Modeling Helix-Turn-Helix Protein-Induced DNA Bending with Knowledge-Based Distance Restraints

2 **AUTHORS**, INCLUDING:

Wen-Shyong Tzou

National Taiwan Ocean University

53 PUBLICATIONS   **436** CITATIONS

Modeling Helix-Turn-Helix Protein-Induced DNA Bending with Knowledge-Based Distance Restraints

Wen-Shyong Tzou and Ming-Jing Hwang
Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan, ROC

ABSTRACT   A crucial element of many gene functions is protein-induced DNA bending. Computer-generated models of such bending have generally been derived by using a presumed bending angle for DNA. Here we describe a knowledge-based docking strategy for modeling the structure of bent DNA recognized by a major groove-inserting α-helix of proteins with a helix-turn-helix (HTH) motif. The method encompasses a series of molecular mechanics and dynamics simulations and incorporates two experimentally derived distance restraints: one between the recognition helix and DNA, the other between respective sites of protein and DNA involved in chemical modification-enabled nuclease scissions. During simulation, a DNA initially placed at a distance was "steered" by these restraints to dock with the binding protein and bends. Three prototype systems of dimerized HTH DNA binding were examined: the catabolite gene activator protein (CAP), the phage 434 repressor (Rep), and the factor for inversion stimulation (Fis). For CAP-DNA and Rep-DNA, the root mean square differences between model and x-ray structures in nonhydrogen atoms of the DNA core domain were 2.5 Å and 1.6 Å, respectively. An experimental structure of Fis-DNA is not yet available, but the predicted asymmetrical bending and the bending angle agree with results from a recent biochemical analysis.

INTRODUCTION

A "stereospecific" nucleoprotein complex composed of proteins and bent DNA is usually required for DNA replication, transcription, and recombination. In this complex, not only does DNA bend in the right direction (protein-induced DNA bending); the right sequence of DNA is recognized by the binding protein as well (DNA-binding specificity). Our knowledge about these two aspects of protein-DNA association has been greatly enhanced by analyses of several high-resolution structures of such association (reviewed by, for example, Harrison and Aggarwal, 1990; Pabo and Sauer, 1992; Travers, 1992). However, although structures resolved at the atomic level of many DNA-binding proteins alone are available, their DNA-bound complexes are not; hence computer-generated models for the latter have often been relied upon to interpret biochemical data and to suggest further experiments. Such modeling constitutes a molecular docking problem whose solution has been extensively researched for protein-drug and protein-protein associations; in contrast, few studies have attempted to dock a DNA substrate to binding proteins (Lengauer and Rarey, 1996; Sternberg et al., 1998).

Structure modeling of protein-bent DNA complexes dates back more than 10 years in studies of l cro (Matthew and Ohlendorf, 1985) and catabolite gene activator protein (CAP) (Weber and Steitz, 1984; Warwicker et al., 1987) systems. In those studies a pre-bent DNA was docked to the binding protein by interactive molecular graphics, and electrostatic complementarity was considered. Recently, Sandmann et al. (1996) described a systematic method for docking a pre-bent DNA to twofold symmetrized proteins of a HTH (helix-turn-helix) motif. In this method, the most optimal structure was searched by varying relative distances and orientation angles between protein and a pre-bent DNA through an energy minimization procedure assisted by solving a linearized Poisson-Boltzmann equation. The method was tested on the known complex structure of the CAP and the phage 434 cro system and then used to predict the unknown structure of the Fis-enhancer complex. Another protein-DNA docking modeling was a Monte Carlo simulation study by Knegtel et al. (1994a,b), who allowed the flexibility of DNA and incorporated experimentally determined contacts as energy bonuses in the simulation. This method was successfully applied to the complex structure of a 434 cro, lac, and gal repressor headpiece bound with a half-site of their respective DNA substrate, but suffered from the limitation that the initial position of DNA could not deviate much from its x-ray-determined protein-bound structure (no more than 1 or 2 bp up or down the "correct" basepair position), or DNA would start to "curl around the protein." More recently, Aloy et al. (1998) demonstrated that a global search method starting from the coordinates of unbound protein and a standardized B-DNA model is capable of producing native-like complex structures of several repressor systems, especially when some experimentally derived distance restraints were incorporated.

We present here a molecular simulation protocol for docking a straight DNA, i.e., without pre-bending, to a dimerized HTH protein from a distant position. The docking was "steered" by two types of distance restraint derived from experimental data from the region of the HTH binding and from sites of nuclease scission. The crystal structures of CAP-DNA and Rep-DNA were used as tests of the simu-

---

lation method, and the unknown complex structure of Fis-DNA was subsequently predicted. The simulation results were evaluated by examining the effects of the restraints imposed on the predicted models, and by comparison to models of previous predictions. In addition, differences in the DNA bending modes of the three systems and the organization of their dimerized pair of recognition helix were analyzed, from which a structural role of conserved TG dinucleotide steps is proposed.

MATERIALS AND METHODS

All of the calculations and molecular structure manipulations, unless otherwise noted, were carried out with the Discover/InsightII molecular simulation and modeling package from Molecular Simulation (MSI 95 release; San Diego, CA) on various models of Silicon Graphics workstations (SGI, Mountain View, CA).

Initial structures for protein and DNA

X-ray-determined protein structures were extracted from their DNA-bound form for CAP and Rep (PDB entry code 1cgp for CAP, Schultz et al., 1991; 2or1 for Rep, Aggarwal et al., 1988). For Fis (PDB 1fia, Kostrewa et al., 1991, 1992), the predicted flap-hinge model (Tzou and Hwang, 1997) for its N-terminal domain was adopted to make a complete protein structure. In this model, the N-terminal domain, which was not resolved crystallographically for the wild type (Kostrewa et al., 1991, 1992; Yuan et al., 1991), protrudes away from the C-terminal DNA-binding domain, and therefore should not have consequences for the present simulations (the predicted protruding of a hairpin Fis N-terminus was observed in a mutant structure (Safo et al., 1997) solved after the completion of the present work). For DNA, canonical B-form structures were constructed in InsightII, setting parameters slide and shift to 0 Å, rise to 3.4 Å, twist angle to 36°, and roll and tilt angle to 0°. Analyzed by program CURVES 5.1 (Lavery and Sklenar, 1988, 1989), the constructed B-DNA yielded the following parameters for Watson-Crick base pairs: opening −4°; propeller twist 4°; buckle 0°; shear, stretch, and stagger all 0 Å; and for DNA backbone torsions: α 313°, β 214°, γ 36°, δ 156°, ϵ 155°, ζ 265°, χ 262°, and pseudorotation phase angle P 192° (for definitions of this nucleic acid nomenclature, see Saenger, 1984; Diekmann, 1989). The derived B-DNA has its phosphate backbone in BI and its sugar pucker in C2′-endo conformation. All hydrogens were added according to the standard assigning scheme of InsightII. The binding nucleotide sequences of CAP and Rep were chosen to be those of the crystal structures. The Rep DNA sequence was extended by 4 bp at one end and 5 bp at the other, using the O_R1 sequence of phage 434 (Bushman, 1993). This extension consequently enclosed the Rep DNA sequence of the x-ray structure into a core and placed the geometric center of the DNA on its dyad axis, thereby allowing balanced restraint forces to be applied on the two half-sites (see below). For the DNA substrate of Fis, we used the distal Fis-binding domain of *hin*, denoted as *hin-D* (Bruist et al., 1987). The lengths of the three DNA fragments simulated are roughly the same: 30 bp for CAP, 28 bp for Rep, and 29 bp for Fis. All three DNA fragments were simulated as continuous, although the DNA in the crystal structure of the CAP-DNA complex is nicked (Schultz et al., 1991; Parkinson et al., 1996).

Initial configurations for each of the three systems were generated by the following procedures:

1. Both DNA and protein were superimposed on their geometric centers, twofold rotation axes, and long axes (the axes were calculated from these molecules' principal moments of inertia). A Cartesian coordinate system was defined such that the twofold axis of protein is the z axis, the long axis of protein is the x axis, and the cross-product of z and x is the y axis (Fig. 1). During the axial alignment, the minor groove at the pseudodyad was chosen to face the binding protein.

2. DNA was translationally moved from protein along the superimposed twofold axes (z axis) in the direction of minus z by 140 Å for CAP, 100 Å for Rep, and 120 Å for Fis (the differential distances took into account the protein size, although for such a large separation, slight variation in the exact value should not affect the simulation results).

3. At this configuration, DNA is regarded as being in the 0° reference state (i.e., the long axis of DNA and that of protein, or the x axis, are parallel). For Fis, its DNA was further rotated 60° about the twofold axis (z axis) such that in the docking simulation DNA would approach the protein's recognition helix (RH) from the C-terminal end. Test runs showed that for Fis, but not for CAP or Rep, DNA approaching RH from the N-terminal resulted in unsatisfactory results, most likely due to the much larger inclination angle of its RH with respect to the x-y plane (~29° for Fis versus ~10° for CAP and ~13° for Rep). As an illustration, the initial configuration for the CAP system is depicted in Fig. 1.

Definition of protein-DNA interaction sites and distance restraints

To guide the initially distant and unbent DNA into docking with and bending toward the binding protein in the simulation, we relied on "steering forces" provided by knowledge-based distance restraints made on two protein-DNA contact sites, i.e., that between the RH of protein and its receiving major groove (MG) of DNA, and that between a protein residue whose chemical modification possesses nuclease activity and the cleavage site of DNA. The two sites are denoted by PC (protein nuclease cleaving residue) and NC (nucleotide cleavage site), respectively. The locations of these "interaction" sites, shown in Fig. 1 and Fig. 2, *A* and *B*, and their distances were derived based on experimental data, as described below.

1. CAP, Rep, and Fis belong to a well-characterized DNA-binding protein family consisting of an HTH motif that interacts with DNA's major groove. Structural analysis has revealed that six amino acid residues on the second helix of this motif are most important for such protein-DNA recognition (Suzuki and Gerstein, 1995). The geometric center of the Cα atoms of these six amino acid residues was therefore selected as the "interaction" site of RH (Figs. 1 and 2 *A*).

2. Chemical interference experiments have shown that methylation of particular bases or ethylation of particular phosphates can abolish protein-DNA binding (Steitz et al., 1983; Bushman et al., 1985; Bruist et al., 1987). The derived chemical modification sites are consistent with the protein-DNA contacts observed in crystallographic structures (Bushman et al., 1985). As shown in Fig. 2 *B*, the sites derived by interference experiments were used to define the "interaction" site for the MG of DNA, with which the binding protein contacts directly. A segment of 10 bp centered on the defined "interaction" site was then selected to represent the MG.

3. Some DNA-binding proteins can be transformed into site-specific nucleases when specific amino acid residues are mutated into cysteines and then conjugated with synthetic nucleolytic agents (Pan et al., 1994b; Bastia, 1996). The modified protein cleaves DNA through an oxidative attack at C-1′H of particular nucleotides. The DNA scission reaction thus provides information about the proximity between the mutated protein residue and the nicked DNA site. Consequently, as shown in Fig. 1 and Fig. 2, *A* and *B*, we defined Cα of the mutated protein residue as the interaction site for the protein (PC), and C1′ of the cleaved nucleotide as the interaction site for DNA (NC).

Identification of these protein-DNA "interaction" sites yields two distance data per half-site: RH-MG and PC-NC. The former has been shown to be in the range of 8–10 Å for a variety of protein-DNA complex structures involving an RH (Suzuki and Gerstein, 1995). We used an upper limit of 11 Å to restrain this distance in the simulation. The latter can be estimated based on the extended arm length of the nucleolytic agents. One such agent for CAP is acetylglycylamino-1,10-phenanthroline-copper (Pendergrast et al., 1994), and one for Fis is acetamido-1,10-phenanthroline-copper (Pan et al., 1994a). These agents have been estimated to have an arm length of 17 Å and 13 Å, respectively. These values were then chosen to be the upper limit of PC-NC restraint for their respective systems. In addition, runs with a different limit of PC-NC distance were

FIGURE 1 The initial protein-DNA configuration for the docking simulation, as illustrated by the CAP-DNA system. The recognition α-helix (RH) and a segment of 10 bp encompassing the major groove (MG) that receives RH are boxed. The "interaction" sites derived from nuclease scission experiments are represented by filled balls. Arrows indicate that in the simulation DNA was drawn toward protein by forces of distance restraints imposed on these "interaction" sites. Axes X, Y (not shown), and Z are defined in Materials and Methods.
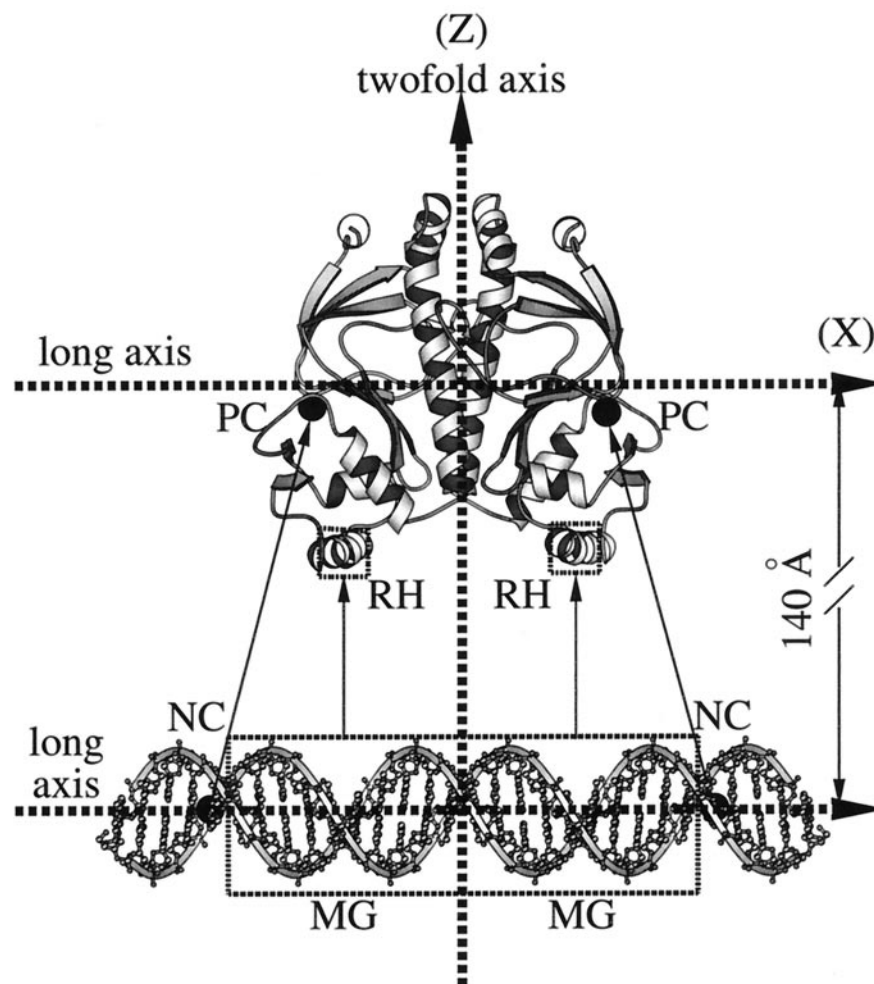
carried out to investigate the effect of this restraint on the simulation results. No PC-NC restraint was applied to Rep because no DNA scission studies on regions outside the protein-binding core sequences are available.

B-DNA restraints

Restraints on the interproton distances deduced from a canonical B-DNA (chosen for its relevance in biology) were imposed by a flat-bottom quadratic penalty function to be within ± 0.1 Å of all distances less than 5 Å. This allowed flexibility for DNA to move and bend while maintaining its basic B-DNA conformation. In all, there were 3959, 3735, and 3840 interproton distances imposed for the DNA of CAP, Rep, and Fis, respectively. To investigate the effect of the B-DNA restraints on the final structure of the simulation, we carried out additional simulations in which the number of interproton distance restraints was systematically reduced. From these simulations, we were able to reduce the number by about two-thirds (from 3959 to 1342 for CAP, 3735 to 1256 for Rep, and 3840 to 1300 for Fis) without significantly degrading the simulation results (see Results). This reduction brought the number of interproton distance restraints from $\sim$66 per nucleotide to $\sim$22 per nucleotide for the three systems. The reduced number is compatible with those used to determine solution DNA structures by nuclear magnetic resonance methods (e.g., $\sim$27 per nucleotide in Weisz et al., 1994; $\sim$24 per nucleotide in Chou et al., 1996; $\sim$30 per nucleotide in Dornberger et al., 1998). The reduction arose from removing the interproton distances that came from those involving the hydrogen(s) other than the first (chosen arbitrarily) of a carbon or a nitrogen, those within the same deoxyribose ring, and those having the shortest distance between a deoxyribose and a nucleic acid base.

Simulation protocol and computational details

Table 1 summarizes the simulation protocol used in the present work. The protocol consists of a series of energy minimization stages and includes a molecular dynamics run of 5 or 10 ps at 300 K. In the simulation, proteins were fixed at their crystallographic coordinates while their substrate DNA was brought into contact from a distance, as described above. In the first two minimization stages, DNA came into contact with the binding protein by the steering restraint forces, adopting in the meantime a correct orientation relative to the protein. Two stages were used because test runs indicated that premature onset of PC-NC restraints resulted in a DNA bent too early, consequently preventing it from properly presenting its MG to RH. The following dynamics at 300 K allowed some steric hindrance to be overcome, as well as RH-MG and PC-NC distances to be forced to move within the imposed upper limits. The simulation was completed by two more stages of energy minimization, with the final stage removing all of the RH-MG and PC-NC restraints but keeping a minimal force to maintain an essentially B-DNA conformation. Protein side chains of Fis were allowed to move during the first four stages of the simulation because, unlike CAP and Rep, the Fis structure did not come from a DNA-bound structure, and test runs showed that, in the case of Fis, absence of side-chain flexibility occluded MG from interacting with RH in an intimate fashion similar to that found in other HTH protein-DNA complexes. On the other hand, allowing side-chain flexibility significantly increased the variation in structures resulting from different simulation runs (see Results).

MSI's CFF force field (Hwang et al., 1994, 1998; Maple et al., 1994, 1998) was used for energy calculations. The suitability of this force field for simulating nucleic acid molecules was demonstrated by its ability to
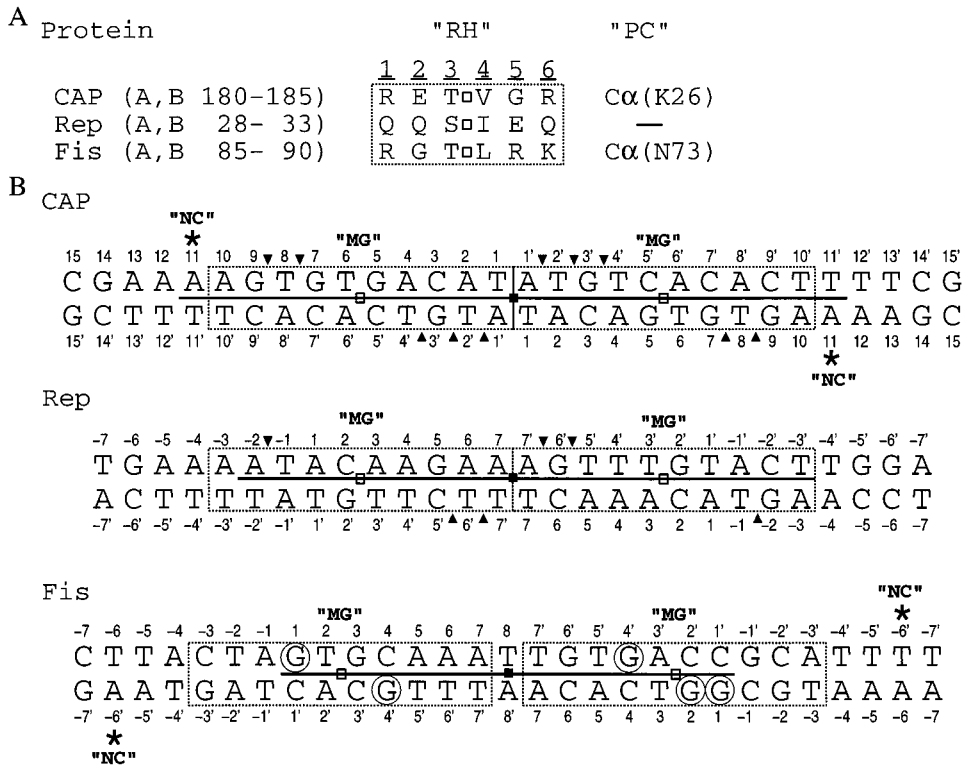
FIGURE 2   Definition of protein-DNA "interaction" sites. (A) RH (recognition helix): the geometric center (*small open square*) of the six most important amino acids of the second helix of the HTH motif involved in protein-DNA recognition; PC (protein nuclease cleaving residue): the Cα of the amino acid whose chemical modification renders nuclease scission activity. (B) MG: the geometric center (*small open square*) of a major groove (*dashed box*) into which RH inserts for protein-DNA recognition (this geometric center is essentially the symmetrical point for the sites of methylation (*circled base*) and ethylation interference (*filled triangle*) found experimentally (Steitz et al., 1983; Bushman et al., 1985; Bruist et al., 1987)); NC (nucleotide cleavage site): C1′ of the cleaved nucleotide, which is marked by an asterisk. The dyad of these sequences is marked by a filled square; to its left is the left half-site, and to its right is the right half-site. The convention used by other researchers to number these sequences was followed. Dark lines between the two complementary strands of DNA represent the nucleotide sequences of the "core" region; those outside the core region are regarded as "flanking sequences." For Rep DNA, the "core" region coincides with the sequence of the x-ray structure (Aggarwal et al., 1988).

reproduce in molecular dynamics simulations the structures of two dinucleotide crystals to a root mean square deviation (rmsd) of 0.4 Å, and that of a hexameric d(CGCGCG)$_2$ Z-DNA complex to 0.8 Å (C. S. Ewig, personal communication). In our simulations all nonbonded interactions within 13 Å were calculated, and those outside were gradually turned off by a fifth-order polynomial switching function of a 1.5-Å spline width. Group-based summation of nonbonded interactions was employed, and a list of interacting atoms was constantly updated, using a buffer width of 0.5

Å added to the cutoff. A distance-dependent dielectric constant $\epsilon(r) = r$ (r is the separation distance) was used for calculating Coulombic interactions. The total charge for a nucleotide is $-1.0e$. Test runs with the CAP system showed that reducing the phosphate charge from $-1.2e$ to $-0.4e$ resulted in unsatisfactory results—coordinates rmsd worsened from 3.3 ± 0.1 Å to 5.8 ± 0.1 Å (mean ± standard deviation from five independent runs). The conjugate gradient method was used for energy minimization, and a Verlet velocity integrator with a time step of 1 fs was used for molecular

TABLE 1   Simulation protocol

Stage	1	2	3	4	5
Method	Min	Min	Dyn (at 300 K)	Min	Min
Steps or energy gradient (kcal/mol-Å)	2000	1000	5 ps (10 ps for Fis)	1000	<0.5
Fixed protein atoms					
CAP and Rep	←		all	→	backbone
Fis	←		backbone	→	
Force constants (kcal/mol-Å^2) for distance restraints					
B-DNA	200	200	200	200	0.1
RH-MG	10	10	100	100	0
PC-NC (for CAP and Fis)	0	10	100	100	0

Min, Minimization; Dyn, molecular dynamics.

dynamics. Initial velocities of atoms were randomly assigned according to a Maxwell-Boltzman distribution at 300 K. The temperature was maintained within a window of 10 K by velocity scaling. During the simulation, the interaction sites of RH and MG (defined above) were treated as pseudoatoms. Restraint energy and energy derivatives owing to the pseudoatoms were computed according to the penalty function imposed on them; these energy derivatives were then converted by chain rules onto the constituent real atoms of the pseudoatoms to provide restraint-guiding forces. The coordinates of the pseudoatoms and their associated energies and energy derivatives were frequently updated—every 50 energy minimization iterations or 50 fs of molecular dynamics. For each system, five independent runs employing the same initial configuration and the same protocol as presented in Table 1 were performed by assigning a different random number to generate different initial atomic velocities.

Structural analyses

1. Structural comparison. The simulation-resulting models were compared to the x-ray structures by superimposing the protein backbone and then calculating the rmsd values in nonhydrogen coordinates of the predicted DNA in the complex structure.

2. DNA parameters. DNA parameters were measured by the program CURVES 5.1 (Lavery and Sklenar, 1988, 1989). They included groove widths, sugar-phosphate backbone torsional angles (α, β, γ, δ, ϵ, ζ, χ, and phase angle P), base pair-step parameters (twist, roll, tilt, rise, shift, slide), and individual base pair parameters (opening, propeller twist, buckle, shear, stretch, stagger). To measure the DNA groove widths for the crystal structure of the CAP-DNA complex (Schultz et al., 1991; PDB entry code 1cgp), two missing phosphate groups were filled based on the coordinates of a second CAP-DNA complex structure with different nick sites (Parkinson et al., 1996; PDB entry code 1ber).

3. DNA orientation and bending angles. The orientation angle of DNA relative to protein was measured as the cross-angle between the long axis of protein (x axis) and that of DNA in their projection onto a plane perpendicular to the twofold axis (x-y plane) of the complex. The sign of the DNA orientation was assigned according to the right-hand rule, with

the thumb pointing to the positive of the z axis (Fig. 1). Bending angles were measured for both (left and right) half-sites of the DNA; their sum then represents the overall bending. The bending angle for each DNA half-site was defined as the angle made by two axes, the long axis of the core sequence and that of the flanking sequence. The long axis of a DNA sequence was determined as that of a standard B-DNA best fitting the sequence by least squares superposition. For the superposition, 10 bp for CAP and Rep and 9 bp for Fis centering on the dyad (Fig. 2 B) were chosen as the core sequence, and 10 bp for CAP and Fis but only 8 bp for Rep (because its x-ray structure contains only the core; Fig. 2 B) as the flanking sequence. The choice of 10 bp to measure the bending angle, as opposed to 5 bp used by other researchers (e.g., Schultz et al., 1991), reduced the influence on the measurement that could be quite significant from just a few nucleotides at the end of the DNA. Our measurement for the reported x-ray structure of CAP DNA (84°, Table 2) is in close agreement with the ~85° value recently suggested to be a better estimate from crystal data for what might be observed in solution (Lutter et al., 1996).

4. Hydrogen bond contacts. All N-H···O or O-H···O contacts between protein and DNA with a H···O distance of less than 3.0 Å and an O(N)-H···O angle larger than 90° were regarded as hydrogen bonded, but a specific hydrogen bond was considered "real" only when it was present in at least three out of five independently simulated structures.

RESULTS

The predicted protein-DNA complex structures

The simulation-predicted protein-DNA complex structures were assessed as shown in Table 2 for a number of geometric parameters. They included 1) the rmsd of DNA heavy atoms in the complex form between the model and the x-ray structure (calculated upon superposition of protein backbones) for the two test systems (CAP and Rep), and among models themselves for the predicted system (Fis); 2)

TABLE 2　Models versus x-ray on selected geometric parameters

Parameters*		X-ray	CAP Model[#] A	B	B1	B2	B3	Rep X-ray	Rep Model[#] A	A1	A2	A3	Fis Model[#] A	B	A1	A2	A3
RMSD (Å)	Core	—	2.6_1	2.5_1	2.5_1	2.5_1	2.4_2	—	1.6_1	2.1_1	1.5_1	1.5_1	2.2_7	2.8_6	1.3_3	2.8_{12}	3.1_9
	All		4.6_2	3.3_1	4.0_1	3.6_1	3.5_2						2.3_5	3.0_6	1.5_4	2.6_8	3.5_{10}
Protein-DNA																	
distance (Å)[§]		23.3	24.4_1	23.7_1	23.6_1	23.8_1	22.9_2	16.2	16.4_1	16.3_1	16.0_1	16.0_1	19.4_2	20.0_2	19.0_1	19.7_2	19.1_2
initial angle (°)[¶]		—	0	0	20	−20	0	—	0	20	−20	0	60	60	40	80	60
final angle (°)		−17	$−22_1$	$−21_1$	$−22_1$	$−21_1$	$−22_5$	−5	$−6_1$	$−8_1$	$−5_1$	$−5_1$	23_1	24_2	23_1	24_1	19_4
RH-MG (Å)	L	8.5	8.2_1	8.0_1	7.8_1	7.9_1	8.0_1	8.0	7.9_1	7.8_1	7.8_1	7.8_1	10.2_3	10.5_2	10.0_2	10.2_3	9.9_3
	R	8.6	8.5_1	8.5_1	8.5_1	8.6_1	8.9_1	7.9	7.9_1	8.5_1	7.8_1	7.8_1	9.4_3	10.6_4	8.8_1	10.0_5	9.4_6
PC-NC (Å)	L	15.5	15.9_2	13.0_2	13.0_1	13.2_1	12.6_3	—					13.5_4	11.3_2	14.0_6	13.2_5	12.2_5
	R	14.8	16.6_1	16.4_2	15.5_{13}	16.7_1	14.7_5						14.3_1	11.3_4	14.1_3	13.7_4	12.5_9
DNA bending (°)	L	39	33_2	35_1	36_1	33_1	46_1	11	10_1	10_1	10_1	9_1	42_1	42_1	43_1	39_3	40_3
	R	45	35_1	42_1	42_1	42_1	46_1	16	18_1	17_1	17_1	17_1	28_2	39_1	25_1	31_3	30_7
	L+R	84	68_2	77_2	78_1	74_1	92_2	27	28_1	27_1	28_1	26_2	70_2	81_1	68_2	70_1	70_4

*Subscripts are standard deviations to the last digit in five independently simulated model structures. Some comparatively larger standard deviations for the same parameter of different models are underlined. L, left half-site; R, right half-site. The bending angle of Fis-bound DNA estimated from experiments of comparative gel electrophoresis is 74 ± 4° (Pan et al., 1996).

[#]Models A and B differ in the PC-NC restraint distance used. A, 17 Å and B, 15 Å for CAP-DNA; A, 13 Å and B, 11 Å for Fis-DNA. For subsequent simulations, model B for CAP-DNA (B1–B3) and model A for Fis-DNA (A1–A3) were chosen (see text). Simulations for model B3 of CAP-DNA and A3 of Rep-DNA and Fis-DNA imposed a reduced set of B-DNA interproton distance restraints (Materials and Methods).

[§]Measured between the geometric center of protein and that of DNA. Initial distance for the simulation was 140 Å for CAP, 100 Å for Rep, and 120 Å for Fis (Materials and Methods).

[¶]Measured between the long axis of protein and that of DNA in their projection to a plane perpendicular to the twofold axis (Materials and Methods).

the relative distance and relative angle between protein and bound DNA; 3) the resulting RH-MG and PC-NC distances upon which a restraint force was introduced in the simulation and removed only in the final stage of energy minimization; and 4) the DNA bending angle. The standard deviations in these parameters resulting from five different simulation trajectories were quite small, indicating good conformity in the model structures (the comparatively larger variations in the DNA coordinates of the Fis system were consequences of allowing its protein side chains to be flexible in the simulation; see Materials and Methods). For CAP-DNA and Fis-DNA, two sets of model structures were presented; the two differ only in the value of the restrained PC-NC distance. The structures of model A adopted the estimated arm length of the nuclease scission agent, 17 Å for CAP (Pendergrast et al., 1994) and 13 Å for Fis (Pan et al., 1994a), whereas those of model B used a distance 2 Å shorter. Model B for CAP-DNA and model A for Fis-DNA were chosen for subsequent analyses because they have a smaller coordinate rmsd. When compared with a second crystal structure of CAP-DNA complex (Parkinson et al., 1996), essentially the same coordinates rmsd results were obtained (3.3 Å for the whole DNA fragment and 2.6 Å for the core). For each of the three systems, the predicted complex structures were superimposed (on protein backbone) and shown in Fig. 3, A–C.

DNA parameters

To understand the conformational change of DNA induced by protein binding, DNA parameters were analyzed and are shown in Figs. 4 and 5 for the three HTH systems modeled.

Fig. 4 shows that for both the x-ray and the model structures, conformations of the sugar-phosphate backbones were almost all located in the most populated regions derived statistically from x-ray-observed B-DNA structures (Berman, 1997; Schneider et al., 1997). The C2′-endo sugar pucker and BI phosphate backbone were also commonly maintained in the model structures. The distribution of these parameters was more restricted in the model structures, which was expected because of the B-DNA interproton distance restraints.

As a result of bending, the major and minor grooves of DNA compress and expand, giving rise to a pattern of different groove widths at different nucleotide positions. As shown in Fig. 5, A and B, for CAP-DNA and Fig. 5, E and F for Rep-DNA, the groove-width patterns of the two known structures were generally reproduced by the model structures, although not to the exact values. Correspondingly, there were peaks in the positive roll angle at a position ~5 bases from the pseudodyad to both its right and left, i.e., junction 6/5 for CAP DNA (Fig. 5 C) and junctions 2/3 and 3/4 for Rep DNA (Fig. 5 G), and a negative value
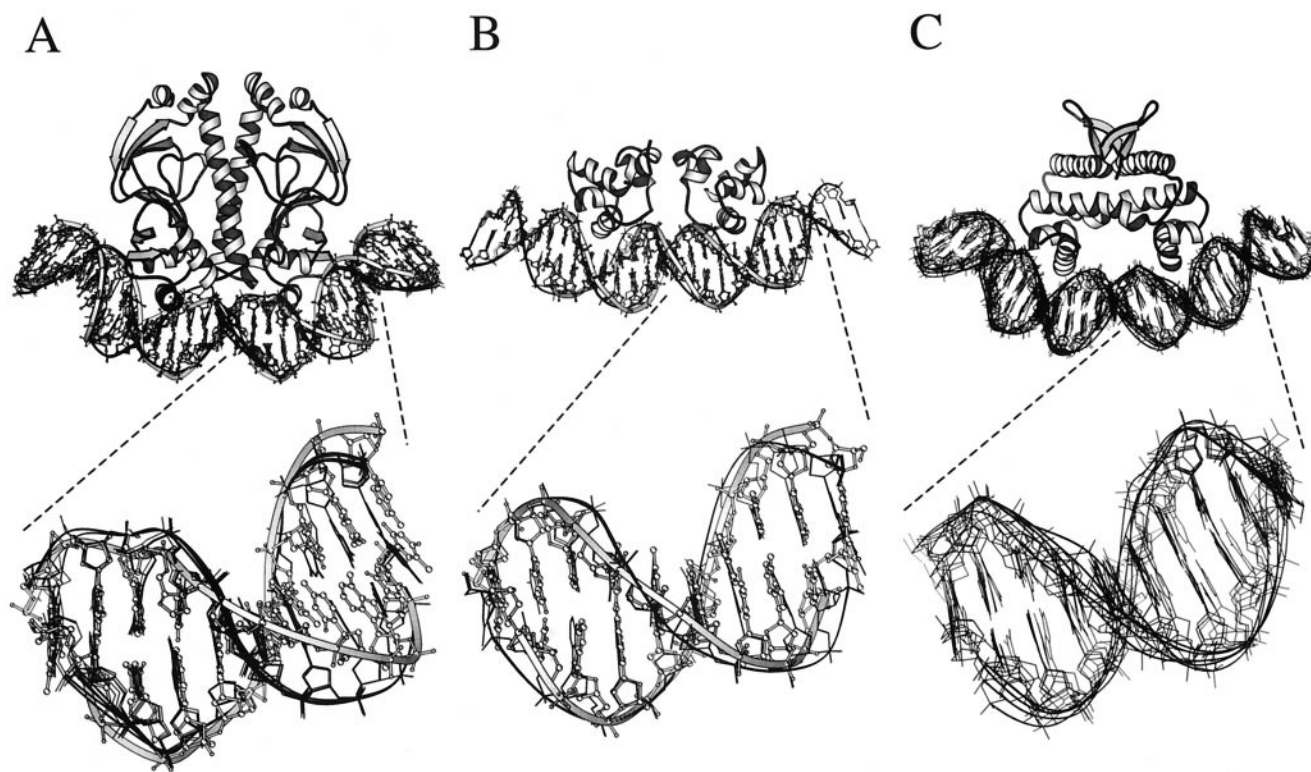

FIGURE 3 The simulation-resulting DNA model structures (*black thin line*) superposed on the x-ray structure (*gray stick and ball*) in the complex form. (*A*) CAP-DNA; (*B*) Rep-DNA; (*C*) Fis-DNA. The superposition was done on the backbone atoms of the protein. Close up is a 10-bp DNA segment of the right half-site. For clarity, proteins are shown by ribbons, and hydrogen atoms of DNA are omitted. These figures were prepared using Molscript (Kraulis, 1991).

FIGURE 4 The sugar-phosphate backbone angle parameters of the DNA in complex with CAP, Rep, and Fis. Parameters of the x-ray and model structures are shown separately. The full range of the angle values is represented by a dial circle, with 0°/360° in the north, 180° in the south. The most populated regions for each parameter as observed in experimentally resolved B-DNA structures (Berman, 1997; Schneider et al., 1997) are depicted by gray areas (α 270°-330°; β 130°-200°; γ 20°-80°; δ 70°-180°; ϵ 160°-210° (BI), 210°-270° (BII); ζ 150°-210° (BII), 230°-300° (BI); χ 200°-300°; phase angle P 340°-60° (C3' endo), 90°-190° (C2' endo)). For comparison, the initial values for each parameter used in the simulations are indicated by dashed lines.
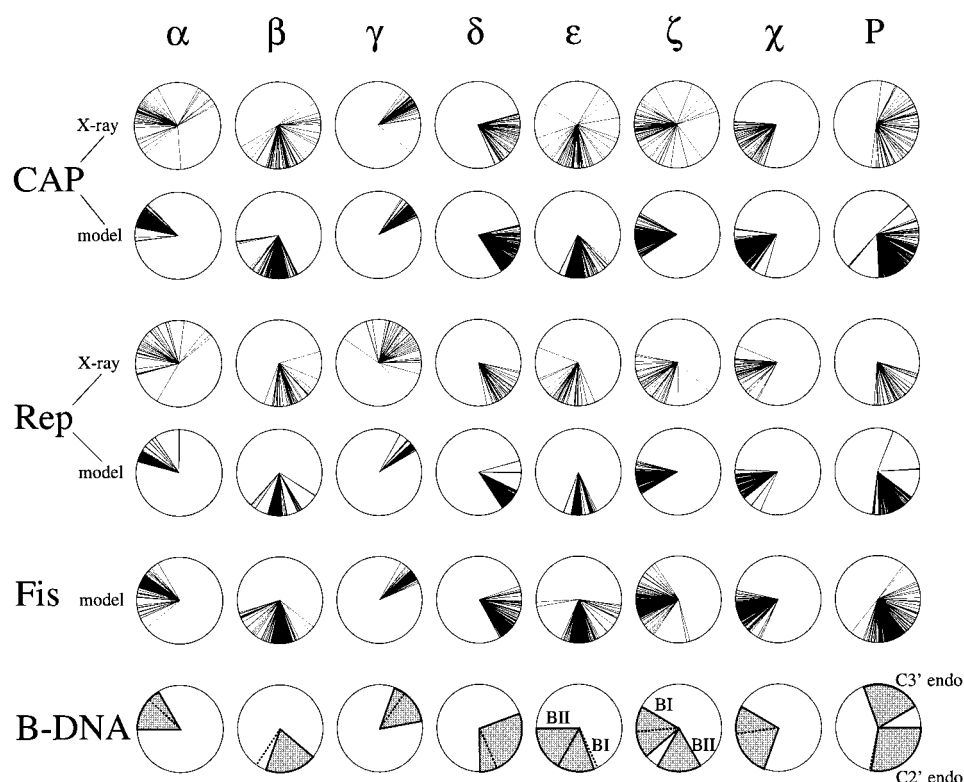
at the pseudodyad. These peaks, which were qualitatively reproduced in the model structures, represent in-phase roll rotation and largely account for the axial bending of DNA toward protein (see, e.g., Calladine and Drew, 1997; Dickerson, 1998). Tilt rotation (Fig. 5, *D* and *H*), which is reported to be both energetically (Zhurkin et al., 1991) and statistically (Grzeskowiak et al., 1993; Suzuki et al., 1996, 1997; El Hassan and Calladine, 1997; Dickerson, 1998) less favored than roll rotation, has a much smaller effect on DNA axial bending, and its variation should therefore be of minor significance.

The pseudodyad of Fis DNA is located at a base pair (a T-A base pair at position 8), not between base pairs, as in the DNA bound by CAP or Rep. If Fis DNA bends in a fashion qualitatively similar to that seen above in the CAP and Rep complexes, i.e., complying with an ideal way of in-plane axial bending (Calladine and Drew, 1997; Dickerson, 1998), one expects its roll angles to be negative at or near the pseudodyad (position 8), positive ∼5 bases from it (positions 3L and 3R), and negative again 5 bases further away (positions −3L and −3R). Although an experimental structure is not yet available to confirm that Fis DNA adopts such bending, the proposition is supported by the positions of DNA groove compression/expansion deduced from methylation, footprinting, and phasing experiments (Bruist et al., 1987; Pan et al., 1996; Perkins-Balding et al., 1997). Consistent with the proposition, negative roll angles at junctions 7L/8, 7R/8, −3L/−2L, and −4R/−3R and positive roll angles in the neighborhood of 3L were predicted in our model (Fig. 5 *K*). In contrast, the large positive roll

angles predicted at junctions 2L/3L and 3L/4L were diminished to almost nil at their symmetrical positions (2R/3R and 3R/4R), and a positive, rather than negative, roll angle at junction −3R/−2R was calculated. In turn, these "nonideal" roll angles gave rise to a statistically smaller bending angle for the right half-site than for the left half-site of the simulated *hin-D* sequence (Table 2). That is, the predicted bending of *hin-D* by Fis was asymmetrical, which is, in fact, in agreement with the observation of biochemical experiments (Pan et al., 1996).

Additional parameters such as twist, slide, buckle, propeller twist, etc., used to measure the extent a B-DNA structure deviates from its canonical reference state, were also analyzed. The results, which are summarized and compared in Table 3, further demonstrate the overall capability of the present simulations to reproduce experimentally observed DNA parameters. For example, statistical analyses of crystal structures of B-DNA have revealed that positive rolling is often accompanied by untwisting and negative sliding (Gorin et al., 1995; El Hassan and Calladine, 1997; Suzuki et al., 1997), and this correlation of parameters describing dinucleotide geometries was evident in our predicted models—namely, roll parameters of the model structures were negatively correlated with twist and slide, and twist and slide were positively correlated (Table 3). Specifically, twist angles smaller than 36°, indicative of untwisting and negative slides, were predicted at junction 6/5 for CAP DNA (indicated in Fig. 5, *M* and *N*, respectively). Similarly, untwisting and negative sliding were predicted at junctions 2/3 and 3/4 for Rep DNA and 2L/3L for Fis DNA
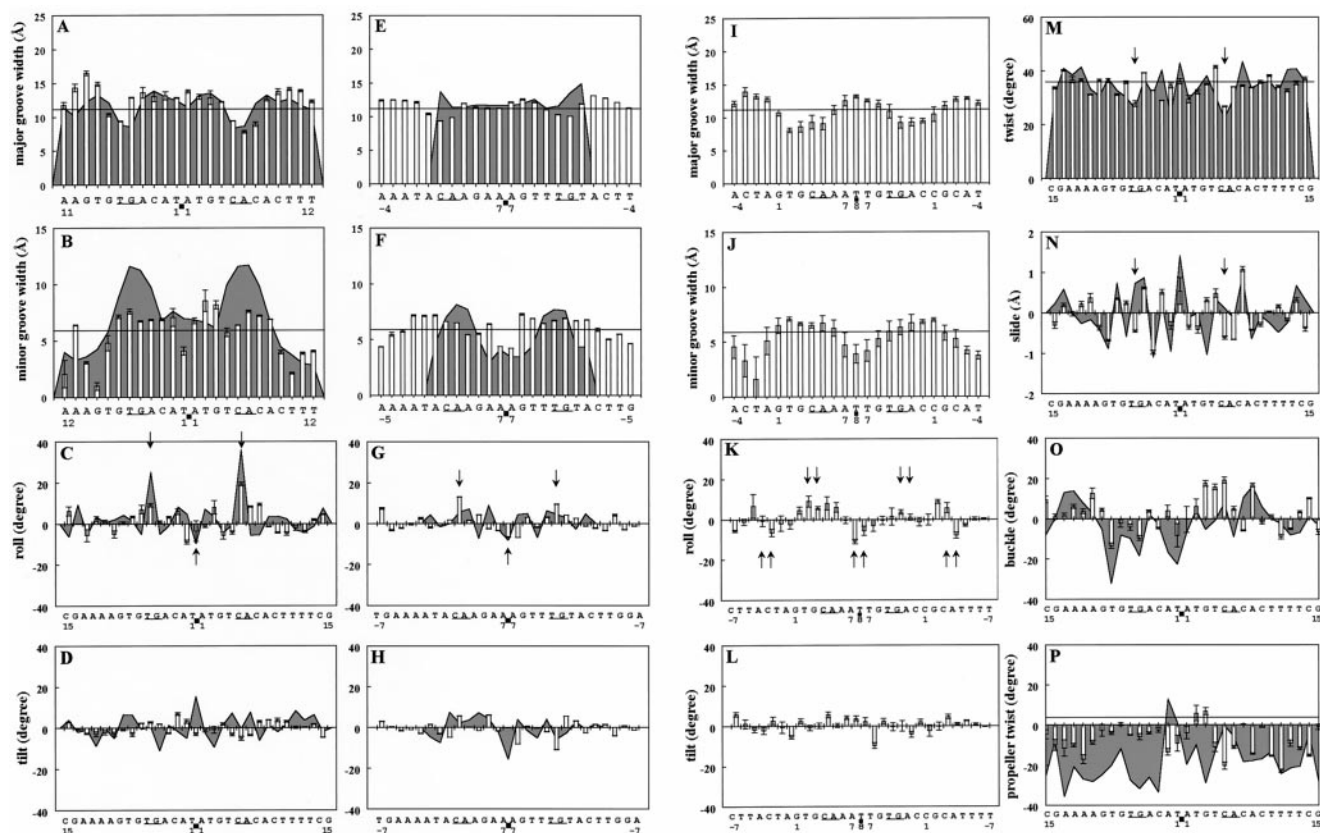
FIGURE 5 Groove width and roll and tilt angle parameters of the DNA in complex with CAP (*A–D*), Rep (*E–H*), and Fis (*I–L*), as well as twist, slide, buckle, and propeller twist of the DNA in complex with CAP (*M–P*). (*A, E, I*) Major groove widths; (*B, F, J*) minor groove widths; (*C, G, K*) roll angle; (*D, H, L*) tilt angle; (*M*) twist angle; (*N*) slide; (*O*) buckle angle; (*P*) propeller twist angle. Averages of the parameter resulting from five simulation-predicted structures are represented by vertical bars, with standard deviation shown. The corresponding values of these parameters for the x-ray structures are shadowed in the background, with the horizontal lines indicating the initial values of a straight B-DNA (11.2 Å and 5.9 Å for the widths of the major groove and the minor groove, respectively; 0° for roll, tilt, and buckle; 36° and 4° for twist and propeller twist, respectively; 0 Å for slide). The position of the pseudodyad is marked by a filled square; the nucleotide sequence to the left of the pseudodyad is the left half-site, and that to the right is the right half-site (also see Fig. 2 *B*). The most conserved TG steps are underlined. Arrows point to the ideal positions for positive (*down arrow*) or negative (*up arrow*) roll angles for in-plane axial bending (*C, G, K*) and to the corresponding positions of the former for the twist and slide parameters of the CAP-bound DNA (*M, N*) (see text). Note that the x-ray structure of the Rep-DNA complex encompasses only the core sequences, and an experimental DNA structure is absent from the Fis-DNA complex.

(data not shown). However, although these predictions are consistent with the findings of statistical analyses, they failed to reproduce quantitatively, although not qualitatively, certain specific parameters that deviate significantly from those of a standardized B-DNA, most notably the large positive slide of CAP DNA occurring at the same position where a large positive roll angle due to a major kink is observed in the crystal structure (indicated in Fig. 5, *C* and *N*). The comparisons of single base pair parameters (Table 3, as well as Fig. 5, *O* and *P*, for, respectively, the buckle and propeller twist parameters of CAP DNA) suggest that the overall features of distortions within a base pair were also qualitatively reproduced in the simulation models.

Protein-DNA hydrogen bond contacts

Using the criteria defined in Materials and Methods, all of the hydrogen bond contacts in both the experimental and

model structures were identified as shown in Fig. 6. Twenty-seven of 31 CAP-DNA and 24 of 25 Rep-DNA hydrogen bond contacts of the x-ray structures were reproduced by the predicted model complexes. Many additional contacts were predicted, probably because of an overly favored electrostatic interaction in the absence of the water solvent. There were 34 hydrogen bond contacts (four base and 30 backbone contacts) predicted for the Fis-DNA complex.

Results using different initial protein-DNA relative orientation angles

To investigate how would the stability of the predicted protein-DNA complex structures be influenced by the initial setting, additional simulations were carried out in which the DNA was rotated 20° from the original 0° (CAP and Rep) and 60° (Fis) used (Materials and Methods). The results are compared to those of the original simulations in Table 2.

TABLE 3 Models versus x-ray on parameters of basepairs and basepair steps

Parameters	CAP			Rep			Fis	
		Model			Model		Model	
	X-ray	B	B3	X-ray	A	A3	A	A3
Average value and standard deviation								
Twist (36°)	35_6	34_4	34_4	35_5	36_4	36_4	37_3	37_4
Roll (0°)	2_9	1_6	3_7	1_5	1_5	1_6	0_6	0_7
Tilt (0°)	1_6	0_3	0_4	0_6	-1_4	-1_4	1_4	0_4
Rise (3.4 Å)	3.5_4	3.4_3	3.4_4	3.4_4	3.3_2	3.3_2	3.3_2	3.3_3
Slide (0 Å)	0.0_6	0.0_5	-0.1_6	0.0_4	0.0_4	0.0_5	0.0_3	0.0_5
Shift (0 Å)	0.0_5	0.0_6	0.0_6	0.0_4	0.0_4	0.0_5	0.0_5	0.0_6
Opening ($-4°$)	2_7	-1_4	0_4	-1_5	-2_3	-2_3	-3_4	-3_5
Propeller (4°)	-18_{11}	-8_7	-7_9	-11_{10}	-7_6	-10_5	-9_7	-11_8
Buckle (0°)	-3_{11}	2_9	0_{11}	-3_8	1_6	2_6	0_8	0_8
Stagger (0 Å)	0.0_5	0.1_3	0.1_3	-0.2_7	0.1_3	0.0_3	0.2_3	0.2_3
Stretch (0 Å)	-0.1_1	-0.1_1	0.0_1	0.0_3	-0.1_1	-0.1_1	-0.1_1	-0.1_1
Shear (0 Å)	0.0_4	0.0_3	0.0_4	0.0_5	0.0_3	0.0_3	-0.1_3	-0.1_3
Correlation								
Roll-twist	-0.49	-0.54	-0.37	-0.23	-0.39	-0.45	-0.23	-0.39
Slide-twist	0.33	0.31	0.34	0.31	0.26	0.50	0.24	0.16
Roll-slide	0.03	-0.16	-0.16	-0.11	-0.33	-0.37	-0.21	-0.14

Subscripts are standard deviations to the last digit in five independently simulated model structures. Initial values for each parameter are listed in parentheses. Models B, B3, and A, A3 are as described in Table 2.

Except for a few large standard deviations, all parameters were quite similar to those of the original simulations. Note that in all of these simulations the final protein-DNA relative orientation angles, which differ significantly from the starting values, converged to essentially the same values. Such a convergence did not prevail (data not shown), however, if initially the DNA was rotated too much (e.g., 180°), indicating that the DNA must be properly oriented immediately before contact with the binding protein for the present docking simulation protocol to be successful.

Results using different numbers of B-DNA interproton distance restraints

In Table 2 we also compare the results of simulations in which the number of B-DNA interproton distance restraints was reduced to only about one-third of the original (Materials and Methods). For these simulations, comparisons of parameters of base pairs and base pair steps were also made (Table 3). These comparisons revealed that the DNA structures resulting from using the full set of B-DNA distance restraints and those from using the reduced set were similar, with the latter structures having only a marginal increase in flexibility, as indicated by their somewhat larger standard deviations in most of the structural parameters measured. Comparisons of individual DNA parameters such as those shown in Fig. 5 also revealed no major differences between the two different sets of simulations. However, reducing the number of interproton distance restraints further (for example, from 1342 to 1252 for CAP DNA) led to highly distorted DNA structures (data not shown).

DISCUSSION

Evaluation on the docking strategy and the effects of distance restraints

The B-DNA interproton distance restraints were required to sustain a B-DNA conformation (Figs. 4 and 5; without them a grotesquely distorted DNA structure resulted) without hindering DNA bending (Table 2). Evaluations of the resulting DNA parameters (Figs. 4 and 5; Table 3) indicate that the B-DNA restraints were effective. For example, in the model structures the distribution of the sugar-phosphate backbone parameters spread inside the most populated B-DNA conformation observed in experimental structures (Berman, 1997; Schneider et al., 1997), despite the fact that the initial values of some of the parameters (β, ϵ, and phase angle P) were outside the most populated region. Likewise, groove widths and geometric parameters of Watson-Crick base pairs and of base pair steps changed from their initial values and adopted a general trend of variations similar to that exhibited in the experimental structures (Fig. 5 and Table 3). Furthermore, these geometric variations as well as the final structures appear to be relatively independent of the number of B-DNA distance restraints imposed, provided that they are sufficient to sustain a B-DNA conformation during the dynamics simulations. This can be seen from the comparisons made in Tables 2 and 3 between models derived from the full set of B-DNA restraints (model B for CAP DNA and model A for Rep and Fis DNA) and those derived from the reduced set (model B3 for CAP DNA and model A3 for Rep and Fis DNA). This suggests that many interproton distance restraints of the original, nondiscriminating set are redundant, even though they did not seem to
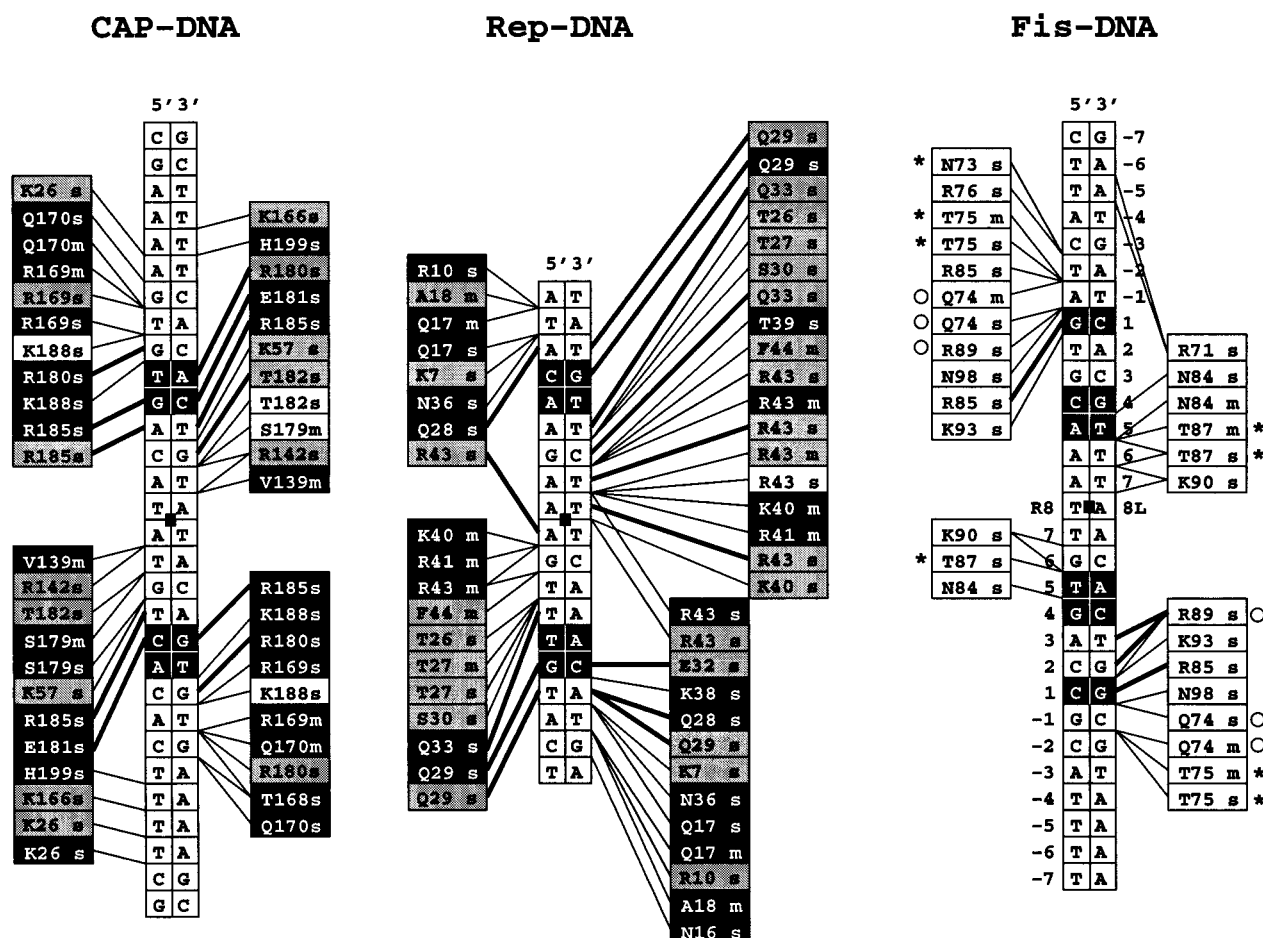
FIGURE 6 Protein-DNA hydrogen bond contacts. Contacts with DNA bases are denoted by a thick line linking the contacting amino acid and the DNA base. Contacts with DNA backbones are denoted by a thin line linking the amino acid and the base-connecting backbone, indicated by box joints. For each amino acid residue, s denotes a side chain and m the main chain. For CAP-DNA and Rep-DNA, the amino acids involving the contacts observed in both the x-ray and the model structures are shown in black boxes; those involving the contacts found only in the model structures are in gray boxes, and those involving the contacts found in the crystal structure but not in the model structure are in blank boxes. For Fis-DNA, the asterisked amino acids indicate that their mutations can cause a significant decrease in both the DNA binding and bending abilities of Fis, whereas the mutation of those marked by an open circle abolishes DNA binding but retains the DNA bending ability of Fis (Pan et al., 1996). The most conserved TG steps in the DNA bound by CAP, Rep, and Fis and a highly conserved G base of Fis-bound DNA are highlighted in inverted-video modes. The sequence of Fis-bound DNA is numbered as in Fig. 2 *B*.

elicit a significantly higher degree of hindrance of the flexibility of the DNA. Energy analyses (Table 4) indicated that there is some compensation between the number of B-DNA distance restraints imposed and the violation energy these restraints caused, in that an approximately threefold reduction in the former yielded only an approximately two-fold reduction in the latter. Consequently, more (percentage-wise) and somewhat severer violations were utilized by the simulations with the reduced set to maintain a B-DNA conformation, which required an energy cost amounting to ~5% of the protein-DNA interaction energy (Table 4). An evident deficiency of the B-DNA conformation restraints, however, was the consequence of a smoother DNA structure resulting from the simulation, which, although it led to an excellent fit for the Rep DNA, inevitably contributed to the larger rmsd for the CAP DNA (Table 2), and the lack of a prediction for the unusually large roll (Fig. 5 *C*) and pro-

peller twist (Fig. 5 *P* and Table 3) angles of this particular system.

As mentioned above, previous structural models of protein-induced DNA bending have relied on a pre-bent DNA. In such models one must decide a priori where to bend the DNA and how. This may not be much of a problem for sequences such as those recognized by CAP and Rep, because they exhibit a conserved TG dinucleotide step (the TG step is known to confer a propensity for bending; Suzuki et al., 1995; Dickerson, 1998) at the right position (i.e., 5 bases from the pseudodyad or, equivalently, the geometric center of the major groove involving recognition), which is bent by a dimerized HTH protein. However, when this general rule does not apply, as in the case of Fis-binding sites where the conserved TG step is only three bases away from the dyad center (position 4/5 on either side; Pan et al., 1996; Figs. 2 *B* and 5 *K*), pre-bending by assumption will present consid-

TABLE 4   Violations of B-DNA interproton distance restraints

Parameters	CAP		Rep		Fis	
	B	B3	A	A3	A	A3
Protein-DNA interaction energy (kcal/mol)	-1373_{30}	-1460_{35}	-1043_2	-1082_7	-1059_{31}	-1104_{50}
No. of restraints	3959	1342	3735	1256	3840	1300
Violation percentage (%)	64_1	76_1	62_1	73_1	62_1	73_1
Violation energy (kcal/mol)	77_2	45_5	58_1	30_1	53_1	33_4
Average violation energy per restraint (kcal/mol)	0.019_{47}	0.034_{77}	0.015_{45}	0.024_{62}	0.014_{36}	0.026_{63}
Average violation distance (Å)	0.3_4	0.4_5	0.2_3	0.3_4	0.2_3	0.3_4

Subscripts are standard deviations to the last digit in five independently simulated model structures. Models B, B3, and A, A3 are as described in Table 2.

erable uncertainties for modeling. Our knowledge-based docking simulation lessened this uncertainty somewhat by employing two experimentally derived distance restraints, RH-MG and PC-NC, as well as B-DNA conformation restraints, to substitute for the pre-bending employed in previous approaches. However, it is clear from the results shown in Table 2 that the predicted bending angle depended on the value of the PC-NC restraint distance, which can only be crudely estimated (see Materials and Methods). To a large extent, this dependency is attributable to the inadequate but practical choice of a distance-dependent dielectric constant and cutoff scheme for treating electrostatic interactions. Recent successes in the simulation of DNA molecules achieved by using explicit solvents, along with the particle-mesh Ewald summation method (e.g., York et al., 1995; Cheatham and Kollman, 1997; Duan et al., 1997; Young et al., 1997), promise to render these distance restraints unnecessary, because, as the power of computers continues to improve, the use of explicit solvent treatments will become practical for simulating a series of protein-DNA complexes.

In contrast to the uncertainties of PC-NC restraints, the RH-MG distances of the CAP- and Rep-DNA crystal structures were reproduced remarkably well in the models (Table 2), considering the fact that the restraint is an upper limit of a much longer distance (11 Å). For the Fis-DNA complex model, it is of interest to note that the RH-MG distance predicted ($\sim$10 Å) is close to those of the eukaryotic homeodomain, $\sim$9.8 Å (Billeter et al., 1993), and evidently longer than those of prokaryotic DNA-binding proteins with an HTH motif, $\sim$8.5 Å (Suzuki and Gerstein, 1995), of which Hin (Feng et al., 1994) is an exception with an RH-MG distance of 9.4 Å (Suzuki and Gerstein, 1995).

Despite only up to four distance restraints (two PC-NC and two RH-MG) being used to associate protein and DNA, improved values of rmsd from experiments were achieved here over those of Sandmann et al. (1996), who employed a pre-bent DNA—3.3 Å versus 3.7 Å for CAP and 1.6 Å for Rep versus 1.9 Å for phage 434 cro, which is a system similar to Rep. The well-predicted bending angles for both half-sites of the DNA bent by CAP and Rep (Table 2, model B for CAP), qualitatively reproduced DNA parameters (Figs. 4 and 5 and Table 3), which include roll-twist-slide correlations (Table 3), and hydrogen bond contacts (Fig. 6), suggest that the present simulation protocol is capable of capturing the essential conformational properties of the DNA in these protein-DNA complexes.

Features of different Fis-DNA models

The structure of the Fis-DNA complex was previously predicted by two research groups, both employing a pre-bent DNA. The DNA bending angle in Sandmann's model (Sandmann et al., 1996) was 90°, and that in Pan's model (Pan et al., 1996) was 52° or 92° when additional bending by the flanking sequences of *hin-D* was considered. These models assumed a symmetrical bending from both half-sites of the DNA. Our simulation for the Fis-induced *hin-D* bending resulted in a 70° overall bending, which is the sum of two asymmetrically bent half-sites (Table 2). This prediction is in close agreement with the asymmetrical nature and bending angle, 74° $\pm$ 4°, deduced from comparative gel electrophoresis (Pan et al., 1996). However, this agreement needs to be viewed from the perspective that different methods for measuring DNA bending angles often yield different values (Lutter et al., 1996). All of these models of Fis-DNA are similar in regions where the minor groove faces protein, producing negative roll angles at junctions $-4/-3$, $-3/-2$, and 7/8. For regions where the major groove faces protein, in Sandmann's model positive roll angles are concentrated at the conserved TG step (junction 4/5) and junctions adjacent to the conserved guanine at position 1, whereas in our model and in Pan's model positive roll angles mostly occur in junctions between the two conserved regions of the DNA sequences.

In the present study, compared to CAP-DNA and Rep-DNA, the predicted protein-DNA hydrogen bond contacts for Fis-DNA were much less extensive (Fig. 6). Similar results were obtained in Sandmann's and Pan's models. This reflects a sharply curved shape of the DNA bent by Fis and, in our model, a larger RH-MG distance ($\sim$10 Å versus $\sim$8.5 Å) as discussed above. In our and Sandmann's model, only R85 and R89 make base contacts, mainly with the conserved guanine at position 1 (Fig. 6 and Sandmann et al.,

1996). In Pan's model, an additional base contact was provided by K90 with the guanine at position 4. This additional contact is consistent with the result of a methylation interference experiment (Bruist et al., 1987) but is difficult to reconcile with the ineffective mutation of K90A (Pan et al., 1996). R85 and R89 are two Fis residues whose mutants can greatly reduce DNA binding—by more than 100-fold over that of wild type (Osuna et al., 1991). It is interesting to note that point mutations for several of the amino acids that were predicted to interact with DNA in the vicinity of the R85 and R89 contacts (i.e., near the conserved guanine at position 1) affect the protein's ability to bind but not its ability to bend DNA, whereas the DNA contacts of these amino acids are flanked by those of the amino acids (N73, T75, T87) whose mutations affect both DNA binding and bending (see Fig. 6). This pattern, together with the observation that N73, T75, and T87 are located at two opposing edges of an extruding surface formed primarily by the two helices of the HTH motif (figure not shown), may suggest specific surface complementation between protein and DNA in this particular system.

Roles of RH-pair configuration and TG dinucleotide step

Protein-induced DNA bending may be considered to be a process by which a DNA is recruited by a protein to mold to its exterior surface (Härd and Lundbäck, 1996; Rhodes et al., 1996; Allemann and Egli, 1997). Such a view supports the notion that, in addition to sequence-specific contacts, the specificity of nucleotide sequences stems from a sequence-encoded propensity for adopting a DNA conformation that can match well with the exterior surface of the binding protein. The TG step appears to be suitable for this task because it is among the most flexible dinucleotide steps (Schultz et al., 1991; Gorin et al., 1995; Suzuki et al., 1996, 1997; El Hassan and Calladine, 1997; Hunter and Lu, 1997; Dickerson, 1998). For an HTH protein dimer, which is ideal for binding palindromic DNA sequences, the surface that contacts DNA is convex, contributed mostly by a pair of extruding helices (i.e., RH). Structural organization of the RH pair with respect to the bound DNA (Fig. 7) and properties of the binding DNA sequences, such as the location of the conserved TG steps (Table 5; Berg and von Hippel, 1988; Barber and Zhurkin, 1990; Pan et al., 1996), may thus confer structural specificities on HTH dimerized protein-DNA complexes.

The most conserved TG steps for the CAP- and Rep-binding sequences are located at a position ideal for DNA bending (Fig. 5, *C* and *G*), and, by adopting in-phase roll angles, they serve as a bending point for necessary conformational changes of DNA. In addition, for a DNA to wind around and embrace the RH with most of the bending coming from a TG step, one can imagine that the distance between the two rolled TG steps, which are dyad-related,
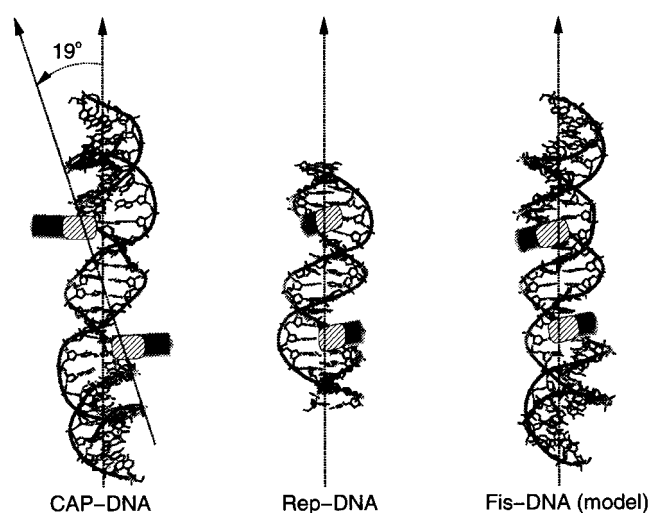

FIGURE 7 The organization of the dimerized RH pair with respect to DNA, as viewed through the twofold axis and with the DNA bent toward the viewer, for CAP-DNA, Rep-DNA, and Fis-DNA (model). The location of the most important six amino acids of the RH for protein-DNA recognition (Suzuki and Gerstein, 1995) is indicated by a slashed box, which is embedded within a complete α-helix. The RH-RH vector was drawn through the geometric center of the six amino acids of each RH (see Fig. 2 *A*).

needs to be at least equal to the distance between the RH pair. This was observed in the experimental structure of CAP-DNA and Rep-DNA complexes (Table 5). It is noteworthy that the TG⋯TG distance for CAP-DNA is several angstroms longer than for Rep-DNA. This difference appears to be necessary to compensate for the displacement of the RH pair from the DNA axis, or, correspondingly, for a much larger (19° versus 0°) RH pair-DNA relative orientation angle for CAP-DNA (Fig. 7). It has been shown that orientations between the two helices of the RH pair play an important role in changing the helical twist of protein-bound DNA (Suzuki et al., 1995); our analysis further suggests that the orientation of the RH pair with respect to the bound DNA could also be significant.

In concert with the smaller number of base pair steps (7 versus 10) separating the two dyad-related TG steps of the Fis-binding sites, the TG⋯TG distance is quite short in the Fis-DNA complex model (shorter than the RH⋯RH distance observed in the x-ray structure of the protein dimer). Therefore, it is likely that the TG step of Fis DNA plays a role somewhat different from those bent by CAP and Rep, in that for Fis the TG step may not necessarily be a bending point. In our model, positive roll angles were predicted for several consecutive dinucleotide steps encompassing the TG step (much more significantly for the left half-site; see Fig. 5 *K*), but the largest roll angle appears not at this step (position 4/5), but instead at position 2L/3L. This prediction—that the major bending point is not at the conserved TG step but at a nearby step—is in accord with the 2/3 step being suggested to have a positive roll angle based on experimentally determined DNase I hypersensitive sites (Bruist et al., 1987;

TABLE 5 Some geometric properties regarding the conserved TG dinucleotide and protein recognition helices

| Complex | TG $\cdots$ TG* | | RH $\cdots$ RH distance (Å) | Orientation angle (°) of the RH pair relative to DNA[#] |
	Base pair steps	Distance (Å)		
CAP-DNA	10	37.0	30.8	19
Rep-DNA	10	33.4	29.3	1
Fis-DNA[§]	7	23.4_3	24.3	0_1

*Number of base pair steps separating the conserved TG step on both half-sites and the distance between them in the complex structure (x-ray for CAP-DNA and Rep-DNA, model for Fis-DNA).

[#]Measured between the long axis of DNA and the line connecting the RH pair in their projection onto a plane perpendicular to the twofold axis of the complex (see Fig. 7).

[§]Model structures from five independent docking runs, with standard deviations to the last digit in subscripts.

Pan et al., 1996). Moreover, it should be noted that this result for Fis is in contrast to those for CAP and Rep, where, despite the fact that the same docking strategy was employed, the largest roll angle peak was indeed predicted to be at the TG step (Fig. 5, *C* and *G*). Thus, both the simulation and biochemical experiment suggest that the TG step of Fis DNA may be more of a bending facilitator than of a bender itself. In the same context, one may speculate that a second TG step at position 7/8 of the CAP-binding sequences, which is only slightly less conserved than the TG step at position 5/6 (Berg and von Hippel, 1988; Barber and Zhurkin, 1990), may also play a helper role in assisting the 5/6 TG step, its neighbor, to produce a drastic kink for accommodating the unusually organized RH pair of the CAP protein (discussed above). This speculation is in line with the suggestion that roll bending of pyrimidine-purine (YR, the dinucleotide step class to which TG belongs) steps can be facultative (Dickerson, 1998).

In conclusion, the present study has provided a first step toward developing an objective methodology for predicting complex structures of protein-DNA associations. We have demonstrated that a rather simple docking approach guided by only a few knowledge-derived distance restraints can be used to satisfactorily model the DNA structure bound and bent by three prototypical HTH protein dimers. The flexibility of DNA, which is a major problem for predicting protein-DNA complex structures, was curbed significantly by a set of interproton distance restraints derived without discrimination (except that they are all within 5 Å) from a standard B-form DNA. Interestingly, DNA bendability was not lost under these restraints. Furthermore, simulations with a much reduced set of B-DNA distance restraints yielded very similar structures with comparable variations in structural parameters that are in qualitative agreement with those observed crystallographically. These results indicate that a controlled conformational flexibility of DNA was achieved to render the simulation reasonably successful, although at the expense of a compromised capability for quantitative modeling of large DNA kinks, such as the one induced by CAP. More accurate treatment of the electrostatics and solvation should improve the objectivity and accuracy of the present method for protein-DNA docking studies.

REFERENCES

Aggarwal, A. K., D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science.* 242:899–907.

Allemann, R. K., and M. Egli. 1997. DNA recognition and bending. *Chem. Biol.* 4:643–650.

Aloy, P., G. Moont, H. A. Gabb, E. Querol, F. X. Aviles, and M. J. E. Sternberg. 1998. Modelling repressor proteins docking to DNA. *Proteins Struct. Funct. Genet.* 33:535–549.

Barber, A. M., and V. B. Zhurkin. 1990. CAP binding sites reveal pyrimidine-purine pattern characteristic of DNA bending. *J. Biomol. Struct. Dyn.* 8:213–232.

Bastia, D. 1996. Structural aspects of protein-DNA interactions as revealed by conversion of the interacting protein into a sequence-specific cross-linking agent or a chemical nuclease. *Structure.* 4:661–664.

Berg, O. G., and P. H. von Hippel. 1988. Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.* 200:709–723.

Berman, H. M. 1997. Crystal studies of B-DNA: the answers and the questions. *Biopolymers.* 44:23–44.

Billeter, M., Y. Q. Qian, G. Otting, M. Müller, W. Gehring, and K. Wüthrich. 1993. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J. Mol. Biol.* 234:1084–1093.

Bruist, M. F., A. C. Glasgow, R. C. Johnson, and M. I. Simon. 1987. Fis binding to the recombinational enhancer of the Hin DNA inversion system. *Genes Dev.* 1:762–772.

Bushman, F. D. 1993. The bacteriophage 434 right operator. Roles of O_R1, O_R2 and O_R3. *J. Mol. Biol.* 230:28–40.

Bushman, F. D., J. E. Anderson, S. C. Harrison, and M. Ptashne. 1985. Ethylation interference and X-ray crystallography identify similar interactions between 434 repressor and operator. *Nature.* 316:651–653.

Calladine, C. R., and H. R. Drew. 1997. Understanding DNA. Academic Press, London.

Cheatham, T. E., III, and P. A. Kollman. 1997. Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[AC-CCGCGGGT]$_2$ in the presence of hexaamminecobalt(III). *Structure.* 5:1297–1311.

Chou, S.-H., L. Zhu, Z. Gao, J.-W. Cheng, and B. R. Reid. 1996. Hairpin loops consisting of single adenine residues closed by sheared A.A and G.G pairs formed by the DNA triplets AAA and GAG: solution structure of the d(GTACAAAGTAC) hairpin. *J. Mol. Biol.* 264:981–1001.

Dickerson, R. E. 1998. DNA bending: the prevalence of kinkness and the virtues of normality. *Nucleic Acids Res.* 26:1906–1926.

Diekmann, S. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787–791.

Dornberger, U., J. Flemming, and H. Fritzsche. 1998. Structure determination and analysis of helix parameters in the DNA decamer d(CATG-GCCATG)$_2$-comparison of results from NMR and crystallography. *J. Mol. Biol.* 284:1453–1463.

Duan, Y., P. Wilkosz, P. Crowley, and J. M. Rosenberg. 1997. Molecular dynamics simulation study of DNA dodecamer d(CGCGAATTCGCG) in solution: conformation and hydration. *J. Mol. Biol.* 272:553–572.

El Hassan, M. A., and C. R. Calladine. 1997. Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behavior of dinucleotide steps. *Philos. Trans. R. Soc. Ser. A.* 355:43–100.

Feng, J.-A., R. C. Johnson, and R. E. Dickerson. 1994. Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science.* 263:348–355.

Gorin, A. A., V. B. Zhurkin, and W. K. Olson. 1995. B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* 247:34–48.

Grzeskowiak, K., D. S. Goodsell, M. Kaczor-Grzeskowiak, D. Cascio, and R. E. Dickerson. 1993. Crystallographic analysis of C-C-A-A-G-C-T-T-G-G and its implications for bending in B-DNA. *Biochemistry.* 32:8923–8931.

Härd, T., and T. Lundbäck. 1996. Thermodynamics of sequence-specific protein-DNA interactions. *Biophys. Chem.* 62:121–139.

Harrison, S. C., and A. K. Aggarwal. 1990. DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.* 59:933–969.

Hunter, C. A., and X.-J. Lu. 1997. DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide x-ray crystal structures. *J. Mol. Biol.* 265:603–619.

Hwang, M.-J., T. P. Stockfisch, and A. T. Hagler. 1994. Derivation of class II force fields. 2. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules. *J. Am. Chem. Soc.* 116:2515–2525.

Hwang, M.-J., X. Ni, M. Waldman, C. S. Ewig, and A. T. Hagler. 1998. Derivation of class II force fields. 4. Carbohydrate compounds and anomeric effects. *Biopolymers.* 45:435–468.

Knegtel, R. M., J. Antoon, C. Rullmann, R. Boelens, and R. Kaptein. 1994a. MONTY: a Monte Carlo approach to protein-DNA recognition. *J. Mol. Biol.* 235:318–324.

Knegtel, R. M., R. Boelens, and R. Kaptein. 1994b. Monte Carlo docking of protein-DNA complexes: incorporation of DNA flexibility and experimental data. *Protein Eng.* 7:761–767.

Kostrewa, D., J. Granzin, C. Koch, H.-W. Choe, S. Raghunathan, W. Wolf, J. Labahn, R. Kahmann, and W. Saenger. 1991. Three-dimensional structure of the *E. coli* DNA-binding protein FIS. *Nature.* 349:178–181.

Kostrewa, D., J. Granzin, D. Stock, H.-W. Choe, J. Labahn, and W. Saenger. 1992. Crystal structure of the factor for inversion stimulation FIS at 2.0 Å resolution. *J. Mol. Biol.* 226:209–226.

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

Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* 6:655–667.

Lengauer, T., and M. Rarey. 1996. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.* 6:402–406.

Lutter, L. C., H. R. Halvorson, and C. R. Calladine. 1996. Topological measurement of protein-induced DNA bend angles. *J. Mol. Biol.* 261:620–633.

Maple, J. R., M.-J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig, and A. T. Hagler. 1994. Derivation of class II force fields. 1. Methodology and quantum force field for the alkyl functional group and alkane molecules. *J. Comput. Chem.* 15:162–182.

Maple, J. R., M.-J. Hwang, K. J. Jalkanen, T. P. Stockfisch, and A. T. Hagler. 1998. Derivation of class II force fields. 5. Quantum force field for amides, peptides, and related compounds. *J. Comput. Chem.* 19:430–458.

Matthew, J. B., and D. H. Ohlendorf. 1985. Electrostatic deformation of DNA by a DNA binding protein. *J. Biol. Chem.* 260:5860–5862.

Osuna, R., S. E. Finkel, and R. C. Johnson. 1991. Identification of two functional regions in Fis: the N-terminus is required to promote Hin-mediated DNA inversion but not l excision. *EMBO J.* 10:1593–1603.

Pabo, C. O., and R. T. Sauer. 1992. Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61:1053–1095.

Pan, C. Q., J.-A. Feng, S. E. Finkel, R. Landgraf, D. S. Sigman, and R. C. Johnson. 1994a. Structure of the *Escherichia coli* Fis-DNA complex probed by protein conjugated with 1,10-phenanthroline copper(I) complex. *Proc. Natl. Acad. Sci. USA.* 91:1721–1725.

Pan, C. Q., S. E. Finkel, S. E. Cramton, J.-A. Feng, D. S. Sigman, and R. C. Johnson. 1996. Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J. Mol. Biol.* 264:675–695.

Pan, C. Q., R. Landgraf, and D. S. Sigman. 1994b. DNA-binding proteins as site-specific nucleases. *Mol. Microbiol.* 12:335–342.

Parkinson, G., C. Wilson, A. Gunasekera, Y. W. Ebright, R. E. Ebright, and H. M. Berman. 1996. Structure of the CAP-DNA complex at 2.5 angstroms resolution: a complete picture of the protein-DNA interface. *J. Mol. Biol.* 260:395–408.

Pendergrast, P. S., Y. W. Ebright, and R. H. Ebright. 1994. High-specificity DNA cleavage agent: design and application to kilobase and megabase DNA substrates. *Science.* 265:959–962.

Perkins-Balding, D., D. P. Dias, and A. C. Glasgow. 1997. Location, degree, and direction of DNA bending associated with the Hin recombinational enhancer sequence and Fis-enhancer complex. *J. Bacteriol.* 179:4747–4753.

Rhodes, D., J. W. R. Schwabe, L. Chapman, and L. Fairall. 1996. Towards an understanding of protein-DNA recognition. *Philos. Trans. R. Soc. Ser. B.* 351:501–509.

Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York.

Safo, M. K., W.-Z. Yang, L. Corselli, S. E. Cramton, H. S. Yuan, and R. C. Johnson. 1997. The transactivation region of the Fis protein that controls site-specific DNA inversion contains extended mobile b-hairpin arms. *EMBO J.* 16:6860–6873.

Sandmann, C., F. Cordes, and W. Saenger. 1996. Structure model of a complex between the factor for inversion stimulation (FIS) and DNA: modeling protein-DNA complexes with dyad symmetry and known protein structures. *Proteins Struct. Funct. Genet.* 25:486–500.

Schneider, B., S. Neidle, and H. M. Berman. 1997. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers.* 42:113–124.

Schultz, S. C., G. C. Shields., and T. A. Steitz. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science.* 253:1001–1007.

Steitz, T. A., I. T. Weber, D. Ollis, and P. Brick. 1983. Crystallographic studies of protein-nucleic acid interaction: catabolite gene activator protein and the large fragment of DNA polymerase I. *J. Biomol. Struct. Dyn.* 1:1023–1037.

Sternberg, M. J. E., H. A. Gabb, and R. M. Jackson. 1998. Predictive docking of protein-protein and protein-DNA complexes. *Curr. Opin. Struct. Biol.* 8:250–256.

Suzuki, M., N. Amano, J. Kakinuma, and M. Tateno. 1997. Use of a 3D structure data base for understanding sequence-dependent conformational aspects of DNA. *J. Mol. Biol.* 274:421–435.

Suzuki, M., and M. Gerstein. 1995. Binding geometry of α-helices that recognize DNA. *Proteins Struct. Funct. Genet.* 23:525–535.

Suzuki, M., N. Yagi, and J. T. Finch. 1996. Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.* 379:148–152.

Suzuki, M., N. Yagi, and M. Gerstein. 1995. DNA recognition and superstructure formation by helix-turn-helix proteins. *Protein Eng.* 8:329–338.

Travers, A. A. 1992. DNA conformation and configuration in protein-DNA complexes. *Curr. Opin. Struct. Biol.* 2:71–77.

Tzou, W.-S., and M.-J. Hwang. 1997. A model for Fis N-terminus and Fis-invertase recognition. *FEBS Lett.* 401:1–5.

Warwicker, J., B. P. Engelman, and T. A. Steitz. 1987. Electrostatic calculations and model-building suggest that DNA bound to CAP is sharply bent. *Proteins Struct. Funct. Genet.* 2:283–289.

Weber, I. T., and T. A. Steitz. 1984. Model of specific complex between catabolite gene activator protein and B-DNA suggested by electrostatic complementarity. *Proc. Natl. Acad. Sci. USA*. 81:3973–3977.

Weisz, K., R. H. Shafer, W. Egan, and T. L. James. 1994. Solution structure of the octamer motif in immunoglobulin genes via restrained molecular dynamics calculations. *Biochemistry*. 33:354–366.

York, D. M., W. Yang, H. Lee, T. Darden, and L. G. Pedersen. 1995. Toward the accurate modeling of DNA: the importance of long-range electrostatics. *J. Am. Chem. Soc*. 117:5001–5002.

Young, M. A., G. Ravishanker, and D. L. Beveridge. 1997. 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys. J*. 73:2313–2336.

Yuan, H. S., S. E. Finkel, J.-A. Feng, M. Kaczor-Grzeskowiak, R. C. Johnson, and R. E. Dickerson. 1991. The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. *Proc. Natl. Acad. Sci. USA*. 88:9558–9562.

Zhurkin, V. B., N. B. Ulyanov, A. A. Gorin, and R. L. Jernigan. 1991. Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA*. 88:7046–7050.