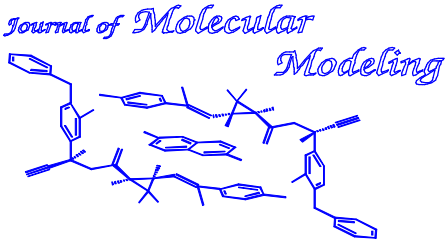
J. Mol. Model. **1996**, *2*, 427 – 445



A Molecular Dynamics Study of the Repressor/Operator(OR1,OR3) Complexes from Bacteriophage 434

Laurent David and Martin J. Field

Laboratoire de Dynamique Moléculaire, Institut de Biologie Structurale – Jean-Pierre Ebel, 41, Avenue des Martyrs, 38027 Grenoble Cedex 1, France (ldavid@ibs.fr and mjfield@ibs.fr)

Received: 21 August 1996 / Accepted: 2 October 1996 / Published: 11 November 1996

Abstract

We have performed three molecular dynamics simulations using the CHARMM molecular modeling program to study the repressor protein from bacteriophage 434 complexed with DNA operators of two different sequences. Two approaches to the modeling of the solvent were used. In the first method, applied to the R1-69/OR1 truncated complex, water molecules were included explicitly in conjunction with a stochastic boundary force to solvate the complex. In the second approach, used for simulations of the R1-69/OR1 and the R1-69/OR3 complexes, the solvent was omitted and implicitly represented by using a distance-dependent dielectric constant and a scaling of the charges on the exposed residues. The simulation with the model which explicitly includes the solvent serves as a validation of the simulations using a simpler solvent representation. In our discussion of the results we focus upon the important interactions between the DNA binding motif of the 434 repressor (motif helix turn helix) and the operators and how the structures of the complexes change with time.

Keywords: Protein-DNA complex, repressor protein, bacteriophage 434, molecular dynamics, solvation models

Introduction

The regulation of gene activity in cells from a wide variety of organisms has been the subject of intense study. Much of the basis for this regulation is due to the specificity of the interaction of various proteins with DNA. Early studies which gave insight into these effects, were based primarily upon biochemical techniques [1, 2, 3]. More recent investigations have made increasing use of structural data obtained from X-ray crystallography.

Systems which have been a focus of investigation are the λ, P22 and 434 bacteriophages [4, 5, 6] in which the regulation is mediated by two proteins, cro and a repressor, which bind to the DNA. Three structures of the 434 repressor complexed to DNA are available from the Protein Data Bank [7]. They are the R1-69/OR1 [8], the R1-69/OR2 [9] and the R1-69/OR3 [10] complexes. These structures have all been determined at 2.5 Å resolution. Their availability provides a very powerful resource for investigating and understanding the crucial interactions that occur between the protein and different sequences of DNA. However, in spite of the high resolution structures, the relative affinities of the 434 repressor for its different operators are not fully understood. Thus, a comparison of the complexes reveals structural differences which do not fully explain the variation in the affinities.

Biochemical studies based on the existing structures have clarified the effect of some specific DNA bases or protein residues in the complexation and highlight some major contacts. To support the first crystallographic low resolution

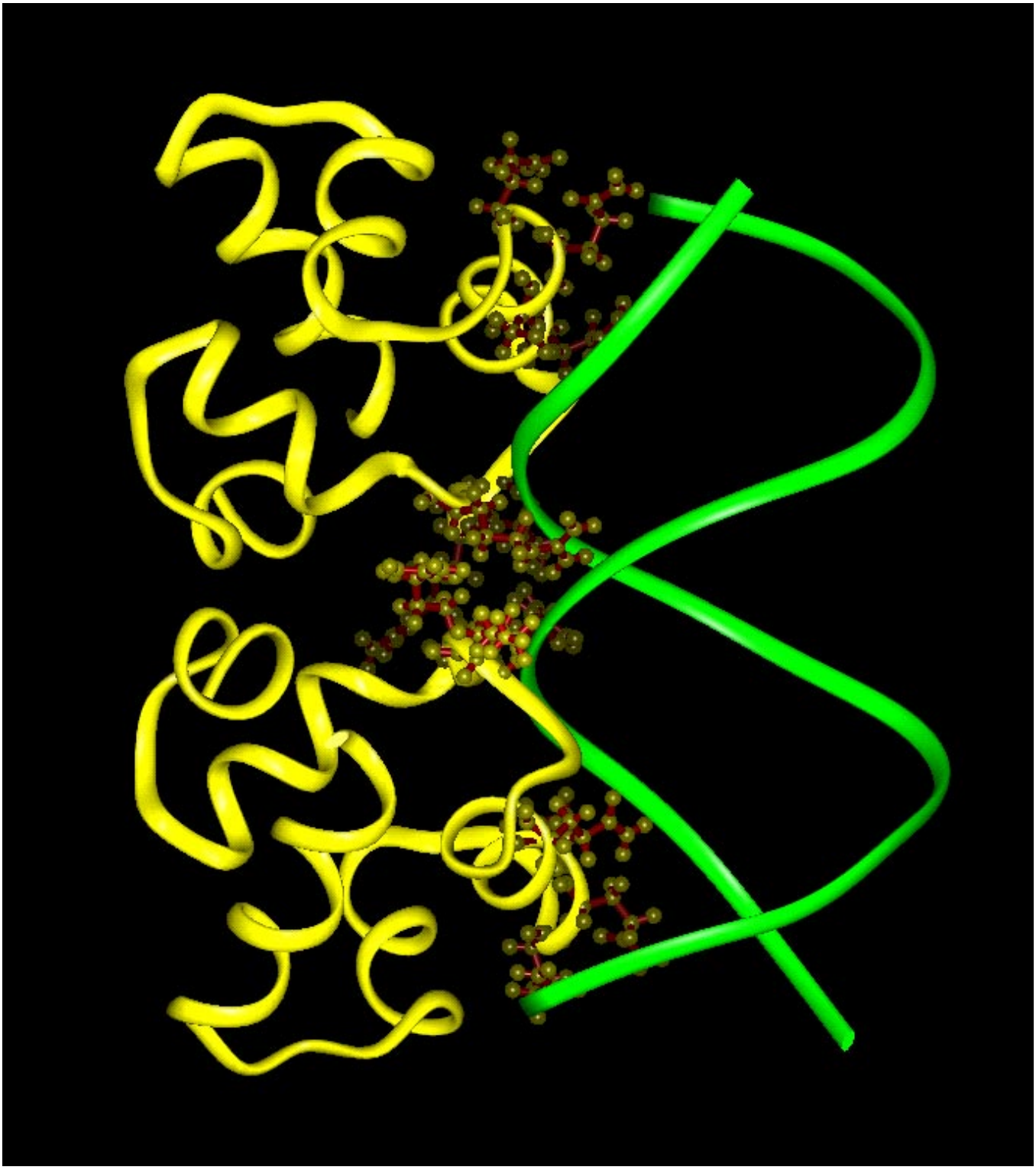


Figure 1. *An image of the protein-DNA complex. The DNA is green, the protein backbone yellow and the atoms of the amino acids which interact with the DNA strand are shown with a ball and stick representation.*

model of the complex [11], the ethylation interference method [12, 13] was used to determine protein-DNA contacts between the 434-repressor and OR1. This study showed that the truncated 434 repressor protein, called R1-69, had the same interaction at the phosphate level with the OR1 operator as with the entire protein.

Table 1. *Calculated pK_a values of the ionizable groups in the protein. The values in the left half of the protein are listed first and then the right half.*

residue	N-term	arg 5	lys 7	lys 9	arg 10	glu 19	lys 23	glu 32
pK_a	5.5	12.9	10.6	10.2	12.9	3.8	10.7	3.6
residue	glu 35	lys 38	lys 40	arg 41	arg 43	glu 47	asp 57	C-term.
pK_a	2.1	10.2	9.7	12.9	11.5	3.7	3.6	3.2
residue	N-term.	arg 5	lys 7	lys 9	arg 10	glu 19	lys 23	glu 32
pK_a	5.5	12.9	10.3	10.4	12.9	4.0	9.6	2.9
residue	glu 35	lys 38	lys 40	arg 41	arg 43	glu 47	asp 57	C-term
pK_a	2.3	11.1	9.8	12.5	11.4	2.4	3.4	3.6

In more recent work, Koudelka and Lam [14] and Bell and Koudelka [15, 16] have worked on differentiating the affinity of the 434 repressor for the OR1 and the OR3 operators. They made mutants of the OR1 operator to transform it, step by step, to OR3 [15]. Bell and Koudelka [16] showed a relation between the intrinsic structure of the operator and the affinity of the operator for the 434 repressor by mutating residues suspected of contributing to the specificity of the interactions between R1-69 and OR1 [14, 17]. Another study of the dependence of the DNA structure on its composition by time-resolved fluorescence polarization anisotropy revealed that the deformation in the DNA depended only slightly on its sequence [18]. As a follow up to this work, Koudelka and Carlson [19] showed that there is a relationship between the intrinsic twist of an operator and its affinity for repressor.

Molecular dynamics simulation is an important tool that can provide information characterizing the specific interactions that occur in complex formation. In this paper, we present the results of three molecular dynamics simulations which were performed with the aim of understanding the specificity of the interactions between R1-69 and its different operators. The first simulation is a 0.5 nanosecond (ns) molecular dynamics simulation of the explicitly hydrated complex R1-69/OR1. The objective was to analyze the different interactions which occur between the protein and a part of the DNA and to show the effect of water on this complexation. The second part of the study is a comparison of two 0.5 ns simulations, carried out on the R1-69/OR1 and the R1-69/OR3 complexes. These simulations use a simplified representation of the electrostatic interactions which allows us to simulate the complete complexes whereas in the first simulation we truncated the operator sequence.

The outline of this paper is as follows. In section the next section, we detail the techniques used to perform and analyze the simulations. we than present a discussion of the results followed by our conclusion.

Methods

The initial structures for all three simulations were obtained from the Brookhaven Protein Data Bank (PDB) [7]. These were the crystal structure of Aggarwal and co-workers for the R1-69/OR1 complex [8] and the structure of Rodgers and co-workers for the R1-69/OR3 complex [10]. The preparation of both structures for simulation is described in the following sections.

Determination of the Protonation States of the Ionizable Groups in R1-69

For the simulations, it was necessary to choose the protonation states of the various ionizable amino acid groups in the protein that were appropriate for the conditions used in the crystallization of the complex (pH=6 for R1-69/OR1 and pH=5.5 for R1-69/OR3). Four types of residue were considered to have variable protonation states – arginine, aspartate, glutamate and lysine as well as the C and N-terminal ends of the protein chain (there are no histidines in the protein).

Recent progress in theoretical techniques has shown how to choose the protonation states of amino acid residues to agree with experiment [20, 21, 22]. The new methods are based upon the resolution of the Poisson-Boltzmann equation which determines the change in electrostatic potential at each residue due to changes in the protonation state of the protein. These changes are directly related to the pK_a of the ionizable groups.

We used these methods to determine the charges of ionizable groups for the simulation of the R1-69 protein. Our calculations combined the cluster method described by Gilson [23] with the electrostatic free energy calculation method described by Antosiewicz et al. [22]. The Poisson-Boltzmann equation was solved using the package UHBD [24]. A Richards probe accessible surface definition with a dielec-

tric constant of 80 for water and 20 for the protein was used. Calculations were performed with a probe sphere radius of 1.4 Å, a Stern layer of 2.0 Å and an ionic strength of 200 mM for the solvent surrounding the complex.

In Table 1, the different pK_a values obtained for the ionizable groups in the uncomplexed R1-69 repressor protein are listed. No calculations of the protein complexed with the OR1 operator were performed because the parameters (charges and radius) [24] have not been optimized for DNA. The pK_a values of the acidic groups are all less than 4 and so they are all charged at pH 5.5 and 6 (the pH of crystallization of the complexes). We protonate the N-termini, which have the same probability of being charged as uncharged. Similarly, the basic groups are all protonated under the same conditions because they all have pK_a greater than 9. Because of the approximate two-fold symmetry, the pK_as are very similar for both subunits of the protein.

Parameters and Protocols

The molecular modeling program CHARMM23 (version F2) was used for all the simulations [25] along with the most recent all-atom CHARMM force field (version 22) [26]. For both types of simulation, we used the Verlet algorithm with a 1 fs time step for the integration of Newton's equations. All simulations were run for 0.5 ns (≡ 500 picoseconds (ps)) and coordinate sets were saved each 0.1 ps giving 5000 structures for each trajectory. The initial velocities were assigned to the atoms from a Maxwell-Boltzmann distribution with a temperature of 300 K.

Molecular Dynamics Simulation with Explicit Solvent. We took the crystal structure of R1-69/OR1 [8] and deleted the three bases at the ends of each piece of DNA to obtain a complex, R1-69/OR1, close to the one described by Anderson et al. [11]. Seventeen sodium ions were added to the system. Each position was chosen so that it lay between 3 and 4 Å from the phosphorus atom of a phosphate group along the bisector of the two free oxygen atoms [27]. The resulting system had a total charge of -3 comprised as follows: -28 (OR1), + 8 (R1-69), +17 (Na^+). To hydrate the system, a sphere with a radius of 30 Å containing water molecules of type TIP3P [28] was superimposed upon the complex. All water molecules closer than 2.8 Å from a protein or a DNA atom, or 2 Å from a Na^+ ion, were deleted.

The system was subdivided into two regions, a sphere of radius 27 Å where molecular dynamics was performed, and a shell between 27 and 30 Å where Langevin dynamics was performed. To mimic the effect of the environment outside the sphere, a stochastic boundary approximation was used which adds a term to the energy function that approximates the mean-field interaction due to the solvent [29].

One hundred steps of minimization, using a conjugate gradient algorithm, were performed on the system with all non-water molecules kept fixed. To fill up any holes created by the reorganization of the water molecules, a sphere of water was superposed with the system and all the water molecules of the sphere closer than 2 Å from an atom of the system were removed. The remaining water molecules were then added to the system. This "two step" cycle of minimization and superposition was repeated three times, until a correct hydration of the system was deemed to have occurred. The final system had 12765 atoms. The cutoff distance for the calculation of the non-bonded (electrostatic and Lennard-Jones) interactions was taken to be 13 Å and a switching function was applied between 12 and 13 Å. All the atoms of the protein and the DNA were free to move during the simulation.

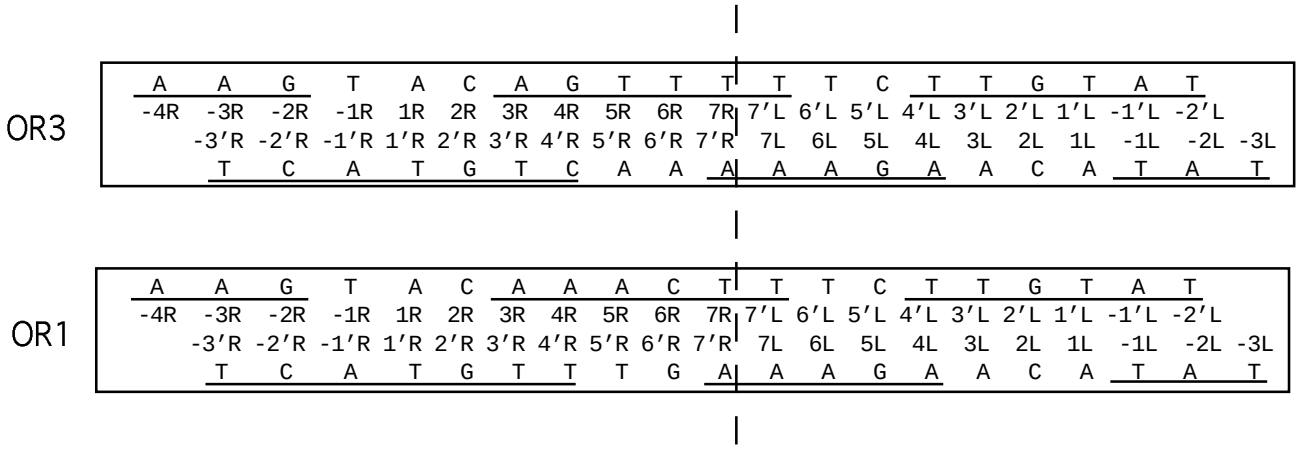


Figure 2. *Sequence of bases for the DNA operators OR1 and OR3. The first sequence represents the OR3 operator and the second the OR1 operator. The numbering of the DNA is the crystallographic numbering with the operator consisting of right (R) and left (L) parts. The bases whose exposed charges were scaled by a factor of 0.3 for the simulations with the implicit solvent model are underlined.*

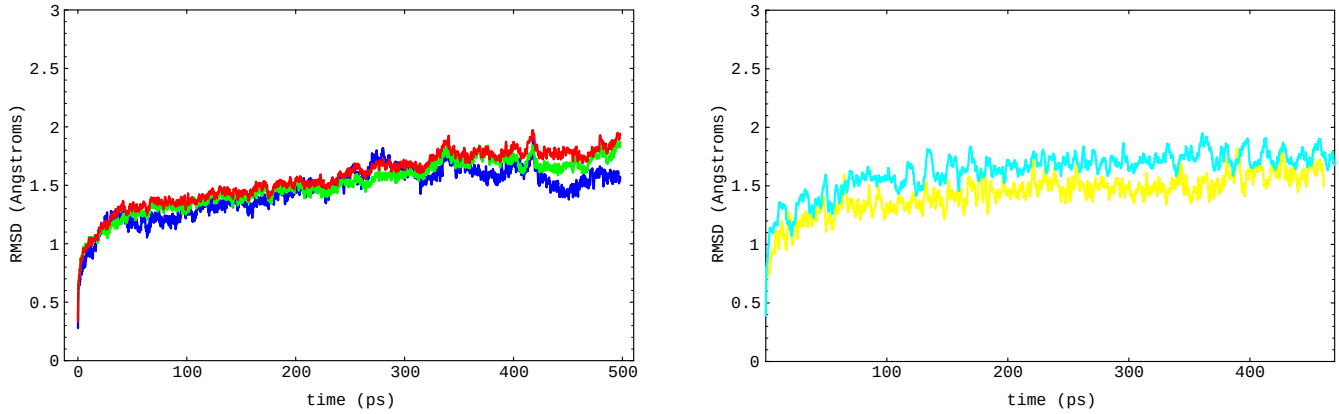


Figure 3. *RMS coordinate deviations for the complexes from the simulations: (a) for the protein (green), DNA (blue) and the complex (red) from the explicit simulation; (b) for the complexes from the implicit simulations, OR1 (blue) and OR3 (yellow).*

Molecular Dynamics Simulation with Implicit Solvent. To run the simulations with an implicit model for the solvent, we modified our treatment of the electrostatic interactions used in the previous simulation. On the basis of previous studies, the charges of the charged protein residues pointing toward the solvent were reduced [30]. For the DNA, we reduced the charges of the sugar-phosphate backbone which were not protected by the protein. This reduction was done by multiplying each charge by 0.3 and mimics two major physical effects – the reorientation of the water molecules, which has a shielding effect, and the effect of the counterion layer, which is not explicitly simulated. We used a van der Waals radius representation of the atoms in the complex to help us choose if the residue or base charges should be scaled. The sequence of the bases in the operators and a list of those bases whose charges were scaled as a result of this procedure are shown in Figure 2. It is to be noted that although the complexes were not explicitly solvated in the simulations, the crystallographic waters (44 for OR1 and 40 for OR3) **were** included in the simulation as these have important structural roles.

The structures were minimized until convergence, which was taken to be when the root mean square gradient fell below 0.1 kcal $mol^{-1} \cdot Å^{-1}$. The non-bonded interaction cutoff distance was 15 Å. The extremities of the DNA were fixed during the entire simulation (last 3 bases). A distance-dependent dielectric was used so that the electrostatic interactions were calculated using the following expression

$$E_{elec} = \sum_{i<j<N} \frac{q_i q_j}{\varepsilon r_{ij}^2} \qquad (1)$$

ε was set to 1 for all simulations. Guenot and Kollmann have obtained a reasonable representation of various dynamical properties, including the RMS coordinate deviations and fluctuations of proteins, by using an equivalent model [31].

Each simulation was preceeded by 20 ps of heating and 30 ps of equilibration (in which the temperature of the system was constrained to be between 290 and 310 K) before data collection began.

Results and Discussion

In this section, we describe the results of all three simulations. Firstly we present a general analysis of the results for all the simulations before we compare the differences between them in detail.

Behaviour of the Dynamics. To characterize the general behaviour of the dynamics, we calculated the root mean square coordinate deviations (RMSD) between the starting structures of each simulation and the structures from the dynamics trajectories. All the structures collected were first superimposed upon the initial structure using the algorithm due to Kabsch [32]. The RMS coordinate differences were then calculated between the starting and the reoriented structures. The results for the three dynamics simulations are shown in Figure 3. The RMSD between the crystallographic structures and the equilibrated structures (structures at t=0 in the dynamics) are, 0.7Å for the OR1 complex with explicit water, 1.8 Å for the OR1 complex with implicit water and 1.4 Å for the OR3 complex respectively.

For the simulation with explicit water we show the RMSD value for the operator, the repressor and the full complex. The RMSD values for the DNA are slightly inferior to the other RMSD values. At about 300 ps the three curves seem to reach a plateau value of about 1.6 Å for the protein and slightly higher for the full complex. The RMSD curves for the simulations with the implicit solvent model reach their stable values more quickly than those for the explicit solvent. For both implicit simulations the RMSD values are rela-

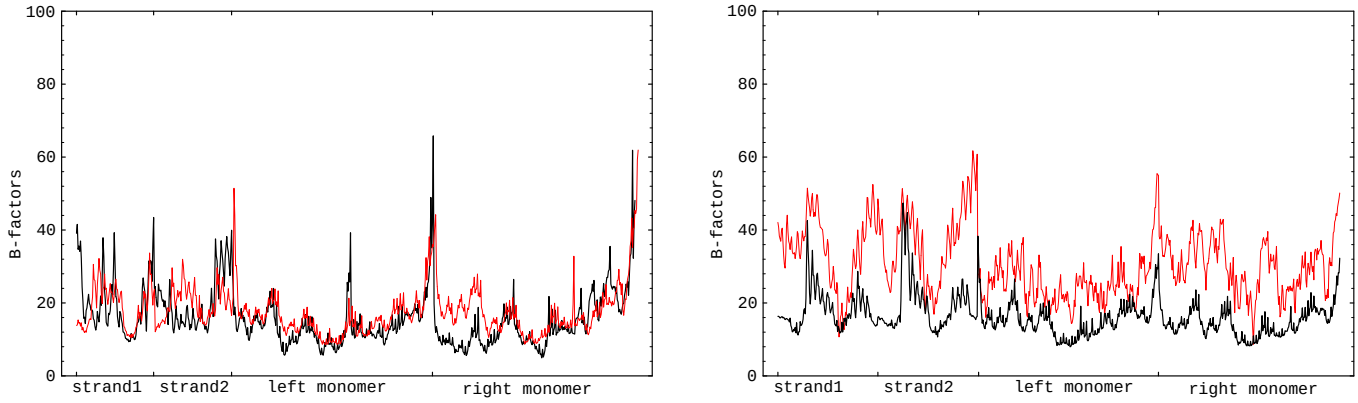


Figure 4. *The B-factors (Å^2) for the atoms of the complex for each of the three simulations: (a) the calculated B-factors for the OR1 complex are superposed with the results for the explicit simulation in black and the implicit simulation in red; (b) the calculated B-factors for the OR3 complex from the implicit simulation (black) are compared to the experimentally obtained ones (red).*

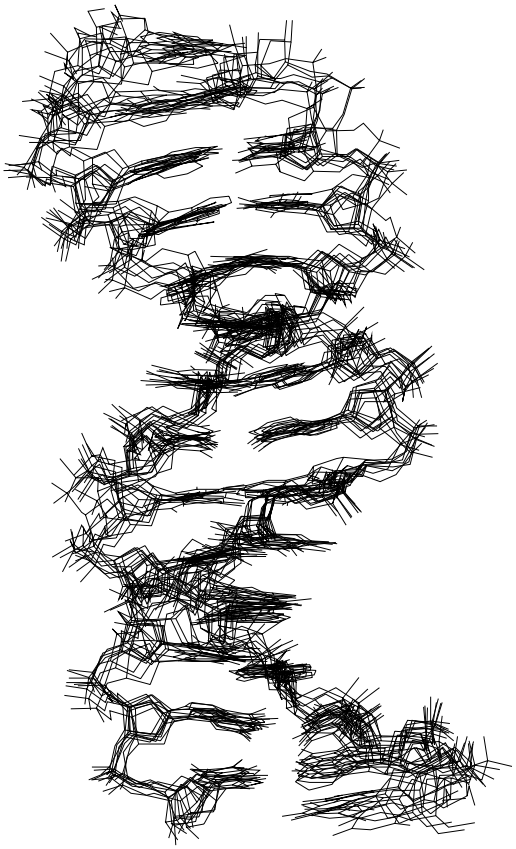


Figure 5. *Superposition of nine structures of the OR1 operator taken from the simulation with explicit solvent. All hydrogens are deleted.*

tively stable throughout the length of the simulation and stay around 1.6 Å for both complexes.

The mean coordinate fluctuations for the atoms in the complexes are shown in Figure 4. The fluctuations were computed for the backbone heavy atoms (N, CA, C and O) of the protein and for the P, O'x and C'x atoms of the DNA. The values are multiplied by 26.32 (= 8 π^2 / 3) so as to compare them directly with the crystallographic B-factors [33].

There is a reasonable correlation between the fluctuations calculated from the simulation with the OR3 operator and the implicit solvent model and the crystallographic B-factors. The calculated fluctuations are smaller than the experimental values (by about a factor of 2) but the higher mobility regions correspond in both sets of data. For the simulations with the OR1 operator a comparison with the B factors is more difficult as the latter values are very different from those for the OR3 complex. However, a comparison between the two simulations of this complex (one with implicit and one with explicit solvent) shows very similar behaviour with the largest fluctuations in roughly the same regions of the operator and the repressor. The results obtained by Arnold and Ornstein, in their study of different solvent models [34], showed greater values with the implicit simulations than with the explicit ones. A comparison of the fluctuations in the OR1 and OR3 complexes shows that while the fluctuations in the DNA are similar the fluctuations in the C-terminal regions of the protein chains are much larger.

Structure of the Hydrated R1-69/OR1 Complex

The DNA Structure. We superimposed 9 OR1 operator structures obtained at 50 ps intervals in the range 100 – 500 ps. They are displayed in Figure 5. The structure is well maintained even at the ends where there are no constraints. In contrast, structures from the first 100 ps do not superimpose very well due, in large part, to a shortening in the DNA strands which takes 100 ps to complete.

To obtain a clearer idea of structural changes, the widths of the minor and major grooves were calculated at various points along the operator. These were estimated as the minimum distance between two phosphates of each strand (a more

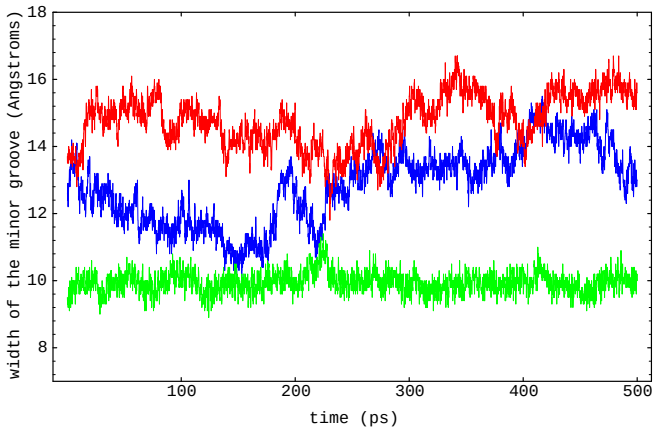


Figure 6. *The width of the minor groove of the OR1 operator DNA taken from the simulation with explicit solvent (green for the center, and blue and red for the extremities of the operator).*

rigorous analysis using the program Curves (see below) gives the same results). The major groove width is approximately constant throughout the simulation along the entire sequence. Its mean width is approximately 17.5 Å which compares with the value for a B-DNA strand of about 17.3 Å.

The minor groove shows more variation. This is shown in Figure 6. For B-DNA the average minor groove width is 11.5 Å while for the crystallographic R1-69/OR1 structure it is 8.8 Å in the middle of the operator and 14 Å at its ends. This minor groove compression at the center of OR1 is due to a curvature in the DNA produced by interactions between the sugar phosphate backbone and helices 3 and 4 of the repressor [8]. During the simulation the minor groove width at the center of the operator stays constant with a mean value between 9 and 10 Å. In contrast, the width of the grooves at the ends of the chain varies much more, although its average value is about 15 Å.

Parameters characterizing the local DNA structure were calculated with the Curves program [35] for 11 structures taken at 50 ps intervals. These parameters are given in Tables 2a, 2b and 2c. The average helical twist is 34.5° and it differs by less than 3 % from the crystallographic average. The average rise per base pair is 3.25 Å and is the same as the one in the crystallographic structure. The average propeller twist angles show the same deformation of the DNA during the simulation as those in the crystallographic model, particularly the one due to the non-Watson-Crick hydrogen bond formed between O4 of T7R and N6 of A7L and the one between N2 of G5L and O2 of T4'L. The sugar puckers are predominantly in the forms C1'exo and C2'endo except for the bases T4'R, G6'R, A4L and T3'L. These sugars are in contact with the protein and consequently more likely to be subject to conformational changes.

The R1-69 Structure. The 434 repressor has a strong homology with the λ [36] phage repressor at the DNA-binding helix turn helix motif (HTH). The RMS coordinate deviations (calculated using the Cα atoms of the motif residues only) between both subunits of the crystallographic structures of the 434 repressor and the λ repressor vary between 0.48 Å and 0.59Å (depending on which monomers of each dimers are paired).

Table 2(a). *Angle variables of the DNA during the explicit simulation.*

Base Pair	Buckle		Propel		Base Pair	Rise		Twist	
	mean	var.	mean	var.		mean	var.	mean	var.
A2-T14	5.59	5.84	-2.72	5.81	A2-C3	-3.05	.23	-35.28	2.45
C3-G13	19.61	9.33	-17.55	4.85	C3-A4	-3.85	.39	-26.49	2.45
A4-T12	-3.10	7.17	-12.27	8.34	A4-A5	-2.95	.23	-34.67	2.07
A5-T11	4.35	5.78	-12.73	6.43	A5-A6	-3.19	.20	-34.60	3.25
A6-T10	2.25	6.95	-16.42	3.74	A6-C7	-3.20	.20	-38.94	4.08
C7-G9	3.87	9.18	-11.94	7.81	C7-T8	-3.27	.32	-33.00	3.24
T8-A8	6.58	10.13	-22.74	7.19	T8-T9	-3.31	.32	-35.79	2.38
T9-A7	-.32	5.23	-28.06	6.71	T9-T10	-3.22	.24	-38.55	3.63
T10-A6	-9.77	6.88	-24.97	4.59	T10-C11	-3.37	.22	-43.17	3.57
C11-G5	-17.95	7.36	-14.38	5.26	C11-T12	-3.12	.18	-29.83	2.53
T12-A4	-9.97	6.07	-17.47	4.43	T12-T13	-3.20	.17	-38.56	3.67
T13-A3	-8.54	7.91	-2.76	4.73	T13-G14	-3.31	.35	-25.41	4.85
G14-C2	-2.02	7.50	-1.49	11.77					

Table 2(b). *Sugar pucker conformations of the DNA during the explicit simulation. The number of structures (out of 11 total) at each conformation are listed.*

C4'exo	C3'endo	O1'endo	C2'endo	C1'exo			C1'exo	C2'endo	O1'endo	C3'endo	C4'exo
						T1	6	4	1	0	0
						A2	4	3	3	1	0
1	1	4	0	5	G13	C3	3	8	0	0	0
0	0	1	3	7	T12	A4	4	1	6	0	0
2	1	1	3	4	T11	A5	7	4	0	0	0
0	0	0	8	3	T10	A6	5	6	0	0	0
4	0	6	0	1	G9	C7	9	2	0	0	0
0	0	0	4	7	A8	T8	4	7	0	0	0
0	2	2	4	3	A7	T9	2	2	7	0	0
0	0	0	11	0	A6	T10	6	5	0	0	0
0	0	0	8	3	G5	C11	4	7	0	0	0
3	2	1	2	3	A4	T12	8	1	2	0	0
0	0	1	2	8	A3	T13	0	1	1	7	2
0	0	2	6	3	C2						
0	0	2	2	6	A1						

Table 2(c). *Non-Watson-Crick hydrogen bonds for the OR1 and OR3 operator sequences for the simulations with explicit and implicit solvent models. Only non-Watson-Crick bonds which occur in about 20% or more of the structures are listed.* d_{min} *is the minimum distance between the non-hydrogen atoms that define the bond during the simulation.*

		OR1 explicit		OR1 implicit		OR3 implicit	
Base (atom)	Base (atom)	d_{min}	Freq.	d_{min}	Freq.	d_{min}	Freq.
A1R(N6)	G2'R(O6)	2.7	19.7	2.6	29.8	2.7	18.5
A3R(N1)	G2'R(N1)	2.9	23.9	2.9	11.0		
A4R(N6)G(N2)	T5'R(O4)T3'R(O2)	2.7	22.6	2.7	19.2	2.6	61.5
G4R(O6)	A5'R(N6)					2.6	66.3
A5R(N6)T(O4)	G6'R(O6)A6'R(N6)	2.6	65.3	2.5	66.2	2.6	44.0
C6R(N4)T(O4)	A7'R(N6)			2.8	38.2	2.7	16.9
T7R(O2)	G6'R(N2)	2.6	77.1	2.6	87.0		
T7R(O4)	A7L(N6)	2.6	58.8	2.6	40.6	2.6	57.2
T7'L(O4)	A6L(N6)	2.6	56.7	2.7	12.0	2.7	17.8
T6'L(O4)	G5L(O6)	2.5	76.8	2.6	44.7		
C5'L(N4)	A4L(N6)	2.8	39.9	2.9	4.1	2.8	31.0
T4'L(O2)	G5L(N2)	2.6	68.6	2.6	73.7	2.6	70.9
G2'L(N1)	A3L(N1)	2.8	33.1	2.8	14.8	2.8	29.9

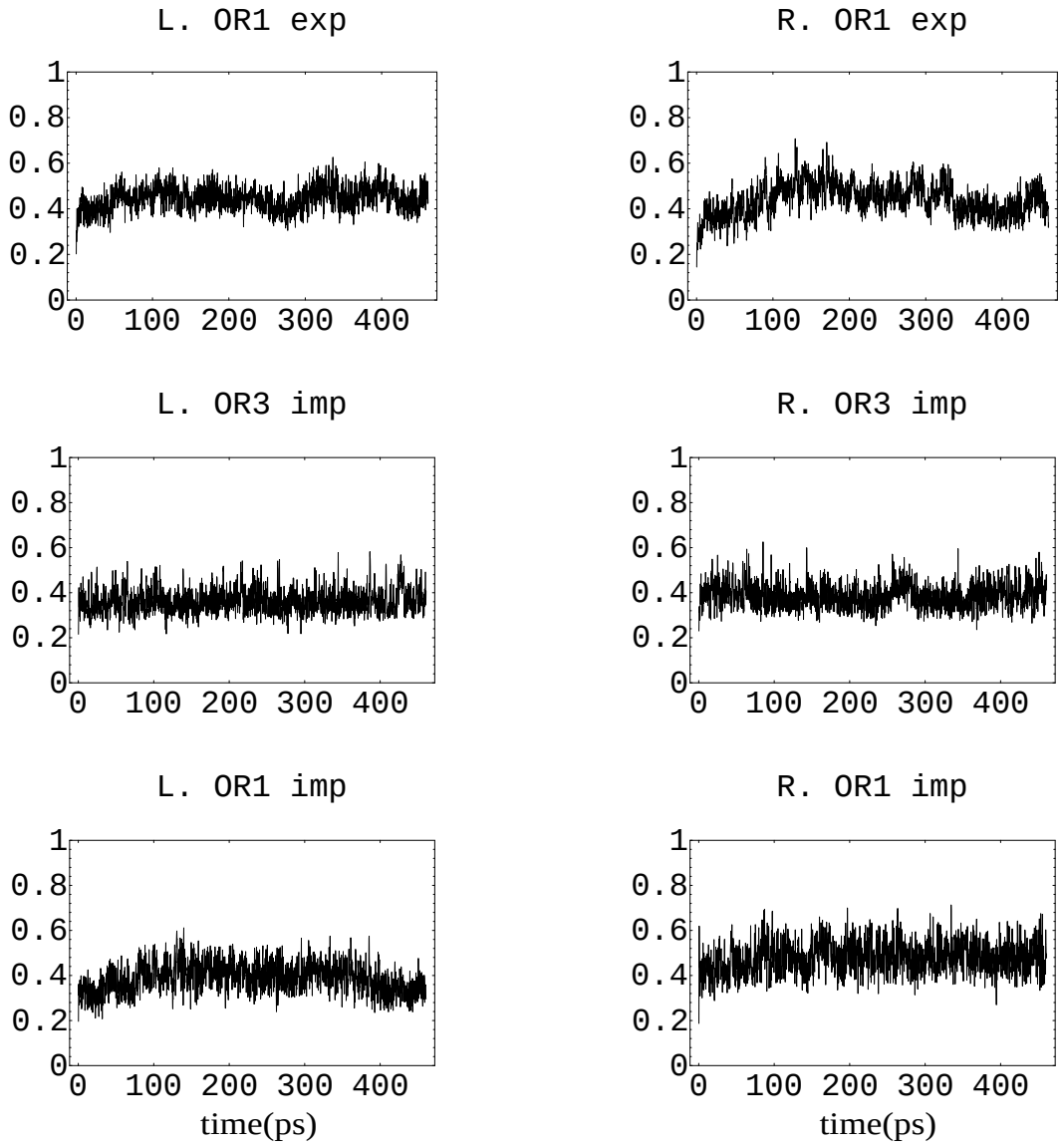


Figure 7. *The RMS coordinate deviations of the Cα atoms in the repressor helix-turn-helix motif for each of the three simulations.*

The calculated RMS coordinate deviations between the initial and intermediate trajectory structures during the simulations for this motif are shown in Figure 7 for the Cα atoms. It can be seen that, compared to the total protein RMS coordinate deviations, the HTH motif deviations are smaller and reach a plateau much more rapidly. If we compare the RMS coordinate deviations taking account of all the atoms for this motif, we obtain the same variations but with a slightly higher value.

Behaviour of the Sodium Ions. The self diffusion coefficient, D, for the sodium ions, was calculated using the Einstein relation:

$$D = \frac{1}{6}\lim_{t\to\infty}\frac{\partial}{\partial t}\left\langle \left(R(t) - R(0)\right)^2 \right\rangle \qquad (2)$$

Figure 8 shows a plot of the right hand side of equation 2 for the Na^+ ions in the explicit simulation. The diffusion coefficient is equal to $0.18(\pm 0.01) \times 10^{-9}\cdot m^2\cdot s^{-1}$. This is lower than the values calculated by other workers. For example, van Gunsteren et al. [37] determined values in the range $[0, 5 \times 10^{-9}]$ $m^2\cdot s^{-1}$ for a B-strand of DNA in water and Norbert and Nilsson [38] found a value of $1.3 \times 10^{-9}\cdot m^2\cdot s^{-1}$ for a dinucleotide in water. In our simulation, the sodium ions are all located near a phosphate, and they stayed in the vicinity of these groups during the simulation. After 500 ps, only two

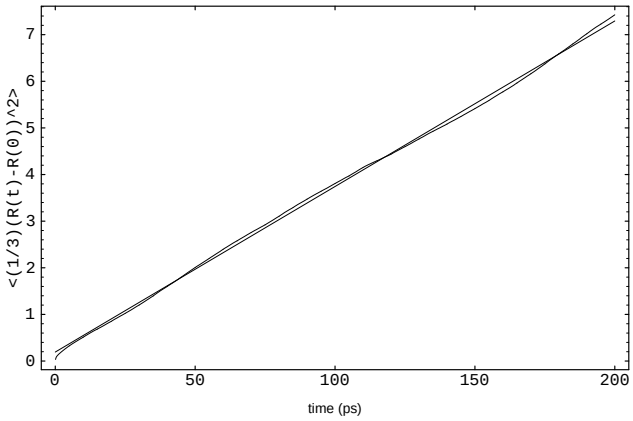


Figure 8. *A plot of the time-dependent mean square displacement (Å^2) for the sodium cations from the simulation with the explicit solvent model.*

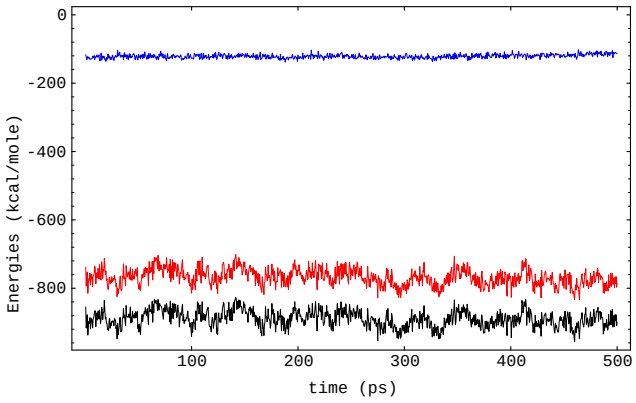


Figure 9. *The non-bonded interaction energies between the protein and the DNA throughout the simulation using the explicit water model. The van der Waals energy is in blue, the electrostatic is in red and the total non-bonded energy is in black.*

ions are located at a distance greater than 10 Å from a phosphate. This, together with the fact that the volume of the simulation system is relatively small, helps to explain the low value of the diffusion coefficient.

DNA-Protein Contacts. Crystallographic Data: Analysis of the crystallographic complex R1-69/OR1 reveals that the R1-69 homodimer binds OR1 by an interaction of helix 3 with the major groove and by contacting the sugar phosphate backbone with two NH_2 groups at the N-terminal ends of helices 2 (Asn16,Gln17) and 4 (Arg43). In the following, we only consider the right half of the complex as the left side is the same by symmetry. A1R(N6,N7) has a double hydrogen bond with Gln28, and G2'R(O4,O6) with NE2 of Gln29. Base pair 3 does not have polar contacts with R1-69, in spite of the presence of the side chain of Glu32 which points toward it. However, a van der Waals pocket is created between Thr27, Gln29 and T3'R. The fourth base pair (A-T) has a hydrogen bond with Gln33 (T4'R O4) and a van der Waals contact with Gln29 and Ser30. The other hydrogen bonds between DNA and R1-69 are with a phosphate group–T1R with Arg10 and Gln17, the NH groups of Lys40 and Arg41 with G6'R and the NH of Arg43 with T5'R. A1R has a hydrogen bond with Gln17 and Asn36, and G2R with Asn16. These last three interactions are not detected by the ethylation interference experiment [8]. There are many van der Waals contacts between the bases T5'R, G6'R, T-1R and A1R and R1-69.

Simulation Results: The electrostatic and van der Waals energies between OR1 and R1-69 are shown in Figure 9. The van der Waals energy lies between about -100 and -140 kcal·mol^{-1} throughout the simulation whilst the electrostatic interaction energy is about six times larger in absolute value and fluctuates much more.

A major part of the electrostatic interaction is represented by the hydrogen bond energy. The criterion that we used for the existence of a hydrogen bond was to say that a bond existed if the distance between the donor (D) and the acceptor (A) was less than 3.2 Å. In the case of multiple hydrogen bonds involving the same donor or acceptor, we calculated the angle D-H··· A and kept the bond with the largest angle (i.e. nearest 180°). In Table 3, we list the hydrogen bonds between the DNA and the protein which appeared in more than 10% of the structures of a trajectory. Most of the hydrogen bonds that can be seen in the crystallographic model are conserved during the simulation. The two hydrogen bonds between Lys40L (resp. Lys40R) with O1P from 6'L (or 6'R)

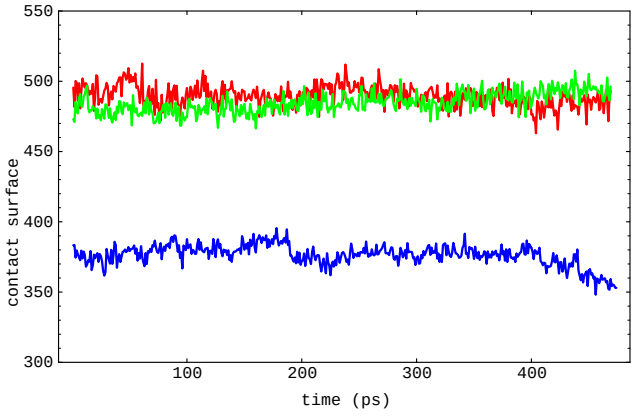


Figure 10. *The contact surface between the protein and the DNA (Å^2) for the three simulations: OR1 explicit solvent (blue); OR1 implicit solvent (red); OR3 implicit solvent (green).*

Table 3: *The hydrogen bonds between R1-69 and OR1 in the simulation with explicit water. R1 is the first strand of DNA and R2 is its complement. SL is the left subunit and SR the right. The base and residue numberings are the same as in the crystallographic model [8]. d_min is the minimum distance between the non-hydrogen atoms that define the bond during the simulation. An asterisk next to a bond means that the bond exists in the crystallographic structure.*

R1		SL				R1		SR			
Base	atom	residue	atom	d_{min}	Freq	Base	atom	residue	atom	d_{min}	Freq
7'L*	O3'	R43	NH2	2.76	49.	1R*	O5'	Q17	OE1	2.52	10.2
6'L*	O1P	K40	N	2.57	34.6	1R*	O1P	N36	ND2	2.51	47.3
6'L*	O2P	R41	O	2.68	21.4	1R*	N7	Q28	NE2	2.72	68.9
6'L*	O2P	K40	N	2.63	37.7	1R*	N6	Q28	OE1	2.56	79.8
6'L*	O2P	R41	N	2.50	99.8	1R	N6	Q29	NE2	2.83	18.7
5'L*	O2P	R43	N	2.53	94.9	2R*	O1P	K38	NZ	2.45	99.5
5'L	O2P	F44	N	2.63	34.5	6'L*	O2P	R43	NH2	2.46	95.9
5'L	O1P	S30	OG	2.47	62.4						
5'L	N4	Q33	OE1	2.57	83.9						
4'L*	O4	Q33	NE2	2.57	85.7						
3'L	O4	Q29	NE2	2.66	21.6						
2'L*	O6	Q29	NE2	2.55	70.4						
R2		SL				R2		SR			
Base	atom	residue	atom	d_{min}	Freq	Base	atom	residue	atom	d_{min}	Freq
1L*	N7	Q28	NE2	2.69	59.8	6'R	O5'	K40	N	2.79	14.4
1L*	N6	Q28	OE1	2.50	91.0	6'R*	O1P	K40	N	2.63	17.5
1L	N6	Q29	NE2	2.83	10.5	6'R*	O2P	K40	N	2.61	10.3
6'R*	O3'	R43	NH2	2.70	11.	6'R*	O2P	R41	N	2.50	98.6
6'R*	O2P	R43	NE	2.63	15.8	6'R*	O2P	R41	O	2.63	28.7
						5'R*	O2P	R43	N	2.48	99.8
						4'R*	O4	Q33	NE2	2.63	65.0
						4'R	O1P	T26	OG1	2.40	82.8
						2'R*	O6	Q29	NE2	2.53	97.4
						1'R	O4	Q29	NE2	2.62	29.4

are broken quickly, although we find that the hydrogen bond on the left half of the site is found in the average structure for the simulation, in agreement with the crystallographic structure.

Bases -2L and -1L are not in our model and the phosphate group of T-1R is not represented, so there is no possibility of van der Waals contacts here. Moreover, as no constraints were applied to the DNA strand during the dynamics, the hydrogen bonds at the end of the operator with Asn16, Gln17 and Asn36 disappear early in the simulation. Two hydrogen bonds not described in the crystallographic model, however, are observed. These are C5'L (O1P) with Ser30L (Oγ) and C5'L (N4) with Gln33L (Oε1). In the crystallographic model [8] they underline the proximity of the Oε1 atom and the C5 atom (3.6 Å) of C5'L. In the right half, these hydrogen bonds cannot occur because of the methyl group of the thymine that replaces the cytosine.

Another way to characterize the interactions between the repressor and the operator is by looking at the contact surface between them. This surface was calculated for the structures from the three simulations and is displayed in Figure 10. We compute these surfaces without taking account of the water molecules using the following expression

$$S_{contact} = \frac{1}{2}\left(S_{ADN} + S_{protein} - S_{complex}\right) \quad (3)$$

where S is the van der Waals surface calculated with a probe of radius 1.6 Å. The contact surface values for the crystallographic complexes are 362 Å^2 for the R1-69/OR1 truncated complex used in the explicit simulation, 420 Å^2 for the entire R1-69/OR1 complex and 439 Å^2 for the R1-69/OR3 crystallographic complex. The difference between the contact surfaces of the truncated and the crystallographic structures at time t=0 shows the importance of the bases that are not explicitly part of the operator sequence. In the explicit simulation, the contact surface is around 370 Å^2, close to the crystallographic value. It decreases slowly at the end of the simulation. This may be explained by the lack of the van der Waals contacts at the end of the DNA. In the implicit simulation, the surface contacts seem more stable, although there are more of them than in the crystallographic structure. This is due to a lack of water molecules, which is compensated by a closer contact between the DNA and the protein.

Interactions with Water Molecules. Forty four water molecules are present in the crystallographic structure of the complex between the repressor and the OR1 sequence. They contribute to multiple bridging interactions between the protein and DNA through hydrogen bonds. Table 4 lists those bridges observed in the simulation with lifetimes of more than 30 ps. Those between Ser30L and the phosphates of 5'L and 4'L are conserved in the simulation. The three bridges between Gln33, the DNA (C5'L and A3R) and a water molecule observed in the crystallographic structure are slightly modified in the simulation. The lifetimes are short (around 35 ps) for A4R(N6)–(Oε_2) in the right site while in the left site Gln33 contacts A3L(N6) and A4L(N6) for a large portion of the simulation. The hydrogen bonds mediated by a water mol-

Table 4. *Indirect hydrogen bonds between R1-69 and OR1 in the simulation with explicit solvent. The notation used is the same as in Table 3. <T> and n are the mean lifetime of the bridge (in ps) and the number of different intermediate water molecules which create a bridge between the protein and the DNA.*

R1		SS3				R1		SS4			
Base	atom	residue	atom	<T>	n	Base	atom	residue	atom	<T>	n
T4'L	O1P	T26	OG1	61	1	A1R	O1P	Q17	NE2	126	1
T4'L	O1P	T27	OG1	193	1	A1R	O1P	E32	O	83	1
T4'L	O1P	S30	OG	42	1	A1R	O1P	E32	OE1	80	2
C5'L	O5'	T26	OG1	117	1	A1R	O1P	E35	OE2	39	1
C5'L	O5'	S30	OG	151	1	A1R	O1P	N36	ND2	99	1
C5'L	O1P	S30	O	180	1	C2R	N4	E32	OE1	301	1
C5'L	O1P	S30	OG	32	2	A4R	N6	Q33	OE2	34	1
C5'L	O1P	T39	OG1	397	1						
T6'L	O1P	K38	O	126	2						
R2		SS3				R2		SS4			
Base	atom	residue	atom	<T>	n	Base	atom	residue	atom	<T>	n
C2L	O1P	E32	OE1	168	3	G2'R	N7	Q29	NE2	93	1
C2L	N4	E32	OE2	134	1	T4'R	O1P	T26	OG1	72	1
C2L	O1P	N36	ND2	53	1	T5'R	O1P	S30	O	94	1
A3L	N7	E32	OE2	114	1	T5'R	O1P	T39	OG1	399	1
A3L	N6	E32	OE2	282	1	T5'R	O3'	R43	NH1	33	1
A3L	N6	Q33	NE2	42	1	G6'R	O2P	R41	NH2	43	1
A4L	N6	Q33	OE1	97	1	G6'R	O1P	K40	NZ	46	2
A7'R	O4'	R43	NH1	48	1	G6'R	O1P	K38	O	41	3
A7'R	N3	R43	NH2	111	1	G6'R	O2P	R41	NE	167	2
A7'R	N3	R43	NH1	37	2	G6'R	O3'	R43	NH2	151	1

ecule between Glu35L–A1L and between Gln29L–C2L are found in the left half site in the crystallographic structure, but are located uniquely in the right half site of our simulation.

The following interactions exist in the simulation but not in the crystallographic structure. In the left half there are many indirect contacts between Ser30, Thr26 and Thr27 which do not occur in the right half. In contrast, Lys40 and Arg41 interact with the base A7'R in the right half but not in the consensus left half. Four main interactions are conserved in both halves: Glu32R with C2R and A1R, Lys38R (O) with G6'R and Thr39R with T5'R. The greater number of indirect interactions in our simulations is due to the larger number of water molecules in the simulation than in the crystallographic model and is evidence of the importance of these types of interaction for stabilizing the complex.

Comparison of the Different Simulations

The structures from the three simulations were compared using the programs Curve and Pcurve [35, 39]. The structures of the DNA and the protein are described using the helical parameters of Sklenar et al. [39] for the protein, and those given at an EMBO workshop for the DNA [40].

The global dynamical behaviour of the simulations with the implicit solvent model have already been discussed in a previous section. In Figure 3, we saw that the RMS coordinate deviations for the implicit solvent model reach a plateau after about 200 ps whose values are of the same order of magnitude as those obtained by Falsafi and Reich on their study of B-DNA [41]. The DNA's RMSD values are smaller than the protein's in the implicit model as is also the case for the explicit simulation. The RMSD values for the HTH DNA-binding motif show a variation in its structure which is slightly greater than the one from the explicit model. This fact can be explained by the lack of solvating water molecules which creates a reorganization of some hydrogen bonds.

Hydrogen Bonds. An analysis of the hydrogen-bond bridges between the protein and the DNA permits an understanding of some of the structural changes in the different simulations (see Figure 11).

There are two different interactions between R1/69 and the bases 4' and 5' in both halves of the operator OR1. The interaction in the left half is hydrophilic but hydrophobic in the right half. The hydrogen bonds between Ser30L and the base 5'L oscillate between being direct and mediated by a water molecule in both OR1 simulations, but this interaction is always mediated by a water molecule in the crystallographic structure and in the OR3 simulation. The residues Thr26L and Thr27L have direct and water-mediated hydrogen bonds in the implicit simulation with OR1 and only water-mediated contacts in the other simulations. In the right site, hydrophobic contacts replace the multiple hydrogen bonds which occur between the protein, the water and the DNA. These produce a contact between Thr26R and the base 4'R that allows the formation of a hydrogen bond. We find in both simulations with OR1 that there is a bridging interaction between T4'R(O1P)-Thr26R which is not found in the crystallographic model. This interaction is also found in the simulation of the OR3 complex. In the left half site the interactions between Thr26 and Thr27 and the DNA are either direct or mediated by a water molecule in the implicit simulation with OR1. In the other two simulations, these contacts are always via a bridging water.

The hydrogen bond between Gln33R and T4'R in both OR1 simulations does not exist in the OR3 simulation, although it was described in the crystallographic complex [10]. The water bridges Q33(NE2)–A3R and Q33(OE1)–A5'R (between R1-69 and OR3) replace these direct and specific contacts. This is the major difference between the implicit OR1 and OR3 complexes.

The interaction between Lys38L(NZ), in the implicit simulations, disappears in the explicit one. Interactions with the side chains of Arg43 are slightly different in both simulations of OR1, although the OR3 arrangement is identical to the one found in the OR1 explicit simulation.

Interactions at the end of the DNA cannot be compared with the explicit model because it was truncated. Asn16R has a weak hydrogen bond at the beginning of the simulation which disappears in the OR3 simulation. This interaction does not exist in the OR1 simulation, although it was present in the crystallographic model but not defined by the ethylation experiment [8]. The side chains of Gln17 and Asn36 have some hydrogen bonds with the DNA despite the fact that the ethylation experiment gave no indication of such an interaction. Many hydrogen bonds are created with water molecules in the three models, although some of these interactions are replaced by direct contacts between protein and DNA in the simulations with implicit solvent models. These crucial interactions are made with Glu32L(R) and Asp36L with A1L(R) and C2L and correspond to the major part of the new hydrogen bonds created in the simulations with the implicit solvent model.

Comparison of the Three Simulations with the Curve Program. The DNA Structures: The Curves program [35] calculates the global and local geometrical properties of DNA using a comprehensive set of parameters. We computed these parameters for the structures of the OR1 and OR3 operators obtained every ten picoseconds from the three simulations, leading to more than 160,000 parameters. We represent these data using a curve, dial and window representation that was inspired by the work of Swaminathan et al. [42]. To reduce the very large amount of data, we superposed all graphs obtained for OR1 in the implicit simulation, with, in a first step, those obtained for OR3 and in a second step, with those obtained for the explicit simulation. For a useful schematic representation of the helicoidal parameters, readers are referred to the paper by Lavery and Sklenar [43].

The first comparison concerns the geometry of the minor and major grooves, and the bending of the two operators.

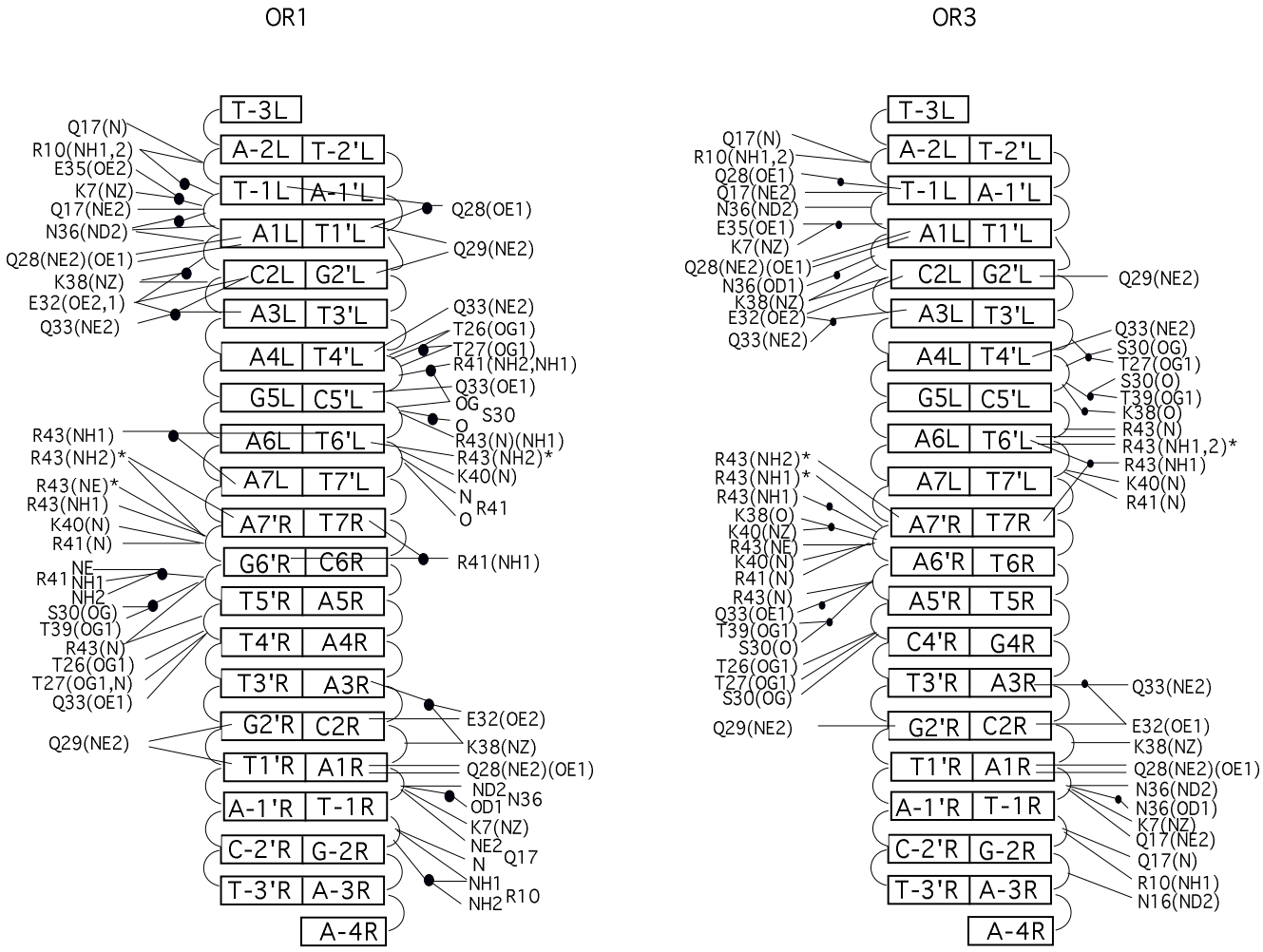


Figure 11. *Schematic representation of the different hydrogen bonds between the R1/69 protein and the DNA for the implicit simulations.*

The work of Fujimoto and Schurr [18], which measured DNA torsion constants from time-resolved fluorescence polarization anisotropy, did not show a large correlation between the sequence of the operator and its structure. Koudelka and Carlson [19] showed that there is a relationship between the sequence of the central bases of an operator and its intrinsic twist. A theoretical study of multiple sets of different sequences of DNA made by Poncin et al. [43] improved the understanding of the role of base sequence on the DNA conformation using the Curves program. They showed that many substates exist which depend on the sequence of the DNA.

The structure of the DNA is strongly linked to the state of its grooves. In Figure 12, density plots are used to show the differences in the widths of the major and minor grooves between the OR1 and OR3 sequences during the implicit simulations and between the OR1 (OR3) implicit and OR1 explicit simulations [44]. The darker the color, the lower the density, so the density is positive or negative if the color is white or black respectively.

There is a strong correlation between the depth and the width of the grooves and so we only describe the width of the major and of the minor grooves. The greatest variation is seen around the mutation triplet (AAC) for the implicit simulations. The width of the minor groove is larger for OR3 than for OR1 throughout the simulation. The depth of the major groove at the site A3R–A4R is also greater for OR3 than for OR1, which accounts for the larger accessibility of the base 4'R in OR1 to the protein than in OR3. It also explains the disappearance of the direct hydrogen bond between Q33R (NE2) and T4'R in the simulation with OR3. This is replaced by an indirect water-mediated interaction between Q33R (NE2) and T3'R. The greatest variation in the DNA structures between the explicit and implicit OR1 simulations is also found around the triplet (AAC). The major groove's width is greater for the explicit than for the implicit simulation making the base 4'R more accessible in the former.

The global curvature of DNA can be evaluated by comparing both the end-to-end distance, and the helix axis path length described by Ravishanker et al. [45]. This calculation

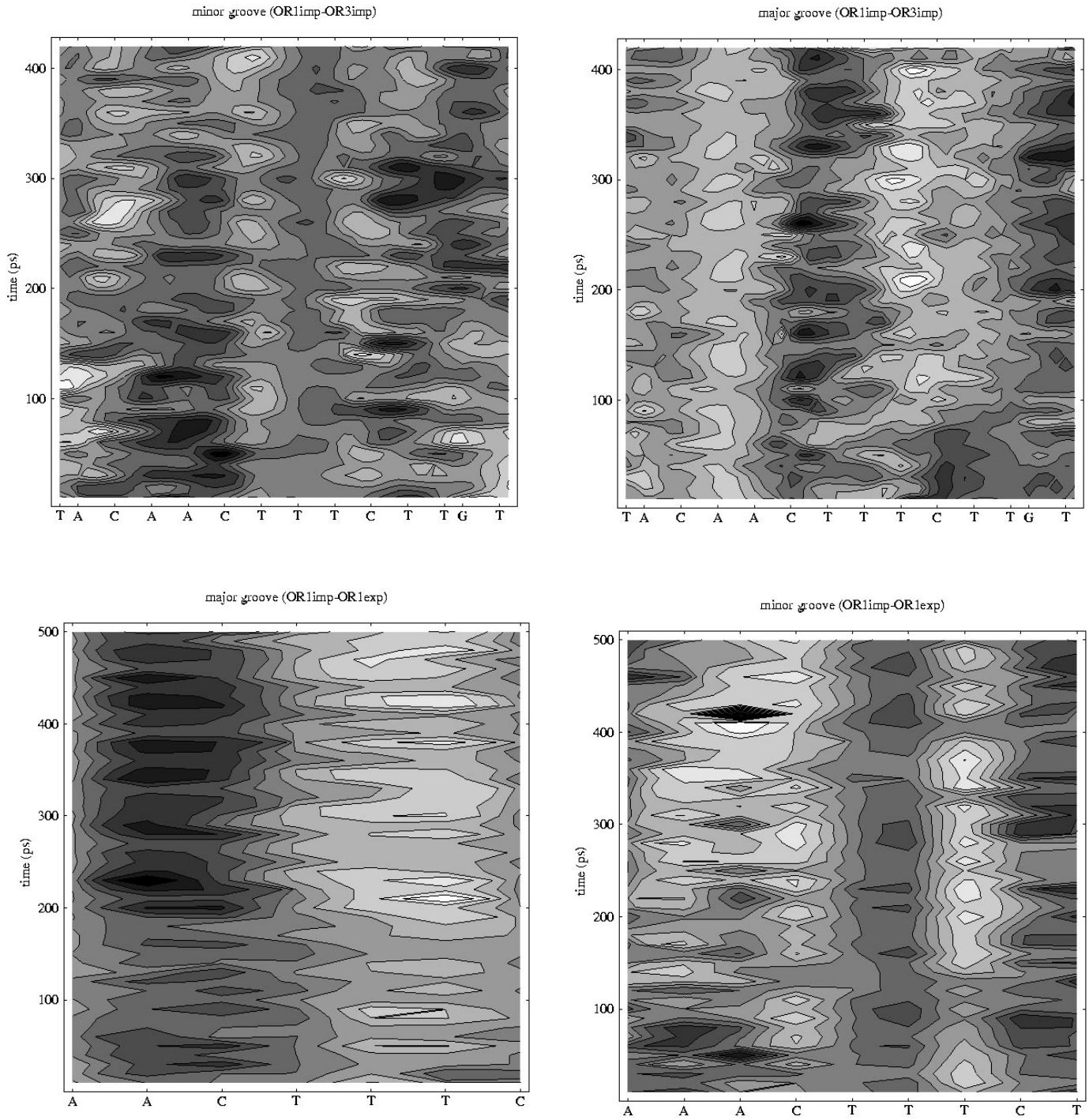
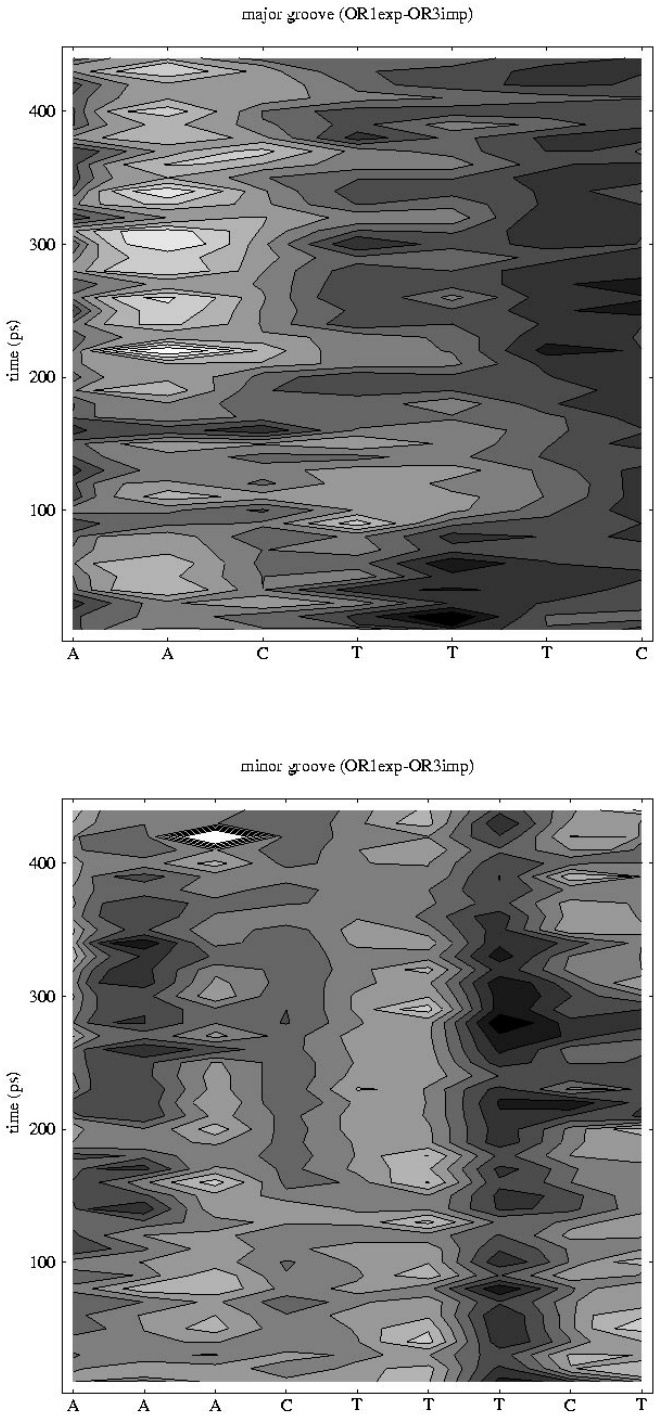


Figure 12. *(continues next page) Density plots to illustrate the difference in the width of the minor and major grooves for the OR1 and OR3 sequences from the simulations with the three models. A zero difference is denoted by a medium grey colour, a positive difference is darker and a negative difference lighter.*

was done only for the implicit simulations as these both have the full 20 base pairs of DNA. The curves showing these quantities reveal that there is no difference between the two global curvatures. The maximum variation for both ratios, of helix axis path length and end-to-end distance, between OR1 and OR3 in both implicit simulations is lower than 2 %. Because of their similarity these data are not shown.



The angle parameters, C1'-C2' and C2'-C3' are shown in Figure 13 for the implicit simulations. They are different for OR1 and OR3 in both the consensus left and the non-consensus right halves of the site. We obtain the same variations with the torsion parameters α, β, γ, ε and ζ – these differences exist between both halves. They are mainly located near the regions A3L, A4L and A3R, C4'R.

A comparison of the global axis curvature parameters (not shown) from both implicit simulations shows differences for

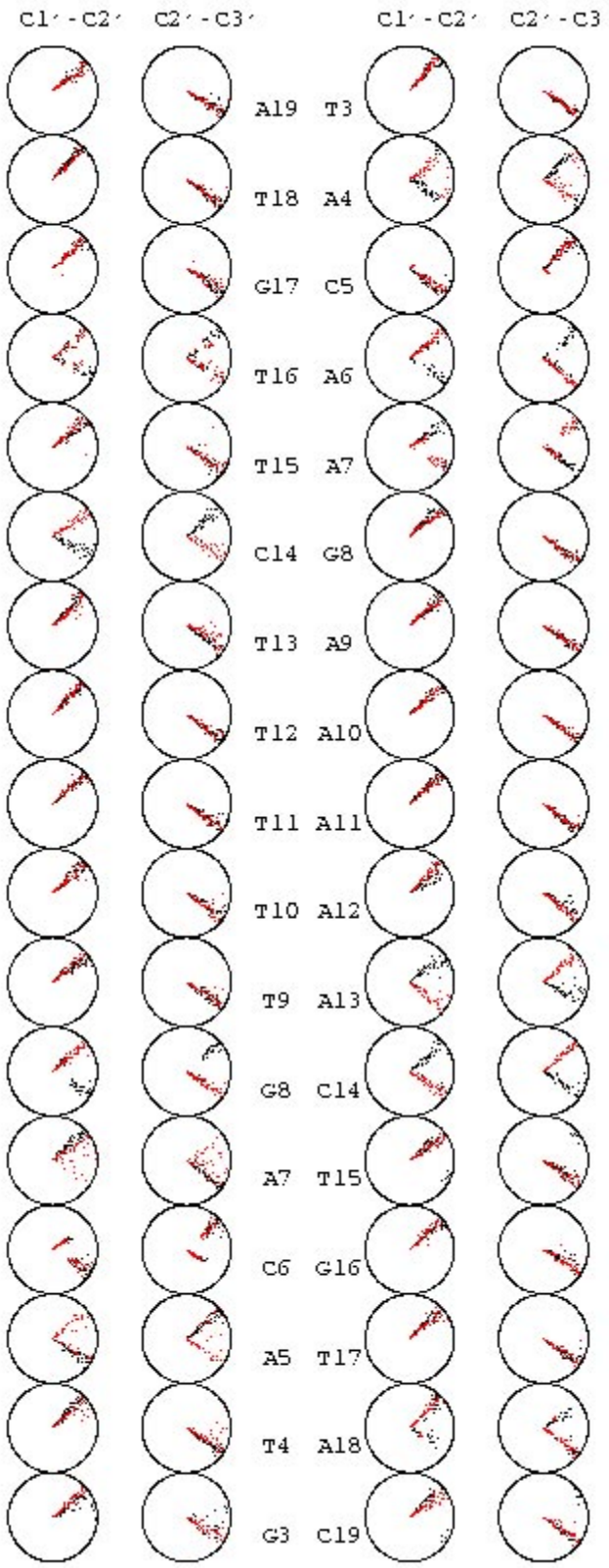


Figure 13. *Analysis of the DNA structure: dihedral angles of the sugar for both implicit simulations using the Curve program (OR1 in red and OR3 in green).*

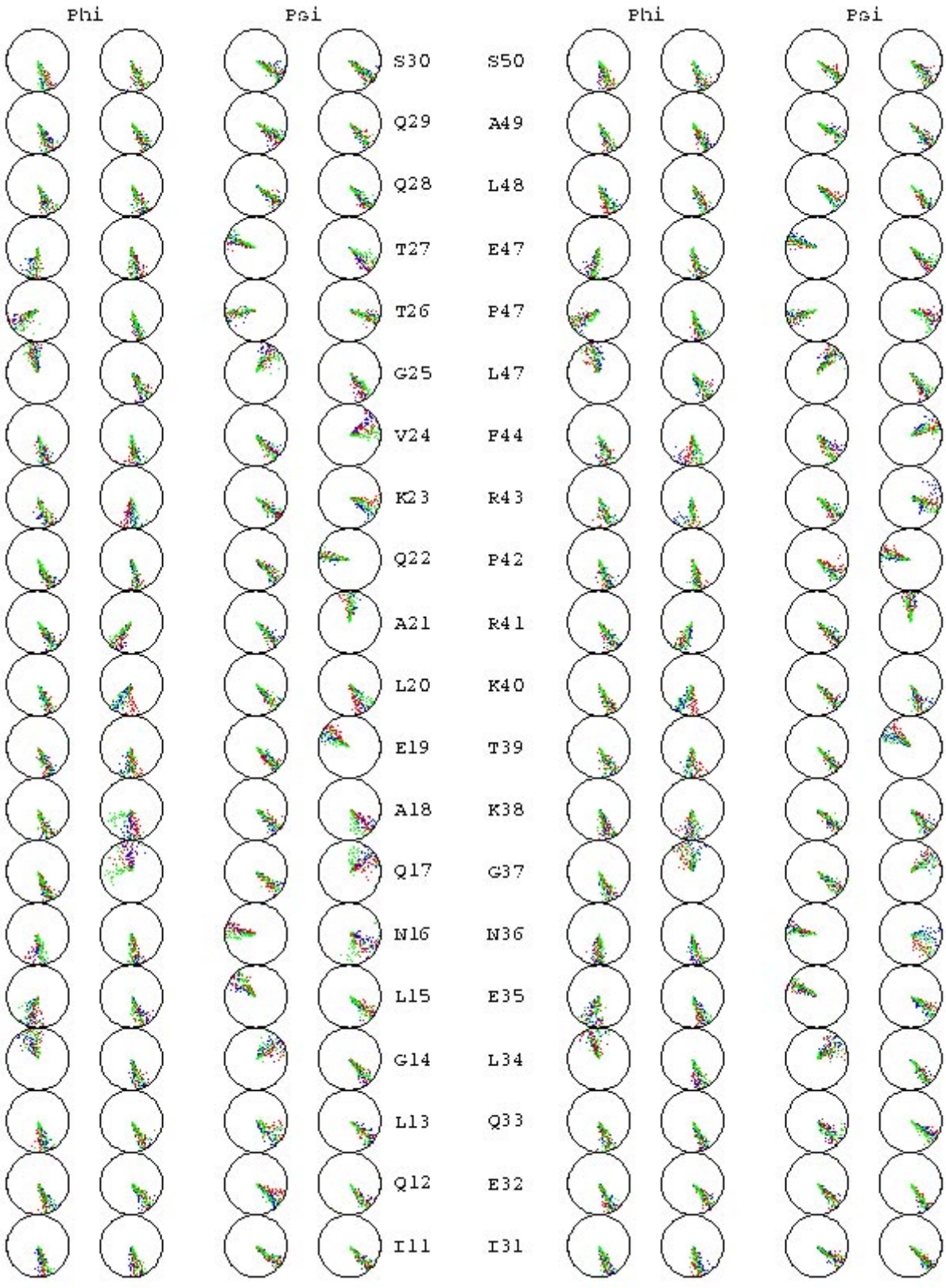


Figure 14. *The angles Φ and Ψ for the R1/69 homodimer for the three simulations using the PCurve program ((OR1 implicit solvent in black, OR3 implicit solvent in red and OR1 explicit solvent in green).*

the angle parameters to the bases C2R-A3R and A3R-G4R, for the axis X-displacement parameter of T7R-T7'R and for the axis tip parameter for A3R-G4R and T6'L-C5'L. The variation in the other parameters are too scattered to see trends between the values of both simulations. There are no significant changes for the inter-base pair parameters between the three simulations and the small variations that do appear are not localized at a specific spot.

The Protein Structure: An analysis of the three structures of R1-69 in the implicit and explicit simulations was made with the Pcurve program [39]. In the same way as for the DNA analysis, the protein structure is described using helicoïdal parameters. These parameters are displayed with dials and windows and superposed for the three simulations. The major changes between the simulations for the global peptide-axis parameters occur for the residues L13, G14 and E35, N36, G37, K38, R43, F44. There are two groups of residues (13–15 and 35–44) where the differences are important. These form turns linking helix 1 with helix 2 and helices 2 and 3. These variations exist in both halves of the site, but they are greater in the non-consensus right site. The values of the angles Φ and Ψ are shown in Figure 14. The Ω angle parameters are not shown since they change very little. Both halves of the site for the OR3 simulation have the same profile, which means that the effect of the mutation does not change the structure of the protein backbone. The greatest variations for these angles are located in both the turns described previously.

Conclusions

Three simulations have been performed of the 434 repressor protein complexed to two different operator sequences of DNA, OR1 and OR3. The major goal of this work was to improve the understanding of the interaction between R1/69 and its different operators by using simulations to obtain dynamical views of the complexes. All the simulations display reasonable behaviour despite the simplifications of the shortening of DNA for the explicit simulation and a use of a "distance-dependent dielectric" for the implicit simulations.

The first point to be underlined deals with the local structure of the DNA, which can be analyzed by characterizing the topology of its grooves. One of the major factors influencing the different specificity of the protein for the two operators is that the minor groove of the OR1 operator at the AAC triplet is shallower and the major groove is narrower than the corresponding grooves at the GTT triplet of the OR3 operator. This effect appears to be due to the closer contact that occurs between the base and the sugar-phosphate backbone in the OR1 sequence and affects the accessibility of the bases for interactions with the protein.

Huang et al. [17] showed that the helical motif Gln28-Gln29-Ser30-X31-X32-Gln33 is the unique sequence which gives a specific binding to 434 repressor for OR1. The strong preference for Gln amino acids is explained by their ability to hydrogen bond and to form van der Waals contacts with DNA. The important roles of Ser30L and Ser30R is surprising. In the crystallographic structure they have only water mediated contacts with DNA while in both OR1 simulations Ser30L has direct hydrogen bonds with DNA. The residue Thr27 is not important [17] and can be replaced by others without affecting the specificity. In the explicit simulation, both Thr27s have no direct hydrogen bonds with DNA. In the implicit simulation, the contacts that are observed can be explained by the reorganization of the side chains due to the lack of additional solvating water.

There are no significant changes in the base-base parameters during the simulation. The backbone parameters are very different in both consensus sites of OR1 and OR3. This shows that the DNA structure has a strong dependence on external parameters, such as its sequence or a complexed molecule. For the protein, the major differences between the structures occur in the turn regions – the remainder of the backbone stays essentially similar – and implies that the protein can be well approximated as a relatively rigid entity outside of the turns.

An implicit representation of the solvent, which uses a distance dependent dielectric with a reduction in the accessible charges, is an inexpensive method with which to model the solvation of macromolecules. It gives results which compare favourably with those obtained from a simulation with explicit waters. For example, the RMS coordinate displacement for both implicit and explicit simulations reaches a plateau around 1.8 Å and the heavy atom fluctuations of the R1-69/OR1 complexes are very similar in both types of simulations. Of course, an implicit model has some limitations. For example, the absence of water results in a partial reorganization of the DNA and the repressor, although this can be avoided by adding some explicit water molecules at the DNA-protein interface in addition to the ones defined crystallographically. In conclusion, for the simulations performed here, it appears that the implicit representation of solvent gives results of reasonable accuracy while enabling simulations to be performed with much less computational expense. The approximations introduced by the use of the implicit model also seem to be less drastic than the one resulting from the truncation of the three bases at the edge of the DNA which was necessary in the simulations with the explicit solvent model.

Acknowledgements The authors would like to thank Dr. Patricia Amara for her comments on the manuscript and the Institut de Biologie Structurale – Jean-Pierre Ebel, the Commissariat à l'Energie Atomique and the Centre National de la Recherche Scientifique for support of this work.

References

1. Wharton, R.P.; Brown, E.L.; Ptashne, M. *Cell* **1984**, *38*, 361.

2. Koudelka, G.B.; Harbury, P.; Harrison, S.C.; Ptashne, M. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 4633.
3. Koudelka, G.B.; Harrison S.C.; Ptashne, M. *Nature* **1987**, *326*, 886.
4. Bushman, F.D. *J. Mol. Biol.* **1993**, *230*, 28.
5. Harrison, S.C.; Aggarwal, A.K. *Annu. Rev. Biochem.* **1990**, *59*, 933.
6. Harrison, S.C. *Nature* **1991**, *353*, 715.
7. Bernstein, F.C.; Koetzle, T.F; Williams, G.J.B.; Mayer, E.F.; Brice, J.M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
8. Aggarwal, A.K.; Rodgers, D.W.; Drottar, M.; Ptashne, M.; Harrison, S.C. *Science* **1988**, *242*, 899.
9. Shimon, L.J.W.; Harrison, S.C. *J. Mol. Biol.* **1993**, *232*, 826.
10. Rodgers, D.W.; Harrison, S.C. *Structure* **1993**, *1*, 227.
11. Anderson, J.E.; Ptashne, M.; Harrison, S.C. *Nature* **1987**, *326*, 846.
12. Bushman, F.D.; Anderson, J.E.; Harrison, S.C.; Ptashne, M. *Nature* **1985**, *316*, 651.
13. Bushman, F.D.; Ptashne, M. *Proc. Natl. Acad. Sci. USA* **1983**, *83*, 9353.
14. Koudelka, G.B.; Lam, C-Y. *J. Biol. Chem.* **1993**, *268*, 23812.
15. Bell, A.C.; Koudelka, G.B. *J. Biol. Chem.* **1995**, *270*, 1205.
16. Bell, A.C.; Koudelka, G.B. *J. Mol. Biol.* **1994**, *234*, 542.
17. Huang, L.; Sera, T.; Schultz, P.G. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3969.
18. Fujimoto, B.S.; Schurr, M. *Nature* **1990**, *344*, 175.
19. Koudelka, G.B.; Carlson, P. *Nature* **1992**, *355*, 175.
20. Bashford, D.; Karplus, M. *Biochem.* **1990**, 29, 10219.
21. Bashford, D.; Karplus, M. J. Phys. Chem. 1991, *95*, 9556.
22. Antosiewicz, J.; McCammon, J.A.; Gilson, M.K. *J. Mol. Biol.* **1994**, *238*, 415.
23. Gilson, M.K. *Proteins: Structure, Function, and Genetics* **1993**, *15*, 266.
24. Davis, M.E.; Madura, J.D.; Luty, B.A.; McCammon, J.A. *Comp. Phys. Comm.* **1991**, *62*, 187.
25. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.
26. MacKerell Jr., A. *Developmental Residue Topology File for Proteins Using All Hydrogens.*
27. Seibel, G.L.; Singh, U.C.; Kollman, P.A. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6537.
28. Jorgensen, W.L.; Chandraseklar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. *J. Chem. Phys.* **1983**, *79*, 926.
29. Brooks III, C.L.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312
30. Mouawad, L.; Perahia, D. *J. Mol. Biol.* **1996**, *258*, 393.
31. Guenot, J.; Kollman, P.A. *Protein Sci.* **1992**, *1*, 1185.
32. Kabsch, W. *Acta Crystallog. Sect. A* **1978**, *34*, 827.
33. McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, 1987.
34. Arnold, G.E.; Ornstein, R.L. *Proteins: Structure, Function, and Genetics* **1994**, *18*, 19.
35. Lavery, R.; Sklenar, H. *J. Biomol. Str. Dynam.* **1988**, *6*, 63.
36. Pabo, C.O.; Aggarwal, A.K.; Jordan, S.R.; Beamer, L.J.; Obeysekare, U.R.; Harrison, S.C. *Science* **1990**, *247*, 1210.
37. van Gunsteren, W.F.; Berendsen, H.J.C.; Geurtsen, R.G.; Zwinderman, H.R.J. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 287.
38. Norberg, J.; Nilsson, L. *Chem. Phys. Lett.* **1994**, *224*, 219.
39. Sklenar, H.; Etchebest, C.; Lavery, R. *Proteins: Structure, Function, and Genetics* **1989**, *6*, 46.
40. Dickerson, R.E.; Bansal, M.; Calladine, C.R.; Diekmann, S.; Hunter, W.N.; Kennard, O.; Lavery, R.; Nelson, H.C.M.; Olson, W.K.; Saenger, W.; Shakked, Z.; Sklenar, H.; Soumpasis, D.M.; Tung, C-S.; von Kitzing, E.; Wang,A.H-J.; Zhurkin, V.B. *EMBO J.* **1989**, *8*, 1; *J. Biomol. Struct. Dynam.* **1989**, *6*, 627; *J. Mol. Biol.* **1989**, *205*, 787.
41. Falsafi, S.; Reich, N.O. *Biopolymers* **1993**, *33*, 459.
42. Swaminathan, S.; Ravishanker, G.; Beveridge, D.L.; Lavery, R.; Etchebest, C.; Sklenar, H. *Protein: Structure, Function, and Genetics* **1990**, *8*, 179.
43. Poncin, M.; Hartmann, B.; Lavery, R. *J. Mol. Biol.* **1992**, *226*, 775.
44. Stofer, E.; Lavery, R. *Biopolymers* **1994**, *34*, 337.
45. Ravishanker, G.; Swaminathan, S.; Beveridge, D.L.; Lavery, R.; Sklenar, H. *J. Biomol. Str. Dynam.* **1989**, *4*, 669.
46. Lavery, R.; Sklenar, H. *J. Biomol. Str. Dynam.* **1989**, *6*, 655.