PROTEINS: Structure, Function, and Genetics 6:104–127 (1989)

An NMR-Based Molecular Dynamics Simulation of the Interaction of the *lac* Repressor Headpiece and Its Operator in Aqueous Solution

J. de Vlieg, H.J.C. Berendsen, and W.F. van Gunsteren
Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

ABSTRACT **The results of a 125 psec molecular dynamics simulation of a *lac* headpiece–operator complex in aqueous solution are reported. The complex satisfies essentially all experimental distance information derived from two-dimensional nuclear magnetic resonance (2-D-NMR) studies. The interaction between *lac* repressor headpiece and its operator is based on many direct- and water-mediated hydrogen bonds and nonpolar contacts which allow the formation of a tight complex. No stable hydrogen bonds between side chains and bases are found, while specific contacts occur between both nonpolar groups and, to a lesser extent, through water-mediated hydrogen bonds. The simulated complex structure in water is intrinsically stable without application of nuclear Overhauser effect (NOE) distance restraints, while being compatible with most of the available biochemical, genetic, and chemically induced dynamic nuclear polarization (CIDNP) data.**

Key words: ***lac*** **repressor; *lac* operator, *lac* headpiece–operator complex, protein–DNA specificity, molecular dynamics, computer simulation**

INTRODUCTION

Repressor–Operator Interactions

Accurate recognition of specific base pair sequences within DNA by regulatory proteins is a prerequisite of accurate regulation of transcription. A general mode of interaction presumed to occur in many DNA–repressor proteins is the recognition of structural details in the major groove of B-DNA by an α-helix. This recognition helix is part of a helix-turn-helix motif which has been found in many repressor and analogous proteins with known three-dimensional structure.[1–8] Most of the repressor proteins have been crystallized in the absence of DNA, however, and the general features of protein–DNA interactions have been deduced on the basis of model building studies.[2–4,7–9] and genetic studies.[10]

Recently the electron density maps of cocrystals of the DNA-binding domain of phage 434 repressor with a synthetic 14 base pair operator at 3.2 to 4.5 Å resolution[11] and a very similar complex with a 20 base pair operator fragment at 2.5 to 3.0 Å resolution[12] have been described. Other complexes involving the DNA-binding domains of λ-repressor at 2.5 Å resolution[13] and the *trp* repressor complexed with its operator at 2.4 Å resolution[14] have been determined as well. In their general features, the observed repressor–operator complexes support the models for DNA–protein interaction that were earlier developed from model building and genetic studies. As predicted, the repressor proteins contain the helix-turn-helix motif with the second α-helix located in the major groove of the DNA. The recognized DNA segment retains its B-type conformation, although some deformation occurs relative to the ideal B-DNA structure.

In spite of the important insights that these X-ray results have provided, a chemical model which could explain the specificity of the interaction between repressor and operator on an atomic level is still lacking. In principle, the base pair sequence can be directly "read" by the repressor through complementary hydrogen bonding to the donor and acceptor groups of the base pairs exposed within the grooves (so called "direct readout"). The methyl group on the thymine provides an additional sequence recognition element, which can be identified from the major groove. However, in order to unambiguously read the identity of a base pair in the major groove a minimum of two hydrogen bonds is necessary, as was pointed out by Seeman and colleagues.[15]

In the high-resolution structure of the complex of 434 repressor[12] the repressor makes a number of hydrogen bonds to the phosphates of the backbone, primarily with peptide amino groups and neutral side chains. There are van der Waals contacts to deoxyribose sugars as well and water molecules are involved in bridging protein–DNA contacts. The three Gln residues of the recognition helix make the

Received March 27, 1989; revision accepted June 20, 1989.

Address reprint requests to J. de Vlieg, Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

primary contacts to the bases exposed in the major groove by making hydrogen bonds as well as nonpolar interactions. Only two base pairs are contacted by the minimum of two hydrogen bonds, of which one is probably a less unique bidentate hydrogen bond joining a guanine and a terminal NH_2 of glutamine. The central base pair, which is known to be important for the recognition, is contacted only through nonpolar interactions to the methyl group of thymine and the outermost base pairs in the major groove are either not linked or linked only by one hydrogen bond to the repressor. Other DNA-contacting residues of the recognition helix, such as 27 Thr and 30 Ser, are involved in van der Waals contacts only.

In spite of the essential differences between the 434 and λ complex, the general characteristics of the interaction appear to be the same. The λ complex is stabilized by an extensive network of hydrogen bonds between the protein and the sugar-phosphate backbone. Nonpolar contacts appear to be as important as hydrogen bonds to the specificity of binding and several side chains seem to "cooperate" in the recognition of one single base pair.

In the crystal structure of *trp* repressor at atomic resolution[14] no direct hydrogen bonds or nonpolar contacts to the bases are observed that can explain the repressor's specificity for the operator sequence in simple chemical terms. In some cases bases are hydrogen bonded to the repressor via water molecules, which explain part of the specificity, but in other cases they are exposed to solvent. Thus, the *trp* repressor–operator complex presents an example of so called "indirect readout": the operator sequence is recognized indirectly through induced changes in the geometry of the phosphate backbone, which, in turn, permit the formation of a tight complex with the protein. The primary role of the DNA sequence is to permit a conformation that substantially enhances attractive interactions with its repressor.[14]

The X-ray complex structures of 434, λ, and *trp* repressors make it clear that a simple "recognition code" in the sense of one-to-one interactions between amino acids and base pairs does not exist and that the specific recognition probably depends on a very complex combination of so called direct and indirect readout principles and solvent effects.

lac Repressor

The *lac* repressor from *E. coli* is a tetrameric protein with identical domains, which binds the operator very tightly, and with that blocks the expression of structural genes of the lactose operon.[16] The *lac* repressor belongs to the class of DNA recognizing proteins which have a helix-turn-helix structural motif as suggested on the basis of sequence homology.[2] The determinants for DNA binding are mainly localized in four small N-terminal domains or "headpieces" which can be cleaved by proteolytic enzymes from the tetrameric core.[17] These headpieces retain their original three-dimensional structure and their ability to recognize DNA specifically: a half *lac* operator on each site of the center of symmetry is recognized by one headpiece.[18]

lac Repressor Headpiece Conformation

In the case of the *lac* repressor no suitable crystals exist for X-ray analysis and the three-dimensional structure of headpiece comes from two-dimensional ^1H NMR studies.[19] On the basis of a set of 169 NOE distance constraints[20] a 3-D structure of headpiece has been derived by an iterative procedure involving model building and restrained molecular dynamics (MD) techniques in vacuo.[21–23] The final structure[23] obtained in this way corresponds closely to the best refined headpiece structures generated by distance geometry calculations.[24] Recently this structure has been studied using a restrained molecular dynamics simulation in aqueous solution.[25] The overall structure shows a helix-turn-helix motif which, together with a third helix, forms a hydrophobic core.

2-D NMR of *lac* Headpiece and Its Operator

The *lac* headpiece structure formed the basis for further 2-D NMR studies aimed at determining the way in which the *lac* repressor recognizes its operator. From the 2-D NOE spectra of the complex of *lac* headpiece with a 14 base pair operator fragment corresponding with the left half of the *lac* operator, 24 NOEs between protein and headpiece were detected.[26,27]

On the basis of these intermolecular NOEs a low-resolution structural model for the complex could be derived using docking studies based on minimizing NOE distance constraints.[28,29] This model showed that the recognition helix of the *lac* repressor headpiece makes contact in the major groove of the operator. Interestingly, the helix orientation is opposite to that observed in all other repressor studies.

The NMR experiments do not allow conclusions to be drawn about specific hydrogen-bond contacts or specific conformational changes upon complexation. The docking studies based on minimizing the protein–DNA NOE distance constraints therefore did not predict unambiguous hydrogen-bonding patterns. In addition, these calculations were done with rigid conformations for headpiece and operator. In order to predict intermolecular contacts we performed a detailed molecular dynamics simulation of the complex in aqueous solution.

Prediction of Interactions for Low-Resolution Repressor–Operator Structure

As became clear from recent complex crystal structures,[11–14] the specificity of the recognition between protein and DNA cannot be explained by

simple complementarity models by which an amino acid side chain recognizes a base pair primarily through hydrogen bonds. Static docking techniques based on considerations such as maximizing the number of hydrogen bonds, optimizing electrostatic interactions, or satisfying postulated hydrogen bonds are therefore insufficient and presumably only confirm a priori assumptions. Moreover, using these techniques, it is very hard to simultaneously take into account other important factors such as shape complementarity, hydrophobic and dipolar interactions, and the role of solvent. In particular, the absence of solvent will introduce serious errors. The binding constant of the complex is directly related to the free energy of interaction and the surrounding solvent will have a considerable influence on its magnitude through entropic contributions. Moreover, water-mediated hydrogen bonds appear to contribute part of the specificity between repressor and operator.[14]

Prediction of Molecular Dynamics (MD) Simulation in Aqueous Solution

A molecular dynamics computer simulation,[30–34] using a sufficiently accurate atomic force field which describes the total system of complex and solvent, will in principle objectively express all factors that influence the binding. During an MD simulation the presence of motional freedom corresponding to room temperature allows the system to scan that part of configurational space that is accessible to the protein/DNA molecules and surrounding solvent within the time period of simulation. The enormous size of the configurational space can never be searched completely, but fortunately it can be reduced to relevant parts by using an additional energy term in the force field representing the experimental NOEs[35] and choosing a plausible initial structure.

At the present state of the art MD simulations are not reliable enough to serve as an unequivocal source of structural predictions. Therefore experimental tests are still mandatory. We have performed an extensive simulation of the *lac* repressor headpiece–operator complex in aqueous solution, using NOE data in the initial procedure, but finally performing an unconstrained run to test the method per se. The first test concerns the question whether the NOE constraints remain satisfied. The second test concerns a comparison of the results with other experimental data as NMR, CIDNP, genetic, and biochemical observations. The third test concerns the future experimental verification of the effects of new mutants predicted by this simulation. The fourth and final test will concern a comparison with structural X-ray data which are not yet available. In the meanwhile an analysis of the results gives insight into possible mechanisms of specific repressor–operator interactions.

MODEL AND COMPUTATIONAL PROCEDURE

The Headpiece and Operator Structures Used in Our Models

Since neither headpiece nor operator undergoes large conformational changes upon binding, we took the 14 base pair operator fragment in standard B-DNA conformations as given by Arnott and Hukins[36] and the free headpiece structure as determined from 2-D-NMR as initial building blocks of the structures of the complex. The headpiece structure is based on a set of 169 NOEs[20] and was obtained by a combination of model building techniques and restrained molecular dynamics calculations. The structure has a low potential energy and satisfies all distance constraints within the accuracy of the experimental data set.[23]

Experimental Distance Constraints

During generation of the initial models of the complex and during part of the MD simulation of the complex in solvent the following distance constraint data sets were used:

Headpiece

The 169 NOEs as determined by Zuiderweg and co-workers[20] in free headpiece were converted to upper bound distance constraints of 0.35 nm corresponding to the distances between C_α protons of residue i and amide protons of both residue $i+1$ and $i+3$ in regular α-helices that give just visible NOE peaks in 2-D NOE spectra with a mixing time of 50 msec. Although not essential, but to be compatible with earlier studies,[21–25] 17 attractive constraints (r_0 = 0.25 nm) between the amide proton of residue i and the carbonyl oxygen of residue $i-4$, with $i = 10–13, 21–25, 38–45$, were added to the 169 attractive NOE constraints.

Operator

Although it is quite clear from NMR experiments that the structure of the *lac* operator in the complex is close to the B-DNA conformation it was impossible to translate the NOE intensities to proton–proton distances due to spin diffusion combined with the use of long mixing times.[26–28] The stabilities of the hydrogen bonds between base pairs appear to be sensitive to details in the interaction function as noticed before in earlier studies of B-DNA in aqueous solution.[37] Therefore we restrained the 36 hydrogen bonds between the base pairs. In order to avoid the distortion of the DNA structure during the generation of initial structures using MD in vacuo, we also restrained the positions of the backbone atoms P, O5′, C5′, C4′, C3′, and O3′ of the DNA molecule.

Headpiece–operator complex

A number of 24 NOEs between protons of the *lac* operator and protons of headpiece has been

identified.[26–28] A distinction is made between unambiguous and probable NOEs. Eleven NOEs between headpiece and operator occur at unique and unambiguous resonance positions of headpiece and operator, respectively. The other NOEs occur in crowded regions of the 2-D NOE spectra, where a potential overlap exists with resonances from other residues. However, many of these NOEs occur between residues for which unambiguous NOEs are found as well. A considerable amount of spin diffusion must be present due to long mixing times in the 2-D NOE spectra. Therefore it will be clear that the obtained NOEs are not very precise and it is not possible to solve the complete three-dimensional structure of the protein–DNA complex from these experimental data alone. However, by formulating weak distance constraints between atoms on the protein and DNA from the observed NOEs it is still possible to propose initial models for the *lac* headpiece–operator complex, which can be used as initial structures for a detailed MD simulation including solvent. A full discussion of these protein–DNA NOEs is given in Refs. 26–28.

For most of the 24 NOEs an upper distance bound of 0.6 nm was taken, while for some stronger NOEs a proton–proton distance of 0.4 nm was chosen. Corrections had to be applied to these upper distance constraints in those cases where protons could not be distinguished in the NMR spectra, either because of the lack of stereospecific assignments or because of rapid dynamic processes. In these cases each constraint refers to an average position (pseudo-atom) and the upper limit bound was corrected for the difference between the proton position and the actually used pseudo-position. For the hydrogen atoms that are attached to carbon atoms the concept of virtual atoms was used, because these hydrogens are incorporated into the carbon atoms in our protein and DNA model. A full description of the concept of the use of pseudo and virtual atoms is given in Refs. 35 and 38. The set of distance constraints derived from the protein–DNA NOEs is listed in Table I.

TABLE I. Intermolecular Distance Constraint List Derived From NOEs Between *lac* Repressor Headpiece and 14 Base Pair *lac* Operator Fragment*

Protein	DNA	Upper limit distance
	Unambiguous†	
7 Tyr Hε1,2	10 Gua H8	0.61
7 Tyr Hε1,2	10 Gua H1′	0.81
7 Tyr Hε1,2	10 Gua H3′	0.61
7 Tyr Hε1,2	9 Cyt H5	0.61
7 Tyr Hε1,2	9 Cyt H6	0.61
6 Leu H_3δ1,2	9 Cyt H5	0.82
17 Tyr Hε1,2	9 Cyt H5	0.81
17 Tyr Hδ1,2	9 Cyt H5	0.81
17 Tyr Hε1,2	9 Cyt H6	0.81
17 Tyr Hδ1,2	9 Cyt H6	0.81
17 Tyr Hε1,2	8 Thy H6	0.81
17 Tyr Hδ1,2	8 Thy H6	0.81
29 His Hε1	2 Ade H8	0.60
	Probable†	
5 Thr H_3γ2	10 Gua H8	0.70
5 Thr H_3γ2	10 Gua H3′	0.50
6 Leu H_3δ1,2	9 Cyt H5	0.82
6 Leu H_3δ1,2	9 Cyt H6	0.82
6 Leu H_3δ1,2	9 Cyt H3′	0.82
6 Leu H_3δ1,2	8 Thy H6	0.82
17 Tyr Hε1,2	8 Thy H_37	0.71
17 Tyr Hδ1,2	8 Thy H_37	0.71
21 Ser Hα	8 Thy H_37	0.70
21 Ser H_2β	2 Ade H_37	0.79
29 His Hε1	2 Ade H3′	0.40
29 His Hε1	2 Ade H4′	0.60
29 His Hε1	2 Ade H_25'	0.69

*The NOEs were detected by Boelens and colleagues and taken from Refs. 26–28.

†The unambiguous distance constraints are derived from unambiguous NOEs which were assigned at unique resonance frequencies, while the probable distance constraints are derived from NOEs which could overlap with resonances of other protons.

Docking of Headpiece on the Operator

To investigate whether the NMR distance constraints would produce similar complex structures from different starting structures, four complex models were generated using different techniques. The first model (model I) was made on a graphical display system and refined using restrained molecular dynamics. Model I was subsequently used by Rullmann et al.[29] as an initial starting structure for a more objective docking approach by means of the ellipsoid algorithm developed by Billeter et al.[39,40] For this purpose the MD-refined complex structure was disrupted by an arbitrary displacement and reorientation of the protein. Hereafter headpiece and operator were docked again by means of the ellipsoid algorithm, while keeping the internal geometries of both molecules unchanged. These calculations were repeated three times, after which the generated structures were refined by restrained MD in vacuo (Models II–IV). The MD-refined complex structure with both a low distance restraint energy and a low headpiece–operator interface energy was chosen as the initial structure for the molecular dynamics simulation in solvent.

Molecular Model and Force Field

The protein *lac* repressor headpiece consists of 399 heavy atoms and the DNA fragment of 568 heavy atoms. The DNA fragment has the following sequence and numbering scheme.

```
5′- G- G- A- A- T- T- G- T- G- A- G- C- G- G
3′- C- C- T- T- A- A- C- A- C- T- C- G- C- C
   −2 −1 1 2 3 4 5 6 7 8 9 10 11 12
```

No phosphate groups are placed at the ends of the DNA strands, which have O5′ and O3′ as terminal atoms. Hydrogen atoms attached to carbon atoms

are incorporated into the latter, forming united atoms, whereas the other hydrogen atoms, which may form hydrogen bonds, are explicitly treated (for headpiece 98 and for DNA 62 polar hydrogens). The potential-energy function describing the interaction between the headpiece and operator, water and ions is described in Ref. 41 and is composed of terms representing bond angle bending, harmonic (out-of-plane, out-of-tetrahedral configuration) dihedral bending, sinusoidal dihedral torsion, van der Waals, and electrostatic coulomb interactions. During distance-restrained MD an additional energy term is added to the potential energy function that represents the set of distance constraints, and which forces the molecule to satisfy the observed atom–atom distances from 2-D NMR experiments.

$$V_{dc}(r_{nm},r^0_l,r^1_l)$$
$$= 0 \qquad \text{if } (0 \le r_{nm} \le r^0_l)$$
$$= \tfrac{1}{2} K_{dc}(r_{nm} - r^0_l)^2 \qquad \text{if } (r^0_l \le r_{nm} \le r^1_l)$$
$$= \tfrac{1}{2}K_{dc}(2r_{nm} - r^0_l - r^1_l)(r^1_l - r^0_l) \text{ if } (r_{nm} \ge r^1_l)$$

where $r^0{}_l$ is the value for the basic upper limit constraint l plus the correction term for pseudo-atoms,[35] and r_{nm} is the actual distance between atoms n and m involved in the specific constraint indicated by l. The potential energy term V_{dc} (r,r^0,r^1) is taken linear beyond $r^1{}_l$ in order to avoid the occurrence of very large restraining forces when r becomes large. In this work we used a value of $r^1{}_l = r^0{}_l + 0.1$ nm. The force constant K_{dc} can be gradually changed during the refinement. During the MD refinements of the complex structures in vacuo K_{dc} values up to 4000 kJ mol^{-1} nm^{-2} were used. For the distance-restrained part of the simulation of the complex in aqueous solution (0–70 psec) different K_{dc} values were used for the inter- and intramolecular NOE and hydrogen bond distance restraints (see above): for 186 headpiece distance restraints K_{dc} = 4000 kJ mol^{-1} nm^{-2}, for 36 operator restraints K_{dc} = 500 kJ mol^{-1} nm^{-2}, and for 26 headpiece–operator restraints K_{dc} = 2500 kJ mol^{-1} nm^{-2}. During the second part of the MD run (70–125 psec) the distance restraints were gradually switched off at a rate of 750 kJ mol^{-1} nm^{-2}. psec^{-1} except for the 36 "hydrogen bond" distance restraints used in the operator which were kept at K_{dc} = 500 kJ mol^{-1} nm^{-2}.

Atomic partial charges[42] were used without modification and a dielectric constant of 1 was applied. However, during the simulations in vacuo we have neutralized the charged atom groups in order to compensate for the missing shielding effect of the solvent.[35]

In order to evaluate electrostatic interactions with sufficient accuracy, we used a long cutoff radius of $R_2{}^c$ = 1.2 nm. A twin-range method was used: the total nonbonded interactions within a short cut-off range R_1 = 0.8 nm are evaluated every MD time-step of Δt = 0.002 psec, while the longer ranged (Coulomb) interactions within $R_2{}^c$ = 1.2 nm were evaluated less frequently (only every 10 time steps) and were kept fixed between updates. We note that the cutoff radii $R_1{}^c$ and $R_2{}^c$ are both applied to the centers of geometry of neutral atom groups (charge groups) in the protein or DNA and to the oxygen atoms of the water molecules; this avoids the breaking of charge neutrality of a group that occurs in the case that an atom–atom cutoff is applied. The covalent bond lengths were kept rigid during the simulation by application of the SHAKE method[43,44] and the water molecules were modeled by a simple rigid three-point charge (SPC) model.[45] The C_6 and C_{12} van der Waals parameters of Na^+ and Cl^- were taken from Refs. 37 and 46.

Simulation in Aqueous Solution

The best MD refined (in vacuo) complex structure obtained as described above was taken as the initial conformation for the simulation in aqueous solution. This structure was placed in the center of a box of which the dimensions (a = 4.002 nm, b = 4.999 nm, c = 6.758 nm) were chosen such that the minimum distance of any protein or DNA atom to the wall of the box was 0.67 nm. Water molecules were inserted in the box by immersing it in an equilibrium configuration of bulk SPC water[42] and subsequently removing all water molecules that are outside the box or of which the oxygen lies within R^i = 0.23 nm of a nonhydrogen protein or DNA atom. In this way 3950 water molecules were inserted into the box, yielding a density of ρ = 1.05 g cm^{-3}. The high nonbonded interaction energy between protein or DNA and H_2O, and between H_2O–H_2O mutually, which is due to the small value of R^i or to the periodic boundary condition, was relaxed by performing steepest-descents energy minimization (EM) while the protein and DNA atoms were positionally restrained.

The total charge of the complex adds up to −24 (headpiece +2, operator −26). In order to obtain a neutral system but also to bring the salt concentration to approximately the same value (0.2–0.3 M) as used during the NMR measurements of the complex,[26] we added 34 Na^+ and 10 Cl^- ions to the system. The Na^+ and Cl^- ions were inserted as follows: For the configuration of headpiece–operator complex and 3950 water molecules, the electrical potential was calculated at all water oxygen positions and the water molecule at the lowest or highest potential was replaced by an Na^+ or Cl^- ion, respectively. This procedure was repeated 44 times and then followed by EM of the whole system while keeping the protein and DNA atoms positionally restrained in order to relax the ion–solute and ion–water interactions. We note that the simulation

TABLE II. Results of Restrained MD Refinement In Vacuo of the Initial Complex Structures of Headpiece and Operator*

	Initial			After restrained MD		
Structure	Σd_{viol}	d_{viol}^{max}	HP–OPR	Σd_{viol}	d_{viol}^{max}	HP–OPR
I	1.68	0.29	$>10^{12}$	0.32	0.10	−712
II	0.71	0.25	−227	0.06	0.03	−672
III	0.56	0.26	−287	0.06	0.04	−643
IV	1.03	0.32	−246	0.15	0.07	−662

*The sum of the violations of the 26 intermolecular NOE distance constraints between headpiece and operator is denoted by the symbol Σd_{viol} and given in nm. The largest single violation is denoted by d_{viol}^{max}. The sum of the electrostatic and the van der Waals interface energy between headpiece and operator is denoted by the symbol HP–OPR and given in kJ mol^{-1}. Due to differences in the force field used during MD simulations in vacuo and in solvent these values cannot be compared to the energy values found in Table IV.

time is too short to allow these ions to realize all possible configurations and thus to represent the average screening resulting from an ionic strength. The chosen configuration represents a statistical sample only. The influence of these ions on the stability of the complex will be limited since no ions are placed in the contact region between DNA and headpiece.

Thus, the system of the headpiece–operator complex in aqueous solution contains 12,889 atoms.

Other parameters that could influence the results of the simulation in aqueous solution were chosen equal to the values used in the simulation of the free headpiece.[25] In this way we ensure that structural differences found in the headpiece are due to binding to the DNA molecule.

Initial velocities for the atoms were taken from a maxwellian distribution at 300 K. The system was weakly coupled to a thermal bath of T_0 = 300 K, and to a pressure bath of P_0 = 0.06102 kJ mol^{-1} nm^{-3} (= 1 atm), when integrating the equations of motion with time step Δt = 0.002 psec. This was done by applying the algorithm given by Berendsen et al.[47] with temperature relaxation time τ_T = 0.1 psec (0.01 psec during the first 5 psec of the run) and pressure relaxation time τ_p = 0.5 psec (0.05 psec during the first 5 psec). These values make the coupling to the temperature and pressure baths weak enough to avoid any significant effect on the atomic properties of the system.[47] The value for the compressibility of the system was chosen as β = 7.48 10^{-4} kJ mol^{-1} nm^{-3}. The pressure scaling of the periodic box was done independently along the three edges of the box.

The MD run covered a time span of 125 psec consisting of a distance-restrained part (0–70 psec) and an unrestrained part (70–125 psec), which took a total of about 1750 central processing unit (CPU) hours on a CONVEX C1-XP computer. Another 250 CPU hours was needed for analyzing the 967 Mbyte of data. The time spans from 45 to 70 and 100 to 125 psec, at a time resolution of 0.04 psec, were used for calculating time averages of various quantities for the restrained and unrestrained MD simulation, respectively.

RESULTS

Comparison of the Initial Complex Models

The total violation distance and the largest violation for the headpiece–operator distance constraints and the interaction energy between both molecules of the four generated models (I–IV) before and after restrained molecular dynamics (MD) refinement are shown in Table II. The total violation distance and the largest violation of the model-built (I) and the complex structures generated by the ellipsoid algorithm (II–IV) are brought down considerably during the restrained MD procedure. All the MD refined structures, although to a lesser extent in the case of the model-built structure, fit the intermolecular distance constraints within the accuracy of the NMR experiments.

The interface energies of the three models docked by using the ellipsoid algorithm (III–IV) are brought down by the MD refinement to approximately the same value of ~700 kJ mol^{-1}. Also the high value of the interface energy of the model-built structure, which is due to this way of structure generation, is lowered to approximately that value. The small variations in the interface energies suggest a similar position of the protein with respect to the DNA.

In order to test whether the use of the NMR distance constraints would lead to similar complex structures we compared the positions of the recognition helix (residues 17–25) with respect to the operator before and after MD refinement. From Table III it becomes clear that the position of the recognition helix relative to the operator converges during the MD refinement. The root mean square (rms) difference between the position of this helix averaged over all four structures is 0.15 nm before and 0.09 nm after the restrained MD refinement. Model II was chosen as an initial structure for a detailed MD simulation including solvent.

Molecular Dynamics Simulation of the Complex in Aqueous Solution

The simulation of the complex covered a total of 125 psec. However, in order to test the ability of the simulated complex in solvent we switched off the

TABLE III. Root Mean Square Differences Between the Relative Position of the Recognition Helix of the Initial Headpiece–Operator Complex Structures Before and After MD Refinement In Vacuo*

	I	II	III	IV
I		0.22	0.27	0.21
II	0.14		0.06	0.03
III	0.16	0.05		0.10
IV	0.09	0.06	0.07	

*Root mean square differences in nm between the positions of the C_α atoms of the recognition helix (residues 17–25) and the operator between the various complex structures. Upper right-hand triangle: before MD refinement. Lower left triangle: after MD refinement. To measure only differences in relative positions of the helix and the operator and not mutual differences of DNA and helix structures an average helix and an average DNA structure were fitted to the complex structures under comparison. Hereafter the atoms of the average DNA structure were used for fitting the complex structures and the positions of the C_α atoms in the average helices were compared.

NOE distance constraints in the second part of the simulation (70–125 psec). The total potential energy of the system and the rms deviations of the actual atomic positions for the C_α and C1′ atoms of headpiece and DNA, respectively, are depicted in Figure 1. The high initial potential energy, which is due to the randomly placed solvent molecules, drops rapidly during the first picoseconds and stabilizes after 20 psec. However, the rms deviations take 40 psec to stabilize. It is striking that during the unrestrained part of the MD run the deviations from the initial structures again become smaller. We decided to use the time spans from 45 to 70 and 100 to 125 psec for calculating time averages of various quantities for the restrained and unrestrained MD simulation respectively.

From Table IV, which shows the various energy terms of the system and the volume of the computational box, it can be observed that the interaction energies of headpiece and operator with the surrounding solvent are in magnitude much more important than the internal nonbonded energies and the binding energies between the two molecules. Clearly it is essential to consider the solvent in order to perform a detailed simulation of the binding of the two molecules. If we compare the time-averaged energy values over the unrestrained period (100–125 psec) with the values found for the restrained period (45–70 psec), we observe a decrease of the interaction energy of DNA and headpiece, and a decrease of the interaction energy of headpiece with solvent. This is concomitant with an increase of the internal nonbonded energies of both headpiece and operator and the interaction energy of operator with solvent.

The value of the distance-restraint energy term oscillates about a low average value of 60 kJ mol^{-1} during the restrained part, which is very small in proportion to the other potential energy values and indicates the satisfaction of essentially all distance restraints during this part of the simulation.

The average volume of the box differs by less than 1% from the initial value, which means that the minimum solute–water distance of $R^i = 0.23$ nm for eliminating water molecules was appropriately chosen.

Reduction of the Mobility of the Headpiece Side Chains in the Complex

In Figure 2 the rms fluctuation of the backbone atoms and side chain atoms of uncomplexed restrained headpiece (A) and restrained and unrestrained complexed headpiece (B, C) are compared. The rms fluctuations of the free headpiece have been obtained from a distance-restrained MD simulation of the headpiece in aqueous solution under conditions similar to those used in this work.[25] From Figure 2 it is observed that the rms fluctuations of most of the backbone and side chain atoms in the complexed headpiece (B) are reduced compared to free headpiece (A), which means that the headpiece surface becomes more rigid in the complexed state.

The reduction of the mobility is also observed when comparing uncomplexed headpiece (A) with unrestrained complexed headpiece (C). This means that the reduction does not arise from the use of the intermolecular NOEs during the restrained MD run, but is caused by the complexation with DNA. The reduction is most clear for the side chain atoms of residues 6, 7, and 11 of helix I, residues 16, 18, 22, and 25 of the recognition helix II, and residue 29 of the loop region. The loss of mobility of the surface residues 7 Tyr and 29 His is in good agreement with the fact that the NOE intensity of the aromatic protons of these residues increased within the bound headpiece compared to the free headpiece.[28] A similar reduction of the flexibility of the backbone and side chain atoms of the complexed *trp* repressor compared to uncomplexed repressor was observed by Otwinowski et al.[14] in their recent X-ray crystallography study of the *trp* repressor operator complex at atomic resolution.

RMS Difference Between the Complexed and Uncomplexed *lac* Repressor Headpiece Structure

Figure 3a shows the rms differences per residue between the average atomic positions of backbone and side chain atoms of free and complexed headpiece structure (both restrained). From Figure 3a it becomes clear that a number of side chains of the *lac* repressor headpiece undergo a large conformational change as a result of the complex formation. Interestingly, in most cases these large conformational changes are not accompanied by large changes in the corresponding backbone atoms. The residues belonging to the hydrophobic core (residues 4, 6, 10,

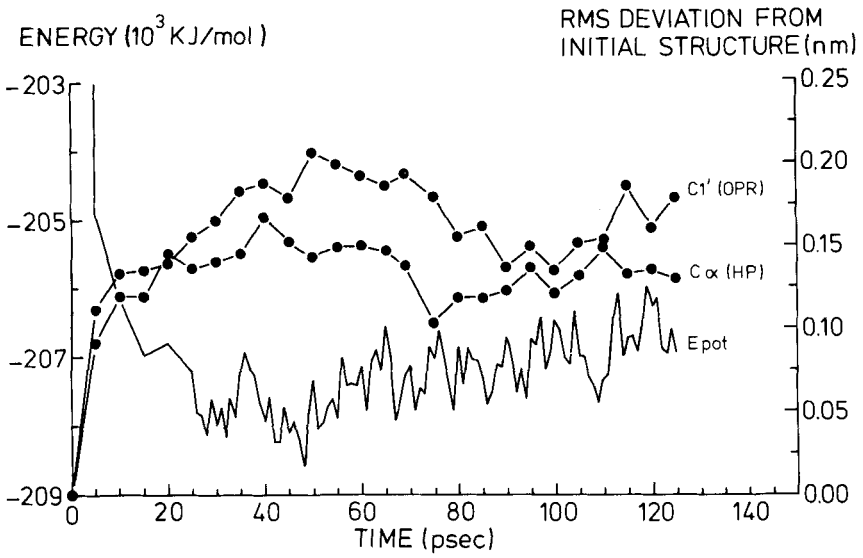


Fig. 1. Total potential energy of the complex in solvent and root mean square deviation of the simulated *lac* repressor headpiece and operator structures as a function of time and averaged over the protein C_α and operator C1′ atoms, respectively.

20, 23, 24, 38, 41, 42, 47), which strongly determine the overall conformation and relative orientations of the three helices of the *lac* repressor headpiece, hardly rearrange upon binding. Comparison of complexed and free DNA binding domain of phage 434 repressor shows a similar result: significant local adjustments of side chain configurations but no large-scale conformational changes.[12] We note that in spite of the important reorientations of some side chains the complexed and uncomplexed *lac* repressor headpiece structures both satisfy the distance constraints within the accuracy of the experimental data set. Significant reorientations (omitting the N- and C-terminal residues) are observed for the side chains of residues 7 Tyr, 11 Glu, 12 Tyr, 18 Gln, 25 Asn, 29 His, 33 Lys, 36 Glu, 44 Glu, and 46 Asn. The side chains of residues 11, 12, 33, 36, and 44 are not at the DNA side of the molecule and therefore not directly involved in protein–DNA contacts. This means that their reorientations must be induced as a side effect of other structural changes. Figure 4 displays the significant changes between the superimposed average structures of free headpiece and headpiece as seen in the interface of the repressor–operator complex. Interestingly, the phenyl rings of 7 Tyr and 17 Tyr which are both involved in protein–DNA interaction are observed to be in a stacked configuration in the complexed headpiece while they are not stacked in the free headpiece.

Photo-CIDNP experiments show that residues 7 Tyr, 17 Tyr, 29 His, and 47 Tyr are covered by DNA, but that residue 12 Tyr remains accessible upon complex formation,[48] and this is confirmed in our simulation. The very small changes of backbone as well as side chain atom positions for residues 6 Leu, 9 Val, and 47 Tyr in the free and complexed headpiece (Fig. 3a) are in agreement with the insensitivity of the chemical shifts for these residues to complex formation.[26]

An interesting conformational change is observed for the side chain of residue 18 Gln, which forms an extremely stable intraresidue hydrogen bond between its main chain HN and Oε1 in the complexed headpiece (exists for 86% and 96% during the restrained and unrestrained part of the simulation, respectively). Such an interaction, occasionally observed with glutamine at amino termini of α-helices,[49] was never observed during the simulation of free headpiece and is therefore a result of DNA binding. The intraresidue hydrogen bond fixes the orientation of this second N-terminal residue of the recognition helix and its Cβ, Cγ, and Cδ atoms are placed in front of the 5-methyl group (C7) of 6 Thy (see section "Nonpolar Contacts and Specificity") while its Nε donor group makes a direct hydrogen bond with 5 Gua O5′ and a water-mediated hydrogen bond with 6 Thy O2P (see sections "Direct Hydrogen Bonds" and "Hydrogen Bonds Via Water Molecules"). An astonishing resemblance is observed with 28 Gln in the specific repressor–operator complex of bacteriophage 434, the structure of which has been refined to 2.5 to 3.0 Å resolution.[11,12] Also this second N-terminal residue of the recognition helix of phage 434 repressor makes an intraresidue Oε-HN hydrogen bond and its Cβ and Cγ atoms are in van der Waals contact with a thymine 5-methyl group of base pair of 3 of the 434 operator. However, its Nε donor group can hydrogen bond to guanine instead of to the sugar-phosphate backbone.

TABLE IV. Potential Energy of *lac* Repressor Headpiece and Its Operator at Constant Temperature and Pressure*

	Potential energy (kJ/mol)									Volume (nm^3)
			Nonbonded							
	Total	Bonded	Hp	Opr	Hp–Opr	Hp–Slv	Opr–Slv	Slv	Dstres	
Initial complex		4382	−3132	−3883	−568				47	
Initial + 3950 H_2O	$>10^6$	4382	−3132	−3883	−568	2454	2904	$>10^6$	47	
EM (95 steps)	−1580	3125	−3399	−4224	−3357	−1259	−3646	−145450	51	
Initial + 3906 H_2O + 34 Na^+ + 10 Cl^-	−174272	3107	−3395	−4218	−3355	−1886	−24461	−140308	52	
EM (88 steps)	−177190	3028	−3028	−4212	−3354	−2353	−24327	−147151	63	135.2
MD (restrained)										
10 psec	−206231	4265	−2926	−4322	−3372	−7641	−28400	−163905	70	134.8
20 psec	−206804	4212	−3189	−4822	−2911	−8116	−28521	−163555	98	134.3
30 psec	−208135	4280	−3266	−5144	−3545	−7353	−28758	−164406	57	134.1
40 psec	−207583	4144	−3015	−5615	−2694	−8175	−28300	−163986	58	134.1
50 psec	−207674	4224	−2864	−5007	−3490	−8431	−28153	−164009	56	134.2
60 psec	−206783	4098	−3032	−5002	−3162	−7895	−27292	−164559	61	134.1
70 psec	−207402	4170	−3275	−5043	−2892	−8168	−27871	−164376	44	134.5
⟨45–70 psec⟩	−207560	4170	−2980	−5165	−3041	−8431	−27771	−164394	54	133.9
	(676)	(74)	(140)	(247)	(188)	(348)	(368)	(523)	(10)	(0.4)
MD (unrestrained)										
80 psec	−206335	4184	−2757	−4856	−3121	−8057	−27576	−164157	4.6	134.0
90 psec	−206881	4265	−2798	−4705	−3194	−8852	−28148	−163455	6.4	133.6
100 psec	−205867	4199	−2556	−4417	−3584	−9181	−27524	−162808	4.1	134.2
110 psec	−207626	4170	−2723	−4890	−3757	−8861	−26612	−164949	4.4	133.9
120 psec	−205262	4062	−2595	−4511	−3420	−8652	−26098	−164051	2.9	134.1
125 psec	−206585	4079	−2737	−5138	−3537	−8307	−26131	−164814	5.1	134.0
⟨100–125 psec⟩	−206825	4172	−2751	−4817	−3583	−8636	−26886	−164328	3.4	134.0
	(714)	(79)	(154)	(227)	(194)	(358)	(399)	(555)	(1.5)	(0.3)

*Total denotes the sum over all potential energy terms of the system and bonded stands for the sum of the bond-angle bending, harmonic dihedral bending, and sinusoidal dihedral torsion terms of both the headpiece and operator molecule. The energy term for bond length has no meaning because the bond lengths are kept rigid by applying the SHAKE method.[43,44] The nonbonded energy is the sum of the electrostatic and van der Waals interactions. The nonbonded potential energy contributions of *lac* repressor headpiece, operator, and the interface energy between them are denoted by Hp, Opr, and Hp–Opr. The interaction energy of headpiece and operator with the surrounding solvent and solvent mutually are denoted by Hp–Slv, Opr–Slv, and Slv, respectively. The solvent consists of H_2O molecules and Na^+ and Cl^- ions. Dstres denotes the distance restraint energy. During the restrained MD run (0–70 psec) 248 intra- and intermolecular NOE and hydrogen bond distance restraints are used (see text). During the second part of the MD (70–125 psec) run these distance restraint energy terms are switched off with the exception of the 36 base pair hydrogen bond distance restraints. The volume of the computational box is denoted by volume. The time-average is denoted by ⟨ ⟩ using a time resolution of 0.04 psec. The standard deviations are given in parentheses.

We note again that the orientation of the recognition helix of the *lac* repressor headpiece in the major groove of DNA with respect to the twofold axis of symmetry at GC 11 is opposite[26–28] to what is found for other repressors. This means that in the *lac* repressor operator complex the N-terminus of helix I is near the twofold axis, while the N-terminus of the first helix for other repressors is distant from this axis. The same reversed orientation has been observed for the simultaneous binding of two headpieces to the full operator.[50]

In Figure 3b the rms differences for the average atomic positions of backbone and side chain atoms per residue are shown for restrained and unrestrained headpiece both complexed to the operator. Although a high force constant of K_{dc} = 4000 kJ mol^{-1} nm^{-2} was used during the restrained part of the simulation of the complex (0–70 psec), the headpiece conformation is not strongly changed by switching off these constraints (70–125 psec). For many residues the changes are even smaller for the side chain atoms than for the backbone atoms. This observation is also true for residues not involved in DNA binding.

Although the rms difference between the restrained and unrestrained headpiece in the complex is small, distance constraint violations are observed for a few residues during the unrestrained part of the simulation (Table V). However, only 9 out of the 212 switched-off distance constraints are violated by more than 0.1 nm with the largest occurring violation at 0.24 nm. The pattern of the repressor–operator interactions does not change in an essential way upon release of the NOE restraints and this release significantly affects only one nonpolar interaction, between 25 Asn and the operator, which is lost during the unrestrained part (see section "Nonpolar contacts and specificity"). Interestingly, the stabilization of a water-mediated hydrogen bond between 29 His and 1 Ade during the unrestrained run (it exists for 24% and 81% during the restrained and unrestrained part of the MD simulation, respec-

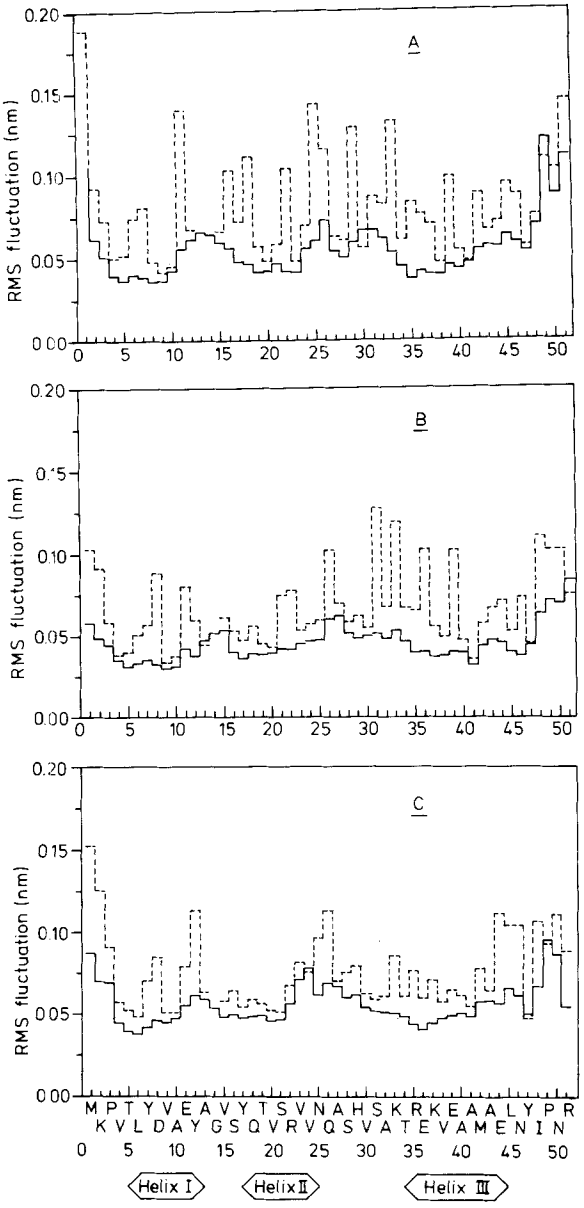


Fig. 2. Root mean square fluctuations of the N, C_α, C, and O backbone atoms (solid lines) and side chain atoms (dotted lines) averaged over each residue of the *lac* repressor headpiece in **A:** NOE distance restrained free headpiece in solution; **B:** restrained headpiece complexed to operator; and **C:** unrestrained headpiece complexed to operator.

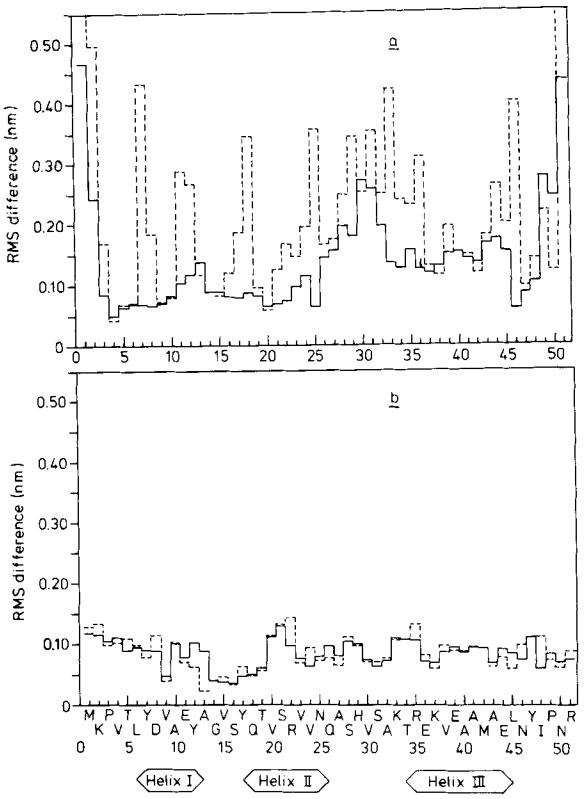


Fig. 3. Root mean square differences between the time-averaged positions of *lac* repressor N, C_α, C, and O backbone atoms (solid line) and side chain atoms (dotted line) averaged over each residue for **a:** free headpiece and headpiece complexed to the operator both in solution and NOE distance restrained and **b:** restrained and unrestrained headpiece in the complex.

tively) results in two operator–headpiece NOE distance constraint violations between 29 His Hε1 and the H8 and H3′ protons of 2 Ade (Table V).

Structure of the Operator

In Figure 5 the operator as seen in the complex is compared to the initial uniform B-DNA conformation. Although the operator deviates clearly from uniform B-DNA, it is still essentially in B-DNA form. In the view looking parallel to the axis of the recognition helix there is a slight narrowing of the major groove introducing a small bend of the DNA helix by which the recognition helix becomes more enclosed by the DNA. The bend is sharpest between base pairs 7 and 8. The minor groove is slightly wider at the position where it interacts with the C-terminal part of headpiece. The total deviation between uniform B-DNA and the average structure is due to the accumulation of small local distortions which are distributed over the total operator. These distortions are induced by the complexation to the repressor and by normal fluctuations; however, because we do not have information on the behavior of the free (simulated) operator it is difficult to distinguish changes caused by binding from changes caused by normal fluctuations.

Deviations from uniform B-DNA are observed for the *trp* operator[14] and synthetic 434 operator DNA[11,12] complexed to their repressors, while DNA complexed to the λ repressor corresponds more closely to regular B-DNA.[13] From Table VI it is shown that the average values and the variations of the helix parameters of the simulated *lac* operator structures are comparable to the values found for

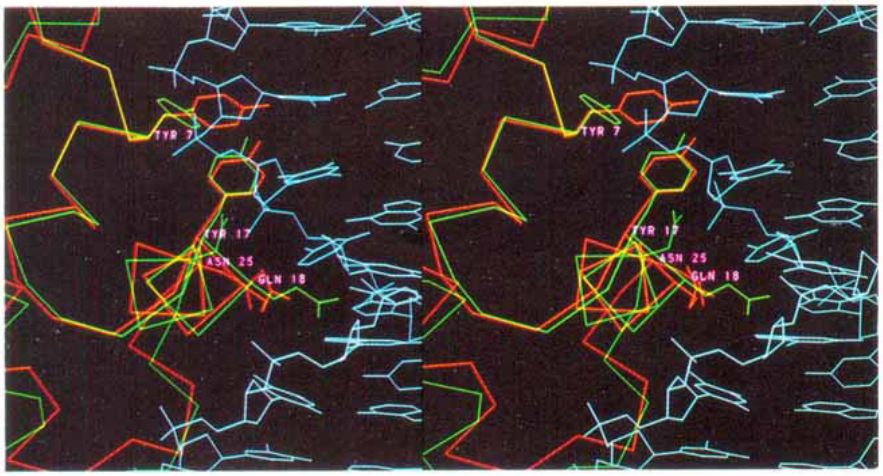


Fig. 4. Stereo view in which the α-carbon trace of helix I, helix II (recognition helix), and a part of the loop region between helix II and III of complexed (gold) and free headpiece (green) are compared. A part of the operator (major groove) is shown in white. Only significantly rearranging residues in the center of the repressor–operator interface are displayed (7 Tyr, 17 Tyr, 18 Gln, 25 Asn). As a result of reorientations of the side chains of 7 and 17 Tyr their phenyl rings are forming a stacked configuration in the complexed headpiece. The side chain of 18 Gln adopts a folded configuration in the complex, which is identical to the one observed in the second Gln residue of the recognition helix in the 434 complex.[11,12] In this way its Cβ, Cγ, and Cδ atoms are brought in front of the 5-methyl group of 6 Thy (see also Fig. 7b).

TABLE V. Violations of NOE Distance Constraints During the Unrestrained Part of the Simulation of Complex in Aqueous Solution*

NOE distance constraint		Upper limit distance	Actual average distance	Violation distance
Headpiece				
12 Tyr Hδ1,2	45 Leu $H_3\delta2$	0.66	0.79	0.13
22 Arg Hα	25 Asn $H_1\beta$	0.35	0.57	0.22
22 Arg Hα	25 Asn $H_2\beta$	0.35	0.59	0.24
23 Val $H_2\gamma1$	38 Val Hβ	0.45	0.56	0.11
34 Thr Hα	37 Lys H	0.40	0.53	0.13
35 Arg H	30 Val $H_3\gamma1$	0.50	0.62	0.12
35 Arg H	30 Val $H_3\gamma2$	0.50	0.72	0.22
Headpiece–Operator				
29 His Hε1	2 Ade H8	0.60	0.83	0.23
29 His Hε1	2 Ade H3′	0.40	0.64	0.24

*The violations are obtained by time averaging over the period 100–125 psec. Distances are given in nm and only violations larger than 0.10 nm are listed.

the *trp* operator fragment in the complex and another crystalline uncomplexed 12 base pair B-DNA fragment.[51] The average helix repeat values for both *lac* operator and *trp* operator are smaller than for uncomplexed B-DNA. The very similar helix parameter values found for the restrained and unrestrained *lac* operator indicate that the conformation is hardly influenced by the intermolecular distance constraints. We note that the hydrogen bonds of the base pairs are restrained in both cases.

Protein–DNA Contacts

The repressor–operator contacts are classified into direct hydrogen bonds, water-mediated hydrogen bonds, and nonpolar contacts, as described in detail below. A stereo view of the contacts between protein and DNA is displayed in Figure 6. The polypeptide chain of the *lac* repressor headpiece (residues 1–51) is folded such that the α-helical axis of helix I (residues 6–13) is at an angle of approximately 35° to the global DNA axis bringing its N-terminus at a hydrogen bond distance of the phosphate group of 9 Cyt. Helix II (residues 17–25) is located in the major groove with its α-helical axis almost perpendicular to the global DNA axis, which implies that the recognition helix is not parallel to the major groove. The N-terminal end of helix II points somewhat into the groove toward position 4, but there are no direct interactions from the N-terminus to the operator. Both sides of the major groove are contacted by the backbone and side chain atoms of helix II. The loop region (residues 26–33) joining helix II and III runs along the sugar-phosphate backbone. The α-helical axis of helix III (residues 34–45) and the global DNA axis lie in one plane and make an angle of approximately 20°. The N-terminus of helix III is

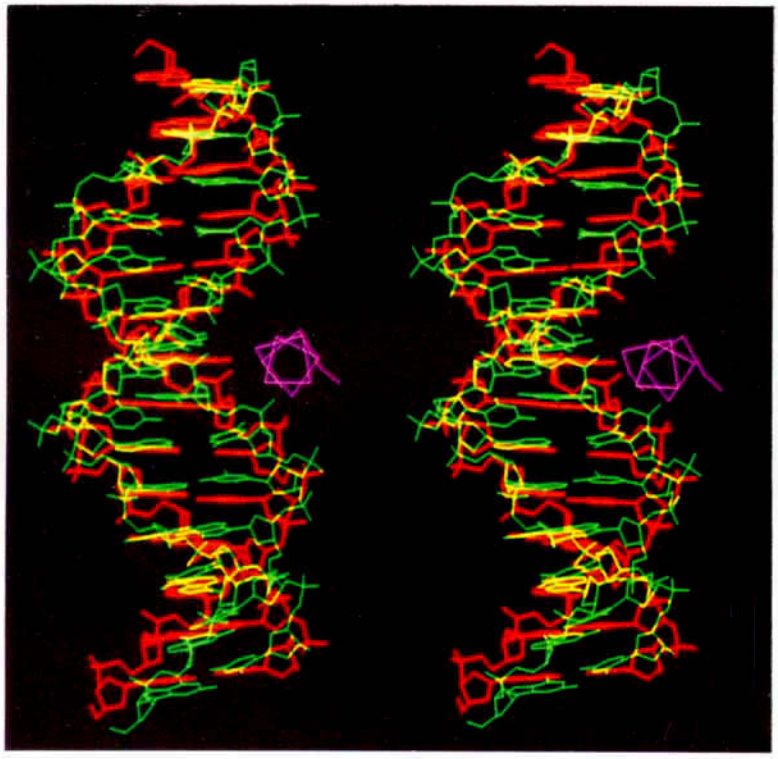


Fig. 5. The DNA structure. Uniform B-DNA structure (red) is superimposed to the average operator structure ((45–70 psec)) as seen in the complex (green). An α-backbone trace of the recognition helix is shown in violet.

TABLE VI. Absolute Value of Helix Parameters for Various Operator Structures Complexed to Headpiece and Averaged Over the Base Pairs*

	Angle between bases of a pair	Base pair roll	Base pair tilt	Helix repeat
Initial structure (B-DNA)	1.2	0.2	4.6	36.0
After MD refinement in vacuo (structure II)	21.2 (8.0)	10.1 (6.6)	7.8 (8.8)	34.9 (8.8)
Average structure in solution				
24–45 psec	18.3 (7.6)	9.1 (5.9)	10.5 (7.8)	33.2 (5.4)
45–70 psec	18.5 (6.7)	7.4 (5.2)	8.4 (5.2)	32.7 (5.1)
100–125 psec	18.1 (6.7)	6.6 (4.2)	7.2 (3.7)	32.7 (5.8)
trp operator fragment in complex†		4.3 (3.5)	2.2 (1.1)	33.8 (5.3)
Crystalline B-DNA‡	14.2 (4.8)	4.9 (3.1)	4.7 (4.0)	36.1 (3.9)

*The values are given in degrees. The standard deviations are given in parentheses.
†Taken from Ref. 14.
‡See Ref. 51.

directed toward the operator, but similar to helix II, is too far away to make direct interactions. For all three helices the location of the N-terminus is closer to the negatively charged phosphate groups than the C-terminus. The C-terminus of the *lac* repressor headpiece ends in the minor groove.

Direct Hydrogen Bonds

The direct hydrogen bonds between the *lac* repressor headpiece and its operator are listed as a function of simulation time in Table VII. The nine most important hydrogen bonds (protein–DNA) can be divided into three backbone–backbone and six side chain–backbone bonds which are stable in the restrained and unrestrained part of the simulation. We did not find stable direct hydrogen bonds to the hydrogen bond donor and acceptor groups of the bases.

The stability of the hydrogen bond between the N-terminus of α-helix I (6 Leu H) and the phosphate oxygen 9 Cyt O2P which exists almost 100% of the simulation time is expected to be a result of oriented

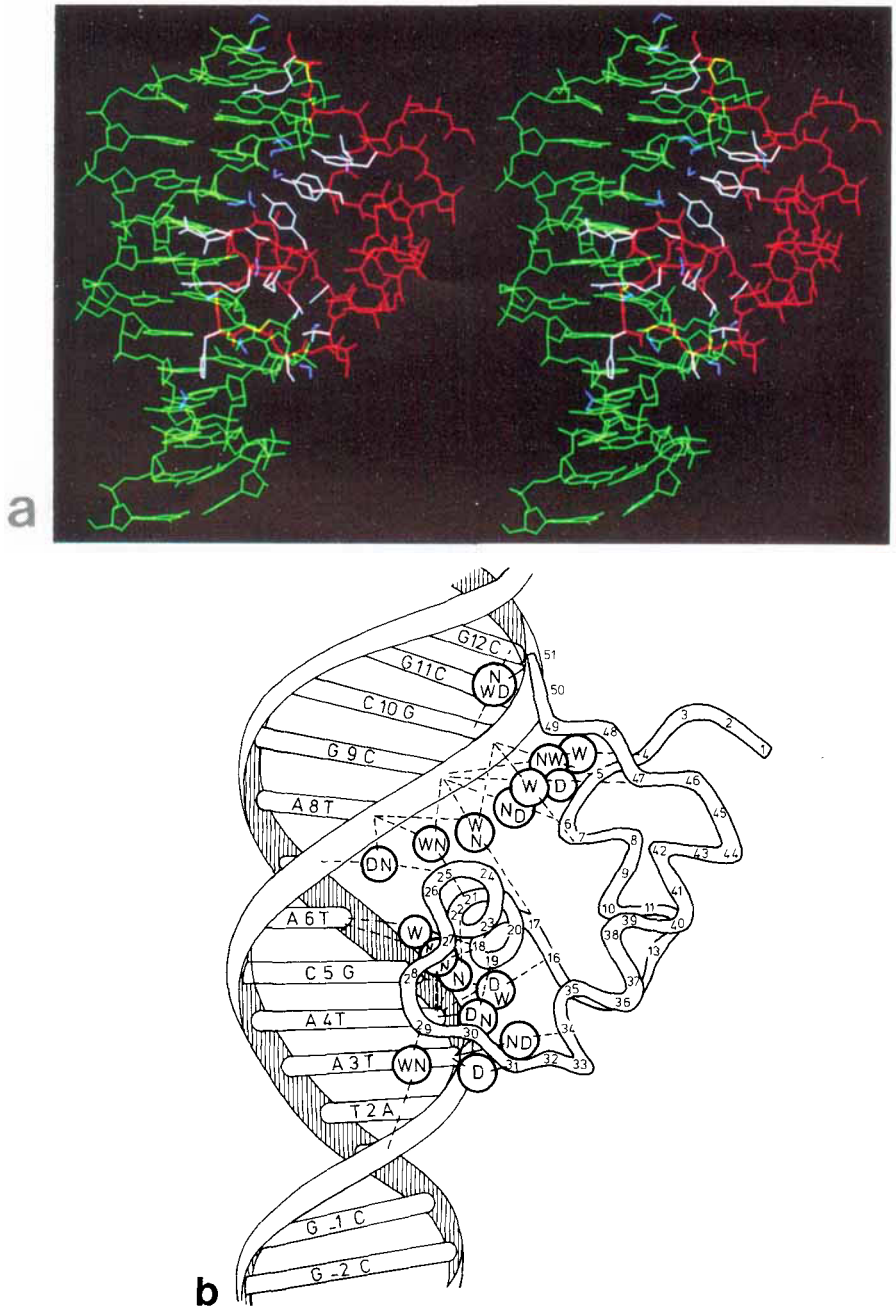
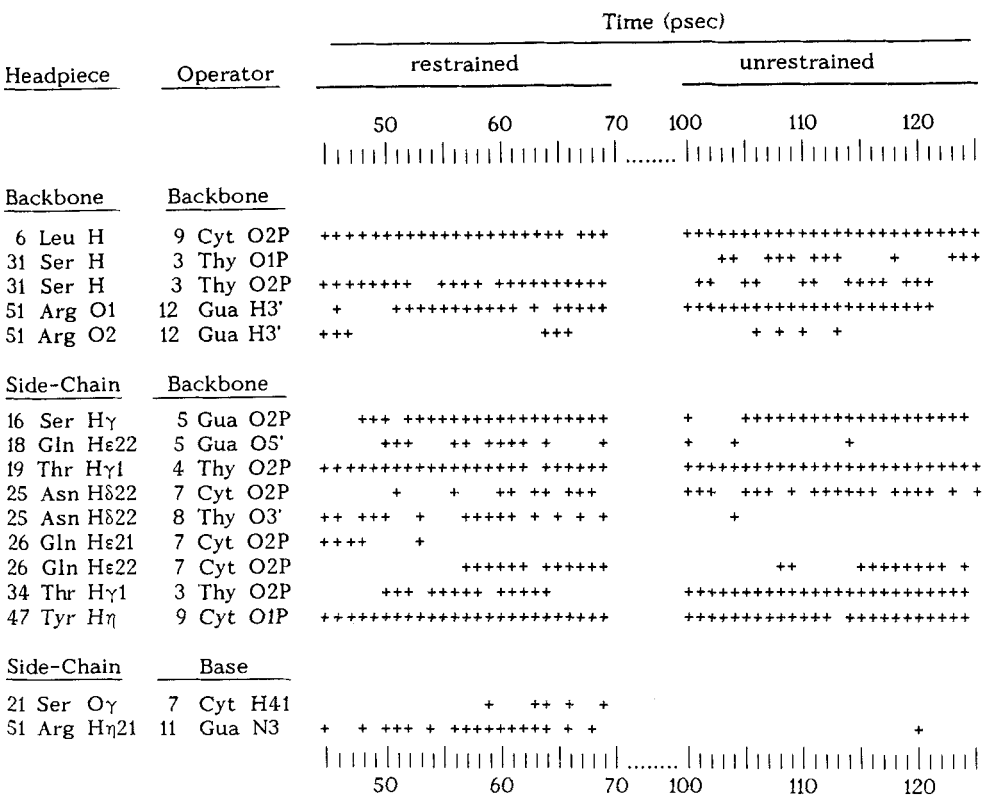


Fig. 6. **(a)** Stereo view of the contacts between protein and DNA as seen in a snapshot structure at t = 57 psec. The N, Cα, Cβ, C, O atoms trace of the repressor is in red. For clarity only side chains of residues involved in contacts (direct- and water-mediated hydrogen bonds or nonpolar) with the DNA are displayed in white: 5 Thr, 6 Leu, 7 Tyr, 16 Ser, 17 Tyr, 18 Gln, 19 Thr, 21 Ser, 22 Arg, 25 Asn, 26 Gln, 31 Ser, 34 Thr, 47 Tyr. Water molecules simultaneous hydrogen bonded to both protein and DNA are shown in blue. In this shapshot structure these water molecules are bridging the following protein–DNA contacts: 5 Thr Hγ1 . . . 10 Gua O1P, 7 Tyr Hη . . . 1O Gua O2P, 16 Ser H . . . 4 Thy O2P, 17 Tyr Hη . . . 9 Cyt O2P, 17 Tyr Oη . . . 8 Thy O4, 18 Gln Hε21 . . . 6 Thy O2P, 21 Ser Hγ . . . 9 Cyt O3′, 22 Arg Hη22 . . . 5 Gua O6, 29 His Hε2 . . . 1 Ade O5′, 34 Thr H . . . 3 Thy O2P, 47 Tyr Oη . . . 8 Thy O1P, 51 Arg O2 . . . 12 Gua O1P (see Table VIII). **(b)** A sketch of the protein–DNA contacts during the unconstrained part of the simulation (100–125 psec). The view is approximately corresponding to the view in **(a).** The topology of the contacts (not their exact positions) is indicated by circles: D = direct hydrogen bond, W = water-mediated hydrogen bond, N = nonpolar contact. The hydrogen bonds (D,W) are indicated when observed for more than 50% of the unconstrained simulation period (100–125 psec). The hydrogen bond criteria are given in the footnote to Table VII. The numbers in the protein *indicate Cα positions.*

TABLE VII. Hydrogen Bonds Between the *lac* Repressor Headpiece and the Operator as a Function of Simulation Time*



*Hydrogen bond criteria: hydrogen-acceptor distance ≤0.25 nm; donor-hydrogen-acceptor angle ≥135°. A (+) sign indicates that at the time point (full psec) between the 45th and 70th psec (restrained MD, see text) and the 100th and 125th psec (unrestrained MD, see text) a H-bond exists. Only hydrogen bonds are listed that exist in total four or more times.

peptide dipoles in the α-helix.[52] Analogous interactions between the N-termini of the first helix of the helix-turn-helix motif and the backbone phosphate group occur also in the 434, λ, and *trp* repressor–operator complexes.[11–14] This interaction can probably be seen as a common feature which anchors each repressor tightly to the DNA at a fixed position.

The very stable hydrogen bonds between the amide nitrogens on the peptide backbone of 6 Leu and 31 Ser and the phosphate groups on DNA strongly constrain the fit and orientation of DNA versus protein and at the same time reduce the flexibility of the backbone of headpiece. The positively charged side chain of 22 Arg is not only anchored by water-mediated hydrogen bonds, but will also contribute to the interaction by longer ranged Coulombic interactions.

Unexpected is the very stable hydrogen bond between Hη of 47 Tyr and 9 Cyt O1P. 47 Tyr is an important residue of a hydrophobic core involving residues of all three helices. This hydrophobic core strongly constrains the relative orientation of these helices and thereby the relative position of the polar 47 OH group. Therefore it will be likely that the strong hydrogen bond between 47 Tyr and 9 Cyt in turn stabilizes the exact orientation of the three helices. The role of the 47 Tyr OH group is probably twofold: (1) it can be seen as a fixed polar recognition point in repressor–operator binding and (2) its strong hydrogen bond with the operator will indirectly constrain the specific helix-turn-helix motif in the complex. We note that these observations are in agreement with photo-CIDNP experiments which showed that 47 Tyr is inaccessible to a dye in the headpiece–operator complex but partly accessible in the free headpiece.[48] The fact that the *lac* repressor binding is sensitive to substitutions of 47 Tyr[53] (among which a Tyr → Leu replacement) is obviously not only caused by a distortion of the tertiary structure as suggested before[26–28] but also by the loss of an important direct interaction.

It is observed that amino acid side chains of residues which can be replaced without loss of repressor activity[53] (residues 2–4, 11, 12, 36, 39, 44) appear not to be involved in interactions with the operator. An exception is found for residue 26 Gln, which can be substituted without significant alteration of the repressor operator recognition,[53] but nevertheless forms a moderately stable hydrogen bond between its side chain and the backbone of DNA.

Hydrogen Bonds via Water Molecules

The water-mediated hydrogen bonds between the *lac* repressor headpiece and its operator are listed in Table VIII. On the average 11–13 water molecules are involved in bridging protein–DNA contacts. The backbone atoms of 4 Val, 16 Ser, and 51 Arg and the side chain atoms of residues 5 Thr, 7 Tyr, 16 Ser, 17 Tyr, 18 Gln, 21 Ser, 29 His, and 47 Tyr are involved in stable water-mediated hydrogen bonds to the O3′, O5′, and phosphate oxygens of the repressor. Residues 16 Ser and 47 Tyr are also involved in stable direct hydrogen bonds. As a consequence these two residues are probably tightly bound to the operator and must be important for DNA recognition or DNA–protein fit. We note that the repressor is sensitive to substitutions of both residues.[53]

The increase in the pK_a of 29 His by 0.5 unit when bound to the operator is probably caused by a water-mediated ionic interaction between 29 His Hε2 and 1 Ade O5′/O1P, rather than a direct ionic interaction as was suggested before.[54]

The seven observed water-mediated hydrogen bonds between the side chain of residues 17 Tyr, 21 Ser, and 22 Arg and base pairs within the major groove are not very stable over the total simulation time and one may argue whether these contacts can explain the specificity of the repressor. The direct contact between 21 Ser and 7 Cyt during the restrained part (Table VII) is replaced by a stable water-mediated contact during the unrestrained part (Table VIII). Residue 22 Arg makes alternating water-mediated hydrogen bonds to atoms of base pairs at positions 5, 6, and 4 in order of importance.

The contacts between 51 Arg and 11 Gua within the minor groove are consistent with the proposal that the minor groove at base pairs 9–11 is covered by residues 51–59 of the full repressor.[55] Interestingly, a corresponding case in which an arginine is bridged by water molecules to the minor groove is observed in the 434 complex.[12]

From the total of 43 more or less stable water-mediated hydrogen bonds between the repressor and its operator approximately 40% are observed during both the restrained and unrestrained part of the MD simulation. For the more stable ones this percentage is almost 80%.

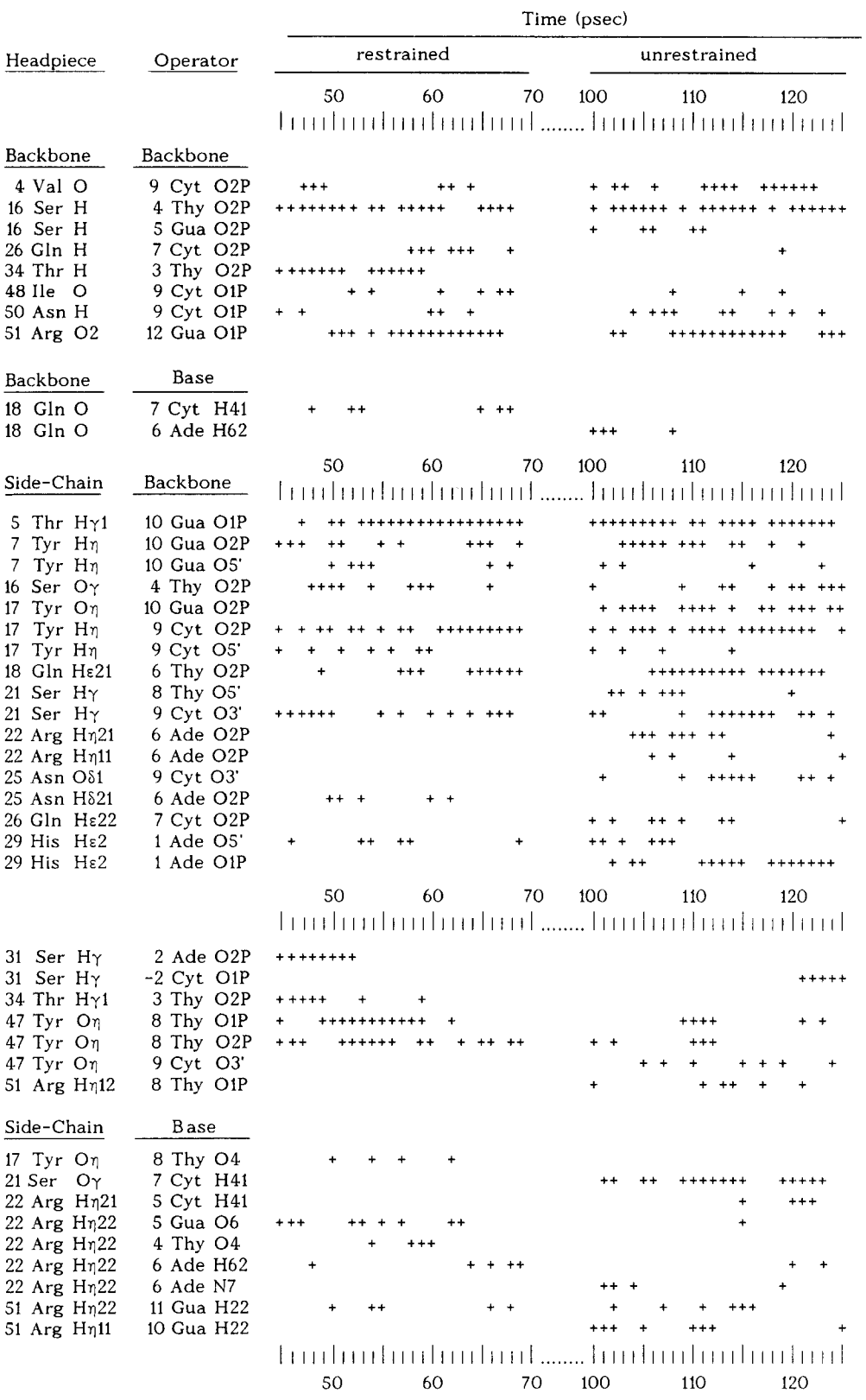
Nonpolar Contacts and Specificity

In Table IX the nonpolar contacts between repressor and its operator which are close enough to exclude the surrounding solvent in the interface between both molecules are listed. The many van der Waals contacts are uniformly distributed over the surfaces of repressor and operator and are obviously important for the fit and the recognition between repressor and headpiece.

The contacts involving the thymine 5-methyl groups (C7) are potential recognition sites where pyrimidines can strongly be distinguished from purines in the major groove since the purine N7 can accept a hydrogen bond at this site while the pyrimidine C7 group is unable to form this bond and also protrudes about 0.1 nm further into the wide groove.[15] It is interesting that the tight protein backbone–base contacts between 18 Gln C and 19 Thr Cα with 4 Thy C7 accompanied by side chain contacts of these residues with sugar backbone atoms of 4 Thy and 3 Thy forming a small hydrophobic pocket. This protein backbone to thymine C7 contact is likely to be quite exact due to the restricted flexibility of the main chain atoms of a folded protein. Other small hydrophobic pockets are formed by the side chains of 17 Tyr and 21 Ser receiving the methyl group of 8 Thy (Fig. 7a) and carbon atoms of 18 Gln receiving the methyl group of 6 Thy (Fig. 7b). Such nonpolar contacts between a methyl group of thymine and—in the first two cases—two side chains forming a small hydrophobic pocket are also observed in the 434 and λ complex.[11–13] Although these contacts essentially discriminate between purines and pyrimidines, they possibly distinguish between the nonpolar C5 hydrogen of cytosine and the 5-methyl group of thymine (C7) at positions characterized by exact fit between repressor and operator. The latter can for example be caused by stable backbone–backbone or backbone–base interactions as observed for the main chain atoms of 6 Leu, 31 Ser, 18 Gln, and 19 Thr. Thus, these nonpolar contacts with the methyl group of thymine are probably very important for the specific recognition between headpiece and operator.

The importance of the 5-methyl group of 8 Thy—in the simulation embedded between 17 Tyr Cε2–Cζ and 21 Ser Cβ (Fig. 7a)—was illustrated by Caruthers[56] using the following evidence. Removal of the methyl group by an AT → AU replacement dramatically reduced the affinity of repressor for operator. A similar loss was observed for the AT → GC mutation. However, insertion of 5-methylcytosine for cytosine stabilized the repressor–operator interaction to the same extent as the naturally occurring adenine–thymine base pair and the reduction of the stability of the complex by using 5-bromouracil (U^+) instead of uracil for an AT → AU^+ replacement is also much less dramatic. The thymine 5-methyl group (C7) occupies sterically the same position as does this methyl group in 5-methylcytosine or the bromine in 5-bromouracil. This means that the 5-

TABLE VIII. Hydrogen Bonds Between the *lac* Repressor Headpiece and the Operator That Are Bridged by a Water Molecule as a Function of Simulation Time*



*See the footnote to Table VII for details.

TABLE IX. Nonpolar Contacts Between Headpiece and Operator Observed in the Average Complex Structures*

Headpiece	Operator	Distance (nm) Restrained	Distance (nm) Unrestrained
Backbone	Base		
18 Gln C	4 Thy C7	0.40	0.40
19 Thr Cα	4 Thy C7	0.40	0.40
Side Chain	Backbone		
5 Thr Cγ2	10 Gua C5′	0.38	0.39
5 Thr Cγ2	10 Gua C4′	0.43	0.42
5 Thr Cγ2	10 Gua C3′	—	0.43
6 Leu Cδ1	9 Cyt C3′	0.38	0.38
6 Leu Cδ1	9 Cyt C5′	0.43	—
18 Gln Cγ	4 Thy C2′	0.37	0.40
19 Thr Cγ2	3 Thy C3′	0.43	0.39
19 Thr Cγ2	3 Thy C2′	0.43	0.38
25 Asn Cβ	8 Thy C2′	0.40	—
25 Asn Cβ	8 Thy C3′	0.42	—
25 Asn Cγ	8 Thy C2′	—	0.41
29 His Cδ2	1 Ade C3′	—	0.39
29 His Cε1	1 Ade C2′	0.39	—
29 His Cε1	1 Ade C3′	0.39	0.42
29 His Cε1	2 Ade C3′	0.40	—
29 His Cε1	2 Ade C2′	0.40	—
34 Thr Cγ2	3 Thy C5′	0.38	0.40
51 Arg Cβ	12 Gua C4′	0.40	0.42
51 Arg Cβ	12 Gua C5′	0.41	0.43
51 Arg Cδ	10 Gua C1′	0.41	0.41
51 Arg Cδ	10 Gua C3′	0.42	—
Side Chain	Base		
17 Tyr Cε2	8 Thy C7	0.43	0.40
17 Tyr Cζ	8 Thy C7	0.39	0.41
18 Gln Cβ	5 Gua C8	0.35	0.38
18 Gln Cβ	6 Thy C7	0.42	0.43
18 Gln Cγ	5 Gua C8	0.37	0.36
18 Gln Cγ	4 Thy C6	0.43	0.41
18 Gln Cδ	6 Thy C7	0.41	0.40
18 Gln Cδ	5 Gua C8	0.43	0.40
21 Ser Cβ	8 Thy C7	0.40	0.43
22 Arg Cβ	4 Thy C7	0.42	0.42
22 Arg Cγ	4 Thy C7	0.43	0.43
25 Asn Cβ	7 Cyt C5	0.42	—
25 Asn Cβ	7 Cyt C6	0.43	—
25 Asn Cγ	7 Cyt C6	0.36	—
25 Asn Cγ	7 Cyt C5	0.39	—

*Only carbon–carbon pairs with a contact distance ≤ 0.43 nm are listed.

methyl group of 8 Thy is recognized by the *lac* repressor and that the remainder of either base pair is of no consequence. The latter is consistent with the absence of (stable) direct or water-mediated hydrogen bonds to base pair atoms of AT 8 in our simulation. However, other atomic details of a base pair appear to be important as well, as illustrated by numerous biochemical studies.[18,56,57] It is questionable whether the specificity of the *lac* repressor headpiece/operator can be explained by the observed combinations of hydrogen bonds and nonpolar contacts alone. Nowhere in the simulation is a base pair simultaneously recognized by two or more hydrogen bonds as needed for a unique identification.[15] This does not exclude the possibility that the specific hydrogen bonds are interchanging rapidly in a dynamic specific hydrogen bond network between repressor and operator which can be recognized only over a much longer simulation time.

The absence of direct hydrogen bonds to the bases that explain recognition is not completely unexpected. The crystal structure of *trp* repressor/operator complex at 2.4 Å resolution[14] shows no direct hydrogen bonds to the bases that can explain the repressor's specificity for the operator sequence in spite of an extensive contact surface, including 24 direct and 6 solvent-mediated hydrogen bonds to the phosphate groups. Otwinowski and colleagues[14] argue that the *trp* repressor provides an example of indirect readout. In other words, B-DNA exhibits a large degree of sequence-dependent structural variation which makes it the primary function of the

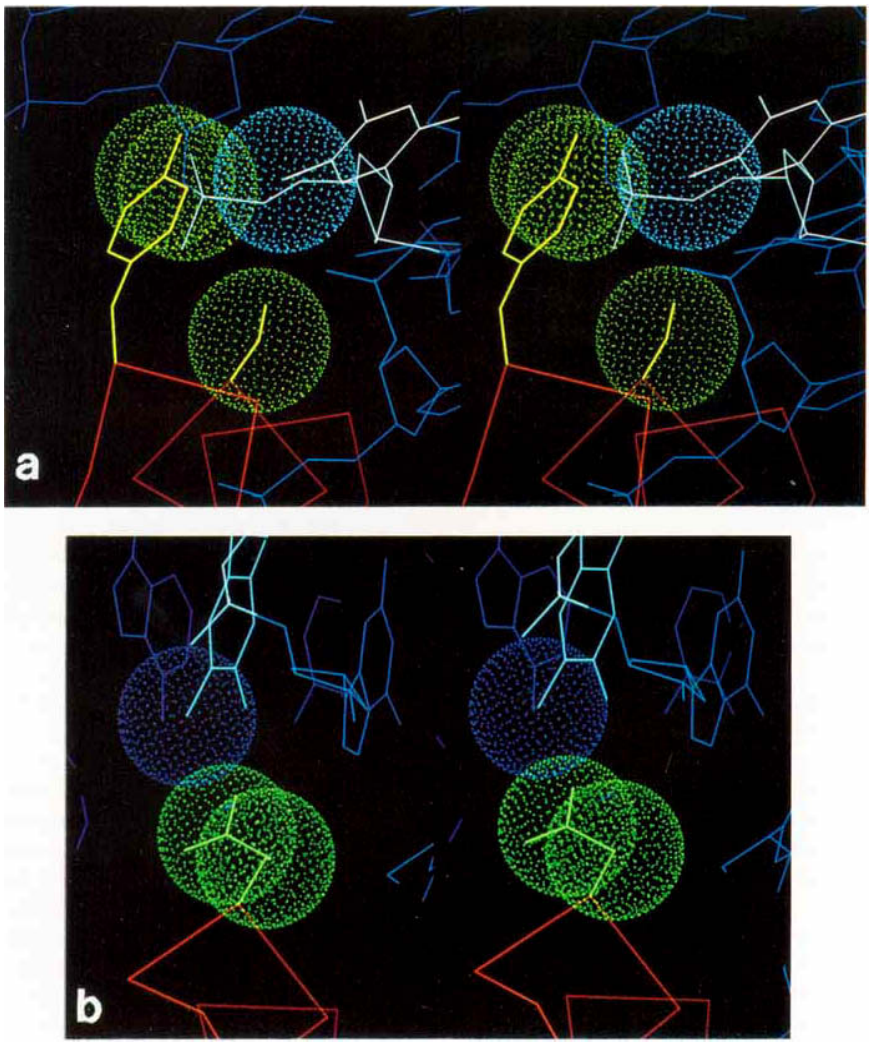


Fig. 7. **(a)** Stereo view of a small hydrophobic pocket formed by 17 Tyr Cε2, Cζ and 21 Ser Cβ receiving the 5-methyl group of 8 Thy in a dot surface representation. 17 Tyr Oη, Hη and 21 Ser Oγ, Hγ are involved in direct or water-mediated hydrogen bonds (not indicated). **(b)** Stereo view of the nonpolar interaction between the Cβ, Cδ atoms of 18 Gln and the 5-methyl group of 6 Thy. The adopted side chain configuration of 18 Gln is fixed by an intraresidual HN–Oε1 hydrogen bond. The color code in **(a)** and **(b)** is as follows: backbone protein, red; side chains of 17 Tyr and 21 Ser **(a)** and 18 Gln **(b)**, green; DNA, blue; 8 Thy **(a)**, 6 Thy **(b)**, white. The dot surfaces of the 5-methyl groups of 8 Thy **(a)** and 6 Thy **(b)** are blue. For clarity the van der Waals radii in the dot representation of the carbon atoms are chosen approximately 10% smaller than the actual values used in the simulation.

DNA sequence to permit the formation of a tight complex with the protein. A corresponding example is the absence of direct hydrogen bonds between the central base pair 3 of 434 operator and its repressor, although the binding affinity is lowered considerably by changes in this central base pair.[11,12] The number of hydrogen bonds to base pairs 4 and 5 of 434 operator is insufficient to uniquely identify a particular base pair.[15] This means that the recognition is not based on direct hydrogen bonding alone.

We expect that the many van der Waals contacts are important for binding and recognition and with exception of the 25 Asn Cβ/Cγ–7 Cyt C5/C6 and the 29 His Cε1–2 Ade C2′/C3′ contacts the van der Waals contacts are essentially stable for both the restrained and unrestrained parts of the MD run (Table IX). The uniqueness of the van der Waals contacts is enhanced by the fact that the surfaces of both protein and DNA at the interface are still extensively covered by water in spite of the many close contacts between repressor and operator (see below). This will restrict the exposed area of methylene and methyl groups in the interface and as a result enlarge the specificity of the van der Waals interactions. On the other hand, the importance of interactions through hydrogen bonds will be lowered in an aqueous environment, because when protein–DNA hydrogen bonds are lost, protein–water and DNA–water bonds are gained.

TABLE X. Average Total Number of Hydrogen Bonds of *lac* Headpiece*

	Average number of hydrogen bonds	
	Free	Complex
Hydrogen bonds between		
Headpiece	27.9	25.9
Headpiece–solvent	83.1	92.5
Headpiece–operator	—	9.6
Total number	111.0	128.0

*The average number of hydrogen bonds is calculated using a time resolution of 0.04 psec and the same hydrogen bond criterion as in the footnote to Table VII.

Differences in Hydration of Complexed and Uncomplexed Headpiece

In Table X the average total numbers of hydrogen bonds between headpiece–solvent, headpiece–DNA, and headpiece internally are shown for uncomplexed and complexed headpiece. Surprising, but at second thought not unexpected, is the fact that the average number of repressor–solvent hydrogen bonds is significantly increased during complexation, which will be an important compensation for the loss of mobility (entropy) in the complexed headpiece structure. Including the headpiece–DNA hydrogen bonds the total number of hydrogen bonds is increased from 111.0 in the free to 128.0 in the complexed headpiece.

Figure 8 shows the average percentage of hydrogen-bonded water to the backbone and side chain atoms per residue of complexed (top) and uncomplexed (bottom) headpiece. High-solvation values are observed for most of the residues directly involved in the interaction between headpiece and operator. This means that large parts of the binding site surfaces of both headpiece and DNA are still covered with water.

The *lac* repressor is thought to slide along unspecific DNA to explain the fact that *lac* repressor finds its operator target at a higher rate than expected for a simple diffusion-controlled reaction.[58–60] We note that if sliding from one binding site to the next does not require stripping of (many) water molecules, the movement will certainly have a lower activation energy and thereby a higher sliding rate, i.e., the water can act as a lubricant!

Although many residues are still hydrated in the complex, some residues dehydrate to some extent. The percentages of water hydrogen bonded to the backbone (bb) and/or side chain (sc) atoms of 6 Leu (bb), 19 Thr (sc), 25 Asn (bb), 28 Ser (sc), 30 Val (bb), and 31 Ser (bb) are zero or very low in the complex, but are moderate in the free headpiece (Fig. 8). For residues 6 Leu, 19 Thr, and 31 Ser these decreases are due to direct hydrogen bonds to the operator while for the other residues these decreases are due to the increase of hydrogen bonds with neighbor residues in the complexed headpiece.

The opposite effect is observed for residues 5 Thr, 15 Val, 34 Thr, 35 Arg, 37 Lys, and 46 Asn for which the hydration is increased upon complexation as a result of a decrease of the percentage of intramolecular hydrogen bonding. For example, the percentage of the intramolecular hydrogen bonding of 5 Thr is decreased from 348% to 178% while the percentage of hydrogen bonded water to this residue is increased from 33% to 179% in free and complexed headpiece, respectively. We note that in the complex 5 Thr is involved in a strong water-mediated hydrogen bond and in a side chain–backbone van der Waals contact (Table VIII and IX). Figure 9 shows that the nonpolar contact between 5 Thr Cγ2 and 10 Gua C3′, C4′, C5′ will position the N-terminus of helix I at the correct distance to make the extremely stable hydrogen bond between 6 Leu H and 9 Cyt O2P (see section "Direct Hydrogen Bonds"). The nonpolar contact at this position will reduce the accessibility of water molecules that would otherwise decrease the stability of this hydrogen bond by serving as competitive hydrogen bond donors and acceptors. The strong water-mediated hydrogen bond between 5 Thr Hγ1 and 10 Gua O1P will not only stabilize the interaction between operator and headpiece but will also constrain the orientation of the side chain of 5 Thr. This water-mediated hydrogen bond will therefore be indirectly important for the stability of the interaction between helix I and the phosphate group of 9 Cyt. Replacement of 5 Thr by Met, Lys, Gly, Leu, Tyr, or Gln results in defective repressors, whereas a Thr-Ser replacement yields only a partially defective repressor.[53] This can structurally be explained by the abilities of both aliphatic hydroxyl side chains to make the water-mediated hydrogen bond without making a bad contact with the sugar carbon atoms. That the affinity of the repressor for the operator is only partially restored by using serine instead of threonine can be explained by the loss of the nonpolar contact between the Cγ2 of threonine and the sugar carbon atoms; thus, as a second effect the replacement by serine will lower the stability of the hydrogen bond between helix I and the phosphate backbone as explained above. Thr 5 is one of the conserved residues between CAP, *gal*, and *lac* repressors.[9]

DISCUSSION

The interaction between the *lac* repressor headpiece and its operator is based on many direct- and water-mediated hydrogen bonds and nonpolar contacts which allow the formation of a very tight complex. Many of the contacts with the DNA are made by residues of the recognition helix, and residues around the N-terminal part of helix I, the loop region between helix II and III, and the C-terminal part of headpiece make other important contacts.

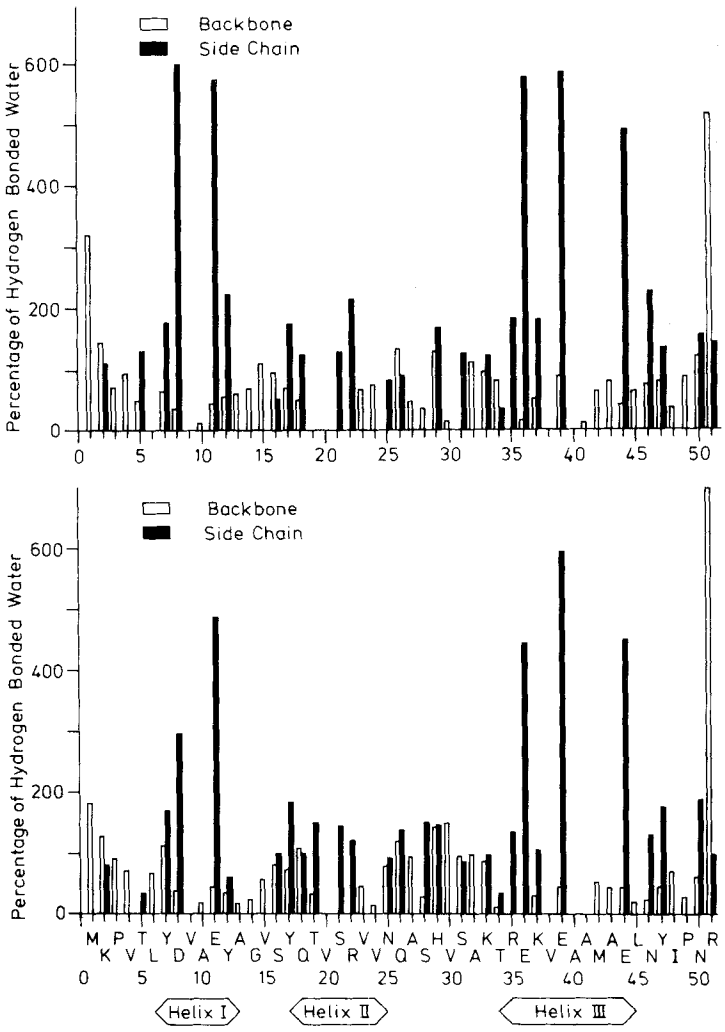


Fig. 8. Percentage of hydrogen bonded water summed over the HN and O backbone atoms (open bars) and side chain atoms (filled bars) averaged over each residue. Top: headpiece in complex; bottom: free headpiece. The percentage of hydrogen bonds is calculated using a time resolution of 0.04 psec and the same hydrogen bond criterion as in the footnote to Table VII.

The positions of all three helices are restrained by stable direct or water-mediated hydrogen bonds between backbone or polar atoms (of mostly short side chains) around the beginning and end of these helices and the operator.

No stable hydrogen bonds between side chains and bases are found, while nonpolar contacts to the 5-methyl group of thymine and, to a lesser extent, water-mediated hydrogen bonds appear to be essential for operator recognition. On the average 11–13 water molecules are involved in bridging protein–DNA contacts. The absence of (stable) direct hydrogen bonds to the bases to contribute part of the specificity is not surprising because the interface is highly covered with water and feasible base–protein hydrogen bonds can easily be replaced by bonds to water. On the other hand, the aqueous environment of the surface will restrict the exposed area of methylene and methyl groups and as a result enlarge the specificity of the nonpolar contacts. The free energy of a van der Waals contact in these cases between methyl groups will be similar or even higher than the free energy of a hydrogen bond. This corresponds with previous estimates for the energy of each type of interaction obtained from protein engineering studies[61]: 1.5–3.4 kcal mol^{-1} for van der Waals contacts and 0.5–1.5 kcal mol^{-1} for leaving an unpaired, uncharged hydrogen bond donor or acceptor.

The results show clearly that there is no simple one-to-one recognition code between amino acids and base pairs. One residue can contact in different ways several backbone and base pair atoms of the DNA. For instance, the side chain of 18 Gln makes nonpolar contacts with base pairs 4–6, a water-mediated hydrogen bond to 6 Thy O2P, and a direct hydrogen bond to 5 Gua O5′. Another example is the

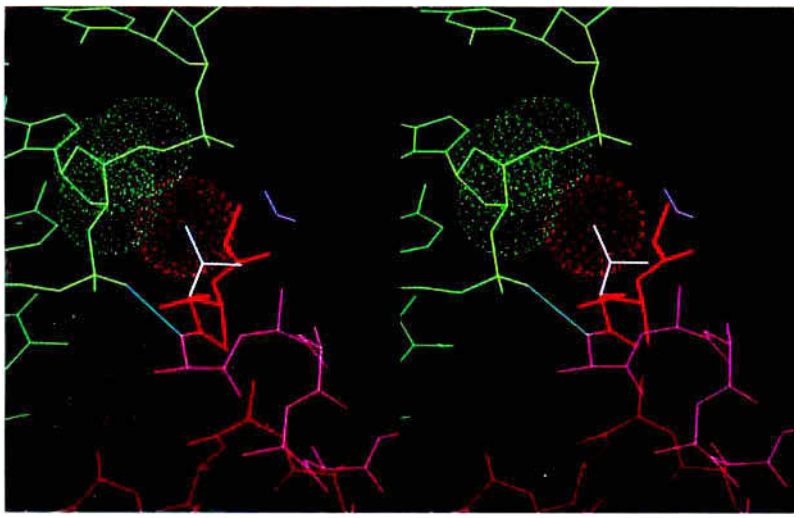


Fig. 9. The nonpolar contact between 5 Thr and 10 Gua reduces the water accessibility at the neighboring hydrogen bond between the N-terminus of helix I and the phosphate group at position 9. The C5′, C4′, and C3′ atoms of 10 Gua (green) and Cγ2 of 5 Thr (red) are represented in a dot surface. The water molecule bridging 5 Thr Hγ1 and 10 Gua O1P is shown in white. This water-mediated hydrogen bond exists ca. 92% of the simulation time (45–125 psec). The DNA is green, α-helix I is violet, while the rest of the repressor is red. The side chain of 5 Thr is white. The hydrogen bond between the N-terminus of helix I (6 Leu N) and the phosphate oxygen 9 Cyt O2P which exists almost 100% of the simulation time is indicated by a light green line.

5-methyl group of 8-thymine, which is buried by contacts with the side chains of two residues (17 Tyr, 21 Ser).

It also becomes clear that some protein–DNA interactions will influence the stability or feasibility of other protein–DNA interactions. This is nicely illustrated by the nonpolar contact between 5 Thr and 10 Gua, which appears to strengthen the stability of the hydrogen bond between the N-terminus of helix I and the phosphate group of 9 Cyt by excluding water.

Although our model can explain most of the available genetic and biochemical data in terms of structural considerations, we also found an exception. A genetic study performed by Ebright[62] showed that the intact *lac* repressor in which 18 Gln is substituted by glycine, serine, or leucine loses the ability to distinguish between base pairs GC, TA, and AT at position 7. In contrast the three substituted variants retain the ability to discriminate at positions 5, 6, and 8–10 by factors between 9 and 37. Therefore, Ebright suggested a direct contact between 18 Gln and base pair 7. In our model we did not observe a direct interaction between base pair 7 and 18 Gln that could explain these results. The lack of such interactions could at first be taken as evidence of the inadequacy of the force field used for the simulation. However, the interaction between 18 Gln and base pair 7 can in principle be very long ranged, by indirectly, but still specifically affecting the equilibrium configurations of the surrounding solvent or making interactions at other repressor and operator positions possible. The difficulty in making a direct link between a genetic result and a structural observation (observed in a simulation that can of course be unreliable) once again underscores the insufficient understanding of the principles of how an amino acid recognizes a particular base pair. A similar difficulty is experienced in the case of the *trp* repressor. Genetic studies, indicating specific interactions between 81 Thr of *trp* repressor and two critical base pairs of its operator,[63] are in contrast with observed interactions between 81 Thr and a phosphate group in the high-resolution X-ray crystallography structure of the complex.[14] Another example is the tight-binding mutant of the λ-repressor in which the glutamic acid is changed into lysine at position 83. The terminal nitrogen of lysine is more than 10 Å away from the DNA, and it is not clear why this mutant binds more tightly.[13]

To understand the principles of the specific interactions between protein and DNA it will be necessary to investigate several individual protein–DNA complexes. Molecular dynamics techniques can be an important tool to understand complex interactions which depend on the simultaneous influence of several forces and entropic effects. A free energy calculation in which a side chain or base is mutated is not only a useful technique to visualize changes explaining long-range influences at the atomic level, but also yields relative free energy differences. In this way experimental information on binding affinity or specificity of mutant studies can be coupled to theoretical calculations.[64] Very interesting free energy calculations would involve mutations of 18 Gln and base pair 7 and repeating the biochemical experiments of Caruthers[56] by mutating the methyl group of 8 Thy into a hydrogen. Such calculations

will be a good test to judge the adequacy of the simulation, but, unfortunately, still require thousands of hours of supercomputer time.

Although the repressor–operator system of the *lac* operon has been extensively studied by genetic analysis of numerous mutants,[53] our work shows that some interesting substitutions are still lacking. An interesting substitution would be the replacement of 47 Thr, an important residue of a hydrophobic core involving residues of all three helices, by Phe, since the strong hydrogen bond between 47 Tyr Hη and 9 Cyt O1P seems to stabilize the orientation of the three α-helices of the *lac* headpiece. In the case of Phe a stable hydrophobic core is expected to form, but the important hydrogen bond will be lacking. Another substitution would be the replacement of 5 Thr by Val, in which case the hydrophobic contact that strengthens the neighboring hydrogen bond between helix I and the phosphate group at position 9 will remain.

The results of the molecular dynamics simulation of the *lac* repressor–operator complex in aqueous solution shows the value of combining computer simulation techniques and experimental NOE distance information. We note that it is essential to perform the molecular dynamics simulations in solvent, because the interface contains a number of significant water molecules and the many different equilibrium configurations of the solvent around the complex will influence the stability of interactions between the repressor and operator. Certainly, static molecular modeling procedures and docking studies performed in vacuo are insufficient to study this kind of complex molecular interactions, while the objectiveness of the simulation techniques allows unprejudiced results. Although these sorts of simulations need a substantial amount of computer time, the rapid developments in hardware and software should make these techniques routine in just a few years.

It is satisfying to observe that the simulated complex in water is intrinsically stable without application of NOE distance constraints. The pattern of repressor–operator contacts does not essentially change upon release of the NOE restraints and only 9 out of 212 switched-off distance constraints are violated by more than 0.1 nm. These results are very encouraging, but it is important to realize that atomic interaction functions are only approximately known, while the time span covered by the MD simulation is limited. This means that MD simulations are not sufficiently reliable to substitute for experimental structure determination methods and that additional experimental tests to confirm the results of the simulations are essential. However, the fact that the observed results are compatible with most of the available experimental data and consistent with the findings of X-ray studies of other repressor–operator complexes suggests that the overall character of the observed interactions between the *lac* repressor headpiece and its operator are essentially correct.

The simulation of the *lac* repressor headpiece–operator complex yields no evidence that a so-called "direct readout" mechanism[14] for recognition is based on direct side chain–base hydrogen bonds. Direct readout occurs rather through nonpolar contacts and water-mediated hydrogen bonds, while the many side chain–backbone contacts support a role of an indirect readout mechanism as well. This means that the mode of interaction between the *lac* repressor headpiece and its operator is somewhere in the middle between the mode used by the 434 and λ complex which display both direct hydrogen bonds and nonpolar contacts to the bases, and the mode used by the *trp* repressor lacking such interactions.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Foundation for Chemical Research (S.O.N.) with financial aid from the Netherlands Organisation for Scientific Research (N.W.O.). The calculations took about 2000 hours on the CONVEX C1-XP mini-supercomputer of the BIOSON institute. We thank Drs. R. Kaptein and R. Boelens for providing us with the NOE distance information and Dr. J.A.C. Rullmann for providing coordinates of complex structures generated by the ellipsoid program. We thank also Drs. J.P.M. Postma and G. Vriend for their assistance in using the model-building programs MIDAS and WHATIF, respectively, and Drs. A.P. Heiner and W. Keegstra for their assistance in using the UNIX operating system.

REFERENCES

1. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M., Pabo, C.O. Homology among DNA-binding protein suggests use of a conserved super-secondary structure. Nature (London) 298:447–451, 1982.
2. Matthews, B.W., Ohlendorf, D.H., Anderson, W.F., Takeda, Y. Structure of the DNA-binding region of *lac* repressor inferred from its homology with *cro* repressor. Proc. Natl. Acad. Sci. U.S.A. 79:1428–1432, 1982.
3. Anderson, W.F., Ohlendorf, D.H., Takeda, Y., Matthews, B.W. Structure of the *cro* repressor from bacteriophage λ and its interaction with DNA. Nature (London) 290:754–759, 1981.
4. Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y., Matthews, B.W. The molecular basis of DNA-protein recognition inferred from the structure of *cro* repressor. Nature (London) 298:719–723, 1982.
5. Shevitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L., Sigler, P.B. The three-dimensional structure of *trp* repressor. Nature (London) 317:782–786, 1985.
6. Weber, I.T., Steitz, T.A. Model of specific complex between catabolite gene activator protein and B-DNA suggested by electrostatic complementarity. Proc. Natl. Acad. Sci. U.S.A. 81:3973–3977, 1984.
7. Pabo, C.O., Lewis, M. The operator-binding domain of λ repressor: Structure and DNA recognition. Nature (London) 298:443–447, 1982.
8. McKay, D.B., Steitz, T.A. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. Nature (London) 290:744–749, 1981.
9. Weber, I.T., McKay, D.B., Steitz, T.A. Two helix DNA binding domain motif of CAP found in *lac* repressor and *gal* repressor. Nucl. Acid. Res. 10:5085–5102, 1982.

10. Pabo, C.O., Sauer, R.T. Protein-DNA recognition. Annu. Rev. Biochem. 53:293–321, 1984.
11. Anderson, J.E., Ptashne, M., Harrison, S.C. Structure of the repressor–operator complex of bacteriophage 434. Nature (London) 326:846–852, 1987.
12. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M., Harrison, S.C. Recognition of a DNA operator by the repressor of phage 434: A view at high resolution. Science 242:899–907, 1988.
13. Jordan, S.R., Pabo, C.O. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor–operator interactions. Science 242:893–898, 1988.
14. Otwinowski, Z., Schevitz, R.W., Zhang, R.-G., Lawson, C.L., *Joachimiak*, A., *Marmorstein*, R.Q., Luisi, B.F., Sigler, P.B. Crystal structure of *trp* repressor/operator complex at atomic resolution. Nature (London) 335:321–329, 1988.
15. Seeman, N.C., Rosenberg, J.M., Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. U.S.A. 73:804–808, 1976.
16. Müller-Hill, B. *Lac* repressor and *lac* operator. Prog. Biophys. Molec. Biol. 30:227–252, 1975.
17. Weber, K., Geisler, N. *Lac* repressor fragments produced in vivo and in vitro: An approach to the understanding of the interaction of repressor and DNA. In: "The Operon," Miller, J., Reznikoff, W. eds. New York: Cold Spring Harbor Laboratory, Cold Spring Harbor, 1978: 155–175.
18. Ogata, R.T., Gilbert, W. DNA-binding site of *lac* repressor probed by dimethylsulfate methylation of *lac* operator. J. Mol. Biol. 132:709–728, 1979.
19. Zuiderweg, E.R.P., Kaptein, R., Wüthrich, K. Secondary structure of the *lac* repressor DNA-binding domain by two-dimensional ^{1}H nuclear magnetic resonance in solution. Proc. Natl. Acad. Sci. U.S.A. 80:5837–5841, 1983.
20. Zuiderweg, E.R.P., Scheek, R.M., Kaptein, R.: Two-dimensional ^{1}H-nmr studies on the *lac* repressor DNA binding domain: Further resonance assignments and identification of nuclear overhauser enhancements. Biopolymers 24: 2257–2277, 1985.
21. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., van Gunsteren, W.F. A protein structure from nuclear magnetic resonance data: *lac* repressor headpiece. J. Mol. Biol. 182:179–182, 1985.
22. Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., Van Gunsteren, W.F., Kaptein, R. Determination of protein structures from nuclear magnetic resonance data using a restrained molecular dynamics approach: The *lac* repressor DNA binding domain. Biochimie 67:707–715, 1985.
23. de Vlieg, J., Boelens, R., Scheek, R.M., Kaptein, R., van Gunsteren, W.F. Restrained molecular dynamics procedure for protein tertiary structure determination from NMR data: A *lac* repressor headpiece structure based on information on J-coupling and from presence and absence of NOE's. Isr. J. Chem. 27:181–188, 1986.
24. de Vlieg, J., Scheek, R.M., Gunsteren, W.F., Berendsen, H.J.C., Kaptein, R., Thomasson, J. Combined procedure of distance geometry and restrained molecular dynamics techniques for protein structure determination from nuclear magnetic resonance data: Application to the DNA binding domain of *lac* repressor from *Escherichia Coli*. Proteins 3:209–218, 1988.
25. de Vlieg, J., Kaptein, R., van Gunsteren, W.F. Molecular dynamics simulation in aqueous solution using NMR data: Application to the DNA binding domain of *lac* repressor from *Escherichia Coli*. In preparation.
26. Boelens, R., Scheek, R.M., Lamerichs, R.M.N.J., de Vlieg, J., van Boom, J.H., Kaptein, R. A two-dimensional NMR study of the complex of *lac* repressor headpiece with a 14 base pair *lac* operator fragment. In: "DNA Ligand Interactions," Guschlbauer, W., Saenger, W. eds. New York: Plenum, 1987: 191–215.
27. Boelens, R., Scheek, R.M., van Boom, J.H., Kaptein, R. Complex of *lac* repressor headpiece with a 14 base pair *lac* operator fragment studied by two-dimensional nuclear magnetic resonance. J. Mol. Biol. 193:213–216, 1987.
28. Boelens, R., Lamerichs, R.M.N.J., Rullmann, J.A.C., van Boom, J.H., Kaptein, R. The interaction of *lac* repressor headpiece and its operator: A NMR view. Protein Sequences Data Anal., in press.
29. Rullmann, J.A.C., Boelens, R., Kaptein, R. NMR based docking studies of *lac* repressor headpiece on a *lac* operator fragment. UCLA Symp., in press.
30. Hockney, R.W., Eastwood, J.W. "Computer Simulation Using Particles." New York: McGraw-Hill, 1981.
31. van Gunsteren, W.F., Berendsen, H.J.C. Molecular dynamics simulations: techniques and applications to proteins. In: "Molecular Dynamics and Protein Structure," Hermans, J. ed. Western Springs, Ill: Polycrystal Bookservice, 1985: 5–14.
32. Ciccotti, G., Hoover, W.G., eds. Molecular-dynamics simulation of statistical-mechanical systems. Proceedings of the International School of Physics "Enrico Fermi," course 97, North-Holland, Amsterdam 1986: 1–610.
33. McCammon, J.A.C., Harvey, S.C. "Dynamics of Proteins and Nucleic Acids." London: Cambridge University Press, 1987: 1–234.
34. Berendsen, H.J.C., van Gunsteren, W.F., Egberts, E., de Vlieg, J. Dynamic simulation of complex molecular systems. In "American Chemical Society Symposium Series 353, Supercomputer Research in Chemistry and Chemical Engineering," Jensen, K.F., Truhlar, D.G., eds. Washington D.C.: Amer. Chem. Soc., 1987: 106–122.
35. van Gunsteren, W.F., Boelens, R., Kaptein, R., Scheek, R.M., Zuiderweg, E.R.P. An improved restrained molecular dynamics technique to obtain protein tertiary structure from nuclear magnetic resonance data. In: "Molecular Dynamics and Protein Structure," Hermans, J., ed. Western Springs, Ill: Polycrystal Bookservice, 1985: 92–99.
36. Arnott, S., Hukins, D.W.L. Optimised parameters for A-DNA and B-DNA. Biochem. Biophys. Res. Commun. 47: 1504–1509, 1972.
37. van Gunsteren, W.F., Berendsen, H.J.C., Geurtsen, R.G., Zwinderman, H.R.J. A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: Comparison with experimental two-dimensional NMR data. Ann. N.Y. Acad. Sci. 482:287–303, 1986.
38. Wüthrich, K., Billeter, M., Braun, W. Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J. Mol. Biol. 169:949–961, 1983.
39. Billeter, M., Havel, T.F., Wüthrich, K. The ellipsoid algorithm as a method for the determination of polypeptide conformations from experimental distance constraints and energy minimization. J. Comp. Chem. 8:132–141, 1987.
40. Billeter, M., Havel, T.F., Kuntz, I.D. A new approach to the problem of docking two molecules: The ellipsoid algorithm. Biopolymers 26:777–793, 1987.
41. van Gunsteren, W.F., Berendsen, H.J.C. GROningen MOlecular Simulation (GROMOS) library manual. Biomos B.V., Nijenborgh 16, Groningen, The Netherlands, 1987: 1–229.
42. Hermans, J., Berendsen, H.J.C., van Gunsteren, W.F., Postma, J.P.M. A consistent emperical potential for water-protein interactions. Biopolymers 23:1513–1518, 1984.
43. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1977.
44. van Gunsteren, W.F., Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraints dynamics. Mol. Phys. 34:1311–1327, 1977.
45. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J. Interaction models for water in relation to protein hydration. In: "Intermolecular Forces," Pullman, B., ed. Dordrecht: Reidel, 1981:331–342.
46. Berendsen, H.J.C., van Gunsteren, W.F., Zwinderman, H.R.J., Geurtsen, R.G. Simulations of proteins in water. Ann. N.Y. Acad. Sci. 482:269–285, 1986.
47. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
48. Stob, S., Scheek, R.M., Boelens, R., Kaptein, R. Photo-CIDNAP study of the interaction between *lac* repressor headpiece and *lac* operator DNA. FEBS Lett. 239:99–104, 1988.
49. Baker, E.N., Hubbard, R.E. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44:97–179, 1984.
50. Lamerichs, R.M.N.J., Boelens, R., Kaptein, R. In preparation.
51. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S.,

Itakura, K., Dickerson, R.E. Structure of a B-DNA dodecamer: Conformation and dynamics. Proc. Natl. Acad. Sci. U.S.A. 78:2179–2183, 1981.
52. Hol, W.G.J., van Duijnen, P.T., Berendsen, H.J.C. The α-helix dipole and the properties of proteins. Nature (London) 273:443–446, 1978.
53. Miller, J.H. Amino acid replacements in the DNA binding domain of the *Escherichia coli lac* repressor. J. Mol. Biol. 180:205–212, 1984.
54. Scheek, R.M., Zuiderweg, E.R.P., Klappe, K.J.M., van Boom, J.H., Kaptein, R., Rüterjans, H., Beyreuther, K. *Lac* repressor headpiece binds specifically to half of the *lac* operator. A proton nuclear magnetic resonance study. Biochemistry 22:228–235, 1983.
55. von Wilken-Bergmann, B., Müller-Hill, B. Sequence of *galR* gene indicates a common evolutionary origin of *lac* and *gal* repressor *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 79:2427–2431, 1982.
56. Caruthers, M.H. Deciphering the protein-DNA recognition code. Acct. Chem. Res. 13:155–160, 1980.
57. Lehming, N., Sartorius, J., Niemöller, M., Genenger, G., von Wilcken-Bergmann, B., Müller-Hill, B. The interaction of the recognition helix of *lac* repressor with *lac* operator. EMBO J. 6:3145–3153, 1987.
58. Richter, P.H., Eigen, M. Diffusion controlled reaction rates in spheroidal geometry: Application to repressor-operator association and membrane bound enzymes. Biophys. Chem. 2:255–263, 1974.
59. Berg, O.C., Winter, R.B., von Hippel, P.H. Diffusion-driven mechanisms of protein-translocation on nucleic acids: Models and theory. Biochemistry 20:6929–6948, 1981.
60. Winter, R.B., Berg, O.G., von Hippel, P.H. Diffusion-driven mechanisms of protein-translocation on nucleic acids: The Escherichia coli *lac* repressor-operator interaction: Kinetic measurements and conclusions. Biochemistry 20:6961–6977, 1981.
61. Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., Winter, G. Hydrogen bonding and biological specificity analysed by protein engineering. Nature (London) 314:235–238, 1985.
62. Ebright, R.H. Evidence for a contact between glutamine-18 of lac repressor and base pair 7 of lac operator. Proc. Natl. Acad. Sci. U.S.A. 83:303–307, 1986.
63. Bass, S., Sorrells, V., Yonderian, P. Mutant *trp* repressors with new DNA-binding specificities. Science 241:240–245, 1988.
64. van Gunsteren, W.F., Weiner, P.K., eds. "Computer Simulation of Biomolecular Systems." Leiden, The Netherlands: Escom, 1989.