432–444 *Nucleic Acids Research, 2020, Vol. 48, No. 1* *Published online 12 November 2019*
doi: 10.1093/nar/gkz1045

Structural insight into the length-dependent binding of ssDNA by SP_0782 from *Streptococcus pneumoniae*, reveals a divergence in the DNA-binding interface of PC4-like proteins

Shuangli Li[1,3,†], Guoliang Lu[2,3,†], Xiang Fang[2,3], Theresa A. Ramelot[4], Michael A. Kennedy[4], Xin Zhou[1], Peng Gong[2], Xu Zhang[1], Maili Liu[1], Jiang Zhu[1,*] and Yunhuang Yang[1,*]

[1]State Key Laboratory of Magnetic Resonance and Atomic Molecular Physics, Key Laboratory of Magnetic Resonance in Biological Systems, National Center for Magnetic Resonance in Wuhan, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan National Laboratory for Optoelectronics, Wuhan 430071, China, [2]Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan 430071, China, [3]University of Chinese Academy of Sciences, Beijing 100049, China and [4]Department of Chemistry and Biochemistry, and the Northeast Structural Genomics Consortium, Miami University, Oxford, OH 45056, USA

Received May 13, 2019; Revised September 30, 2019; Editorial Decision October 20, 2019; Accepted November 07, 2019

ABSTRACT

SP_0782 from *Streptococcus pneumoniae* is a dimeric protein that potentially binds with single-stranded DNA (ssDNA) in a manner similar to human PC4, the prototype of PC4-like proteins, which plays roles in transcription and maintenance of genome stability. In a previous NMR study, SP_0782 exhibited an ssDNA-binding property different from YdbC, a prokaryotic PC4-like protein from *Lactococcus lactis*, but the underlying mechanism remains unclear. Here, we show that although SP_0782 adopts an overall fold similar to those of PC4 and YdbC, the ssDNA length occupied by SP_0782 is shorter than those occupied by PC4 and YdbC. SP_0782 exhibits varied binding patterns for different lengths of ssDNA, and tends to form large complexes with ssDNA in a potential high-density binding manner. The structures of SP_0782 complexed with different ssDNAs reveal that the varied binding patterns are associated with distinct capture of nucleotides in two major DNA-binding regions of SP_0782. Moreover, a comparison of known structures of PC4-like proteins complexed with ssDNA reveals a divergence in the binding interface between prokaryotic and eukaryotic PC4-like proteins. This study provides insights into the ssDNA-binding mechanism of PC4-like proteins, and benefits further study regarding the biological function of SP_0782, probably in DNA protection and natural transformation.

INTRODUCTION

In order to maintain the integrity of genetic information, cells employ many types of single-stranded DNA binding proteins (SSBs) to protect single-stranded DNA (ssDNA) from damage or to recruit repair proteins to the DNA damage site (1–3). The various ssDNA-binding domains engaged by different SSBs can be distinguished according to their structures. The well-studied ssDNA-binding domains are structurally characterized as oligosaccharide/oligonucleotide-binding (OB) folds (4), K homology (KH) domains (5), RNA recognition motifs (RRMs) (6) and whirly domains (7). OB folds usually consist of a five-stranded β-sheet and one α-helix (4), and assemble into multimers, such as the well-known homotetramer formed by the SSB from *Escherichia coli* (EcSSB) (8). There are also OB fold-derived structures that form homodimers to bind with ssDNA, such as those from the positive cofactor 4 (PC4)/Sub1-like proteins (9).

PC4/Sub1 was initially identified as a transcriptional coactivator, and later on its roles in chromatin condensation, maintenance of genome stability and DNA repair were implicated (reviewed in (3)). The C-terminal domain (CTD) of human PC4 is an ssDNA-binding domain that can dimerize to bind with ssDNA or melted double-stranded

*To whom correspondence should be addressed. Tel: +86 27 87197119; Fax: +86 27 87199543; Email: jiangzhu@wipm.ac.cn
Correspondence may also be addressed to Yunhuang Yang. Tel: +86 27 87199541; Fax: +86 27 87199543; Email: yang_yh@wipm.ac.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present address: Guoliang Lu, School of Life Sciences, Fudan University, Shanghai 200438, China.

DNA (dsDNA) in a sequence-independent manner (10,11). In a PC4-CTD dimer, each subunit consists of one α-helix and a four-stranded β-sheet that forms a concave for grasping five nucleotides in one strand of DNA. Two subunits cooperate in binding with each of two opposing DNA strands through hydrogen bonds, hydrophobic interaction and electric charges (9). The ssDNA bound by PC4 should be long enough to bend into two antiparallel strands, as PC4 shows low binding affinity to ssDNA shorter than 10 nt (10). On the other hand, for ssDNA with a length of 28 nt, two dimers of PC4 will bind with it (12,13). PC4 is evolutionarily conserved and present in other eukaryotes, including yeast and *Magnaporthe oryzae*, as well as in prokaryotes and Bacteriophage T5 (14–17). Its homologs MoSub1 in *M. oryzae,* YdbC in *Lactococcus lactis*, SSB in Bacteriophage T5, and BPSL1147 in *Burkholderia pseudomallei*, all exhibit a quaternary structure of the homodimer (15–18). The binding property of YdbC with ssDNA dT19G1 is similar to that of human PC4-CTD, and it can bind with ssDNA of 20 nt in a two-dimer-bound form, although the affinity of the second dimer is much weaker (16). Homolog searching with YdbC identified hundreds of prokaryotic PC4-like proteins with unknown structure and function, which may have distinct ssDNA-binding properties. In fact, EF_3132 from *Enterococcus faecalis* shows dynamics of conformational exchange between DNA-free and DNA-bound states significantly different from that of YdbC (16). Thus, the structure and function of prokaryotic PC4-like proteins is a topic that demands further broad and deep exploration.

Streptococcus pneumoniae is a Gram-positive and significant human pathogenic bacterium. There are two homologs of EcSSB assembling into homotetramers but showing different ssDNA-binding properties in *S. pneumoniae*, designated as SpSsbA and SpSsbB (19). However, the potential homolog of PC4 and YdbC was not well characterized. We previously reported an NMR study on SP_0782 from *S. pneumoniae*, which is 56.2% identical to YdbC in sequence (20). SP_0782 exists as a homodimer in solution, similar to PC4-CTD and YdbC. However, it exhibited a different binding property toward the ssDNA dT19G1 in NMR titration experiments compared to YdbC. To better understand the underlying mechanism for the different binding property, in this study, we investigated the ssDNA binding of SP_0782 in detail, and determined the structures of apo SP_0782 and SP_0782:ssDNA complexes, which reveal a distinctive ssDNA-binding mechanism of SP_0782 from those of PC4 and YdbC, and suggest a variation of PC4-like proteins in the ssDNA-binding interface. Moreover, the DNA-binding patterns of SP_0782 for different lengths of ssDNA were revealed, enlightening the mechanism by which SP_0782 assembles at the exposed single-strand regions of genomic DNA that are variable in length.

MATERIALS AND METHODS

Sample preparation

The expression and purification of SP_0782 (7–79) wild-type and W30A mutant protein were carried out using the method previously described (20). The protein sample was concentrated to 0.7 mM in 10% D_2O/90% H_2O (v/v), 20 mM NH_4OAc, 100 mM NaCl, 10 mM dithiothreitol (DTT), 5 mM $CaCl_2$, at pH 4.5 for NMR experiments. Single-stranded DNA (ssDNA) dT6, dTCTTCC, dT12 and dT19G1 were purchased from Sangon Biotech Co., purified through high performance liquid chromatography (HPLC), and then dissolved with desired buffer.

NMR structure calculation and NMR titration

All NMR data for structure calculation were collected at 293 K using 600 MHz Varian Inova and 850 MHz Bruker Avance NMR spectrometers equipped with 5-mm cryogenic probes, processed using NMRPipe and analyzed with Sparky program. Backbone and side chain resonance assignments were described previously (20) and deposited in the BioMagResBank (BMRB) with an accession number 17175. NOE-based inter-proton distance restraints were determined automatically using CYANA 2.1. Input for CYANA consisted of chemical shift assignments, NOESY peak lists from three NOESY spectra with peak intensities, the restraints for backbone phi (φ) and psi (ψ) torsion angles derived from chemical shifts of backbone atoms using the TALOS+ software program (21). Manual and iterative refinements of NOESY peak picking lists were guided using NMR RPF quality to assess 'goodness of fit' between calculated structures and NOESY peak lists (22). The 20 lowest energy structures calculated by CYANA 2.1 were further refined using restrained molecular dynamics in explicit water CNS 1.2 (23) and the PARAM19 force field, using the final NOE-derived distance constraints and TALOS-derived dihedral angle restraints. The final NMR ensemble of 20 structures was deposited in the Protein Data Bank (PDB, ID: 2L3A). Structural statistics and global structure quality factors were computed using PSVS version 1.5 (24) (Supplementary Table S1). NMR titrations were performed on a 600 MHz Bruker spectrometer at 293 K. The samples containing 0.4 mM SP_0782 and different concentrations of dT6, dT12 or dT19G1 were pre-mixed and allowed to equilibrate for 1 h, and subsequently used for ^{1}H–^{15}N HSQC spectrum collection. Chemical shift perturbations (CSPs) were calculated using the equation proposed in a previous study (25).

Isothermal titration calorimetry (ITC)

ITC was performed on a VP-ITC instrument (GE Healthcare) at 293 K. All samples were prepared in the same buffer as that used in the NMR experiments to avoid any heat change resulted from mixing buffers. Solutions containing 100 μM of dT6, dT12 or dT19G1 were titrated into the reservoir containing 10 μM of SP_0782, respectively. Each titration experiment consisted of a preliminary injection of 2 μl followed by 27 injections of 10 μl, with a duration of 20 s and a delay of 180 s between each injection. Thermogram analysis was performed using MicroCal Origin 7.0 (Microcal Software Inc.) with a one-site binding model. The first data point was not included when the titration curves were fitted. Dilution heat was subtracted with a ctrl titration through injecting the DNA solution into buffer solution alone for each experiment. The ITC experiments were repeated in triplicate.

Electrophoretic mobility shift assay (EMSA)

EMSA experiments between SP_0782 and ssDNA probes with biotin label at 5′-end were carried out using a Chemiluminescent EMSA Kit (Beyotime Biotech Co.). The samples containing 0.1 pmol ssDNA probes (final concentration of 0.01 μM in a 10 μl binding mixture) and different contents of SP_0782 were premixed and incubated for 25 min at room temperature, and subsequently loaded into a 10% native PAGE gel in 0.5× TBE buffer. After separation by electrophoresis, the free probe and the complex of probe and protein were transferred onto a nylon membrane and subjected to UV crosslinking. Subsequently detection of chemiluminiscence was conducted following the protocol of the Chemiluminescent EMSA Kit, and using a Chemi-Doc MP instrument (Bio-rad).

Protein-DNA complex crystallization and structure determination

Octahedron-shaped crystals of four SP_0782:ssDNA complexes were obtained by sitting drop vapor diffusion. Briefly, lyophilized SP_0782 was dissolved in a crystallization buffer (5 mM Tris–HCl, pH 7.4, 100 mM NaCl, 5 mM DTT and 0.02% (w/v) NaN_3) to reach a final protein concentration of 8 mg/ml with ssDNA (dTCTTCC, dT6 or dT12) supplied at a 2:1 molar ratio to protein, and incubated for about 1 h on ice before crystallization. 0.4 μl of each protein–DNA mixture was mixed with an equal volume of a precipitant solution. Crystals appeared after 1 week in precipitant solutions of 0.2 M potassium nitrate and 20% (w/v) polyethylene glycol (PEG) 3350 for SP_0782:dTCTTCC, 0.2 M sodium nitrate, pH 6.8 and 20% (w/v) PEG3350 for SP_0782:dT6, and 0.07 M citric acid, 0.03 M Bis–Tris propane, pH 3.4, and 16% (w/v) PEG3350 for SP_0782:dT6 and SP_0782:dT12. Crystals were transferred to cryo-solutions (precipitant solutions supplemented with 20% (v/v) glycerol) through incremental buffer exchange prior to flash cooling in liquid nitrogen.

Single crystal X-ray diffraction data were collected at 100 K on Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U1 for SP_0782:dTCTTCC, and BL19U1 for SP_0782:dT6 and SP_0782:dT12. At least 180° of data were typically collected in 0.2–0.5° oscillation steps. The data were indexed, merged and scaled with HKL2000 (26). The initial structure solution was obtained using the molecular replacement program PHASER (27) with an apo SP_0782 structure (PDB ID: 3OBH) as the search model. Further model building and structure refinement were done using Coot and Phenix (28,29), respectively. 5000 K composite simulated-annealing (SA) omit $2F_o - F_c$ electron density maps were generated using the Cartesian-Anneal method (30) by Phenix. All structures were superposed using the maximum likelihood based superpositioning program THESEUS (31) unless otherwise indicated. Data collection and refinement statistics are summarized in (Supplementary Table S2) and the atomic coordinates and structure factors have been deposited in the PDB (PDB ID: 5ZKL (SP0782:dT12), 5ZKM (SP0782:dTCTTCC), 6JIQ (SP_0782:dT6, form 1) and 6JIP (SP_0782:dT6, form 2)).

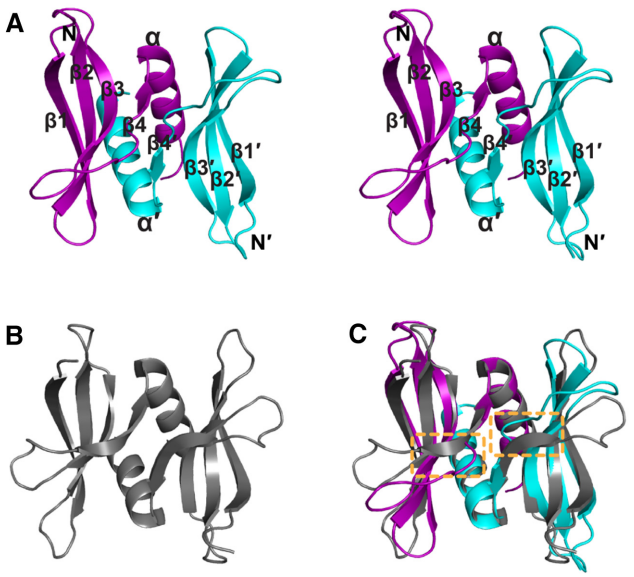

Figure 1. Solution NMR structure of apo SP_0782 from *S. pneumoniae*. (**A**) The stereo-view of the cartoon representation of the SP_0782 homodimer with the lowest energy. Two subunits are colored in violet and cyan, respectively. All secondary structural elements are indicated. (**B**) The crystal structure of SP_0782 (PDB ID: 3OBH). (**C**) Overlay of the solution and crystal structure of SP_0782. The difference in β4 length between the solution and crystal structure is highlighted with a yellow box.

RESULTS

Solution structure of apo SP_0782

Our previous work revealed that apo SP_0782 exists predominantly as a homodimer in solution under the used conditions, suitable for 3D structure determination using NMR spectroscopy (20). We report here the solution NMR structure of apo SP_0782. Each subunit of dimeric SP_0782 contains a curved anti-parallel β-sheet composed of four β-strands in the N-terminus and one 12-residue α-helix in the C-terminus (Figure 1A and Supplementary Figure S1A). All secondary structural elements are well-defined as follows: β1 (F14-N26), β2 (W30-F39), β3 (A44-W50), β4 (I62-L64), and α1 (N66-K77). The two domain-swapped helices arrange with an angle of ∼60° to each other, and pack mainly against the β-sheet of their respective dimeric partner. To date, the crystal structure of SP_0782 (PDB ID: 3OBH) was also solved by scientists from the Northeast Structural Genomics (NESG) consortium (Figure 1B) (32). By comparison, the two structures share highly similar overall folds, except that β4 (M58-L64) in the crystal structure is a little longer than its counterpart in the solution NMR structure, which can be reasonably explained by insufficient observed NOE peaks in this region due to relatively flexible conformational dynamics in solution (Figure 1C).

Sequence alignment shows that SP_0782 shares high similarity to YdbC and EF_3132, two prokaryotic homologs with determined structures, with a sequence identity of 56.2% and 53.4%, respectively (Supplementary Figure S1B). As expected, three proteins exhibit high structural similarity (Supplementary Figure S1C). ConSurf (33) analysis shows that highly conserved residues are mostly located

on the antiparallel β-sheet and the loops connecting the β-strands, except for E68 in the α-helix (Supplementary Figure S2A). Electrostatic surface potential analysis (34) shows that the β-sheet surface consists of many positively charged amino acids, whereas the backside is made up of negatively charged amino acids (Supplementary Figure S2B). Seven positively charged residues located at the β-sheet, including K28, K32, R36, K45, R49, K57 and K60, extend their side chains to participate in forming one face of the SP_0782 dimer with extensively positive charge, similar to those of PC4-CTD and YdbC. In brief, SP_0782 exhibits a PC4-like fold potential for ssDNA binding.

The varied binding patterns of SP_0782 for different lengths of ssDNA

The binding of both YdbC and PC4-CTD with the ssDNA dT19G1 has been well studied (9,16), while the binding of SP_0782 with dT19G1 has previously been preliminarily investigated through NMR titration, which showed that most peaks of SP_0782 disappeared during dT19G1 titration (Supplementary Figure S3) (20). Further NMR experiments at different temperatures revealed that increasing temperature could significantly enhance the peak intensities of SP_0782 titrated with dT19G1 at protein (monomer):DNA ratios of 2:1.2 and 2:0.2, but the observed peaks at the protein:DNA ratio of 2:0.2 were much less than those at the protein:DNA ratio of 2:1.2 (Supplementary Figure S4). As increasing temperature causes more rapid tumbling and may turn the conformational exchange at intermediate rate on the NMR timescale into fast rate (35,36), the peak disappearance of SP_0782 at the protein:DNA ratio of 2:1.2 should mainly result from a conformational exchange between apo and dT19G1-bound states at intermediate rate. On the other hand, at the protein:DNA ratio of 2:0.2, other factors, such as the formation of a high molecular weight (MW) complex, may also contribute to the peak disappearance. In either case, it prevented further NMR study regarding dT19G1 binding. Thus, ssDNA samples other than dT19G1 were employed in NMR titration experiments, wherein dT6 and dT12 that are shorter than dT19G1 did not cause obvious peak disappearance but significant chemical shift perturbations (CSPs) of SP_0782 (Supplementary Figure S3). The binding of SP_0782 was basically saturated at protein(monomer):DNA ratios of 2:2 for dT6 and 2:1 for dT12, respectively. These results confirmed the ssDNA-binding function of SP_0782, and implied varied binding properties of SP_0782 when binding to different lengths of ssDNA.

Since SP_0782 showed different binding properties for dT6, dT12 and dT19G1 in NMR titrations, EMSA experiments were further employed for comparing the binding of SP_0782 with the three ssDNAs. The results showed significantly different binding patterns of SP_0782 for the three ssDNAs (Figure 2). Two kinds of SP_0782:dT12 complexes with different MWs were detected, with the small one formed at low protein:DNA concentration ratios and the large one formed at high protein:DNA ratios. SP_0782 could bind to dT6, forming a complex with a MW similar to that of the small SP_0782:dT12 complex, while the SP_0782:dT6 complex with a MW similar to that of the large SP_0782:dT12 complex was not steadily detected. In contrast, SP_0782 bound to dT19G1, forming a series of complexes with MWs around that of the large SP_0782:dT12 complex, while no SP_0782:dT19G1 complex with a MW similar to that of the small SP_0782:dT12 complex was detected. Moreover, as SP_0782 concentration increased, the MW of the SP_0782:dT19G1 complex increased stepwise, suggesting the formation of various complexes containing different numbers of SP_0782 dimers (Figure 2). Additionally, SP_0782 showed obviously different affinities for dT6 and dT12, as revealed by the concentrations of SP_0782 when the SP_0782:ssDNA complexes were initially detected (Figure 2).

In order to further determine the affinities of SP_0782 for dT6, dT12 and dT19G1 and the stoichiometries in the complexes of SP_0782 with the three ssDNAs, ITC experiments were carried out, and the results are shown in Figure 3 and Table 1. The dissociation constants (K_D) of SP_0782 for dT6, dT12 and dT19G1 are 2.68×10^{-5}, 2.06×10^{-8} and 1.69×10^{-7} M, respectively. The binding affinity for dT6 is much weaker than that for dT12, while the affinity for dT19G1 is also 8.2-fold weaker than that for dT12. The binding stoichiometries of one subunit of the SP_0782 dimer for dT6 and dT12 are 0.81 and 0.49, suggesting that one SP_0782 dimer binds to two dT6 molecules but only binds to one dT12 molecule, consistent with the NMR titration data. Considering the largely higher affinity of SP_0782 for dT12 than dT6, this implies that two subunits in the SP_0782 dimer cooperate to achieve high-affinity binding of ssDNA, similar to PC4-CTD. The stoichiometry of one subunit of the SP_0782 dimer for dT19G1 is close to 3:1 (0.33), suggesting that two dimers bind with one molecule of dT19G1, but among the four subunits of the two dimers, one subunit may not bind with DNA. This may be the reason for the lower affinity of SP_0782 for dT19G1 than dT12. A multi-angle static light scattering assay coupled with size exclusion chromatography (SEC-MALS) showed that the SP_0782:dT19G1 complex containing two SP_0782 dimers and one dT19G1 could steadily exist at the protein:DNA ratios of 2:0.2 and 2:0.1, under conditions similar to those for NMR titrations (Supplementary Figure S5). Thus, based on the EMSA, ITC and SEC-MALS data, one dT19G1 can be bound by two or more SP_0782 dimers at a high protein:DNA ratio, which may lead to a high-density binding. Accordingly, the small SP_0782:dT12 and SP_0782:dT6 complexes detected in EMSA experiments should contain one SP_0782 dimer, while the large SP_0782:dT12 complexes may contain two SP_0782 dimers. To sum up, one SP_0782 dimer could sufficiently bind to ssDNA with a length of ∼12 nt in a high-affinity manner, while at a high protein:DNA ratio, more than one SP_0782 dimer could bind to dT12 and dT19G1, possibly in a high-density manner but with relatively lower affinity.

The ssDNA-binding interface of SP_0782 derived from NMR titrations

The NMR titration data were further analyzed to characterize the interface of SP_0782 for binding with ssDNA, which could be in general derived from changes in peak intensity and position. Figure 4A shows an overlay of ^{1}H-

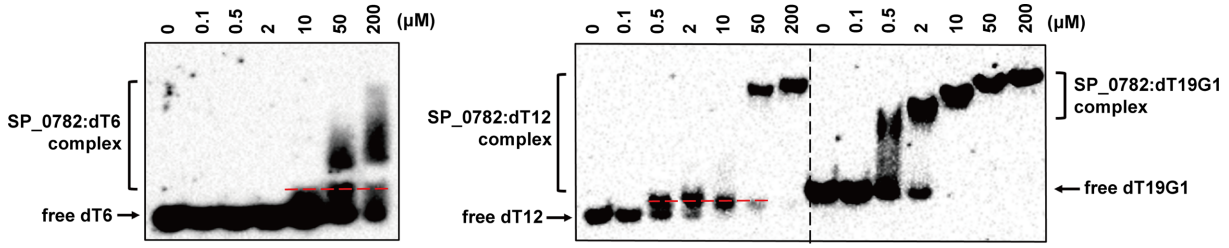

Figure 2. Different binding properties of SP_0782 for dT6, dT12 and dT19G1. EMSA assay using 0.01 μM biotin-labeled ssDNA as a probe to investigate the ssDNA binding of SP_0782. The concentrations of SP_0782 are indicated on top, and the position of ssDNA bound with one SP_0782 dimer is suggested with red dashed lines.

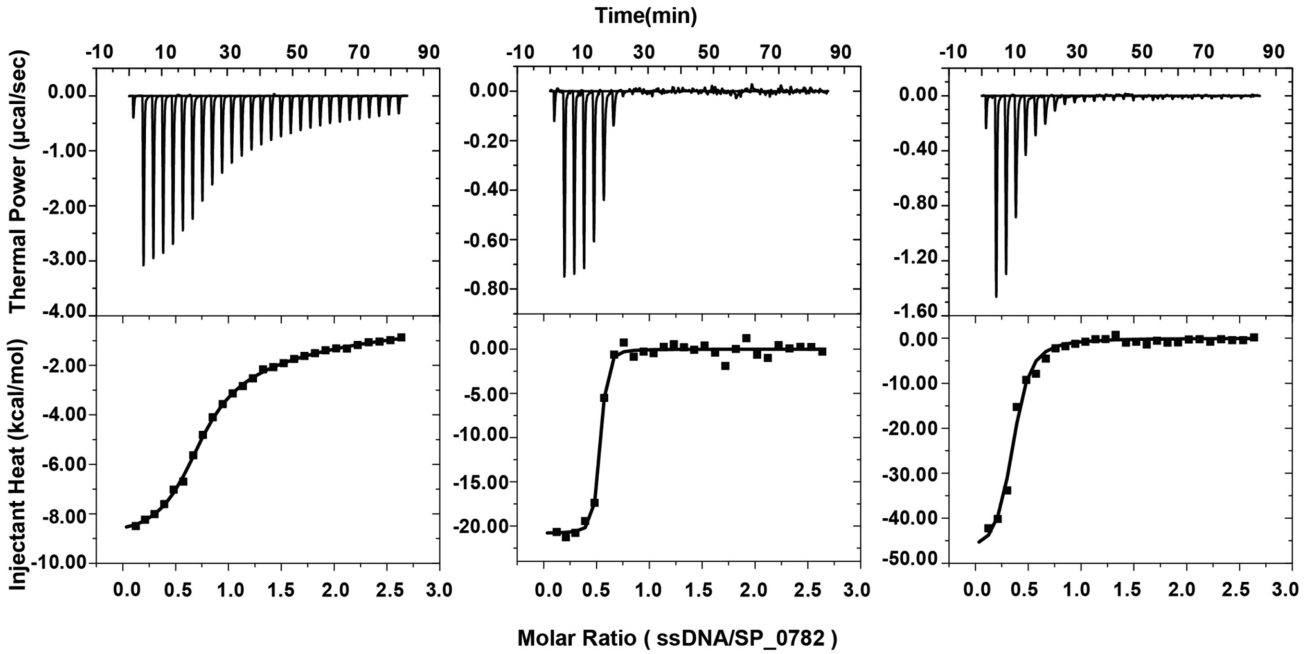

Figure 3. Interactions between SP_0782 and different lengths of ssDNA (dT6, dT12 and dT19G1) studied by ITC experiments. (Top) Thermal power versus time as indicated. (Bottom) Injection heat versus ssDNA:SP_0782 (monomer) molar ratio. The resulting parameters, such as stoichiometric ratio (N) and dissociation constant (K_D), are indicated in Table 1.

Table 1. ITC-derived parameters for the binding of SP_0782 to different lengths of single-stranded DNA (ssDNA)

ssDNA	N	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/mol·K)	K_D (M)
dT6	0.81 ± 0.05	−6.14 ± 1.03	−13.26 ± 1.03	− 243	(2.68 ± 0.34) × 10^{-5}
dT12	0.49 ± 0.02	−10.71 ± 0.37	−20.85 ± 0.37	− 346	(2.06 ± 0.26) × 10^{-8}
dT19G1	0.33 ± 0.01	−9.27 ± 1.97	−48.53 ± 1.97	− 1340	(1.69 ± 0.03) × 10^{-7}

The ITC profiles shown in Figure 3 were fitted with a one-site binding model and independent sets of binding sites. All parameters were allowed to float during the fitting routines. Values for ΔG and ΔS were calculated using the standard formalisms containing the maximum errors as carried through the equations.

^{15}N HSQC spectra of SP_0782 titrated with dT12 at a series of monomeric protein:ssDNA ratios. CSPs are clearly observed, and in terms of conformational exchange dynamics on the NMR timescale, three distinct kinds of CSPs were found, indicating residues undergoing fast exchange such as G41, slow exchange such as S65, and intermediate exchange such as T63 (Figure 4A). The CSP of each cross peak was further analysed, as illustrated in Figure 4B. The residues with CSPs greater than mean + 1σ, plus another six residues (R49, A50, M58, G59, K60 and T63, labeled with asterisks) showing disappeared signals, were selected to define the ssDNA binding interface. When mapped onto the solution structure of apo SP_0782, the dT12 binding interface was found to be mainly located on the β-sheet (Figure 4C). In addition, eight residues (T13, K32, V37, E39, N40, I48, I62 and L64) involved in dT12 binding undergo slow conformational exchange, consistent with the nano-molar affinity for dT12 determined in the ITC experiment.

In the case of dT6 titration, most residues with CSPs undergo fast conformational exchange upon dT6 binding (Supplementary Figure S6), consistent with the micromolar affinity of SP_0782 for dT6. The interface of SP_0782

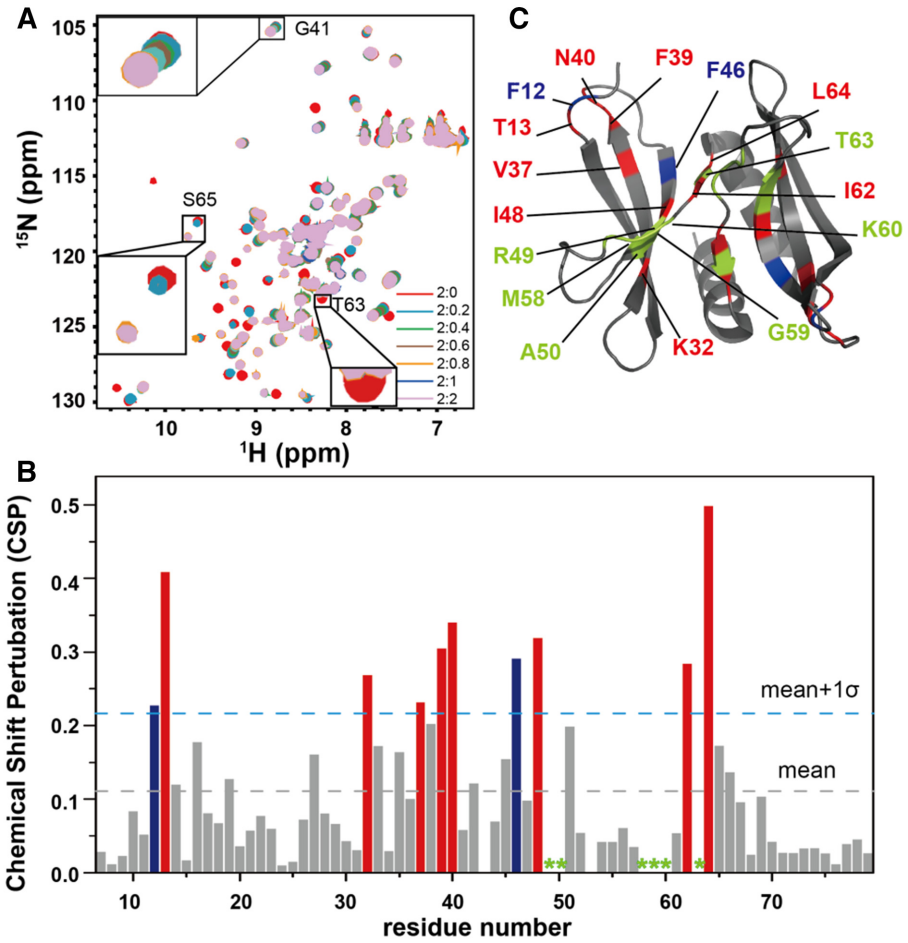

Figure 4. NMR titration of SP_0782 with ssDNA dT12. (**A**) Overlay of a series of ^{1}H–^{15}N HSQC spectra of SP_0782 in the presence of dT12 at different monomeric protein:ssDNA molar ratios, which are colored differently as indicated. Chemical shift perturbations (CSPs) of G41, S65 and T63, which represent three different rates of conformational exchange processes (fast, slow and intermediate) on the NMR timescale between the free and dT12-bound state, are zoomed as insets. (**B**) CSP for each residue is calculated and illustrated. Here, 'mean' stands for mean value and 'σ' for standard deviation. All residues with CSPs greater than 'mean + 1σ', together with six disappeared residues labelled with asterisks, are further mapped onto the NMR structure of SP_0782 as shown in (**C**). Two residues undergoing fast exchange are colored in blue, eight residues undergoing slow exchange are colored in red, and six residues undergoing intermediate exchange are colored in green in both (B) and (C).

for dT6 was derived using the NMR titration data following the same process as that for dT12, and is nearly identical to the interface for dT12 (Supplementary Figure S6C). These results indicate that SP_0782 binds with dT6 and dT12 with very different affinities but similar binding interfaces, and support the proposal that the two subunits of the SP_0782 dimer cooperate to bind long ssDNA and achieve high-affinity binding. The binding interface for dT19G1 could not be derived due to the peak disappearance of SP_0782 in NMR titration.

Crystal structures of SP_0782:ssDNA complexes

In order to further dissect the ssDNA-binding mechanism of SP_0782, we determined four crystal structures of SP_0782 in complex with ssDNA (Figure 5 and Supplementary Figure S7, Table S2). Among these structures, two of them were derived from a SP_0782:dT6 complex with ssDNA resolved to different extent (form 1: two nucleotides; form 2: three nucleotides), and the third one was derived from a SP_0782:dT12 complex. The crystal structure of SP_0782 with dT19G1 that was reported previously (PDB ID: 4G06) (37) was used for comparison with the crystal structures of apo SP_0782, SP_0782:dT6 (both forms), and SP_0782:dT12 (Figure 5A–E). Compared with apo SP_0782 (PDB ID: 3OBH, chain A), the RMSD values of ssDNA-bound SP_0782 are 0.30, 0.32, 0.38 and 0.94 Å for SP_0782:dT6 (form 1), SP_0782:dT6 (form 2), SP_0782:dT12, and SP_0782:dT19G1, respectively. Two major DNA-binding regions (DBRs), DBR1 and DBR2, could be identified in the three complex structures, as indicated in Figure 5F and G. In the SP_0782:dT6 or SP_0782:dT12 structures, 2–4 nucleotides were bound with each subunit of the dimer, occupying both DBRs (Figure 5B–D). In contrast, DNA was present only in one subunit of the dimer in the SP_0782:dT19G1 structure, and only the DBR2 was occupied by a two-nucleotide stretch (Figure 5E).

In the SP_0782:dT12 crystal structure, four nucleotides were bound by each SP_0782 subunit (Figures 5D and 6A).

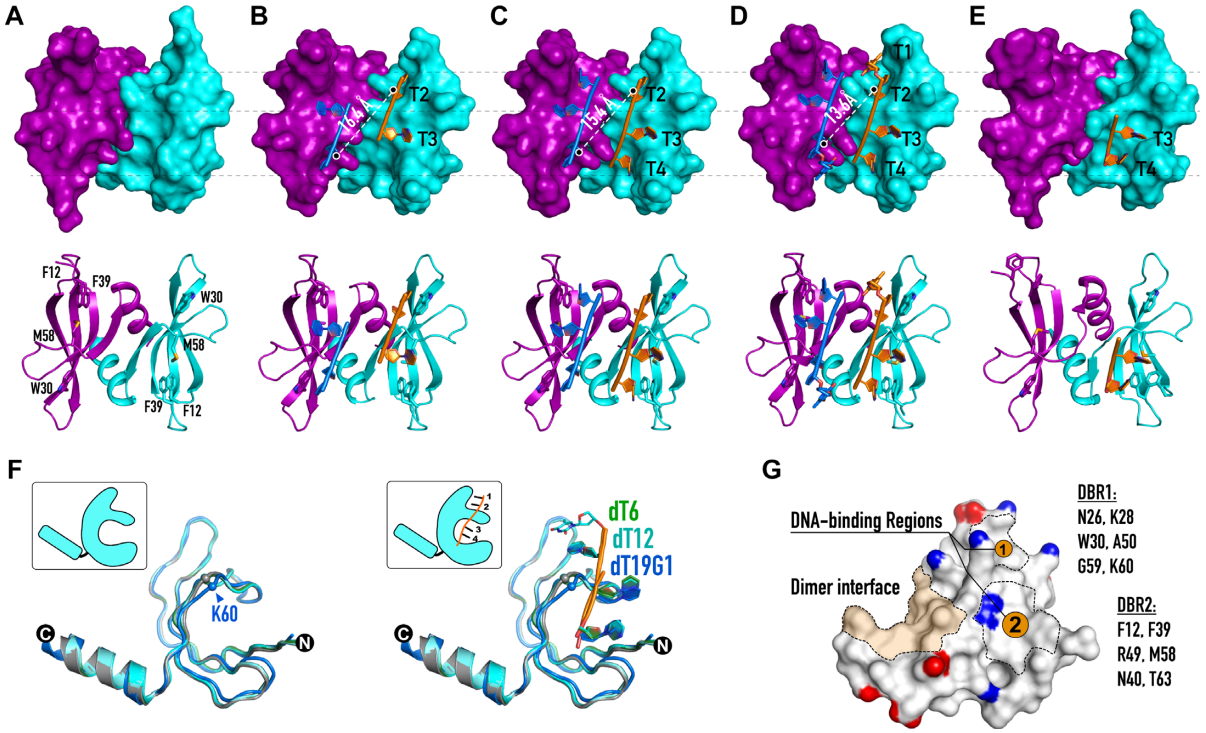

Figure 5. Crystal structures of the apo SP_0782 and its complexes with ssDNAs in different lengths. (**A**) The crystal structure of the apo SP_0782, (B–E) crystal structures of SP_0782 with dT6 form 1 (**B**), dT6 form 2 (**C**), dT12 (**D**) and dT19G1 (**E**), respectively. The chain A (cyan) and chain B (purple) of the homodimer are shown as surface (top) or cartoon (bottom) representations, while the ssDNAs are shown as cartoons and colored in orange (complexed with chain A) and marine (complexed with chain B). (**F**) Structural superpositions of the apo SP_0782 and the ssDNA-bound chain of SP_0782 complexed with dT6 (form 2), dT12, and dT19G1. The α-carbons of K60 are shown as spheres. (**G**) A surface view of the SP_0782 subunit colored in white, with positively and negatively charged residues colored in blue and red. The DNA-binding regions and the dimer interface are highlighted with dashed lines. The residues involved in forming the surfaces of the DNA-binding regions are indicated.

Eleven residues, E11, F12, N26, K28, W30, F39, N40, M58, K60, G61 and T63, are involved in direct or indirect ssDNA binding (Figure 6A), among which W30, F39, N40, M58, K60, G61 and T63 are highly conserved. In DBR1, the bases of T1 and T2, and the indole group of W30 form 'three-layer' hydrophobic stacking interactions. The T1 and T2 bases are sandwiched by the side chains of residues K28 and K60 on their edges, while the T2 base is further stabilized by residues N26/K60 and a water molecule through hydrogen bonding interactions. In DBR2, the ribose of T3 interacts with the side chain of M58, and the ribose and base of T4 pack against the side chain of F39 and F12, constituting the primary hydrophobic interaction network. The base and phosphate of T4 are further stabilized by direct or water-mediated hydrogen bonding interactions with residues E11, N40 and G61, and several polar interactions are also worth noting (Figure 6A). In addition, a fifth nucleotide with only the phosphate group modeled participates in a hydrogen bonding network (Figure 6A). In general, hydrophobic interactions form the core of the protein–DNA interface, with polar interactions stabilizing the peripherals for both DBRs.

In comparison with the NMR titration data, the backbone atoms of F12, F39, N40, M58, K60 and T63 showed marked CSPs (Figure 4), while the side-chain atoms of N26 and W30 also displayed marked chemical shift changes (Supplementary Figure S8), indicating a high consistency between the crystal structure and the NMR data. The backbone atom of K28 did not show any significant CSP, which may be due to the fact that the interaction of K28 with DNA is conducted by its side-chain atoms, similar to N26 and W30. However, the resonances of K28 side-chain atoms cannot be observed in the ^{1}H–^{15}N HSQC spectrum. The contacts of E11 and G61 with DNA are mediated by water molecules revealed by the crystal structure, which may be weak in the solution environment and were not detected in the NMR titration.

Except for fewer nucleotides resolved, the ssDNA–protein interaction details in the two SP_0782:dT6 structures are largely consistent with those in the SP_0782:dT12 complex, with subtle differences such as the loss of three hydrogen bonds from residues E11, K60 and T63 in the form 2 of the SP_0782:dT6 structure (Figure 6B). Interestingly, in the SP_0782:dT19G1 structure, only two nucleotides were resolved in one subunit of the SP_0782 dimer, and the phosphate of T4 establishes a direct hydrogen bond with G61 and a salt bridge with the guanidinium group of R49, resulting in a closer contact between the sugar-phosphate backbone of the T3 and T4 nucleotides and the DBR2. Accordingly, residues 59–61 undergo a β-strand to loop rearrangement, resulting in a conformation similar to that in the apo NMR structure (Figures 5E and 6C). A similar local structure rearrangement was depicted previously between the apo and dT19G1-bound YdbC (16), while the

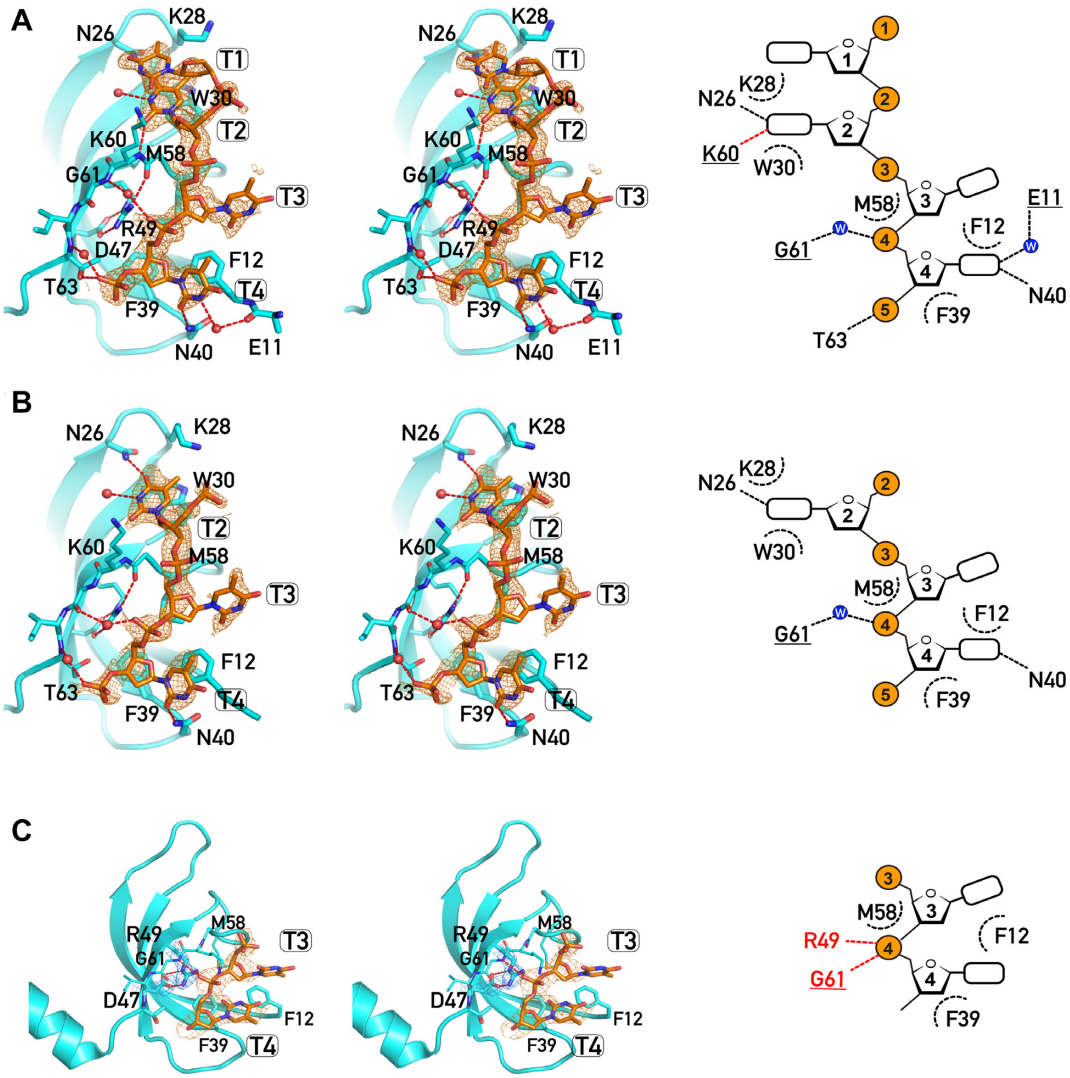

Figure 6. The protein–DNA binding details for SP_0782 structures in complex with different ssDNAs. (A–C) Left: stereo-pair images of 5,000 K composite SA-omit $2F_o$-F_c electron density map (contoured at 1.2 σ) overlaid onto structural models of SP_0782:dT12 (**A**), SP_0782:dT6 form 2 (**B**), or SP_0782:dT19G1 (**C**). The protein (cyan), ssDNA (orange), and water molecules (red) are shown as cartoon, sticks and spheres, respectively. The key residues of SP_0782 involved in ssDNA binding are shown as sticks. Right: cartoon illustrations of key protein-DNA interactions. The straight dashed lines indicate hydrogen bonds or salt bridges, while the dashed arc lines indicate hydrophobic interactions. The residues with the backbone involved in interaction are underlined, and red residues and lines emphasize the interaction details in SP_0782:dT6 and SP_0782:dT19G1 notably different from those in SP_0782:dT12.

PC4 and MoSub1 proteins adopt a loop structure naturally in this corresponding region (9,38) (discussed below in Figure 7). Taken together, these structures of SP_0782:ssDNA complexes reveal that the varied binding patterns and affinities of SP_0782 for different lengths of ssDNA are associated with distinct capture of nucleotides in the two DBRs of SP_0782.

The potential dimer-dimer interaction of SP_0782 in ssDNA binding

As foregoing shown, two or more SP_0782 dimers could bind with dT19G1 and form large complexes in a potential high-density manner. A model for this binding was deduced from the crystal packing of the SP_0782:dT19G1 complex (Supplementary Figure S9A). In this model, direct interaction between neighbouring SP_0782 dimers was suggested, as the β2/β3-turn (residues 40–43) from the ssDNA-bound subunit of a SP_0782 dimer stacks with the W30 residue in the DBR1 from the ssDNA-bound subunit of a downstream SP_0782 dimer, while the α-helix of the same subunit of the upstream SP_0782 dimer stretches into the DBR2 of the ssDNA-unbound subunit of the downstream dimer (Supplementary Figures S9A and S10A). In order to validate the dimer–dimer interaction in the deduced model, several mutants expected to weaken the interaction between SP_0782 dimers were obtained, among which the W30A mutant substituting W30 with an alanine (W30A) showed a significantly smaller size of the dT19G1-bound complex than that of the wild-type SP_0782:dT19G1 complex

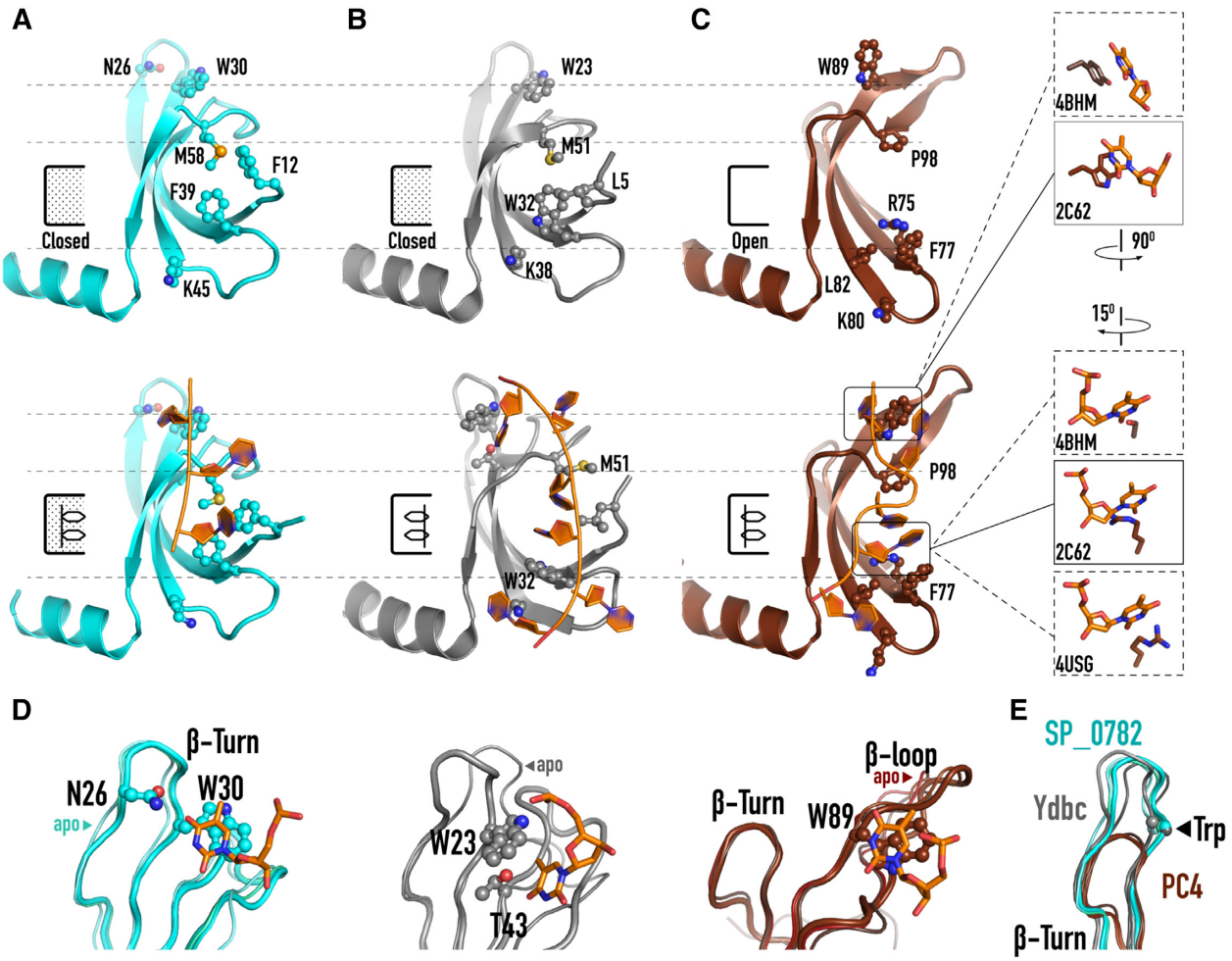

Figure 7. A structural comparison of SP_0782, YdbC, and PC4. (**A**) Crystal structures of the apo SP_0782 (top, PDB ID: 3OBH) and SP_0782:dT12 (bottom, this work). (**B**) NMR structures of apo YdbC (top, PDB ID: 2LTD) and YdbC:dT19G1 (bottom, PDB ID: 2LTT). (**C**) Crystal structures of apo PC4 (top, PDB ID: 1PCF) and PC4:dT19G1 (bottom, PDB ID: 2C62). (**D**) The DBR1-T2 nucleotide binding details of SP_0782:ssDNA complex structures (Left), YdbC:dT19G1 (Center), and PC4:dT19G1 (Right). (**E**) β1/β2-turn comparison among PC4-like proteins. SP_0782 (cyan), YdbC (gray), PC4 (brown), and ssDNAs (orange) are shown as cartoons and the key residues involved in ssDNA binding are shown as ball-and-sticks. The open rectangles with or without black spots indicate the ground closed or open state of DBR2, and the two hexagons indicate the T3 and T4 nucleotides. The right panels in (C) with PDB IDs on the bottom left emphasize the interaction between the paired residue and the nucleotide base.

in EMSA (Supplementary Figure S10B). The W30A mutant existed as a homodimer, similar to wild-type SP_0782 (Supplementary Figure S11A). Thus, the smaller size of the W30A:dT19G1 complex should be due to a less number of W30A dimer binding to dT19G1. Subsequently, the stoichiometry of one subunit of the W30A dimer for dT19G1 was determined to be 0.45 by ITC (Supplementary Figure S11B), suggesting that one dimer binds with one dT19G1 strand. Moreover, only the complex containing one W30A dimer and one dT19G1 was found in the SEC-MALS assay under similar conditions for wild-type SP_0782 (Supplementary Figure S11A). These results indicated that the W30A mutation reduced the formation of large SP_0782:dT19G1 complexes containing two or more dimers, and supported the dimer–dimer interaction of SP_0782 dimers during dT19G1 binding. In addition, the W30A mutant displayed significantly decreased affinity for dT19G1 (4.12×10^{-5} M) in ITC and EMSA experiments (Supplementary Figures S10B and S11B), confirming the important role of W30 in nucleotide binding, as revealed by the crystal structures of the SP_0782:ssDNA complexes,.

DISCUSSION

PC4-like proteins have OB fold-derived structures and form homodimers to bind with ssDNA, and the representative protein, human PC4, was originally identified as a transcriptional coactivator playing a role during transcription (9,11). Subsequent studies have revealed broad roles of PC4 and Sub1, its close homolog from yeast, in chromatin condensation, maintenance of genome stability and DNA repair (3). Moreover, several PC4-like proteins in bacteria have been identified, including YdbC in *L. lactis* and BPSL1147 in *B. pseudomallei* (16,18). However, the structure and function of the bacteria PC4-like proteins are still largely unknown. In this study, we identify and characterize

SP_0782, a novel PC4-like protein in *S. pneumoniae* that has an ssDNA-binding property and mode different from those of PC4 and YdbC, which will facilitate the understanding of the DNA-binding mechanism of PC4-like proteins.

The variation in ssDNA length bound by PC4-like proteins

Human PC4 shows no significant binding for ssDNA shorter than dT10, while 16 or more nucleotides are needed to reach the affinity of nano-molar level (10). If the length of ssDNA extends to 28 nt, two dimers of PC4 would be steadily bound (12,13). These binding properties of human PC4 are consistent with its crystal structure bound with dT19G1, which shows that each subunit of the PC4 dimer binds with five nucleotides (9). YdbC shows an affinity of nano-molar level for dT19G1, whereas two dimers can bind to one dT19G1 with the second dimer showing very weak and micro-molar affinity, which could be explained by its complex structure, showing that each subunit of the YdbC dimer binds with seven nucleotides (16). However, SP_0782 could bind with dT19G1 steadily in a two-dimer bound form and with dT6 in a micro-molar affinity, and could even bind with dT12 in a two-dimer bound form at a high protein:DNA ratio, suggesting that the nucleotides bound by each subunit of the SP_0782 dimer are less than those bound by PC4 and YdbC. The crystal structure of the SP_0782:dT12 complex fully supports this suggestion, wherein four nucleotides bind to the surface of each subunit of the SP_0782 dimer. Thus, it appears that PC4-like proteins are divergent in the ssDNA length of binding.

The variation in DNA-binding details in PC4-like proteins

The detailed structural differences and diverse DNA binding modes between SP_0782 and two representative PC4-like proteins are summarized in Figure 7. Although the overall folds of SP_0782, YdbC and PC4 are similar, the protein:ssDNA binding details are different at the DBR1 and DBR2 regions. In the SP_0782:dT6 and SP_0782:dT12 structures, the T2 base stacks on top of the W30 side chain (Figure 7D, left panel), and the corresponding base inserts between the W23 side chain and the G52 backbone in the YdbC:dT19G1 structure (Figure 7D, central panel), or stacks on top of the W89 side chain in the PC4:dT19G1 structure (Figure 7D, right panel). The W30 in SP_0782 and the W23 in YdbC are located at the beginning of the β2 strand, while the W89 in PC4 and the Y74 in MoSub1 (a W89 equivalent in PC4) are located at the end of the β3 strand. Moreover, the β-turns between β1 and β2 of PC4 and MoSub1 are shorter than those of SP_0782 and YdbC (Figure 7E). Another significant difference among the PC4-like family is the DBR2 region, denoted as 'closed' or 'open' form herein. Different from the naturally open DBR2 of PC4, which could readily intake two adjacent nucleotides between P98 and F77, the closed DBR2 of YdbC switched to open state by the inducement of a single-strand DNA (Figure 7B and C). The DBR2 of SP_0782 is also closed; however, it could bind two nucleotides of various lengths of ssDNAs without large-scale structural rearrangement (Figures 6 and 7A). Thus, it can be concluded that although these PC4-like proteins adopt a conservative overall fold to build the DBRs and utilize similar hydrophobic functional groups for DNA binding, the structural details in each PC4-like protein and the structural rearrangement upon ssDNA binding do vary to some extent. Importantly, two major differences between prokaryotic and eukaryotic PC4-like proteins can be summarized as the location of the key Trp residue in DBR1 and the open/closed state of DBR2 without DNA binding.

The role of DBR1 for ssDNA binding

Among the four resolved SP_0782:ssDNA complex structures, one nucleotide was resolved stacking with residue W30 in the SP_0782:dTCTTCC complex structure, two or three nucleotides were resolved occupying DBR1 and partial DBR2 regions in SP_0782:dT6 complex structures (form 1 and form 2), and four intact nucleotides were resolved occupying both DBR1 and DBR2 in SP_0782:dT12 complex structures (Figure 5B–D and Supplementary Figure S7). In accordance with the observation of lower binding affinity (8.13×10^{-5} M) for SP_0782:dTCTTCC (Supplementary Figure S12), only one nucleotide was resolved stacking with W30 in the structure of the SP_0782:dTCTTCC complex, while two to four nucleotides were modeled stretching from DBR1 to DBR2 in the SP_0782:dT6 and SP_0782:dT12 structures, indicating that DBR1 may serve as an initial ssDNA binding site and DBR2 enhances the ssDNA-SP_0782 binding affinity, subsequently. Even though the overall T2-DBR1 interaction network is shared by the four resolved types of complex structures, the direct distances between T2 and K60 reduced along with the extension of modeled nucleotides, resulting in a transition from water-mediated to direct T2 and K60 interactions (Supplementary Figure S7). The observation of T2 swaying suggests that T2-DBR1 interactions may play key roles in both initializing and accommodating the binding of ssDNA at various lengths.

A potential high-density binding pattern of SP_0782

It is known that PC4 can bind with ssDNA in a two-dimer-bound form (12,13). Furthermore, the crystal packing of the PC4:dT19G1 complex reveals a probable mode of DNA-mediated multimerization of the PC4 dimer, which may be relevant for the unwinding of duplex DNA by PC4 (9). Similarly, in the crystal structure of the MoSub1:dT19G1 complex, two ssDNA strands mediate the multimerization of three MoSub1 dimers (Supplementary Figure S9B) (38). In this study, the formation of large complexes of SP_0782 with ssDNA was found in EMSA and SEC-MALS experiments, and a model of the DNA-mediated multimerization of the SP_0782 dimer was deduced from the crystal packing of the SP_0782:dT19G1 complex (Supplementary Figure S9A). Thus, the DNA-mediated multimerization seems to be a general mechanism for PC4-like proteins, and can be reasonably linked to their functions when the binding of long ssDNA is required.

In the deduced multimerization model from the SP_0782:dT19G1 structure, the distance between the 5′-phosphate of T3 and the 3′-OH of T4 from an adjacent SP_0782:dT19G1 dimer is ∼7 Å, implying a disordered

nucleotide (T2) between the two SP_0782 dimers (Supplementary Figure S9A). Only one subunit of each SP_0782 dimer is occupied by DNA, and each SP_0782 dimer binds with three nucleotides in the DNA strand, displaying a high-density binding. During SP_0782 and dT19G1 binding, the interaction between SP_0782 dimers facilitates the multimerization and high-density binding, suggesting a cooperative mechanism of SP_0782 dimers. Nevertheless, the multimerization models for PC4, MoSub1 and SP_0782 revealed by the crystal packing of the protein:dT19G1 complexes are different to some extent. Whether the conserved Trp residue in DBR1 is also essential for the multimerization of PC4 and MoSub1 requires further investigation.

The potential subunit-occupied modes of SP_0782

The high-density binding model of SP_0782 implies a potential single-subunit-occupied (SSO) mode, wherein only one subunit of the SP_0782 dimer is occupied by DNA, in contrast to the double-subunit-occupied (DSO) mode revealed by the complex structures of SP_0782:dT6 and SP_0782:dT12. The length-dependent DNA-binding patterns of SP_0782 can be summarized and classified into the SSO and DSO modes, respectively (Supplementary Figure S13). One SP_0782 dimer can bind with one (SSO) or two dT6 oligos (DSO) in a micro-molar affinity, while one SP_0782 dimer bind with one dT12 (DSO) in a nano-molar affinity likely achieved by a cooperation between the two subunits in the same dimer. Moreover, two SP_0782 dimers may bind with one dT12 (SSO) at high SP_0782:ssDNA molar ratios in a potential high-density manner, similar to the binding of dT19G1 by SP_0782 (Supplementary Figure S9A). The situation for dT19G1 is more complicated. Besides the binding pattern containing two SP_0782 dimers and one dT19G1 (a hybrid of SSO and DSO) detected in ITC and SEC-MALS experiments, other patterns, including one SP_0782 dimer versus one dT19G1 (DSO), three SP_0782 dimers versus two dT19G1 oligos (DSO), and three SP_0782 dimers versus one dT19G1 (SSO), could also be adopted.

The SSO and DSO modes may undergo transitions when the SP_0782:ssDNA ratio changes (Supplementary Figure S13), as implied by the EMSA experiments. This is reminiscent of the transition of EcSSB between the $(SSB)_{35}$ and $(SSB)_{65}$ modes during ssDNA binding. EcSSB binds 35 nt ssDNA with two subunits of its tetramer ($(SSB)_{35}$ mode) and 65 nt ssDNA with all four subunits of its tetramer ($(SSB)_{65}$ mode), but at a high protein:DNA ratio, EcoSSB could also bind with 65 nt ssDNA with two tetramers bound in $(SSB)_{35}$ mode to achieve high-density binding (reviewed in (39)). Only two subunits of each EcSSB tetramer bind with DNA, and EcSSB tetramers interact with each other to achieve high-density binding in a cooperative manner (reviewed in (39)). Therein, the unoccupied DNA-binding sites in one tetramer interact with the C-terminal conserved nine-amino acid tip (TIP) from a neighboring tetramer (40), analogous to the direct contacts between neighbouring SP_0782 dimers deduced in the high-density binding model (Supplementary Figure S9A). Nevertheless, the hypothesis of SSO mode and the transition between SSO and DSO modes of SP_0782 demands further demonstration in future.

The biological function of SP_0782

Broad roles of PC4 and Sub1 in transcription and maintenance of genome stability have been implicated. The role of PC4 in maintenance of genome stability depends on its CTD (41). In yeast and in the context of ectopic expression in *E. coli*, eukaryotic PC4 could protect chromosomal DNA against oxidative damage via a mechanism that strictly depends on the ssDNA-binding domain (42,43). Thus, it is expected that SP_0782 may be involved in DNA protection and repair under oxidative stress in *S. pneumoniae*, which are important for pathogenic bacteria survival, especially during the infection process, because host immune cells usually produce an oxidative burst aimed to kill the pathogens. Meanwhile, it is worth noting that SP_0782 may be involved in natural transformation in *S. pneumoniae*. Natural transformation is a biological process in which bacteria intake exogenous DNA and incorporate it into their genomes by recombination, which facilitates genome evolution and fitness (44,45). SpSsbB has been shown to be involved in the natural transformation of *S. pneumoniae* through binding exogenous ssDNA, and may facilitate the ssDNA binding by SpDprA and then loading to recombination-related protein RecA (46–49). A pioneering study showed that the expression level of the *SP_0782* gene was up-regulated eight-fold during competence induction, which is a status of *S. pneumoniae* cells for natural transformation (50). However, the deletion of the *SP_0782* gene did not significantly change the competence induction of *S. pneumoniae* cells. Considering that the ssDNA binding function of SP_0782 was not revealed until our study, more well-designed experiments may be required to figure out whether SP_0782 is associated with the natural transformation of *S. pneumoniae*.

CONCLUSION

In summary, SP_0782 represents a novel PC4-like protein that binds with ssDNA in *S. pneumoniae*. The SP_0782 dimer has a conserved PC4-like fold, and binds to two antiparallel ssDNA strands with each subunit. The two subunits in one SP_0782 dimer bind with ssDNA cooperatively to achieve a high affinity. The ssDNA length fully occupied by SP_0782 is shorter than those occupied by PC4 and YdbC, which is associated with the different structures of the ssDNA-binding interface. SP_0782 exhibits varied binding patterns for different lengths of ssDNA, which result from distinct capture of nucleotides in the two major DNA-binding regions in SP_0782. Furthermore, SP_0782 tends to form a large complex with ssDNA at a high SP_0782:ssDNA molar ratio, and a high-density binding model involving cooperation between SP_0782 dimers during ssDNA binding is proposed. According to the binding patterns of SP_0782 dependent on DNA length, two different subunit-occupied modes of SSO and DSO are suggested, along with a hypothesis of the transition of SP_0782 between the SSO and DSO modes due to the change in SP_0782:ssDNA molar ratio. In addition, the biological roles of SP_0782, probably in DNA protection and repair, as

well as natural transformation are discussed. Our study provides insights into the DNA-binding mechanism of SP_0782 and even PC4-like proteins, and benefits further studies regarding the DNA-binding dynamics and the biological function of SP_0782.

DATA AVAILABILITY

The atomic coordinates and NMR restraints for the reported NMR structure have been deposited at the Protein Data Bank under accession number 2L3A. The atomic coordinates and structure factors for the reported crystal structures have been deposited at the Protein Data bank under accession numbers 5ZKL, 5ZKM, 6JIP, and 6JIQ.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the scientists in the Rutgers University Protein Production Facility for construct design and sample preparation of SP_0782 for NMR structure determination. We thank synchrotron SSRF (beamlines BL17U1 and BL19U1, Shanghai, China) for access to beamlines.

FUNDING

National Key R&D Program of China [2016YFA051201, 2018YFA0507200]; National Natural Sciences Foundation of China [21575155, 31670154, 31802147]; National Institute of General Medical Sciences of USA [U54-GM094597]. Funding for open access charge: National Key R&D Program of China [2016YFA051201]; National Natural Science Foundation of China [21575155].
Conflict of interest statement. None declared.

REFERENCES

1. Shereda,R.D., Kozlov,A.G., Lohman,T.M., Cox,M.M. and Keck,J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 289–318.
2. Dickey,T.H., Altschuler,S.E. and Wuttke,D.S. (2013) Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure*, **21**, 1074–1084.
3. Garavis,M. and Calvo,O. (2017) Sub1/PC4, a multifaceted factor: from transcription to genome stability. *Curr. Genet.*, **63**, 1023–1035.
4. Theobald,D.L., Mitton-Fry,R.M. and Wuttke,D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 115–133.
5. Valverde,R., Edwards,L. and Regan,L. (2008) Structure and function of KH domains. *FEBS J.*, **275**, 2712–2726.
6. Clery,A., Blatter,M. and Allain,F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
7. Cappadocia,L., Marechal,A., Parent,J.S., Lepage,E., Sygusch,J. and Brisson,N. (2010) Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. *Plant Cell*, **22**, 1849–1867.
8. Raghunathan,S., Kozlov,A.G., Lohman,T.M. and Waksman,G. (2000) Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat. Struct. Biol.*, **7**, 648–652.
9. Werten,S. and Moras,D. (2006) A global transcription cofactor bound to juxtaposed strands of unwound DNA. *Nat. Struct. Mol. Biol.*, **13**, 181–182.
10. Werten,S., Langen,F.W., van Schaik,R., Timmers,H.T., Meisterernst,M. and van der Vliet,P.C. (1998) High-affinity DNA binding by the C-terminal domain of the transcriptional coactivator PC4 requires simultaneous interaction with two opposing unpaired strands and results in helix destabilization. *J. Mol. Biol.*, **276**, 367–377.
11. Brandsen,J., Werten,S., van der Vliet,P.C., Meisterernst,M., Kroon,J. and Gros,P. (1997) C-terminal domain of transcription cofactor PC4 reveals dimeric ssDNA binding site. *Nat. Struct. Biol.*, **4**, 900–903.
12. Werten,S., Stelzer,G., Goppelt,A., Langen,F.M., Gros,P., Timmers,H.T., Van der Vliet,P.C. and Meisterernst,M. (1998) Interaction of PC4 with melted DNA inhibits transcription. *EMBO J.*, **17**, 5103–5111.
13. Jonker,H.R., Wechselberger,R.W., Boelens,R., Kaptein,R. and Folkers,G.E. (2006) The intrinsically unstructured domain of PC4 modulates the activity of the structured core through inter- and intramolecular interactions. *Biochemistry*, **45**, 5067–5081.
14. Knaus,R., Pollock,R. and Guarente,L. (1996) Yeast SUB1 is a suppressor of TFIIB mutations and has homology to the human co-activator PC4. *EMBO J.*, **15**, 1933–1940.
15. Steigemann,B., Schulz,A. and Werten,S. (2013) Bacteriophage T5 encodes a homolog of the eukaryotic transcription coactivator PC4 implicated in recombination-dependent DNA replication. *J. Mol. Biol.*, **425**, 4125–4133.
16. Rossi,P., Barbieri,C.M., Aramini,J.M., Bini,E., Lee,H.W., Janjua,H., Xiao,R., Acton,T.B. and Montelione,G.T. (2013) Structures of apo- and ssDNA-bound YdbC from Lactococcus lactis uncover the function of protein domain family DUF2128 and expand the single-stranded DNA-binding domain proteome. *Nucleic Acids Res.*, **41**, 2756–2768.
17. Huang,J., Zhao,Y., Huang,D., Liu,H., Justin,N., Zhao,W., Liu,J. and Peng,Y. (2012) Structural features of the single-stranded DNA-binding protein MoSub1 from Magnaporthe oryzae. *Acta Crystallogr. D. Biol. Crystallogr.*, **68**, 1071–1076.
18. Werten,S., Kohler,C., Bayer,N.J., Steinmetz,I. and Hinrichs,W. (2016) Structural analysis and knock-out of a Burkholderia pseudomallei homolog of the eukaryotic transcription coactivator PC4. *Gene*, **577**, 140–147.
19. Grove,D.E., Willcox,S., Griffith,J.D. and Bryant,F.R. (2005) Differential single-stranded DNA binding properties of the paralogous SsbA and SsbB proteins from Streptococcus pneumoniae. *J. Biol. Chem.*, **280**, 11067–11073.
20. Li,S., Ramelot,T.A., Kennedy,M.A., Liu,M. and Yang,Y. (2016) Chemical shift assignments of the homodimer protein SP_0782 (7-79) from Streptococcus pneumoniae. *Biomol. NMR Assign.*, **10**, 341–344.
21. Shen,Y., Delaglio,F., Cornilescu,G. and Bax,A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
22. Huang,Y.J., Powers,R. and Montelione,G.T. (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.*, **127**, 1665–1674.
23. Linge,J.P., Williams,M.A., Spronk,C.A., Bonvin,A.M. and Nilges,M. (2003) Refinement of protein structures in explicit solvent. *Proteins*, **50**, 496–506.
24. Bhattacharya,A., Tejero,R. and Montelione,G.T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins*, **66**, 778–795.
25. Williamson,M.P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.*, **73**, 1–16.
26. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
27. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
28. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2126–2132.
29. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 213–221.

30. Hodel,A. and S.-H.K.a.A.T.B. (1992) Model bias in macromolecular crystal structures. *Acta Crystallogr. A*, **A48**, 851–858.
31. Theobald,D.L. and Wuttke,D.S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22**, 2171–2172.
32. Kuzin,A., Abashidze,M., Lew,S., Seetharaman,J., Patel,P., Xiao,R., Ciccosanti,C., Lee,D., Everett,J.K., Nair,R. *et al.* (2010) X-ray crystal structure of protein SP_0782 (7-79) from Streptococcus pneumoniae. Northeast Structural Genomics Consortium Target SpR104.
33. Ashkenazy,H., Abadi,S., Martz,E., Chay,O., Mayrose,I., Pupko,T. and Ben-Tal,N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, **44**, W344–W350.
34. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. U.S.A.*, **98**, 10037–10041.
35. Kwan,A.H., Mobli,M., Gooley,P.R., King,G.F. and Mackay,J.P. (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.*, **278**, 687–703.
36. Demers,J.P. and Mittermaier,A. (2009) Binding mechanism of an SH3 domain studied by NMR and ITC. *J. Am. Chem. Soc.*, **131**, 4355–4367.
37. Kuzin,A.P., Su,M., Seetharaman,J., Patel,P., Xiao,R., Ciccosanti,C., Lee,D., Everett,J.K., Acton,T.B., Montelione,G.T. *et al.* (2012) Crystal structure of protein SP_0782 (7-79) from Streptococcus pneumoniae complexed with ssDNA. Northeast Structural Genomics Consortium (NESG) target SPR104.
38. Huang,J., Zhao,Y., Liu,H., Huang,D., Cheng,X., Zhao,W., Taylor,I.A., Liu,J. and Peng,Y.L. (2015) Substitution of tryptophan 89 with tyrosine switches the DNA binding mode of PC4. *Sci. Rep.*, **5**, 8789.
39. Antony,E. and Lohman,T.M. (2019) Dynamics of E. coli single stranded DNA binding (SSB) protein-DNA complexes. *Semin. Cell Dev. Biol.*, **86**, 102–111.
40. Su,X.C., Wang,Y., Yagi,H., Shishmarev,D., Mason,C.E., Smith,P.J., Vandevenne,M., Dixon,N.E. and Otting,G. (2014) Bound or free: interaction of the C-terminal domain of Escherichia coli single-stranded DNA-binding protein (SSB) with the tetrameric core of SSB. *Biochemistry*, **53**, 1925–1934.
41. Mortusewicz,O., Evers,B. and Helleday,T. (2016) PC4 promotes genome stability and DNA repair through binding of ssDNA at DNA damage sites. *Oncogene*, **35**, 761–770.
42. Wang,J.Y., Sarker,A.H., Cooper,P.K. and Volkert,M.R. (2004) The single-strand DNA binding activity of human PC4 prevents mutagenesis and killing by oxidative DNA damage. *Mol. Cell Biol.*, **24**, 6084–6093.
43. Yu,L., Ma,H., Ji,X. and Volkert,M.R. (2016) The Sub1 nuclear protein protects DNA from oxidative damage. *Mol. Cell Biochem.*, **412**, 165–171.
44. Straume,D., Stamsas,G.A. and Havarstein,L.S. (2015) Natural transformation and genome evolution in Streptococcus pneumoniae. *Infect. Genet. Evol.*, **33**, 371–380.
45. Lin,J., Zhu,L. and Lau,G.W. (2016) Disentangling competence for genetic transformation and virulence in Streptococcus pneumoniae. *Curr. Genet.*, **62**, 97–103.
46. Attaiech,L., Olivier,A., Mortier-Barriere,I., Soulet,A.L., Granadel,C., Martin,B., Polard,P. and Claverys,J.P. (2011) Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. *PLos Genet.*, **7**, e1002156.
47. Lisboa,J., Andreani,J., Sanchez,D., Boudes,M., Collinet,B., Liger,D., van Tilbeurgh,H., Guerois,R. and Quevillon-Cheruel,S. (2014) Molecular determinants of the DprA-RecA interaction for nucleation on ssDNA. *Nucleic Acids Res.*, **42**, 7395–7408.
48. Mortier-Barriere,I., Velten,M., Dupaigne,P., Mirouze,N., Pietrement,O., McGovern,S., Fichant,G., Martin,B., Noirot,P., Le Cam,E. *et al.* (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell*, **130**, 824–836.
49. Quevillon-Cheruel,S., Campo,N., Mirouze,N., Mortier-Barriere,I., Brooks,M.A., Boudes,M., Durand,D., Soulet,A.L., Lisboa,J., Noirot,P. *et al.* (2012) Structure-function analysis of pneumococcal DprA protein reveals that dimerization is crucial for loading RecA recombinase onto DNA during transformation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2466–E2475.
50. Peterson,S.N., Sung,C.K., Cline,R., Desai,B.V., Snesrud,E.C., Luo,P., Walling,J., Li,H., Mintz,M., Tsegaye,G. *et al.* (2004) Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Mol. Microbiol.*, **51**, 1051–1070.