Cell, Vol. 85, 1057–1065, June 28, 1996, Copyright ©1996 by Cell Press

Hydration and DNA Recognition by Homeodomains

Martin Billeter, Peter Güntert, Peter Luginbühl, and Kurt Wüthrich
Institut für Molekularbiologie und Biophysik
Eidgenössische Technische Hochschule-Hönggerberg
CH-8093 Zürich
Switzerland

Summary

A 2-nanosecond molecular dynamics (MD) simulation of an Antennapedia homeodomain–DNA complex in explicit solvent water at ambient temperature and pressure was performed to supplement experimental nuclear magnetic resonance (NMR) data on the structure and dynamics of this complex. In addition to direct protein–DNA contacts, the MD trajectory attributes an essential role for specific DNA recognition to hydration water molecules that mediate intermolecular contacts. The simulation provides a detailed description of the pathways of hydration water molecules exchanging in and out of the protein–DNA interface and indicates that the residence times of these "interior" waters are on the nanosecond time scale, near the lower end of the range determined by NMR.

Introduction

The two prevalent techniques for atomic resolution structure determination of biological macromolecules, X-ray diffraction in crystals (Drenth, 1994) and nuclear magnetic resonance (NMR) spectroscopy in solution (Wüthrich, 1986, 1995), also provide information on hydration water molecules associated with the macromolecules. Both in crystals and in solution, the hydration water is predominantly on the macromolecular surface, but a small number of water molecules may be located in the interior of globular macromolecular structures. These interior water molecules can generally be assigned structural roles in the macromolecular architecture (e.g., Thanki et al., 1988; Zhang and Matthews, 1994).

A special situation is encountered in protein–DNA complexes, where hydration water in the protein–DNA interface is observed in the functional form of the macromolecules (e.g., Otwinowski et al., 1988; Billeter et al., 1993; Qian et al., 1993a; Hirsch and Aggarwal, 1995; Wilson et al., 1995). These water molecules may have structural or functional roles, or both (Wüthrich, 1993), and are therefore an ideal ground for detailed investigations on possible roles of solvent water in biological recognition processes. In this paper we describe the combined results of NMR studies and long-time molecular dynamics (MD) simulations in explicit water of a solvated Antennapedia homeodomain (Antp HD) complex with an operator DNA duplex which represents the BS2 binding site.

X-ray crystallography and NMR provide complementary data on hydration water molecules. X-ray diffraction in single crystals can yield precise information on the location of individual water molecules, but the technique records only the population of a hydration site and is insensitive to rate processes, such as exchange of water molecules in and out of the site (e.g., Frey, 1994; Karplus and Faerman, 1994). In contrast, NMR is sensitive to such rate processes and can within certain ranges (see below) be used for quantitative measurements of their frequencies (Otting et al., 1991a). Lifetimes shorter than 1 nanosecond (ns) are typically observed on the surface of proteins and in the major groove of DNA, while lifetimes longer than 1 ns have been observed in the interior of globular proteins, in cavities near the protein surface, in the minor groove of DNA duplexes, and in the interface of protein–DNA complexes (for a recent review, see Billeter, 1995). Residence times of hydration water molecules longer than milliseconds have so far not been observed in any protein or protein–DNA complex.

NMR studies of the rate processes with interior hydration water are incomplete in the sense that only bounds on frequency ranges rather than precise frequency values have been reported. For example, for solvent-inaccessible interior hydration water molecules in the outstandingly stable protein basic pancreatic trypsin inhibitor (BPTI), an upper bound on the lifetime with respect to exchange in and out of a given hydration site was determined with different techniques (Otting et al., 1991b; Denisov and Halle, 1995; Dötsch and Wider, 1995), and all the data are in agreement with an upper limit of 200 μs. The lower bound is about 1 ns (Otting et al., 1991a), so that the actual values may vary over several orders of magnitude. The presently described long-time MD simulation complements the NMR data in two ways: it provides an indication of whether the actual lifetimes of the hydration water molecules in the protein–DNA interface are nearer to the upper or lower limit of the experimentally defined range, and it supplements the measurement of exchange rates with a description of the associated reaction pathways. The MD simulation is sufficiently long to observe specific events, such as conformational changes of amino acid side chains and penetration of water molecules into the protein–DNA interface. In the Discussion section, these results will be evaluated in light of the NMR structure and of genetic and biochemical data.

Results

The Aqueous Solution of the Antp HD–DNA Complex Used for the MD Simulation

The system selected for the present investigation is closely related to the aqueous solution of the Antp HD–DNA complex used previously for the NMR structure determination (Otting et al., 1990; Billeter et al., 1993; Qian et al., 1993b), which consisted of a 68-residue polypeptide containing the mutant Antp(C39S) HD in positions 1–60 and a 14–base pair DNA duplex representing the BS2 binding site, with the α-strand sequence d-$G_1A_2A_3A_4G_5C_6C_7A_8T_9T_{10}A_{11}G_{12}A_{13}G_{14}$ (the two DNA

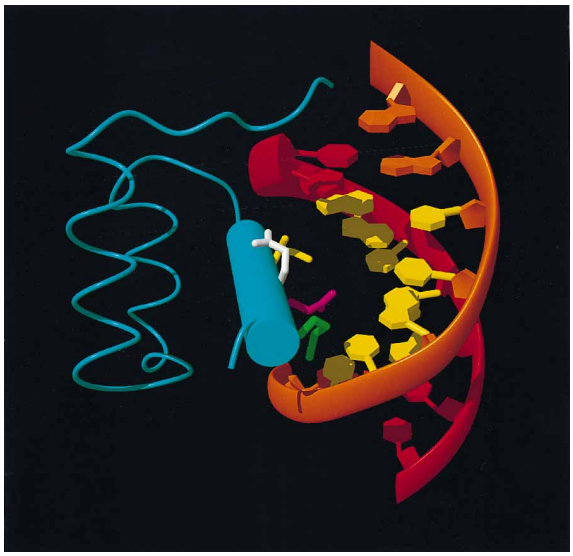

Figure 1. View of the DNA Complex with the Antennapedia Homeodomain That Was Studied in the Present Molecular Dynamics Simulation

In this schematic drawing of the macromolecular components of the system, the DNA backbone is represented by ribbons, with the α strand colored orange and the β strand red. The central 9 base pairs of the 14–base pair DNA duplex used in the NMR structure determination to represent the BS2 binding site are retained (see text), with the α-strand sequence d-$A_4G_5C_6C_7A_8T_9T_{10}A_{11}G_{12}$. Bases with contacts to the protein are shown in yellow. For the homeodomain the residues 3–56 are retained. The polypeptide backbone is colored cyan, and the recognition helix with residues 43–54 is shown as a cylinder carrying heavy atom representations of those side chains that contact DNA bases: Ile-47 (yellow), Gln-50 (pink), Asn-51 (gray), Met-54 (green). This drawing was prepared with the atom coordinates of the complex (Billeter et al., 1993) that was used at the start of the MD simulation. This drawing and Figures 2–5 were prepared with the program MOLMOL (Koradi et al., 1996).

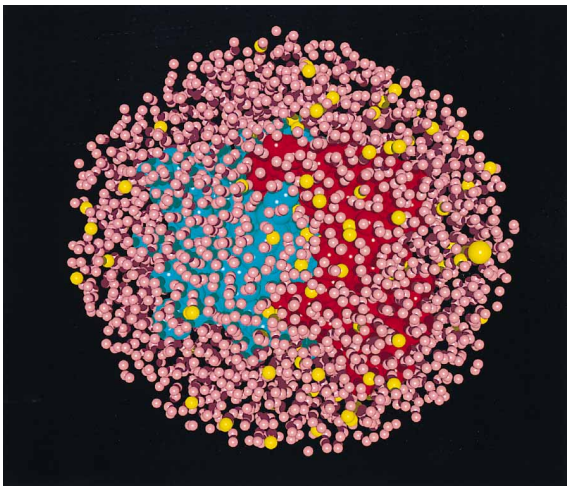

Figure 2. The Hydrated Antp HD–DNA Complex

In the starting configuration of the complete system used for the MD simulation, the Antp HD-DNA complex of Figure 1 is surrounded by 2714 water molecules, with heavy atoms represented by spheres. The protein is colored in cyan, and the two DNA strands are red. The yellow spheres identify those water molecules that enter the protein–DNA interface and are located near Gln-50 at some instant during the simulation. The large yellow sphere on the right represents the starting position of the water molecule characterized in Figures 5, 7D, and 7E. The oxygen atoms of the other waters are colored magenta. To afford a view of interior parts, the radius of the water molecules is smaller than what would correspond to the actual relative size.

strands are denoted α and β; the base pairs are identified by their position in the 14–base pair duplex, with numeration of the α strand in the direction 5′-3′). In keeping with the main focus on protein–DNA contacts and hydration in the major groove, we eliminated those parts of the polypeptide and the DNA duplex that are neither essential for the global architecture nor involved in major groove contacts (Billeter et al., 1993). The resulting complex, with the polypeptide fragment 3–56 and the central nine base pairs 4–12 (Figure 1), was placed in an ellipsoid-shaped water bath in such a way that the minimal distance from the surface of the complex to the outer boundary of the water bath was 6 Å (Figure 2). The atomic coordinates for the starting structure of the complex were taken from the NMR structure (Billeter et al., 1993). With this "truncated" complex, and using the new program OPAL (P. Luginbühl, P. Güntert, M. Billeter and K. Wüthrich, unpublished), which is optimized for this type of calculation on supercomputers (for details, see Experimental Procedures), a 2 ns MD simulation could be performed with reasonable use of computing time.

Survey of the 2-Nanosecond MD Trajectory

Figure 1 shows that the DNA-bound protein is folded into a globular structure consisting of three helical regions connected by short loops and an N-terminal arm interacting with the DNA minor groove (Otting et al., 1990; Billeter et al., 1993). In the interface between the recognition helix (third helix of the homeodomain, shown as a cyan-colored cylinder in Figure 1) and the central DNA base pairs with the α-strand sequence d-$C_6C_7A_8T_9T_{10}$ (yellow bases in Figure 1), the amino acid side chains of Ile-47 (yellow), Gln-50 (pink), Asn-51 (gray), and Met-54 (green) contact DNA bases. In the MD simulation, which started from the configuration of the complete protein–DNA–water system shown in Figure 2, equilibration by adjustment of the macromolecular structure to the new environment characterized by the water bath and the force field used (see Experimental Procedures) was reached within 100 picoseconds (ps). The structure of the complex taken after 100 ps differs from the starting structure by a root mean square deviation (rmsd) of 2.3 Å for all heavy atom positions and by 1.8 Å for the heavy atom positions of residues 43–54 in the recognition helix and the central five base pairs 6–10. Motions of this interface during the 1.9 ns trajectory following the equilibration period are illustrated in Figure 3, which shows a superposition of 20 snapshots taken at intervals of 100 ps. The fluctuations about the mean positions of the heavy atoms exceed 1.0 Å in only a few instances. The largest motions occur for the C-terminal turn of the recognition helix and for base pair 6. Although no experimental constraints were imposed during the simulation, only 60 out of the 1140 distance constraints that were used for the determination of the NMR structure of this complex (Billeter et al., 1993) are on a time average

Figure 3. Superposition of 20 Snapshots Taken at Intervals of 100 ps during the 2 ns MD Simulation

The recognition helix is shown in cyan, with the side chains of Ile-47 (yellow), Gln-50 (pink), Asn-51 (gray), and Met-54 (green), and the bases of the five central DNA base pairs (α-strand sequence d-$C_6C_7A_8T_9T_{10}$) with colors ranging from yellow to blue. The view is perpendicular to the DNA double helix axis, approximately along the axis of the recognition helix.

violated by more than 1 Å; six of these are intermolecular protein–DNA distance constraints. For the following discussions on protein–DNA interactions we assume that the protein–DNA complex is long-lived compared to the simulation period of 2 ns (Otting et al., 1990) and that the overall relative orientation of the polypeptide backbone in the recognition helix and the d-$C_6C_7A_8T_9T_{10}$ region of the DNA is fixed. Superimposed onto this stable scaffold, the individual amino acid side chains undergo larger-amplitude motions, which lead to exchange between different contacts with the DNA. In Figure 3 this is clearly visible for Gln-50, a residue for which a key role in the DNA recognition has been established by genetic and biochemical experiments (e.g., Gehring et al., 1994). The MD simulation also indicates that the population of individual amino acid side chain–DNA contacts may be as high as 70%, with maximal lifetimes for individual interactions of up to 280 ps (Table 1).

In the MD trajectory, different individual water molecules exhibit different behavior depending on their interactions with the macromolecules. Most water molecules remain in the bulk solvent, with occasional short-lived contacts to the surface of the protein–DNA complex. To obtain reference values for the interpretation of the data on Gln-50 in the protein–DNA interface, we performed an analysis of the residence times of hydration water molecules located near the amide groups of Gln-22, Gln-28, Asn-39, and Gln-50, which are all located on the surface of the complex (Billeter et al., 1993). Out of a total number of 1604 water contacts with these four amino acid side chains, 53% have residence life times between 1 and 5 ps, and 98.9% are shorter than 10 ps. Compared with experimental data measured at 4°C with different proteins (Otting et al., 1991a; Wüthrich, 1993) these residence times of surface hydration water molecules, which were calculated at a temperature of 27°C, are about 10-fold shorter.

Table 1. Populations and Maximal Lifetimes of Selected Direct Protein–DNA Contacts in the Antp HD–DNA Complex Observed during the 2 ns MD Simulation[a]

Protein[b]	DNA[c]	P_{rel} (%)[d]	t_{max} (ps)[e]
Ile-7			
Cγ2	A_{10} C8 (β)	38	30
Cγ2	A_9 C8 (β)	73	177
Cγ1	T_8 C7 (β)	28	16
Cδ1	T_8 C7 (β)	56	36
Gln-50			
Oε1	C_6 N4 (α)	34	216
Oε1	C_7 N4 (α)	60	281
Nε2	C_7 N4 (α)	64	208
Nε2	A_8 N6 (α)	26	38
Nε2	T_8 O4 (β)	38	111
Asn-51			
Nδ2	A_{10} N6 (β)	72	165
Oδ1	A_{10} N7 (β)	49	52
Nδ2	A_{10} N7 (β)	72	65
Met-54			
Cε	T_9 C7 (α)	35	16

[a]Direct contacts are interatomic distances shorter than 4.0 Å either between different carbon atoms (hydrophobic interactions for Ile-47 and Met-54) or between oxygen and nitrogen atoms (hydrogen bonds for Gln-50 and Asn-51). Only contacts with a population exceeding 25% out of the total of 1900 snapshots taken at intervals of 1 ps during the MD simulation from 100 to 2000 ps are listed.

[b]The residue type, the sequence position, and the atom considered are identified with the standard notation. The interactions with Ile-47 and Met-54 are hydrophobic; those with Gln-50 and Asn-51 are hydrogen bonds.

[c]The subscripts refer to the base pair numeration in the 14-mer BS2 sequence used for the NMR structure determination of the Antp HD–DNA complex; the atom identification is according to the standard numeration of heavy atoms in the bases; location in the α or β strand is indicated in parentheses.

[d]Relative population, expressed as the percentage of the 1900 snapshots in which this contact is observed.

[e]Maximal lifetime, taken as the longest continuous set of snapshots during which this interaction is present.

After the equilibration of the starting system (Figure 2) during the initial 100 ps of MD simulation, the protein–DNA interface in the major groove contained several hydration water molecules in all instances. Overall, during the 2 ns of MD simulation ~5% of all water molecules penetrated into the protein–DNA interface. The snapshot of Figure 4 shows a structure with six water molecules in the interface near the side chains of Ile-47, Gln-50, Asn-51, and Met-54. These interior water molecules remain only for a limited time before leaving the interface, either along the same way they came in or by a different pathway. Lifetimes for water molecules in the positions indicated in Figure 4 may be as long as 1300 ps. Figure 5 illustrates the pathway of a water molecule that was initially located well outside of the complex (big yellow sphere in Figure 2), then penetrated into the interface, spent time near the amino acid side chains of Ile-47 and Gln-50, and left the interface again after ~600 ps. The residence time is thus about 100 times longer than for surface hydration waters near chemically identical side chains. An additional important observation on the hydration of the complex is shown in Figure 2: in the starting configuration, the water molecules that enter the protein–DNA interface during the 2 ns of MD simulation

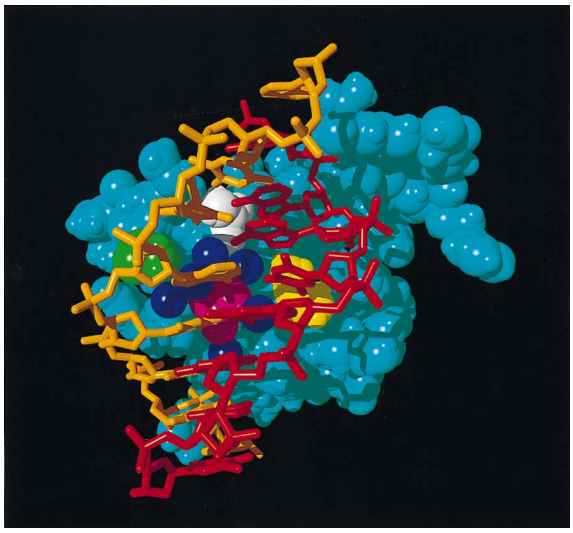

Figure 4. Snapshot of the Protein–DNA Interface after 1148 ps of MD Simulation of the Antp HD–DNA Complex

The protein is represented by a space-filling model, with all atoms in cyan except for the side chains of Ile-47 (yellow), Gln-50 (pink), Asn-51 (gray), and Met-54 (green). The DNA is shown as a wireframe model with the same coloring as the backbone ribbons in Figure 1. The positions of interfacial hydration water molecules are given by dark blue spheres. The water molecule characterized in Figures 7D and 7E is located between Ile-47 and Gln-50.

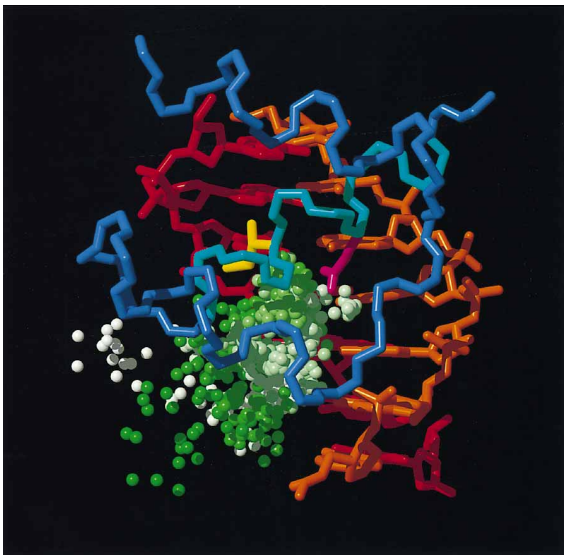

Figure 5. Diffusion Pathway of the Water Molecule Identified by a Big Yellow Sphere in Figure 2 during the Time Period from 840 to 1500 ps of the MD Simulation

The DNA is shown as an orange and red wire frame model. For the protein, the backbone is shown as a dark blue tube, except for the cyan-colored recognition helix, the yellow side chain of Ile-47, and the pink side chain of Gln-50. The water molecule is represented by a series of spheres with variable color as a function of simulation time, i.e., from green at 840 ps to white at 1500 ps, with a time resolution of one sphere per picosecond.

are quite uniformly distributed over the water bath, showing that the length of the simulation in the present study was adequately chosen. It is intriguing to rationalize that all these yellow water molecules are at some instant involved in specific DNA recognition by the homeodomain.

In the remainder of this paper, we shall first give a more quantitative description of the MD trajectories and thereby focus primarily on the amino acid side chains in the major DNA groove (Figure 1) and on interfacial hydration. The results from the MD simulation will then be evaluated in light of the experimental data available on the same system.

The Homeodomain Recognition Helix

As an illustration for the state of the recognition helix during the 2 ns trajectory, Figure 6 shows variations in the dihedral angles of two of the helical residues, Ile-47 and Gln-50, as a function of the simulation time. Only minor fluctuations are observed for the backbone dihedral angles φ and ψ, which are of the same order as those of the "rigid" angle ω. The hydrophobic side chain of Ile-47 remains also mostly rigid, with the exception of two 120° jumps of the value of χ^{21}, which move the δ-methyl group along the DNA major groove. The side chain of Gln-50 undergoes more frequent structural changes. Both the χ^1 and the χ^3 angles flip several times by $\geq$ 90° during the 2 ns time range reported in Figure 6. The variations of χ^1 are reflected in the two groups of different conformations of this side chain seen in Figure 3. These changes affect the spatial locations of hydrogen bond donor and acceptor groups of Gln-50 and thus lead to time-dependent variations of the network of protein–DNA interactions.

It is on the basis of this type of data that we arrived at the conclusion, previously mentioned, to consider the backbone of the recognition helix as part of an invariant scaffold that supports the more flexible atom groups that are directly involved in intermolecular contacts. The nearly static nature of this scaffold, which includes also the central base pairs of the DNA with the α-strand sequence d-$C_6C_7A_8T_9T_{10}$, is further documented by the time course of the rmsd between the mean structure of the trajectory and the 2000 snapshots taken at intervals of 1 ps (Figure 7A). The mean structure used as a reference was obtained by optimally superimposing the snapshots from 101 to 2000 ps onto the equilibrated structure at 100 ps and then averaging the coordinates. This rmsd only rarely exceeds a value of 1.0 Å, and from inspection of Figure 3 it is apparent that the major contributions to this averaged quantity come from the peripheral base C_6 and the last turn of the recognition helix—emphasizing that for the remainder of the structural entity considered, the time variations are indeed very small.

Direct Contacts between Amino Acid Side Chains of the Recognition Helix and DNA Bases

Direct protein–DNA interactions include hydrogen bonds from the side chains of Gln-50 and Asn-51 to DNA bases, as well as hydrophobic contacts between the side chain of Ile-47 or Met-54 and DNA bases (Figure 8; Table 1). Gln-50 forms hydrogen bonds with four bases in three

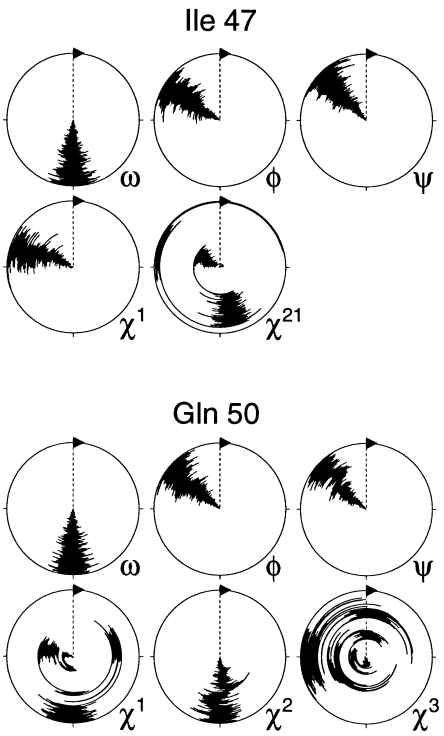

Figure 6. Variation of the Dihedral Angles of Ile-47 and Gln-50 in the Antp HD–DNA Complex during the 2 ns MD Simulation

Dihedral angle values are with respect to the vertical dotted line, with positive angles in the direction of the arrow. The time axis runs from 0 ns in the center of each circle to 2 ns at the periphery. ω, φ, and ψ are the common polypeptide backbone dihedral angles; χ^1–χ^3 and χ^{21}, the dihedral angles about the side chain single bonds.

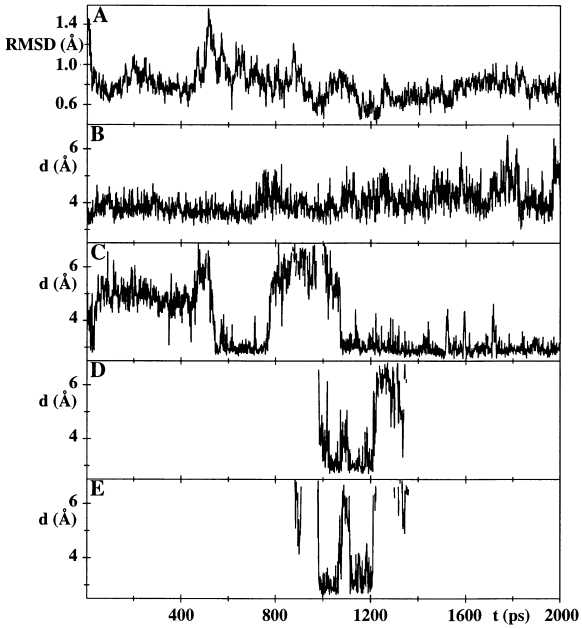

Figure 7. Variation of Selected Atom Positions and Intermolecular Distances during the 2 ns MD Simulation

(A) Root mean square deviation (rmsd) with respect to the mean structure calculated for the backbone heavy atoms of residues 43–54 in the recognition helix and the central five DNA base pairs (α-strand sequence d-$C_6C_7A_8T_9T_{10}$).

(B) Distance (d) between the δ-methyl carbon of Ile-47 and the methyl carbon of T_8.

(C) Distance between the ϵ-oxygen of Gln-50 and the nitrogen at position 4 of C_7.

(D) Distance between the ϵ-nitrogen of Gln-50 and the oxygen atom of the water molecule represented as a big yellow sphere in Figure 2.

(E) Distance between the oxygen atom at position 4 of T_8 and the oxygen atom of the same water molecule as in (D). The distances plotted in (D) and (E) are always >10 Å during the time periods 1–800 and 1400–2000 ps.

different base pairs, Ile-47 contacts three bases of the β strand, and Asn-51 and Met-54 interact with one base each. In Table 1 these interactions are characterized by their populations, P_{rel}, and their maximal lifetimes, t_{max} (see footnotes to Table 1 for the precise definitions). Hydrophobic interactions occur with comparable populations as the hydrogen bonds, but they are typically very short lived. The only exception is that the contact of the γ-methyl group of Ile-47 with the C8 atom of A_9 has a t_{max} value of almost 200 ps. In contrast, more than half of the hydrogen bonds have t_{max} values longer than 100 ps, with a maximum of almost 300 ps for the interaction between the side chain oxygen of Gln-50 and the N4 atom of C_7.

The time courses of two of the contacts in Table 1 during the 2 ns MD simulation are shown in Figure 7. The hydrophobic interaction between the δ-methyl group of Ile-47 and the methyl group of T_8 (Figure 7B) is an example of a highly populated contact with short t_{max}. For almost the entire duration of the simulation, this contact distance is between 3.5 and 5.5 Å, i.e., it represents a rather rigid local structure. The hydrogen bond between the side chain of Gln-50 and the N4 atom of C_7, which is a highly populated contact with rather long t_{max}, shows motions with amplitudes of several angstroms and periods of several hundred picoseconds (Figure 7C), which represent more pronounced local flexibility than in the example of Figure 7B.

Analysis of the data in Table 1 does not lead to a clear-cut correlation between P_{rel} and t_{max}: while all but one of the contacts with P_{rel} > 60% have t_{max} values longer than 150 ps, there are two interactions with P_{rel} < 40% and t_{max} > 100 ps. Possible P_{rel} − t_{max} correlations for direct protein–DNA contacts are further investigated in Figure 9A. While many of these contacts are "nonspecific" in the sense that they are both very short lived and have low populations (lower left corner of Figure 9A), other "specific" contacts are formed repeatedly and thus reach a high population. For comparison, data are also given for salt bridges between arginines and lysines of the recognition helix and DNA phosphate groups. The observed trend to longer lifetimes is consistent with the notion that these salt bridges undergo a more stable type of interaction than the amino acid side chain–DNA base contacts.

Hydration Water and Water-Mediated Protein–DNA Interactions

In the MD simulation, the water–water hydrogen bonds are very short lived, in the range of the time resolution given by the sampling of the trajectory at 1 ps intervals. Water molecules that interact with the surface of the

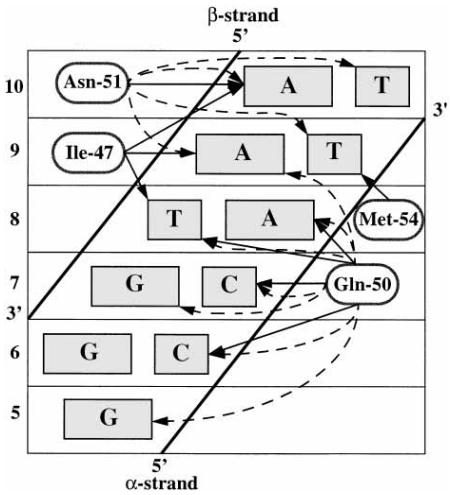

Figure 8. Contact Map with Direct and Water-Mediated Protein–DNA Interactions in the Antp HD–DNA Interface

Direct interactions (solid arrows) are those identified in Table 1, where multiple interactions between a given amino acid–base combination are represented by a single arrow. Each of the water-mediated hydrogen bonds (dashed arrows) between Gln-50 or Asn-51 and a DNA base includes two simultaneous contacts shorter than 4.0 Å, one between the nitrogen or oxygen atom of the amino acid side chain and a water molecule, and the other one between the same water molecule and either a nitrogen or an oxygen atom of a DNA base. Water-mediated interactions are only included if their relative population, P_{rel} (see footnotes to Table 1), exceeds 25%. Note that all the interactions shown here do not need to be present simultaneously, since the populations are typically well below 100%.

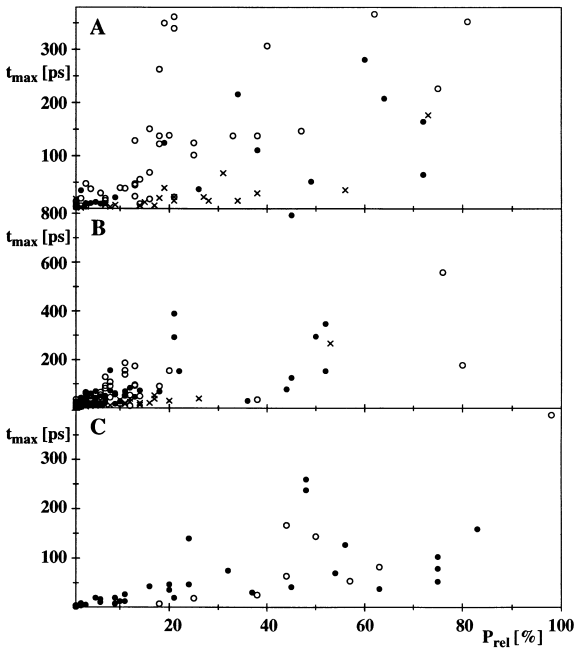

Figure 9. Correlation of Maximal Lifetimes and Relative Populations for Selected Short Interatomic Contacts in the Antp HD–DNA–Water System

The maximal lifetime, t_{max}, and the relative population, P_{rel}, are defined in the footnotes to Table 1. (A) Direct intermolecular contacts shorter than 4.0 Å between pairs of heavy atoms in the Antp HD–DNA complex, where one atom is located in the recognition helix and the other one in a base of the DNA. Filled circles, hydrogen bonds; crosses, hydrophobic interactions. For comparison, salt bridge contacts of DNA phosphates with Arg and Lys of the Antp HD recognition helix are represented by empty circles. (B) Contacts shorter than 4.0 Å between any of the hydration water molecules and any of the heavy atoms of the following residues: crosses, Ile-47; closed circles, Gln-50; open circles, Asn-51. (C) Water-mediated hydrogen bonds between Gln-50 (closed circles) or Asn-51 (open circles) and DNA bases as defined in the caption to Figure 8. For the evaluation of P_{rel}, the contributions from contacts between a given pair of atoms that are mediated by different hydration water molecules were added up.

complex have short residence times (see above), which is in good qualitative agreement with both the results of a previous MD simulation of surface hydration of the small protein BPTI (Brunne et al., 1993) and NMR measurements with oxytocin and BPTI (Otting et al., 1991a). Superimposed onto this background of nonspecific water–water and macromolecular surface–water contacts there are longer-lived interactions with water molecules that penetrate the protein–DNA interface, which are further analyzed in the following.

As a complement to Figure 5, which visualizes the pathway of a specified water molecule (large yellow sphere in Figure 2), Figures 7D and 7E show the time evolution of hydrogen bonds formed simultaneously by a given water molecule with atom groups of both the protein and the DNA. Similar to the direct protein–DNA contacts, rapid fluctuations on the picosecond time scale are superimposed on slower motions in the 100 ps range. A structure where a water molecule mediates a protein–DNA interaction is manifested in a snapshot taken after 1061 ps, where the two distances from the water oxygen to the side chain nitrogen of Gln-50 and to the oxygen at position 4 of T_8 are 3.1 and 2.8 Å, respectively. Figure 9B correlates maximal lifetimes and corresponding populations of contacts between water molecules and selected amino acid side chains in the protein–DNA interface (Figure 1). It is seen that "specific" contacts have typical populations of 20%–50%, and typical maximal lifetimes of 100–400 ps, with extreme values of $P_{rel} = 80\%$ and $t_{max} = 800$ ps. Figure 9C presents a corresponding statistical analysis for water molecules that mediate protein–DNA hydrogen bonds. For water-mediated protein–DNA interactions the values for both the maximal lifetimes and the populations are comparable to those for direct hydrogen bonding interactions between the macromolecules, but the values for P_{rel} and t_{max} are higher than for direct hydrophobic contacts (Figure 9A). The contact map of Figure 8 shows that water-mediated hydrogen bonds connect Gln-50 and Asn-51 to the same DNA bases as the direct contacts but also extend the interaction network to additional bases. Clearly, all the interactions indicated in Figure 8 do not occur simultaneously, but they rather form a network that is fluctuating with time. For clarity, it should also be added that the observation of outstandingly high populations in Figure 9, which correspond to up to 1600 out of the total of 1900 snapshots, and rather short maximal lifetimes reflects also movements of water molecules within the protein–DNA interface, rather than exclusively multiple entries of individual water molecules into the interface.

Discussion

Specific interactions between different macromolecules, such as the recognition of specific DNA sequences by proteins, rely largely on the complementarity of the interacting surfaces. In this context, complementarity of two molecules is not restricted to steric matching, but more generally includes favorable intermolecular interactions that may result from a particular surface distribution of charges, matching regions of different polarity, and distribution of donor and acceptor groups for hydrogen bonds. The assumption of rigid intermolecular interfaces in a complex, as described by models referred to as "lock-and-key," or by the final result of an "induced fit" process, may often not take proper account of entropy considerations. The increase in free energy that would result from complete immobilization of amino acid side chains on the protein surface, upon interaction of this surface with another molecule, can be a high price to pay for ideal geometry of the intermolecular interactions. Indications of residual mobility of amino acid side chains in intermolecular interfaces emerge from the ability of certain proteins to adapt to a range of different interaction partners, as has been demonstrated in protein–DNA complexes (e.g., Gewirth and Sigler, 1995) as well as in protein–protein combinations (e.g., Braun et al., 1995). A particularly interesting illustration is the observation of three locally different comparably populated conformations of the specificity-conferring residue Gln-50 in the crystal structure of the even-skipped homeodomain complex with its target DNA (Hirsch and Aggarwal, 1995). It is now of keen interest to complement these indications of conformational isomerism by investigations that can combine atomic spatial resolution with temporal resolution over a wide range of frequencies, as is the case with NMR spectroscopy and MD simulations.

With the Antp HD–DNA complex, NMR has been used for both the structure determination and for studies of dynamic aspects of protein–DNA recognition. The 16 conformers selected to represent the NMR solution structure of the Antp HD–DNA complex exhibited 16 different combinations of intermolecular interactions, including different hydration of the intermolecular interface (Billeter et al., 1993). Direct NMR evidence was obtained for internal mobility of the side chain of Asn-51 on the millisecond time scale (Billeter et al., 1993; Qian et al., 1993a). Finally, NMR experiments yielded approximate upper and lower bounds of 20 ms and 1 ns, respectively, for the residence times of hydration water molecules located in the protein–DNA interface (Figure 4). Considering that the NMR structure of the Antp HD–DNA complex (Billeter et al., 1993) and more recently the crystal structures of DNA complexes with the paired homeodomain (Wilson et al., 1995) and the even-skipped homeodomain (Hirsch and Aggarwal, 1995) all demonstrate the importance of water-mediated hydrogen bonds among the protein–DNA interactions, the NMR data on the short-lived nature of these water-mediated structures are particularly direct evidence for local rate processes involved in DNA recognition by homeodomains and other DNA-binding proteins.

The model of the Antp HD–DNA complex that emerges from the analysis of the presently described 2 ns MD trajectory is consistent with the NMR data, and in addition provides a picture of the reaction pathways linked to the frequencies measured by NMR. It further gives some leads for the interpretation of the range between the upper and lower bounds on the lifetimes of the hydration waters in the protein–DNA interface defined by the NMR measurements. The implication from the MD simulation is that the actual lifetimes of these waters are rather close to the lower limit of 1 ns. This high mobility leads to a constantly ongoing interchange between a large number of different interaction networks, and we have to conclude that the observed specificity of the intermolecular recognition is in part the result of the ensemble of these rapidly interconverting nonbonding structures. The special role of the solvent must be largely based on the small size of the water molecule and its ability to form up to four hydrogen bonds. Water molecules appear to act not only as building blocks to improve the complementarity of the interaction surfaces of the protein and the DNA, but also as a lubricant to reduce entropic costs arising when a dense network of interactions is required for highly specific macromolecular recognition. This would also be supported by the conclusion of Dunitz (1994) that most surface hydration water molecules associated with proteins or nucleic acids have similar behavior to that of liquid water. These general considerations can be expected to apply to a wide variety of protein–DNA systems. For the presently considered Antp HD–DNA complex, where genetic and biochemical experiments have convincingly identified a pivotal role of residue 50 for the specificity of the DNA recognition (Hanes and Brent, 1989; Treisman et al., 1989; Percival-Smith et al., 1990), Figures 4 and 8 very nicely illustrate that Gln-50 has a central role in the intermolecular interaction network formed by the two macromolecules and several water molecules. Considering that Gln-50 interacts with four base pairs and bases in both DNA strands (Figure 8; Table 1), it is difficult to speculate on the potential interactions to be expected when Gln-50 is replaced by a different amino acid. In particular, it will be interesting to compare the results of the present MD simulation with those from a future similar MD simulation with a homeodomain containing Lys in position 50 and a DNA α strand containing cytosines in positions 6 and 7, as is found in the bicoid HD–DNA system (Hanes and Brent, 1989).

Experimental Procedures

For the MD simulation, the program OPAL (P. Luginbühl, P. Güntert, M. Billeter, and K. Wüthrich, unpublished) was used. The main goal in the design of OPAL was to combine efficiency in the calculation of MD trajectories for large molecular systems with the flexibility and ease of use of a macro language (Güntert et al., 1992). Thus, during an MD simulation, OPAL computes with an average of 1.5 GFlops on a NEC SX-3 supercomputer and may perform a complete MD run with less than a dozen commands. The AMBER force field (Weiner et al., 1986) was used for the definition of the energetic interactions.

The coordinates of the starting structure for the MD calculations (Figure 1) were those of one of the 16 conformers that represent the NMR solution structure of the Antp HD–DNA complex (Billeter et al., 1993). As described in detail in the Results section, 3 N-terminal and 11 C-terminal residues of the protein, as well as a total of 5 base pairs near the 2 ends of the DNA duplex, were

removed in order to avoid unnecessary and costly calculations of fragments with poorly defined structure that are well separated from the region of interest, i.e., the recognition helix in the DNA major groove. In the truncated complex containing the polypeptide chain with residues 3–56 of the Antp HD and a DNA double helix with the α-strand sequence d-$A_4G_5C_6C_7A_8T_9T_{10}A_{11}G_{12}$, all protein–DNA contacts observed for the HD recognition helix in the NMR structure are present.

A spatial boundary surrounds the 1007 protein atoms and the 572 DNA atoms that form the macromolecular complex; this boundary was shaped as an ellipsoid with an initial minimal distance of 6.0 Å from any atom of the macromolecules. The ellipsoid was filled with 2714 water molecules, for which the positions and orientations were equilibrated by a separate MD run prior to the addition of the complex. The random addition of the water molecules resulted in the initial placement of one water molecule inside the protein–DNA interface; this particular water molecule left the interface after ~500 ps. Atoms were confined to the interior of the ellipsoidal "container" (Figure 2) by a potential rising from zero to infinity within a boundary layer of 1.0 Å thickness. No counter-ions were used, since an aqueous mixture with ions would on its own take several nanoseconds of MD simulation to equilibrate (Schiffer et al., 1995). A trajectory covering a time range of 2 ns, including an equilibration phase of ~100 ps, was then performed, with integration steps of 2.5 femtoseconds (fs). The coordinates of the protein, the DNA, and the water atoms were recorded in intervals of 1 ps. SHAKE was applied to constrain all covalent bond lengths (Ryckaert et al., 1977). The temperature and the pressure were kept at ambient values, i.e., 300 K and 1 bar, by scaling of the velocities and the coordinates, respectively (Berendsen et al., 1984). Recording of these quantities in intervals of 25 fs yielded average values and standard deviations of $T = 299.4 \pm 2.2$ K and $p = 1.0 \pm 0.2$ bar, respectively. No cutoff for nonbonding interactions was applied. Two nanoseconds of simulated trajectory time required 335 hr of CPU time on an NEC SX-3 supercomputer.

Analyses of the MD trajectory were performed with TRAJEC, a supplementary program to OPAL, and with the molecular display and analysis software MOLMOL (Koradi et al., 1996). With TRAJEC, all 2000 snapshots of the trajectory including all atoms of the protein, the DNA, and the water can be read, and for example, interatomic interactions exceeding a specified lifetime can be listed. TRAJEC was further used to trace selected dihedral angles, interatomic distances, and the locations of specified water molecules.

Acknowledgments

We thank the Centro Svizzero di Calcolo Scientifico for use of the NEC SX-3 computer, the Schweizerischer Nationalfonds for financial support (project 31.32033.91), and R. Marani for the careful processing of the text.

Received February 28, 1996; revised April 25, 1996.

References

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. J. Chem. Phys. *81*, 3684–3690.

Billeter, M. (1995). Hydration water molecules seen by NMR and x-ray crystallography. Prog. NMR Spectr. *27*, 635–645.

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

Braun, W., Kallen, J., Mikol, V., Walkinshaw, M.D., and Wüthrich, K. (1995). Three-dimensional structure and actions of immunosuppressants and their immunophilins. FASEB J. *9*, 63–72.

Brunne, R.M., Liepinsh, E., Otting, G., Wüthrich, K., and van Gunsteren, W.F. (1993). Hydration of proteins. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J. Mol. Biol. *231*, 1040–1048.

Denisov, V.P., and Halle, B. (1995). Protein hydration dynamics in aqueous solution: a comparison of bovine pancreatic trypsin inhibitor and ubiquitin by oxygen-17 spin relaxation dispersion. J. Mol. Biol. *245*, 682–697.

Dötsch, V., and Wider, G. (1995). Exchange rates of internal water molecules in proteins measured using pulsed field gradients. J. Am. Chem. Soc. *117*, 6064–6070.

Drenth, J. (1994). Principles of Protein X-ray Crystallography (New York: Springer).

Dunitz, J.D. (1994). The entropic cost of bound water in crystals and biomolecules. Science *264*, 670.

Frey, M. (1994). Water structure associated with proteins and its role in crystallization. Acta Crystallogr. *D50*, 663–666.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211–223.

Gewirth, D.T., and Sigler, P.W. (1995). The basis for half-site specificity explored through a non-cognate steroid receptor–DNA complex. Struct. Biol. *2*, 386–394.

Güntert, P., Dötsch, V., Wider, G., and Wüthrich, K. (1992). Processing of multidimensional NMR data with the software PROSA. J. Biomol. NMR *2*, 619–629.

Hanes, S.D., and Brent, R. (1989). DNA specificity of the bicoid activation protein is determined by homeodomain recognition helix residue 9. Cell *57*, 1275–1283.

Hirsch, J.A., and Aggarwal, A.K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. EMBO J. *14*, 6280–6291.

Karplus, P.A., and Faerman, C. (1994). Ordered water in macromolecular structure. Curr. Opin. Struct. Biol. *4*, 770–776.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *27*, in press.

Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J., and Wüthrich, K. (1990). Protein–DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. EMBO J. *9*, 3085–3092.

Otting, G., Liepinsh, E., and Wüthrich, K. (1991a). Protein hydration in aqueous solution. Science *254*, 974–980.

Otting, G., Liepinsh, E., and Wüthrich, K. (1991b). Proton exchange with internal water molecules in the protein BPTI in aqueous solution. J. Am. Chem. Soc. *113*, 4363–4364.

Otwinowski, Z., Schevitz, R.W., Zhang, R.-G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., and Sigler, P.B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. Nature *335*, 321–329.

Percival-Smith, A., Müller, M., Affolter, M., and Gehring, W.J. (1990). The interaction with DNA of wild-type and mutant *fushi tarazu* homeodomains. EMBO J. *9*, 3967–3974. Corrigendum EMBO J. *11*, 382.

Qian, Y.Q., Otting, G., and Wüthrich, K. (1993a). NMR detection of hydration water in the intermolecular interface of a protein–DNA complex. J. Amer. Chem. Soc. *115*, 1189–1190.

Qian, Y.Q., Otting, G., Billeter, M., Müller, M., Gehring, W.J., and Wüthrich, K. (1993b). NMR spectroscopy of a DNA complex with the uniformly ^{13}C-labelled *Antennapedia* homeodomain and structure determination of the DNA-bound homeodomain. J. Mol. Biol. *234*, 1070–1083.

Ryckaert, J., Ciccotti, G., and Berendsen, H.C.J. (1977). Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. J. Comp. Phys. *23*, 327–341.

Schiffer, C., Dötsch, V., Wüthrich, K., and van Gunsteren, W.F. (1995). Exploring the role of the solvent in the denaturation of a protein: a molecular dynamics study of the DNA binding domain of the 434 repressor. Biochemistry *34*, 15057–15067.

Thanki, N., Thornton, J.M, and Goodfellow, J.M. (1988). Distribution

of water around amino acid residues in proteins. J. Mol. Biol. *202*, 637–657.

Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., and Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell *59*, 553–562.

Weiner, S.J, Kollman, P.A., Nguyen, D.T., and Case, D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. *7*, 230–252.

Wilson, D.S., Guenther, B., Desplan, C., and Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. Cell *82*, 709–719.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. (New York: Wiley).

Wüthrich, K. (1993). Hydration of biological macromolecules in solution: surface structure and molecular recognition. In DNA and Chromosomes, Cold Spring Harbor Symp. Quant. Biol. *58*, 149–157.

Wüthrich, K. (1995). NMR in Structural Biology. (Singapore: World Scientific).

Zhang, X.-J., and Matthews, B.W. (1994). Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. Protein Sci. *3*, 1031–1039.