

Toxicity assays were performed by transforming 100 ng of pBAD plasmids containing different P12 variants ($P12_{(\text{wild-type})}$, $P12_{(1–148)}$, $P12_{(1–140)}$, $P12_{(1–122)}$, $P12_{(F44A)}$, and $P12_{(F44A+K10A)}$) into electrocompetent BL21 (DE3) *E. coli* cells. Cells were recovered for 1.5 h at 37°C after electroporation before being plated on LB-agar plate containing 25 µg/mL tetracycline and either 1% D-glucose or 0.2% L-arabinose to repress or induce protein expression respectively. Plates were incubated at 37°C before counting colonies.

Fold classification and structural comparisons

To evaluate whether the P12 fold belongs to any already classified protein fold, our crystallographic model of P12 was first used as query for structural searches against the ECOD [29], CATH [30], and PDB [26] databases. Searches against ECOD were performed using the 'protein structure search' method in the ECOD webserver, which employs TM-align [31] as the search method. Searches against CATH were carried out using the Foldseek [32] webserver, and against the PDB using both the DALI [33] and Foldseek webservers. All searches were carried out using default settings.

To put P12 in context of the matches obtained, a comprehensive set of experimental and predicted structural homologs of P12 was collected and a structure similarity network constructed. Starting from our cryo-EM model of P12 and the closest PDB match found with DALI (the DNA binding domain of Tequatrovirus T4 gp32 protein, PDB ID: 8GME, chain A, residues 24–193), structural searches against the following pre-computed Foldseek databases were carried out with the Foldseek API with default parameters using the 3Di mode: the AlphaFold database v4 [34] filtered to 50% sequence identity (afdb50 v4), the ESM Metagenomic Atlas [35] filtered to 30% sequence identity (mgnify_esm30 v1), and the PDB filtered to 100% sequence identity (pdb100 20 240 101). All matches were collected and processed for protein domains with Chainsaw [36] and only the domains aligning with at least one of the queries were considered further. For the search starting from the P12 model, a further Foldseek search round was carried out with each domain matched with a probability of 80% using the same approach, and the matches were further enriched by a similar but independent search over the Big Fantastic Virus Database (BFVD 2023_02) [37].

To reduce redundancy, the resulting set of collected domains was then filtered to a maximum of 50% sequence identity with Foldseek easy-cluster, which automatically groups sequences into clusters of identical proteins and selects a representative for each group. These representatives were then complemented with the structure models of 33 non-redundant (maximum sequence identity of 100%) P12 proteins from Quinones-Olvera *et al.* [9], which we generated with AlphaFold2 [18], as well as the single-stranded DNA-binding domains of Hafnia phage Enc34 ORF6 protein (PDB ID: 5ODL, chain A, residues 1–177) and *Enterococcus faecalis* PrgE protein (PDB ID: 8S4T, chain B, residues 1–136).

A protein structure similarity network of the resulting 14 986 domain representatives was computed by all-against-all Foldseek pairwise comparisons and layouted with Cosmograph [38] until equilibrium. For layouting, each edge was given a weight proportional to the *E*-value of the match, and a maximum of 10 outbound edges were considered per node, except for the P12 proteins from Quinones-Olvera *et al.* 2024, where all edges with an *E*-value better than 1×10^{-4} were kept. Clusters were automatically identified and annotated iteratively, by searching for connected components at a given *E*-value threshold. For each connected component in the map, the median and median absolute deviation of the *E*-values were computed at each iteration and the *E*-value threshold set to $median - 0.5 \times MAD$. This threshold was set to find further connected components in the next iteration until no more connected components with a minimum of 10 nodes were defined. Cluster identification was carried out for a maximum of five iterations.

Results

In vitro characterisation of specific ssDNA binding by P12

We set out to characterise the binding specificity of P12 to various nucleic acid substrates by EMSAs (Fig. 1 and Supplementary Fig. S2). ssDNA substrates of varying lengths, ranging from 20 to 80 nt, were incubated with saturating concentrations of P12 and then analysed by EMSA. P12 consistently bound ssDNA strands with lengths of 40 nt and above (Fig. 1A). The binding affinity of P12 appears to increase with the length of the ssDNA substrates, as indicated by the appearance of sharp bands. This further suggests that cooperativity may be a feature of P12, with adjacent subunits favourably interacting along the ssDNA strand.

We investigated the sequence-specificity of P12 by testing its binding to 80 nt homopolymers (dA), (dC), and (dT), as well as an 80 nt ssDNA strand with a non-repetitive sequence

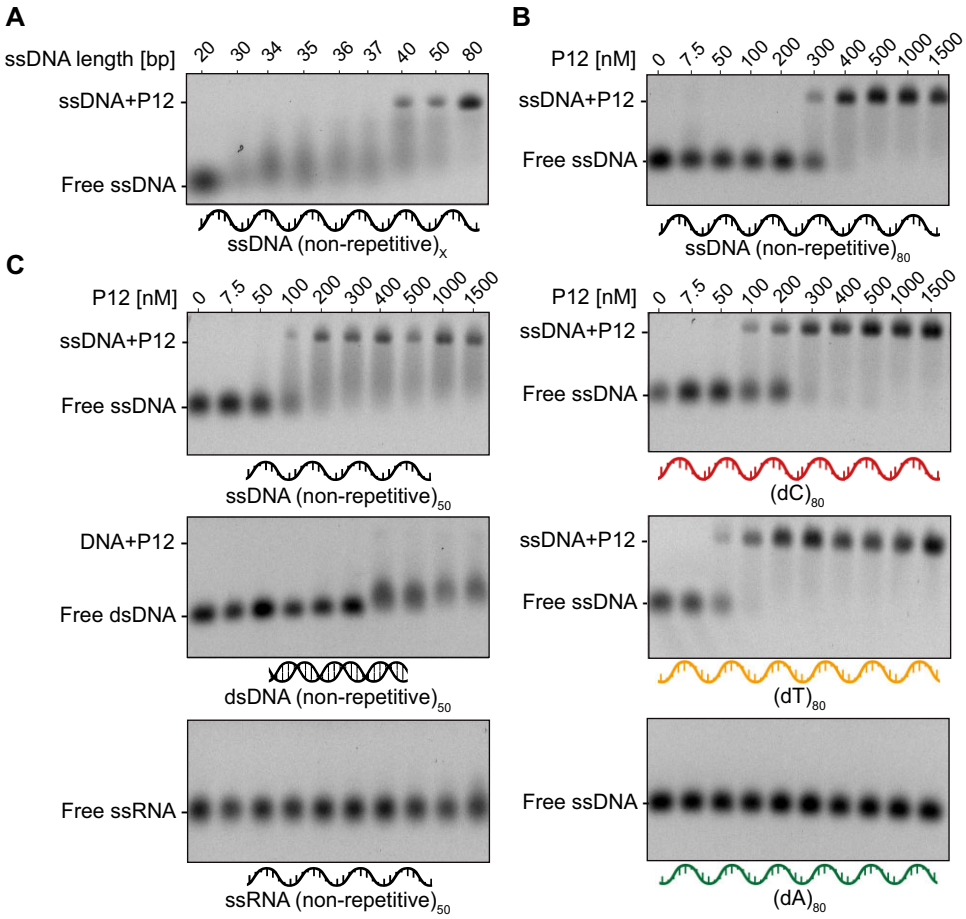


Figure 1. P12 specifically binds to single-stranded DNA. EMSAs of P12 in the presence of different nucleic acid substrates. (**A**) 7.5 nM of Cy5-labelled ssDNA of variable lengths incubated with 1 μM of P12. (**B**) 7.5 nM of different 80 nt long Cy5-labelled ssDNA substrates (ssDNA[non-repetitive]$_{80}$, [dC]$_{80}$, [dT]$_{80}$, [dA]$_{80}$) incubated with increasing concentrations of P12. (**C**) 7.5 nM of different 50 nt long Cy5-labelled nucleic acid substrates (ssDNA[non-repetitive]$_{50}$, dsDNA[non-repetitive]$_{50}$, and ssRNA[non-repetitive]$_{50}$) incubated with increasing concentrations of P12.

(Fig. 1B). P12 bound the non-repetitive substrate, and both poly(dC) and poly(dT) tightly at a ssDNA concentration of 7.5 nM and 40 fold stoichiometric excess of P12 (Fig. 1B). Notably, the two homopolymers were bound with higher affinity than the non-repetitive substrate. In contrast, the poly(dA) substrate showed a significantly reduced affinity for P12 (Fig. 1B). We furthermore assessed the specificity to other nucleic acid forms. P12 does not bind to RNA and only weakly interacted with dsDNA across a broad concentration range, in contrast to its pronounced affinity for ssDNA substrates of equivalent length (Fig. 1C). P12 is thus a specific ssDNA binder, efficiently distinguishing from dsDNA or RNA, without requiring specific sequence elements for binding.

Structure determination of P12–ssDNA filaments

We employed single-particle cryo-EM microscopy to determine the structure of P12–ssDNA complexes (Fig. 2). Initial micrographs of P12 in complex with 80mer ssDNA with a defined, non-repetitive sequence showed filamentous assemblies, from which portions were manually picked and subjected to 2D classification (Fig. 2A). The resulting low-resolution templates guided particle picking, followed by iterative 2D classification to generate high-resolution class averages (Fig. 2B). Using these particle stacks, we performed *ab initio* model generation and non-uniform refinement without symmetry constraints. By estimating the rise and twist of the helical assembly through Fourier–Bessel indexing and 3D reconstruction, we refined the helical parameters significantly, improving the map quality to 3.71 Å resolution (Fig. 2C and Supplementary Fig. S3A-C). P12 assembled as a long and continuous, homooligomeric filament along one ssDNA strand, and two such ssDNA–P12 filaments annealed to a double filament in anti-parallel fashion. The ssDNA was poorly resolved in the density map, likely due to variability introduced by the non-repetitive ssDNA sequence causing the density to blur during the averaging process. Nonetheless, this map allowed us to estimate that each P12 protomer in the filament binds a 5–7 nt long segment of the ssDNA.

To improve the resolution, we switched to repetitive ssDNA sequences. EMSA experiments showed that substrates with 6 or 7 nt repeat sequences exhibited a reduced binding affinity compared to the 5 nt repeat sequence, presumably due to the formation of secondary structures (Supplementary Fig. S2). Using this 5 nt repeat sequence in the same cryo-EM workflow resulted in a map with a substantially improved resolution of 3.1 Å (Supplementary Fig. S3D and E). Despite the improved resolution, the map unambiguously revealed a 6 nt long ssDNA segment bound per P12 protomer, and thus a remaining mismatch between the 5 nt repeat used as ssDNA substrate and the available binding sites per P12 protomer in the filament (Fig. 2D). Therefore, we next used an 80 nt

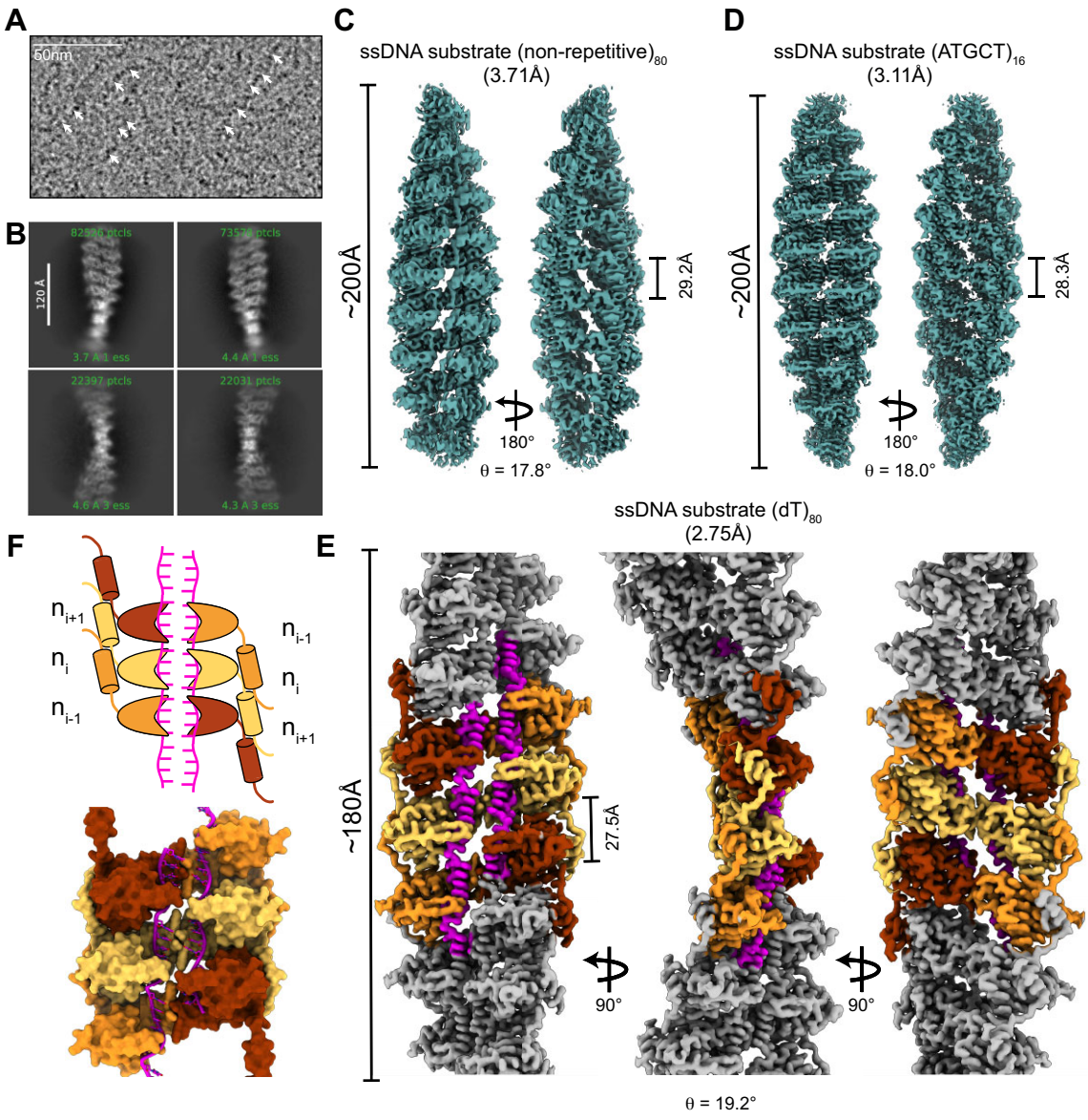


Figure 2. P12 forms a cooperative filament to bind ssDNA. (**A**) Cryo-EM micrograph showing filamentous assemblies of P12 bound to a ssDNA substrate (white arrows). (**B**) Representative 2D class averages of the P12–ssDNA complex. (**C**) 3D reconstruction of P12 bound to ssDNA (non-repetitive)$_{80}$. (**D**) 3D reconstruction of P12 bound to a five-repeat ssDNA substrate (ATGCT)$_{16}$ at 3.07 Å resolution. (**E**) 3D reconstruction of P12 bound to (dT)$_{80}$ at 2.75 Å resolution. (**F**) Modelled structure of the anti-parallel double-filament of the P12–ssDNA complex, highlighting the arrangement of P12 subunits (coloured by protomer) and the ssDNA threaded through the central cleft of each filament (pink).

long (dT) sequence devoid of any secondary structure formation as a substrate. The single-particle cryo-EM structure of poly(dT) bound to P12 was exceptionally well resolved, achieving a global resolution of 2.75 Å for both the 16 kDa P12 protomers forming the filament and the ssDNA (Fig. 2E and Supplementary Fig. S4A–D). Importantly, the structure retained the same anti-parallel double filament arrangement observed with both the 5 nt repeat and the non-repetitive ssDNA substrates, indicating that the ssDNA binding mode of P12 is not sequence specific.

Model building of the P12 protomer with bound ssDNA into the high-resolution map (2.75 Å) revealed specific features. Successive P12 protomers are arranged in a repetitive manner along the ssDNA strand, forming a continuous filament. In this filament, P12 protomers bound to one ssDNA strand connect via loose contacts to an opposing filament of the same composition, forming a double filament (Fig. 2E and F). This arrangement seems facilitated by the avidity originating from the repetition of weak pairwise interaction between the two P12 filaments (Supplementary Fig. S5A). Notably, the two ssDNA strands remain spatially separated by approximately 6–7 Å in this configuration and do not make direct contact (Supplementary Fig. S5B).

Structural determinants of ssDNA binding

The high-resolution map revealed further features underlying ssDNA binding by P12 (Fig. 3). The N-terminal domain of P12 comprises five anti-parallel β-strands and an α-helix, resembling the Greek key motif typically found in oligonucleotide/oligosaccharide-binding (OB) folds, a prominent class of ssDNA-binding proteins. However, P12 adopts

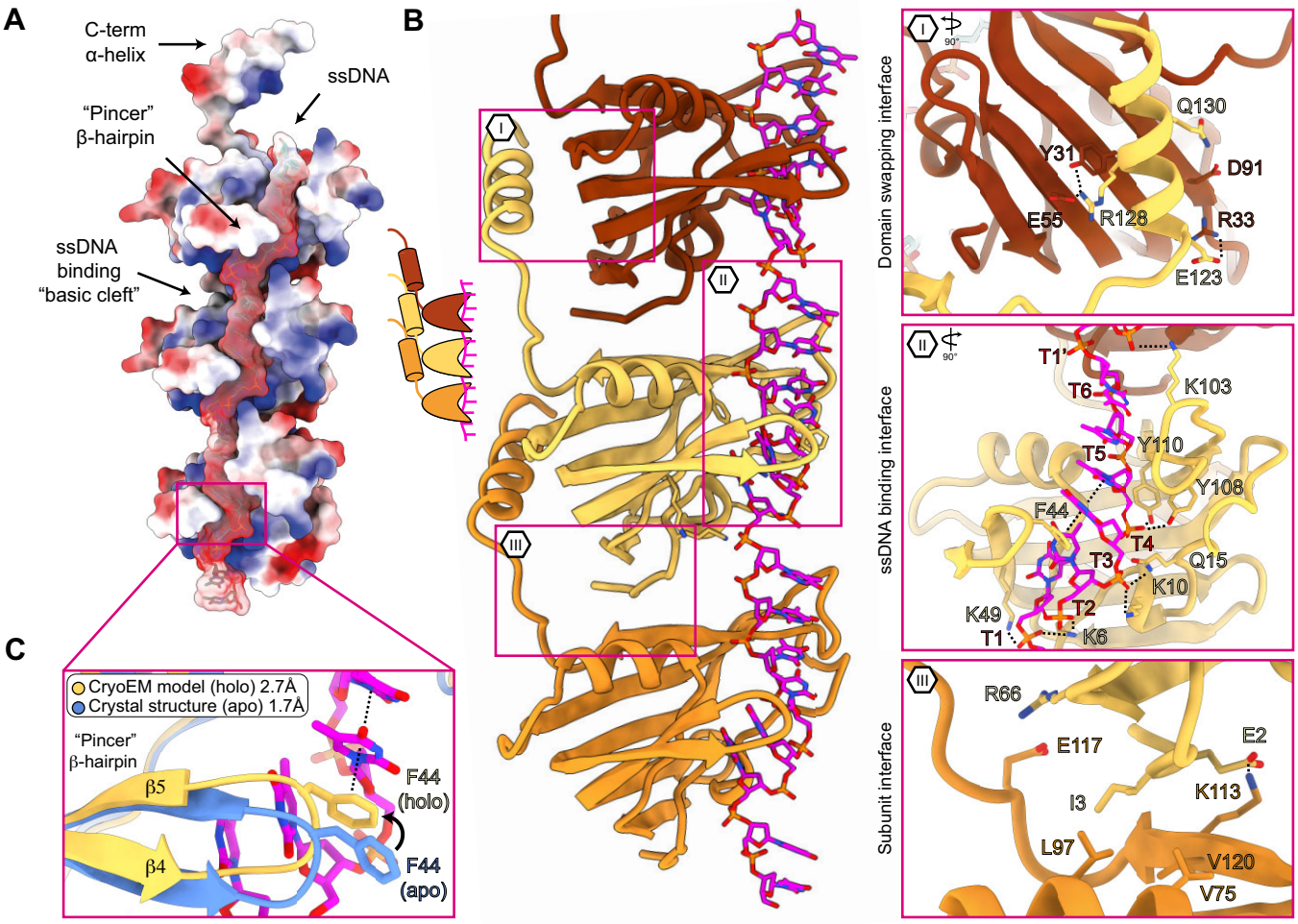


Figure 3. Key structural interfaces mediating P12–ssDNA interactions. (**A**) Surface electrostatic representation of P12 upon ssDNA binding. The ssDNA threads through the positively charged cleft formed by the pincer β-hairpin and the core of each P12 protomer. (**B**) Structural representation of the P12 filament bound to ssDNA. Three key interaction sites are marked with pink boxes: (**I**) Domain-swapping interface between P12 protomers. (**II**) ssDNA-binding interface showing interactions between P12 and the ssDNA backbone. (**III**) Subunit interface illustrating interactions between neighbouring P12 protomers. (**C**) Overlay of the crystal structure (apo, blue) and the cryo-EM model (holo, yellow), illustrating the position of the F44 residue at the tip of the pincer loop between β-strands 4 and 5.

a unique pincer-shaped conformation, forming a positively charged binding cleft, thereby significantly diverging from the classic closed β-barrel commonly found in OB-folds. Upon filament assembly, this charged cavity accommodates the ssDNA (Fig. 3A). Cooperativity between P12 protomers is mediated by the C-terminal α-helix of each protomer, which extends to the adjacent protomer, where it stacks onto the β-sheet core (Fig. 3B). The key polar residues responsible for this cooperative binding are R128, E123, and Q130 of the C-terminal α-helix, interacting with Y31 and E55, R33, and D91 of the adjacent protomer, respectively (Fig. 3B I). Within the ssDNA-binding cleft, the first two nucleobases of the 6 nt binding segment are oriented inwards, interacting with the core of the P12 protomer. The remaining four nucleobases face outward, with their sugar-phosphate backbone anchored within the cavity (Fig. 3B II). The primary interactions between P12 and ssDNA are thus mediated by hydrogen bonds with the sugar-phosphate backbone, involving residues K6, K10, Q15, K49, K103, Y108, and Y110. A key stabilising interaction in the 'pincer' β-hairpin involves F44, creating π–stack interactions with the ssDNA nucleobases.

Direct hydrogen bonding with the nucleobases is minimal, consistent with the lack of sequence specificity in ssDNA binding. Although purine bases can fit into the binding pocket, we did not observe binding for poly(dA) substrates. Poly(dA) ssDNA strands have been shown to adopt a helical conformation promoted by the efficient base stacking of adenine nucleobases [39]. This distinctive structural feature likely interferes with the ability of this substrate to interact effectively within the P12 binding cleft (Fig. 1B). Lastly, two salt bridges present between the N-terminus of P12 and the α-helix of the preceding protomer might also contribute significantly to the stability of the filament (Fig. 3B III). The C-terminal tail of P12 beyond the α-helix responsible for cooperative binding was predicted to be intrinsically disordered and, as anticipated, was not resolved in our cryo-EM density map.

To independently validate the P12 structure, we crystallised a truncated form of P12 lacking the C-terminal tail, in the absence of ssDNA. The resulting X-ray crystal structure (1.7 Å resolution) closely matched the filament protomer structure, with an RMSD of 0.87 Å (Supplementary Fig. S5C). Thus, the core fold of P12 is a stable feature of the protein and is

not perturbed by binding of ssDNA. Notably, the most significant structural deviation between the apo-state of P12 and the holo-state within the P12 filament occurs in the pincer region. Specifically, F44, located at the tip of the pincer loop between β-strands 4 and 5, undergoes a well-resolved 5 Å shift in the ssDNA bound state (Fig. 3C). This conformational change enables F44 to engage in π-stacking with the nucleobases at positions T3 and T4, effectively stabilising the interaction by locking the ssDNA in place.

Role of the ssDNA interface and C-terminus in ssDNA interaction

To further validate our structural model and to investigate the functional contributions of key regions of P12, we performed a series of EMSAs and toxicity tests to assess the effects of specific mutations and truncations (Fig. 4). First, we introduced mutations at two key residues of the ssDNA-binding interface. F44 stabilises ssDNA through π-stacking with nucleobases, while K10 forms hydrogen bonds with the sugar-phosphate backbone (Fig. 3B II and C). Consistent with our structural model, we observed a strong reduction of ssDNA binding for the mutant F44A and a complete loss upon introduction of the additional mutation K10A (Fig. 4A). We next set out to characterise the role of the C-terminal tail, which is comprised of an α-helix spanning residues 123–134 and an unstructured tail from residues 135–160. Our cryo-EM structure shows that the C-terminal α-helix binds the adjacent P12 subunit in the filament, thereby providing an element of cooperativity for filament formation (Fig. 3B I). In contrast, the C-terminal tail was not resolved in the density. We purified two truncated versions of P12, $P12_{(1–140)}$, and $P12_{(1–122)}$ and assessed their affinity to ssDNA (Fig. 4B). $P12_{(1–140)}$ lacks the disordered C-terminal tail after the α-helix. Intriguingly, this truncation leads to a significant reduction in DNA binding compared to full-length P12, thus indicating that the intrinsically disordered tail significantly contributes to cooperativity even though it is not resolved in the cryo-EM map. Truncating the tail *N*-terminally of the α-helix ($P12_{(1–122)}$) led to a total loss of detectable ssDNA binding, validating the essential role of the α-helix for cooperative binding predicted from our cryo-EM structure (Fig. 3B I).

We then quantified the binding characteristics of P12 using Hill modelling. We quantified band intensities in EMSAs to assess ssDNA-binding affinity and cooperativity across various substrates. The Hill coefficient (n) and dissociation constant (K_D) varied significantly depending on the substrate (Fig. 4C). For $(\text{non-repetitive})_{80}$ ssDNA, a high Hill coefficient ($n = 5.3$) indicated strong cooperative binding, with a K_D of 350 nM. In contrast, the $(dT)_{80}$ substrate exhibited moderate cooperativity ($n = 2.8$) and stronger binding affinity ($K_D = 63$ nM). These findings demonstrate that the observed variability in the binding dynamics of P12 is influenced by both substrate length and sequence with cooperativity being mediated by the C-terminal α-helix, which plays a critical role in the stabilisation of ssDNA (Fig. 4D).

Next, we investigated whether the ability of P12 to bind ssDNA would be connected to the reported P12 toxicity [12–14]. In addition to the ssDNA-binding deficient mutants F44A and F44A+K10A and the truncations $P12_{(1–122)}$ and $P12_{(1–140)}$ (Fig. 4A and B), we designed one additional truncated version of P12, $P12_{(1–148)}$. We tested the toxicity of each of these P12 variants in *E. coli* cells (Fig. 4E). BL21 (DE3) cells were transformed with a plasmid containing one of the P12 variants and then plated onto two different LB-agar plates, with expression-inducing or -repressing conditions. The difference in colony-forming units (CFUs) between the two plates serves as an indicator of the toxicity associated with each variant. The results show that the full-length, wild-type protein is highly toxic to *E. coli*, effectively killing the cells. Toxicity decreases with increasing deletion of the C-terminal tail and establishes a link between the DNA binding capacity of P12 (Fig. 4B) and its associated toxicity (Fig. 4E). Similarly, we observe that the two mutants with reduced ssDNA-binding capacity *in vitro* (Fig. 4A) also have a strongly decreased toxicity (Fig. 4E). The data thus clearly connect the toxic effect of P12 to its ssDNA-binding activity.

P12 protects ssDNA from nucleolytic degradation

To further explore the biological role of P12, we examined its efficiency to protect bound ssDNA from enzymatic degradation [12–14]. To this end, we tested whether saturating concentrations of P12 could shield ssDNA from degradation by Exonuclease I from *E. coli in vitro*. The degradation of ssDNA was followed and quantified by acrylamide gel electrophoresis (Fig. 5). The results show that P12 efficiently protects ssDNA from nucleolytic degradation, increasing the half-life from a few seconds in the absence of P12 to around 5 min, i.e. several orders of magnitude, when P12 was present (Fig. 5A). Notably, the protective effect of the filament was not complete, as we observed a slow ssDNA decay over time in the presence of P12 (Fig. 5B). This suggests a dynamic binding/unbinding equilibrium of P12 and ssDNA that allows the nuclease to access and degrade the ssDNA strand.

P12 adopts a new fold within the OB-like structure landscape

Intrigued by the distinct pincer-shaped structure of P12, we sought to determine how this novel fold positions itself within the broader protein fold space. To do so, we conducted structural searches across multiple fold classification databases, including ECOD [29] and CATH [30], to explore potential relationships between P12 and previously characterised protein folds (Fig. 6 and Supplementary Fig. S6). These searches did not identify any high-confidence or full-length matches to known protein folds, confirming the singularity of the P12 fold. We further tried to relate the P12 fold to pre-existing structures using the DALI server [33]. This search revealed local structural similarities to the single-stranded DNA binding domain of the gp32 protein from the Tequatrovirus T4 (PDB ID: 8GME, chain A, residues 24–193, Z-score 3.8), which has been described as displaying an OB-fold [40]. This finding prompted us to further examine the relationship between the P12 fold and the OB- and OB-like fold families, which notably include a variety of ssDNA-binding proteins. We therefore constructed a structure similarity network incorporating 117 experimental and 14 869 predicted structural domains (Fig. 6A and Supplementary Fig. S6A). This network reveals a diverse OB-like structure landscape with multiple clusters, representing groups of protein domains with similar structural features that may indicate shared functions or evolutionary origins. We specifically highlighted clusters containing experimentally solved structures from the PDB, offering validated reference points within the landscape.

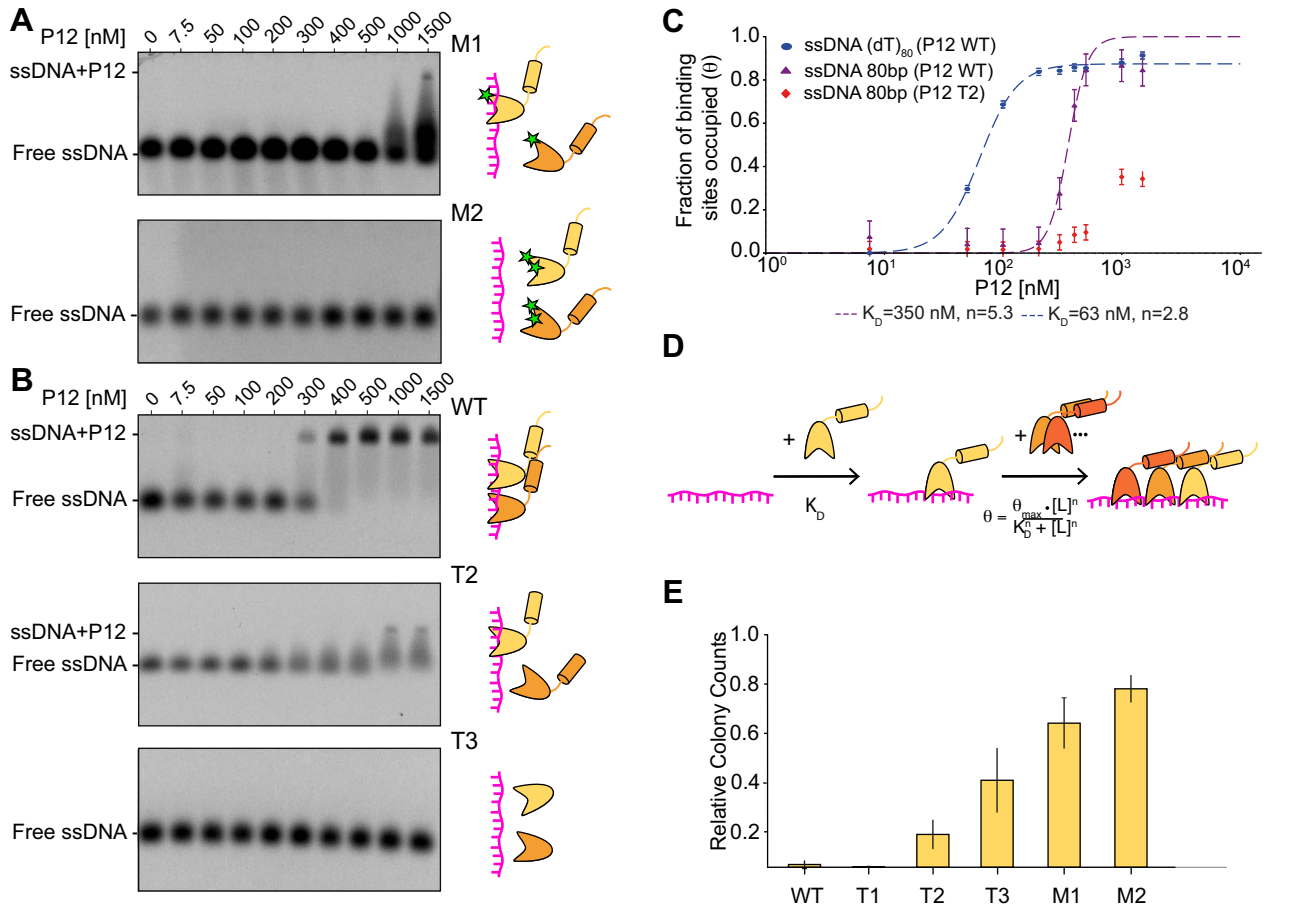


Figure 4. Cooperativity and role of the P12 C-terminus. (**A**) 7.5 nM of Cy5-labelled ssDNA(non-repeptitive)$_{80}$ substrate was titrated with increasing concentrations of P12 binding interface mutants. M1 = P12$_{(F44A)}$, M2 = P12$_{(F44A+K10A)}$. (**B**) 7.5 nM of Cy5-labelled ssDNA(non-repetitive)$_{80}$ substrate was titrated with increasing concentrations of C-terminally truncated variants of P12. WT = wild-type, T2 = P12$_{(1–140)}$, T3 = P12$_{(1–122)}$. (**C**) Data from **A** and **B** and Fig. 1B analysed with Hill binding model. The fraction of occupied binding sites is plotted against the protein concentration. Binding curves are shown for P12 WT with ssDNA(non-repetitive)$_{80}$ and (dC)$_{80}$, and for P12 T2 with ssDNA(non-repetitive)$_{80}$. (**D**) Schematic illustration of the Hill binding model used. (**E**) Relative colony counts of CFUs between induced and repressed plates reflecting the toxicity of the different P12 variants. WT = wild-type, T1 = P12$_{(1–148)}$, T2 = P12$_{(1–140)}$, T3 = P12$_{(1–122)}$, M1 = P12$_{(F44A)}$, M2 = P12$_{(F44A+K10A)}$). All experiments were conducted in triplicate. Data are presented as mean ± standard deviation.

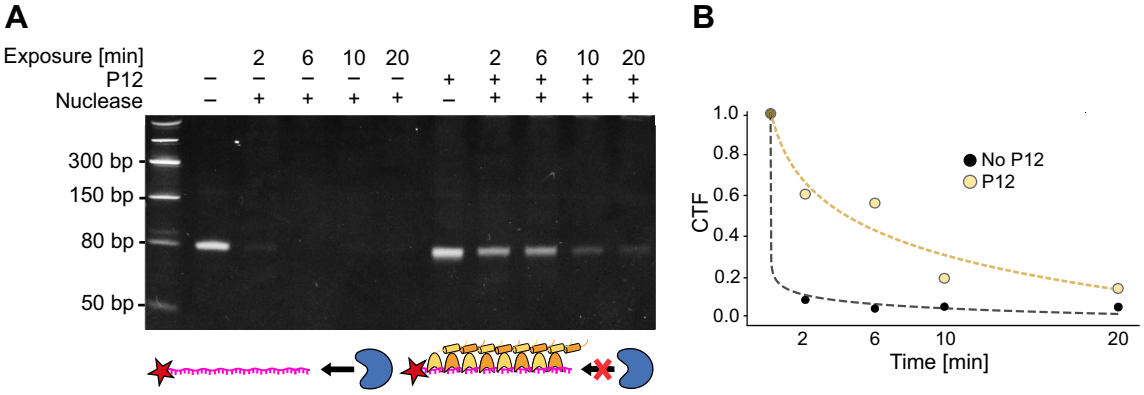


Figure 5. Functional role of P12 DNA-binding filaments. (**A**) Nuclease protection assay: 20 nM of Cy5-labelled ssDNA(non-repetitive)$_{80}$ substrate was exposed to Exonuclease I (*E. coli*) for varying incubation times, in the absence (left) and presence (right) of 50 μM of P12. (**B**) Quantification of intact ssDNA substrate over time. CTF = Corrected Total Fluorescence.

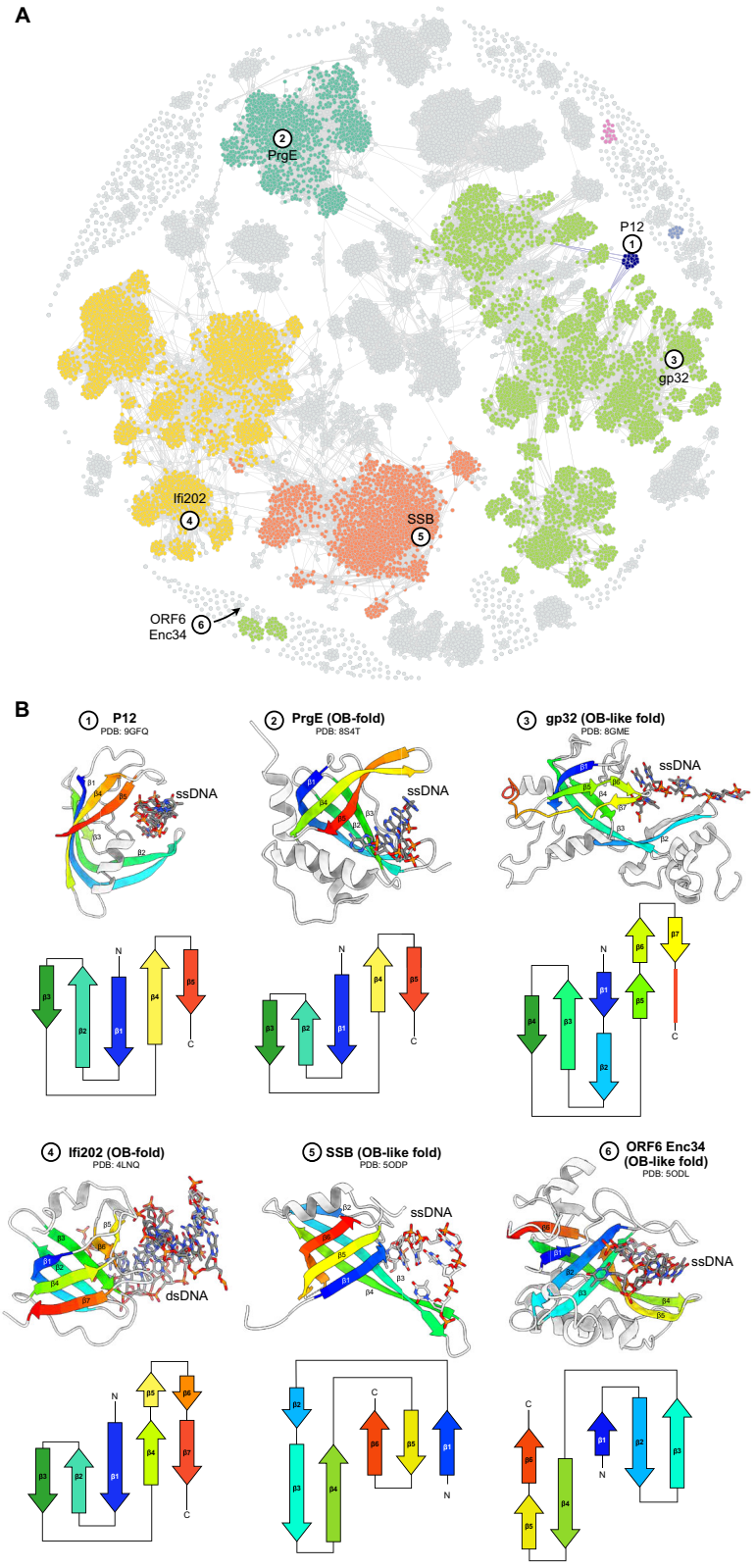


Figure 6. Structural and topological comparisons of the P12 fold. (**A**) Structure similarity network representing the fold-space of 14 986 P12-like domains with a maximum sequence identity of 50%, computed with Foldseek and Cosmograph. Points represent groups of homologous domains with a minimum sequence identity of 50% and gray lines Foldseek matches between their representatives at an *E*-value $< 1 \times 10^{-5}$. Matches at *E*-value $> 1 \times 10^{-5}$ involving the proteins in the P12 cluster are shown. Automatically identified clusters with at least one representative in the PDB are coloured, with key representatives highlighted for the clusters containing PDB models bound to nucleic acids. The two coloured clusters with PDB models lacking nucleic acids correspond to PDB entries 5M3K (chain E) and 1U5K (chain B). An interactive and annotated version of the network is available at: P12 structure space, and a further annotated version is displayed in Supplementary Fig. S6A. (**B**) Comparison of topological representations of P12 with five structurally similar representatives of DNA-binding domains with experimentally determined structures, selected from the clusters highlighted in **A**.

As can be seen in our interactive network (P12 structure space), P12 and its 33 homologs from a recently expanded collection of *Alphatectivirus* phages [9], along with 5 uncharacterised metagenomic proteins, form an independent cluster. As anticipated by DALI, the nearest cluster to the P12 cluster is that of the gp32-like folds. However, the connections to this hub are of low significance (E-value $> 1 \times 10^{-5}$), reflecting the fact that this cluster relies on two predicted, low-confidence metagenomic models with an average pLDDT of below 70 (Supplementary Fig. S6B). To further characterise the position of P12 within this protein fold landscape, we conducted a topological analysis, comparing the structure of P12 with known experimental structures representative of the various clusters within the network (Fig. 6B). From a topological perspective, the P12 fold bears a global similarity to several experimentally characterised single-stranded and double-stranded DNA-binding OB-like domains, including the plasmid-encoded PrgE [41], the Ifi202 HINa domain [42], and the ORF6 from phage Enc34 [43]. The P12 core β-sheet follows the canonical Greek key motif found in OB-folds [44]. However, the P12 fold differs markedly from these structures, as it does not form a closed β-barrel. In canonical β-barrels, binding occurs on the barrel surface, while in the P12 fold, a coiled β-sheet creates a distinct ssDNA-binding cleft. Such a cleft is also observed in gp32, but whereas gp32 binds to the nucleobases, P12 engages with the sugar-phosphate backbone. Overall, within the context of the millions of predicted structures currently available, our observations allow us to conclude that P12 represents a distinct and divergent OB-like fold family, exhibiting similarities to, yet significant differences from, gp32 and other known ssDNA-binding domains. The taxonomic conservation observed among PRD1-like phages indicates a unique structural and functional adaptation, presumably tailored for the specific purpose of protein-primed DNA replication in bacteriophages.

Discussion

SSBs are essential in all kingdoms of life, supporting critical cellular processes such as DNA replication, recombination, and repair [45]. Our study has resolved the structure of an SBB from a bacteriophage, the protein P12 from phage PRD1 [10]. The P12 core domain features five anti-parallel β-strands with an intervening α-helix that are arranged into a pincer-like conformation with a large binding cleft securing the ssDNA strand via interactions with the phosphate backbone. In addition, P12 possesses a C-terminal a-helix and a flexible tail connecting each protomer to adjacent subunits, thereby forming a filament along the ssDNA strand.

P12 exhibits high cooperativity when binding to ssDNA, as evidenced by a Hill coefficient of up to 5 and the absence of intermediate species observed in EMSA experiments at intermediate concentrations (Fig. 1). SSBs can exhibit varying degrees of cooperative behaviour, from non-cooperative [46] to highly cooperative binders [47–50]. The C-terminal α-helix and its adjacent disordered tail function as a strong element of cooperativity in P12 (Fig. 4). The prevalence of such disordered tails in SSBs has been extensively documented, with studies underscoring their role in cooperativity [51], in regulating interactions with ssDNA [46, 52, 53] and in engaging with proteins involved in DNA metabolism, such as polymerases and helicases [51, 54–56]. While our cryo-EM structure highlights the critical role of the C-terminal α-helix in filament formation, the mechanisms by which the intrinsically disordered tail contributes to cooperative binding remains unclear (Fig. 4B). Notably, P12 possesses a basic tip at its C-terminus—a feature conserved across all PRD1-related phages and distinctly different from the acidic tips commonly observed in many other SSBs [51, 57, 58]. This unique characteristic raises intriguing questions about its functional significance, including its potential role in mediating dynamic interactions with ssDNA or between P12 units.

A key function of SSBs is their ability to protect ssDNA from nucleolytic degradation [10, 45] contributing to essential genetic processes. P12 filaments protect ssDNA against nuclease degradation (Fig. 5), suggesting a role in stabilising ssDNA during the infection of *E. coli* by bacteriophage PRD1. The protective role of P12 in PRD1 is consistent with that of other bacteriophage SSBs involved in protein-primed replication, such as those from Φ29 [59]. Given the strong preference of P12 for ssDNA over other nucleic acid substrates (Fig. 1), it is likely involved in organising and protecting ssDNA produced during the replication of the phage genome. This function resembles that of SSBs from bacteriophages Φ29, Nf, and GA-1, which bind along displaced ssDNA during protein-primed replication [59]. These SSBs primarily stimulate dNTP incorporation by stabilising ssDNA and preventing the formation of secondary structures, thereby reducing the likelihood of non-productive binding of the DNA polymerase to ssDNA.

Importantly, we observe that the ssDNA-binding activity of P12 has detrimental effects on host cell viability (Fig. 4). Both the deletion of the C-terminal tail and mutations in the ssDNA-binding site of P12 alleviate toxicity. This effect correlates with the ssDNA-binding affinity of P12, indicating that P12-induced toxicity is directly tied to its capacity to form ssDNA-binding filaments. A plausible explanation for this observation is that P12 interferes with host DNA metabolism by binding to ssDNA, leading to replication stress and genomic instability. Whether this toxicity serves as an intended functional role in the phage life cycle or is merely a side effect of this essential protein for the phage remains an open question.

From a structural perspective, P12 represents an important exception from the well-characterised OB-fold SSBs (Fig. 6). Our computational and structural analyses suggest that P12 represents a distinct evolutionary branch of ssDNA-binding proteins, emphasising the diversity of DNA-binding strategies. SSBs, most of which display an OB-like fold, interact with ssDNA in various binding arrangements, either forming filamentous structures [41, 43, 60, 61] or organising ssDNA around higher-order oligomeric species [51]. High-resolution models of such complexes are limited to a few structures, including *E. coli* SSB [62], plasmid-encoded PrgE [41], ORF6 from phage Enc34 [43], HaLEF-3 from baculovirus [60], and eukaryotic RAD52 [61].

Overall, our structural and functional studies give rise to the following model for the role of P12 in PRD1 genome replication (Fig. 7). Following the initiation of protein-primed DNA replication, a daughter strand is synthesised in the 5′ to 3′ direction from the double-stranded linear genome with P12 stabilising the emerging parental ssDNA strand. As the replication fork progresses, the daughter strand replaces the parental strand [63–65]. In classical DNA replication, synthesis of the complementary strand proceeds via discontinuous synthesis of Okazaki fragments. In protein-primed DNA replication, we propose that replication of the complementary strand occurs through two potential, nonexclusive mech-

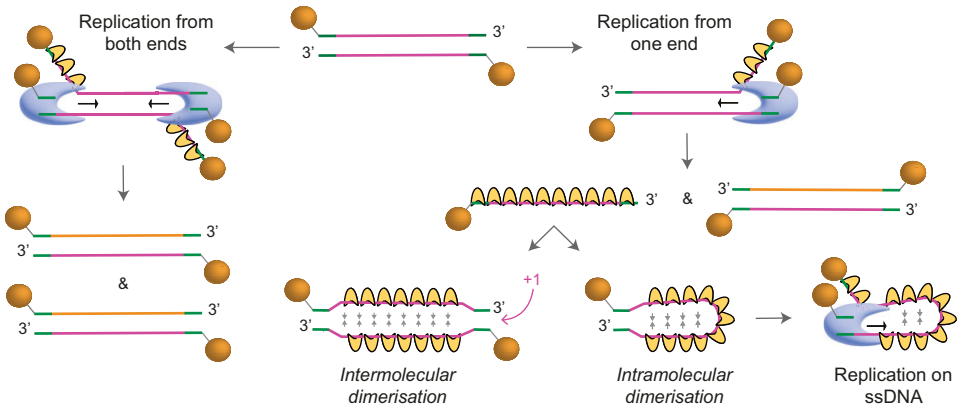


Figure 7. Model for P12-assisted DNA replication. The DNA polymerase (moon shaped, blue) initiates replication of the dsDNA phage genome either from both ends of the genome (left path) or asynchronously (right path), leading to a single DNA strand that is stabilised by P12 (small curved shapes, yellow). P12 then either promotes the intermolecular hybridisation of this strand with a complementary single strand from another replication event (left), or the intramolecular hybridisation to form of a panhandle structure (right). The presence of inverted terminal repeats (ITRs) at the genome ends (green) promotes the formation of such panhandle structures. TPs (balls, orange) are attached to the 5′ end of each DNA strand and promote DNA replication either from dsDNA or panhandle structures.

anisms (Fig. 7). In one scenario, replication initiates simultaneously from both ends, yielding two double-stranded DNA copies. Alternatively, replication may proceed from a single end, producing one double-stranded daughter duplex and a remaining parental ssDNA strand stabilised by P12. The fate of this displaced ssDNA strand remains a subject of ongoing discussions. Research on adenovirus and PRD1 replication has proposed that this displaced strand might undergo intramolecular hybridisation to form panhandle structures, facilitated by the presence of inverted terminal repeats (ITRs) at the ends of the linear genome [63, 65]. Such panhandle structures have not only been a popular model to explain protein-primed replication initiation from a displaced strand but also to describe homology-restoring mechanisms [65, 66]. However, direct evidence of such structures is still lacking. The anti-parallel double filament of P12–ssDNA complexes observed in this study raises the question of the involvement of P12 in the subsequent utilisation of this single ssDNA strand (Fig. 7). The propensity of the P12 filament to dimerise, as evidenced by our cryo-EM structure, could promote the formation of an anti-parallel double filament through interactions between opposing protein units (Supplementary Fig. S5A and B). These weak interactions allow the two strands to diffuse over one another, ultimately leading to the favourable annealing of the ITRs regions. If this mechanism holds true, P12 could be categorised as an annealase, akin to lambda Redβ [67], HSV-1 ICP8 [68], and human Rad52 [61], which catalyse ssDNA annealing without requiring ATP. While annealase activity in P12 remains to be confirmed, such a role would further highlight its role in PRD1 protein-primed DNA replication. P12 might indeed promote intramolecular and intermolecular ITR annealing, mediating the formation of either a double-stranded DNA genome or a panhandle structure, both of which are compatible with subsequent rounds of replication by the PRD1 DNA polymerase. This would enable protein-primed replication to proceed from a single genome end without loss of ssDNA, thereby ensuring efficient duplication. The mechanism by which P12 is removed during replication remains unknown. The disordered nature of the P12 tail could render it particularly susceptible to proteolytic cleavage which, given its crucial role in cooperative binding, may lead to filament destabilisation over time. Filament destabilisation, driven by changes in helical pitch and substrate elongation as seen with gp32 in T4 [53], is also conceivable for P12.

In conclusion, our study provides structural and functional insights into the PRD1 SSB P12, revealing a unique fold and binding mechanism. The cooperative binding of P12 to ssDNA, its role in ssDNA protection, and its associated host toxicity underscore its central importance in PRD1 replication. The structural findings thus lay the foundation for future studies aimed at elucidating the broader regulatory functions of P12, including its potential interactions with other replication proteins, such as the phage DNA polymerase, and its role in the overall bacteriophage–host interaction. More broadly, the relevance of this work is underscored by a recent publication showcasing the development of an orthogonal DNA replication system inspired by the PRD1 replication machinery [4]. This orthogonal DNA replication system, where P12 plays an essential role, has received considerable attention due to its demonstrated ability to enable continuous protein evolution in *E. coli*. The innovative use of this system highlights its transformative potential for synthetic biology, particularly in the development of directed evolution approaches. The structural and mechanistic insights into P12 presented in this manuscript provide a molecular framework to optimise and enhance the capabilities of such systems, paving the way for more reliable and user-friendly platforms, bringing them closer to realising their full potential.

Acknowledgements

We thank S. Panke for helpful discussions and sharing of facilities. We are grateful to P. Spies for his help with titration experiments. We kindly acknowledge P. Afanasyev (ScopeM) and the BioEM Lab of the Biozentrum, University of Basel, for their support. We also thank Torsten Schwede, the members of the Schwede lab and sciCORE at the University of Basel for providing computational resources and support.

Author contributions: L.K.T. and M.D. contributed equally. L.K.T. expressed all proteins used in this study, froze cryo-EM grids, set up crystallisation screens, and conducted EMSAs and

toxicity tests. R.D.T. and L.K.T. collected the crystallographic data and solved the structure. M.D. determined the cryo-EM structure. N.H.-D. cloned initial P12 constructs and replicated toxicity tests. J.D. and J.P. carried out the bioinformatic analysis of the P12 structure space. N.H.-D. supervised L.K.T. and S.H. supervised M.D. and R.D.T.. N.H.-D. defined the direction of research. All authors analysed and interpreted data. L.K.T., M.D., J.P., S.H., and N.H.-D. wrote the paper with input from all authors.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

The authors declare no competing interests.

Funding

This work was funded by a Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNSF) Ambizione grant [PZ00P3_202 090] to N.H.-D. and by the Swiss Nanoscience Institute to S.H. Authors J.P. and J.D. were supported by funding from the SIB—Swiss Institute of Bioinformatics and the Biozentrum, University of Basel. Funding to pay the Open Access publication charges for this article was provided by Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung.

Data availability

Model coordinates for the P12 apo crystal structure have been deposited in the Protein Data Bank (PDB ID: 9GFQ). The cryo-EM map and model for P12 bound to $(dT)_{80}$ ssDNA substrate have been deposited in the PDB and Electron Microscopy Data Bank (EMDB) under accession codes 9I86 and 52708, respectively. Cryo-EM maps for P12 bound to either the $(ATGCT)_{80}$ or $(non\text{-}repetitive)_{80}$ ssDNA substrates have been deposited in the EMDB under accession codes 52709 and 52710, respectively. The metadata associated with the bioinformatic analysis of the P12 structure space is available at Zenodo: https://zenodo.org/records/14500985.

Code availability

The code used in this study has been deposited in GitHub and Zenodo and can be accessed through the following link: https://zenodo.org/records/14500985.

References

1. Salas M. Protein-priming of dna replication. *Annu Rev Biochem* 1991;**60**:39–71. https://doi.org/10.1146/annurev.bi.60.070191.000351
2. Ravikumar A, Arzumanyan GA, Obadi MKA *et al.* Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell* 2018;**175**:1946–57. https://doi.org/10.1016/j.cell.2018.10.021
3. Tian R, Zhao R, Guo H *et al.* Engineered bacterial orthogonal DNA replication system for continuous evolution. *Nat Chem Biol* 2023;**19**:1504–12. https://doi.org/10.1038/s41589-023-01387-2
4. Tian R, Rehm FBH, Czernecki D *et al.* Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. coli*. *Science (1979)* 2024;**383**:421–6.
5. Williams RL, Liu CC. Accelerated evolution of chosen genes. *Science (1979)* 2024;**383**:372–3.
6. Salas M, Vega M . 2016; Chapter four - protein-primed replication of bacteriophage Φ29 DNA. In: Kaguni LS, Oliveira MT (eds.), *The Enzymes*. Academic Press, **39**, 137–67.
7. Hoeben RC, Uil TG. Adenovirus DNA replication. *Cold Spring Harb Perspect Biol* 2013;**5**:a013003. https://doi.org/10.1101/cshperspect.a013003
8. Cronan JE. *Escherichia coli* as an experimental organism. In *eLS*. Chichester, West Sussex, England: John Wiley & Sons, Ltd, 2014. https://doi.org/10.1002/9780470015902.a0002026.pub2
9. Quinones-Olvera N, Owen SV, McCully LM *et al.* Diverse and abundant phages exploit conjugative plasmids. *Nat Commun* 2024;**15**:3197. https://doi.org/10.1038/s41467-024-47416-z
10. Pakula TM, Caldentey J, Serrano M *et al.* Characterization of a DNA binding protein of bacteriophage PRD1 involved in DNA replication. *Nucleic Acids Res* 1990;**18**:6553–7. https://doi.org/10.1093/nar/18.22.6553
11. Pakula TM, Caldentey J, Gutiérrez C *et al.* Overproduction, purification, and characterization of DNA-binding protein P19 of bacteriophage PRD1. *Gene* 1993;**126**:99–104. https://doi.org/10.1016/0378-1119(93)90595-T
12. Mindich L, Bamford D, Goldthwaite C *et al.* Isolation of nonsense mutants of lipid-containing bacteriophage PRD1. *J Virol* 1982;**44**:1013–20. https://doi.org/10.1128/jvi.44.3.1013-1020.1982
13. McGraw T, Yang H-L, Mindich L. Establishment of a physical and genetic map for bacteriophage PRD1. *Mol Gen Genet* 1983;**190**:237–44. https://doi.org/10.1007/BF00330646
14. Gerendasy D, Ito J. Nucleotide sequence and transcription of the right early region of bacteriophage PRD1. *J Bacteriol* 1990;**172**:1889–98. https://doi.org/10.1128/jb.172.4.1889-1898.1990
15. Punjani A, Rubinstein JL, Fleet DJ *et al.* cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 2017;**14**:290–6. https://doi.org/10.1038/nmeth.4169
16. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* 2021;**30**:70–82. https://doi.org/10.1002/pro.3943
17. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;**60**:2126–32. https://doi.org/10.1107/S0907444904019158
18. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
19. Croll TI. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* 2018;**74**:519–30. https://doi.org/10.1107/S2059798318002425
20. Williams CJ, Headd JJ, Moriarty NW *et al.* MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci* 2018;**27**:293–315. https://doi.org/10.1002/pro.3330
21. Potterton L, Agirre J, Ballard C *et al.* CCP4 i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* 2018;**74**:68–84. https://doi.org/10.1107/S2059798317016035
22. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:125–32. https://doi.org/10.1107/S0907444909047337
23. McCoy AJ, Grosse-Kunstleve RW, Adams PD *et al.* Phaser crystallographic software. *J Appl Crystallogr* 2007;**40**:658–74. https://doi.org/10.1107/S0021889807021206
24. Liebschner D, Afonine PV, Baker ML *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct*

Biol 2019;**75**:861–77. https://doi.org/10.1107/S2059798319011471
25. Murshudov GN, Skubák P, Lebedev AA *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 2011;**67**:355–67. https://doi.org/10.1107/S0907444911001314
26. Berman HM, Westbrook J, Feng Z *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000;**28**:235–42. https://doi.org/10.1093/nar/28.1.235
27. Schindelin J, Arganda-Carreras I, Frise E *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;**9**:676–82. https://doi.org/10.1038/nmeth.2019
28. Virtanen P, Gommers R, Oliphant TE *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;**17**:261–72. https://doi.org/10.1038/s41592-019-0686-2
29. Cheng H, Schaeffer RD, Liao Y *et al.* ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol* 2014;**10**:e1003926. https://doi.org/10.1371/journal.pcbi.1003926
30. Orengo CA, Michie AD, Jones S *et al.* CATH – a hierarchic classification of protein domain structures. *Structure* 1997;**5**:1093–109. https://doi.org/10.1016/S0969-2126(97)00260-8
31. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;**33**:2302–9. https://doi.org/10.1093/nar/gki524
32. van Kempen M, Kim SS, Tumescheit C *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 2024;**42**:243–6. https://doi.org/10.1038/s41587-023-01773-0
33. Holm L. DALI and the persistence of protein shape. *Protein Sci* 2020;**29**:128–40. https://doi.org/10.1002/pro.3749
34. Varadi M, Bertoni D, Magana P *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* 2024;**52**:D368–75. https://doi.org/10.1093/nar/gkad1011
35. Lin Z, Akin H, Rao R *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science (1979)* 2023;**379**:1123–30.
36. Wells J, Hawkins-Hooker A, Bordin N *et al.* Chainsaw: protein domain segmentation with fully convolutional neural networks. *Bioinformatics* 2024;**40**:btae296. https://doi.org/10.1093/bioinformatics/btae296
37. Kim RS, Karin EL, Steinegger M *et al.* BFVD—a large repository of predicted viral protein structures. *Nucleic Acids Res* 2025; **53**:D340-7. https://doi.org/10.1093/nar/gkae1119
38. Rokotyan N, Stukova O, Kolmakova D *et al.* Cosmograph: GPU-accelerated Force Graph Layout and Rendering [Computer software]. https://cosmograph.app/
39. Chakraborty S, Sharma S, Maiti PK *et al.* The poly dA helix: a new structural motif for high performance DNA-based molecular switches. *Nucleic Acids Res* 2009;**37**:2810–7. https://doi.org/10.1093/nar/gkp133
40. He X, Yun M-K, Li Z *et al.* Structural and functional insights into the interaction between the bacteriophage T4 DNA processing proteins gp32 and dda. *Nucleic Acids Res* 2024; https://doi.org/10.1093/nar/gkae910
41. Breidenstein A, Lamy A, Bader CPJ *et al.* PrgE: an OB-fold protein from plasmid pCF10 with striking differences to prototypical bacterial SSBs. *Life Sci Alliance* 2024;**7**:12748–62. https://doi.org/10.26508/lsa.202402693
42. Li H, Wang J, Wang J *et al.* Structural mechanism of DNA recognition by the p202 HINa domain: insights into the inhibition of Aim2-mediated inflammatory signalling. *Acta Crystallogr Section F* 2014;**70**:21–9.
43. Cernooka E, Rumnieks J, Tars K *et al.* Structural basis for DNA recognition of a single-stranded DNA-binding protein from Enterobacter Phage Enc34. *Sci Rep* 2017;**7**:15529. https://doi.org/10.1038/s41598-017-15774-y
44. Theobald DL, Mitton-Fry RM, Wuttke DS. Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct* 2003;**32**:115–33. https://doi.org/10.1146/annurev.biophys.32.110601.142506
45. Pestryakov PE, Lavrik OI. Mechanisms of single-stranded DNA-binding protein functioning in cellular DNA metabolism. *Biochem Moscow* 2008;**73**:1388–404. https://doi.org/10.1134/S0006297908130026
46. Hernandez AJ, Richardson CC. Gp2.5, the multifunctional bacteriophage T7 single-stranded DNA binding protein. *Semin Cell Dev Biol* 2019;**86**:92–101. https://doi.org/10.1016/j.semcdb.2018.03.018
47. Ruyechan WT, Wetmur JG. Cooperative binding of the *Escherichia coli* DNA unwinding protein to single-stranded DNA. *Biochemistry* 1975;**14**:5529–34. https://doi.org/10.1021/bi00696a023
48. Kowalczykowski SC, Lonberg N, Newport JW *et al.* Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids: I. Characterization of the binding interactions. *J Mol Biol* 1981;**145**:75–104. https://doi.org/10.1016/0022-2836(81)90335-1
49. Kim C, Paulus BF, Wold MS. Interactions of human replication protein A with oligonucleotides. *Biochemistry* 1994;**33**:14197–206. https://doi.org/10.1021/bi00251a031
50. Cashen BA, Morse M, Rouzina I *et al.* Dynamic structure of T4 gene 32 protein filaments facilitates rapid noncooperative protein dissociation. *Nucleic Acids Res* 2023;**51**:8587–605. https://doi.org/10.1093/nar/gkad595
51. Bonde NJ, Kozlov AG, Cox MM *et al.* Molecular insights into the prototypical single-stranded DNA-binding protein from E. coli. *Crit Rev Biochem Mol Biol* 2024;**59**:99–127. https://doi.org/10.1080/10409238.2024.2330372
52. Pant K, Anderson B, Perdana H *et al.* The role of the C-domain of bacteriophage T4 gene 32 protein in ssDNA binding and dsDNA helix-destabilization: kinetic, single-molecule, and cross-linking studies. *PLoS One* 2018;**13**:e0194357. https://doi.org/10.1371/journal.pone.0194357
53. Cashen BA, Morse M, Rouzina I *et al.* C-terminal domain of T4 gene 32 protein enables rapid filament reorganization and dissociation. *J Mol Biol* 2024;**436**:168544. https://doi.org/10.1016/j.jmb.2024.168544
54. Bianco PR. The tale of SSB. *Prog Biophys Mol Biol* 2017;**127**:111–8. https://doi.org/10.1016/j.pbiomolbio.2016.11.001
55. Cadman CJ, McGlynn P. PriA helicase and SSB interact physically and functionally. *Nucleic Acids Res* 2004;**32**:6378–87. https://doi.org/10.1093/nar/gkh980
56. Glover BP, McHenry CS. The χψ subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template*. *J Biol Chem* 1998;**273**:23476–84. https://doi.org/10.1074/jbc.273.36.23476
57. Kim YT, Tabor S, Bortner C *et al.* Purification and characterization of the bacteriophage T7 gene 2.5 protein. A single-stranded DNA-binding protein. *J Biol Chem* 1992;**267**:15022–31. https://doi.org/10.1016/S0021-9258(18)42141-2
58. Shokri L, Rouzina I, Williams MC. Interaction of bacteriophage T4 and T7 single-stranded DNA-binding proteins with DNA. *Phys Biol* 2009;**6**:025002. https://doi.org/10.1088/1478-3975/6/2/025002
59. Gascón I, Lázaro JM, Salas M. Differential functional behavior of viral ϕ29, nf and GA-1 SSB proteins. *Nucleic Acids Res* 2000;**28**:2034–42. https://doi.org/10.1093/nar/28.10.2034
60. Yin J, Fu Y, Rao G *et al.* Structural transitions during the cooperative assembly of baculovirus single-stranded DNA-binding

protein on ssDNA. *Nucleic Acids Res* 2022;**50**:13100–13. https://doi.org/10.1093/nar/gkac1142
61. Liang C-C, Greenhough LA, Masino L *et al*. Mechanism of single-stranded DNA annealing by RAD52–RPA complex. *Nature* 2024;**629**:697–703. https://doi.org/10.1038/s41586-024-07347-7
62. Shishmarev D, Wang Y, Mason CE *et al*. Intramolecular binding mode of the C-terminus of *Escherichia coli* single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Res* 2014;**42**:2750–7. https://doi.org/10.1093/nar/gkt1238
63. Lechner RL, Kelly TJ Jr. The structure of replicating adenovirus 2 DNA molecules. *Cell* 1977;**12**:1007–20. https://doi.org/10.1016/0092-8674(77)90165-9
64. Yoshikawa H, Garvey KJ, Ito J. Nucleotide sequence analysis of DNA replication origins of the small *Bacillus* bacteriophages: evolutionary relationships. *Gene* 1985;**37**:125–30. https://doi.org/10.1016/0378-1119(85)90264-1
65. Savilahti H, Bamford DH. Protein-primed DNA replication: role of inverted terminal repeats in the *Escherichia coli* bacteriophage PRD1 life cycle. *J Virol* 1993;**67**:4696–703. https://doi.org/10.1128/jvi.67.8.4696-4703.1993
66. Wang K, Xu F-Y, Ahern KG *et al*. Inverted repeats direct repair of adenovirus minichromosome ends. *Virology* 1991;**183**:44–51. https://doi.org/10.1016/0042-6822(91)90116-S
67. Newing TP, Brewster JL, Fitschen LJ *et al*. Redβ177 annealase structure reveals details of oligomerization and λ red-mediated homologous DNA recombination. *Nat Commun* 2022;**13**:5649. https://doi.org/10.1038/s41467-022-33090-6
68. Tolun G, Makhov AM, Ludtke SJ *et al*. Details of ssDNA annealing revealed by an HSV-1 ICP8–ssDNA binary complex. *Nucleic Acids Res* 2013;**41**:5927–37. https://doi.org/10.1093/nar/gkt266

Received: December 12, 2024. **Revised:** February 5, 2025. **Editorial Decision:** February 7, 2025. **Accepted:** February 11, 2025