Design and Characterization of a Multimeric DNA Binding Protein Using Sac7d and GCN4 as Templates

Sz-Wei Wu,[1,2] **Tzu-Ping Ko,**[1] **Chia-Cheng Chou,**[3] **and Andrew H.-J. Wang**[1,2,3*]
[1]*Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan*
[2]*Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan*
[3]*Core Facility for Protein X-ray Crystallography, Academia Sinica, Taipei, Taiwan*

ABSTRACT The protein Sac7d belongs to a class of small chromosomal proteins from the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. Sac7d is extremely stable to heat, acid, and chemical agents. This protein is a monomer and it binds DNA without any particular sequence preference, while inducing a sharp kink in the DNA. By appending a leucine-zipper-like helical peptide derived from the yeast transcriptional activator GCN4 to the C-terminal end of Sac7d, the modified monomers (denoted S7dLZ) are expected to interact with each other via hydrophobic force to form a parallel dimer. The recombinant S7dLZ was expressed in *Escherichia coli* and purified by heating and ion-exchange chromatography. The formation of dimer was detected by gel-filtration chromatography and chemical cross-link. The results of surface plasmon resonance and circular dichroism experiments showed that the DNA-binding capacity was retained. Furthermore, X-ray diffraction analysis of single crystals of S7dLZ in complex with DNA decamer CCTATATAGG showed that the leucine-zipper segments of S7dLZ were associated into an antiparallel four-helix bundle. There are two DNA fragments bound to each S7dLZ tetramer in the crystal. This model works as a successful template that endows protein a new function without losing original properties. Proteins 2005;60:617–628. © 2005 Wiley-Liss, Inc.

Key words: leucine zipper; coiled coil; X-ray diffraction; crystal structure; hyperthermophile; transcription factor; genetic engineering; four helix bundle

INTRODUCTION

The archaebacterium *Sulfolobous acidocaldarius*, which lives in hot springs, is highly resistant to heat and acid. Because the GC content in the DNA of *S. acidocaldarius* is only about 40%, comparable to those of mesophiles, such acidothermoplilic organisms should circumvent the risk of DNA denaturation by certain mechanisms.[1] Five different chromosomal DNA-binding proteins with molecular weights of about 7 kDa have been isolated from *S. acidocaldarius*. They have similar amino acid sequences and are named Sac7a to 7e according to their isoelectric points.[2] The Sac7d protein in this study comprises 66 residues, with a total mass of 7.6 kDa. The solution structure of Sac7d was first determined by NMR spectroscopy,[3] which showed a compact folding with a C-terminal α-helix packed against an incomplete β-barrel consisting of five antiparallel strands. X-ray diffraction analysis of the crystal structure of Sac7d in complex with DNA further showed that the protein binds to DNA through the minor groove, while intercalation of two side chains of Val26 and Met29 into adjacent base pairs produces a sharp kink of about 72° in the double helix of DNA.[4] Sac7d increases the melting temperature of DNA by more than 30°C in vitro with no preference for specific nucleotide sequence. Presumably this will stabilize duplex DNA and help its packaging in the acidothermophilic cells.

Coiled coils, which have been observed in many proteins, are composed of closely packed α-helices with the opposing hydrophobic side chains in a knobs-into-holes arrangement.[5] The leucine zipper is a class of coiled coils in which two parallel helices are juxtaposed via a leucine-rich hydrophobic interface.[6] Such modules are frequently adopted in DNA-binding proteins for gene regulation, as in the Fos-Jun complex.[7] Besides hydrophobic interactions, the interface of Fos-Jun complex also contains a number of salt bridges, which may account for their specificity. Another typical example of leucine zipper is the yeast transcriptional activator GCN4,[8] which contains an eight-turn α-helix with one leucine residue in every seven amino acids. GCN4 has a very high affinity for the AP-1 binding site of DNA, but it functions only as a dimer. Interestingly, the coiled-coil structure of GCN4 is retained even when the sequence is reversed into a retro-peptide.[9] Retro-GCN4 forms a parallel four-helix bundle in the crystal.

Fusion proteins are frequently seen recently as a result of the fast growing technologies of protein engineering. One of the most widely used applications is to help purification of recombinant proteins by appending a $(His)_n$-tag to the protein and by subsequent immobilized metal-ion affinity chromatography.[10] Larger affinity tags such as
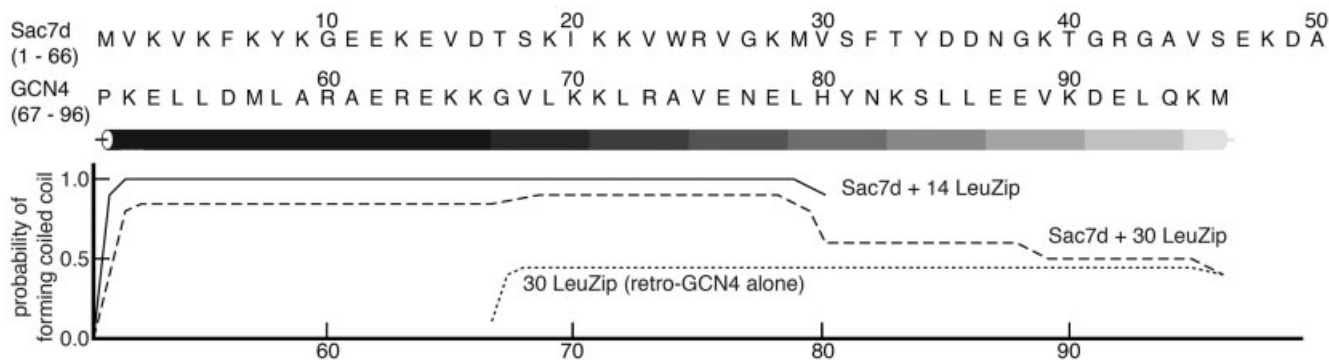
Fig. 1. Sequence and secondary structure prediction of S7dLZ. The original Sac7d sequence (1–66) and the appended retro-GCN4 leucine zipper sequence (67–96) are shown on the top. The C-terminal α-helix in Sac7d is shown as a cylinder below the sequence, and it is extended for possible propagation through the GCN4 peptide. The corresponding probability for coiled coil formation, calculated using MultiCoil, is shown in the plot below the sequence. The chimera protein S7dLZ in this study consists of the entire Sac7d sequence and 14 additional residues (from Gly67 to His80) of the retro-GCN4.

maltose binding protein, thioredoxin or glutathione S-transferase fused to a number of originally insoluble recombinant proteins are also effective in increasing their solubility.[11,12] Furthermore, by incorporating a specific module from a foreign protein, the recombinant protein would be endowed with additional functional properties. An example is the dimerization of an artificial HIV enhancer binding peptide by fusion with a GCN4 leucine zipper.[13] In this paper, we produced a fusion protein of Sac7d and retro-GCN4[9] that was expected to form a similar dimer, which may bind to longer DNA sequences by the two Sac7d moieties and cause consecutive sharp kinks in the DNA. The results turned out to be a dimer in solution and a tetramer in crystal, with an antiparallel four-helix bundle, in which the DNA-binding capacity is retained.

MATERIALS AND METHODS
Design of the S7dLZ Chimera

To estimate the propensity of forming a coiled coil structure after appending a leucine zipper to the C-terminus of Sac7d, the protein sequences were analyzed using the program MultiCoil.[14] We chose to use the retro-GCN4 sequences because it yielded the highest score among several leucine zippers, including the native GCN4. The probability of coiled-coil formation was nearly zero for Sac7d alone (data not shown). For retro-GCN4 the probability was about 0.4, as shown in Figure 1. Interestingly, when residues 6–35 of the retro-GCN4 were added to the 66-residue sequence of Sac7d, a sharp rise of the probability to 0.8 was seen at residue 52, which is the beginning of the C-terminal α-helix in the wild-type Sac7d structure. The high probability of 0.8 remained for 28 residues and declined gradually after this to 0.4 at the end of retro-GCN4 sequence. If only 14 residues (6–19) of the retro-GCN4 were included, the probability increased from zero to about 1.0 at residue 52 of Sac7d and continued through the C-terminus. Prediction using CoilScan of the GCG package (Accelrys) yielded similar results. As shown below, these are consistent with the crystal structure that we

determined. This new protein of residues 1–80 in Figure 1 is named S7dLZ.

The selection of a particular segment of the leucine zipper, which was to be appended to Sac7d, should also take into consideration that the original DNA binding capacity would not be impaired after oligomer formation. Consequently, a model was constructed using the program O[15] in which the known Sac7d–DNA complex (PDB code: 1AZP) was extended with the 14 additional residues of S7dLZ propagating as a continuous α-helix, starting at Lys52. Using the crystal structure of retro-GCN4 (PDB code: 1C94) as a template, the leucine zipper moiety of S7dLZ was superimposed on each subunit of the dimer, while the Sac7d and bound DNA were also moved accordingly. The dimeric model suggested that the S7dLZ molecules would have limited bumps with only a few atoms of the counter subunit, near the two-fold axis. The DNA-binding sites on the S7dLZ subunits would be directed toward opposite sides and would remain capable of binding DNA. They did not involve the C-terminal α-helices and neither were they blocked by the formation of the leucine zipper.

Expression and Purification of the Fusion Protein

The primer elongation protocol of polymerase chain reaction (PCR) was employed to append the DNA sequence of the retro-GCN4 leucine zipper (GVLKKLRAVENELH) to the gene of Sac7d. The plasmid of pET-3b/sac7d[16] that contained the cloned Sac7d gene was used as the template for PCR, and the T7 promoter primer was used as the 5′-end primer. For the 3′-end, a primer of 5′-cgc **ggatcc** *tta* atg cag ttc gtt ttc cac ggc acg cag ttt ttt cag cac acc ttt ctt ctc tct ttc tgc tct tgc-3′ was designed and synthesized, which contained a cleavage site for the restriction enzyme *Bam*H I. The insert gene produced by PCR was ligated with the plasmid pET-11a (Novagen), after treatment with *Bam*H I and *Nde* I. After transformation, the competent cells of BL21 (DE3; Novagen) were grown on agar plates of Luria-Bertani medium that contained 50 μg/mL Ampicillin (LBA medium). A positive colony that hosted the

plasmid with correct gene insert of S7dLZ was verified by sequencing. The selected colony was further propagated in LBA medium and stored at −20°C after addition of 15% glycerol.

To express the recombinant S7dLZ protein in large quantities, 10 mL of an overnight culture was inoculated to 1 L of LBA medium and incubated at 37°C. When the absorbance at 600 nm reached 1.0, IPTG was added to a final concentration of 0.4 mM for induction. The optimal length for maximal yield of induced expression was 2–2.5 h. After harvesting by centrifuge, the *Escherichia coli* cells were resuspended in 20 mM Tris-HCl (pH 7.4, approximately 10 mL for 1 L of broth) and sonicated. The original procedure for purification of the recombinant Sac7d protein includes treatments with heat and acid.[16] Because the recombinant protein was expressed in *E. coli*, these steps were effective in removing most of the proteins that are not resistant to the extreme conditions. Although the S7dLZ protein is also resistant to both heat and acid (data not shown), to obviate subsequent neutralization step for more efficient purification, it was only treated with heat. After a few trials, we determined that the optimal condition was heating at 70°C for 30 min, followed by removal of precipitate with centrifuge.

Using the AKTA Purifier 900 FPLC system (Amersham Pharmacia), the supernatant was loaded directly to an SP-Sepharose ion-exchange chromatography column equilibrated with 20 mM Tris-HCl (pH 7.4) and washed extensively with the same buffer. Then the column was eluted with 15 volumes of 0–1.0 M NaCl gradient in the buffer. The fractions were analyzed by Tris-Glycine SDS-PAGE (Invitrogen) and those containing the 9-kDa recombinant S7dLZ protein were collected and dialyzed against distilled water. The final product was lyophilized and stored at −20°C.

Circular Dichroism Measurements

CD spectra were measured using a JASCO J-720 Spectrophotometer at room temperature. The photocell had a volume of 200 µL and a path length of 1.0 mm. The protein and DNA samples were dissolved in distilled water. The concentration for both Sac7d and the recombinant S7dLZ was 0.5 mg/mL in measurements of the protein-alone spectra. For the DNA complexes, the samples were prepared by mixing equal volumes of 0.5 mg/mL DNA (16 base pairs, CGCGTACGCGTACGCG) solution with either 0.4 mg/mL Sac7d or 0.5 mg/mL S7dLZ. These were incubated for 30 min prior to spectra measurements.

Gel Filtration and Cross-Link Experiments

An HPLC system, including a Waters 1525 binary pump and a Waters 2487 dual λ absorbance detector, and a silica gel column of Shodex KW-802.5 were used in gel-filtration chromatography analysis. The column was equilibrated with phosphate buffered saline (PBS) prior to injection of the samples. The molecular weight markers used were albumin (66 kDa), carbonic anhydrase (30 kDa), cytochrome c (12.4 kDa) and aprotinin (6.5 kDa). Protein samples including S7dLZ were dissolved in PBS with

concentration of 1 mg/mL and injected separately to measure retention time.

To compare oligomeric states of Sac7d and S7dLZ, 20 µL of 2 mg/mL protein solutions that contain 0.25% glutaraldehyde were incubated at room temperature for 1, 5 and 10 minutes. The reaction was terminated by addition of 1 µL 1 M Tris-HCl solution (pH 8.0) and standing for 30 min. After concentration to 10 µL using Microcon YM-3 (Millipore), the samples were then analysed using 4–20% gradient Tris-glycine SDS PAGE. To investigate concentration effect on oligomer formation, solutions containing 10, 5, 2, 1, and 0.5 mg/mL S7dLZ protein and 0.25% glutaraldehyde were incubated at 4°C for 20 min. The reaction was terminated by adding 1 M Tris-HCl (pH 8.0) and, after standing on ice for 30 min, the solution was concentrated and analyzed using 12% Bis-Tris SDS PAGE.

Surface Plasmon Resonance Experiments

For kinetic studies of the DNA–protein interactions, the instrument BIACORE-X and the SA chips (BIACORE) were employed. The biotin-labeled DNA molecules (CGCG-TACGCGTACGCG) were first immobilized via streptavidin onto the SA chip by passing 100 µL of 50 nM solution of the DNA at a rate of 20 µL/min through the flow cell. The charging process continued with 100 µL of 50 nM solution of unlabeled complementary DNA for duplex formation. After each measurement of protein binding, the charging procedure was repeated with 50 µL of the biotin-labeled and unlabeled DNA solutions to ensure sufficient number of duplex DNA molecules on the chip.

The protein solutions of Sac7d and S7dLZ were diluted with 10 mM HEPES (pH 7.4) to 57, 114, 171, and 228 nM before use. Before data acquisition, the flow cell was equilibrated with the buffer at a flow rate of 20 µL/min. This was followed by injecting 50 µL of the protein solution for measurement of association rate, and then by a subsequent pause of 100 sec for measurement of dissociation rate. Finally, the chip was washed with 10 µL of 10 mM HCl solution to remove the bound protein. Data analysis was carried out using the BIA Evaluation Software (v3.01).

Crystallization and X-Ray Diffraction Analysis

Both hanging-drop and sitting-drop vapor diffusion methods were employed in initial crystallization trials of the S7dLZ protein, and a hexagonal crystal form was obtained by using 30% PEG400. However, the crystals diffracted poorly and were not suitable for structural analysis. Because previous Sac7d crystals were grown in the presence of DNA, subsequent trials also included DNA molecules of various lengths and sequences, from 8 to 12 base pairs, in different ratios to the protein. The tetragonal crystals used here were grown in a hanging drop that contained a mixture of 1 µL 5 mM S7dLZ, 1 µL 5 mM DNA (5′-CCTATATAGG-3′) and 2 µL reservoir of 1.6 M sodium citrate at pH 6.5. The drop was equilibrated against 500 µL of the reservoir at room temperature and the crystals grew in two or three weeks to a size of $0.1 \times 0.1 \times 0.2$ mm^3.

Data collection was carried out using the CuKα radiation from a Rigaku MicroMax007 rotating-anode genera-

tor (Molecular Structure Corporation) at $-150°C$. The diffraction images were recorded by using the R-Axis IV[++] system (MSC) and processed by using the program package of XDS.[17] Subsequent structure determination and refinement were carried out using the program CNS.[18] The crystal had unit-cell dimensions of $a = b = 58.65$, $c = 186.68$ Å and belonged to a space group of $P4_x2_x2$. Preliminary estimate by Matthews coefficient[19] suggested that there was more than one S7dLZ–DNA complex in the asymmetric unit. However, molecular replacement searches with the Sac7d model in complex with the same 10-base-pair DNA[20] yielded only one solution in the space group of $P4_122$. No solution could be obtained with other space groups.

An initial electron density map was calculated at 2.6 Å resolution, with an overall R-value of 0.46. The map showed good corresponding densities for the model, including some extra densities for the C-terminal α-helix. Adjacent to this helical region, the map clearly indicated the presence of another α-helix. In addition, a nearby region showed densities for the β-strands of a second Sac7d molecule, and this was fitted manually into the map. The new model, containing two Sac7d molecules and a 10-base pair DNA in an asymmetric unit reduced the R-value to 0.42 after rigid-body refinement. After extension of the C-terminal helices of two S7dLZ molecules and adjustments of the 5′- and 3′-ends of the DNA molecule, the model yielded R and R_{free} values of 0.249 and 0.288 upon simulated annealing and B-value refinements. The refinement continued with addition of water molecules and further adjustment of the protein and DNA models.

Further structural analysis employed the programs O[15] and CCP4.[21] The sequence in Figure 1 was prepared using ALSCRIPT.[22] Figures 5(a–c) 6, and 7(a) were produced using MolScript[23] and Raster3D.[24] Figure 7(b) was generated using the program NUCPLOT.[25]

RESULTS
The Recombinant S7dLZ Protein forms Oligomers Spontaneously

The two-step purification by heating and SP-Sepharose ion exchange chromatography is effective in obtaining the S7dLZ protein in large quantities sufficient for subsequent studies. By obviating acid treatment as used for the wild type Sac7d,[16] the S7dLZ protein can be purified using a suitable salt gradient in chromatography with increased yield. As shown in Figure 2(a), the S7dLZ protein migrated as a single band on the SDS gel, which corresponded to an apparent molecular mass higher than Sac7d. Analysis by Electron-Spray Ionization Mass Spectrometry showed the molecular weight of the protein was 9197, consistent with the calculated value from the amino acid sequence. The circular dichroism (CD) spectra of Sac7d and S7dLZ were very similar, and the slight differences could be attributed to the higher content of helical secondary structure in S7dLZ (data not shown).

Gel filtration experiments showed that the retention time of S7dLZ was slightly shorter than cytochrome c. Plot of retention time against the logarithm of molecular
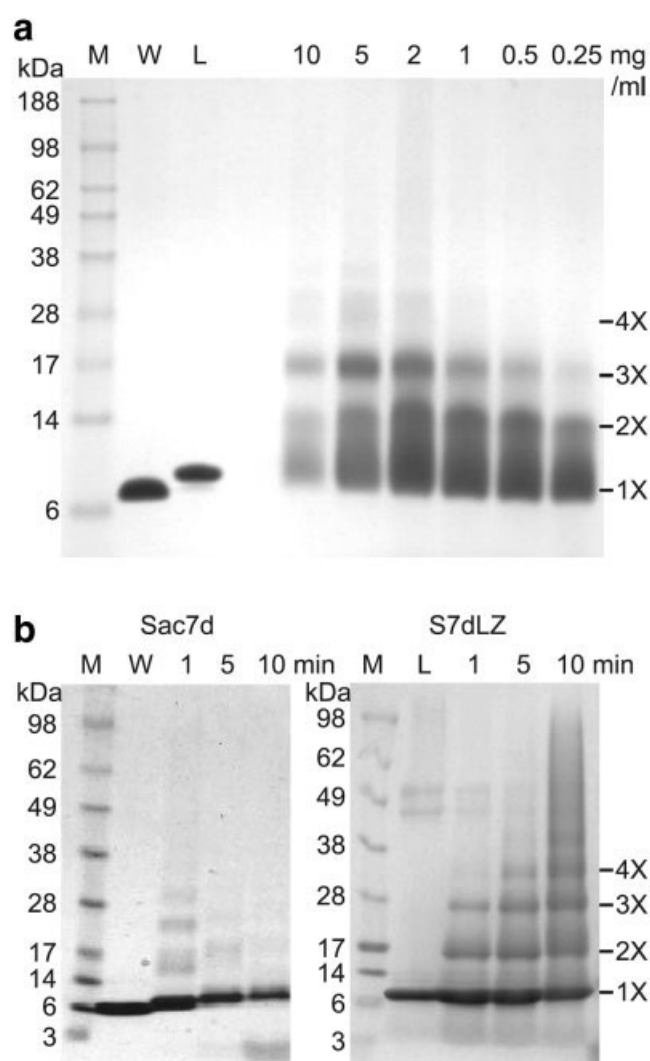
Fig. 2. Chemical cross-linking experiments using glutaraldehyde as a reagent. **a**: Different concentration (10, 5, 2, 1, 0.5, and 0.25 mg/mL) of S7dLZ (with equal amounts of protein, 40 μg) was treated with 0.25% (final concentration) glutaraldehyde for 20 min at 4°C. The cross-linking products were separated using 12% Bis-Tris SDS PAGE and stained with coomassie blue. **b**: Sac7d and S7dLZ proteins at same concentration (40 μg in 20 μL) were reacted with equal amounts of glutaraldehyde (final in 0.25%) for 1, 5, and 10 min at room temperature. The products were separated using 4–20% Tris-Glycine SDS PAGE and stained with coomassie blue. M: marker; W: untreated Sac7d; L: untreated S7dLZ.

weights, with a correlation coefficient of 0.9962, suggested that the apparent molecular weight of S7dLZ was 15.4 kDa, significantly higher than that of a monomer. Therefore, most of the S7dLZ protein should form dimers in 1 mg/mL protein concentration in solution. The longer retention time than expected for an 18.4 kDa dimer was probably caused by polar interactions between the positively charged protein and the silicate material of the column. Such retardation also caused an asymmetric peak profile, whereas other peaks had normal Gaussian distribution. The position and shape of the peak could also be a result of the fluctuating distribution of monomers and dimers in equilibrium.
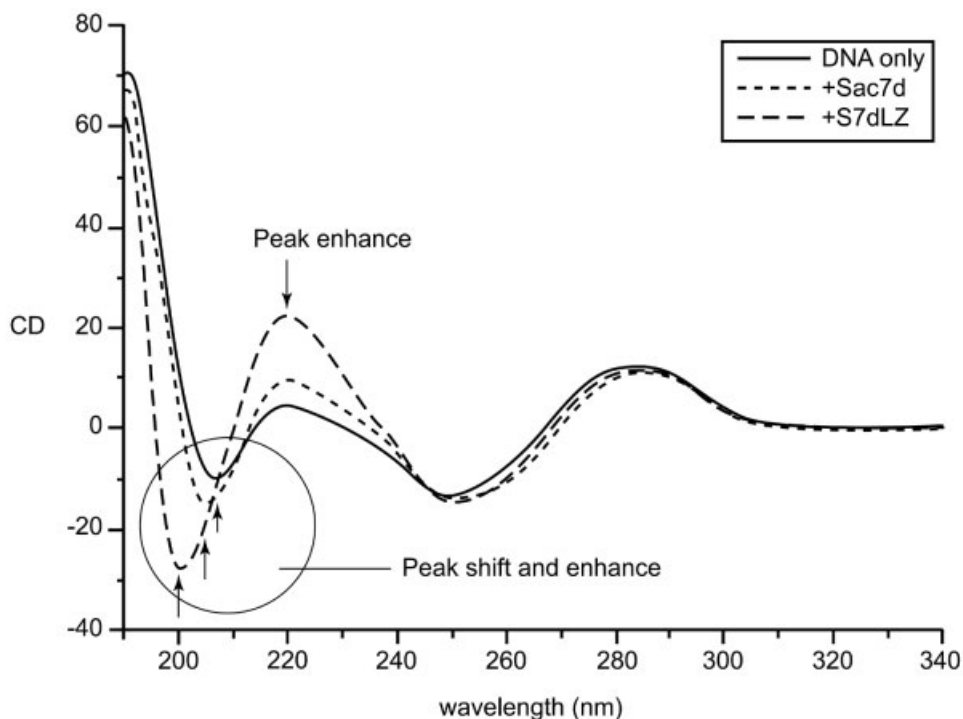
Fig. 3. Comparison of the effects of S7dLZ and Sac7d on dsDNA conformation using CD spectroscopy. The positive peaks of DNA near 220 nm both increased in strength upon addition of S7dLZ and Sac7d, while the strength difference for S7dLZ was about twice as much as that for Sac7d. The negative peaks of DNA at about 205 nm were also enhanced and shifted toward the far-UV region. Again, the shift and enhancement were more prominent for S7dLZ. The peak positions are denoted using arrows.

In Figure 2(b), four bands that corresponded to monomer, dimer, trimer, and tetramer of the S7dLZ protein are seen in the crosslinking experiments. It shows that the extent of cross-link increased with the time of glutaraldehyde treatment. In contrast, the wild type Sac7d protein did not yield similar results at the same protein concentration. Only a few minor bands were observed, probably due to nonspecific interactions. Furthermore, experiments with different protein concentrations [Fig. 2(a)] showed that S7dLZ associated into dimer even after dilution to 0.25 mg/mL. At higher concentrations, bands for trimer and tetramer were also observed. These results suggest that the S7dLZ protein is capable of oligomerization after addition of the C-terminal leucine zipper, and the major form of assembly could be a dimer.

The Recombinant S7dLZ Protein Binds DNA

The CD spectrum of a mixture of the S7dLZ protein and a 16-base pair duplex DNA showed significant differences from that calculated by combining individual spectra of the protein and DNA. Similar results were obtained when the wild-type Sac7d protein was used. The differences in spectra were indicative of structural alterations in the protein and/or DNA molecules when they were mixed, that is, the proteins were bound to the DNA. However, previous results showed that the structure of Sac7d was not altered significantly upon DNA binding.[4] Therefore, by assuming that the changes in CD spectra upon mixing the DNA and

protein solutions were caused by the conformational changes in DNA molecules alone, two spectra were calculated by subtracting the protein spectra from those of the protein–DNA mixtures. As shown in Figure 3, the effects of S7dLZ on the DNA structure were much more than those of Sac7d. The positive peaks of DNA near 220 nm both increased in strength upon addition of S7dLZ and Sac7d, while the strength difference for S7dLZ was about twice as much as that for Sac7d. The negative peaks of DNA at about 205 nm were also enhanced and shifted toward the far-UV region. Again, the shift and enhance were more prominent for S7dLZ.

Surface plasmon resonance (SPR) experiments with immobilized ligand of DNA showed that the resonance units (RU) for the analyte S7dLZ were more than two times as much as those for Sac7d. As shown in Figure 4, analytes of different concentrations up to four-fold yielded similar results. Because the RU of SPR is proportional to the molecular mass of the bound analyte, such observation suggested that the major form of the S7dLZ protein bound to the duplex DNA ligand has a molecular mass of about twice as much as the wild type Sac7d protein. Therefore, S7dLZ probably binds to DNA as a dimer.

Crystal Structure of S7dLZ–DNA Complex

The structure of the tetragonal crystal of the S7dLZ–DNA complex was determinated by molecular replacement. The space group is P4$_1$22 and the asymmetric unit
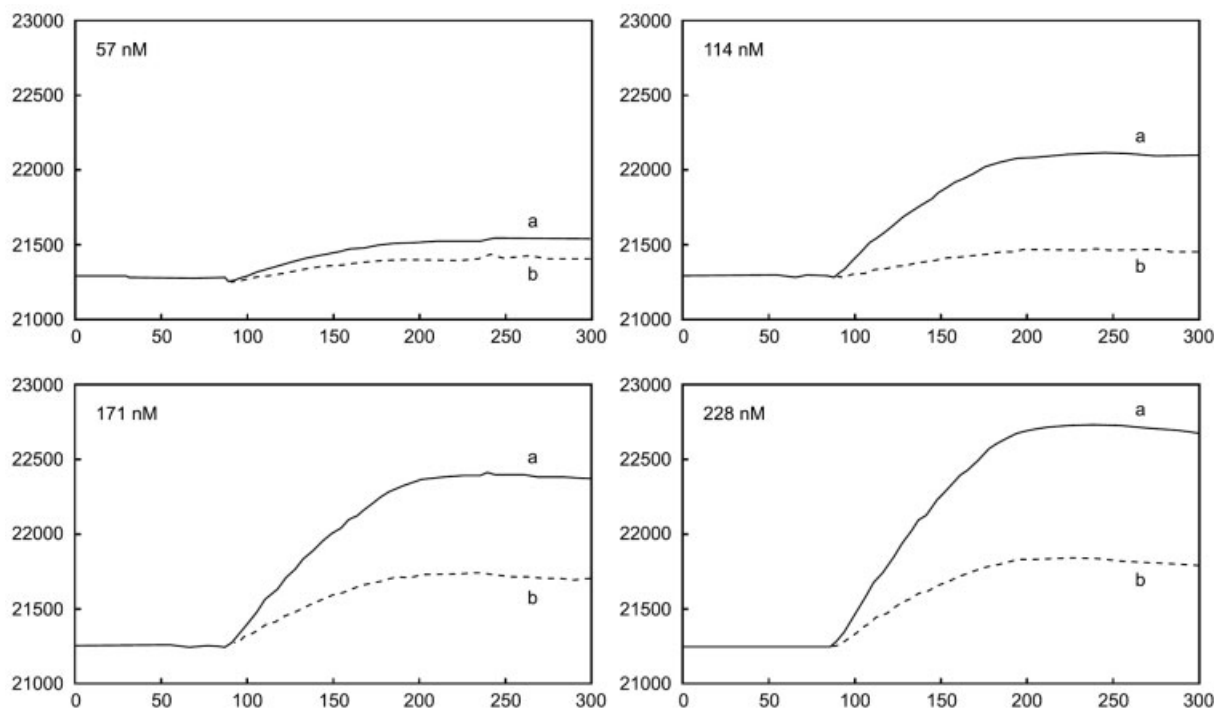
Fig. 4.   Comparison of resonance unit (RU) changes by Sac7d and S7dLZ in SPR. At different concentrations (57, 114, 171, and 228 nM), both proteins interact with the dsDNA ligand on the SPR chip. The perturbations (in RU) by the analyte S7dLZ were more than two times as much as those for Sac7d at all protein concentrations. **a**: S7dLZ; **b**: Sac7d.

contains two S7dLZ molecules and a duplex DNA of 5′-CCTATATAGG-3′. After refinement using 2.6 Å resolution data, the model comprises residues 1–80 for both S7dLZ subunits A and B, residues 102–110 and 111–120 for the DNA molecule, and 146 water molecules. In one of the DNA strands, the 5′ nucleotide did not have clear density in the Fourier map, and could not be modeled with certainty. The final model yielded R and R_{free} values of 0.230 and 0.269, respectively, with acceptable geometry. All of the peptide dihedral (ϕ, ψ) angles are within the allowed regions of Ramachandran plot. Statistics for the diffraction data and the refined model are shown in Table I. The atomic coordinates and structure factors have been deposited with the RCSB Protein Data Bank (accession code 1WVL).

As shown in Figure 5(a), the two S7dLZ molecules in a crystallographic asymmetric unit appeared to be associated via the extended C-terminal helices in a parallel manner into a dimer. The monomer to which the DNA molecule was bound is designated monomer A, and the other monomer B. The dimer was further associated with another dimer into a tetramer, as shown in Figure 5(b). These two dimers were related by a crystallographic dyad axis along the diagonal in the a–b plane of the unit cell, and thus the four α-helices were arranged into an anti-parallel bundle. Figure 5(c) is a different view of the tetramer, which resembles some creature with a wide mouth (the helix bundle) and two big eyes (DNA). The four-helix bundle is presented as helical wheel diagrams in Figure 5(d) and, as will be discussed below, suggests that an antiparallel dimer may be preferred.

TABLE I. Data Collection and Refinement Statistics[†]

Data collection	
Space group	P4$_1$22
Unit cell a, b (Å)	58.65
c (Å)	186.68
Resolution (Å)	20–2.6 (2.70–2.60)
Number of observations	111603 (8224)
Unique reflections	10690 (1110)
R_{merge} (%)	7.7 (22.0)
Completeness (%)	99.5 (100.0)
Refinement	
Number of reflections (>3.0 σ(F))	10550 (1110)
R_{work} (95% data)	0.230 (0.382)
R_{free} (5% data)	0.269 (0.420)
RMSD bond distance (Å)	0.025
RMSD bond angle (degree)	2.48
Ramachandran plot (percent non-Gly and non-Pro residues)	
In most favored regions	87.3
In additional allowed regions	12.7
Average B (Å^2)/Number of non-H atoms	
Protein	56.5/1286
DNA	45.9/389
Water	58.7/146

[†]Numbers in parentheses are for the highest resolution shell.

Crystal Packing Analysis

DNA molecules form a continuous virtual strand in all other known Sac7d crystals.[4,20] Specifically, the DNA fragments are stacked with its neighbors related by crystal-
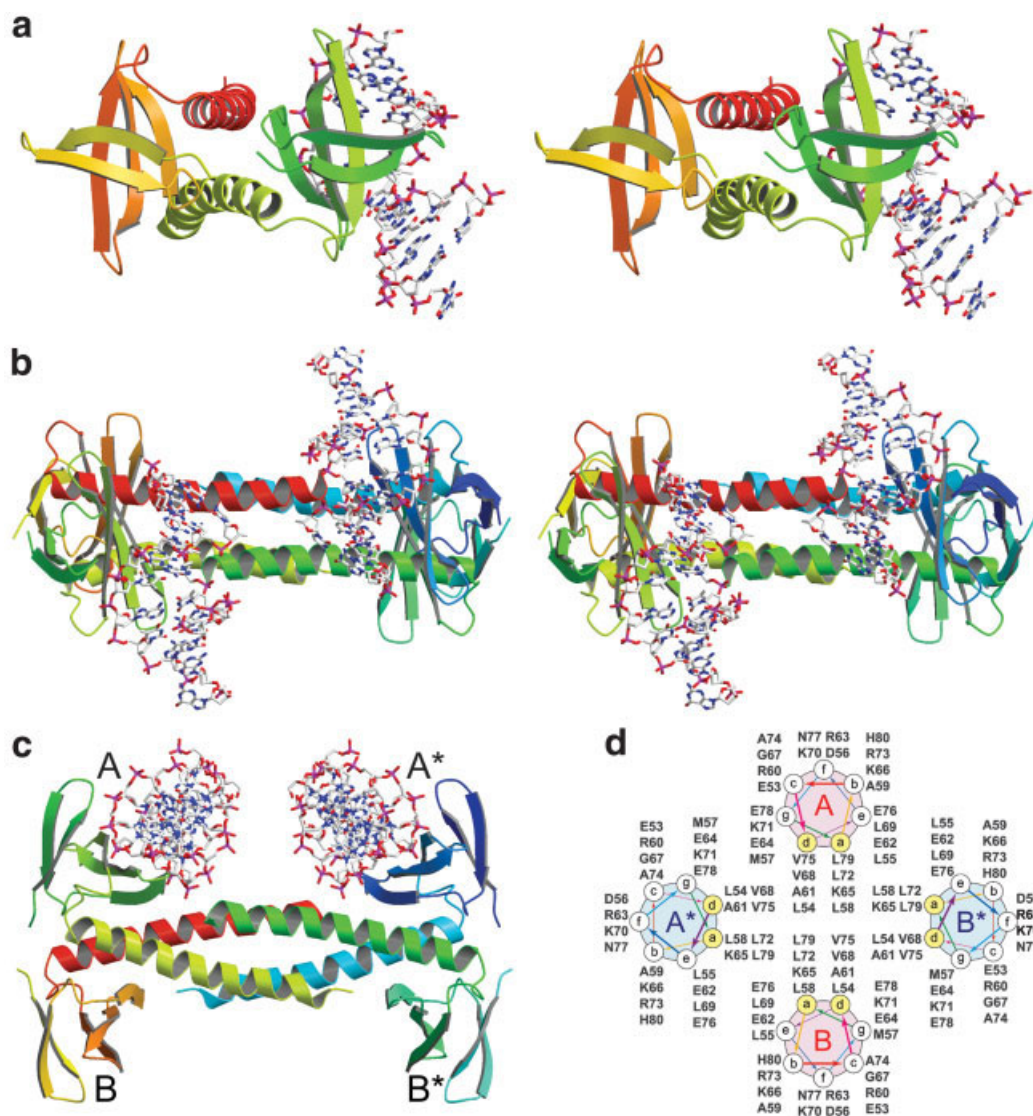
Fig. 5.   Oligomers of S7dLZ. **a**: The dimeric S7dLZ that comprises the asymmetric unit in the tetragonal crystal. **b**: A view of the tetrameric S7dLZ along the crystallographic twofold axis, with an apparent 222 symmetry for the protein molecules. **c**: Another view of the tetrameric S7dLZ. **d**: The four-helix bundle region shown as helical wheels.

lographic 2_1 screw axes, in which the 3′ end of every strand is adjacent to the 5′ end of the opposing strand. The Sac7d molecules are bound to these virtual double helices of DNA in an alternating manner throughout the crystal. In the S7dLZ–DNA crystal studied here, however, the packing is completely different. The DNA fragments do not form virtual strands but are arranged in pairs, with opposite directions. As shown in Figure 6(a), the S7dLZ tetramers make extensive contacts with neighboring molecules related by the crystallographic 4_1 screw axis, with six specific bonds. The surface area covered by this type of lattice contact is 840 Å^2 on each side.

The second type of lattice contact interactions occurs between the protein molecule and the DNA fragment bound to the tetramer related by unit-cell translation in the a–b plane [Fig. 6(b)], with seven specific bonds. The buried surface areas are 458 Å^2 on the protein and 537 Å^2 on the DNA. Furthermore, the DNA fragment is in contact with that bound to another neighboring S7dLZ molecule related by the 4_1 screw axis plus unit-cell translation in the a–b plane. The interactions consist mostly of base stacking between two symmetry-related Cyt102 and triple base pairs as shown in Figure 6(c). Besides the Watson–Crick base pair between Cyt102 and Gua119, the symmetry-related Gua120 makes two additional hydrogen bonds to them. In contrast to the DNA-interactions in other crystal forms, the DNA fragments are arranged head-to-head and related by a crystallographic dyad symmetry axis. This third type of interface buries 405 Å^2 on each DNA fragment.

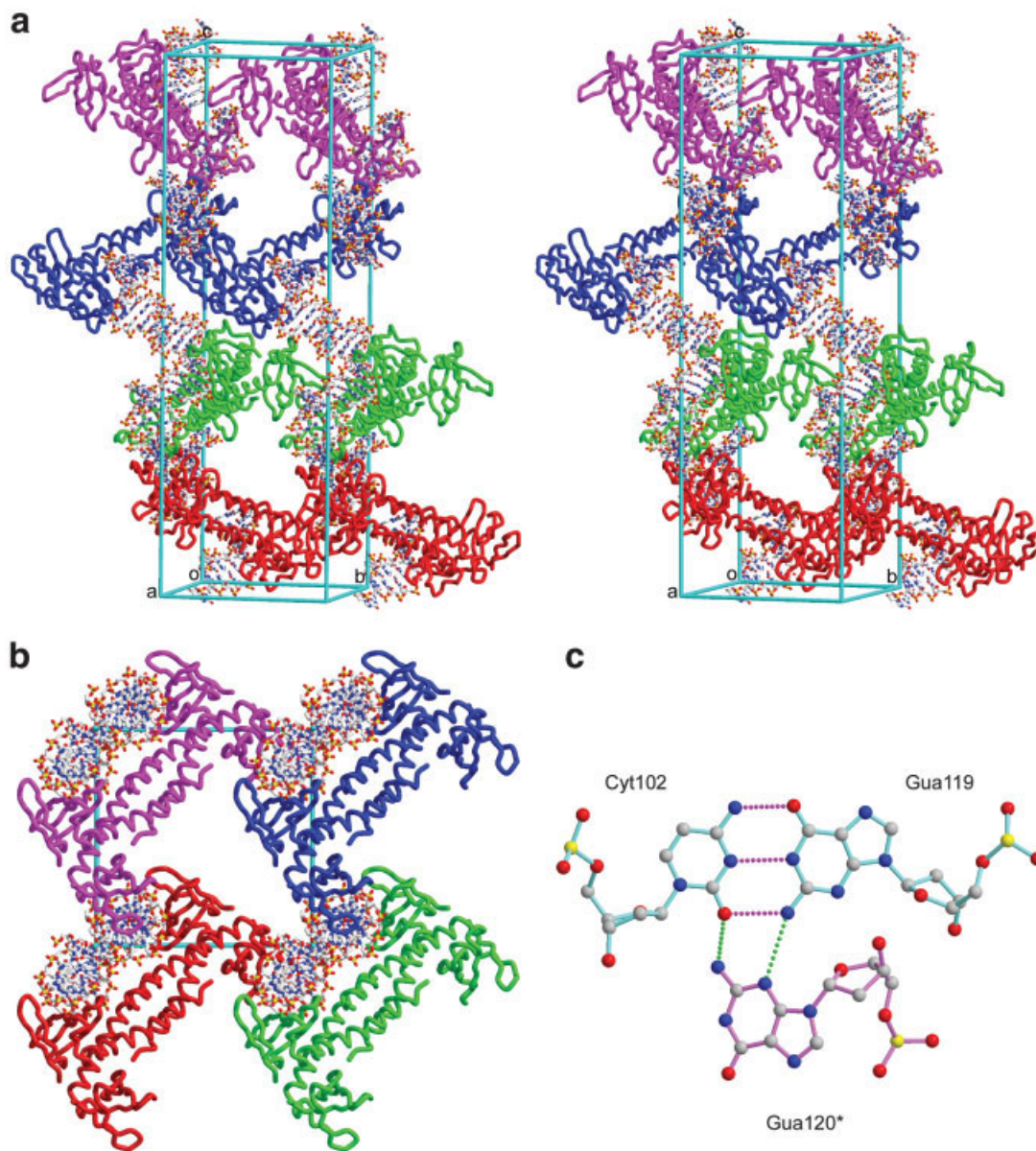In the tetragonal crystal, the DNA binding site of monomer B is occluded by the neighboring molecule,

Fig. 6. Crystal lattice interactions of the tetrameric S7dLZ and DNA. In (**a**) and (**b**) are two views of crystal packing from the side and top of the unit cell. Protein molecules are shown as tubes in different colors, whereas the DNA molecules are shown as sticks. (**c**) Triple base pair of DNA observed in the crystal. Potential hydrogen bonds are indicated by strings of small balls.

specifically, by the N-terminal β-hairpin loop of monomer A [Fig. 6(b)]. However, it does not account for the asymmetric binding of DNA to the S7dLZ tetramer, because the crystal-packing interactions are much weaker than the protein–DNA interactions in general. Instead, the half occupancy of DNA in this crystal is a result of the crystallization condition. Because the DNA should form a duplex before binding to the protein, with equal molar amounts (1 μL, 5 mM) of S7dLZ and DNA in solution, only one half of the protein monomers are bound to the DNA. Crystals suitable for X-ray analysis can only be obtained with such unitary protein–DNA ratio; other protein–DNA ratios should result in higher or lower occupancies of the DNA that would hinder crystallization or lead to crystalli-

zation in other unit cells with different, perhaps less symmetric, space groups.

DISCUSSION
Comparison With Wild Type Sac7d and Protein–DNA Interactions

Although the tetramer of S7dLZ shows an apparent pseudo-222 symmetry [Fig. 5(b)], there are some deviations between the monomers A and B. The two monomers in a parallel dimer [Fig. 5(a)] can be superimposed on each other with a root-mean-square deviation (RMSD) of 2.28 Å between all 160 Cα atoms (2.78 Å for all 1286 nonhydrogen atoms); if individual monomers are superimposed, the RMSD is 1.59 Å for 80 Cα atoms (2.08 Å for 643 non-H
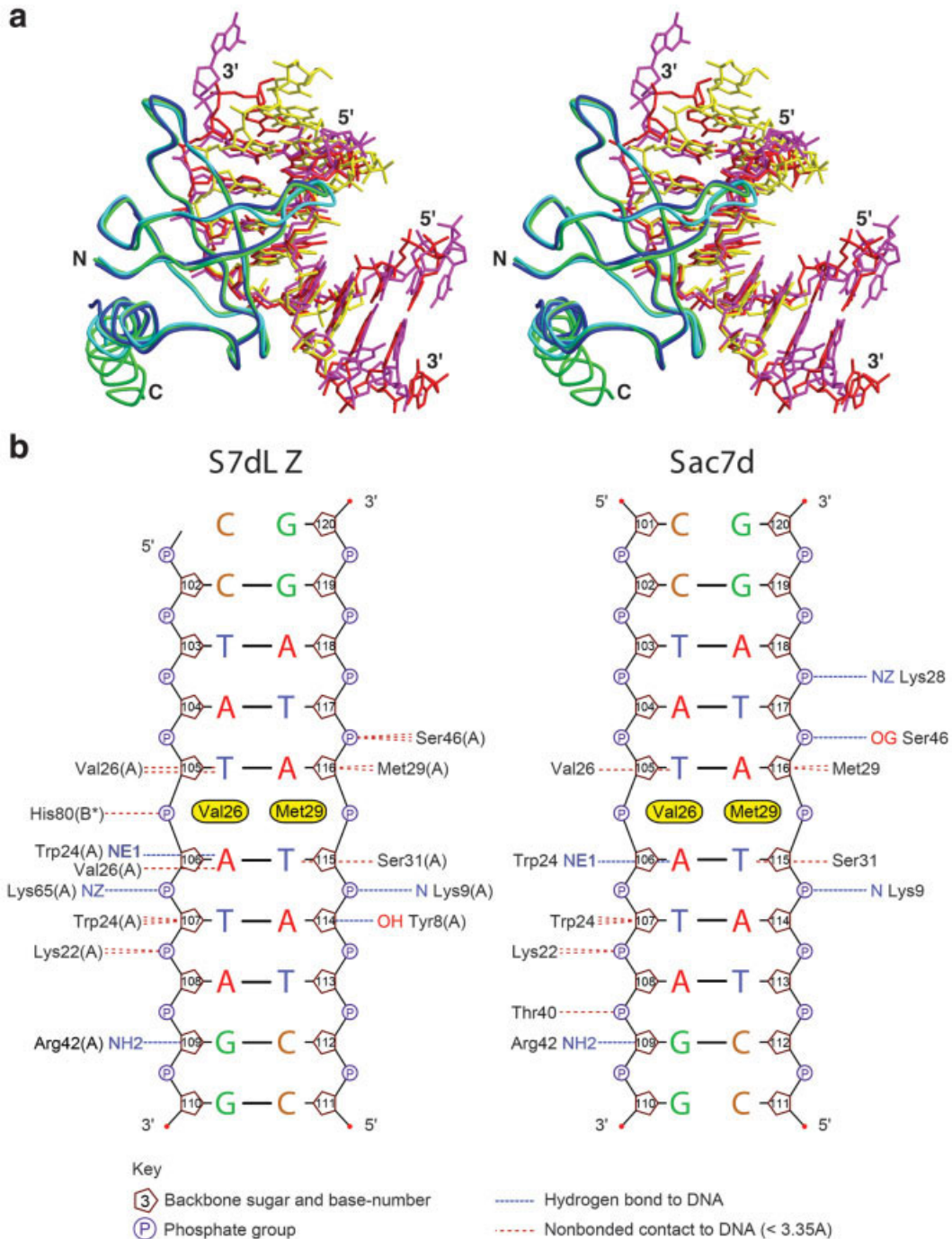
Fig. 7. Structural comparison of S7dLZ and Sac7d bound to DNA. In (**a**) the refined crystal structures of three protein–DNA complexes were overlapped as described in the text. The alpha-carbon tracing models of the wild-type Sac7d bound to DNA octamer (PDB 1AZP) and decamer (1WD0) as well as the engineered S7dLZ are colored in blue, cyan, and green, respectively, with the corresponding DNA molecules in yellow, magenta, and red. In (**b**) the interactions between the DNA molecules of CCTATATAGG decamer and the proteins of S7dLZ and Sac7d are compared in two schematic diagrams.

atoms). Using the least squares procedure of program O, with a matching criterion of 2.0 Å, 108 pairs of Cα atoms are selected in the dimer with an RMSD of 1.15 Å, and it is 0.72 Å between 65 atom pairs in the two individual monomers. In the latter cases, the C-terminal leucine zipper modules do not overlap and only the Sac7d moieties

superimpose well. Excluding the C-terminal residues, both monomers A and B can be superimposed as well on the Sac7d models from two other crystal forms (PDB 1AZP and 1WD0). With the same criterion of 2.0 Å, the RMSD are 0.50–0.67 Å for 61–64 Cα atoms. The C-terminal helices in the S7dLZ dimer are slightly different but these do not

TABLE II. Interactions Between the S7dLZ Molecules Within a Tetramer

Residue 1	Atom 1	Residue 2	Atom 2	Distance (Å)	Comments
Monomers A/B (A*/B*)					
IleA20	O	ArgB60	NH1	2.75	Hydrogen bond; core–helix
LysA21	O	ArgB63	NE	3.47	Hydrogen bond; core–helix
LysA22	NZ	GluB64	OE1	2.65	Salt bridge; Core–helix
ArgA60	NH1	IleB20	O	2.51	Hydrogen bond; core–helix
GluA64	OE1	LysB22	NZ	3.44	Salt bridge; core–helix
LysA65	NZ	GluB64	OE1	3.46	Salt bridge; helix–helix
ValA68	CG1	ValB68	CG1	3.93	Hydrophobic; helix–helix
LeuA72	CD2	LeuB72	CD1	3.80	Hydrophobic; helix–helix
Monomers A/B*(A*/B)					
ValA68	CG1	LeuB79*	CD1	3.95	Hydrophobic; helix–helix
LeuA72	CD1	ValB75*	CG1	4.12	Hydrophobic; helix–helix
ValA75	CG2	LeuB72*	CD2	3.37	Hydrophobic; helix–helix
LeuA79	CD2	LeuB69*	CD1	3.03	Hydrophobic; helix–helix
HisA80	OXT	LysB65*	NZ	2.84	Salt bridge; helix–helix
HisA80	ND1	TrpB24*	CD2	4.05	VDW stacking; core–helix
Monomers A/A*					
LysA71	NZ	GluA78*	OE1	2.72	Salt bridge; helix–helix
LeuA72	CD2	LeuA72*	CD2	3.84	Hydrophobic; helix–helix
GluA78	OE1	LysA71*	NZ	2.72	Salt bridge; helix–helix
Monomers B/B*					
ValB68	CG1	LeuB79*	CD2	4.16	Hydrophobic; helix–helix
LysB71	NZ	GluB78*	OE1	3.02	Salt bridge; helix–helix
ValB75	CG2	ValB75*	CG2	2.77	Hydrophobic; helix–helix
GluB78	OE1	LysB71*	NZ	3.02	Salt bridge; helix–helix
LeuB79	CD2	ValB68*	CG1	4.16	Hydrophobic; helix–helix

incur significant conformational change in the DNA binding module, as shown in Figure 7(a).

The structure of the bound DNA molecule also shows differences from those of PDB 1AZP and 1WD0. The DNA used in crystallization is identical for S7dLZ and 1WD0. The structures have an overall RMSD of 2.63 Å between 389 atoms in 19 nucleotides, whereas the first DNA residue Cyt101 was not observed in the S7dLZ crystal. Such large deviation is due to the open 3′- and 5′-ends in the DNA structure of 1WD0.[20] As shown in Figure 7(b), both proteins bind to the same site of the decamer DNA molecule. If only the eight central base pairs are compared, a much smaller RMSD of 0.84 Å for 328 atoms is obtained. The DNA of 1AZP has a different sequence (GCGATCGC). However, the seven base pairs of 101–107 and 110–116 superimpose well with those of 104–110 and 111–117 in the S7dLZ crystal [Fig. 7(a)]. The RMSD between the 151 corresponding sugar-phosphate backbone atoms is 1.11 Å.

In Figure 7(b), the interactions of S7dLZ and Sac7d with the same DNA molecule are compared. Major interactions between the S7dLZ protein and DNA are similar to those previously observed for the wild-type Sac7d. These include the intercalation of two hydrophobic side chains of Val26 and Met29 into the DNA base pairs to interrupt the base stacking, and thus producing a sharp kink in the DNA double helix. Analysis using the program 3DNA[26] shows that the overall kink angle of DNA produced by S7dLZ is enhanced to 103° (for nine base pairs), with a single-step kink of 54° at the center of the decamer. These values are comparable to those

observed in other Sac7d–DNA complex crystals,[20] with overall kinks of 80–100° (eight base pairs) and single step kinks of 50–60°. The spreading of kinks over several DNA base pairs has been observed previously, as a result of the TA-rich sequence.

Some minor differences in the protein–DNA interactions include the hydrogen bonds of Tyr8 and Lys65 in S7dLZ and those of Lys28 and Thr40 in Sac7d. Adjacent to the intercalation site, His80 of the antiparallel counter monomer B* also interacts with the phosphate group of Ade106. On the other hand, the side chain of His80 of monomer A is stacked with that of Trp24 of monomer B*. The side chain of this Trp24 in both S7dLZ and Sac7d forms a hydrogen bond with the base of Ade106 of the bound DNA [Fig. 7(b)]. Such stacking should be disrupted upon DNA binding in solution.

Protein–Protein Interactions and Oligomerization Mechanism

Interactions between the Sac7d core domains in the parallel S7dLZ dimer [Fig. 5(a)] are more extensive than between the extended C-terminal helices. The total surface areas are 6197 Å^2 and 6298 Å^2 for the monomers A and B, respectively, and the interface buries 643 Å^2 and 704 Å^2. Specific interactions are listed in Table II. There are at least twelve residues from each core domain involved in this interface, but the only two close contacts in the helical regions are between residues Val68 and Leu72. On the other hand, more extensive interactions in the helical regions are observed between monomers A and B of different dimers. The interface involves eleven residues

from each monomer. The buried surface areas are 545 Å² and 572 Å² on monomers A and B, respectively. Most interactions are hydrophobic, including four close contacts between the aliphatic side chains, and are between the antiparallel α-helices, except the stacking of the side chains of HisA80 and TrpB24* and the salt bridge between the C-terminal carboxyl group of HisA80 and the side chain of LysB65* (Table II).

Less extensive interactions are observed between the antiparallel helix pairs of A/A* and B/B* [Fig. 5(c)]. The interfaces have two-fold symmetry as the helices are related by a crystallographic dyad axis. They involve six residues in each monomer of the A/A* and the B/B* pairs, and bury surface areas of 238 Å² and 258 Å², respectively. Specific interactions are also listed in Table II. All interactions are hydrophobic except the two dyad-related salt bridges between LysB71 and GluB78*. In the A/A* pair, similar salt bridges are observed between LysA71 and GluA78*. The helical wheel diagram in Figure 5(d) also shows that the major interactions between the four C-terminal helices in the tetramer are hydrophobic, with some additional salt bridges. Every helix in the four-helix bundle is antiparallel to its neighboring helices and parallel to the diagonal one. Each helix contains at least two heptad repeats (a–g). The core of the four-helix bundle comprises exclusively of hydrophobic residues, including Val and Leu, located in positions a and d inside the bundle, whereas all of the polar or charged residues are distributed about the surface. Thus, the major driving forces in formation of tetramer are hydrophobic interactions between monomers A/B*, A*/B, A/A* and B/B*. Those between monomers A/B and A*/B* have only minor contribution.

Although the interface between monomers A and B in a parallel dimer is the most extensive in the tetramer, it involves more polar bonds in the Sac7d moiety (with corresponding surface areas of 577 Å² and 608 Å², respectively) than hydrophobic bonds in the leucine zipper helices (with areas of 66 Å² and 96 Å²). The Sac7d molecules alone did not show significant oligomerization upon treatment with glutaraldehyde, and oligomerization was only observed after the attachment of the leucine zipper [Fig. 2(b)]. Because the antiparallel dimer A/B* (or A*/B) is associated by the C-terminal helices mostly via hydrophobic interactions, with comparable interface areas as in the parallel dimer A/B, the parallel dimer as shown in Figure 5(a) does not uniquely represent the dominant species in solution, which can be an antiparallel dimer of A/B* as well. It remains unclear whether the engineered leucine zipper favors a parallel or an antiparallel dimer. Further studies by mutagenesis of the interfacial residues should be carried out to resolve this issue.

Concluding Remarks

In this paper we designed a multerimeric DNA binding protein based on the known structures of Sac7d and leucine zipper. We showed that the engineered S7dLZ protein forms dimer and higher oligomers in solution, with dimer being the major species. Because the molar ratio of the protein and duplex DNA is 2:1, although both Sac7d domains in the S7dLZ dimer should retain the DNA-binding capacity, in the tetragonal crystal studied here, only one DNA fragment is bound to each S7dLZ dimer, and the protein molecules further assemble into tetramers, each containing an antiparallel four-helix bundle. The DNA molecules are packed heads on, with inter-duplex base stacking and hydrogen bonds. At the other end they interact with the counter subunits of the S7dLZ tetramer.

In addition, this fusion protein retains most characteristics of Sac7d including thermostability and DNA binding capacity. Similar DNA affinity with K_D of about 10^{-9} M was observed for both proteins using SPR (data not shown). This model works as a successful template that endows the Sac7d protein a new function without losing original properties. Although this is a sequence-general DNA binding protein, it allows further modifications to investigate other possible interactions with the DNA. In a previous study, we showed that Sac7d–K66C mutant linked by disulfide bridge binds to plasmid DNA with a more condensed band.[27] Likewise, since oligomeric S7dLZ can compact and stabilize DNA molecules more efficiently than the wild-type monomeric Sac7d protein, this protein may find applications as DNA carriers or DNA-compacting materials.

REFERENCES

1. Reddy TR, Suryanarayana T. Novel histonelike DNA-binding protein in the nucleoid from the acidothermophilic archeabacterium *Sulfolobous acidocaldarius* that protect DNA against thermal denaturation. Biochim Biophys Acta 1988;949:87–96.
2. Choli T, Wittmann-Liebold B, Reinhardt R. Microsequence analysis of DNA-binding proteins 7a, 7b, and 7e from the archeabacterium *Sulfolobous acidocaldarius*. J Biol Chem 1988;263:7087–7093.
3. Edmondson SP, Qiu L, Shriver JW. Solution structure of the DNA-binding protein Sac7d from the hyperthermophile *Sulfolobous acidocaldarius*. Biochemistry 1995;34:13289–13304.
4. Robinson H, Gao YG, McCrary BS, Edmondson SP, Shriver JW, Wang AHJ. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature 1998;392:202–205.
5. Crick FHC. The packing of α-helices: simple coiled-coils. Acta Crystallogr 1953;6:689–697.
6. Alber T. Structure of the leucine zipper. Curr Opin Genet Dev 1992;2:205–210.
7. Glover JN, Harrison SC. Crystal structure of the heterodimeric bZIP transcription factor c-Fos:c-Jun bound to DNA. Nature 1995;373:257–261.
8. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 1991;254:539–544.
9. Mittl PR, Deillon C, Sargent D, Liu N, Klauser S, Thomas RM, Gutte B, Grutter MG. The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure. Proc Natl Acad Sci USA 2000;97:2562–2566.
10. Crowe J, Dobeli H, Gentz R, Hochuli E, Stuber D, Henco K. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. Methods Mol Biol 1994;31:371–387.
11. Shih YP, Kung WM, Chen JC, Yeh CH, Wang AHJ, Wang TF. High-throughput screening of soluble recombinant proteins. Prot Sci 2002;11:1714–1719.
12. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. Protein Sci 2003;12:1313–1322.
13. Liu N, Caderas G, Gutte B, Thomas RM An artificial HIV enhancer-binding peptide is dimerized by the addition of a leucine zipper. Eur Biophys J 1997;25:399–403.
14. Wolf E, Kim PS, Berger B. MultiCoil: a program for predicting

two- and three-stranded coiled coils. Protein Sci 1997;6:1179–1189.

15. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for the building of protein models in electron density and the location of errors in these models. Acta Crystallogr A 1991;47:392–400.

16. McAfee JG, Edmondson SP, Datta PK, Shriver JW, Gupta R. Gene cloning, expression, and characterization of the Sac7d proteins from the hyperthermophile *Sulfolobus acidocaldarius*. Biochemistry 1995;34:10063–10077.

17. Kabsch W. Automatic indexing of rotation diffraction patterns. J Appl Cryst 1988;21:67–71.

18. Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.

19. Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33:491–497.

20. Ko TP, Chu HM, Chen CY, Chou CC, Wang AHJ. Crystal structures of the hyperthermophilic chromosomal protein Sac7d

in complex with DNA decamers. Acta Crystallogr D Biol Crystallogr 2004;60:1381–1387.

21. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.

22. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 1993;6:37–40.

23. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.

24. Merritt EA, Murphy MEP. Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr D Biol Crystallogr 1994;50:869–873.

25. Luscombe NM, Laskowski RA, Thornton JM. NUCPLOT: a program to generate schematic diagrams of protein-DNA interactions. Nucleic Acids Res 1997;25:4940–4945.

26. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res 2003;31:5108–5121.

27. Yang JM, Wang AHJ. Engineering a thermostable protein with two DNA-binding domains using the hyperthermopile protein Sac7d. J Biomol Struct Dyn 2004;21:513–526.