The Solution Structure of the Sac7d/DNA Complex: A Small-Angle X-ray Scattering Study†

Joanna K. Krueger,‡ Bradford S. McCrary,§ Andrew H.-J. Wang,|| John W. Shriver,§ Jill Trewhella,‡ and Stephen P. Edmondson*,§

Department of Biochemistry and Molecular Biology, School of Medicine, Southern Illinois University, Carbondale, Illinois 62901-4413, Chemical Science and Technology Division, Mail Stop G758, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and Department of Cell & Structural Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801

Received April 5, 1999; Revised Manuscript Received June 1, 1999

ABSTRACT: Small-angle X-ray scattering has been used to study the structure of the multimeric complexes that form between double-stranded DNA and the archaeal chromatin protein Sac7d from *Sulfolobus acidocaldarius*. Scattering data from complexes of Sac7d with a defined 32-mer oligonucleotide, with poly[d(GC)], and with *E. coli* DNA indicate that the protein binds along the surface of an extended DNA structure. Molecular models of fully saturated Sac7d/DNA complexes were constructed using constraints from crystal structure and solution binding data. Conformational space was searched systematically by varying the parameters of the models within the constrained set to find the best fits between the X-ray scattering data and simulated scattering curves. The best fits were obtained for models composed of repeating segments of B-DNA with sharp kinks at contiguous protein binding sites. The results are consistent with extrapolation of the X-ray crystal structure of a 1:1 Sac7d/octanucleotide complex [Robinson, H., et al. (1998) *Nature 392*, 202−205] to polymeric DNA. The DNA conformation in our multimeric Sac7d/DNA model has the base pairs tilted by about 35° and displaced 3 Å from the helix axis. There is a large roll between two base pairs at the protein-induced kink site, resulting in an overall bending angle of about 70° for Sac7d binding. Regularly repeating bends in the fully saturated complex result in a zigzag structure with negligible compaction of DNA. The Sac7d molecules in the model form a unique structure with two left-handed helical ribbons winding around the outside of the right-handed duplex DNA.

Bending and kinking of DNA play an important role in biology. Although some DNA sequences possess an intrinsic curvature, significant bending is normally associated with protein binding (*1*). A number of sequence-specific DNA binding proteins have been shown to induce bends or kinks of up to 65° in duplex DNA, and it is generally believed that such protein-induced bends are involved in the regulation of gene expression (*2*). Proteins that bind nonspecifically to DNA may also induce DNA bending and compaction, as in the nucleosome structure (*3*). In almost every instance studied to date, proteins that induce bending in DNA bind as oligomers.

Sac7d[1] is a small 7.6 kDa protein abundant in the thermophilic archaea *Sulfolobus acidocaldarius* (*4, 5*). The role of Sac7d in vivo is not known with certainty, but it is thought to be a chromatin protein that stabilizes the duplex form of DNA at the high growth temperature of *Sulfolobus* (about 80 °C) (*6*). DNA binding studies indicate that Sac7d binds noncooperatively to poly[d(GC)] with a binding constant of 6.5×10^6 M^{-1} in 50 mM KCl and a binding site size of 4 base pairs (*7*). The binding affinity shows only about a 10-fold variation for different DNA sequences. At the same time, circular dichroism data indicate that Sac7d binding induces a cooperative conformational change in DNA with the extent of cooperativity being sequence dependent (*7*). The binding of a single monomer appears to be sufficient to induce the conformational change in poly-[d(AG)]·poly[d(CT)], while poly[d(IC)] requires binding of adjacent Sac7d monomers.

The crystal structure of a 1:1 complex between Sac7d and an octanucleotide duplex has been solved to 1.6 Å resolution (*8*). The Sac7d in this complex is almost identical to that previously determined for the free protein in solution (*9*) with an overall root-mean-square deviation for the backbone atoms of 1.75 Å. The Sac7d binds to the minor groove of the octanucleotide with the side chains of Val 26 and Met 28 intercalating between two base pairs. The DNA is deformed at this site with one of the largest kinks (about 70°) observed to date in duplex DNA. Although Sac7d leads to a severe distortion of the DNA in the crystal structure of the 1:1 complex, it increases the T_m of duplex DNA in solution by more than 30 °C (*10*).

† This work was supported by the National Institutes of Health (GM 49686) and the APS Project of the Illinois Board of Higher Education (J.W.S. and S.P.E.). Part of this work was performed under the auspices of the Department of Energy under contract to the University of California, and was supported by DOE project KP1101010 (J.T.).

* Corresponding author. Phone: 618-453-6435. Fax: 618-453-6440. Email: sedmondson@som.siu.edu.

‡ Los Alamos National Laboratory.

§ Southern Illinois University.

|| University of Illinois at Urbana−Champaign.

[1] Abbreviations: Sac7d and Sso7d, 7 kDa DNA binding proteins from *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, respectively; CD, circular dichroism; NMR, nuclear magnetic resonance; NSLS, National Synchrotron Light Source.

10.1021/bi990782c CCC: $18.00
Published on Web 07/23/1999

The experiments reported here were aimed at obtaining structural information on multimeric complexes with polymeric DNA. Protein/DNA complexes with long DNA sequences have the potential for interactions that may be absent in the 1:1 crystal structure complex. Such interactions include protein−protein interactions or mutual reinforcement of DNA conformational changes due to contiguously bound protein molecules. These types of interactions could be important in packaging DNA in vivo. Steric hindrances from neighboring sequences also might inhibit large changes in the DNA structure such as the sharp kink that is observed in the crystal structure of the 1:1 Sac7d/DNA complex. This possibility was suggested by NMR data on a 2:1 complex of the homologous Sso7d protein (from *S. solfataricus*) bound to DNA that indicate a considerably less bent DNA structure than was observed in either the Sac7d/DNA or the Sso7d/DNA crystal structures (*11*).

We have used small-angle X-ray scattering to characterize the solution conformations of Sac7d complexed with a 32 base pair duplex oligonucleotide, poly[d(GC)], and *E. coli* DNA. Small-angle X-ray scattering is sensitive to the size and shape of macromolecules in solution and is particularly useful for obtaining information on conformations and higher order structures in DNA/protein complexes (*12*). We show that the Sac7d interactions with the longer oligonucleotides and polynucleotides are such that multimeric complexes are formed with many Sac7d proteins. These complexes have an overall rod-shaped structure with a highly extended DNA, and the Sac7d protein molecules pack around the DNA. The scattering data place constraints on the ways in which the protein monomers can be packed on the DNA, and a molecular model of an extended sequence of DNA saturated with Sac7d is derived. This model indicates that the Sac7d/DNA interaction and the bending of the DNA are consistent with the crystal of the Sac7d/octanucleotide complex.

MATERIALS AND METHODS

Sample Preparation. Recombinant Sac7d protein was overexpressed in *E. coli* and purified as previously described (*10*). Protein samples were dialyzed against 0.01 M KH_2PO_4, pH 7.0, and protein concentrations were determined by UV absorption using $\epsilon_{278} = 1.1$ mL/(mg·cm).

Three different types of DNA were used to prepare complexes with Sac7d for scattering studies: a synthetic 32-mer oligonucleotide of defined sequence, double-stranded poly[d(GC)] (Pharmacia) with an average length of 1300 base pairs, and *E. coli* DNA (Sigma) with an average length greater than 2000 base pairs. DNA samples were dialyzed against the same 0.01 M KH_2PO_4, pH 7.0, solutions used for the Sac7d protein. The complimentary strands of a synthetic 32-mer oligonucleotide sequence:

5′-d(CTGCGCAGCGCGCTGCGCACGCGCACGC-TGCG)-3′

3′-d(GACGCGTCGCGCGACGCGTGCGCGTGC-GACGC)-5′

were purchased from Genosys as HPLC-purified oligomers and dialyzed against 0.01 M KH_2PO_4 buffer, pH 7.0. This sequence was chosen based on its ability to form a unique double-stranded structure at 20 °C, and the slight preference of Sac7d for (G+C)-rich regions. Concentrations were determined from the A_{260} at 80 °C using single-stranded extinction coefficients determined from nearest-neighbor frequencies. The double-stranded oligomer was prepared by mixing equimolar amounts of the complimentary strands at 80 °C and annealed by cooling.

DNA/protein complexes were prepared by titrating the DNA (initial concentrations of 1−2 mg/mL) with protein to the desired protein:nucleotide ratio. The complexes were concentrated to about 15 mg/mL using Centricon 30 concentrators for the polynucleotides and Centricon 10 for the oligonucleotide. Less than 1% of the protein and DNA passed through the filters, indicating the formation of stable complexes under these conditions. No increase in light scattering was detected as a result of concentration, indicating there was minimal aggregation at these concentrations. CD spectra were independent of complex concentration over the measurable range (up to 3 mg/mL in a 0.01 cm path-length cell) and were identical within experimental error to those previously obtained for fully saturated complexes (*7*). The concentrations of the complex were determined from UV absorption spectra using extinction coefficients calculated from those of free protein and DNA.

X-ray Scattering Measurements. X-ray scattering data were collected at 17 °C using the small-angle instrument at Los Alamos National Laboratory that has been described previously (*13*). Preliminary scattering data were obtained at the X12B synchrotron beamline at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, using quartz capillary tubes with the 2D phase-sensitive detector placed 1.0 m from the sample. Scattering data were reduced to $I(Q)$ versus Q and analyzed as previously described (*13*). $I(Q)$ is the scattered X-ray intensity per unit solid angle. Q is the amplitude of the scattering vector and is given by $4\pi(\sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the scattered X-rays (1.54 Å). Scattering data were collected for Q values from 0.01 to 0.25 $Å^{-1}$ at Los Alamos and from 0.01 to 0.22 $Å^{-1}$ at the NSLS. The net scattering from the Sac7d/DNA complexes was determined by subtracting a normalized buffer spectrum measured in the same sample cell. Scattering data for each sample were collected at several concentrations of the protein/DNA complexes ranging from 5 to 20 mg/mL. The effects of interparticle interference due to electrostatic repulsion between molecules that can give rise to suppression of the lowest Q data were corrected by a linear least-squares extrapolation to zero concentration of the experimental scattering functions divided by the concentration (*14*).

An indirect Fourier transform analysis that uses a series expansion of sin x/x functions as described by Moore (*15*) was used to calculate the vector length distribution function, $P(r)$, from the scattering data for Sac7d/oligonucleotide complexes. The GNOM analysis of Svergun and colleagues (*16*) that uses an alternate linear regularization method was used to calculate $P(r)$ functions for the Sac7d/polymer complexes. This latter analysis is preferred for $P(r)$ calculations of complexes with very large dimensions as it is less prone to artifactual high-frequency "ripples" in the calculated $P(r)$ functions for these structures. Both $P(r)$ analyses included corrections for the slit geometry of the scattering instrument that has the effect of smearing the experimental scattering data. $P(r)$ is the frequency of vector lengths

connecting small-volume elements within the entire volume of the scattering particle, and hence goes to zero at the maximum dimension of the particle, d_{max}. The radius of gyration, R_g, can be calculated from the second moment of $P(r)$. Alternatively, R_g and the radius of gyration of cross section, R_c, can be calculated using the Guinier approximations to the lowest Q data for globular and rod-shaped particles, respectively (*17*, *18*). R_g is the root-mean-square distance of all elemental volumes from the center-of-mass of the particle, weighted by their scattering densities. R_c is the weighted, root-mean-square distance of all elemental areas from the center of the cross-sectional area of a rod-shaped particle. The scattering data were corrected for the slit-smearing geometry of the scattering station for all R_c analyses.

Model Calculations. An all-atom procedure was used to calculate $P(r)$ functions for molecular models of the protein/DNA complexes. Points were randomly selected within the van der Waals volume of pairs of randomly selected atoms, and the $P(r)$ distribution function was calculated by summing the distances between the selected points weighted by their atomic scattering power. Intensity curves, $I(Q)$, were calculated from the model $P(r)$ functions using the relation:

$$I(Q) = 4\pi \int P(r) \sin(Q \cdot r)/(Q \cdot r)\, dr \qquad (1)$$

To compare the scattering profiles calculated for the models with the experimental data acquired using the Los Alamos scattering instrument, which has a slit beam geometry, the model data were convoluted with the instrumental slit-smearing function. The NSLS instrument has an approximate point beam geometry, and hence such a correction was not necessary for those data. The experimental scattering data were scaled to the scattering profiles calculated for each model over a Q range of 0.01−0.2 Å^{-1} using a least-squares linear program based on singular value decomposition. The fit between model scattering profiles and experimental data was evaluated using reduced chi-square, χ^2_ν, values.

Molecular models were constructed using MidasPlus 2.0 (*19*) software on Silicon Graphics workstations. Energy minimization and the construction of DNA models were done using the AMBER package (*20*).

RESULTS

Small-Angle Scattering Data and Structural Parameters for the Sac7d/DNA Complexes. Figure 1 shows the scattering data measured for the specific 32-mer oligonucleotide (see Materials and Methods for sequence), poly[d(GC)] DNA, and *E. coli* DNA each saturated with bound Sac7d. The poly-[d(GC)] and *E. coli* DNAs are saturated with one Sac7d protein per four base pairs (*7*). In the case of the 32-mer, the DNA is saturated with an average of seven protein molecules per DNA molecule, presumably due to "end effects" with the short oligonucleotide. Under these conditions, the highly charged DNA is effectively neutralized by the protein as evidenced by the fact that the scattering data show only small concentration-dependent interparticle interference effects at the lowest Q values. These effects were removed by extrapolating the scattering data to infinite dilution (see Materials and Methods).

The $P(r)$ functions calculated from the scattering data for each protein/DNA complex are shown in Figure 2. The $P(r)$ functions are asymmetric with a peak at ~40 Å and a steadily decreasing $P(r)$ as they approached d_{max}. These rodlike characteristics of the $P(r)$ indicate that each complex has an elongated shape. The d_{max} value of the 32-mer complex is 120 Å, which is slightly longer than the value expected for a standard 32-mer B-form duplex DNA structure (~105 Å assuming the canonical 3.4 Å/bp). For the poly[d(GC)] and *E. coli* DNA complexes, d_{max} is ~450 Å. The $P(r)$ function for Sac7d/poly[d(GC)] decreases in an almost linear fashion with increasing vector lengths and as such is the closest to a true linear, rod-shaped particle. The $P(r)$ of the Sac7d/*E. coli* DNA complex exhibits more intermediate length vectors around 120 Å. These differences may reflect differences in the flexibility of these complexes or heterogeneity in the binding of Sac7d protein to *E. coli* DNA. Guinier plots of

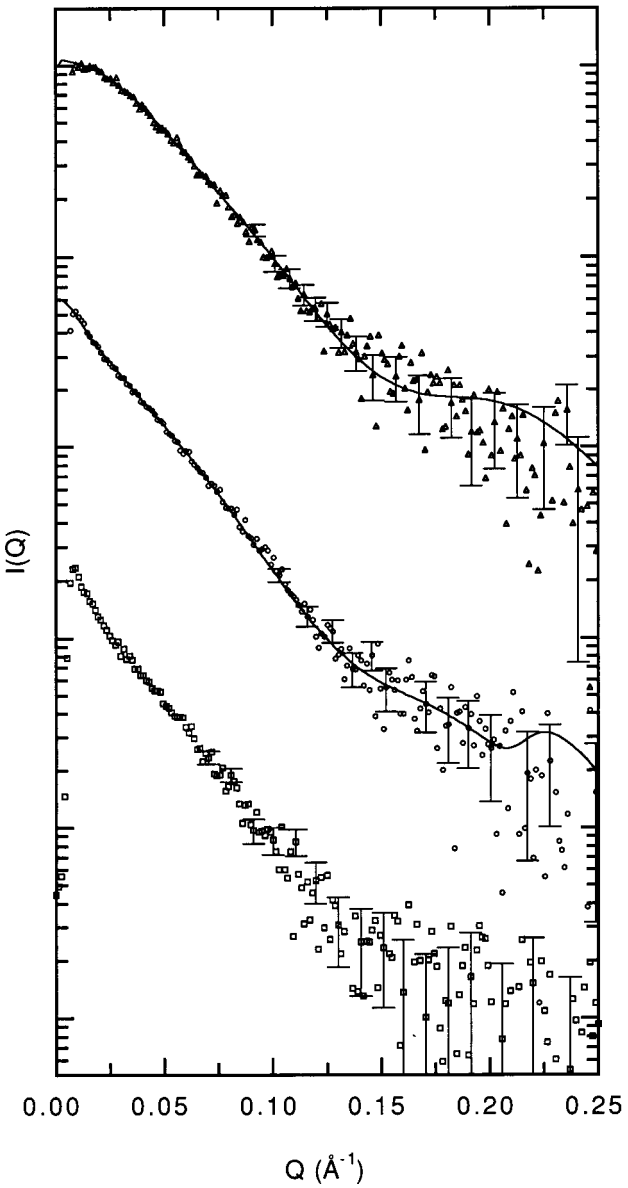

FIGURE 1: Scattering profiles of protein/DNA complexes at a ratio of ~4 base pairs per protein molecule for Sac7d/32-mer (upper curve), Sac7d/poly[d(GC)] (middle curve), and Sac7d/*E. coli* DNA (lower curve). The normalized, buffer-subtracted scattering data were extrapolated to zero concentration. Each scattering profile is displaced for clarity. The error bars represent ±1 SD. The solid lines represent the calculated scattering profiles for the models whose parameters are given in Table 2.

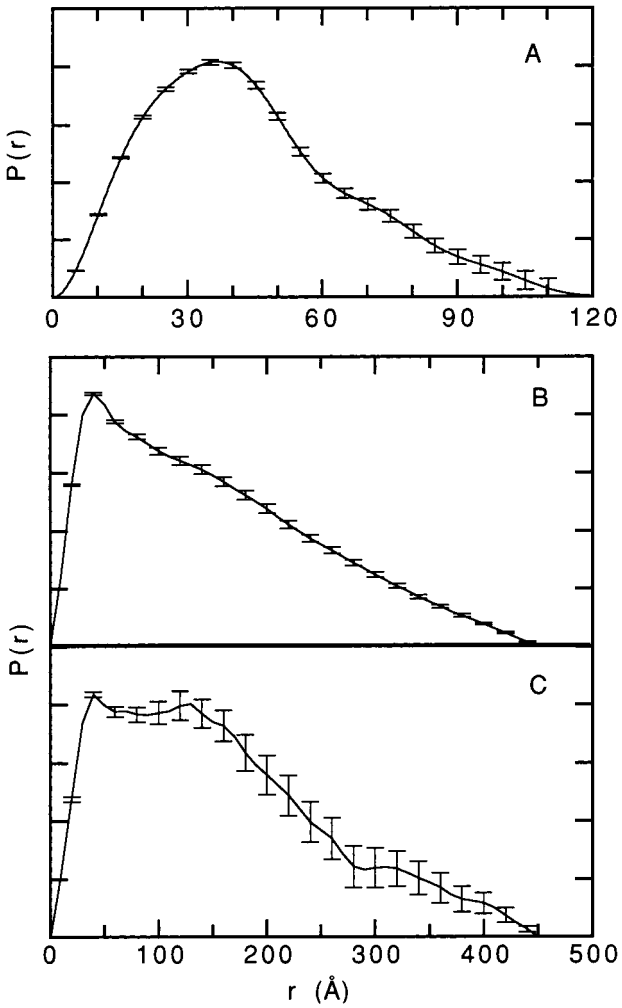

FIGURE 2: $P(r)$ distribution functions of (A) Sac7d/32-mer, (B) Sac7d/poly[d(GC)], and (C) Sac7d/*E. coli* DNA, calculated from the data in Figure 1. The error bars represent ±1 SD.

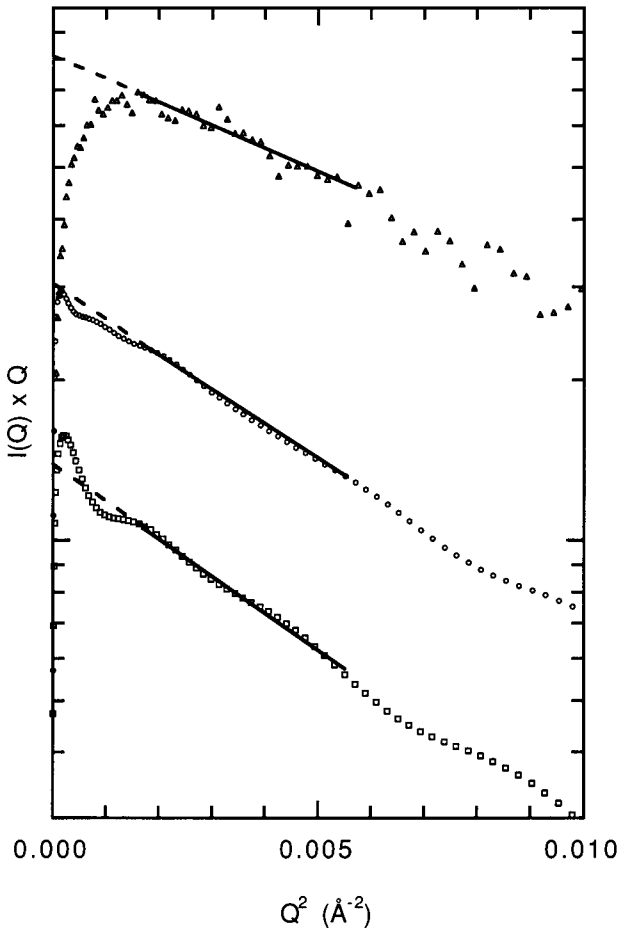

FIGURE 3: Guinier plots of the scattering data used for determining the R_c of rodlike molecules for Sac7d/32-mer (upper curve, raw data), Sac7d/poly[d(GC)] (middle curve, desmeared data), and Sac7d/*E. coli* DNA (lower curve, desmeared data). The solid lines indicate the fitted region, and the dashed lines represent extrapolation of the fits. Slit-smearing corrections, which increase with increasing R_g, were applied to the DNA polymer data but were insignificant for the Sac7d/32-mer complex.

the experimental scattering data for the Sac7d/DNA complexes, assuming rodlike behavior, are shown in Figure 3. They show the expected linear dependence for finite length rod-shaped particles in the Q-range ~0.04−0.07 Å^{-1} and were used to determine R_c values (Table 1).

Table 1 summarizes the structural parameters calculated from scattering data collected for the Sac7d/DNA complexes. In addition, parameters for the free 32-mer DNA and its 1:1 complex with Sac7d (0:32 and 1:32 Sac7d:bp ratio) are included for comparison. The R_g value of the 32-mer duplex DNA is 31.0 Å, with a radius of gyration of (cylindrical) cross section, R_c, of 7.4 Å and a d_{max} of ~100 Å. These values are consistent with those expected for an approximately rod-shaped B-form duplex DNA. Binding of a single Sac7d protein to the 32-mer had no effect (within experimental error) on the measured R_g or d_{max} values, but does increase R_c by about 20%. The R_g value for the fully saturated DNA/protein complex is about 10% greater than that of duplex DNA, while d_{max} increases to 120 Å, indicating that the protein/DNA complex remains extended and does not contract around a protein core as in chromatin. At the same time the R_c value more than doubles upon saturating the DNA with protein, consistent with packing of Sac7d onto an extended DNA structure. The cross-sectional radii of gyration (R_c) of the Sac7d/polymer complexes are 17.6−17.9 Å (Table 1), which is ~10% larger than that measured for the saturated Sac7d/32-mer complex. This relatively small increase may be due to increased flexibility in the longer DNA lengths giving rise to a larger average effective R_c.

If the polymeric DNA molecules adopted a fully extended B-form duplex DNA, they would give d_{max} values of >1000 Å, and we observe values significantly less than this for the complexes. The persistence lengths for the complexes are undoubtedly much less than the contour length of fully extended DNA, and the scattering experiments measure average structural parameters. Thus, we interpret the 450 Å dimension as an average length for rodlike segments in the overall structure. However, it should be noted that our minimum Q value of ~0.01 Å^{-1} does limit the vector lengths that can be determined accurately to a few hundred angstroms. It is therefore possible that the values determined for R_g and d_{max} are underestimated for these large DNA polymer complexes. The analysis of the cross-sectional parameters and the interpretation of the overall rod-shape, however, are unaffected by this low-Q limitation.

Modeling the DNA/Protein Complexes. The structural parameters determined from the scattering data indicate that Sac7d binds to the outer surface of an extended DNA

Table 1: Structural Parameters Determined from the Scattering Data for Sac7D/DNA Complexes[a]

complex	Sac7d:DNA (protein per bp)	R_c (Å)	D (Å)	R_g (Å)	d_{max} (Å)
Sac7d/32-mer	0:32	7.4 ± 0.1	20.9 ± 0.3	31.0 ± 0.3	100
	1:32	8.7 ± 0.1	24.6 ± 0.3	31.0 ± 0.3	100
	7:32	16.1 ± 0.2	45.5 ± 0.5	34.3 ± 0.2	120
Sac7d/poly[d(GC)]	1:4	17.6 ± 0.2	49.8 ± 0.6	129	450
Sac7d/*E. coli* DNA	1:4	17.9 ± 0.5	50.6 ± 1.4	131	450

[a] d_{max} and R_g were determined from indirect Fourier transform, $P(r)$, and Guinier analyses of the experimental scattering data. R_c was determined by linear regression analysis of $I(Q)\cdot Q$ versus Q^2 of the scattering data after correction for the slit beam geometry of the scattering instrument. The diameters were calculated from the relation $D = 2R_c\sqrt{2}$.

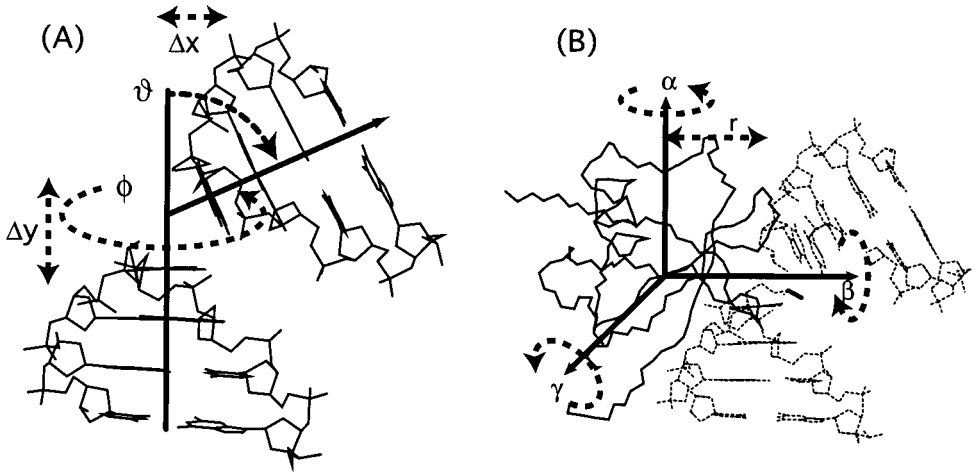

FIGURE 4: (A) Bent DNA was modeled as 4 base pair segments of B-DNA joined end-to-end with a longitudinal displacement Δy, a helical twist ϕ between segments, a radial displacement Δx along the dyad axis of the first base pair after the kink, and a bend ϑ about the guanine N4→cytosine O6 axis of the first base pair after the kink. (B) Sac7d was positioned in the minor groove of the bent DNA according to the X-ray crystal structure (*8*).

structure. We combined this constraint with high-resolution crystal structure and solution binding data to build and evaluate molecular models of the saturated Sac7d/DNA complexes. Model scattering profiles were evaluated against the scattering data in order to find a best-fit model that could provide further structural insights into the Sac7d/DNA complexes. Atomic coordinates of the Sac7d and DNA components were used to calculate the scattering, since they are available and are convenient for defining the molecular shapes to be modeled. Small-angle X-ray scattering does not depend, however, on the high-resolution atomic details of the model. Rather the scattering data are sensitive to the size and shape of the scattering particle, as well as to the relative dispositions of molecular domains within it. For a protein/DNA complex, information on the relative dispositions of the DNA and proteins is accentuated due to the more than 2-fold difference in X-ray scattering contrast between these components. [The scattering signal from a particle is directly proportional to the contrast, which is the square of the difference in the mean scattering density of the solute and solvent. DNA has a significantly greater contrast due to its large number of relatively electron-dense phosphate groups.] Therefore, the X-ray scattering data contain information on the organization of the packing of the protein onto the DNA, while the subtleties of the interactions and minor conformational adjustments at higher resolution within and between the DNA and protein structures require other techniques for analysis.

The complexes were modeled using sequential subunits, each consisting of a 4 base pair segment of canonical B-form DNA, separated by hinges with one Sac7d protein per segment oriented according to the X-ray crystal structure of the Sac7d/octanucleotide complex. The 4 base pair DNA binding site size was previously determined from fluorescence titration data (*7*). No attempt was made to model potential position-dependent conformational variations in the DNA. For simplicity, a $d(GC)_n$-$d(GC)_n$ sequence was used to model both the Sac7d/poly[d(GC)] and Sac7d/32-mer complexes. Conformational space was searched by positioning consecutive segments of B-form DNA end-to-end according to a longitudinal rise (Δy), a helical winding angle (ϕ) between segments, a radial displacement (Δx) along the dyad axis of the first base pair of the newly added segment (in the direction of the major groove), and a bend angle (ϑ) relative to the helix axis of the preceding DNA segment (see Figure 4A). The bending axis was about the guanine N4→cytosine O6 axis, which is approximately parallel to the C1′→C1′ axis. For reference, the values of these parameters for B-form DNA are $\Delta y = 3.4$ Å, $\Delta x = 0.0$ Å, $\phi = 36°$, and $\vartheta = 0°$. Sac7d proteins were placed identically in the minor groove of each 4 base pair segment (see Figure 4B) according to the X-ray crystal structure of Sac7d bound to $d(GTAATTAC)_2$ (PDB accession no. 1AZQ) (*8*). Since Sac7d does not bind symmetrically about the kink, the protein bound to each B-DNA segment interacts primarily with three of the base pairs of that segment plus one base pair of the subsequent DNA segment. The Sac7d/32-mer complex was modeled using 7 bound proteins, and Sac7d/poly[d(GC)] was modeled using 29 proteins bound to 120 base pairs of DNA.

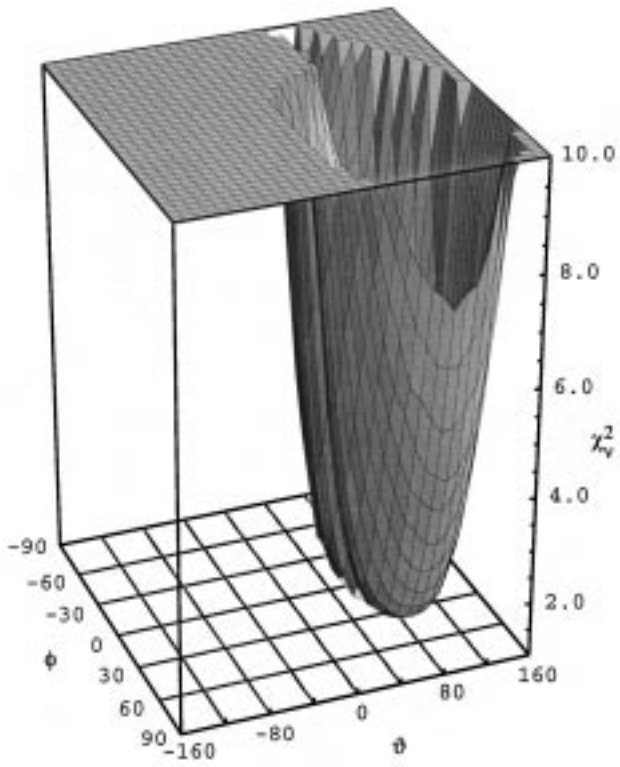

FIGURE 5: Reduced chi-square (χ^2_ν) surface of the fit of the Sac7d/32-mer complex as a function of the twisting angle ϕ and the bending angle ϑ.

The geometric parameters Δx, Δy, ϕ, and ϑ were systematically varied over parameter space using a grid search algorithm. The reduced chi-square, χ^2_ν, values for the fits between calculated and experimental X-ray scattering profiles were determined over a Q range of 0.01−0.20. The χ^2_ν surface of the fit for Sac7d/32-mer is shown in Figure 5 as a function of the helical twist and bending angles ϕ and ϑ. The surface has a well-defined minimum at $\phi = 40°$ and $\vartheta = 80°$, with relatively steep sides around the minimum. It is clear that models with nonperturbed, linear B-DNA ($\phi = 36°$, $\vartheta = 0°$) do not fit the scattering data and can thus be ruled out. Notably, the bottom of the well in the χ^2_ν surface is rather broad in the vicinity of the minimum, with the largest standard deviation being in the determination of the bending angle, ϑ, for the Sac7d/32-mer DNA of approximately ±45°. The χ^2_ν surface for the Sac7d/poly[d(GC)] fits (not shown) has a more narrow distribution about the minimum at $\phi = 30°$ and $\vartheta = 110°$, with a smaller standard deviation in ϑ of approximately ±35° (see Table 2). The fits of the calculated model scattering to the experimental data for Sac7d/32-mer and Sac7d/poly[d(GC)] complexes are shown as the solid lines in Figure 1.

The best-fit models for the Sac7d/32-mer and Sac7d/poly[d(GC)] complexes have bends between B-DNA segments of about 80−110° into the major groove. The helical twist between base pairs is about 30−40° at the site of the DNA kink, close to the value of 36° for canonical B-DNA. The geometric parameters derived for the Sac7d/32-mer and Sac7d/poly[d(GC)] structures are the same within error. This result was confirmed by fitting simultaneously the scattering data for both complexes using identical parameters for both data sets. The helical twist and bending angles that give the best fit simultaneously to both Sac7d/DNA complexes were

Table 2: Fitted Model Parameters[a]

	Δy (Å)	Δx (Å)	ϕ (deg)	ϑ (deg)	χ^2_ν
Sac7d/32-mer	3.3 ± 0.5	3.3 ± 0.8	40 ± 16	80 ± 45	1.07
Sac7d/poly[d(GC)]	3.3 ± 0.5	3.5 ± 0.5	30 ± 12	110 ± 35	1.26
global fit[b]	3.3	3.4	35 ± 12	100 ± 35	1.17

[a] The geometric parameters Δy, Δx, ϕ, and ϑ were systematically varied over ranges of 1.8−3.8 Å in increments of 1.0 Å, 1.0−4.0 Å in increments of 1.0 Å, −90 to 90° in increments of 10°, and −160 to 160° in increments of 10°, respectively. Mean values for Δy and Δx were derived by averaging over solutions with $\chi^2_\nu < 1.5$, and the scattering data were fit again using these constrained Δy and Δx values. The standard deviations in ϕ and ϑ were determined from fits with $\chi^2_\nu < [\chi^2_\nu(\text{min}) + 1]$ (*26*). The fit was over the Q range of 0.01−0.20. [b] The scattering data for both Sac7d/32-mer and Sac7d/poly[d(GC)] were fit globally by minimizing the total χ^2_ν over ϕ and ϑ parameter space with fixed Δy and Δx values.

$\phi = 35°$ (±12°) and $\vartheta = 100°$ (±35°), respectively (Table 2). In fact, the differences in the scattering profiles of the Sac7d/32-mer and Sac7d/poly[d(GC)] complexes are primarily at small Q and can be accounted for entirely by differences in the lengths of the complexes. The $I(Q)$ of the oligonucleotide complex is depressed at low Q ("roll-over"), consistent with the finite length of the Sac7d/32-mer complex (*21*), whereas little roll-over is detected for the much longer Sac7d/poly[d(GC)]. Calculations on different lengths of duplex DNA indicate little change in $I(Q)$ for Q greater than 0.01 in the scattering curves of complexes consisting of more than 125 base pairs.

The DNA component of the Sac7d/DNA complex was modeled using linear segments of B-form DNA with variable bends between each segment, as described above. This approach was suggested by the X-ray crystal structure of a Sac7d/octanuceotide complex which showed a sharp kink between two DNA base pairs. Because of the abrupt bends in the DNA used for fitting the scattering data, the DNA of the fitted models is in a highly strained and unrealistic conformation. Therefore, the DNA component (50 base pairs) of the best-fit Sac7d/poly[d(GC)] model was energy minimized in the absence of protein (with no constraints) to obtain acceptable bond lengths and dihedral angles. The calculated scattering curve of the complex was unaffected by this procedure since X-ray scattering depends on the distribution of protein and DNA scattering domains, which did not significantly change during minimization, and not the atomic details. Although the X-ray scattering data do not give direct information on the local structure of the DNA (or the protein for that matter), it is still reasonable to determine the basic helical parameters that provide the appropriate disposition of molecular scattering units.

The potential energy of the DNA in the Sac7d-bound conformation was −2050 kcal/mol after energy minimization in the absence of protein. This is slightly less than that of canonical B-DNA (−1950 kcal/mol after an identical energy minimization), indicating that the types of kinks induced in DNA by Sac7d binding are stereochemically reasonable. The base pair tilt, measured as the angle between the long axis and the normal of the C1′−C1′ vector, was 35° ± 4° averaged over the 4 base pairs of the central Sac7d binding site. The base pairs were displaced by $D = 3$ Å from the long axis. The sugar puckers retained their original C3′-endo conformation during minimization.

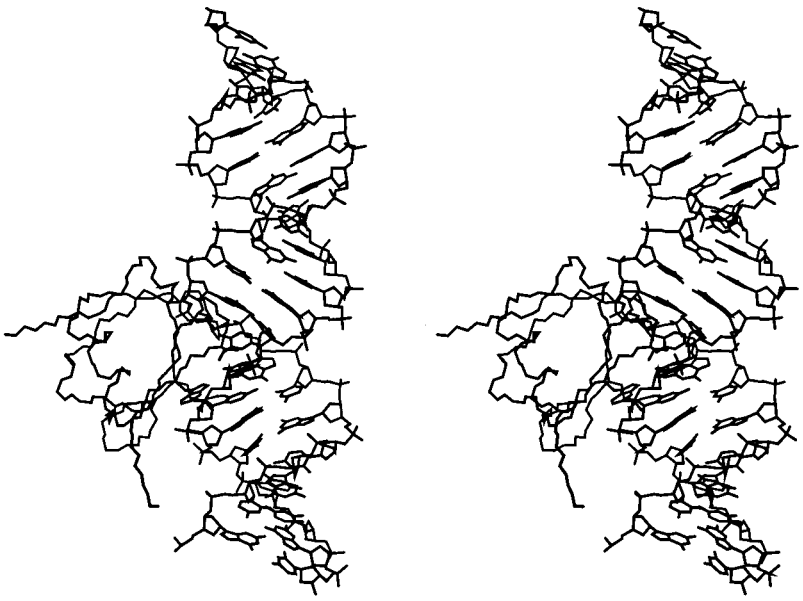

FIGURE 6: Stereo (walleye) diagram showing the structure of the DNA in the fully bound Sac7d/DNA complex after energy minimization of the DNA component. The peptide backbone of a single Sac7d protein (with Trp24, Val26, Met29, and Arg42 side chains) is shown bound on the DNA to illustrate the interactions between Sac7d and DNA.

As expected, energy minimization resulted in more moderate bending angles as the distortions at the kink sites are allowed to influence neighboring base pairs. The resulting roll angle between the base pair planes at the kink site was 62° (compared to the original 100° roll in Table 2), and the roll at adjoining base pairs averaged 10° ± 4°. To quantitate the degree of "bending" in the energy-minimized Sac7d-bound DNA structure, an 8 base pair fragment of DNA was considered in which dinucleotide 1 consists of base pairs 1 and 2, dinucleotide 2 consists of base pairs 2 and 3, etc., and the kink site is at dinucleotide 4 (between base pairs 4 and 5). The local helix axis of each dinucleotide was defined as a vector connecting the center-of-mass of each of its base planes. The overall bend, measured as the angle between helix axes on opposite sides of the kink and averaged over all 9 such dinucleotide pairs, was 70° ± 7°. Although energy minimization decreases the apparent bends used in the modeling simulations from 100° to about 70°, the shapes and disposition of the protein and DNA components remain the same, and there is no effect on the calculated X-ray scattering.

The resulting energy-minimized DNA structure is shown in Figure 6, with a single Sac7d molecule placed at a protein binding site. Regularly repeating kinks every 4 base pairs manifest themselves as an apparent tilting of the bases rather than a noticeable supercoiling of the DNA. The zigzag structure results in little change in the overall linear topology of DNA in the fully saturated complex.

A model of the oligomeric Sac7d/DNA complex based on the best-fits to the scattering data is shown in Figure 7. The Sac7d protein is bound on the outside of the DNA and forms a helical structure with approximately 7 proteins in 3 turns and a pitch of about 95 Å. There are 28 base pairs of DNA within this same distance. Since the DNA binding site size is 4 base pairs, consecutive proteins along a fully saturated DNA sequence lie on opposite sides of the DNA. A consequence of this arrangement is that neighboring proteins are not sequential along the DNA but are at binding sites i and $i+2$. Furthermore, the basic 7_3 helical symmetry of the complex is overshadowed by higher-order symmetry, and the Sac7d appears to form a left-handed double-helical ribbon winding around the right-handed DNA duplex. This type of helical symmetry has not been previously observed in DNA/protein complexes.

DISCUSSION

The scattering data place significant constraints on the structure of the Sac7d complexes with polymeric DNA. The modeling of the scattering data used an extended Q-range ($0.01-0.2$ Å^{-1}), and the χ_ν^2 values showed significant variations over the conformational space searched, indicating that the data are sensitive to the parameters being varied. Within the constraints defined, the parameter search was exhaustive, and a single broad minimum was obtained that included a set of conformationally similar structures. In addition to the scattering data, constraints included the Sac7d protein structure and its orientation with respect to DNA, which were taken from the crystal structure of the 1:1 complex *(8)*. The binding site size was constrained to that determined by DNA binding studies monitored by fluorescence quenching *(7)*. The DNA structure was based on idealized B-form DNA, with rigid segments of DNA linked by variable kink and twist angles which were subsequently conformationally relaxed via an energy minimization step. The model presented here thus combines all the existing structural constraints for the complex and provides important insights into thermophile protein/DNA interactions.

The best-fit models of the Sac7d/DNA complexes have large kinks in the DNA. Importantly, the basic parameters determined from the modeling were the same for the 32-mer DNA and for the poly[d(GC)] DNA. After relaxing the requirement for strict B-DNA conformation by performing an energy minimization on the DNA, the bend angle at the Sac7d-induced kink site was about 70°, which is the same as that observed in the crystal structure (*8*). The scattering data and modeling therefore indicate that the sharp kink of

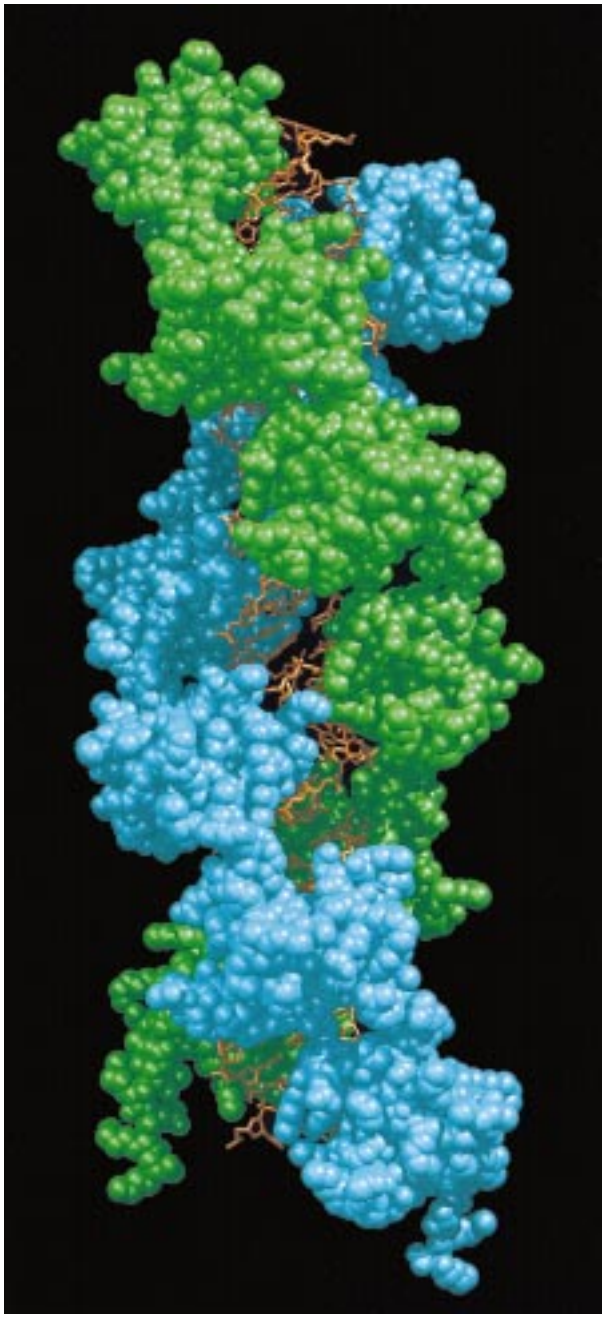

FIGURE 7: Model of the Sac7d/DNA complex showing 12 proteins bound to 50 base pairs of duplex DNA. The model is based on a global fit to the scattering data with parameters $\phi = 35°$ and $\vartheta = 100°$. Sac7d is shown in a space-filling representation, and DNA is a copper wireframe. Alternating proteins along the DNA sequence are colored blue and green, respectively, to emphasize the left-handed double-helical symmetry of the Sac7d protein in the complex.

~70° observed in the crystal structure of the 1:1 complex of Sac7d with an oligonucleotide is not an artifact of the crystal packing forces and is maintained in solution, even for multimeric complexes that form with polymeric DNA.

Though binding of Sac7d leads to a sharp kink in the DNA, sequential binding of the protein to long lengths of DNA at contiguous four base pair sites leads to little change in the overall topology of the DNA because the sequential bends are compensatory. An NMR structure of the homologous Sso7d protein bound to the minor groove of a DNA dodecamer shows a similar compensating effect in a 2:1 protein/DNA complex. However, the DNA bending angle in the NMR strucuture (*11*) was significantly smaller (~30°) than the bends observed in the Sso7d/DNA or Sac7d/DNA crystals (*22*, *8*) and from that obtained by fitting the X-ray scattering data presented here. The NMR structure also differs from the X-ray structures in that there is no apparent intercalation of protein side chains between the DNA bases in the NMR structure. These differences may be attributed to the NMR data being collected under conditions of fast exchange with excess DNA. Because the dodecamer was capable of binding two proteins, there was a mixture of unligated, singly- and doubly ligated DNA under the conditions of the NMR experiment. It would appear that the NMR structure therefore represents an average of at least three species.

Interestingly, the DNA in the complex superficially resembles A-form DNA with the bases tilted by about 35° and displaced 3 Å from the long axis, and the widths of the major and minor grooves being almost equal. The resemblance to A-form DNA may explain the large change observed in the CD of DNA upon Sac7d binding. Sac7d induces an increase in the long-wavelength CD band of DNA (*7*) not too unlike that observed for the B to A conformational transition (*23*).

Electron microscopy of both Sac7d/DNA and the homologous Sso7d/DNA complexes shows highly condensed compact clusters, presumably of protein and DNA with loops of free DNA projecting outward (*24*, *25*). An ability to form clusters and loops could be interpreted in terms of Sac7d possibly containing multiple DNA binding sites per molecule, or existing as higher order oligomers through protein−protein interactions. Neither of these options is supported by the solution scattering data, the DNA binding data, or the X-ray crystal structure. The lack of cooperativity in DNA binding indicates that the protein binds as a monomer, and the X-ray structure indicates that there is only one binding site per protein. The scattering data rule out the formation of compact clusters in pure solutions of DNA saturated with Sac7d. The clusters observed by electron microscopy may therefore be an artifact of specimen preparation or due to the presence of other chromosomal proteins that influence the packing.

Sac7d binding results in negligible compaction of DNA. The R_c, R_g, and $P(r)$ functions for the Sac7d/DNA complexes, as well as the modeling, show that when Sac7d binds to DNA it forms a rod-shaped structure in which the protein monomers pack around an extended DNA. In the case of the complex with the defined-length 32-mer oligonucleotide, the rod-shaped structure is only slightly longer than an ideal B-form DNA duplex with the same number of base pairs. Thus, the Sac7d protein is quite distinct from the histone proteins of the nucleosome particle of chromatin which form a core about which the DNA wraps itself. Compaction of the *Sulfolobus* chromosome must be accomplished with other proteins, such as the 10 kDa DNA-binding protein that has been reported to form regular compact structures with DNA (*24*). The function of Sac7d may then be primarily to protect DNA from degradation and thermal denaturation at the high growth temperature of *Sulfolobus*.

REFERENCES

1. Travers, A. A. (1995) in *DNA−Protein: Structural Interactions* (Lilley, D. M. J., Ed.) pp 49−75, IRL Press, New York.

2. Werner, M., Gronenborn, A. M., and Clore, G. M. (1996) *Science 271*, 778−784.
3. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature 389*, 251−258.
4. Kimura, M., Kimura, J., Davie, P., Reinhardt, R., and Dijk, J. (1984) *FEBS Lett. 176*, 176−178.
5. Choli, T., Wittmann-Liebold, B., and Reinhardt, R. (1988) *J. Biol. Chem. 263*, 7087−7093.
6. Dijk, J., and Reinhardt, R. (1986) in *Bacterial Chromatin* (Gualerzi, C., & Pon, C., Eds.), Springer-Verlag, Berlin.
7. McAfee, J. G., Edmondson, S. P., Zegar, I., and Shriver, J. W. (1996) *Biochemistry 35*, 4034−4045.
8. Robinson, H., Gao, Y.-G., McCrary, B. S., Edmondson, S. P., Shriver, J. W., and Wang, A. H.-J. (1998) *Nature 392*, 202−205.
9. Edmondson, S. P., Qiu, L., and Shriver, J. W. (1995) *Biochemistry 34*, 13289−13304.
10. McAfee, J. G., Edmondson, S. P., Datta, P. K., Shriver, J. W., and Gupta, R. (1995) *Biochemistry 34*, 10063−10077.
11. Agback, P., Baumann, H., Knapp, S., Ladenstein, R., and Härd, T. (1998) *Nat. Struct. Biol. 5*, 579−584.
12. Olah, G. A., Gray, D. M., Gray, C. W., Kergil, D. L., Sosnick, T. R., Mark, B. L., Vaughan, M. R., and Trewhella, J. (1995) *J. Mol. Biol. 249*, 576−594.
13. Heidorn, D. B., and Trewhella, J. (1988) *Biochemistry 27*, 909−915.
14. Pilz, I. (1982) *Small-Angle X-ray Scattering*, Academic Press, New York.
15. Moore, P. B. (1980) *J. Appl. Crystallogr. 13*, 168−175.
16. Svergun, D. I., Semenyuk, A. V., and Feigin, L. A. (1988) *Acta Crystallogr. A44*, 244−250.
17. Guinier, A. (1939) *Ann. Phys. (Paris) 12*, 161−237.
18. Sosnick, T. R., Charles, S., Stubbs, G., Yau, P., Bradbury, E. M., Timmins, P., and Trewhella, J. (1991) *Biophys. J. 60*, 1178−1189.
19. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.
20. Pearlman, D. A., Case, D. A., Caldwall, J. W., Ross, W. S., Cheatham, T. E., Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1995) University of California, San Francisco.
21. Hjelm, R. P. (1985) *J. Appl. Crystallog. 18*, 452−460.
22. Gao, Y.-G., Su, S.-Y., Robinson, H., Padmanabhan, S., Lim, L., Edmondson, S. P., Shriver, J. W., and Wang, A. H.-J. (1998) *Nat. Struct. Biol. 5*, 782−786.
23. Sprecher, C. A., Baase, W. A., and Johnson, W. C., Jr. (1979) *Biopolymers 18*, 1009−1019.
24. Lurz, R., Grote, M., Dijk, J., Reinhardt, R., and Dobrinski, B. (1986) *EMBO J. 5*, 3715−21.
25. Choli, T., Henning, P., Wittmann-Liebold, B., and Reinhardt, R. (1988) *Biochim. Biophys. Acta 950*, 193−203.
26. Bevington, P. R., and Robinson, D. K. (1992) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, Inc., New York.

BI990782C