

Binding of the Estrogen Receptor to DNA. The Role of Waters

Dorina Kosztin, Thomas C. Bishop, and Klaus Schulten

Beckman Institute, Departments of Chemistry and Physics, University of Illinois at Urbana-Champaign, 405 North Matthews, Urbana, Illinois 61801 USA

ABSTRACT Molecular dynamics simulations are carried out to investigate the binding of the estrogen receptor, a member of the nuclear hormone receptor family, to specific and non-specific DNA. Two systems have been simulated, each based on the crystallographic structure of a complex of a dimer of the estrogen receptor DNA binding domain with DNA. One structure includes the dimer and a consensus segment of DNA, ds(CCAGGTCACAGTGACCTGG); the other structure includes the dimer and a nonconsensus segment of DNA, ds(CCAGAACACAGTGACCTGG). The simulations involve an atomic model of the protein–DNA complex, counterions, and a sphere of explicit water with a radius of 45 Å. The molecular dynamics package NAMD was used to obtain 100 ps of dynamics for each system with complete long-range electrostatic interactions. Analysis of the simulations revealed differences in the protein–DNA interactions for consensus and nonconsensus sequences, a bending and unwinding of the DNA, a slight rearrangement of several amino acid side chains, and inclusion of water molecules at the protein–DNA interface region. Our results indicate that binding specificity and stability is conferred by a network of direct and water mediated protein–DNA hydrogen bonds. For the consensus sequence, the network involves three water molecules, residues Glu-25, Lys-28, Lys-32, Arg-33, and bases of the DNA. The binding differs for the nonconsensus DNA sequence in which case the fluctuating network of hydrogen bonds allows water molecules to enter the protein–DNA interface. We conclude that water plays a role in furnishing DNA binding specificity to nuclear hormone receptors.

INTRODUCTION

A central problem in gene regulation is how DNA binding specificity is achieved by regulatory proteins. To understand and explain the molecular basis of DNA binding specificity most studies investigated the nature of the interface of various protein–DNA complexes through both structural and biochemical analysis (Luisi et al., 1991; Schwabe et al., 1993a,b; Rodgers and Harrison, 1993; Rastinejad et al., 1995; Gewirth and Sigler, 1995; Lundbäck et al., 1993, 1994; Ha et al., 1992; Eriksson and Nilsson, 1995).

In the present study we investigate the binding of one member of the nuclear hormone receptor family, the estrogen receptor (ER), to specific and nonspecific sequences of DNA, by means of molecular dynamics simulations. We analyze direct and water-mediated hydrogen bonds at the protein–DNA and protein–protein interface in order to investigate the possible role of water in the mechanism and the energetics of specific site recognition. The role of distortions induced by the protein on the DNA has been studied, on the basis of the same molecular dynamics simulations, in Bishop et al. (1997).

Nuclear hormone receptors are ligand-activated transcription factors that regulate gene expression by binding to specific sequences of DNA, termed response elements (RE), upstream of their target genes (Laudet et al., 1992; Evans,

1988; Parker, 1991). This family of receptor proteins includes the receptors for steroid hormones, like the ER or the glucocorticoid receptor (GR), as well as receptors for thyroid hormones, vitamin D₃ and retinoic acids. The receptor proteins exhibit a high degree of sequence conservation, being composed of several domains which are differentially conserved between the various receptors and have different roles: the N-terminal domain is involved in transcriptional activation, the central domain is involved in DNA binding, and the C-terminal domain serves to bind the hormone (Krust et al., 1986; Kumar et al., 1987; Freedman and Luisi, 1993). The recognition of the specific RE by receptor proteins is mediated through the small, highly conserved, DNA binding domain (DBD) that contains ~65–70 residues. A schematic representation of the tertiary structure of the DNA binding domain of the estrogen receptor is presented in Fig. 1. Domain swap experiments indicate that this domain is necessary and sufficient to impart target specificity (Green and Chambon, 1987; Green et al., 1988).

The isolated DBD of the ER was proven to be monomeric in solution; however, two DBDs bind highly cooperatively as homodimers to the response element (Kumar and Chambon, 1988; Schwabe et al., 1990, 1993a). The ER recognizes its response element by interaction of a “reading helix” with the major groove of the DNA. The cooperative binding between monomers is due to protein–protein contacts in the region, called the D-box, between residues 44 and 58 (Schwabe et al., 1993a).

The consensus estrogen response element (ERE) contains two hexameric half-sites (ERE_{1/2} with the AGGTCA sequence), arranged as a palindrome with a three-basepair spacer (see Fig. 1). It differs from the glucocorticoid response element (GRE) only in the central two basepairs of

Received for publication 19 December 1996 and in final form 9 May 1997.

Address reprint requests to Klaus Schulten, Department of Physics, Beckman Institute 3147, University of Illinois, 405 N. Mathews Ave., Urbana, IL 61801. Tel.: 217-244-1604; Fax: 217-244-6078; E-mail: kschulte@ks.uiuc.edu.

Thomas C. Bishop's present address is University of California at Berkeley, 350 Calvin Laboratory, Berkeley, CA 94720-0001.

© 1997 by the Biophysical Society

0006-3495/97/08/557/14 \$2.00

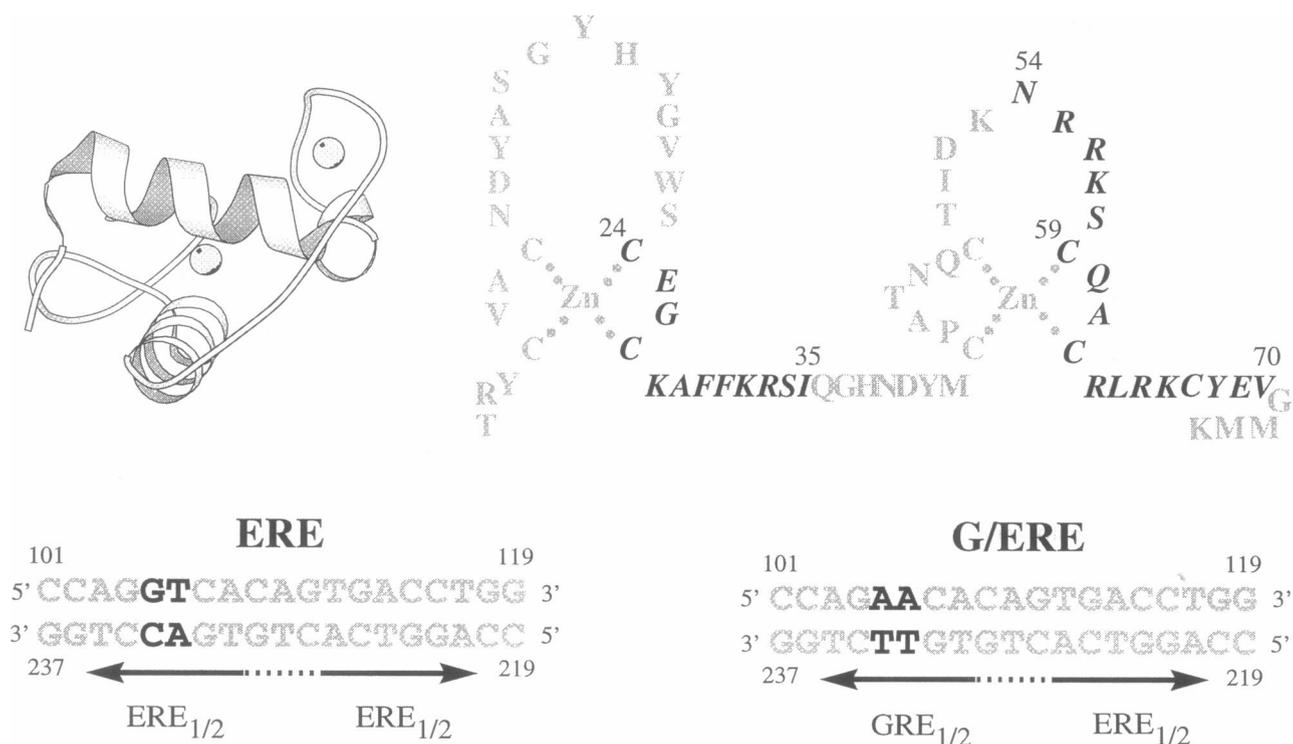


FIGURE 1 Top left, representation of the DNA binding domain of the estrogen receptor. The DNA binding domain has two subdomains, each containing a zinc ion coordinated by four cysteine residues, followed by an α -helix. The two α -helices are almost perpendicular to each other and hide the hydrophobic core of the protein. The third, short α -helix is not always present in the structure (Schwabe et al., 1993). This image was created using Molscript (Kraulis, 1991); right, the numbering for the ERDBD amino acid sequence shown in the classical "zinc-finger" representation. Helical regions are emphasized by boldface italic letters (C24–I35, N54–S58, C59–V70). Bottom, the numbering and sequence for the consensus ERE and nonconsensus G/ERE DNA response elements employed in the *ER-ERE* and *ER-G/ERE* simulations. Basepairs in the GRE_{1/2} half-site that differ from those in the ERE_{1/2} half-site are shown in boldface.

the half-site, the GRE_{1/2} having the sequence AGAACA, and from many other response elements, such as consensus thyroid and retinoic acid response elements, only in the relative spacing and orientation of these sequences. Mutagenesis experiments showed that changing only three amino acids (termed the P-box) Glu-25, Gly-26, and Ala-29 within the first zinc-binding motif of the ER is sufficient to switch the specificity of recognition from ERE to GRE (cf. Fig. 1) (Mader et al., 1989; Umesono and Evans, 1989; Danielsen et al., 1989). The change in binding specificity due to P-box mutations is partly due to changes in the cooperativity of dimerization (Zilliaccus et al., 1995).

The question of how the receptors interact with DNA and how they discriminate between different response elements has been addressed experimentally by determining the structures of protein–DNA complexes. However, in the majority of the resolved structures, the protein was bound to its consensus palindromic binding site. In only three cases were structures determined for nonconsensus DNA sequences that lead to insights in the mechanism of recognition (Rodgers and Harrison, 1993; Schwabe et al., 1995; Gewirth and Sigler, 1995). The crystal structure of the phage 434 repressor DNA-binding domain in complex with different DNA targets (Rodgers and Harrison, 1993) have

revealed extensive differences in the protein–DNA interface at consensus and nonconsensus sequences, showing a shift in the DNA backbone, a small global movement of the entire protein, a rearrangement of several amino acid side chains, as well as a shift in the stacking of some of the bases. The crystal structure of an estrogen receptorlike DBD bound to a glucocorticoid response element (Gewirth and Sigler, 1995) reveals that the basis for receptor target discrimination lies in the slight difference in the helical geometry of the two types of response element half-sites that leads to the inclusion of water molecules in the interface region. The crystal structure of the complex between a dimer of the ER-DBD and a nonconsensus DNA target site (Schwabe et al., 1995), in which there is a single base substitution in one-half of the palindromic binding site, revealed that recognition of the nonconsensus sequence is achieved by the rearrangement of a lysine side chain. Thus, it is now apparent that recognition of different sequences of DNA is a process accompanied by conformational changes, positive or negative interactions which are either direct or mediated by water.

A detailed understanding of the mechanism involved in DNA sequence discrimination for the estrogen receptor requires knowledge of all the components that influence its

specificity. The crystal structure of the DBD dimer of the estrogen receptor complexed with DNA (Schwabe et al., 1993a) is used as a reference structure for the two simulations: in one simulation the protein is bound to a consensus ERE; in the second simulation the protein is bound to a nonconsensus sequence of DNA. The latter nonconsensus sequence was obtained by mutating the central two basepairs in one half-site to correspond to the half-site sequence recognized by GR. The resulting nonconsensus sequence is still biologically active, but the affinity of the protein for the sequence is reduced (Schwabe et al., 1995). The loss of binding affinity for proteins is usually correlated with the failure to expel water molecules from the protein–DNA interface. In case of the ER protein, the crystallographic study of (Schwabe et al., 1995) for single-basepair mutations revealed only a rearrangement of a lysine side chain. The number of water molecules seen at the protein–DNA interface was the same, although the affinity of the protein for the nonspecific DNA sequence is 10-fold lower in this case. Hence, other effects are required in order to explain the loss of binding affinity. By replacing two basepairs involved in contacts with the protein, we significantly alter the protein–DNA interaction as well as the complementarity of shape at the protein–DNA interface. We expect that our comparison of the differences between the recognition of the estrogen receptor to its specific and nonspecific DNA sequence will convey the key determinants of the specificity of binding.

METHODS

All simulations were performed using the molecular dynamics program NAMD (Nelson et al., 1996) and version 22 (MacKerell, Jr., et al., 1995; MacKerell, Jr., and Karplus, 1996) of the CHARMM force field (Brooks et al., 1983). Parameters for the tetrahedral coordination of the zinc ions to cysteine residues were the same as those used by Eriksson et al. (1995). NAMD provides the option of calculating full electrostatic interactions through the use of a multipole expansion algorithm, namely the program DPMTA (Rankin and Board, 1995). This option was used throughout to calculate electrostatic interactions between atoms beyond a cutoff distance of 10.5 Å to an accuracy of 10^{-6} . The computational effort in the calculation of the long-range interactions has been reduced by means of a Verlet I multiple time step option of NAMD (Biesiadecki and Skeel, 1993). Test simulations of the ER-DBD dimer–ERE complex demonstrated that a 1-fs short time step and a 4-fs long time step conserved energy, but a step size above 4 fs did not conserve energy.

The simulations carried out as described below were based on the 2.4-Å resolution X-ray crystallographic structure of the ER-DBD bound to DNA (Schwabe et al., 1993a). In this respect, it must be noted, however, that the crystals of the ER-DBD–ERE contain two DNA duplexes with four copies of the protein. The dimers, referred to as dimer A and dimer B, have the same position relative to the DNA and to its partner protein in the dimer, but a difference exists between the structures of dimer A and dimer B: a short α -helix that forms part of the dimer interface and makes contacts to the phosphate backbone of the DNA is present in both proteins of dimer A and absent in both proteins of dimer B. Only dimer A bound to its cognate DNA was used in our simulations. All the water molecules reported in the crystal structure were included in the simulations. The DNA duplex in the crystal structure was 17 bp long with a single cytosine base overhanging at the 5' end of each strand. The two unpaired C bases were paired with G bases, i.e., the final DNA duplex was 19 bp long with no base overhang as represented in Fig. 1.

Molecular dynamics simulations protocol

Two systems were created starting with the structure described above. The first system, labeled *ER-ERE*, contains the ER-DBD dimer and the DNA duplex, ds(CCAGGTCACAGTGACCTGG). The second system, labeled *ER-G/ERE*, was created using the same crystallographic data, but a different DNA sequence. The two basepairs, ds(GT), at positions 105 and 106 on the 5' strand, were mutated to ds(AA) corresponding to the half-site sequence that GR recognizes. This sequence of DNA is referred to as *G/ERE* since half of it resembles the GRE half-site ($GRE_{1/2}$) and the other half the ERE half-site ($ERE_{1/2}$). The monomer facing the nonspecific half-site ($GRE_{1/2}$) is monomer 1. The protein and DNA sequences and numbering used in both simulations are presented in Fig. 1.

The protocols used for solvating and simulating the two systems are identical. Hydrogen bonds were added using X-PLOR (Brünger, 1992). One thousand steps of minimization in the presence of strong harmonic constraints on all heavy atoms were performed on the complex to remove bad contacts and constraints due to crystal packing. The integrity of the DNA structure depends sensitively on the local environment; accordingly, a water bath together with counterions is required to stabilize the DNA conformation during the simulations (Seibel et al., 1985; Forester and McDonald, 1991; York et al., 1992; Prevost et al., 1993; Miaskiewicz et al., 1993; Schneider et al., 1993; Beveridge and Ravishanker, 1994; Jayaram et al., 1994; Sharp and Honig, 1995). Previous simulations of a GR-DBD dimer–DNA complex (Harris et al., 1994; Eriksson et al., 1995; Bishop and Schulten, 1996) indicated that members of this family of proteins will bend DNA; accordingly, the solvation shells were chosen large enough to permit DNA bending. The complex was immersed in the center of a sphere of radius 45 Å filled with TIP3P-water (Jorgensen et al., 1983). The 45-Å radius sphere of water was constructed by covering the spherical domain with a three-dimensional cubic grid. The grid spacing was 3.1 Å so that one water per cube resulted in a density of 1.0 g H₂O/cc. An oxygen atom and two hydrogen atoms were located within each cube. In order to obtain the proper geometry for each water molecule, the oxygens were held fixed, and 500 steps of energy minimization of the bond and angle energies were performed. All atoms were then released, and the entire system equilibrated at 300 K for 20 ps. At the end of the equilibration procedure, a proper radial distribution function for the oxygen–oxygen and oxygen–hydrogen distances and a stable temperature of 300 K had been achieved. All water molecules having one of the atoms closer than 1.8 Å from the nearest protein, DNA, or crystal water atoms were removed. Each resulting solute–solvent system was then subjected to 5 ps of equilibration during which only water molecules were allowed to move. To achieve electroneutrality for the system, sodium ions were added by replacing 30 water molecules with the highest electrostatic energies of the oxygen atom and located more than 9 Å apart from each other, and 5 Å apart from the protein or DNA atoms. The systems were further equilibrated for 5 ps without any constraints. The total number of atoms in the two resulting systems was 36,284 for the *ER-ERE* system and 36,573 for the *ER-G/ERE* system. A view of the simulated protein–DNA complex, surrounded by water and counterions, is presented in Fig. 2.

The oscillations around constant values of the total potential energy and of the RMS deviations between simulated and crystal structures during the last 5 ps of equilibration indicated that the system had been sufficiently equilibrated and the molecular dynamics simulations were continued for 100 ps. The resulting trajectories were analyzed by several methods described below. A summary of relevant data for each simulated system is presented in Table 1.

Analysis

The atom selection commands and energy routines available in X-PLOR (Brünger, 1992) were used to calculate nonbonded protein–protein and protein–DNA interaction energies every 1 ps for each trajectory. The total interaction energies are given by the sum of electrostatic and van der Waals energies.

The structural deviations of protein and DNA from the initial X-ray crystal structure were assessed on the basis of root-mean-square deviations

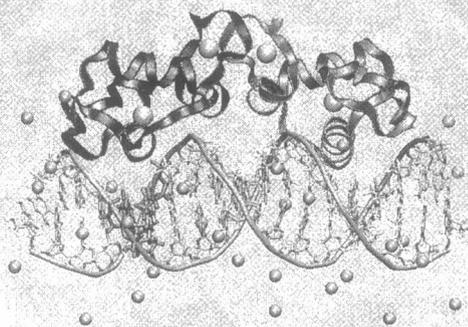


FIGURE 2 Simulated complex of an estrogen receptor (ER) dimer complexed to the estrogen response element (ERE). The structure is oriented such that the viewer is looking along the recognition helices lying in adjacent major grooves. The dimer interface is located between the monomers, above the minor groove of the DNA. The two mutated basepairs in one half-site are emphasized. Shown are also the sodium ions (as van der Waals spheres) and the sphere of water molecules included in the simulations. For a better view of the protein and the DNA, half of the sphere of water was cut. Figure created using VMD (Humphrey et al., 1996).

(RMSD). In computing the rms deviations the overall translational and rotational motions have been removed by superimposing the backbone of the protein or DNA of each configuration in the 100-ps trajectory onto the backbone of the protein or DNA in the crystal structure using a least-square fitting algorithm (Kabsch, 1976). The Debye-Waller factors, or B-factors, provide another important basis for comparing molecular dynamics trajectories with experimental results of X-ray crystallography. The theoretical temperature factors were computed according to:

$$B = \frac{8\pi^2}{3} \langle \Delta r^2 \rangle \quad (1)$$

where $\langle \Delta r^2 \rangle$ is the mean square atomic displacement averaged over the trajectories after rigid body alignment against the coordinates of the starting structure. The average was taken over all atoms in a given amino-acid or nucleotide residue. Solvent accessible surface areas for the proteins and the DNA were calculated using the routine available in X-PLOR (Brünger, 1992).

Direct and water-mediated hydrogen bond interactions between monomers and between protein and DNA were analyzed using the following conventions. Two atoms were considered to form a hydrogen bond (A . . . H-D) if the acceptor donor distance was less than 3.5 Å and if the A-H-D angle was between 120° and 180°. A water bridge was defined as an interaction between any residues that hydrogen bond to a common water molecule. In the Results section are presented and discussed only those hydrogen bonds and those water molecules that occupy the same bridging positions for >30% of the simulation time.

RESULTS AND DISCUSSION

In this section we focus our analysis on the hydrogen bond network at the protein–DNA and protein–protein interface,

on water molecules that place themselves at the interface and mediate protein–DNA interactions, and on the role played by each subunit of the DNA (phosphodiester backbone and bases) in establishing binding specificity.

Energy analysis

In order to study the role of the bases and phosphate backbone of the DNA in furnishing the binding specificity, the interactions between each monomer and DNA were divided into two contributions: interactions of the monomers with the bases of the DNA and interactions of the monomers with the phosphate backbone of the DNA. The interactions between the DNA phosphodiester backbone and the protein align the protein for further interactions with the bases. During the binding of the protein to the DNA the interaction with the bases serves to further stabilize binding and furnish binding specificity. The interaction energies for each contribution are presented as a function of time in Fig. 3 for both simulated systems.

In case of the *ER-ERE* system the total interaction of the two monomers with the bases and backbone of the DNA cannot be distinguished according to either of the two contributions to protein–DNA interactions (see Fig. 3). This is expected since each monomer faces the same DNA sequence.

In case of the *ER-G/ERE* system the mutations introduced into the first half-site yield an imperfect binding site, causing the interactions between DNA bases and monomer 1 to become less favorable than the interactions with the consensus half-site. Interestingly, the interaction with the DNA backbone is stronger for monomer 1 than for monomer 2. This result indicates clearly that the interaction between the protein and the phosphate backbone of the DNA does not impart specificity, but contributes to specificity by positioning the reading helix correctly in the major groove; specificity is ultimately conferred by the interaction of the protein with the bases of the DNA.

The time dependence of the RMSD of the protein heavy atoms and DNA revealed that, in both simulations, the RMSD rises during the first 30 ps of simulation time and then fluctuates around stable values. In the *ER-ERE* system the RMSD value for the DNA stabilizes at ~1.4 Å and the RMSD for the protein stabilized at ~2.0 Å for monomer 1, and 1.8 Å for monomer 2. In case of the *ER-G/ERE* system the RMSD values for the DNA and protein are larger. The DNA stabilized at ~1.9 Å, monomer 1 at ~2.3 Å, while the RMSD reaches a value of 2 Å for monomer 2. It is interesting to note that the rise of the RMSD value for monomer 1 triggers a rise in the RMSD value for monomer 2, obviously due to interactions between monomers. The RMSD values are presented and further discussed in Bishop et al. (1997).

Solvent-accessible surfaces and temperature factors

Thermodynamics studies indicate that sequence-specific DNA binding is usually accompanied by a large and nega-

Table 1 Simulation summary

Simulation	<i>ER-ERE</i>	<i>ER-G/ERE</i>
DNA sequence	CC AGGTA CAG TGACCT GG	CC AGAACA CAG TGACCT GG
Protein		ER DBD dimer
Initial Coordinates	Crystal	Modified
Total no. atoms	36,284	36,573
Total no. waters	10,944	11,040
Total no. Na ⁺	30	30
Simulation program		NAMD
Processor (no. nodes@mem.)		HP735-125MHz (8@128 Mb)
Energy function		All atom, CHARMM22, TIP3P
Integration method		Verlet I multiple time step with switching function
Timestep (short, long)		1 fs/4 fs
Coulomb evaluation		DPMTA
Total simulation time		100 ps
Avg. time/step (s)	18.5	18.7

tive change in heat capacity. This change has been attributed to the release of water molecules from the surface of the complex (or removal of solvent accessible surface from bulk water on complexation) accompanied by a stiffening of molecular vibrations at the interface (Ha et al., 1989; Livingston et al., 1988; Spolar and Record, 1994).

The evolution of buried accessible surface for monomer 1 in both simulated systems was calculated, and the dynamic behavior of this parameter is shown in Fig. 4. As seen in the graph, the accessible surface, buried by the monomer facing the nonconsensus DNA sequence in the *ER-G/ERE* system, decreases during the simulation time, which correlates well with the fact that there are two more water molecules that intrude into the protein–DNA interface as well as with the slight movement of the monomer facing the nonspecific DNA sequence, out of the major groove. However, the number of water molecules and the change in the buried

accessible surface cannot completely account for the large decrease in binding affinity of the protein for this particular DNA sequence.

By computing the Debye-Waller factors (B factors or temperature factors) for all the residues of the protein and bases of the DNA, we try to correlate our results with the experimental results of X-ray crystallography. The crystallographically determined Debye-Waller factors were taken from the latest refinement of the ER DBD–DNA structure (pdb entry 1hcq). The values for the experimentally determined and theoretically calculated B factors of the individual residues of monomer 1 in both simulated systems are presented in Fig. 5. A reduction in the value of the B factor for a residue is indicative of a tightening of the structure. The region of most interest is the “reading helix” (helix 1 in Fig. 5), residues 24 to 34. It is apparent that in the specific *ER-ERE* system there are net reductions in the B factors of

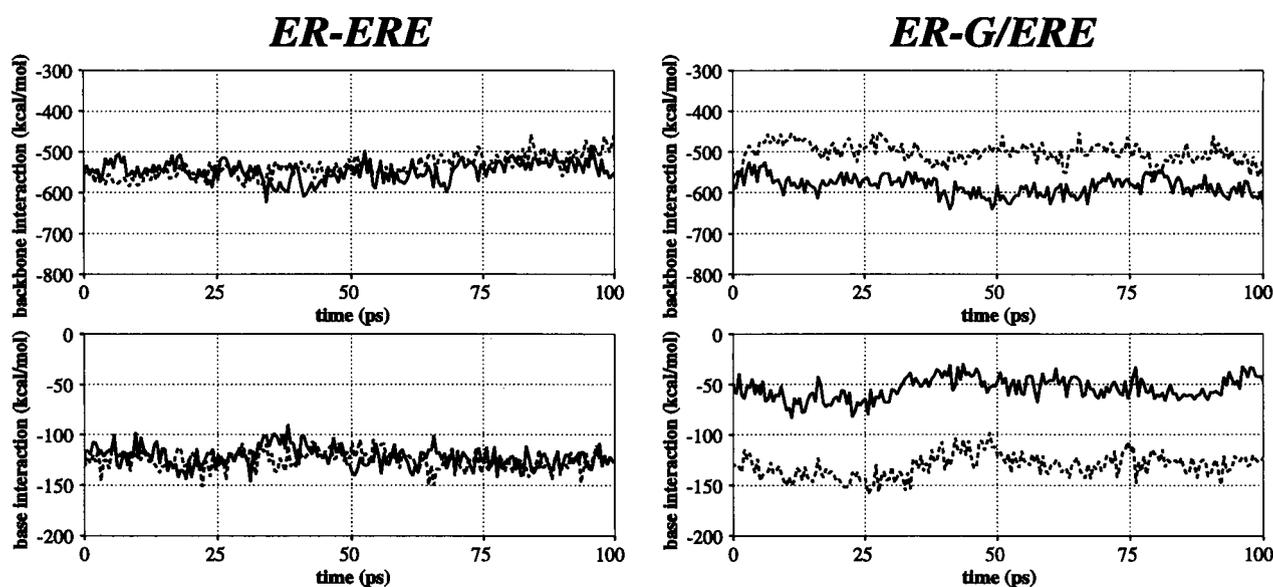


FIGURE 3 Protein–DNA interaction energies in the simulated systems. Solid lines represent data for monomer 1 (the one facing the nonspecific DNA half-site in the *ER-G/ERE* system) and dotted lines represent data for monomer 2. The energy was calculated for two categories of interactions: monomers and DNA backbone (backbone interactions) and monomers and DNA bases (base interactions).

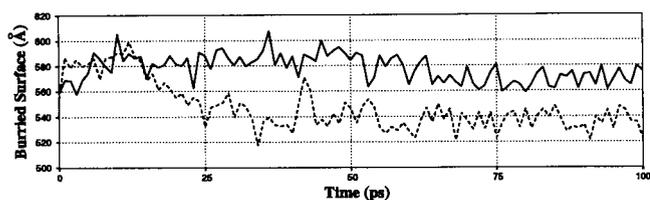


FIGURE 4 Accessible surface area buried by the monomer facing the consensus DNA sequence in the *ER-ERE* system (continuous line) and the nonconsensus DNA sequence in the *ER-G/ERE* system (broken line). The difference between the values for the buried surface area reported in this paper and the ones reported by Gewirth and Sigler (1995) arises from the different ways of calculating solvent-accessible surfaces as well as from the fact that the protein used by Gewirth and Sigler (1995) is a mutant protein (described in text).

residues involved in the interaction with DNA bases: Glu-25, Lys-28, Lys-32, and Arg-33. The restriction of the motion of side chains at the interface will also extend to the water molecules mediating the interaction of these residues with the bases of the DNA. In the nonspecific *ER-G/ERE* system, the four residues mentioned above have higher temperature factors than both in the crystal and the specific *ER-ERE* system. Of these four residues, Lys-28 and Arg-33 have B factors very close to the experimental ones, but Glu-25 and Lys-32 have higher B factor values, which indicate their ability to fluctuate between different contact sites on the DNA. Also, these two residues form a fluctuating network of bonding interactions with DNA and water molecules described in detail below.

For the rest of the protein, the correspondence between the experimental and theoretical values of the B factors suggests that the molecular dynamics trajectories captured some of the qualitative features of the dynamic behavior of the real molecule, of most importance being the correspondence between the high and low mobility regions. The region with the highest B factors in both simulated systems,

as well as in the crystal, is the loop between the reading helix (helix 1) and the D-box of the dimer interface, as well as the short α -helix (helix 2). It was proposed that this short α -helix forms upon binding to DNA (Schwabe et al., 1993a). In the *ER-ERE* system, the B factors for this short α -helix have low values and the helix is quite stable during the simulation time as opposed to the high values of the short α -helix in the *ER-G/ERE* system. The high fluctuations, as well as the beginning of an unwinding of this short α -helix in the nonspecific system, support and give credence to the idea that this region folds upon binding to specific sequences of DNA. Overall, the values for the average B factors correlate well with the secondary structure of the protein. For a detailed discussion on theoretical and experimental B-factors, see Ichiye and Karplus (1988); Post et al. (1989); Brooks et al. (1989); Kuriyan and Weis (1991); Philippopoulos and Lim (1995); and Hunenberger et al. (1995).

Direct and water-mediated hydrogen bonds

The major role in the DNA binding specificity of a protein is played by direct and water-mediated hydrogen bonds that form at the protein-DNA interface. To better understand how ER recognizes different sequences of DNA, we analyzed the pattern of hydrogen bonds at the protein-DNA interface as well as at the dimer interface. Only the hydrogen bonding network for monomer 1 and its DNA half-site are presented for each simulation, since monomer 2, in both systems, faces the same consensus DNA half-site as monomer 1 in the *ER-ERE* system.

Direct contacts

The central elements in the recognition of a consensus DNA sequence by a protein are the hydrogen bond donors and

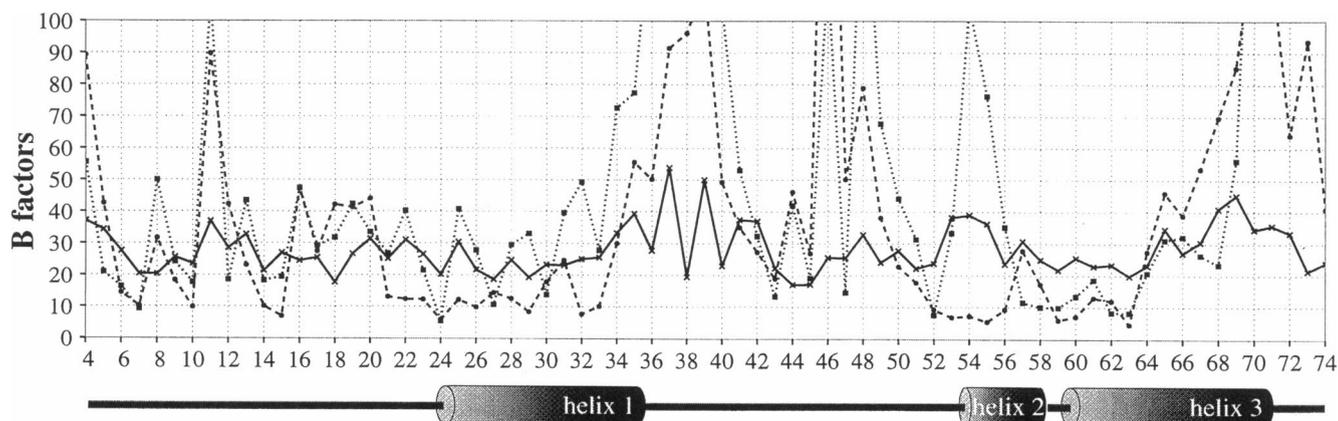


FIGURE 5 Comparison of calculated B factors (in \AA^2) for individual residues (broken line for *ER-ERE* system, dotted line for *ER-G/ERE* system) with experimental values (continuous line) taken from PDB entry 1 hcq. The high temperature factors for the residues 36 to 40 are due to the fact that some of the atoms in the side chains of residues 36, 38, and 40 were missing in the crystal structure and they were modeled. For the rest of the residues the B factor values are discussed in the text. For a discussion on how to interpret the temperature factors see Ichiye and Karplus (1988), Post et al., (1989), Brooks et al. (1989), Kuriyan and Weis (1991); Philippopoulos and Lim (1995), and Hunenberger et al. (1995).

acceptors of the basepairs. Readout of these hydrogen bond recognition interactions could be direct, as well as indirect (involving specific water molecules) and could be facilitated by an induced fit of the DNA or protein that improve the complementarity of the interacting protein and nucleic acid surfaces. It should also be noted that to discriminate between basepairs, a protein does not need to contact both bases in the pair; binding to just one base automatically establishes the other base in conformity with the base-pairing rules.

The direct hydrogen bonds between monomer 1 and the DNA half-site that it binds to are schematically presented in Fig. 6 for both systems. The four residues that play a major role in the binding specificity and stability of the ER are: Glu-25, Lys-28, Lys-32, and Arg-33 (see Figs. 6 and 9). Only one of the three amino acids that form the P-box, Glu-25, makes a direct contact to the DNA. Ala-29, which has only the methyl group in its side chain, can interact strongly only with the T base because it is the only base that has a strongly hydrophobic group. Since the crystallographic study and our simulations indicate that this residue is too distant to make a van der Waals contact with the methyl group of the T bases, another role may be attributed to this residue. Residue Ala-29 may be contributing to specificity not by increasing affinity to specific bases, but by providing effective repulsion to a nonspecific target site

or by sterically permitting another residue, Glu-25 for example, to make a direct side-chain contact with DNA (Zilliaccus et al., 1995). The role of the rest of the residues is discussed separately for each of the simulated systems.

In case of the *ER-ERE* system, the direct contacts between the protein and DNA are stable, and the same as in the crystallographic structure (Schwabe et al., 1993a). Residue Glu-25 has a high affinity for the C4 of C-233, and molecular modeling suggests that this residue has an inhibitory property when the protein binds to a nonconsensus response element (Alroy and Freedman, 1992; Zilliaccus et al., 1995). Residue Lys-28, in addition to making a salt bridge to Glu-25, donates a hydrogen bond to O6 of G-104, Lys-32, which is a residue conserved between different receptors, interacts with the two central bases that are specific to the ERE half-site. In the crystal structure, as well as in our simulation, Lys-32 forms a direct bifurcated hydrogen bond to N7 of G-105 and to O4 of T-106. Residue Arg-33 contacts both a base (N7 of G-231) and a phosphate group through direct hydrogen bonds. All these residues have very low values for the calculated temperature factors, which reflects their stability and tightening during the binding of the protein to the specific DNA sequence.

In the case of the *ER-G/ERE* system the mutation of the two basepairs in the half-site leads predominantly to three changes in the hydrogen bonding pattern at the protein-

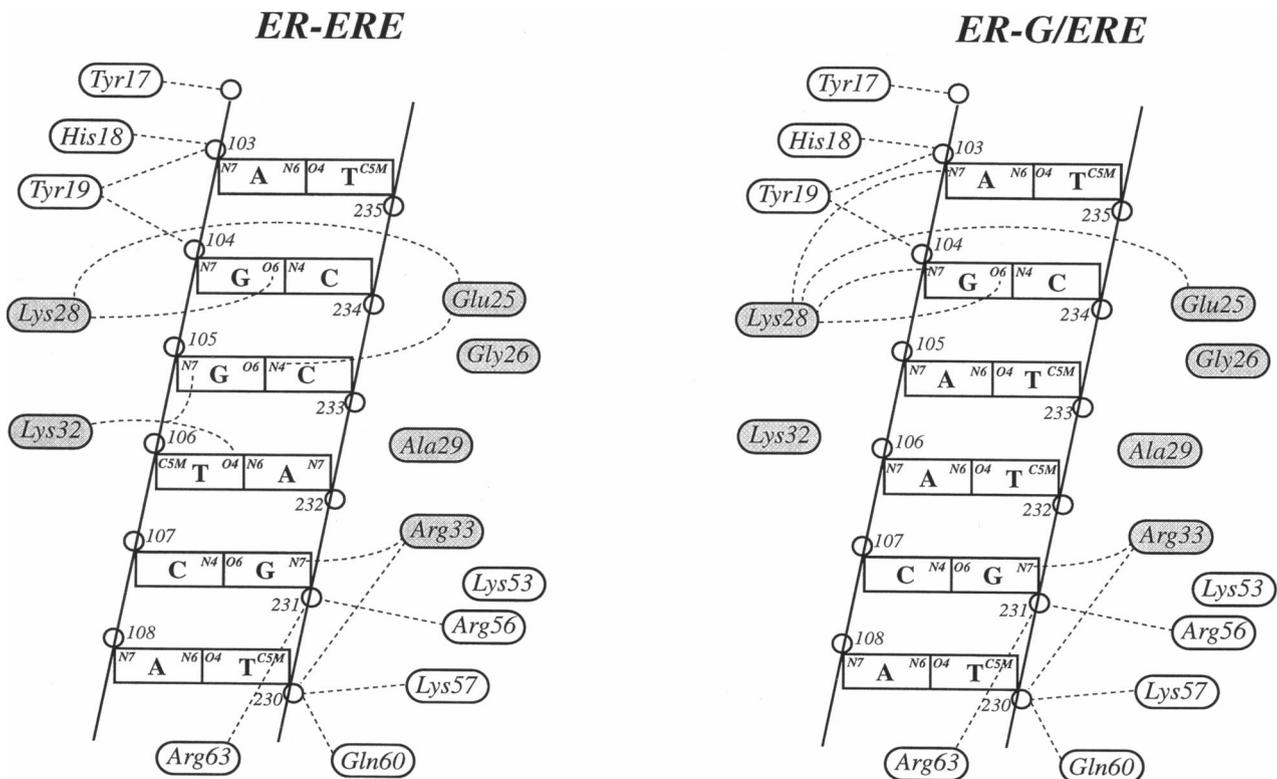


FIGURE 6 Schematic representation of the direct hydrogen bonds in the consensus (left) *ER-ERE* and nonconsensus (right) *ER-G/ERE* systems. The major groove of the DNA is represented as an opened-out helix. The residues that are part of the reading helix are represented by shaded ovals. Bases and sugar rings are represented as boxes and the phosphate backbone by a circle and a solid line.

DNA interface (see Figs. 6 and 9). First, the direct, "specific," hydrogen bond between residue Glu-25 and C4 of C-233 is abolished. Residue Glu-25 and the bulky methyl groups of the two central T bases close the gap (waters cannot enter or escape at the interface) at the amino-terminal end of the reading helix. By losing a direct hydrogen bond, residue Glu-25 becomes more flexible, which is reflected in the high value of the temperature factor for this residue. Also, the interaction energy of this residue with the DNA half-site is much reduced. Second, residue Lys-28 is able to make new hydrogen bonds due to the mutated bases; the basepair mutations alter the conformation of DNA, bending the ds(AA) step through rolling and tilting of the two basepairs. The position of the axis deformation is correlated with the TIP (orientation of the basepairs with respect to the helical axis), ATP (axis tip angle), ROL (roll), and TWS (twist) parameters [for a detailed definition of these parameters see Lavery and Sklenar (1988)]. The time evolution of these parameters in the two simulations is presented in Fig. 7 *a*. It is easy to follow the deviation of

these parameters from the starting structure in the two systems and see that they remain more or less the same in the specific *ER-ERE* system. There is a large increase in the axis tip angle (ATP) for the A106-T232 step to $\sim 15^\circ$ toward the major groove and a big TIP angle of $\sim 8^\circ$ for the A105-T233 step in the *ER-G/ERE* system. A positive value for the roll (ROL) parameter indicates an opening of the angle between basepairs toward the minor groove. In the *ER-G/ERE* system, the roll opens the basepairs step A106-T232 toward the minor groove, and the previous step, A105-T233, toward the major groove. The change in the ROL results in a bend in the DNA helical axis, shown in Fig. 7 *b*. The values for the TWS parameter indicate an unwinding (for values $< 36^\circ$) and an overwinding (for values $> 36^\circ$) at the roll points.

The bending at the ds(AA) site brings the DNA closer to Lys-28 providing additional hydrogen bonds between Lys-28 and the N7 atoms of G-104 and A-103. As seen in Fig. 6, the mutations do not affect the direct hydrogen bonds between Lys-28 and DNA bases that are observed in the crystal structure and are also stable in the *ER-ERE* system. However, the mutations allow the residue to reach and contact the first A base in the DNA half-site. In the non-consensus complex of Schwabe et al. (1995), the Lys-28 residue adopts another conformation due to a steric clash with the mutated basepair which does not arise for the *ER-G/ERE* system. A third change in the nonconsensus *ER-G/ERE* system is the disappearance of the bifurcated hydrogen bond between residue Lys-32 and N7 of G-105 and O4 of T-106. Lysine residues are known to bind to A, T, and G bases; nevertheless, in the present case they bind almost exclusively to G bases, probably due to the electrostatic interaction between the positively charged side-chain of lysines and the negatively charged G bases (Suzuki, 1994). The high temperature factor for Lys-32 residue also reflects the flexibility of this residue that allows it to fluctuate between different positions. The interaction energy of this residue with DNA is lower in the specific *ER-ERE* system than in the nonspecific *ER-G/ERE* system. The position and direct hydrogen bonds for Arg-33, a conserved residue among the receptors, is not perturbed by the mutations induced in the DNA half-site. It should also be noted that the two central A bases that substituted the G-105 and T-106 bases make no direct contacts to any of the protein side chains.

The phosphate backbone of the DNA, on either side of the major groove, is contacted in both systems either through direct hydrogen bonds (see Figs. 6 and 8 *a*) or through water-mediated bridges (see Fig. 9). On one side of the major groove the phosphate backbone is contacted by residues Tyr-17, His-18, and Tyr-19. These residues make direct hydrogen bonds to the phosphate group of A-102, A-103, and G-104 in the crystal structure; hydrogen bonds that are stable during the course of the simulation in both simulated systems. The contacts to phosphates on the other side of the major groove are made by more residues located in the three α -helices of the protein. Residues Arg-56 and

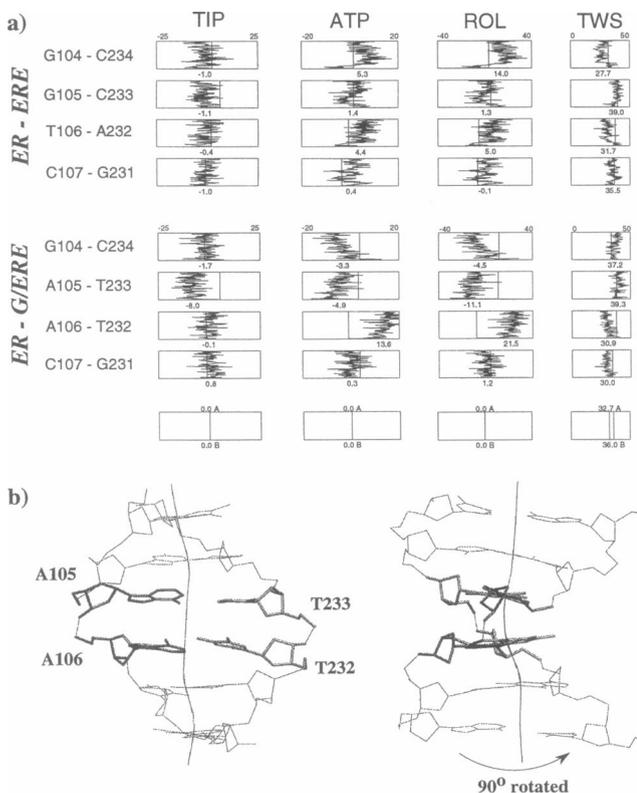


FIGURE 7 *(a)* Time evolution of the position and orientation of the central four basepairs in one half-site relative to the helical axis of the DNA, displayed in "windows" (Lavery and Sklenar, 1988; Ravishanker et al., 1989). The time axis is on the vertical, increasing from bottom to top in these figures. The line through the windows represents the parameters for the starting structure. The numbers under the windows represent the average values of the parameters during the simulations. The windows at the bottom of the figure show the parameters for ideal B-form DNA. *(b)* Snapshot of the DNA half-site in the *ER-G/ERE* system emphasizing the distortions of the two mutated central basepairs A105-T233, A106-T232, and of the DNA axis. Figure created using VMD (Humphrey et al., 1996).

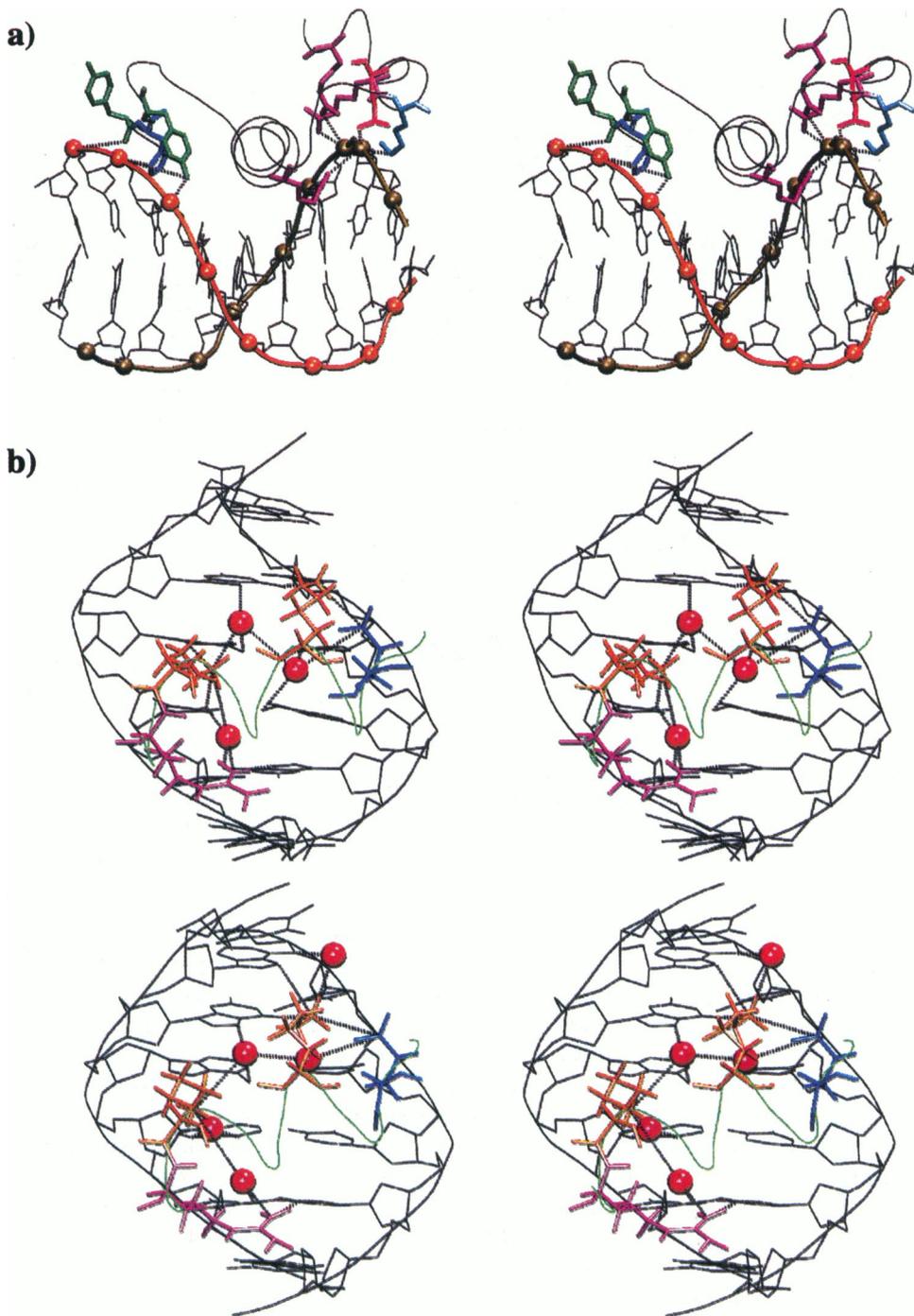


FIGURE 8 (a) Stereo view of the interaction between the protein residues and the phosphodiester backbone of the DNA. Residues Tyr-17, His-18, and Tyr-19 contact the DNA backbone on the left side and residues Arg-56, Lys-57, Gln-60, and Arg-63 contact the DNA backbone on the right side. The DNA backbone is represented as a tube with the phosphorus atoms represented as van der Waals spheres. (b) Top view of the interaction between the DNA half-site and the reading helix in the major groove. The reading helix is drawn as a tube with specific depiction, from left to right, of residues Arg-33, Lys-32, Lys-28, and Glu-25. Water molecules are represented as van der Waals spheres. Only direct and water-mediated hydrogen bonds to the bases of the DNA are presented. Figures created using VMD (Humphrey et al., 1996).

Arg-63 on one side, and Lys-57 and Gln-60 on the other side saddle the DNA backbone, holding it in a strong grip. Direct hydrogen bonds between the protein side chains and the phosphodiester backbone of the DNA are illustrated in Fig. 8 *a*.

Due to this bonding pattern of the phosphate backbone on both sides of the major groove, the protein can measure the width of the major groove. Any change in the groove width should be detected and should influence the binding of the protein.

Water-mediated contacts

The water-mediated hydrogen bonds between monomer 1 and the DNA half-site are schematically presented in Fig. 9 for both simulated systems. The number of water molecules that bridge between DNA bases and protein residue side chains is different in the two systems. Two more water molecules make their way into the major groove of the nonconsensus system *ER-G/ERE*, and it is tempting to speculate that a longer simulation time may allow more water

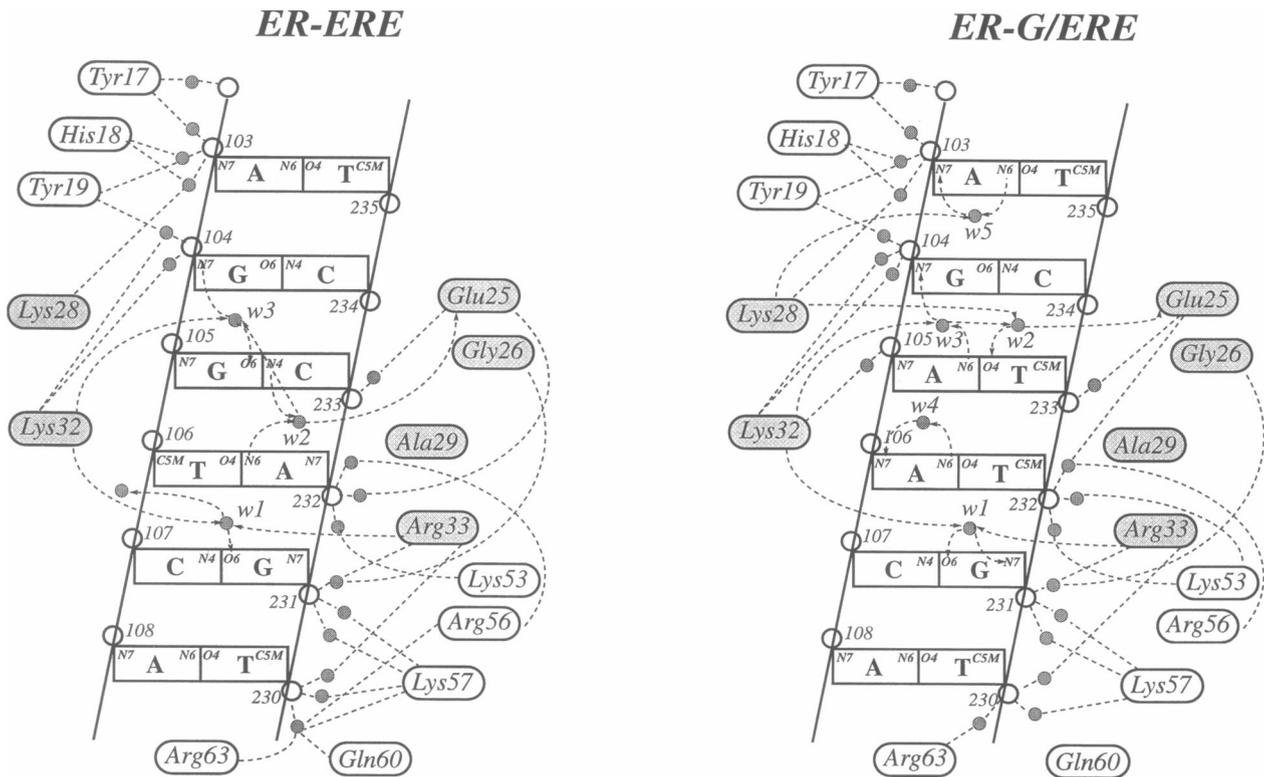


FIGURE 9 Schematic representation of water-mediated hydrogen bonds in the consensus (*left*) *ER-ERE* and nonconsensus (*right*) *ER-G/ERE* systems. Ordered water molecules at the protein–DNA interface are shown as dark circles. The bridging waters between protein side chains and DNA bases are numbered w1 to w5. The direction of the hydrogen bonds made with the bases of the DNA is indicated by arrows going from the donor to the acceptor atoms. All the interactions indicated in this figure do not occur simultaneously, but they form a network that fluctuates with time.

molecules to enter the major groove at the protein–DNA interface.

There are three water molecules that facilitate interactions between protein residues and DNA bases in the crystal structure (Schwabe et al., 1993a). Water molecule w1, which is fully coordinated by its contact with Lys-28, Arg-33, and O6 of G-231, continues to occupy the same bridging position as in the crystal structure during the entire simulation time in both systems. Water molecule w2, which extended the reach of Glu-25 residue to the N6 of A-232 and N4 of C-233 in the crystal and the *ER-ERE* system is displaced by the basepair substitution in the *ER-G/ERE* system, being pushed over the C-233 base toward the C-234 base (see Fig. 9). The orientation of this water molecule is fixed by its interaction with another water molecule, w3, that bridges between Lys-32 and the G-104, A-105 bases. In both systems this water, which allows Lys-32 to contact the central two basepairs in the crystal structure (Schwabe et al., 1993a), moves over the 105 base toward the G-104 base, extending the reach of Lys-32 to N7 of G-104 (see Fig. 9). By doing so, the direction of the hydrogen bond between the two waters changes such that the hydrogen bond balance for the water molecule is still maintained (w3 becomes the donor and w2 becomes the acceptor).

As mentioned above, there are two more water molecules that make their way into the major groove of the DNA and

contact the bases. One water molecule, w5, bridges between Lys-28 and N6 of A-103, thus increasing the recognition of the first basepair in the half-site. The other water molecule, w4, fills the void left by the mutation of the two central basepairs in the *ER-G/ERE* system and the rearrangement of residue Lys-32. This water molecule maintains a stable bond between N7 and N6 of A-106. Experimental studies of the spatial distribution of waters around the DNA spine of hydration (Kim et al., 1993; Feng et al., 1994; Shaked et al., 1994). This water molecule makes it impossible for the Lys-32 residue to contact the A bases. All the water molecules and protein side chains making hydrogen bonds to the bases of the DNA in the two simulated systems are presented in Fig. 8 b.

Water molecules are also seen to cluster around the phosphate backbone of the DNA in both simulated systems. On both sides of the major groove, the number of water molecules bridging between protein residues and DNA phosphates is greater in the nonconsensus *ER-G/ERE* system than in the consensus one. The positioning of these water molecules may be facilitated by the intrusion of the two water molecules into the binding interface between the protein and DNA bases that leads to a looser contact.

A slight movement of the monomer, facing the nonconsensus half-site, out of the major groove is revealed in Fig.

10. The slight movement may account for new water mediated hydrogen bonds that develop between some residues and the corresponding phosphate groups, as well as for the decrease in the buried accessible surface area.

The crystal structure of an estrogen receptorlike DNA binding domain bound to the GRE (Gewirth and Sigler, 1995) revealed an internal cavity at the protein–DNA interface filled with seven water molecules. Five of these water molecules bridge between the protein residues and the two central DNA bases in the half-site recognized by the monomer. The estrogen receptorlike DNA binding domain that Gewirth and Sigler (1995) used in their crystallographic study is a mutationally altered GR DNA binding domain with the P-box of an estrogen receptor and the dimer interface, D-box, of a thyroid receptor. They found that the internal cavity is created by a lack of conformational adjustments by the DNA to close the gap. In our study we also find five water molecules that bridge between the DNA bases and the protein in the altered half-site. However, these water molecules exhibit different placements than those in the crystal structure of Gewirth and Sigler (1995). Also, in the *ER-G/ERE* simulated system the two central basepairs are not directly contacted by the protein but they may, nonetheless, affect the stability of the complex through their effect on the DNA structure. These two basepairs are contacted by five water molecules in Gewirth and Sigler's (1995) structure. The difference in the number and location of the water molecules, as well as in the DNA conformation, between the crystallographic study and our theoretical study may be due to the short simulation time or to the use of a hybrid protein in the crystallographic study. It should also be noted that the amino acids within the P-box of the ER are not in the same relative spatial location as the P-box residues in the GR (Schwabe et al., 1993a).

Dimer contacts

The dimer interface between the two monomers in the protein–DNA complex is formed by the region within the DBD that was found to be distorted in solution (Schwabe et al., 1990). The contacts in the dimer interface are the ones

that impose a requirement of a spacer of three basepairs between the half-sites for the protein to bind DNA. These contacts also assist the protein in binding to nonconsensus sequences of DNA. The strength and stability of these contacts are demonstrated in the crystal structure of the glucocorticoid receptor DBD bound to a response element with four basepairs between the two half-sites. The dimer interaction is so strong that even when it is bound to a DNA sequence with four basepairs between the two consensus half-sites, the dimer is not pulled apart by specific DNA interactions of the monomers (Luisi et al., 1991).

The first part of the dimer interface, Met-42 to Cys-49, is the region with the most contacts between the two monomers. Pro-44 and Ala-45 residues make van der Waals contacts with the same residues in the other monomer. The water molecule that bridges between the Thr-46 residues in the crystal structure relocates during the simulation time, in both simulated systems, to form a bridge between the Ser-58 residues. This leads to a high fluctuation of the Thr-46 residue, reflected in the high values for the B factor of this residue. The two water molecules that bridge residue Met-42 of monomer 1 and residue Ser-58 of monomer 2 (and conversely) remain at the same bridging position during the entire course of the simulation of both systems.

The second part of the dimer interface, Thr-50 to Ser-58, comprises the short α -helix present only in dimer A of the crystal structure (see Methods). This short α -helix has very high temperature factors (see Fig. 5) as a consequence of an unwinding event that can be observed in the nonspecific *ER-G/ERE* system. A direct, symmetrical hydrogen bond between Ser-58 residues, located at the carboxy terminus of the short α -helix, develops during the course of the simulation in both simulated systems. The symmetrical hydrogen bonds between Pro-44 residue in monomer 1 and Thr-50 residue in monomer 2 (and conversely) remain strong during almost the entire simulation time. Also, new symmetric hydrogen bonds develop between residues Cys-43 and Arg-55 of each monomer.

A few interstitial water molecules are found at the dimer interface. There are many water molecules that hydrogen-bond to residues in the dimer interface, since on the surface of the protein water molecules are more easily exchanged with the surrounding bulk water than at the more buried dimer interface. However, only two (in the *ER-G/ERE* system) and one (in the *ER-ERE* system) new water molecules were found to make stable bridges during the simulation. In the *ER-ERE* system one water bridges between Ala-45 of monomer 1 and Thr-50 of monomer 2 and the other water molecule bridges between Asn-54 of monomer 1 and Met-42 of monomer 2. The new water found in the *ER-G/ERE* system bridges between Ala-45 residue of monomer 1 and Arg-55 residue of monomer 2. A representation of all the residues and water molecules that interact at the dimer interface is provided in Fig. 11.

It should also be mentioned that there are several residues in the second part of the dimer interface contacting the phosphate backbone of DNA through direct or water-mediated

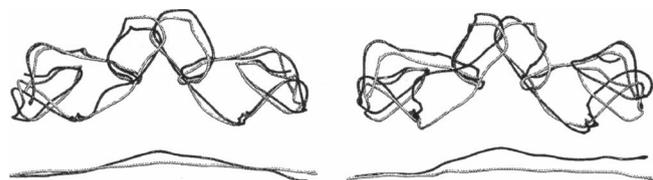


FIGURE 10 String representation of the simulated protein–DNA systems (left: *ER-ERE* system; right: *ER-G/ERE* system). The molecular axis for protein and DNA were determined using “Molecular Dynamics Analysis Toolchest” (Sklenar et al., 1989; Ravishanker et al., 1989) and displayed using the visualization program VMD (Humphrey et al., 1996). The initial structure is colored gray and the final structure (after 100 ps) is colored black. In each structure monomer 1 (that faces the nonspecific half-site) appears on the left and monomer 2 appears on the right side.

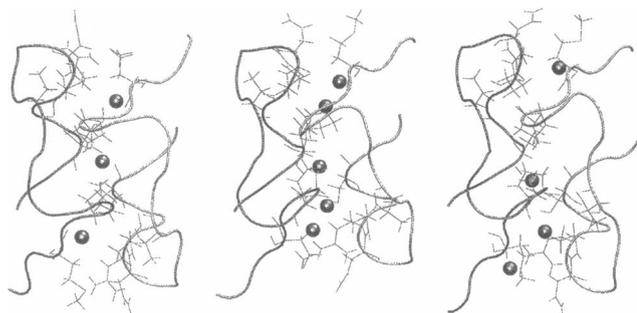


FIGURE 11 Structure of the dimer interface in the crystal structure (*left*), *ER-ERE* system (*center*) and *ER-G/ERE* system (*right*) viewed from the top as defined in Fig. 3. The symmetry of the interactions between monomers is broken in the *ER-G/ERE* system (*right*). Figure created using VMD (Humphrey et al., 1996).

ated hydrogen bonds. Contacts made by these residues, Lys-53, Arg-56, and Lys-57, are presented in Figs. 6 and 9. Thus, it may be stated that binding to the DNA not only correctly aligns the dimer interface, but it also helps to stabilize it.

CONCLUSIONS

Recognition of DNA sequences requires the formation of specific contacts between the protein and the DNA. Multiple specific protein–DNA base contacts, both direct or water mediated, should establish a DNA sequence specificity for the steroid hormone receptors. These proteins exhibit, however, a flexibility in recognizing DNA sequences and also accept a variety of amino acid substitutions in their reading helix without abolishing binding (Zilliaccus et al., 1995). How can this flexibility in DNA sequence recognition be correlated with the specific base contacts seen in the crystal structures?

The results of our study indicate that the binding specificity of the ER-DBD to different sequences of DNA is mostly determined by three factors:

1) *The surface distribution of donor and acceptor groups for the hydrogen bonds at the protein–DNA interface.* When the donor and acceptor groups are altered, the side chains of the residues reaching into the major groove of the DNA rearrange themselves. There are approximately the same number of hydrogen bonds at the protein–DNA interface in the two simulated systems, but a significant difference in the interaction energy of protein–DNA base contacts establishes itself between the consensus and nonconsensus complexes that may arise from the less favorable geometry of the hydrogen bonds in the mutated structure. Glu-25 and Lys-32 side chains rearrange as a result of the mutation; both make new contacts with the bases or the phosphodiester backbone of the DNA. Also, the two residues form a fluctuating network of hydrogen bonds that may lead to a lower stability of the protein–DNA system.

2) *The number and position of the water molecules found at the protein–DNA interface.* The small size of the water

molecule and its ability to form up to four hydrogen bonds allows it to fit in some of the gaps at the protein–DNA interface as well as to satisfy some of the hydrogen donor or acceptor groups of the protein or DNA. The relocation of the water molecules that extend the reach of Lys-32 and Glu-25, and also bond together, illustrates best how easily the protein adapts its hydrogen bonding network to the changes in DNA sequence. The direction of the hydrogen bond between the two water molecules changes in the two systems in such a way that the hydrogen balance for the water molecules is maintained. The gap left by the movement of the Lys-32 side chain, as it rearranges itself as a result of the mutations, is filled by a water molecule that is well positioned, bridging between the N7 and N6 groups of the A106 base.

3) *The altered conformation of the nonspecific half-site imposed by the protein.* The mutated basepairs ds(AA) in the *ER-G/ERE* system alter the conformation of the DNA through their roll and tilt angles. These deformations allow residue Lys-28 to make new contacts with the first adenine and second guanine base in the half-site, thus forcing the DNA to maintain this conformation. The axis of the mutated DNA half-site in the *ER-G/ERE* system is bent as a consequence of new roll and tilt angles of the two central basepairs.

We conclude that the marked weakening (or instability) of the hydrogen bonding network at the protein–DNA interface, as well as the inadequacy of the protein to expel the fixed water molecules from the interface, may reduce the affinity of the estrogen receptor protein for the nonspecific DNA sequences. Together with previous data (Zilliaccus et al., 1995; Schwabe et al., 1995; Gewirth and Sigler, 1995) our results contribute to a better understanding of the role played by the residue side chains and water molecules in the DNA binding specificity of the estrogen receptor protein. Specificity is achieved through a delicate balance of energetic contributions, where a small difference in the energy required to rearrange some of the residues or to bend DNA leads to significant differences in the binding specificity of ER. The surface of the DNA is defined by sequence-specific variations in the functional groups of the bases as well as by the local deformability of the DNA. The binding protein has to complement this surface and water molecules are proven to be the suitable intermediary agents that help the protein to match its surface to the surface of the DNA.

The authors thank J. Schwabe for making the coordinates of the crystallographic structure of the estrogen receptor DNA binding domain dimer–DNA complex available to us (Schwabe et al., 1993a); R. Skeel, M. Nelson, R. Brunner, and A. Gursoy for help with NAMD, and Alexander Balaeff for helpful discussions.

This work was supported by National Institutes of Health Grant PHS 5 P41 RR05969-04, the National Science Foundation Grant BIR-9318159, the MCA 93S028P computertime Grant at Pittsburgh Supercomputing Center, and the Roy J. Carver Charitable Trust.

REFERENCES

- Alroy, I., and L. P. Freedman. 1992. DNA binding analysis of glucocorticoid receptor specificity mutants. *Nucleic Acids Res.* 20:1045–1052.
- Beveridge, D. L., and G. Ravishanker. 1994. Molecular dynamics studies of DNA. *Curr. Opin. Struct. Biol.* 4:246–255.
- Biesiadecki, J. J., and R. D. Skeel. 1993. Danger of multiple-time-step methods. *J. Comput. Phys.* 109:318–328.
- Bishop, T. C., D. Kosztin, and K. Schulten. 1997. How hormone receptor-DNA binding affects nucleosomal DNA. The role of symmetry. *Bio-phys. J.* 72:2056–2067.
- Bishop, T., and K. Schulten. 1996. Molecular dynamics study of glucocorticoid receptor-DNA binding. *Proteins Struct. Funct. Genet.* 24: 115–133.
- Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.
- Brooks, C. L., M. Karplus, and B. M. Pettit. 1989. Proteins: a theoretical perspective of dynamics, structure and thermodynamics. In *Advances in Chemical Physics*, Wiley, New York. 71.
- Brünger, A. T. 1992. X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University.
- Danielsen, M., L. Hinck, and G. M. Ringold. 1989. Two amino acids within the knuckle of the first zinc finger specify DNA response element activation by the glucocorticoid receptor. *Cell.* 57:1131–1138.
- Eriksson, M., T. Härd, and L. Nilsson. 1995. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.* 68:402–426.
- Eriksson, M. A. L., and L. Nilsson. 1995. Structure, thermodynamics, and cooperativity of the glucocorticoid receptor DNA-DNA-binding domain in complex with different response elements. Molecular dynamics simulation and free energy perturbation study. *J. Mol. Biol.* 253:453–472.
- Evans, R. M. 1988. The steroid and thyroid hormone receptor superfamily. *Science.* 240:889–895.
- Feng, J., 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.
- Forester, T. R., and I. R. McDonald. 1991. Molecular dynamics studies of the behaviour of water molecules and small ions in concentrated solutions of polymeric B-DNA. *Mol. Phys.* 72:643–660.
- Freedman, L. P., and B. F. Luisi. 1993. On the mechanism of DNA binding by nuclear hormone receptors: a structural and functional perspective. *J. Cell. Biochem.* 51:140–150.
- Gewirth, D. T., and P. Sigler. 1995. The basis for half-site specificity explored through a non-cognate steroid receptor-DNA complex. *Nat. Struct. Biol.* 2:386–394.
- Green, S., and P. Chambon. 1987. Oestradiol induction of a glucocorticoid-responsive gene by a chimaeric receptor. *Nature.* 325:75–78.
- Green, S., V. Kumar, I. Theulaz, W. Wahli, and P. Chambon. 1988. The n-terminal DNA binding “zinc-finger” of the oestrogen and glucocorticoid receptors determines target gene specificity. *EMBO J.* 7:3037–3044.
- Ha, J.-H., M. W. Capp, M. D. Hohenwarter, W. Baskerville, and M. T. Record. 1992. Thermodynamic stoichiometries of participation of water, cations, and anions in specific and non-specific binding of *lac* repressor to DNA. *J. Mol. Biol.* 228:252–264.
- Ha, J.-H., R. S. Spolar, and M. T. Record. 1989. Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J. Mol. Biol.* 209:801–816.
- Harris, L. F., M. R. Sullivan, P. D. Popken-Harris, and D. F. Hickok. 1994. Molecular dynamics simulations in solvent of the glucocorticoid receptor protein in complex with a glucocorticoid response element DNA sequence. *J. Biomed. Struct. Dyn.* 12:249–270.
- Humphrey, W. F., A. Dalke, and K. Schulten. 1996. VMD—Visual molecular dynamics. *J. Mol. Graphics.* 14:33–38.
- Hunenberger, P. H., A. E. Mark, and W. F. van Gunsteren. 1995. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* 252:175–194.
- Ichiye, T., and M. Karplus. 1988. Anisotropy and anharmonicity of atomic fluctuations in proteins: implications for x-ray analysis. *Biochemistry.* 27:3487–3497.
- Jayaram, B., N. Aneja, E. Rajasekaran, V. Arora, A. Das, V. Ranganathan, and V. Gupta. 1994. Modeling DNA in aqueous solutions. *J. Sci. Indust. Res.* 53:88–105.
- Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
- Kabsch, W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A.* 32:922–923.
- Kim, J. L., D. B. Nikolov, and S. K. Burley. 1993. Cocystal structure of TBP recognizing the minor groove of a TATA element. *Nature.* 365: 520–527.
- Kraulis, P. 1991. MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.
- Krust, A., S. Green, P. Argos, V. Kumar, P. Walter, J. Bornert, and P. Chambon. 1986. The chicken oestrogen receptor sequence: homology with *v-erbA* and the human oestrogen and glucocorticoid receptors. *EMBO J.* 5:891–897.
- Kumar, V., and P. Chambon. 1988. The oestrogen receptor binds tightly to its responsive element as a ligand induced homodimer. *Cell.* 55: 145–156.
- Kumar, V., S. Green, G. Stack, M. Berry, J. R. Jin, and P. Chambon. 1987. Functional domains of the human estrogen receptor. *Cell.* 51:941–951.
- Kuriyan, J., and W. I. Weis. 1991. Rigid protein motion as a model for crystallographic temperature factors. *Proc. Natl. Acad. Sci. USA.* 88: 2773–2777.
- Laudet, V., C. Hänni, J. Coll, F. Catzeflis, and D. Stéhelin. 1992. Evolution of the nuclear receptor gene superfamily. *EMBO J.* 11:1003–1013.
- 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.
- Livingstone, J. R., R. S. Spolar, and M. T. Record. 1988. Contribution of the thermodynamics of protein folding from the reduction in water-accessible non-polar surface area. *Biochemistry.* 30:4237–4244.
- Luisi, B. F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, and P. B. Sigler. 1991. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature.* 352:497–505.
- Lundbäck, T., C. Cairns, J. A. Gustafsson, J. Carlstedt-Duke, and T. Härd. 1993. Thermodynamics of the glucocorticoid receptor-DNA interaction: binding of wild-type GR DBD to different response elements. *Biochem. J.* 32:5074–5082.
- Lundbäck, T., J. Zilliacus, J. A. Gustafsson, J. Carlstedt-Duke, and T. Härd. 1994. Thermodynamics of sequence-specific glucocorticoid receptor-DNA interactions. *Biochem. J.* 33:5955–5965.
- MacKerell, Jr., A., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975.
- Mader, S., V. Kumar, H. de Verneuil, and P. Chambon. 1989. Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid-responsive element. *Nature.* 338: 271–274.
- Miaskiewicz, K., R. Osman, and H. Weinstein. 1993. Molecular dynamics simulation of the hydrated d(CGCGAATTCGCG)₂ dodecamer. *J. Am. Chem. Soc.* 115:1526–1537.
- Nelson, M., W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. D. Skeel, and K. Schulten. 1996. NAMD—A parallel, object-oriented molecular dynamics program. *J. Supercomput. App.* 10:251–268.
- Parker, M. G., editor. 1991. Nuclear Hormone Receptors. Academic Press, San Diego, CA.
- Philippopoulos, M., and C. Lim. 1995. Molecular dynamics simulation of *E. coli* ribonuclease h₁ in solution: correlation with NMR and x-ray data and insights into biological function. *J. Mol. Biol.* 254:771–792.
- Post, C. B., C. M. Dobson, and M. Karplus. 1989. A molecular dynamics analysis of protein structural elements. *Proteins Struct. Funct. Genet.* 5:337–354.
- Prevost, C., S. Louise-May, G. Ravishanker, R. Lavery, and D. L. Beveridge. 1993. Persistence analysis of the static and dynamical helix deformations of DNA oligonucleotides: applications to the crystal structure

- and molecular dynamics simulation of d(CGCGAATTCGCG)₂. *Biopolymers*. 33:335–350.
- Rankin, W., and J. Board. 1995. A portable distributed implementation of the parallel multipole tree algorithm. *IEEE Symposium on High Performance Distributed Computing*. In press. [Duke University Technical Report 95-002].
- Rastinejad, F., T. Perlmann, R. M. Evans, and P. Sigler. 1995. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature*. 375:203–211.
- Ravishanker, G., S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar. 1989. Conformational and helicoidal analysis of 30 psec of molecular dynamics on the d(CGCGAATTCGCG) double helix. *J. Biomol. Struct. & Dyn.* 6:669–699.
- Rodgers, D. W., and S. C. Harrison. 1993. The complex between phage 434 repressor DNA-binding domain and operator site O_R3: structural differences between consensus and non-consensus half-sites. *Structure*. 1:227–240.
- Schneider, B., D. M. Cohen, L. Schleifer, A. R. Srinivasan, W. Olson, and H. M. Berman. 1993. A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. *Biophys. J.* 65:2291–2303.
- Schwabe, J. W. R., L. Chapman, J. T. Finch, and D. Rhodes. 1993a. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*. 75:567–578.
- Schwabe, J. W. R., L. C. Chapman, J. T. Finch, D. Rhodes, and D. Neuhaus. 1993b. DNA recognition by the estrogen receptor: from solution to the crystal. *Structure*. 15:187–204.
- Schwabe, J. W. R., L. C. Chapman, and D. Rhodes. 1995. The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure*. 3:201–213.
- Schwabe, J. W. R., D. Neuhaus, and D. Rhodes. 1990. Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature*. 348:458–461.
- Seibel, G. L., U. C. Singh, and P. A. Kollman. 1985. A molecular dynamics simulation of double-helical B-DNA including counterions and water. *Proc. Natl. Acad. Sci. USA*. 82:6537–6540.
- Shaked, Z., G. Guzikovich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P. B. Sigler. 1994. Determinants of repressor/operator recognition from the structure of the *trp* operator binding site. *Nature*. 368:469–473.
- Sharp, K. A., and B. Honig. 1995. Salt effects on nucleic acids. *Curr. Opin. Struct. Biol.* 5:323–328.
- Sklenar, H., C. Etchebest, and R. Lavery. 1989. The definition of generalized helicoidal parameters and of axis of curvature for irregular nucleic acids. *Proteins Struct. Funct. Genet.* 6:46–60.
- Spolar, R. S., and M. T. Record. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science*. 263:777–784.
- Suzuki, M. 1994. A framework for DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure*. 2:317–326.
- Umesono, K., and R. M. Evans. 1989. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell*. 57:1139–1146.
- York, D. M., T. Darden, and D. Deerfield II. 1992. The interaction of Na(I), Ca(II), and Mg(II) metal ions with duplex DNA: a theoretical modeling study. *Int. J. Quantum Chem. Symp.* 19:145–166.
- Zilliacus, J., A. P. H. Wright, J. Carlstedt-Duke, L. Nilsson, and J. Gustafsson. 1995. Modulation of DNA-binding specificity within the nuclear receptor family by substitutions at a single amino acid position. *Proteins Struct. Funct. Genet.* 21:57–67.