

Structure, Interaction, Dynamics and Solvent Effects on the DNA-EcoRI complex in Aqueous Solution from Molecular Dynamics Simulation

Srikanta Sen and Lennart Nilsson

Center for Structural Biochemistry, Karolinska Institute, Department of Biosciences, Huddinge, Sweden

ABSTRACT A 0.7-ns molecular dynamics simulation of the DNA-EcoRI complex in a 7.0-Å solvent shell indicated a stable behavior of the system. No significant evaporation or smearing of the solvent's outer boundary occurred. The structure and the intermolecular interactions were found to be well maintained during the simulation. The interaction pattern in the simulation was found to be very similar to that in the crystal structure. Most of the specific interactions between the DNA and the protein were found to be enhanced in the simulation compared to that in the crystal structure as a result of improved interaction geometry. The nonspecific interactions were found to be stronger than the specific ones. The specific interactions between the N7 atoms of Gua⁴ or Ade⁵ or Ade⁶ and the protein were found to be present over almost the entire time of the simulation, whereas hydrogen bonds involving the amino groups of the Ade⁵ and Ade⁶ with the protein were found to be relatively weaker, with lower probability and shorter lifetime. The time evolution of the root mean square deviations of the DNA and the protein were highly correlated even at the later part of the simulation, showing the tight binding between them. Several long-lived water bridges were found between the DNA backbone atoms and the protein and also between the two protein monomers, which increased the overall stability of the complex. The two protein monomers were found to interact strongly with each other. The energy of the DNA kink deformation was estimated as approximately 31 kcal/mol.

INTRODUCTION

One of the central issues of modern molecular biology and biophysics is to understand the interactions stabilizing biomolecular complexes in aqueous solution, particularly how a small sequence change can lead to a significant difference in affinity, either directly or through structural rearrangements. Knowledge of these aspects is essential for designing simple artificial biomolecules having desired properties, which can be used for controlling various cellular processes. From the biological point of view, DNA-protein interactions provide the most important class of biomolecular complexes for studying such issues. There has been considerable experimental work in this field (Draper, 1993; Rosenberg, 1991; Kim et al., 1990; Luisi et al., 1991; Newman et al., 1994; Lesser et al., 1990, 1993; Robinson and Sligar, 1993, 1994; Billeter et al., 1993; Qian et al., 1993; Schwabe et al., 1993; Venclovas et al., 1994; Hirsch and Aggarwal, 1995; Eriksson et al., 1995; Szezelkun et al., 1997; Jeltsch et al., 1994; Misra et al., 1994). Although characterization of the interactions between the DNA and the protein in a DNA-protein complex is possible through x-ray crystallography, nuclear magnetic resonance (NMR), or some indirect biochemical experiments, the detailed characterization of the interaction strength and the dynamic features at the atomic level in such cases has not yet proved to be very straightforwardly obtained from experiments. Molecular modeling and molecular dynamics (MD) simulation tech-

niques, through which individual interactions and dynamics at the atomic level can be probed explicitly, provide another way to look into such details of molecular structure and interactions (McCammon and Harvey, 1987; Brooks et al., 1989; van Gunsteren, 1993). Moreover, as most biomolecular interactions take place in an aqueous environment, it is also very important to consider the effects of aqueous solvent, including, among others, competition for hydrogen bonding, screening of electrostatic interactions, and favoring structural rearrangements with minimal exposure of hydrophobic groups to aqueous solvent. There are also cases where water molecules have been seen to supply crucial and specific contacts between protein and DNA (Billeter et al., 1996; Otwinoski et al., 1988; Qian et al., 1993; Karplus and Fearman, 1994; Eriksson et al., 1995; Eriksson and Nilsson, 1995; Kosztin et al., 1997). Again, it is not always possible to characterize the solvent molecules in terms of localization and dynamics in x-ray or NMR experiments. On the other hand, the individual solvent molecules can be followed rather easily in MD simulation studies and can be characterized in full detail (Eriksson et al., 1995; Eriksson and Nilsson, 1995; Kosztin et al., 1997; Alexander et al., 1998). Thus, the results from MD simulations can complement the corresponding experimental studies in improving our insights about structure, dynamics, interactions, and solvent effects involving biomolecular complexes at the atomic level. The aim of the present study is to use molecular modeling and MD simulation techniques for investigating the details of the crucial interactions (both direct and water-mediated) that are responsible for the stability and high specificity in the DNA-EcoRI complex and its dynamics in aqueous solution. Several works (Eriksson et al., 1995; Eriksson and Nilsson, 1995; Bishop et al., 1996, 1997; Kosztin et al., 1997) in this direction involving other

Received for publication 4 December 1998 and in final form 13 July 1999.

Address reprint requests to Lennart Nilsson, Center for Structural Biochemistry, Karolinska Institute, Department of Biosciences, Hälsovägen 7, Floor 7, S-141 57 Huddinge, Sweden. Tel.: 46-8-608-9228; Fax: 46-8-608-9290; E-mail: lennart.nilsson@csb.ki.se.

© 1999 by the Biophysical Society

0006-3495/99/10/1782/19 \$2.00

DNA-protein complexes have been published in recent years. In our present work we have selected DNA-*EcoRI* complex as a model system for such studies because a large body of experimental biochemical and biophysical data is readily available (Draper, 1993; Rosenberg, 1991; Kim et al., 1990; Newman et al., 1994; Lesser et al., 1990, 1993; Robinson and Sligar, 1993, 1994; Venclovas et al., 1994; Misra et al., 1994; Szczelkun et al., 1997; Jeltsch et al., 1994) for this system, which can be of great use in evaluating the reliability of the results obtained by the present modeling and simulation studies.

EcoRI is a well known type II restriction endonuclease that binds to DNA, specifically to the base sequence containing the stretch $d(G^{\downarrow}AATTC)_2$, and cuts the DNA strands at the position marked by the arrow by hydrolyzing the sugar-phosphate backbone in the presence of Mg^{+2} ion as a cofactor. Endonucleases are unique in that they can cleave efficiently small (4–6 bp long) recognition sites on foreign DNA but avoid potential cleavage of the cellular DNA at sites that differ by as little as a single basepair. Fig. 1 represents the crystallographic structure of the DNA-*EcoRI* complex. *EcoRI* is a dimeric protein. Each monomer contains 277 amino acid residues and binds to the above base sequence on each DNA strand to form the complex. Binding of the protein monomers introduces bends in the structures of both the DNA strands, making the DNA kinked (Rosenberg, 1991; Kim et al., 1990). Protein-DNA interactions are often classified into specific and nonspecific contacts (Billeter, 1996). The interactions between the protein and the DNA bases are known as the specific interac-

tions, and those formed between the protein and the DNA backbone are termed nonspecific interactions because the sugar-phosphate backbone parts are the same for all DNA nucleotides. Clearly, the specific contacts play the major role in the selection of a certain DNA base sequence, whereas the nonspecific ones serve mostly to increase the overall stability of the complex. Structural studies on the DNA-*EcoRI* complex by crystallography also indicate that the sequence specificity is determined in part by the direct contacts between the protein and the DNA bases. It is found that both the strands are recognized by hydrogen bonds with each purine base and contacts to the pyrimidines and the tight complementarity of the surfaces of the DNA and the protein at the recognition site also provide extensive interactions between the protein and the DNA backbone which further increases the overall stability of the complex (Rosenberg, 1991; Kim et al., 1990; Newman et al., 1994; Lesser et al., 1990, 1993; Venclovas et al., 1994). It is generally believed that specific contacts play the major role in the selection of a certain DNA base sequence and the nonspecific ones serve mostly to increase the overall stability of the complex. There is, however, also the possibility of indirect recognition through the local structural properties of DNA. Such indirect mechanisms may play an important role in protein-DNA complexes like the *EcoRI*-DNA complex, where protein binding induces large deformations in the DNA.

In the present work, we have performed MD simulation studies of the *EcoRI*-DNA complex in aqueous solution, starting from the 2.5 Å resolution x-ray crystallographic

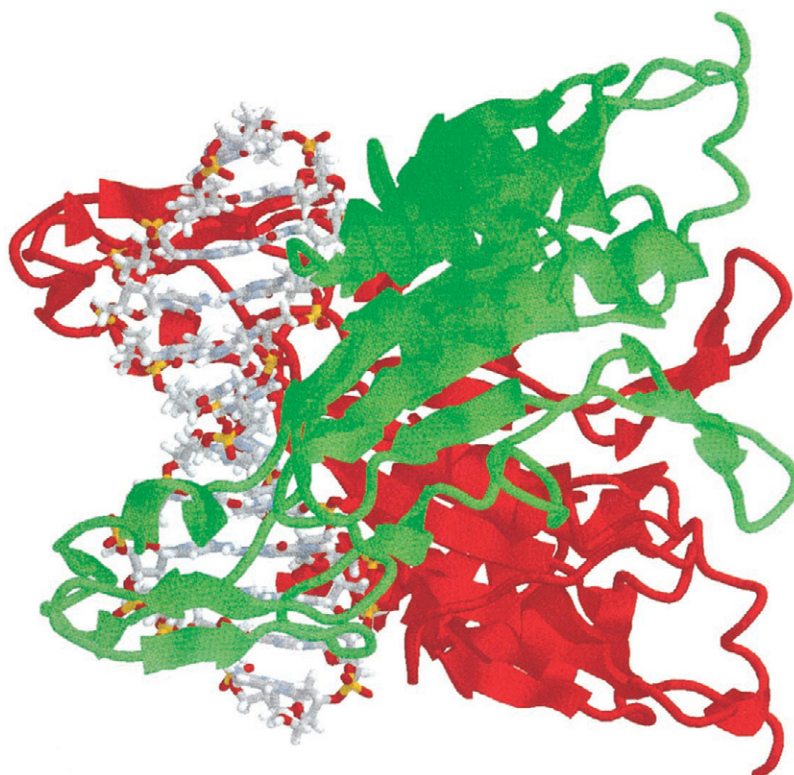


FIGURE 1 Crystal structure of the *EcoRI*-DNA complex showing the relative positions of the DNA strands and the protein monomers (colored green and red).

structure of the complex (Rosenberg, 1991; Kim et al., 1990). However, as the complex itself is a huge one (~ 9000 atoms), use of a solvent box with periodic boundary condition or a solvent sphere with stochastic deformable boundary (SDB) condition to solvate the system makes the size of the overall system quite large, and a MD simulation of reasonable length becomes computationally highly expensive. We therefore have solvated the whole complex by a water shell 7 Å thick around it. The advantage of using a solvent shell is that it solvates the whole complex and at the same time keeps the overall system size reasonably moderate. As no artificial constraints are imposed on the solvent system, the complex is also free to bend for proper relaxation. Such a solvent shell has been used before (Kandadai and Reddy, 1996; McCammon and Harvey, 1987), but its properties have not yet been fully characterized. In the present work we have analyzed some properties of this setup to check whether it can provide a realistic solvent environment and support correct dynamics of the macromolecular complex. We have also compared the time evolutions of a smaller system (*EcoRI* complex with only the first protein monomer) between MD simulations in a solvent shell setup and in a more accurate SDB setup and found that the behaviors are quite similar, justifying the use of the solvent shell setup for performing MD simulations of large systems, at least for the time scale of several hundreds of picoseconds.

From the 700-ps dynamics trajectory thus obtained, we have compared the details of the direct specific and non-specific interactions between the DNA and the protein in the simulation structure in aqueous solution with those in the crystal structure. Most of the important specific and non-specific interactions in the crystal structure are present in the simulated structure in solution and remain well maintained during the dynamics. It is seen that in general, the specific interactions are enhanced in the simulation compared to the crystal structure as a result of either improved H-bonding geometry or the development of a few new additional interactions. We have also analyzed the water bridges in the complex to investigate their role on the specificity and the overall stability of the complex. The time evolution of different properties, such as the interactions between different molecular components, have been investigated, and the observed results have been discussed in terms of solvent interactions. Finally, we have analyzed the energetics of the kink introduced in both the strands of the DNA duplex in the complex. For this analysis, we have allowed the kinked DNA isolated from the complex to relax by MD simulation in aqueous solvent and have checked if the kink is still there in the relaxed average structure of the DNA. We have compared the self energy of the relaxed free DNA in solution with that of the kinked DNA in the complex to have an estimate of the kink energy. Free energy perturbation calculations and ordinary MD simulations for several mutants of the complex have been discussed in detail in the accompanying article (Sen and Nilsson, 1999).

METHODS

System setup and solvation of the complex

We have used the x-ray crystallographic coordinates of the *EcoRI*-DNA complex at 2.5 Å resolution (Rosenberg, 1991; Kim et al., 1990) as the starting model structure with an all-atoms representation. The complex system includes the whole of the dimeric protein (except the first 16 amino acid residues in each monomer as obtained from the x-ray crystallographic coordinates) and the associated dodecamer DNA duplex of base sequence $d(\text{CGCGAATTCGCG})_2$ containing the recognition sequence $d(\text{GAATTC})_2$ as the central part of the DNA. In the complex are five histidine residues for each protein monomer, which we have chosen to be neutral, with the NE2 atom protonated, because the experimental pH was about 7 (Rosenberg, 1991; Lesser et al., 1993). First, we have minimized the energy of the crystallographic structure of the complex in vacuum for 500 steepest descent steps to remove any residual bad van der Waal contacts or other strains in the structure and then solvated the complex. Because the complex itself is very large (~ 9000 atoms and dimensions are $x = \pm 36.9$ Å, $y = \pm 37.7$ Å, and $z = \pm 30.0$ Å) we have solvated the molecular complex by a 7.0-Å aqueous solvent shell prepared by immersing the complex in a large sphere of TIP3P water (Jorgensen et al., 1983) and deleting any water molecule which had its oxygen atom at a distance either less than 2.8 Å or more than 7 Å from a non-hydrogen atom of the solute. This keeps the overall size of the solvated system moderate (18,000 atoms). On the other hand, solvating the complex in a water sphere with a water layer of at least 7 Å should have required a radius of about $38 + 7 = 45$ Å, and a periodic box would have been even larger. The total charge of the complex is -26 units, so we have included 26 Na^+ counterions to make the system electrically neutral. The counterions were placed by replacing 26 water molecules with highest electrostatic energies of the oxygen atoms in the solvent shell and care was taken such that no two counterions were placed closer than 5 Å. Further energy minimization of the whole solvated system was done by 500 steepest descent steps, and this solvated system was used to perform subsequent MD simulation as discussed in the protocol section. For the SDB setup we have immersed the DNA-protein complex with only the first monomer of the protein at the center of a sphere of radius 36 Å filled with TIP3P water molecules and all the water molecules which were closer than 2.8 Å to any atom of the complex were removed. The water molecules in the sphere were experiencing the action of a deformable boundary force arising from mean field interactions of water molecules beyond the 36-Å boundary (Brooks and Karplus, 1983). To make the system electrically neutral 24 sodium counterions were placed as described before and energy minimization was performed as described above.

Dynamics simulation protocol

The dynamics simulation in the case of solvent shell was performed by integrating Newton's equations of motion by the leap-frog algorithm with an integration time step of 2 fs, using the program package CHARMM version 25 along with the combined protein and nucleotide topology and parameters 22 (Brooks et al., 1983; MacKerell et al., 1995, 1998). We used 13 Å as the cutoff value for generating the nonbonded pair list and updated the list every 20 steps. For the nonbonded interactions, force shift option was used, causing the interaction energies and the forces to vanish smoothly at a distance of 12.0 Å. All bond lengths involving hydrogens were kept fixed using the SHAKE algorithm (Ryckaert et al., 1977). During simulation, we first heated the system to 300K over 2 ps and subsequently equilibrated at 300K over 2 ps by assigning velocities from a Gaussian distribution of velocities at 300K. The simulation was then continued and the temperature was checked every 200 steps. If the average temperature was outside the window of ± 10 K, the velocities of the atoms were rescaled accordingly to maintain the temperature around 300K. During the simulation, the base pairing at either end of the DNA duplex was constrained by harmonic constraint on the hydrogen bonds to minimize possible end effects. The simulation was continued for a total of 700 ps and

the trajectories were saved every 100 steps for analysis. The trajectories over the last 300 ps were used to calculate most of the various average structural, interactional, and dynamic properties of the system as described in Analysis (below). The average structure of the complex was calculated over the last 300 ps of the trajectory, taking all the frames into consideration, and the potential energies of the average coordinates of the complex were then energy-minimized by 100 steepest descent steps in vacuum. In the SDB setup the atoms in the region between the radii 34 Å and 36 Å (the buffer region) interacted with a stochastic heat bath at 300K and dissipative forces (Brooks and Karplus, 1983) and followed Langevin dynamics algorithm while, for the atoms inside the 34-Å radius, Newton's equation of motion for ordinary MD simulation was applied. Other aspects were the same as in solvent shell simulation. SDB dynamics simulation was performed for 200 ps.

Procedure for estimating the energetics of the DNA kinks

For investigating the energetics of the kinked DNA, we have isolated the kinked DNA from the crystal structure of the complex and minimized its energy by 300 steepest descent steps and then placed 22 Na⁺ ions as counterions, each at a distance of 3.5 Å from the phosphorus atom on the bisector of the angle O1P-P-O2P of the respective phosphate group. This DNA with the counterions was then solvated by a 7-Å water shell and the overlapping water molecules were removed. MD simulation of this solvated kinked DNA was then performed, following the protocol described above, in order to allow it to relax into its unstrained conformation in the absence of the protein counterpart in the complex. We monitored the time evolution of the RMSD of the DNA to confirm its relaxation. When the average RMSD was found to be stable over a substantial period of simulation, we assumed that the DNA had reached an average relaxed structure. Then we calculated the average self energy $\langle E_{\text{relax}} \rangle$ and its fluctuations for the free DNA from the time series obtained from the simulation trajectory over the last 300 ps. Similarly, we have calculated the average self energy $\langle E_{\text{kinked}} \rangle$ and its fluctuations of the DNA in the complex over the last 300 ps of the simulation, over which the structure is most stable, and have compared it with $\langle E_{\text{relax}} \rangle$. Using these two average self energies, we finally calculated $\Delta E = \langle E_{\text{kinked}} \rangle - \langle E_{\text{relax}} \rangle$, which gives a reasonable estimate of the kink deformation energy in the DNA.

Analysis

Structural deviations of the protein-DNA complex from the initial energy-minimized x-ray crystallographic structure were assessed on the basis of the RMSD. In computing the RMSD the overall translational and rotational motions have been removed by superimposing each configuration of the complex from the trajectory at intervals of 100 steps onto the initial structure.

The mean square positional fluctuations $\langle \Delta r^2 \rangle$ of the atoms with reference to the average structure in solvent, averaged over the trajectories after rigid body alignment against the coordinates of the starting structure, were calculated for comparing the relative dynamic fluctuations of the different residues of the protein and DNA in the complex. The average was taken over all atoms except the hydrogens in a given amino acid residue or nucleotide.

The conformations of the DNA part of the complex in the crystal and the average structure in solution were characterized by the helicoidal parameters calculated by using the software package Curves 5.1 (Lavery and Sklenar, 1988; Ravishanker et al., 1989).

The correlation coefficients between pairs of dynamical quantities were computed to investigate the correlated motions present in the dynamics of the complex. The linear cross-correlation coefficient between two dynamical quantities x and y is given by the expression

$$C_{xy} = \langle \Delta x_i \cdot \Delta y_i \rangle / [\langle \Delta x_i^2 \rangle \cdot \langle \Delta y_i^2 \rangle]^{1/2},$$

where $\langle \dots \rangle$ denotes an average over the trajectory and $\Delta x_i = x_i - \langle x \rangle$, $\Delta y_i = y_i - \langle y \rangle$.

The diffusion coefficients for the water molecules and the sodium counterions in the systems were estimated from the mean square displacement, by using the Einstein relation (Allen and Tildesley, 1987)

$$\lim_{t \rightarrow \infty} \partial \langle [R(t) - R(0)]^2 \rangle / \partial t = 6D$$

where $\langle \dots \rangle$ denotes the average over all the water molecules except the evaporated ones, $R(t)$ is the position vector of a particular water molecule at time t , and D is the diffusion coefficient.

Solvent-accessible surface area for the whole complex and its different molecular components (DNA, protein) were calculated using the method of Lee and Richards (1971) with a test particle diameter of 1.6 Å.

Direct and water-mediated H-bonding interactions between the monomers and between the DNA and the protein were analyzed using the following criterion. Two atoms are considered to form a H-bond (A—H—D) if the acceptor-hydrogen distance is < 2.4 Å and the acceptor-hydrogen donor angle is $> 135^\circ$. A water bridge is defined as an interaction between residues that makes hydrogen bonds to a common water molecule. The number of average hydrogen bonds or water bridges made by an atom is calculated by dividing the sum of the number of H-bonds or water bridges made by the atom in each frame by the total number of frames considered for analysis. Similarly, the average lifetime of a H-bond or a bridge is calculated by summing the duration of all H-bond or bridging events and dividing by the number of such events for any given donor-acceptor pair. In these calculations we used the coordinate frames from the trajectory at an interval of 1 ps and, in the calculations of averages, considered only events with lifetimes > 5 ps.

RESULTS AND DISCUSSION

Properties of the solvent shell and comparison between shell and SDB simulations

First we have checked whether a 7-Å solvent shell behaves properly during the period of simulation to mimic the solvation of the complex in a realistic fashion. Because in such a setup, in the absence of the periodic boundary condition or the boundary force as used in a SDB setup, there is the possibility that the outer solvent surface may diffuse out or the water molecules may evaporate, we have looked into these aspects as follows. We have monitored the number of evaporated water molecules against the simulation time. When the least distance of the oxygen atom of a water molecule from any of the atoms in the solute complex becomes very large (> 50 Å), then we consider that particular water molecule to be an evaporated one. Fig. 2 *a* indicates that even in this setup, not many water molecules are really evaporated, although the number increases in a fairly linear fashion with the increase in the length of the simulation period. Considering the actual number (~ 3000) of water molecules present in the system, the number (39) of evaporated water molecules at the end of the simulation over 700 ps is insignificant.

Fig. 2 *b* shows the time evolution of the number of water molecules residing outside the 7-Å boundary limit. It is found that during the first few tens of picoseconds the number increased very rapidly and then reached an apparent steady state when the number did not change significantly over the simulated time period. The fluctuating nature of these data points is due to the fact that under the action of

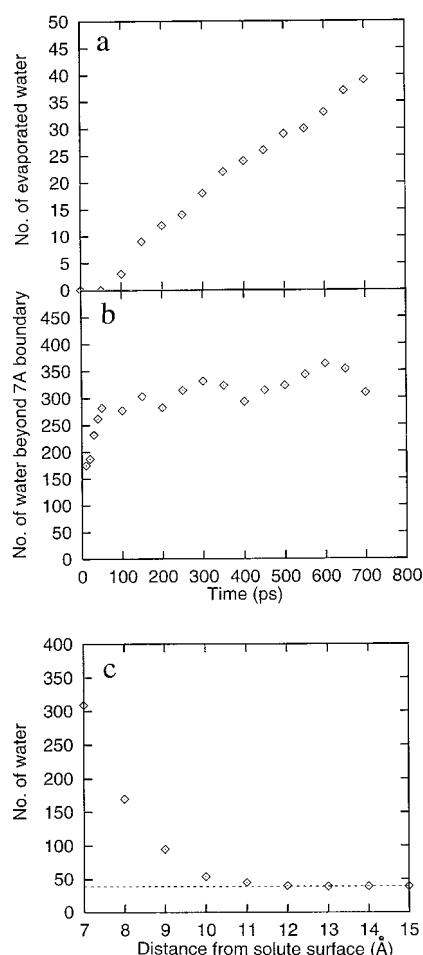


FIGURE 2 (a) Time evolution of the number of evaporated water molecules from the solvent shell during MD simulation. (b) Time development of the number of water molecules which are diffused out of the 7-Å-thick initial solvent boundary due to free dynamics. (c) The distribution of diffused water molecules outside the 7-Å-thick layer after 700 ps dynamic simulation of the system. The dotted line indicates the number of evaporated water molecules.

the mutual interactions among the nearby water molecules and its dynamics, some water molecules go out crossing the 7-Å boundary at the same time some others cross into the boundary; a transient imbalance between these two processes causes the fluctuation. The steady average value of the number of water molecules outside the 7-Å boundary is negligible compared to the total number of water molecules in the system, indicating insignificant smearing of the solvent boundary over the simulation time period.

We have also plotted the number distribution of the water molecules beyond the 7-Å limit of the solvent shell against the distance from the solute surface in the frame of the 700th ps in Fig. 2 c. The result indicates that there is indeed a smearing out of the boundary, but the spreading is limited within only a few angstroms from the initial 7 Å boundary. The dashed line represents the number of evaporated water molecules in that frame.

For further characterization of the dynamic properties of the solvent in the shell, we have calculated the diffusion

coefficients of water (D_{wat}) and the ions (D_{ion}) using the mean squared displacement obtained from the dynamic trajectories and the standard Einstein relation. Estimates were made over a 500-ps time scale for three different time origins and their average was calculated. The estimated values $D_{\text{wat}} = (2.04 \pm 0.02) \times 10^{-9} \text{ m}^2/\text{s}$ and $D_{\text{ion}} = (0.39 \pm 0.05) \times 10^{-9} \text{ m}^2/\text{s}$ from the present simulation data are in agreement with earlier studies with TIP3P water model (Eriksson and Nilsson, 1995) and also for the sodium ions (Eriksson and Nilsson, 1995; Koneshan et al., 1998). In this estimation we have excluded the evaporated water molecules. The corresponding experimental value is $D_{\text{wat}} = 2.4 \times 10^{-9} \text{ m}^2/\text{s}$. However, recent simulation with pure TIP3P water yields the diffusion coefficient in the range $D_{\text{wat}} = (5.2\text{--}5.5) \times 10^{-9} \text{ m}^2/\text{s}$ (van der Spoel et al., 1998), which is higher than our estimated value. Compared to the self-diffusion coefficient of pure TIP3P water (without any solute inside), a smaller value of the estimate is expected in the present case because the DNA-protein complex has a large surface in contact with water and a large number of water molecules interact with it, contributing less to the calculated mean squared displacement value. It may also be noted that as the thickness of the solvent shell is only 7 Å, the motion of the water molecules over longer time scale may be a little bit restricted (the component of motion normal to the solute surface) and hence, calculation of diffusion coefficients using large values of t (i.e., 500 ps) seems to give an incorrect estimate. On the other hand, for shorter time scales the validity of the diffusion equation breaks down. Thus, as previously suggested (Lindahl and Edholm, 1998), we calculated the diffusion coefficient for water over an intermediate time interval (100 ps) and obtained a value of $(3.29 \pm 0.49) \times 10^{-9} \text{ m}^2/\text{s}$. Use of a much shorter time interval (10 ps) yields a higher value ($4.5 \times 10^{-9} \text{ m}^2/\text{s}$) for D_{wat} approaching the value obtained for TIP3P in an unrestrained system (van der Spoel et al., 1998). The variation of the values estimated over different time lengths is due to the restricted degrees of freedom in the shell. We have not calculated the radial distribution function of water molecules in the solvent shell for comparison because it appears that the water shell is not thick enough to produce a meaningful radial distribution function in the present case.

Fig. 3 compares the time evolution of the RMSD of the complex where only one monomer (the first one) is present and dynamic simulations are performed in two completely independent setups, the SDB setup and the 7-Å solvent shell setup. We removed the second monomer from the crystal structure and used this modified system as the test system because, if there is any significant difference between the performance of the two different setups, the difference could appear more pronounced for this non-native system than in the case of the original complex with both the monomers. The comparison shows that there is no significant difference in the behavior of the time evolution of RMSD. All these results together indicate that this 7-Å solvent shell setup is reasonably good for studying the

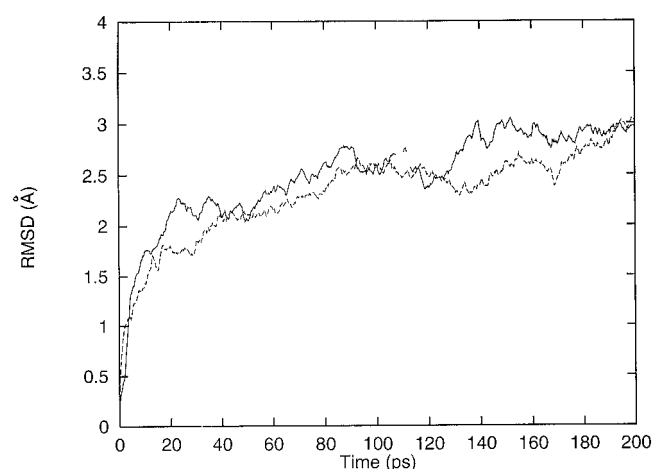


FIGURE 3 Time development of the RMSD of the complex with a single protein monomer in solvent shell setup (*solid line*) and SDB setup (*dashed line*).

solvent simulation of such large systems over time periods of considerable length.

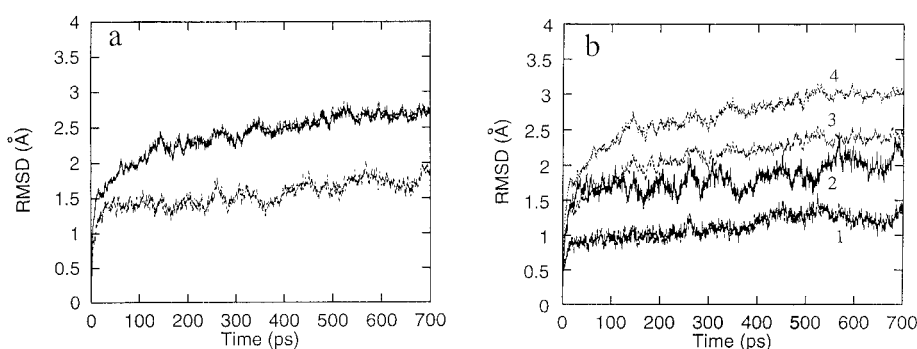
Time evolution of the RMSD of different molecular components and the RMSD and fluctuation patterns of the different residues during dynamics

In Fig. 4 *a*, the solid line represents the time evolution of the RMSD of the non-H atoms of the whole protein-DNA complex over the entire period of simulation. It is clearly seen that the average values of the RMSD of the whole complex as well as the molecular components have become fairly stable over a substantial period of the simulation, indicating that a stable structure of the complex in solution has been reached. However, considering the fact that we have used a 2.5-Å resolution crystal structure for the starting coordinates, the average RMSD (2.57 ± 0.11 Å over the last 400 ps of the trajectory) of the complex compared to the starting structure does not indicate a very significant conformational deviation of the complex in solution. Comparison between the time evolution of the RMSD of the DNA (*dotted line*) and the protein (*dashed line*) indicates that more structural rearrangements occurred for the protein

compared to the DNA duplex. The average RMSD of the DNA duplex is 1.89 ± 0.15 Å, whereas for the protein it is 2.63 ± 0.12 Å. The relatively larger fluctuations in the RMSD values for the DNA as seen in Fig. 4 *a* seem to be due to the ends of the DNA duplex (presence of more free ends exposed outside the complex) and also to the smaller number of atoms in it. The average RMSD of the DNA without the end bases is obtained as 1.66 ± 0.13 Å. It is also noticeable that since the protein is about 12 times larger compared to the DNA duplex involved, the time evolution of the whole complex is governed mainly by the behavior of the protein part.

Calculation of the linear cross-correlation coefficient between the two time series of the RMSD of the DNA and the protein yields a value of 0.9 for the linear correlation coefficient over the first 100 ps of the dynamics trajectory, indicating that they are highly correlated. On the other hand, the correlation coefficient is found to be only 0.4 over the last 100 ps of the trajectory, implying a loss of correlation. During the first 100 ps there is a common drift in the RMSD time series of the DNA and the protein that may be the cause of the high correlation over this period. Moreover, analysis of the solvent accessible surface area (discussed below) indicates that in the crystal structure of the complex, a large fraction of the DNA surface is covered by the protein and thus, during the initial dynamics, structural change in most parts of the DNA can occur only when it is coupled with a structural change in the protein in the same locality, which allows the DNA to change its local conformation. This may also be partially responsible for the initial high correlation between the two time series. As dynamics goes on, the complex relaxes in solvent and gradually becomes slightly loose, allowing the DNA more freedom to move independently of the protein; thus, the motion between them becomes less correlated. However, the correlation coefficient over the last 100 ps of the trajectory still shows the existence of a significant degree of correlation, indicating tight binding between the DNA and the protein even in the relaxed complex. Fig. 4 *b* indicates that more rearrangements occurred for the backbone atoms of the DNA compared to the bases, whereas for protein, more structural changes occurred for the side chains than the backbone atoms.

FIGURE 4 Time development of the RMSD of (a) the (i) whole complex (*solid line*), (ii) the protein (*dashed line*, almost superimposed on the *solid line*) and (iii) the DNA duplex (*dotted line*), (b) of the base atoms (line 1), backbone atoms (line 2) of the DNA, and the backbone atoms (line 3) and the side chains (line 4) of the protein.



In the case of DNA the fluctuations are smaller compared to the protein and are similar over the nucleotides, except at the ends (data not shown). Fig. 5 represents the plots of the RMSD (Fig. 5 *a*) and the positional fluctuations of the atoms (Fig. 5 *b*) in a residue averaged over all the atoms in the residue of monomers *EcoRIa* (solid line) and *EcoRIb* (dashed line) in the complex in aqueous solution, with reference to the average structure in solution. Comparison between the fluctuation pattern of the two monomers of the protein shows striking similarity, indicating that the two monomers exhibit very similar dynamics in solution. The regions of the protein monomers that interact with the DNA strand are found to exhibit relatively smaller fluctuation compared to the other parts of the monomer. It is seen that larger changes occurred for the residues that mainly represent the different loops of the protein. The largest change occurred in the residue range Asn¹²¹ to Ala¹⁴², which corresponds to the longest loop in the structure of the protein.

The time evolution of the radius of gyration of the complex (data not shown) does not show any significant change throughout the simulation, indicating that the complex maintained its overall shape and size during the simulation in aqueous solvent.

Characterization of the direct interactions between the DNA and the protein

Fig. 6 *a* compares the specific interaction pattern of the DNA bases of the first strand (DNA1) with the protein dimer of the complex in the crystal structure (dashed line) and the dynamic average structure in aqueous solution (solid line) against the DNA nucleotide number. Fig. 6 *b* represents the same for the other strand (DNA2) of the DNA. It is seen that in both the strands, all six bases in the

recognition site interact with the protein in different ways, showing their different roles in providing the specificity and stability of binding. Interestingly, we found that both in the crystal structure and in the dynamics average in solution, some additional bases in the 5' side of the DNA strand outside the recognition site interact appreciably with the protein. These interactions may have an important role in the initial binding to some extent to the nonspecific region and then moving to the recognition part by a sliding mechanism, as suggested (Szczelkun et al., 1997; Jeltsch et al., 1994).

It is clearly seen that the interaction pattern for the two DNA strands (Fig. 6, *a* and *b*) is very similar to that in the crystal structure, indicating that most of the specific interactions between the DNA and the protein as found in the crystal structure are also present in the dynamic average structure and the interaction strengths, in general, are relatively enhanced in aqueous solution. These plots also indicate that, as in the crystal structure, the most important specific interactions of DNA in the simulation average structure come from Gua⁴ and Ade⁶ and the interaction strength is much enhanced in the simulated average structure in solution. The reason for the large enhancement of the specific interaction energy in the case of Gua⁴ in both strands was found to be closer H-bonding contact resulting from relaxation of the structure in the simulation. The distance between the acceptor and the hydrogen was 3.2 Å in the crystal structure, whereas it was less than 2.4 Å in solution structure.

Fig. 6, *c* and *d*, compares the nonspecific interaction pattern of the two DNA strands with protein in the crystal structure and in the average structure. Here also the pattern of the nonspecific interaction between the DNA nucleotide and the protein dimer in the simulation was found to be very

FIGURE 5 (*a*) The average RMSD of the protein residues against the residue number of the monomers *EcoRIa* (thick line) and *EcoRIb* (thin line) and (*b*) the average positional fluctuations of the protein residues against the residue number of the monomers *EcoRIa* (thick line), *EcoRIb* (thin line), and the temperature factors of the crystal structure, converted to positional fluctuations (dotted line).

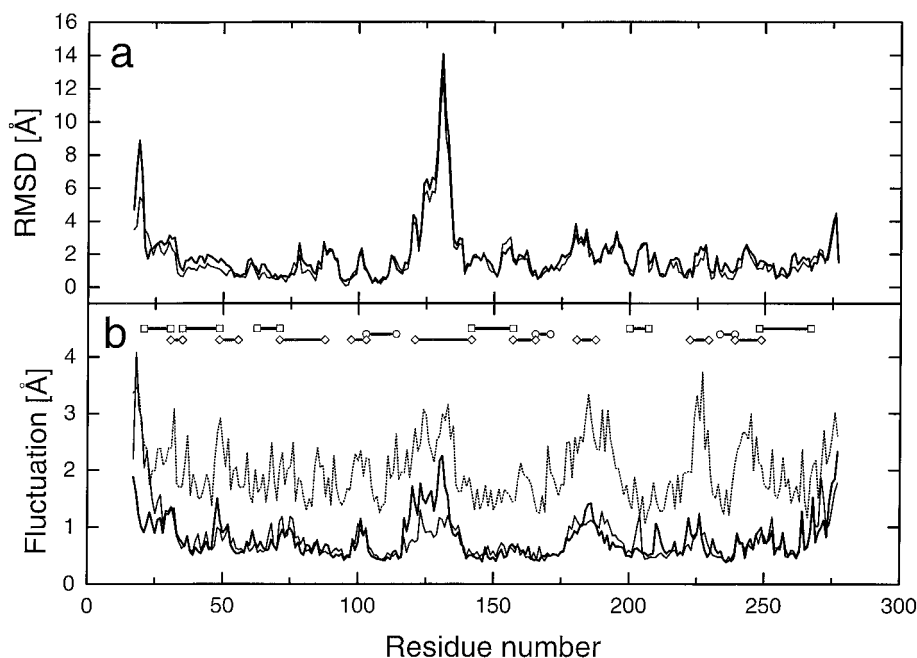
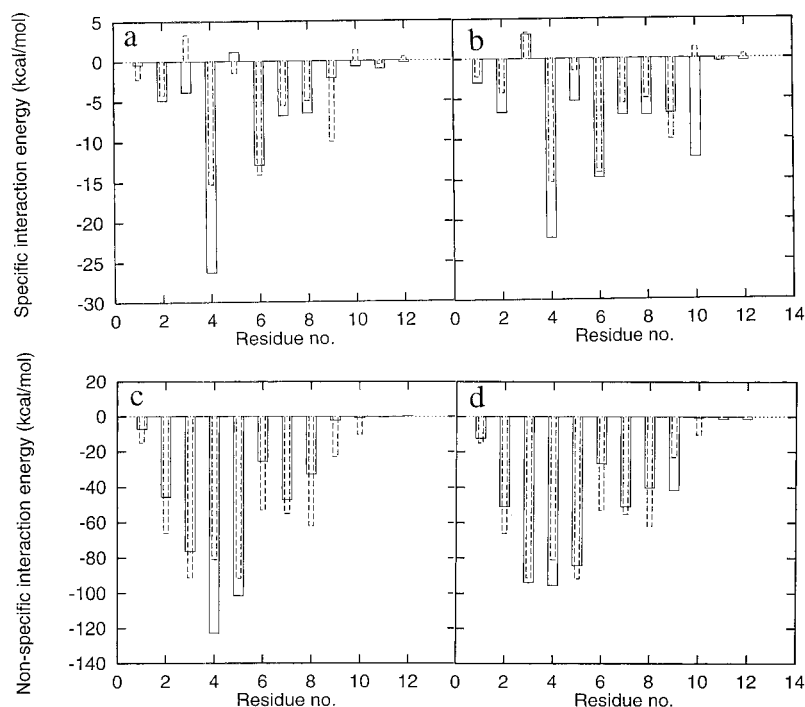


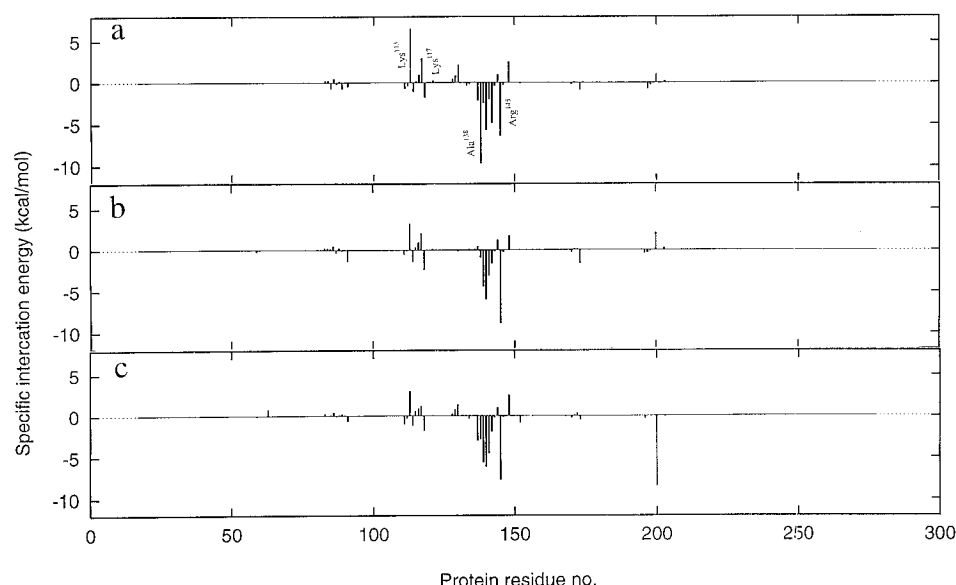
FIGURE 6 The interaction energies of the individual DNA residues with the whole protein in crystal (*dashed line*) and in the energy minimized dynamic average structure (*solid line*) against the DNA nucleotide number. (a) and (b) represent the cases of specific interaction between the protein and the DNA strand 1 and the DNA strand 2, respectively. (c) and (d) represent the nonspecific interactions for the DNA strands 1 and 2, respectively.



similar to that seen in the crystal structure. The nonspecific interaction strengths in the dynamics average were found to be slightly enhanced for some nucleotides compared to the crystal structure, whereas for others it was found to be slightly reduced. The strongest nonspecific interactions of DNA came from Cyt³, Gua⁴, and Ade⁵. Specificity is determined by the overall difference between the total DNA-protein complex and a DNA-protein interaction energies of the wild-type complex and a mutant variant. When the difference is taken, it is expected that the common part coming from the nonspecific interaction should be cancelled out, making the specificity dependent mainly on the difference in specific interaction energies. However, the nonspecific interaction actually enhances the overall stability of the complex and thus, a large nonspecific interaction energy only shifts the equilibrium concentration of the complex toward higher value in the complex formation reaction. Fig. 7 *a* represents the specific-interaction energies between the whole DNA and *EcoRIa* (the first protein monomer) in the crystal structure against the protein residue number. A clearly distinct pattern of interaction is observed, indicating the relative strengths of the interactions with which the different bases of the DNA and the protein residues are interacting in the complex. The number of each residue that interacts considerably with the DNA (as evidenced by this energy calculation) is indicated at the top of the corresponding histogram bar. The interaction of the DNA is found to occur with several different highly local regions on the one-dimensional sequence of the protein. Fig. 7 *b* represents the specific-interaction pattern between the whole DNA and *EcoRIa* in the average structure in aqueous solution. Fig. 7 *c* represents the same for the whole DNA and *EcoRIb* (the second monomer of the protein). In a similar fashion Fig. 8

a represents the nonspecific interaction patterns between the whole DNA and *EcoRIa* in the crystal structure. Fig. 8, *b* and *c*, represents the same for the non-specific interaction between the DNA and *EcoRIa* (Fig. 8 *b*) and DNA and *EcoRIb* (Fig. 8 *c*) in the average structure in solution. Comparison of the interaction strengths for the different residues gives a quantitative measure of the relative importance of the different residues of the protein in the process of molecular recognition and stabilization of the complex. From the plots in Figs. 7 and 8, it is evident that the interaction patterns are very similar to those in the crystal structure. It was also found that the most important interactions for specificity seem to involve the protein residues Ala¹³⁸, Gly¹⁴⁰, Asn¹⁴¹, and Arg¹⁴⁵, in which any alteration is expected to result in a considerably altered specificity of DNA sequence, whereas the most important nonspecific interactions come from the residues Lys¹¹³, Lys¹¹⁷, Arg¹⁴⁵, and Lys¹⁴⁸, where mutation should most affect the overall stability of the complex. However, the relative importance of the interactions of these residues in aqueous solution are found to be slightly different than those in the crystal structure. Variation of the interaction strengths in both the cases of specific and nonspecific interactions are mainly found to occur due to improved H-bonding geometry and the development of a few more interactions between the two interacting residues (Table 1). It may be mentioned here that in the crystal there are contacts between the primary protein-DNA complex and the neighboring complexes. In solution environment such contacts are lost. Thus, this physical difference may also be partly responsible for some of the differences in the structural and interactional properties of the complex observed here.

FIGURE 7 Specific interactions of the whole DNA duplex with (a) the protein monomer *EcoRIa* in crystal structure, (b) the protein monomer *EcoRIa* in the average structure in aqueous solution, and (c) the protein monomer *EcoRIb* in the average structure in aqueous solution, against the protein residue number.



Further calculations of the nonbonded interaction energies show that, as in the crystal structure, each DNA strand is not only interacting mostly with the corresponding protein monomer (DNA1-*EcoRIa* and DNA2-*EcoRIb*) but also cross-interacting with the other monomer (DNA1-*EcoRIb* and DNA2-*EcoRIa*) with comparable energies in solution simulation (data not shown). Only a few bases in the recognition site are involved in these cross-interactions and the most important interactions come from the protein residues Arg²⁰⁰. Table 1A gives a detailed comparison of the DNA-protein H-bonding interaction pattern in solution simulation with those in crystal structure. It was found that the specific interactions involving the N7 atom of the purine bases Gua⁴, Ade⁵, and Ade⁶ were maintained throughout the simulation, whereas the specific interactions involving the amino groups of the Ade⁵ and Ade⁶ were found to be less stable with less H-bonding time and shorter lifetime during the

later part of the present simulation. Table 1B gives the comparison of the H-bonding interaction pattern between the two monomers of the protein in crystal and in the average simulation structure.

It is clearly seen from Fig. 5, *c* and *d*, that the nonspecific interactions of Ade⁶ and Thy⁸ in both DNA strands with the protein are decreased in the simulated average structure. We looked into the local structure and the time development of these interactions and found that each of these nucleotides interacts with a few residues of the protein. During dynamics, in the case of Ade⁶, the side chain of the interacting protein residue Lys¹¹³ moved away from its phosphate oxygen; as a result, the overall interaction was reduced. In a similar way the side chain of the residue Lys¹¹⁷, which was interacting with the DNA nucleotide Thy⁸ in the crystal structure, moved out due to a rotation of the side chain. As these reductions in interactions happened in both strands of

FIGURE 8 Nonspecific interaction of the whole DNA duplex with (a) the protein monomer *EcoRIa* in crystal structure, (b) the protein monomer *EcoRIa* in the average structure in aqueous solution, and (c) the protein monomer *EcoRIb* in the average structure in aqueous solution, against the protein residue number.

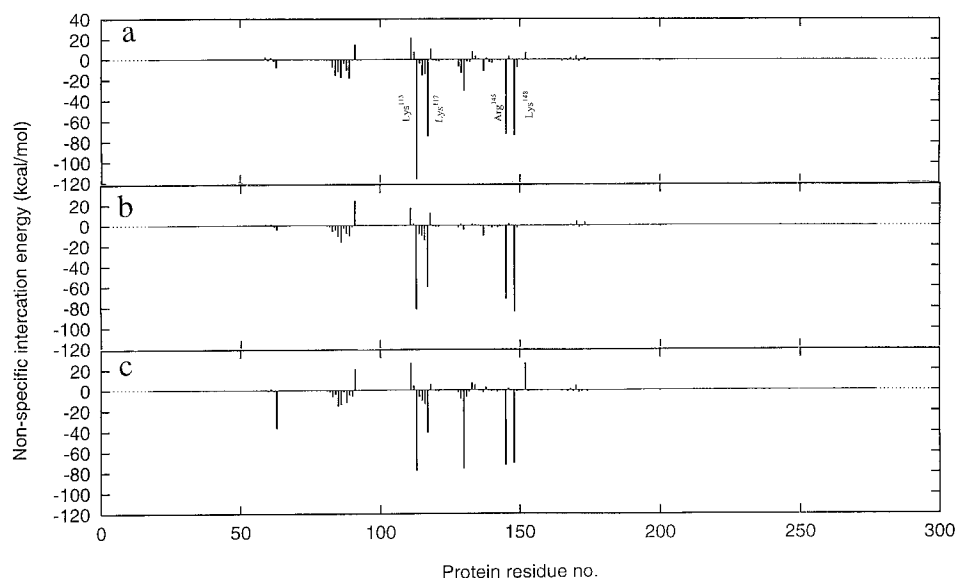


TABLE 1A H-bonding interaction between the DNA and protein

Atoms bonded by a H-bond	Total residence time (ps)	Aver. lifetime (ps)
DNA1:2G-O1P ····· HN-Gly196:EcoRIb*	149	12.8
DNA1:2G-O1P ····· HG1-Ser195:EcoRIb	88	29.3
DNA1:3C-O2P ····· HN-Ser87:EcoRIa*	99	11.0
DNA1:4G-O1P ····· HD22-Asn141:EcoRIb	53	6.6
DNA1:4G-O1P ····· HH11-Arg203:EcoRIb	207	12.9
DNA1:4G-O2P ····· HZ2-Lys48:EcoRIa	297	74.2
DNA1:4G-N7 ····· HH22-Arg200:EcoRIb	261	23.6
DNA1:5A-O5' ····· HH11-Arg145:EcoRIa	225	13.2
DNA1:5A-O1P ····· HH22-Arg145:EcoRIa*	297	297.0
DNA1:5A-O2P ····· HZ1-Lys113:EcoRIa	298	74.5
DNA1:6A-O1P ····· HN-His114:EcoRIa	297	59.4
DNA1:6A-N7 ····· HH12-Arg145:EcoRIa*	298	74.5
DNA1:7T-O1P ····· HN-Gly116:EcoRIa*	269	22.4
DNA1:8T-O2P ····· HZ2-Lys1117:EcoRIa	81	20.3
DNA1:9C-H41 ····· O-Ala139:EcoRIa†	92	15.3
DNA2		
DNA2:2G-O1P ····· HN-Gly196:EcoRIa*	208	13.8
DNA2:3C-O1P ····· HH11-Arg203:EcoRIa	125	25.0
DNA2:3C-O2P ····· HZ2-Lys63:EcoRIb*	92	46.0
DNA2:4G-O1P ····· HH12-Arg203:EcoRIa	238	17.0
DNA2:4G-O1P ····· HH21-Arg203:EcoRIa	229	45.8
DNA2:4G-O2P ····· HZ2-Lys148:EcoRIb	291	97.0
DNA2:4G-N7 ····· HH12-Arg200:EcoRIa	121	13.4
DNA2:4G-O6 ····· HH12-Arg200:EcoRIa	258	23.4
DNA2:5A-O5' ····· HH11-Arg145:EcoRIb	242	20.1
DNA2:5A-O1P ····· HH22-Arg145:EcoRIb†	299	99.6
DNA2:5A-H61 ····· OD1-Asn141:EcoRIa	299	99.6
DNA2:6A-N7 ····· HH12-Arg145:EcoRIb†	145	10.3
DNA2:6A-O1P ····· HN-His114:EcoRIb*	299	99.6
DNA2:7T-O1P ····· HN-Gly116:EcoRIb*	219	18.2
DNA2:9C-O1P ····· HZ2-Lys130:EcoRIb	119	14.8

Total residence time and average lifetime over the last 300 ps of the dynamic trajectory considering frames every 500 steps. Only hydrogen bonds with total residence time >50 ps are considered.

* H-bonds which are also present in the crystal structure.

† H-bonds include the same atom in DNA as in crystal structure, but the atom of the protein involved is different from the crystal structure.

the DNA, it seems to be a natural consequence of the solvent environment. It was also found that the moving out of Lys¹¹⁷ is the result of hydrophilic effect for both strands, which indicated an increase in their interaction energy with water. On the other hand, in the case of Ade⁶, we found that a water molecule penetrated into the local structure and screened the interaction between Ade⁶ and Lys¹¹³. Moreover, in the H-bond list, it may be noted that a crucial specific interaction (between H61 of Ade⁵-DNA1 and OD1 of Asn¹⁴¹ of *EcoRIb*) is absent. To find the reason, we looked into the time evolution of the distance between these two atoms and found that this H-bond was well maintained up to the first 375 ps of the dynamic simulation and then suddenly disrupted due to a rotation of the side chain of the Asn¹⁴¹ of *EcoRIb* (Fig. 9). However, the same interaction in the other strand was properly maintained throughout the entire simulation period (Fig. 9). As the other important interactions between the DNA and the protein were also maintained, we consider this event an isolated accidental

TABLE 1B H-bonds between residues of monomers

Atoms bonded by a H-bond	Total residence time (ps)	Average lifetime (ps)
EcoRIa:63Lys-HZ1 ····· OD2-Asp202:EcoRIb	166	55.3
EcoRIa:63Lys-HZ2 ····· OD2-Asp202:EcoRIb	87	43.5
EcoRIa:80Thr-HN ····· O-Gly210:EcoRIb	236	13.8
EcoRIa:80Thr-HG1 ····· O-Gly210:EcoRIb	123	8.8
EcoRIa:80Thr-OG1 ····· HG1-Thr205:EcoRIb	211	14.0
EcoRIa:82Phe-HN ····· OD1-Asp202:EcoRIb	211	16.2
EcoRIa:83Val-HN ····· OD1-Asp202:EcoRIb	298	99.3
EcoRIa:84Ser-HN ····· O-Asn194:EcoRIb	126	12.6
EcoRIa:86Ser-HG1 ····· OD1-Asp202:EcoRIb	229	19.0
EcoRIa:144Glu-OE2 ····· HH21-Arg145:EcoRIb	301	301
EcoRIa:145Arg-HH21 ····· OE2-Glu144:EcoRIb	300	150
EcoRIa:148Lys-HZ1 ····· OE2-Glu144:EcoRIb	157	11.2
EcoRIa:154Ala-O ····· HN-Tyr209:EcoRIb	266	22.1
EcoRIa:194Asn-O ····· HN-Ser84:EcoRIb	101	9.1
EcoRIa:202Asp-OD1 ····· HG1-Ser86:EcoRIb	72	36.2
EcoRIa:202Asp-OD2 ····· HN-Val83:EcoRIb	253	25.3
EcoRIa:209Tyr-O ····· HN-Gly62:EcoRIb	125	17.8
EcoRIa:221Lys-HZ1 ····· OD1-Asp74:EcoRIb	75	18.7
EcoRIa:221Lys-HZ2 ····· OD1-Asp74:EcoRIb	140	46.6
EcoRIa:221Lys-HZ2 ····· OD1-Asp76:EcoRIb	58	11.6
EcoRIa:229Ser-O ····· HN-Met231:EcoRIb	301	301
EcoRIa:231Met-HN ····· HN-Met231:EcoRIb	150	11.5
EcoRIa:269Asp-OD2 ····· HZ3-Lys226:EcoRIb	137	19.5

List of H-bonds between the residues of the two monomers (*EcoRIa* and *EcoRIb*) of the protein, their total residence time and average lifetime over the last 300 ps of the dynamic trajectory considering frames at an interval of 1 ps. Only hydrogen bonds with residence times >50 ps are considered.

case that does not affect the general validity of the results. Because this conformational change occurred only once, we can say only that this is a possible motion and cannot estimate the likelihood that it will happen.

Calculation of the non-bonded interaction energies is also quite useful for identifying the protein residues in the interface of the DNA and protein. Consideration of only those protein residues whose energies of interaction with the DNA are >0.5 kcal/mol indicates that parts of the β sheets β I, β II, β^2 , β^3 , and β^4 , the α -helices α^1 and α^4 , and the loops between them form the DNA-protein interface. Only those protein residues within the cutoff for evaluation of non-bonded interactions (12 Å) are included in these calculations, and, because DNA is a charged molecule, there is small but finite possibility that a charged amino acid residue at a larger distance could also have an interaction energy >0.5 kcal/mol. The names of the β sheets and α helices are according to the crystal structure (Rosenberg, 1991; Newman et al., 1994).

Characteristics of the DNA in the average structure of the complex in solution

To see if any significant difference occurs to the structure of the kinked DNA of the complex in aqueous solution, we looked into the helicoidal parameters of the DNA in the average structure of the complex obtained from simulation in water and compared them with the corresponding values

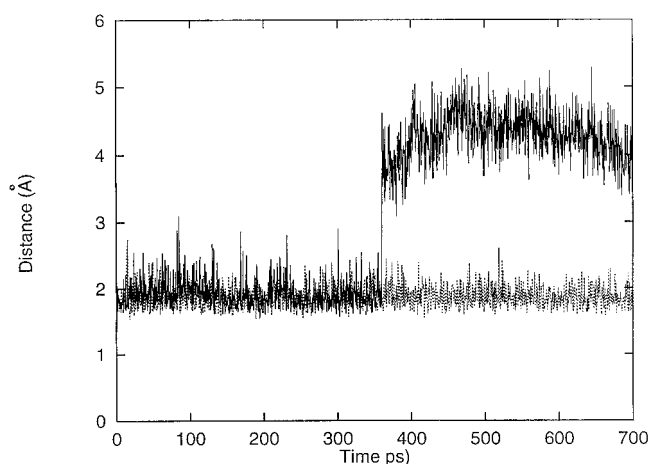


FIGURE 9 The time evolution of the distance between the atom H61 of Ade5 of the first strand of the DNA and the atom OD1 of Asn141 of the second monomer of the protein (*solid line*) and the distance between the atom H61 of Ade5 of the second strand of the DNA and the atom OD1 of Asn141 of the first monomer of the protein (*dotted line*).

in the crystal structure. Fig. 10 shows the summary of the comparison of some of the helicoidal parameters between the crystal structure and the solution average structure. A large difference is observed in the average inclination value for the bases in the DNA in simulated average structure of the complex compared to the crystal structure. The Tilt values were also found to be more irregular in the simulated structure than in the crystal structure. The observed differences may be due to the fact that the contacts of the complex with the other neighboring units in the crystal are lost in the solution environment, allowing enhanced local effects. The indirect effects of A-helix bias of CHARMM parameters may also be partially responsible (Yang and Pettitt, 1996; MacKerell, 1997; Feig and Pettitt, 1997; Sprous et al., 1998). Looking at the different helicoidal parameter values of the crystal structure also shows a few unusually large local values (rise, etc.), indicating that the strong interaction between DNA and the protein introduce strong local effects which may be more pronounced in solution. The data also indicate that the strong binding also cause the parameters to be significantly irregular over the entire DNA duplex in the complex. However, the average values over all the nucleotides of Rise (3.69 in the crystal and 3.74 in the simulation) and Twist (31.2 in the crystal and 28.7 in the simulation) show close similarities between crystal structure and the average structure. The average values of the different backbone torsional angles were mostly similar in the crystal structure and in the simulation average structure, except in nucleotides number 8 and 9, for which the torsional angles δ , ϵ , and ξ were found to be significantly different.

The time evolution of the interactions between DNA and protein

The time evolutions of the interactions between the component molecules are also important quantities for better char-

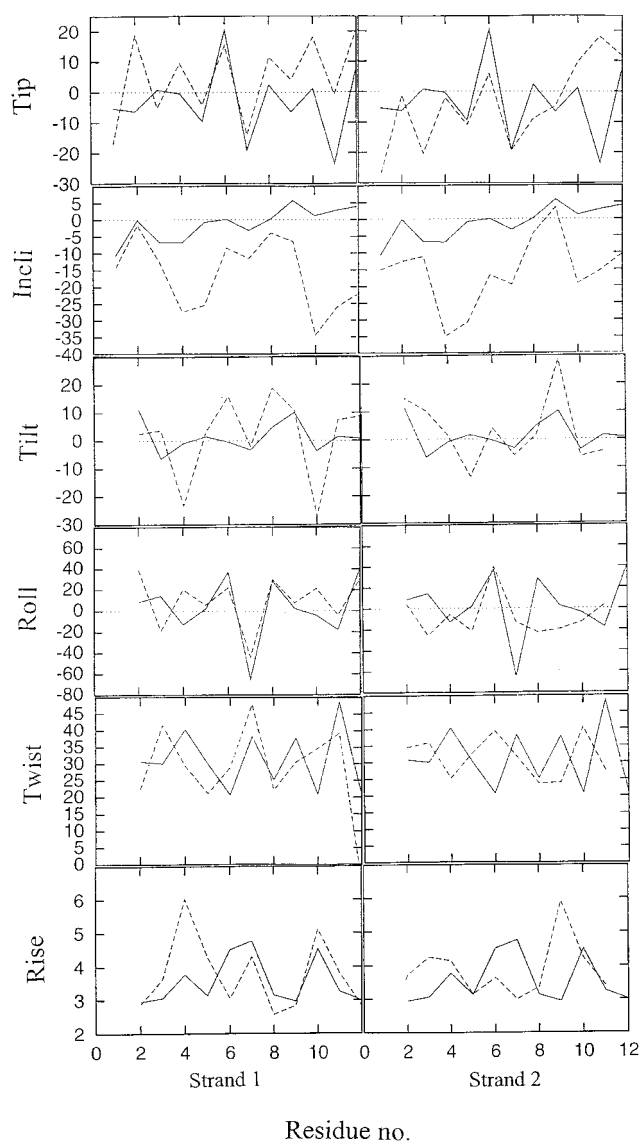


FIGURE 10 Summary of the comparison between the helicoidal parameters for the DNA in the EcoRI complex in crystal structure (*solid line*) and the average structure in aqueous solution (*dashed line*) against the DNA nucleotide number. The left column represents the first strand of DNA (DNA1) and the second column represents the other strand.

acterization of the behavior of the complex in aqueous solution. The average values evolution of the interactions of DNA1-EcoRIa and DNA2-EcoRIb over the last 300 ps of the trajectory were obtained as -351 ± 16 kcal/mol and -395 ± 21 kcal/mol, respectively. The averages were computed as the mean of the interaction energies in frames at an interval of 2 ps obtained from the last 300 ps of the dynamic trajectory. The linear correlation coefficient between these two time series was obtained as 0.34, indicating a significant degree of correlation between the dynamics of the direct interaction of these two pairs of systems in the complex. The average values for the DNA1-EcoRIb and DNA2-EcoRIa, calculated in the same way are found to be -138 ± 13 and -133 ± 11 kcal/mol, respectively. The linear cor-

relation coefficient between these two time series was found to be 0.04, indicating practically uncorrelated dynamics. Calculation also showed that the two protein monomers interact favorably with each other and their mutual interaction contributes considerably to stabilizing the complex. For the time evolution of the interaction between the two monomers of the protein it was found that during the early part of the simulation a more favorable interaction was grown, which gradually disappeared, and the interaction reached a steady state with an average interaction strength -808 ± 49 kcal/mol, which is only slightly less favorable than in the initial state. The physical reason behind this behavior seems to be the interaction with water molecules. During initial dynamics only the local interactions between the monomers dominate and grow faster. As the dynamic simulation continues, the complex relaxes and more and more water molecules get access to the protein and adjust themselves around the solute, making favorable interactions with the monomers. This is evident from the time evolution of the interaction energy of the protein monomers with water (discussed below), at the cost of the direct interactions between the two monomers. The time evolution of these energies and the values of the associated fluctuations clearly show that these interactions are well maintained throughout the dynamic simulation. Calculation of the linear correlation between the time series for the interaction between the two monomers of the protein and the DNA1-*EcoRIa* interaction show a linear correlation coefficient of -0.51 and for the other pair as -0.43 indicating that in each case the interactions are considerably anti-correlated to a similar degree. The DNA1-*EcoRIb* and DNA2-*EcoRIa* show correlation with the monomer-monomer interaction of the protein with correlation coefficients 0.61 and 0.17, respectively. The large difference between the correlation coefficients in these two cases is not quite clear. The overall DNA-protein interaction energy is anti-correlated with the dynamics of the monomer-monomer interaction of the protein by a correlation coefficient -0.29 . The anti-correlated nature of the time series of the individual DNA-protein interaction and the protein-protein interaction simply implies that during the equilibrium dynamics, the variation in the DNA-protein interaction energy grows or decays partially at the cost or gain of the protein-protein interaction. However, the average interaction energy between the two protein monomers was found to be stronger than any of the average interactions between DNA1 and *EcoRIa*, DNA2 and *EcoRIb*, DNA1 and *EcoRIb*, or DNA2 and *EcoRIa*, though altogether the DNA-protein interaction is stronger than the monomer-monomer interaction of the protein. Fig. 11 represents the interaction map between the two protein monomers indicating the mutually interacting regions. Only interactions >0.5 kcal/mol in magnitude were considered. The monomer-monomer interface thus consists of selected regions of the monomers. It was found that their interaction consists mostly of the H-bonding interactions and van der Waal interaction also constitutes a significant contribution. In some cases strong interaction of as much as -74.4

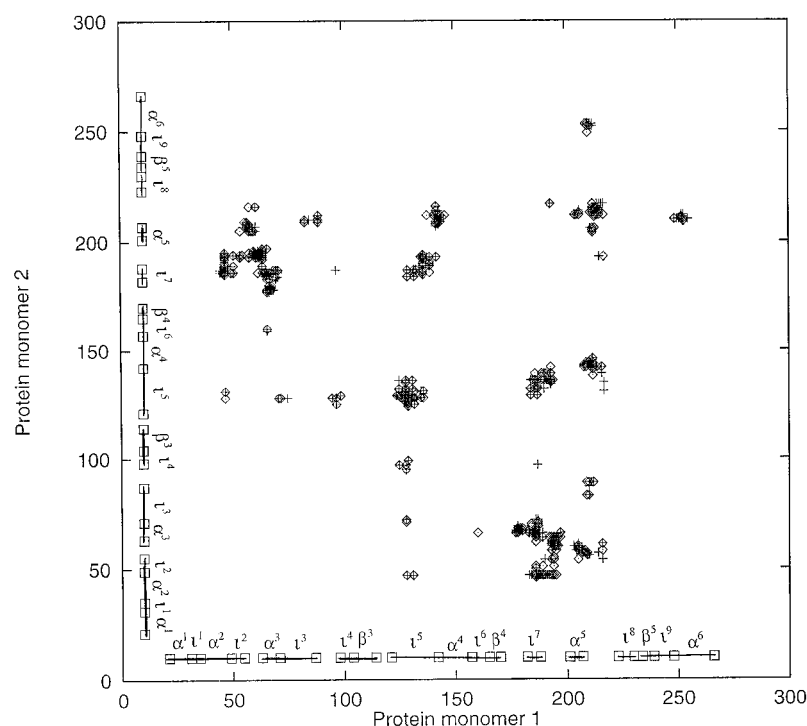
kcal/mol between two residues of the two monomers were observed. Throughout the dynamic simulation the interaction energy between the two strands of the DNA remained quite stable, with an average value of -1045 ± 24 kcal/mol.

Solvent interactions: solvent-accessible surface area, H-bonding, and water bridging

Considering the time developments of the interactions of the DNA and the protein separately with water (Fig. 12) it is seen that in both cases favorable interaction with water grew considerably with time. The interaction energy reached a stable average value for the DNA very rapidly, whereas for the protein the growth of interaction occurred in two phases: a first phase of large growth in interaction occurred very fast, then saturation was reached slowly. A double exponential ($E = E^\infty + A_1 e^{-t/t_1} + A_2 e^{-t/t_2}$) fit of the time evolution of the DNA-water interaction gives the first time constant, t_1 , as 7.2 ps and the second time constant, t_2 , as large as 1933.5 ps, indicating that the variation in the second phase is very slow over the time period of our simulation. Similarly, for the time evolution of the protein-water interaction, t_1 was obtained as 4.2 ps and t_2 as 309.4 ps. Because water molecules are small, their reorientational relaxation time is also small. Thus it seems that the variation in the DNA-water interaction energy is due mainly to the reorientations of the nearby water molecules. The time evolution of the RMSD and the self energy of DNA in the complex also indicated very little change in its structure. On the other hand, it appears that the rapid increase in the protein-water interaction in the early part of the simulation occurs due to the reorientational relaxation of the nearby water molecules, and the slow growth part results due to the relaxation of the protein monomers themselves to enhance their interaction with water. This physical picture is consistent with the above fitting results.

In order to investigate how these interactions of the solvent with the solute are related to the solvent accessible surface area (ASA) of the solute complex, we analyzed different aspects of the ASA value of the whole complex and also separately for the molecular components. Fig. 13 *a* represents the time evolution of the ASA value of the whole complex. It is clearly seen that during the first 100 ps of dynamics in solution, the complex was gradually loosened up with a considerably larger overall ASA value of the complex and maintained that state throughout the simulation. We also compared the solvent-accessible surface area of the molecular components of the complex between its initial crystal structure and the average values calculated from different frames at an interval of 1 ps from the last 200 ps of the dynamic trajectory. Table 2 summarizes the results for comparison. It was found that the ASA for both the protein and the DNA have increased significantly in the simulation compared to that in the crystal structure. It is also seen from Table 2 that major contributions to the enhance-

FIGURE 11 Comparison of the residue-residue maps of the interaction between the two monomers (*EcoRIa* and *EcoRIb*) of the protein in crystal structure (\dagger) and in simulation average structure (+). The major secondary structures are also indicated.



ment in the protein's ASA is caused by the increased solvent exposure of the charged, polar and the glycine units of the protein while the non-polar residues showed a significant decrease in ASA value due to hydrophobic effect. Calculation also shows that the hydrophobic residues have the least average ASA (19.7 \AA^2) per residue and the charged residues have the largest value (67.5 \AA^2). An overview of the difference in solvent exposure of the individual protein residues of the monomer *EcoRIa* between the crystal structure and the solution simulation is given in Fig. 13 *b* as a histogram plot of the difference ($\Delta\text{ASA} = \text{ASA}_{\text{ave}} - \text{ASA}_{\text{crys}}$) in the residue-wise ASA values of the *EcoRIa* protein monomer in crystal (ASA_{crys}) and in the average

solution structure (ASA_{ave}). A positive bar for a residue in this plot represents an enhancement of the solvent exposure of that residue while a negative bar indicates a reduction of solvent exposure. Considerable enhancement of exposure for some residues are seen clearly, while for some other residues reduction of solvent exposure can be observed. The residues in the interior of the complex have zero exposure and do not show any change. Thus, the ASA calculations also are consistent with the idea that the growth of favorable solvent interaction is partly due to some structural changes of the protein.

To have an idea about the amount of surface area of the DNA duplex actually accessible to the solvent, we compared the ASA values of the DNA duplex when it is in the complex and when it is isolated from the complex. The results for the crystal structure as represented in Table 3 indicate that although a large fraction of the DNA surface is covered by the protein, a significant amount is still accessible to the solvent. The average values calculated from the dynamic average (Table 2) show a significant increase in the ASA value for the DNA duplex and this allows the DNA to form considerable number of H-bonds with water and also water bridges with protein. Analysis of H-bonds of water with the DNA indicates that strong H-bonds are mostly formed with the phosphate group oxygens and occasionally with the O5' and O3' atoms of the DNA backbone. However, in some cases the O₂, H42, of Cyt, O₂ of Thy, N3 of Ade, and N3, H22, O6, and N7 atoms of Gua bases were found to participate in H-bonding with water. Hydrogen bonds between water and the protein occur extensively with both the peptide backbone and the side chains.

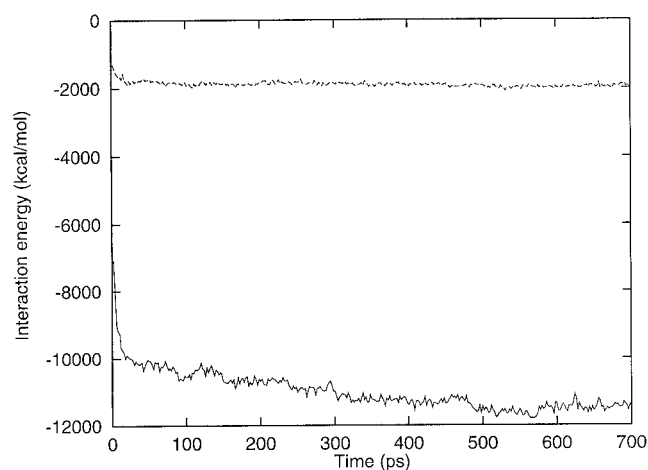


FIGURE 12 Time evolutions of the interactions of water with the protein (solid line) and the DNA (dashed line).

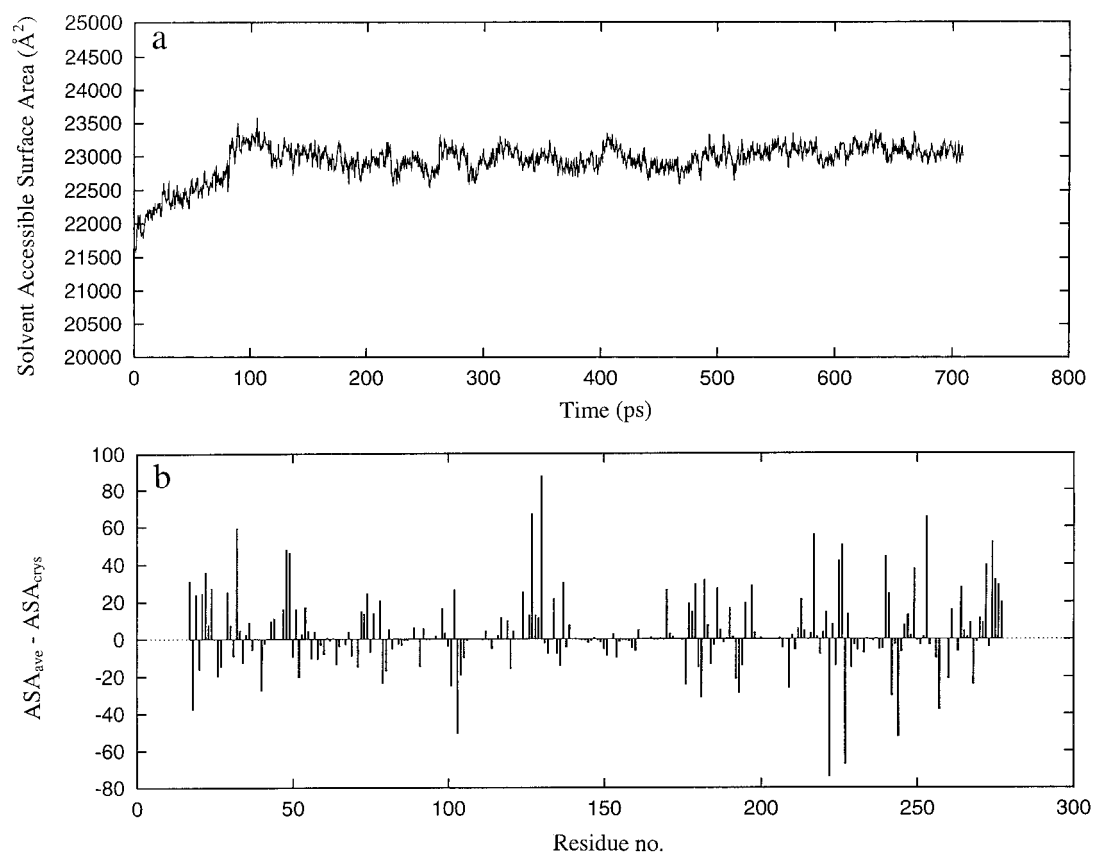


FIGURE 13 (a) Time evolution of the solvent-accessible surface area of the whole complex and (b) plot of the difference (DASA = $ASA_{ave} - ASA_{cryst}$) between the ASA values of the individual residues of the protein monomer *EcoRIa* in crystal (ASA_{cryst}) and in solution average structure (ASA_{ave}).

We have separately analyzed the water bridges between DNA and protein and those between protein and protein. The results are summarized in Table 4, A and B. We found that the DNA formed a considerable number of water bridges with the protein, which are quite long-lived, as seen in Table 4 A. We have also found several strong water bridges between the two monomers of the proteins. All these water bridges increase the overall stability of the entire complex. It may be pointed out that in the present study we did not find any water bridge between the DNA and the protein involving the bases which could enhance the specificity of the DNA-protein interaction. It is clearly seen from Table 4 A that water bridges are formed mainly involving the oxygen atoms of the phosphate groups and the O3' atoms of the DNA backbone. We have not found any

water bridge between the two strands of the DNA, which may be a consequence of the fact that the DNA is substantially buried into the protein, but have found some water bridges between different atoms of the same DNA strand, particularly at the ends of the helix. In some cases it is observed that depending on the geometric criteria (i.e., distance and angle) for H-bonds used, calculation shows breaks in the bridging interaction with one atom pair while at the same time developing a new bridge between another atom pair in the same locality. The water molecule only changes its orientation and, as a result, remains trapped in that locality for quite a long time, but the average lifetime for the bridges are less compared to the overall residence time of a water molecule in that locality. In most cases a specific atom pair is bridged by a water molecule over a

TABLE 2 Solvent-accessible surface areas

System	Total (Å ²)	DNA (Å ²)	Protein (Å ²)	Hydro (Å ²)	Charged (Å ²)	Polar (Å ²)	Gly (Å ²)
Crystal	21642.7	2406.9	19235.7	4358.8	8521.3	5607.0	748.6
Trajectory	23032.8	2617.3	20474.9	4029.3	8970.9	6498.6	976.1
average	±139.1	±38.2	±112.3	±60.1	±102.0	±90.9	±28.9

Comparison of the solvent-accessible surface area of the different molecular components of the complex and different residue types of the protein between the crystal structure and the trajectory averages over the last 200 ps taking every frame at an interval of 1 ps.

TABLE 3 Comparison of solvent-accessible surface area values for the DNA in the DNA-*EcoRI* complex

Residue no.	DNA in the complex		DNA isolated from complex	
	Whole residue	Base	Whole residue	Base
1	226.5 (152.3)	98.3 (72.2)	234.8 (238.2)	100.6 (105.2)
2	42.4 (54.4)	21.4 (25.4)	203.2 (208.4)	38.6 (47.5)
3	25.8 (6.0)	12.3 (3.8)	200.8 (186.2)	40.3 (26.2)
4	27.2 (7.8)	3.3 (1.1)	164.1 (162.9)	22.5 (20.4)
5	36.9 (28.9)	8.5 (9.8)	148.9 (150.4)	22.8 (24.6)
6	58.7 (39.6)	0.8 (0.8)	193.7 (192.5)	47.5 (53.7)
7	45.1 (24.6)	1.9 (2.4)	188.1 (179.1)	52.7 (51.4)
8	86.5 (94.2)	10.7 (7.8)	187.1 (209.8)	39.2 (50.4)
9	147.5 (123.5)	14.5 (12.9)	197.8 (204.9)	42.5 (40.8)
10	162.3 (167.3)	16.2 (20.7)	162.6 (173.3)	16.2 (28.7)
11	161.2 (186.7)	16.5 (46.3)	163.7 (187.2)	17.9 (46.4)
12	336.2 (338.4)	112.2 (100.4)	335.5 (336.6)	111.7 (98.6)
Total	1356.3	317.8	2377.3	549.5

Values in parentheses are for the other strand.

substantial part of the trajectory where the bridging water molecule is some times exchanged with another water molecule or the same water molecule renews the bridge after a break making the average lifetime shorter.

TABLE 4A DNA-protein water bridges

Atom pairs on DNA and protein connected by water bridges	Total residence time (ps)	Average lifetime (ps)
DNA1:2G-O2P ······ HN-Asn85:EcoRIa	109 (1)	8.6
DNA1:2G-O3' ······ HG-Ser87:EcoRIa	63 (4)	9.0
DNA1:2G-O3' ······ O-Asn85:EcoRIa	63 (4)	9.0
DNA1:2G-O1P ······ O-Asn195:EcoRIb	109 (1)	10.9
DNA1:3C-O1P ······ O-Asn199:EcoRIb	257 (2)	16.0
DNA1:3C-O2P ······ HE-Arg203:EcoRIb	147 (1)	18.3
DNA1:3C-H42 ······ O-Gly196:EcoRIb	29 (1)	7.2
DNA1:5A-O2P ······ O-Lys89:EcoRIa	255 (1)	17.0
DNA1:5A-O2P ······ HZ3-Lys148:EcoRIa	183 (1)	13.0
DNA1:5A-O2P ······ HD22-Asn149:EcoRIa	59 (1)	7.3
DNA1:6A-O1P ······ OD1-Asp91:EcoRIa	197 (3)	17.9
DNA1:6A-O2P ······ ND1-His114:EcoRIa	56 (7)	8.0
DNA1:7A-O1P ······ OD1-Asn173:EcoRIa	117 (1)	14.6
DNA1:7A-O2P ······ O-Gly116:EcoRIa	196 (1)	14.0
DNA2:2G-O1P ······ O-Asn194:EcoRIa	199 (1)	12.4
DNA2:2G-O1P ······ O-Asn194:EcoRIb	263 (1)	18.7
DNA2:3C-O1P ······ HE-Arg203:EcoRIa	143 (1)	9.7
DNA2:3C-O1P ······ O-Asn199:EcoRIa	181 (2)	15.0
DNA2:4G-O3' ······ OD2-Asp91:EcoRIb	272 (1)	27.2
DNA2:4G-O3' ······ HN-Asp91:EcoRIb	139 (1)	10.6
DNA2:6A-O1P ······ ND1-His114:EcoRIb	52 (1)	8.6
DNA2:6A-O2P ······ ND1-His114:EcoRIb	225 (5)	15.0
DNA2:7T-O2P ······ O-Gly116:EcoRIb	113 (2)	14.2
DNA2:8T-O1P ······ HZ2-Lys117:EcoRIb	67 (1)	13.4
DNA2:8T-O1P ······ HZ3-Lys117:EcoRIb	49 (2)	12.2
DNA2:8T-O1P ······ HN-Lys117:EcoRIb	49 (1)	9.8

List of water bridges formed between the DNA and the protein, their total residence time and average lifetime over the last 300 ps of the dynamic trajectory considering each frame at an interval of 1 ps. The number within the parentheses represents the number of water molecules participated in the bridging by replacing one another during the period. Only cases with residence time >50 ps are considered.

Interactions and distribution of counterions

Calculation of the average interaction energies of the counterions with the DNA and separately with the protein indicates that the overall ion-protein interaction energy (-539.5 ± 56 kcal/mol) is stronger than the ion-DNA interaction energy (-342.4 ± 33 kcal/mol). The reason seems to be the relative size of the molecular components. As the protein is very large compared to the DNA and in the complex, it covers a large part of the DNA double helix, the ions, like the water molecules, cannot get much access to the DNA; rather, they get much more access to the protein. As a result the overall interaction energy with protein is found to be more favorable than that with the DNA. However, the interaction energy per residue indicates much stronger interaction with the DNA than with the protein, as expected. Fig 14 shows the superimposed positions of the counterions over the last 200 ps of the dynamics trajectory at an interval of 1 ps obtained after removing the overall rotational and translational motions of the solute from their coordinates. It was found that the ions are mobile but always stay close to the surface of the complex and there are several local spots on the molecular surface where the ions stay most of the time with occasional moving over to a nearby similar spot. These local sites correspond mainly to the phosphate groups of the DNA exposed to the solvent and the negatively charged groups of the protein residues accessible to the solvent.

Relaxation and energetics of the kink in the DNA

The time evolution of the RMSD of the kinked DNA isolated from the DNA-*EcoRI* complex with respect to its starting kinked structure shows a rapid growth of the RMSD value over the first 200 ps of dynamic simulation and then the average value is stabilized around 5.5 Å (Fig. 15). This

TABLE 4B Water bridges between protein monomers

Atom pairs of the protein monomers forming water bridges	Total residence time (ps)	Average lifetime (ps)
EcoRIa:74Asp-OD2 ······ O-Phe222:EcoRIb	70 (4)	8.5
EcoRIa:76Asp-OD1 ······ HZ1-Lys221:EcoRIb	51 (1)	10.2
EcoRIa:85Asn-HN ······ O-Asn194:EcoRIb	63 (1)	8.0
EcoRIa:87Ser-O ······ HE-Arg203:EcoRIb	93 (1)	11.6
EcoRIa:159Ser-OG ······ ND1-His225:EcoRIb	29 (1)	9.6
EcoRIa:194Asn-O ······ HN-Asn85:EcoRIb	177 (1)	11.0
EcoRIa:209Tyr-HH ······ HE21-Gln233:EcoRIb	205 (1)	15.7
EcoRIa:209Tyr-HH ······ OD1-Asn155:EcoRIb	192 (1)	10.6
EcoRIa:224Asn-HN ······ O-Ile73:EcoRIb	57 (1)	11.4
EcoRIa:224Asn-HN ······ OD2-Asp74:EcoRIb	57 (1)	11.4
EcoRIa:227Asp-OD2 ······ HG1-Ser159:EcoRIb	65 (1)	8.1
EcoRIa:228Lys-O ······ O-Leu158:EcoRIb	190 (1)	19.0
EcoRIa:229Ser-HN ······ O-Ser229:EcoRIb	199 (1)	18.1

List of water bridges between the protein monomers, their total residence time and average lifetime over the last 300 ps of the dynamic trajectory. We have considered only those bridges whose total residence time is >50 ps.

FIGURE 14 Plot of the superimposed positions of the counterions obtained from snapshots at every 2 ps from the last 200 ps of the trajectory after removing the overall rotational and translational motions from them. The average structure of the complex is also shown. The two protein monomers are represented in red and green and the ion positions are represented in yellow.



RMSD behavior clearly indicates that the initial distorted structure of the kinked DNA isolated from the complex has been relaxed properly during the MD simulation in aqueous solution. Similar structural studies on the *EcoRI* recognizing kinked DNA duplex by MD simulation have been performed in a recent work (Duan et al., 1997) and have found that the kinked DNA is relaxed into a straight B-DNA like structure without any kink. X-ray crystallographic studies also indicate that the DNA base sequence adopts a straight B-like structure in crystalline environment (Dickerson and Drew, 1981; Drew and Dickerson, 1981; Wing et al., 1980). All these results together suggest that the kink in the DNA is not an intrinsic structural property of the DNA as a consequence of the specific base sequence but is induced into the structure due to the binding of the protein monomers. It should be pointed out here that as we have used the CHARMM22 parameter set (MacKerell et al., 1995), which has a bias toward the A-form of DNA (Yang and Pettitt, 1996; MacKerell, 1997; Feig and Pettitt, 1997; Sprous et al., 1998), we have also found the relaxed DNA in this case in a form closer to the canonical A-form than to the B-form. The RMSD of the energy-minimized average structure of the free DNA in water with reference to the canonical B-form DNA is 5.2 Å, whereas it is only 2.3 Å with reference to its canonical A-form. On the other hand, it is now known that this base sequence has a B-like structure (Dickerson and Drew, 1981; Drew and Dickerson, 1981; Duan et al., 1997). Thus, the A-form bias of CHARMM22 parameter set may affect the present estimation of kink

energy. To minimize the error coming from this factor we have compared the self energies of only the recognition part of the DNA duplex, where the kink is actually situated. The average values and the fluctuations of the self energies in the two cases (free DNA and DNA in the complex) are estimated from the 100 frames at 2-ps interval from the last 200 ps of their respective dynamic trajectories as described in the method section, as $\langle E_{\text{kinked}} \rangle = -364 \pm 15$ kcal/mol and $\langle E_{\text{relax}} \rangle = -395 \pm 16$ kcal/mol and the difference in the

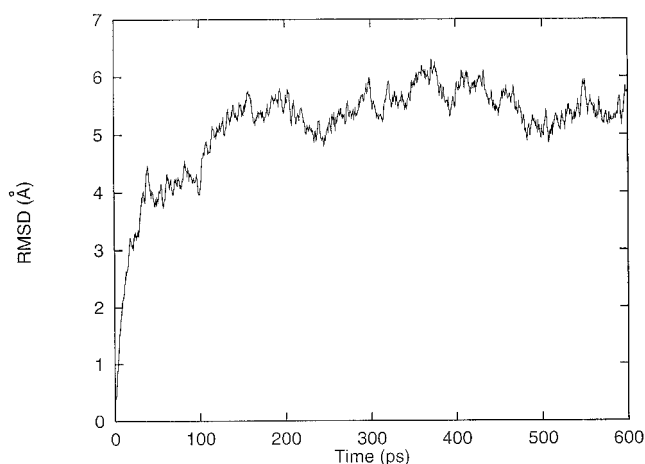


FIGURE 15 Time development of the RMSD of the DNA in the dynamic simulation of the kinked DNA isolated from the DNA-EcoRI complex.

TABLE 5 Comparison between the total and the individual energy components of the recognition part of the DNA for the complex and for the free DNA averaged over the last 200 ps of the respective trajectories considering each frame at interval of 2 ps.

	E_{tot}	E_{bond}	E_{ang}	E_{vdw}	E_{elec}	E_{dih}
E_{comp}	-363.7 ± 15.6	104.8 ± 9.7	214.7 ± 11.0	-34.0 ± 7.6	-846.1 ± 9.4	101.3 ± 5.0
E_{free}	-395.4 ± 15.8	102.7 ± 8.6	212.7 ± 10.9	-45.5 ± 8.5	-854.9 ± 8.9	94.3 ± 5.1
$E_{\text{comp}} - E_{\text{free}}$	31.7	2.1	2.0	11.5	8.8	7.0

All energies are expressed as kcal/mol.

average self energies $\Delta E = \langle E_{\text{kinked}} \rangle - \langle E_{\text{relax}} \rangle$ over the time period is found to be approximately 31 kcal/mol. Details of the relative contributions of the different individual energy components in the average self energies of the kinked DNA in the complex and in the relaxed free DNA are given in Table 5, showing that the major contribution in the DNA kink energy comes from the E_{vdw} , E_{elec} , and E_{dih} components. Considering the dynamic fluctuations in each of the quantities $\langle E_{\text{kinked}} \rangle$ and $\langle E_{\text{relax}} \rangle$ and the fact that in the present simulations the relaxed DNA is neither in canonical B-form nor in canonical A-form, the difference due to the A-helix bias does not seem to be very significant and the present estimation thus gives a reasonable idea of the kink energy. However, it must be emphasized here that, as the DNA kink formation is associated with the binding of the protein monomers to the DNA, not only enthalpic but also unfavorable entropy changes are involved in this binding process (Duan et al., 1996). So the energetics of the DNA kink can be described properly only in terms of the free energy difference, and here we have made an estimation of only the enthalpic part of this free energy. The estimation of the free energy difference should be treated as a separate issue for this type of complex case.

SUMMARY AND CONCLUSIONS

We have performed a detailed MD simulation study of the interaction, structure, dynamics, and solvent interactions of the DNA-*EcoRI* complex in aqueous solvent. The main results and conclusions from the present study may be summarized as follows.

1. A 7-Å solvent shell seems to provide a reasonably good and realistic solvent environment for performing the MD simulation studies of large molecular systems (like the present *EcoRI*-DNA complex), in order to investigate their structural, interactional, and dynamic aspects and also their explicit interactions with water molecules. No significant evaporation of water molecules or smearing of the outer boundary of the solvent shell takes place over the 0.7-ns dynamic simulation. The diffusion coefficients of water in the shell was found to be similar to the standard values for TIP3P water model and the diffusion coefficient for the Na^+ counterions was also found to be consistent with other studies based on more detailed setups. Time evolution of the global properties of the system were found to be very similar to that obtained in the more general spherical deformable boundary setup.

2. The overall structure and shape of the complex was well maintained during dynamic simulation in the solution and the average structure is not very different from the crystal structure. Very little structural change occurred for the DNA ($\text{RMSD} = 1.89 \pm 0.15$ Å), whereas relatively more structural rearrangements took place for the protein, as indicated by a RMSD value of 2.63 ± 0.12 Å. It was found that the time series of the RMSD of the DNA and the protein over the last 100 ps of the trajectory are significantly correlated with a linear correlation coefficient of 0.4, indicating that even in the relaxed complex, the DNA and the protein are tightly bound to each other. It was also found that most structural rearrangements occurred in the protein's largest loop, Asn¹²¹ to Ala¹⁴².

3. The major interactions (both specific and nonspecific) observed in the crystal structure of the *EcoRI*-DNA complex were also observed to be present in the solution average structure and are found to be maintained throughout during the 700-ps-long dynamic simulations. Most of the specific interactions were found to be enhanced in the simulation, due either to improved interaction geometry or development of additional interactions. It was identified that as in the crystal structure, the residues Gua⁴ and Ade⁶ of the DNA provide the most important specific interactions and the nucleotides Cyt³, Gua⁴, and Ade⁵ constitute the strongest nonspecific interactions for the DNA in the average structure obtained from the simulation. Similarly, the residues Ala¹³⁸, Gly¹⁴⁰, Ala¹⁴², and Arg¹⁴⁵ of the protein are the most strongly interacting units, contributing specificity in the complex, and Lys¹¹³, Lys¹¹⁷, Arg¹⁴⁵, and Lys¹⁴⁸ are the protein residues that provide the most prominent nonspecific interactions, making the complex most stable. Thus, any alteration in these specific residues of the DNA and the protein should directly cause the most severe effect on the structure and stability of the complex.

Some additional nucleotides outside the recognition sequence in the 5' side of the strand interact appreciably with the protein both in crystal structure and in solution simulated structure. These interactions may have some role on the binding and sliding mechanism of the complex.

4. Comparison of the specific and nonspecific interaction energies between the DNA and the protein indicates that the nonspecific interaction energy is considerably larger than the specific interactions both in the crystal structure and in the average structure in solution simulation. The range of the most prominent specific interaction is only 5 to 10 kcal/mol in the crystal structure and 5 to 35 kcal/mol in the

simulation average structure, while that for the non-specific interaction is 60 to 90 kcal/mol in the crystal structure and 60 to 120 kcal/mol in simulation average structure.

5. The interaction pattern of the protein with the DNA is highly local in the one-dimensional sequence of the protein. Each protein monomer interacts with both strands of the DNA. The protein monomers also interact favorably with each other and contribute strongly to the overall stability of the complex. All these intermolecular interactions are well maintained throughout the dynamic simulation. DNA-protein and protein-protein interactions were found to be anti-correlated. Favorable interactions of the protein monomers with water molecules grows very fast during the early part of the simulation, which corresponds to the interaction developed due to the proper orientation of the water molecules around the protein, whereas the slow growth of the interaction during the later part of the simulation is due to the structural relaxation of the complex providing further interactions with the solvent.

The DNA-protein interactions were found to remain quite stable over the entire simulation time.

6. Analysis of the solvent-accessible surface area indicates that in crystal structure the DNA double helix was substantially covered by the protein, but during dynamics the ASA of the DNA and that of the whole complex were increased significantly and a considerable number of strong, long-lived water bridges were formed between the DNA backbone and the protein. A significant number of water bridges between the two protein monomers was also found. All of these water bridges increase the overall stability of the entire complex. However, no water bridge involving the DNA bases and the protein was found in the present simulation.

7. Comparison of the interactions of the DNA and the protein with the counterions shows more favorable interaction energy with the protein than with the DNA due to its larger size and more access to the ions. The counterions were found to stay close to the surface of the complex and there are several spots on the molecular surface where the ions reside most of the time during dynamics; occasionally, the ions move from one spot to another.

8. The average DNA kink energy was approximately estimated to be 31 kcal/mol and it was found that the energy components E_{vdw} , E_{elec} , and E_{dih} in order, contribute most to the kink energy. The A-form bias of CHARMM22 parameter set relaxed the free DNA into a structure close to the A-form of the DNA in the present simulation. However, as in the present simulation the relaxed DNA is neither in the canonical B-form nor the canonical A-form and the quantities $\langle E_{\text{kinked}} \rangle$ and $\langle E_{\text{relax}} \rangle$ show considerable fluctuations, the effect of the A-helix bias does not seem to be very significant in this respect.

REFERENCES

- Alexander, M., J. J. Bonvin, M. Sunnerhagen, G. Otting, and W. F. van Gunsteren. 1998. Water molecules in DNA recognition II: a molecular dynamics view of the structure and hydration of the *trp* operator. *J. Mol. Biol.* 282:859–873.
- Allen, M. P., and D. J. Tildesley. 1987. *Computer Simulations of Liquids*. Oxford Science, Oxford.
- Billeter, M. 1996. Homeodomain-type DNA recognition. *Prog. Biophys. Molec. Biol.* 56:211–225.
- Billeter, M., P. Guntert, P. Luginbuhl, and K. Wuthrich. 1996. Hydration and DNA recognition by homeodomains. *Cell.* 85:1057–1065.
- Billeter, M., Y. Q. Qian, G. Otting, M. Muller, W. J. Gehring, and K. Wuthrich. 1993. Determination of the nuclear magnetic resonance structure of an Antennapedia homeodomain-DNA complex. *J. Mol. Biol.* 234:1084–1093.
- Bishop, T., D. Kosztin, and K. Schulten. 1997. How hormone receptor-DNA binding affects nucleosomal DNA. *The role of Symmetry. Biophys. 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. Com. Chem.* 4:187–217.
- Brooks, C. L., M. Karplus, and B. M. Pettitt. 1989. *Proteins: a theoretical perspective of dynamics, structure and thermodynamics*. In *Advances in Chemical Physics*. Wiley, New York. 61.
- Brooks, III, C. L., and M. Karplus. 1983. Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.* 79:6312–6325.
- Draper, D. E. 1993. Protein-DNA complexes: the cost of recognition. *Proc. Natl. Acad. Sci. USA.* 90:7429–7430.
- Dickerson, R. E., and H. R. Drew. 1981. Structure of a B-DNA dodecamer II. Influence of base sequence on helix structure. *J. Mol. Biol.* 149:761–786.
- Duan, Y., P. Wilkosz, and J. M. Rosenberg. 1996. Dynamic contributions to the DNA binding entropy of the *EcoRI* and *EcoRV* restriction endonucleases. *J. Mol. Biol.* 264:546–555.
- Duan, Y., P. Wilkosz, M. 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.
- 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., and L. Nilsson. 1995. Structure, thermodynamics and cooperativity of glucocorticoid receptor DNA-DNA-binding domain in complex with different response elements: molecular dynamics and free energy perturbation study. *J. Mol. Biol.* 253:453–472.
- Feig, M., and B. M. Pettitt. 1997. Experiment vs. force fields: DNA conformation from molecular dynamics simulations. *J. Phys. Chem. B.* 101:7361–7363.
- Hirsch, S. D., and A. K. Aggarwal. 1995. Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.* 14:6280–6291.
- Jeltsch, A., J. Alves, H. Wolfes, G. Maass, and A. Pingoud. 1994. Pausing of the restriction endonuclease *EcoRI* during linear diffusion on DNA. *Biochemistry.* 33:10215–10219.
- 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.
- Kandadai, N. S., and M. R. Reddy. 1996. Solution structure of papin as studied by molecular mechanics and molecular dynamics techniques. *J. Comp. Chem.* 17:1328–1338.
- Karplus, P. A., and C. Fearman. 1994. Ordered water in macromolecular structure. *Curr. Opin. Struct. Biol.* 4:770–776.
- Kim, Y., J. C. Grable, R. Love, P. J. Green, and J. M. Rosenberg. 1990. Refinement of *EcoRI* endonuclease crystal structure: a revised protein chain tracing. *Science.* 249:1307–1309.

- Koneshan, S., J. C. Rasaiah, R. M. Lynden-Bell, and S. H. Lee. 1998. Solvent structure, dynamics, and ion mobility in aqueous solution at 25°C. *J. Phys. Chem. B* 102:4193–4204.
- Kosztin, D., T. C. Bishop, and K. Schulten. 1997. Binding of the estrogen receptor to DNA: the role of waters. *Biophys. J.* 73:557–570.
- Lavery, R., and H. Sklenar. 1988. The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.
- Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
- Lesser, D. R., M. R. Kurpiewski, T. Waters, B. A. Connolly, and L. Jen-Jacobson. 1993. Facilitated distortion of the DNA site enhances *EcoRI* endonuclease-DNA recognition. *Proc. Natl. Acad. Sci. USA* 90:7548–7552.
- Lesser, D. R., M. R. Kurpiewski, and L. Jen-Jacobson. 1990. The energetic basis of specificity in the *EcoRI* endonuclease-DNA interaction. *Science* 250:776–785.
- Lindahl, E., and O. Edholm. 1998. Solvent diffusion outside macromolecular surfaces. *Phys. Rev. E* 57:791–796.
- 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.
- MacKerell, A. D. 1997. Influence of magnesium ions on duplex DNA structural, dynamic, and solvation properties. *J. Phys. Chem. B* 101:646–656.
- MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkerich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics. Studies of proteins. *J. Phys. Chem. B* 102:3586–3616.
- 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.
- McCammon J. A., and S. C. Harvey. 1987. *Molecular Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, New York.
- Misra, V., J. L. Hecht, K. A. Sharp, R. A. Friedman, and B. Honig. 1994. Salt effects on protein-DNA interactions: the λ cl repressor and *EcoRI* endonuclease. *J. Mol. Biol.* 238:264–280.
- Newman, M., T. Strzelecka, L. F. Dörner, I. Schildkraut, and A. K. Aggarwal. 1994. Structure of restriction endonuclease BamHI and its relationship to *EcoRI*. *Nature* 368:660–664.
- Otwinoski, Z., R. W. Schvitz, R. G. Zhang, C. L. Lawson A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335:321–329.
- Qian, Y. Q., G. Otting, and K. Wuthrich. 1993. NMR detection of hydration water in the intermolecular interface of a protein-DNA complex. *J. Am. Chem. Soc.* 115:1189–1190.
- Ravishanker, G., S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar. 1989. Conformational and helicoidal analysis of 30ps of molecular dynamics on the d(CGCGAATTCGCG) double helix. *J. Biomol. Struct. Dyn.* 6:669–699.
- Robinson, C. R., and S. G. Sligar. 1993. Molecular recognition mediated by bound water: a mechanism for star activity of the restriction endonuclease *EcoRI*. *J. Mol. Biol.* 234:302–306.
- Robinson, C. R., and S. G. Sligar. 1994. Hydrostatic pressure reverses osmotic pressure effects on the specificity of *EcoRI*-DNA interactions. *Biochemistry* 33:3787–3793.
- Rosenberg, J. M. 1991. Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* 1:104–113.
- Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* 23:327–341.
- Schwabe, J. W., R. L. Chapman, J. T. Finch, and D. Rhodes. 1993. The crystal structure of the estrogen receptor DNA-binding domain bound to the DNA: how receptors discriminate between their response elements. *Cell* 75:567–578.
- Sprou, D., Ma. A. Young, and D. L. Beveridge. 1998. Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: theoretical considerations of the A 219 B transition. *J. Phys. Chem. B* 102:4658–4667.
- Szczelkun, M. D., P. Janscak, K. Firman, and S. E. Halford. 1997. Selection of non-specific DNA cleavage sites by the type IC restriction endonuclease *EcoRI*. *J. Mol. Biol.* 271:112–123.
- van der Spoel D, P. J. van Maaren, and H. J. C. Berendsen. 1998. A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *J. Chem. Phys.* 108:10220–10230.
- van Gunsteren 1993. Molecular dynamics and stochastic dynamics simulation: a primer. In *Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications*, Vol 2. ed. by W. F. van Gunsteren, P. K. Weiner and A. J. Wilkinson, Leiden, The Netherlands ESCOM. 3–36.
- Venclovas, C., A. Timinskas, and V. Syksnys. 1994. Five stranded β -sheet sandwiched with two α -helices: a structural link between restriction nucleases *EcoRI* and *EcoRV*. *Proteins Struct. Funct. Genet.* 20:279–282.
- Wing, R., H. R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1980. Crystal structure analysis of a complete turn of B-DNA. *Nature* 287:755–758.
- Yang, Y., and B. M. Pettitt. 1996. B to A transition of DNA on the nanosecond time scale. *J. Phys. Chem. B* 100:2564–2566.