JOURNAL OF MOLECULAR RECOGNITION
J. Mol. Recognit. 2004; **17**: 120–131
DOI:10.1002/jmr.658

The role of flexibility and hydration on the sequence-specific DNA recognition by the Tn916 integrase protein: a molecular dynamics analysis

Alemayehu A. Gorfe, Amedeo Caflisch and Ilian Jelesarov*
Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

The N-terminal domain of the Tn916 integrase protein (INT-DBD) is responsible for DNA binding in the process of strand cleavage and joining reactions required for transposition of the Tn916 conjugative transposon. Site-specific association is facilitated by numerous protein–DNA contacts from the face of a three-stranded β-sheet inserted into the major groove. The protein undergoes a subtle conformational transition and is slightly unfolded in the protein–DNA complex. The conformation of many charged residues is poorly defined by NMR data but mutational studies have indicated that removal of polar side chains decreases binding affinity, while non-polar contacts are malleable. Based on analysis of the binding enthalpy and binding heat capacity, we have reasoned that dehydration of the protein–DNA interface is incomplete. This study presents results from a molecular dynamics investigation of the INT–DBD–DNA complex aimed at a more detailed understanding of the role of conformational dynamics and hydration in site-specific binding. Comparison of simulations (total of 13 ns) of the free protein and of the bound protein conformation (in isolation or DNA-bound) reveals intrinsic flexibility in certain parts of the molecule. Conformational adaptation linked to partial unfolding appears to be induced by protein–DNA contacts. The protein–DNA hydrogen-bonding network is highly dynamic. The simulation identifies protein–DNA interactions that are poorly resolved or only surmised from the NMR ensemble. Single water molecules and water clusters dynamically optimize the complementarity of polar interactions at the 'wet' protein–DNA interface. The simulation results are useful to establish a qualitative link between experimental data on individual residue's contribution to binding affinity and thermodynamic properties of INT–DBD alone and in complex with DNA. Copyright © 2004 John Wiley & Sons, Ltd.

Keywords: **molecular dynamics; protein–DNA binding; thermodynamics; conformational flexibility; hydration**

Received 26 June 2003; revised 9 November 2003; accepted 12 November 2003

INTRODUCTION

The Tn916 conjugative transposon spreads resistance to tetracycline among pathogenic bacteria (Scott and Churchward, 1995). The transposon is encoding for the integrase protein that performs strand cleavage and joining reactions during transposition (Clewell and Fllanagan,1993; Clewell *et al.*, 1995). The N-terminal integrase domain (INT-DBD) is responsible for specific DNA binding. Structural, biochemical and thermodynamic studies have provided insight into the energetics of site-specific recognition at the macroscopic level (Connolly *et al.*, 1998, 2000; Milev *et al.*, 2003a,b; Wojciak *et al.*, 1999). However, a comprehensive understanding of the binding process requires microscopic structural details about the dynamic properties of the protein–DNA complex and its components in the unbound state. Global flexibility, local dynamics of discrete protein–DNA contacts and interfacial water molecules may be important determinants of association.

Sequence-specific protein–DNA binding is often accompanied by structural adaptation (Dyson and Wright, 2002; Jen-Jacobson *et al.*, 2000). Disorder-to-order conformational transitions of the protein upon DNA binding are typical. However, the opposite has also been observed. For example, the p66 fingers and thumb sub-domains of HIV-1 reverse transcriptase undergo a structural transition from a closed to an open conformation upon binding to DNA. This was shown to originate from increased flexibility induced by DNA (Madrid *et al.*, 2001). The solution structure of INT-DBD has been solved in both the DNA-bound and DNA-free forms (Connolly *et al.*, 1998; Wojciak *et al.*, 1999). Comparison of the structures reveals distinct differences (Fig. 1). The major structural transition involves turn T2 connecting β-strands 2 and 3. It is pulled away from the C-terminal α-helix in the bound form. Although turn T2 moves as a unit with preserved local structure, the edge of the hydrophobic core becomes disrupted and the C-terminal α-helix is two amino acids shorter in the bound protein.

*Correspondence to: I. Jelesarov, Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.
E-mail: iljel@bioc.unizh.ch
Contract/grant sponsors: Swiss National Science Foundation; Swiss National Center for Competence in Research in Structural Biology.

Abbreviations used: β, β-strand; ΔG, binding free energy; $\Delta\Delta G$, change of the binding free energy upon X-to-Ala mutation; INT-DBD, the N-terminal DNA-binding domain of the Tn916 integrase protein (residues 2–74); L, loop; RMSD, root mean square deviation; RMSF, root mean square fluctuation; T, turn.

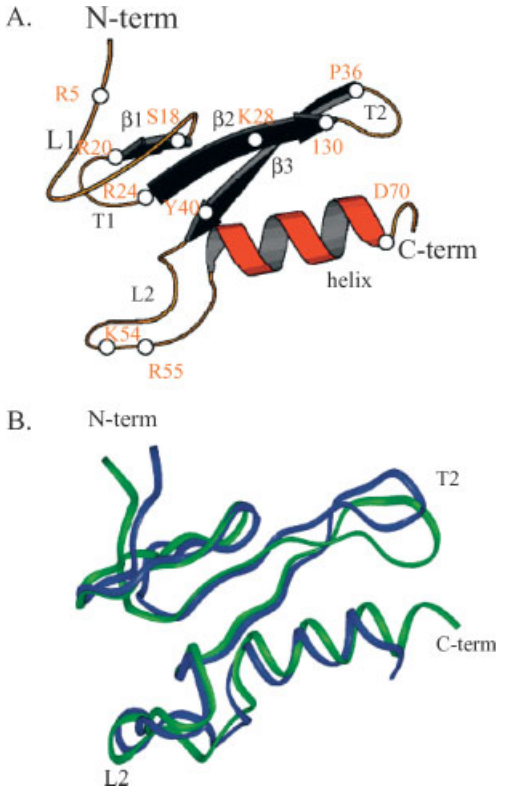

Figure 1. NMR structure of the DNA-binding domain of integrase Tn916 (INT-DBD). (A) Free form of INT–DBD. The secondary structure elements are labeled. L, loop; T, turn; β, beta strand; α-helix. The approximate location of some of the residues that are discussed in the text is indicated. (B) Comparison of the free (in green) and bound (in blue) forms of the protein. The superposition of the Cα backbone atoms of the three-stranded β-sheet and the α-helix only emphasizes the different positioning of turn T2 and loop L2, which is the major structural rearrangement induced upon binding.

Hinge-like movements may be indicative for 'built-in' flexibility. Intrinsic thermal flexibility in some parts of the protein has recently been suggested based on our calorimetric and spectroscopic studies (Milev *et al.*, 2003a,b). We found the apparent molar heat capacity of the protein to display non-linear temperature dependence before the onset of the main unfolding transition. Furthermore, INT–DBD is marginally stable and may thus be susceptible to local unfolding due to accumulation of thermal fluctuations. Although the order-to-disorder transition appears to be concomitant with the formation of INT–DBD–DNA contacts, it is not *a priori* known whether binding *per se* induces structural changes in a flexible molecule. Alternatively, the free and bound conformations of the protein may be separated by a low energy barrier and exist in equilibrium before binding has taken place, so that the DNA duplex 'selects' and stabilizes the binding-competent form.

INT–DBD recognizes DNA by a rare structural motif, a three-stranded β-sheet. Two unambiguous base-specific intermolecular hydrogen bonds involving Tyr40 and Lys28 can be inferred from NMR data. A third base-specific hydrogen bond is likely made by the side chain of Arg20. In addition, van der Waals contacts and non-specific hydrogen bonds with the DNA backbone appear to help anchoring the β-sheet in the major groove. On the other hand, biochemical data have indicated that several side chains, which are disordered in the NMR structure, are energetically important (Connolly *et al.*, 2000). The lack of (abundant) NOEs for such residues suggests formation of highly dynamical intermolecular contacts but the exact mode of interaction is not easily discernible.

It has long been recognized that water molecules play an important role in protein–DNA association (Schwabe, 1997). Structural and computational analysis have identified water molecules bridging protein and DNA groups and often participating in hydrogen bonding networks (Billeter *et al.*, 1993; Chillemi *et al.*, 2001; Duan and Nilsson, 2002; Lynch *et al.*, 2002; Reddy *et al.*, 2001). Recently, Duan and Nilsson have performed MD simulations on wild-type and several mutant homeodomain–DNA complexes to analyze the role af Lys50 and the behavior of water molecules at the interface (Duan and Nilsson, 2002). MD simulations at ambient and elevated pressures have been used to reveal details about the direct and water-mediated interactions between *Bam*HI and its cognate DNA (Lynch *et al.*, 2002). From a structural analysis of a number of complexes, Ready and Jayaram have concluded that water molecules also serve as electrostatic screen between like charges, in addition to bridging protein–DNA hydrogen atoms (Reddy *et al.*, 2001). In the case of the INT–DBD–DNA complex, strong evidence for the presence of trapped water comes from packing density calculations and thermodynamic data (Milev *et al.*, 2003a,b). Using the NMR structures of complex, protein and DNA, we have identified a total of 140 ± 40 Å^3 of 'empty space' distributed over six or seven cavities at the complex interface. These cavities are large enough to contain together about 10 water molecules that are inaccessible to the bulk solvent. The magnitude of the binding heat capacity changes and of the binding enthalpy changes can be rationalized if one assumes that the protein–DNA interface is only partially dehydrated (Milev *et al.*, 2003a,b).

The issues mentioned above are central for an in-depth understanding of the binding process, yet they are difficult or impossible to approach by experiment. Here we perform explicit water molecular dynamics simulations aimed at elucidating the intermolecular contacts in the protein–DNA complex and the role of water molecules at atomic level of detail. Specifically, we address the following questions: (i) are there inherently flexible regions in the free protein and how does association alter the global dynamics of INT–DBD? (ii) Can we give a detailed atomic description of protein–DNA contacts, consistent with NMR and biochemical data? (iii) How does water participate in binding?

METHODS

Structures

The average energy minimized NMR structures of the free protein and of the complex with a 13 base-pair DNA duplex were obtained from the Brookhaven Protein Data Bank (accession codes 2bb8 and 1b69, respectively; Connolly *et al.*, 1998; Wojciak *et al.*, 1999). Residues that are not defined in the structures (Ser2, His72, Asp73, Gly74 in the complex; Ser2 and Gly74 in the free protein) were manually

built and bad contacts were removed by 100 steps of steepest descent minimization.

MD protocol

The MD simulations were carried out with the CHARMM program (Brooks *et al.*, 1983) using the all atom parameter set of CHARMM27 (MacKerell *et al.*, 1998). In all calculations, CHARMM charges were assigned assuming neutral pH. The solutes were immersed in a box of appropriate dimensions (in Ångströms: $60 \times 60 \times 60$, protein; $72 \times 48 \times 48$, DNA; and $74 \times 60 \times 58$, complex) containing pre-equilibrated TIP3P waters (Jorgensen *et al.*, 1983), as well as sodium and chloride ions to neutralize the total charge of the macromolecular system (+5, −24 and −19 for the protein, DNA and protein–DNA complex, respectively). In all cases the minimal distance from any solute atom to the edge of the box was 10 Å. The total number of atoms was 18 099 and 18 185 (simulations started from the free or from the bound conformation of the protein, respectively), 16 813 (simulations of DNA) and 24 923 (simulations of the protein–DNA complex). All subsequent calculations were performed under periodic boundary conditions. The system was first relaxed by 50 steps of steepest descent minimization, followed by 200 steps of the adopted-basis Newton–Raphson (ABNR) algorithm. During minimization all water oxygen atoms and all protein heavy atoms were harmonically constrained by a force constant of 1.0 and 2.0 kcal mol^{-1} Å^{-2}, respectively. The system was heated to the target temperature in 15 ps. The harmonic constraint force constant was progressively decreased from 2.0 to 0.0 kcal mol^{-1} Å^{-2} in the initial 10 ps and all atoms were free to move during the last 5 ps of the heating process. The fully unconstrained system was then subjected to a 20 ps equilibration at constant temperature and volume using Gaussian distribution for the assignment of atomic velocities. The production simulations were performed at 278, 298 and 328 K for the bound and free forms of the protein and at 298 K for the protein–DNA complex. A shift function was employed with a cutoff at 12 Å for the van der Waals interactions. The particle mesh Ewald method was used to treat the long-range electrostatics (Darden *et al.*, 1993). The real space contribution was truncated at 14 Å. The number of grid points for the fast Fourier transformations was 81, 64 and 64 for the *x*, *y* and *z* directions, respectively. The simulations were performed in the isothermal, isobaric ensemble using the leapfrog integrator (Bogusz *et al.*, 1998). The pressure and temperature were kept constant using a Langevin piston of mass 600 amus and a Hoover thermostat with a thermal piston of mass 1000 kcal ps^2. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm (Rykaert *et al.*, 1977). The integration time step was 2 fs. Structures were collected every 1 ps for analysis.

Simulations of the protein started from the structure of the free form are termed 'free simulations' (f). Trajectories initiated from the INT–DBD structure, which was extracted from the protein–DNA complex by removing the DNA are termed 'bound simulations' (b). The trajectory of the protein–DNA complex is referred to as 'complex simulation' (c). The temperature of the simulation is explicitly stated in the trajectory name. For example, 278f, 328b and 298c represent the free simulation at 278 K, the bound simulation at 328 K and the simulation of the complex at 298 K, respectively.

Analysis of the trajectories

Protein and DNA atoms were considered hydrogen-bonded if the distance between hydrogen and acceptor atoms was less than 2.6 Å and the donor–hydrogen–acceptor angle was greater than 120°. Hydrophobic interaction was defined when the distance between a pair of carbon atoms was less than 4 Å. Interfacial hydration water molecules were identified by applying a cut-off of $\leq$2.6 Å for the distance of water to protein and DNA atoms simultaneously.

CPU requirements

A multi-processor SGI machine and a Beowulf cluster were used for the calculations. Typically, a 1 ns simulation of the protein–DNA complex ($\sim$25 000 atoms) requires about 30 days on eight R10 000 195 MHz processors or 28 days on a dual Athlon 1.4 GHz computer.

RESULTS AND DISCUSSION

Trajectories were started from the energy-minimized average NMR structures of the free protein, of the bound protein after removing the DNA duplex from the coordinate file, and of the protein–DNA complex. In the following we refer to these experiments as to 'free simulations', 'bound simulations' and 'complex simulations', respectively. The free protein was simulated at 278 K, close to the temperature of maximal stability; at 298 K, where binding affinity is fully preserved but the protein undergoes a subtle conformational change; and at 328 K, 10° above the midpoint of thermal denaturation (Milev *et al.*, 2003a,b). The bound protein was simulated at 298 K. In addition, a trajectory at 328 K was started with the bound INT–DBD to speed up possible conformational transitions. The INT–DBD–DNA complex was simulated at 298 K. For brevity, the trajectories are given names consisting of a number indicating the simulation temperature followed by the letters b (for 'bound simulation'), f (for 'free simulation'), and c (for 'complex simulation'). For example, 298b and 278f are the trajectory at 298 K started from the bound conformation and the trajectory at 278 K started from the free conformation, respectively. The time courses of the C_α positional root mean square deviations (RMSD) from the starting structures are shown in Fig. 2. In the free simulations, the low and ambient temperature trajectories (278f and 298f) were stable after 350 ps; in the simulation at 328 K (328f) the C_α RMSD reached a plateau at 3 Å after about 1000 ps. The bound simulations, 298b and 328b, became stabilized after $\sim$600 ps. The Cα RMSD in the 3 ns complex trajectory (298c) reached a plateau already at 200 ps. For the purpose of comparison of different conformations, it is more appropriate to compare trajectories within a time window of the same size (1–2 ns). Protein–DNA interactions were studied using the last 2 ns of the complex trajectory.

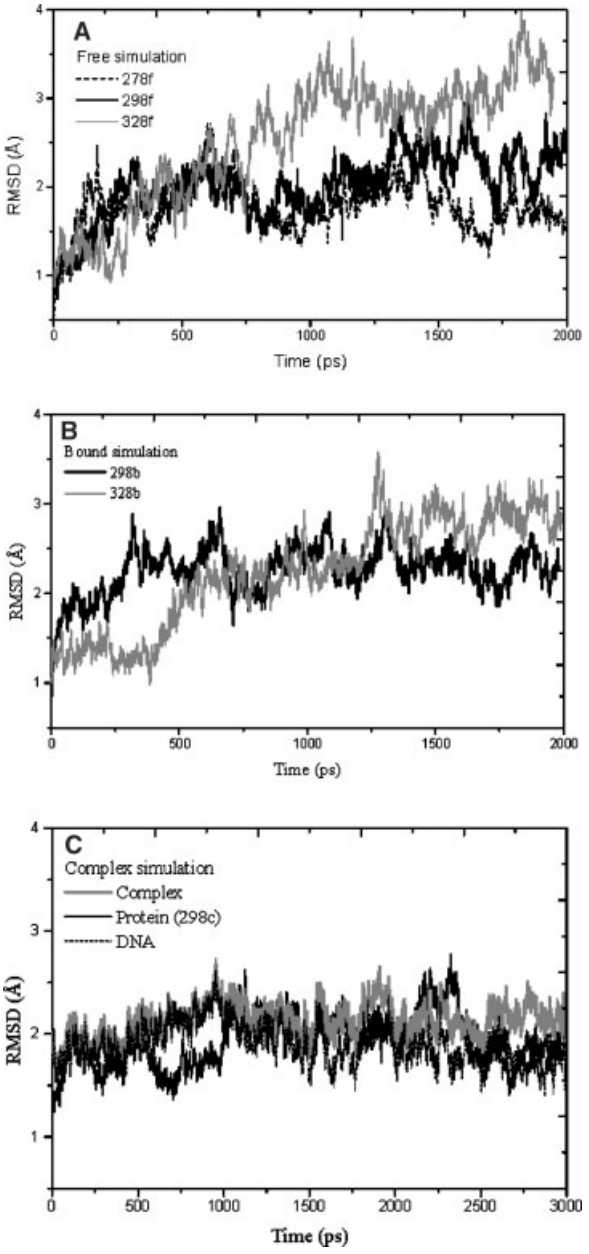

Figure 2. Time evolution of the root mean square deviation (RMSD) measured from the corresponding starting structure. (A) Simulations at 278 K (dashed line), 298 K (black heavy line) and 328 K (gray line) starting from the average NMR structure of free INT–DBD. (B) Simulations at 298 K (black heavy line) and 328 K (gray line) starting from the average NMR structure of the bound protein. (C) Simulation of the INT–DBD–DNA complex at 298 K. Dashed line, DNA backbone; heavy black line, protein Cα atoms; gray line, protein–DNA complex.

Structural properties of INT–DBD

Global behavior and structural disorder. As a measure of the global structural properties of the protein, we compare C_α RMSD, radius of gyration (R_g) and water accessible surface (ASA) averaged over 1000 snapshots. The results are shown in Table 1. Overall, we observe the typical behavior expected for low and high temperature simulations. With increasing temperature, the protein moves further away from the corresponding starting structure, slightly expands and exposes more molecular surface. Obviously, the overall thermal motions are intensified at higher temperatures. According to NMR data, bound INT–DBD is less structured than the free protein (Connolly *et al.*, 1998). The simulations are fully compatible with this observation. In the trajectory of the complex, the β-sheet is substantially less structured than in the trajectory of the free protein at the same temperature. Especially affected are strands β2 and β3, which contribute the most important protein–DNA contacting residues. Also the C-terminal α-helix is persistently shorter by one or two residues as compared to the free protein. Furthermore, INT–DBD appears less compact in the complex, as far as the radius of gyration increases and more molecular surface is exposed to the solvent (Table 1). Even when simulated at 328 K, the free form is still as compact as the bound form at 298 K. Interestingly, ASA of the protein calculated from the complex trajectory at 298 K (298c) is larger than ASA of both the free and bound protein at the same temperature, and is almost as large as ASA of the bound INT–DBD at 328 K.

Comparison of the distance deviations from the starting structure for the non-loop secondary structure elements is instructive. Since DNA binding necessarily restricts molecular motions, the protein in the complex trajectory shows the lowest average RMSD over all the non-loop structural elements (an average of 1.7 Å compared with 2.1 Å for the free protein and 2.2 Å for the bound conformation; see Table 1). However, β-strands β2 and β3, which are involved in sequence-specific DNA contacts at the center of the binding site, deviate more in the complex simulation than in the bound and free simulations. The turns, loops and the C-terminal α-helix, which either participate in interactions with the DNA backbone or do not contact DNA at all, stay closer to the NMR structure. Taken together, these observations emphasize the role of protein–DNA contacts as an important factor inducing and/or maintaining the structural changes observed in the complex.

Intrinsic and DNA-induced flexibility. Distance deviations from the starting structure may not necessarily mirror the mobility of structural elements. A more comprehensive information on flexibility is achieved by comparing root mean square fluctuations (RMSF, see Table 1). Excluding the first and last three residues, C_α-RMSF for all simulations scatter within 0.2 Å around an average of 1.08 Å, indicating that the global mobility is not much dependent on the presence or absence of DNA. This is in accord with the properties of the NMR ensemble (RMSF of 0.55 and 0.52 Å for the free protein and the protein bound to DNA, respectively). Nonetheless, a modest increase of the global fluctuations of the free protein is evident for a temperature difference of 20 °, as expected from general considerations. The same trend was obtained from analysis using shorter blocks of the trajectories (1–1.5 and 1.5–2 ns) and therefore the results do not seem to be affected by statistical factors.

Averaged RMSF may dump local differences in the mobility when comparisons are done between different conformations. To identify flexible regions in the molecule and to facilitate comparisons, we plot C_α-RMSF on a per residue basis (Fig. 3). Excluding the chain termini, it springs to the eye that turn T2 (residues 31–35) and the middle part

Table 1. Global structural characteristics of INT-DBD in the simulations

Simulation[a]	Average root mean square deviation (RMSD, Å)[b]								R_g (Å)[c]	ASA (Å^2)[d]	Fluctuations[e] (RMSF, Å)
	Total	β1	β2	β3	T1	T2	α-Helix	Average			
278f	1.8	1.7	1.0	0.9	2.6	3.0	2.5	1.95	13.1	6120	0.9
298f	2.3	1.0	1.3	1.0	1.9	3.6	3.7	2.1	13.1	6140	1.1
328f	3.1	2.5	1.9	1.7	2.2	6.3	3.5	3.0	13.4	6170	1.2
298b	2.3	1.9	1.1	1.0	2.9	4.0	2.4	2.2	13.4	6210	1.0
328b	2.7	1.0	1.0	1.5	3.6	4.6	4.4	2.7	13.8	6750	1.2
298c	2.1	0.8	1.6	2.1	1.6	2.3	2.0	1.7	13.3	6530	1.1

[a] For description of simulation names see Results. [b]Average over the MD ensemble calculated after optimal superposition of all Cα atoms (total) or Cα atoms of secondary structure elements (strands β1, β2 and β3, turns T1 and T2, and the α-helix). [c]Radius of gyration. [d]Atomic surface accessibility calculated with NACCESS (Hubbard and Thornton, University College London, 1993). [e]RMSF of C_α atoms from the average MD structure in the last 1 ns of the simulation.

of loop L2 (residues 49–55) show distinctly large RMSF in all simulations, but there are variations among the trajectories. The segments anchoring loop L2 to the body of the protein (residues 43–48 and 56–60) and turn T1 (residues 20–23) are moderately flexible. The least fluctuating segments are loop L1 (residues 4–17, except for Lys 14), strand β1(residues 18–19), strand β2 (residues 25–30), the C-terminal part of strand β3 (residues 37–41), and most of the α-helix (except for the C-terminus). This pattern is not surprising since many residues located in strands β2 and β3, in loop L1, and in the α-helix contribute to the hydrophobic core. Since the distribution of mobile and less mobile segments along the chain is very similar in the free protein and in the protein–DNA complex, we conclude that DNA binding does not much alter the structural flexibility of INT–DBD. In fact, the fluctuations of turn T2 and loop L2 are reduced upon binding. Hence, high flexibility in these two regions appears to be an intrinsic, 'built-in' property of the protein. We note that all regions, which are more flexible than the average RMSF of 1.1 Å, participate in one way or another in binding. The 1.8 Å displacement of turn T1 facilitates hydrogen bonds via Arg20 and Lys21. Van der Waals contacts of Ile30 and Pro36, both located at the base of turn T2, are possible because this loop is displaced by 4.5 Å toward the duplex. Lys54 and Lys55, which contribute $\sim$1.3 kcal mol^{-1} to binding are located in the middle of loop L2. DNA binding slightly intensifies the thermal motions of strand β3 as compared with both the free and bound simulations (Fig. 3). This strand also has a larger RMSD from the starting structure in 298c (Table 1). As β3 contributes some of the critical residues for affinity (e.g. Tyr40), its flexible nature may point to a dynamic local interaction. Protein–DNA interactions via residues located in flexible segments may be entropically less expensive since the entropic penalty of backbone freezing is smaller. The increased flexibility of the last five residues (Asp69–Gly74) which are not involved in binding can be explained with the disruption of the edge of the hydrophobic core at the end of the α-helix.

Effect of temperature on the flexibility. Previous calorimetric results are consistent with a minor conformational transition of INT-DBD occurring above $\sim$15 °C, at temperatures where the protein is globally folded and binds to DNA with the same affinity as at lower temperatures (Milev *et al.*, 2003a,b). Unfolding is often initiated in flexible regions of the molecule (Pedersen *et al.*, 2002; Rader, *et al.*, 2002). The increase in RMSD, R_g and ASA suggests that the protein becomes relaxed at higher temperature. The most obvious increase of flexibility upon an increase of the simulation temperature is seen in the C-terminal half of the protein while the fluctuation in turn T1 and loop L1 are not very temperature-dependent. Therefore, the MD simulation supports the calorimetric observation of a structural rearrangement below the temperature range of the main two-state unfolding transition.

In conclusion, our analysis of the structural properties of the protein indicates that INT–DBD is generally more disordered as a result of the conformational change upon DNA binding. However, this disorder appears to be due to the existence of intrinsically flexible regions in the free protein. The C-terminal half of the protein becomes more flexible at increased temperature, suggesting that partial loss of structure may occur in this region before the onset of unfolding.

Protein–DNA interactions

As shown in Fig. 2, the RMSD of both the protein and DNA in the complex measured from the average NMR starting structure reached a plateau after 200 ps, with an overall average of 1.9 Å (DNA backbone), 1.9 Å (protein C_α atoms) and 2.1 Å (complex). For analysis, we consider the trajectory after the first nanosecond; reference is made to the whole trajectory where relevant. The DNA duplex in complex with INT–DBD is very stable throughout the entire simulation. Hydration and helical properties indicate that the DNA remains B-like in both the free and complexed form, although the major groove is widened as the result of binding. The DNA displays no special behavior along the trajectory and, therefore, the properties of the duplex will not be discussed in detail.

Direct and water-mediated protein–DNA contacts. Protein–DNA complexes are held together by van der Waals contacts and hydrogen bonds, which can be either direct or mediated by water molecules (Billeter *et al.*, 1993, 1996;

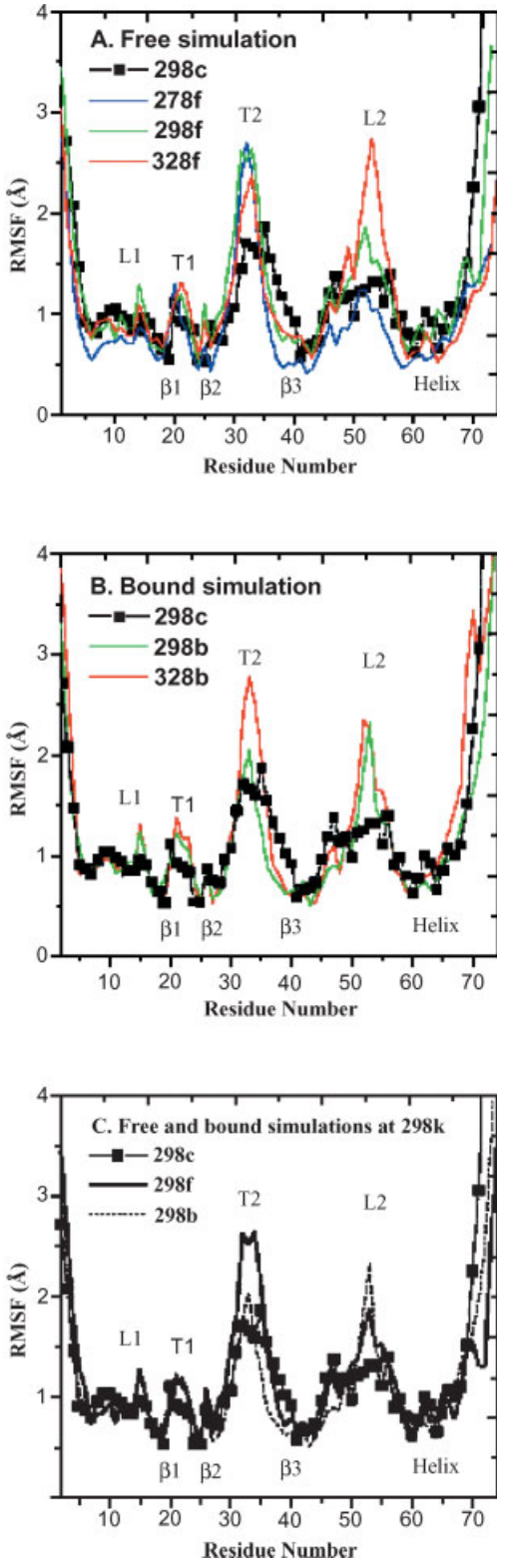

Figure 3. Root mean square fluctuation (RMSF) of the Cα atoms around the average MD structure. (A) Simulations of the free protein at 278 K (blue), 298 K (green) and 328 K (red). (B) Simulations of the bound protein at 298 K (green) and 328 K (red). (C) Comparison of RMSF measured for the free (continuous line) and bound (dotted line) INT-DBD at 298 K. In all panels RMSF measured for the protein simulated in complex with DNA at 298 K is shown to serve as a reference (-■-).

Davey *et al.*, 2002; Gruschus and Ferretti, 2001; Reddy *et al.*, 2001; Sen and Nilsson, 1999). In a few cases, like for example in the *trp* repressor–operator complex, no direct base-specific protein–DNA contacts are observed, and specificity is renderd through water-mediated polar contacts (Otwinowski *et al.*, 1988). Several studies have identified stable (long-lived) water molecules bridging protein and DNA (Chillemi *et al.*, 2001; Labeots and Weiss, 1997; Sunnerhagen *et al.*, 1998; Tsui *et al.*, 2000), but fast-exchanging (transient) waters have also been found to serve as bridges (Otting *et al.*, 1997; Tsui *et al.*, 2000). The presence of water at the INT–DBD–DNA interface has been surmised from analysis of the magnitude of the enthalpy and heat capacity changes associated with binding (Milev *et al.*, 2003a,b). Moreover, voids and cavities can be identified in the NMR structure of the complex and thus bound water molecules are expected to participate in one way or another in formation of the interface.

Table 2 presents the direct and water-mediated hydrogen bonds extracted from the trajectory. Based on biochemical data and structural analysis of the NMR ensemble, 11 polar or charged residues were shown to make hydrogen bonds with the bases and DNA backbone polar atoms (Connolly *et al.*, 2000; Wojciak *et al.*, 1999). All these interactions are maintained during the simulations, albeit with varying population. The majority of hydrogen-bonded side chains are simultaneously involved in hydrophobic contacts with DNA. An average of 28 ± 3 waters form a characteristic 'wet' interface (Janin, 1999) in the simulation. In agreement with statistical analysis of high-resolution protein–DNA structures predicting 15 ± 7 tightly bound waters to be present at a protein–DNA interface of 2300 Å^2 (Nadassy *et al.*, 1999) about half of them mediate protein–DNA contacts, mostly through bridging side chains to DNA, with only few backbone amide-to–DNA bridges. However, the majority of water molecules is mobile and exchanges fast with the bulk. We therefore describe the participation of water at the interface from the perspective of individual side chains, as non-separable part of the bonding pattern observed in discrete 'sites'. In the following we discuss the bonding interactions observed in the simulation with reference to the NMR structure of the INT–DBD–DNA complex (Wojciak *et al.*, 1999) and to an extensive mutational study aimed at elucidating the role of discrete contacts to affinity (Connolly *et al.*, 2000). Although changes in overall affinity upon mutation can contain contributions that are not directly caused by the loss of specific contacts, the pattern of interactions observed in the simulation might help to rationalize the experiments.

All the structural elements except turn T2 and the α-helix contribute residues that interact with DNA. The major protein–DNA contacts occur at the largely hydrophobic interface formed by turn T1 and strands β2 and β3 (Fig. 1). Four residues from strand β3 (Pro36, Phe38, Tyr40 and Trp42) interact with DNA atoms both in the NMR and MD ensemble. The van der Waals packing interactions of Pro36 and Phe38 with either side of Thy22 methyl group are well defined during the simulation. An additional contact between these residues and Thy4 was also observed in the early stages but was broken after 1 ns. The Pro36–Thy22 interaction persists for $\sim$40% of the simulation time. Elimination of the methyl group of Thy22 by

Table 2. Protein–DNA contacts observed in the simulation[a]

	Direct hydrogen bond		Van der Waals		Water-mediated hydrogen bond[b]			
Residue[e]	DNA[f]	Percentage of population	DNA[e]	Percentage of population	DNA[e]	Percentage of population	$\Delta\Delta G$ (kcal mol^{-1})[c]	NMR[d]
Tyr40 (β3)	Ade20(b) Cyt21(b)	60	Thy19(s,b)	100	Ade5(b) Ade20(b) Cyt21(b)	25	1.55 ± 0.04	BHB
Tyr40,HN					Thy19(p) Ade20(p)	36		
Lys28 (β2)	Gua3(b)	35	Ade2(b) Gua3(b) Thy4(b)	25	Ade2(b,p) Gua3(b,p)	90	1.42 ± 0.02	BHB
Lys54 (L2)	Thy18(p,s)	100	—	—	Thy17(**p**,s) Thy18(**p**,s)	44	1.37 ± 0.09	PHB
Arg24 (T1)	Thy19(p)	64	Thy18(b)	25	Thy18(b,p) Thy19(b,p)	100	1.26 ± 0.03	PHB
Arg55 (L2)	Thy18(s) Thy19(p)	36	Thy18(s) Thy19(s)	63	Thy18(**p**,s) Thy19(**p**,s)	100	1.21 ± 0.07	PHB
Arg5 (L1)	Gua3(p) Thy4(p)	60	Gua3(s)	36	Ade2(p,s) Gua3(p,s) Thy4(p,s)	100	0.79 ± 0.13	PHB
Lys21 (T1)	Ade5(p) Gua6(p)	25	—	—	Ade5(p) Gua6(p)	60	0.74 ± 0.01	PHB
Trp42 (β3)	Thy19(p)	60	Thy18(s)	30	Thy17(p) Thy18(p)	40	0.49 ± 0.05	PHB
Arg20 (T1)	Ade5(p) Gua6(b,p)	100	Thy4(s) Ade5(b)	100	Ade5(b,p) Gua6(b,p)	90	0.47 ± 0.07	BHB
Gly16,HN (L1)	—	—	—	—	Gua3(p)	46	n.d.	
Ser18 (β1)	—	—	Gua3(s)	25	Gua3(**p**,s) Thy4(**p**,s)	50	-0.13 ± 0.07	PHB
Leu26 (β2)	—	—	Thy4(b)	33	—	—	-0.20 ± 0.09	BvdW
Pro36 (β3)	—	—	Thy22(b)	36	—	—	n.d.	
Gln37 (β3)	—	—	—	—	Cyt21(p)	32	n.d.	
Phe38 (β3)	—	—	Cyt21(b) Thy22(s)	100			-0.28 ± 0.13	BvdW
Phe38, NH	—	—	—		Ade20(p) Cyt21(p)	32		

[a] Only contacts populated $>25\%$ in the last 2 ns of the trajectory are listed. For the definition of hydrogen bonds see Methods. Residues are listed according to their apparent importance to binding derived from an experimental alanine scanning mutagenesis study (Connolly *et al.*, 2000). [b]Only water bridges populated for at least 500 ps are shown. [c]Changes in binding free energy upon replacement of the side chain by alanine: $\Delta\Delta G = \Delta G_{mut} - \Delta\Delta G_{wt}$ (from Connolly *et al.*, 2000). [d]The last column indicates the type of contact eliminated by mutation to alanine according to analysis of the NMR structure (Connolly *et al.*, 2000). BHB, base hydrogen bond; PHB, phosphate hydrogen bond; BvdW, base van der Waals contact. [e]Secondary structure elements to which a given residue belongs are given in parentheses. Interactions via the backbone amide are labeled by NH. [f]The nucleotide names are followed in parentheses by the letters b, p and s, indicating that the protein group contacts base atoms, phosphate oxygens or sugar atoms, respectively. When multiple contacts are observed, the dominating interaction is given in bold.

substitution to uracil had no effect on binding-affinity and hence this interaction is not energetically important. On the other hand, the side chain of Phe38 interacts more persistently with the nucleotide ring of Cyt21. This interaction has not been observed by NMR experiments, but is consistent with the energetic role of this base inferred from mutational experiments. It should be noted, however, that replacement of Phe38 by Ala is itself slightly stabilizing the complex ($\Delta\Delta G = 0.3$ kcal mol^{-1}). Thus the energetic role of the Phe38–Cyt21 van der Waals contacts is not clear. Cyt21 participates in multiple contacts to other side-chain atoms. Among those is the hydrogen bond interaction with Tyr40, which is located at the center of the binding site and contributes 1.5 kcal mol^{-1} to binding affinity. In the first half of the trajectory, Tyr40 interacts mainly with Ade20 accepting a hydrogen bond from N6 and donating a hydrogen bond to N7 but occasionally contacts Cyt21 as well (the average distance between Tyr40:O_η and Cyt21:N4 is 3.5 Å). After $\sim$2 ns, the Tyr40$\sim$Ade20 bonds are broken and are replaced by a hydrogen bond with the N4 atom of Cyt21. In the NMR ensemble of 20 structures, Tyr40 accepts a hydrogen bond from N4 of Cyt21 (65%) and donates a hydrogen bond to N7 of Ade20 (10%). The switch in interaction of Tyr40 from Ade20 to Cyt21 in the course of

the simulation thus identifies the dynamic character of this very important interaction. In total, Tyr40 remains directly hydrogen bonded to a base for three-quarters of the total simulation time. In about 25% of the conformers, a single water molecule serves as a bridge between the hydroxyl group and DNA. In addition, the main-chain amide of this residue also contacts DNA (Ade20 phosphate oxygen) through a water molecule. Interestingly, the lost hydrogen bond between Tyr40 and Ade20 is 'replaced' by otherwise infrequent direct or water-mediated interaction between $N_{\varepsilon}1$ atom of Trp42 and the phosphate oxygen atoms of Thy19. Details about the structural bases of the interaction network involving Tyr40 and Trp42 will be discussed in the following section. Additional waters bridge the backbone amides of Gln37 and Phe38 to Thy18 and Ade20/Cyt21, respectively, thus contributing to the formation of a significant contact interface between strand $\beta 3$ and DNA.

Contacts from strand $\beta 2$ include two hydrophobic residues (Leu26 and Ile30) and a charged residue (Lys28). As observed in the NMR ensemble, Leu26 is involved in a van der Waals contact with the nucleotide ring of Thy4. The interaction persists throughout the 3 ns of simulation time but is lowly populated (35%). The rare interaction of Ile30 with Gua1 at the 5′ end of the duplex disappears completely in the last 1 ns. The instability of these two hydrophobic contacts can explain why substitution of Leu26 and Ile30 by alanine has no effect on affinity. In contrast, Lys28 was shown to contribute significantly to the binding free energy by both experimental and computational mutagenesis (Connolly *et al.*, 2000; Gorfe and Jelesarov, 2003) and is present in about half of the NMR conformers. In the course of the simulation, the base specific hydrogen bond of Lys28 with N7 or O6 of Gua 3 is almost exclusively mediated by water (90% of the time). A second, direct hydrogen bond is observed mainly in the first half of the simulation (35% of the conformers). Additionally, the Lys28 side chain contacts apolar groups of Ade2 and Gua3, which interaction is weakened after $\sim$1.5 ns; at later stages a van der Waals contact with the base of Thy4 is populated.

Stable and persistent network of protein–DNA hydrogen bonds and packing interactions from turn T1 involving Arg20, Lys21 and Arg24 are observed along the trajectory. Arg20 is poorly defined in the NMR structure, yet it closely approaches several hydrogen bond acceptors. The simulation identifies direct hydrogen bonds with O6 and N7 of Gua6 in the major groove. However, the base specific interactions were broken around 1.4 ns, concomitantly with those of Tyr40 and Lys28. In the remaining part of the trajectory direct contacts with the phosphate oxygen atoms of Ade5, but also a rare hydrogen bond to the sugar oxygen O5′ are present. The side chain is bridged through water with several polar groups in the major groove and participates in hydrophobic interactions with the nucleotide ring of Ade5 in the first half of the trajectory, and with the sugar carbons of Thy4 in the second half. The conformational change at 1.4 ns also involves the solvent exposed and flexible side-chain of Lys21, whose interaction with Gua6 breaks at that time point. Lys21 makes stronger, water-mediated hydrogen bonds to base and backbone atoms of Ade5 and Gua6. One of the major contributors to binding affinity, Arg24, is located in turn T1. The 35° bend of the DNA duplex toward INT–DBD in the complex facilitates a network of hydrogen bonds with phosphates located away from the β-sheet–DNA interface (Thy18 and Thy19). However the direct interaction with Thy18 is stable only until about 800 ps, afterwords Arg24 makes water-mediated polar contacts with both Thy18 and Thy19 and a low-populated apolar contact with the base of Thy18.

Replacement by alanine of Arg5 and Arg55 from loops L1 and L2, respectively, decreases INT-DBD binding affinity. The side chains are disordered in the NMR structure, possibly due to the flexibility of the loops. In the simulation Arg5 is hydrogen bonded to Thy4 and to a lesser extent to Gua3. The direct hydrogen bond interaction is significantly populated (>50%). However, water bridges with phosphates of Ade2 and Gua3 persist for more than 85% of the simulation and thus appear to be the more important interaction. The side chain is very flexible and substantially surface-exposed. In the case of Arg55, direct hydrogen bonding is not observed until about 2.5 ns, except for the occasional interactions with O5′ of Thy18 and phosphates of Thy19. A strong hydrogen bond with the phosphate of Thy18 and a less frequent contact with O3′ of Thy17 appear later in the simulation. Also, highly populated water bridges are observed with backbone atoms of Thy18 and Thy19. In addition, Arg55 participates in dynamic hydrophobic contacts with the sugar carbon atoms of Thy11, Thy18 and Thy19. All these interactions may rationalize the free energy loss of 1.2 kcal mol^{-1} upon removal of the side chain of Arg55. Loop L2 also provides another critical residue, Lys54. Its hydrogen bond to a phosphate of Thy18 is one of the most significantly populated (96%) in the simulation, while the van der Waals contacts with the same base become weaker with time. Occasionally, a water-mediated bond with Thy17 is seen.

Additional protein–DNA contacts involving a cluster of three residues located in the C-terminal part of loop L1 and the adjacent strand $\beta 1$ were inferred from the NMR structure. Thr15 is positioned in the NMR ensemble as to make a van der Waals contact with a ribose. Such interaction is not populated to any significant extent along the trajectory. It is likely, that it is also not present in solution, since binding of the Thr15-to-Ala mutant is unchanged compared to wild type INT-DBD. A hydrogen bond between the main-chain amide of Gly16 and phosphates of Gua3 (present in 95% of the NMR conformers) is observed in the initial 250 ps but is then lost until it reappears with a modest population ($\sim$20%) around 2 ns. In contrast, the Ser18 hydroxyl-to-Thy4 phosphate hydrogen bond, which is populated to only 10% in the NMR ensemble, is detected only in the final 0.5 ps, consistent with the lack of effect of the Ser18-to-Ala mutation. However, these residues contact DNA through water molecules. We observe no interactions of Gln19 with DNA. Therefore, contacts from strand $\beta 1$ seem not to contribute much to the stability of the complex as inferred from mutational results.

The observed interactions are in a good agreement with NMR data. The protein–DNA contact site is defined by 74 inter-molecular NOEs (Wojciak *et al.*, 1999). In total, 20 of the experimental NOEs are violated in the simulation. However, 13 violations involve only two residues, Thr15 and Lys21. For the remaining contacts, the NOE violation is 14%. Comparison of the violations in the first and in the second half of the simulation, or in 1 ns blocks shows that the

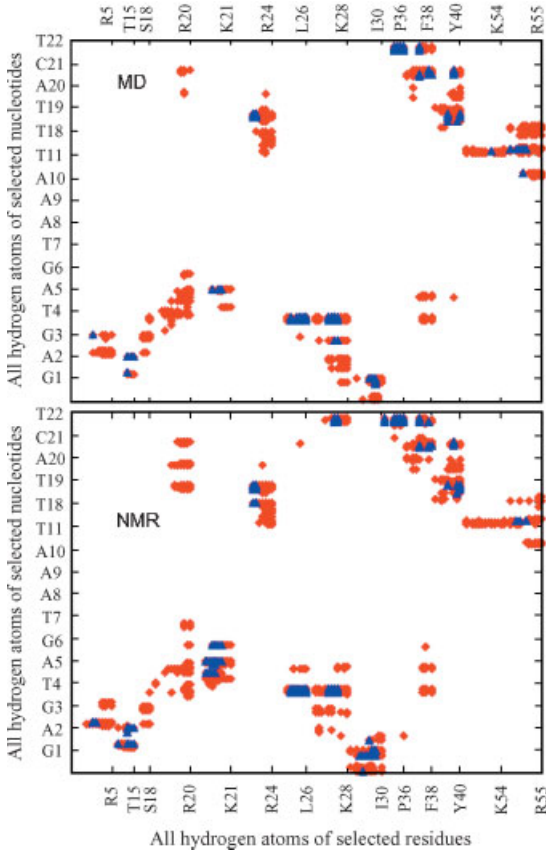

Figure 4. Contact map plots of selected protein–DNA interactions. Shown are only residues which are seen to interact in the NMR structure of the complex (Wojciak *et al.*, 1999). Hydrogen-hydrogen distances shorter than 5 Å are plotted in red. The experimental NOEs are plotted in blue.

protein–DNA interactions are generally the same throughout (but see the next section). Figure 4 shows a contact map plot for selected protein–DNA contacts observed in the NMR and MD ensembles. Although some contacts are lost during the simulation, some others are strengthened and the overall number of interactions changes little.

Overall, the simulation correctly captures the important protein–DNA interactions. The frequency of occurrence of a particular contact is not much different from its population derived from the NMR ensemble, and roughly matches the relative energetic contribution of the corresponding side chains. Most importantly, certain energetically important direct protein–DNA contacts that are difficult to determine from the NMR data alone due to side chain disorder have been identified (such as Arg5–Thy4, Arg20–Ade5/Gua6 and Arg55–Thy18/Thy19). Residues that were found to be energetically less important (e.g. Thr15 and Ser18) are found to make insignificant interactions, thus validating the mutational study. The role of water at the 'wet' interface appears significant. There are as much water-mediated contacts as direct protein–DNA contacts, as found before in other systems (Chillemi *et al.*, 2001). The pronounced dynamics of protein–water–DNA networks may help to lower the unfavorable entropy from freezing of side chain motions and tightly trapping of water molecules at the complex interface.

Structural re-organization at the major protein–DNA recognition site. Perhaps the most interesting feature of the simulation is the observed concerted shift of several protein–DNA and protein–water–DNA interactions taking place in the time interval 1.4–2.0 ns. In short, the following major changes in the hydrogen-bonding pattern are observed (Table 3). (i) The direct hydrogen bonds of Tyr40 with Ade20 are broken until new interactions eventually are formed with Cyt21. (ii) The Lys28–Gua3 base-specific hydrogen bond is temporarily disrupted. (iii) The base-specific side-chain interactions of Arg20 with Gua6 are replaced by side-chain-phosphate interactions with Ade5. (iv) The hydrogen bond involving Trp42 is strongest within the time window in which the base-specific hydrogen bond of Tyr40 with Ade20 is lost. (v) The hydrogen bond between Lys21 and Gua6 is broken after 1.4 ns. Although all the interactions that appear before or after the conformational change are also seen in (or inferred from) the NMR ensemble, the observations (i)–(v) merit explanations. There are two possible reasons. One is a major change in the secondary and/or tertiary interactions of the protein, DNA or both if protein–DNA contacts introduce conformational strain. Another possibility may be that subtle local movements occur, allowing penetration of water molecules replacing in part the direct protein–DNA contacts. We investigated both scenarios.

Table 3. Protein–DNA base-specific contacts and the number of 'cavity' and 'long-life' water molecules in time segments along the trajectory[a]

	Nucleotide			
Residue	1–1.5 ns	1.5–2.0 ns	2.0–2.5 ns	2.5–3.0 ns
Tyr40	Ade20 (Cyt21)	—	Cyt21	Cyt21
Trp42	Thy18/Thy19	Thy19	Thy19	—
Lys28	Gua3	Gua3	Gua3	Gua3
Arg20	Gua6	Ade5	Ade5	Ade5
Lys21	Gua6	—	—	—
Type of water				
'Cavity'	6.2 ± 3.8	5.0 ± 3.5	6.6 ± 2.8	6.1 ± 2.8
'Long-life'	4.5 ± 1.8	7.1 ± 2.0	6.3 ± 1.3	5.3 ± 1.3

[a] 'Cavity' waters are disconnected to the bulk, as defined in Billeter *et al.*, 1996.
'Long-life' waters appear in the interface for a total of at least 500 ps of the last 2 ns.

Structure of strand β and base-specific hydrogen bonds. Along the trajectory, no major structural reorganization of DNA occurs. The protein–DNA complex also shows no sign of a global structural change as judged from RMSD and R_g (Fig. 2 and Table 1). However, the secondary structure of strand β3 is partially disrupted early in the simulation, and reappears after about 1.4 ns. Therefore, we probed whether there is any correlation between the rearrangement in the side-chain-base hydrogen bond network and the structural integrity of the β3 strand. To this end, a series of constrained MD trajectories (a total of 15) was started either from snapshots where the protein–DNA hydrogen bonds are as in the starting structure (the NMR structure) but β3 is unstructured, or snapshots where the protein–DNA hydrogen bond pattern has rearranged and β3 strand is restructured (at 1 and 1.5 ps, respectively). A harmonic distance constraint with a force constant of 50 kcal mol^{-1} Å^{-2} was applied between the hydrogen-bonding atoms of Arg20, Lys21, Lys28, Tyr40 and Trp42 and DNA atoms individually or in several combinations. The distance was smoothly increased or decreased during 50 ps, so that the targeted hydrogen bonds formed or where disrupted. This was followed by a productive sampling for 50–400 ps. In two other simulations a similar procedure was applied upon the β2–β3 bridge formed by the backbone of Leu26 (β2) and Phe38 (β3), all protein–DNA interactions being unconstrained. Within the time scales tested, no relationship between the structure of the β3-sheet and the base-specific hydrogen bond pattern could be observed. Since the β2 and β3 strands are fully structured in all NMR conformers of the free protein, while the β2/β3 contacts are partially or even fully disrupted in about 20% of the protein–DNA conformers, we conclude that the abrupt rearrangement of the interactions around 1.4 ns thus appears to originate from local side-chain movements and is most likely linked to changes in the hydration pattern.

Hydration of the interface. As mentioned above, water molecules hydrating the INT–DBD–DNA interface are exchanging fast in general. We observe, however, 'long-life' waters as well. Arbitrarily, we classify a particular water molecule to be 'long-life' if it appears in the interface for a total of at least 500 ps of the last 2 ns. With this criterion, 14 water molecules are 'long-life', of which 6 ± 2 (average $\pm$ standard deviation) reside in the interface at any instance of the simulation (depicted as blue balls in Fig. 5; see also Table 3). An average of 6 ± 3 waters are buried in cavities in the interface of the protein–DNA complex at a given time (shown as red balls in Fig. 5; see also Table 3). These 'cavity' waters are only temporarily disconnected from the bulk; due to structural fluctuations the cavities occasionally 'open' and communicate with the bulk through a chain of waters. The number of waters in cavities and 'long-life' waters is approximately conserved throughout the simulation. However, their distribution qualitatively differs in the time window where the conformational switch takes place.

Up to about 1.4 ns simulation time the waters solvating the interface are mainly located at the periphery of the binding site; only few penetrate deeply into the major groove. Their structural role may vary. Some connect the backbones of protein and DNA, others solvate protein or DNA groups. Networks of hydrogen bonds are rare. The only side chain participating in extensive contacts with interfacial waters is Lys28, which reaches out to the phosphates of Ade2 and to Gua3 via two-water bridges. During the conformational change, clusters of water molecules form and penetrate the center of the major groove. Direct base-specific hydrogen bonds are being replaced by single or multiple water-mediated interactions. Water chains or bifurcated clusters are often observed. Most interestingly, in many snapshots INT–DBD side chains are simultaneously water-bridged to two DNA groups representing the hydrogen bond partners mostly populated before and after the structural transition. It appears that the switch of the interaction partners is a stepwise process involving partial preservation of both contacts through water networks. Such a 'compensation' mechanism would presumably confer binding energy during the dynamic switch between alternating, stronger direct hydrogen bonds. In the last portion of the simulation waters are expelled from the center of the binding site toward the edge of the major groove and play qualitatively the same role as in the early stages before the re-arrangement.

The seemingly coupled changes (fluctuations) of protein–DNA interactions and interface hydration are intriguing. It is likely that the observed rearrangement mirrors 'real' fluctuations between isoenergetic states of the solvated protein–DNA complex. Throughout the simulation there is no detectable change in any of the global structural characteristics of the complex, such as solvent accessibility, radius of gyration, RMSD from the starting structure and energy. Furthermore, the bonding patterns in the initial and final parts of the trajectory do not conflict with NMR and biochemical data. For example, the major contributor to

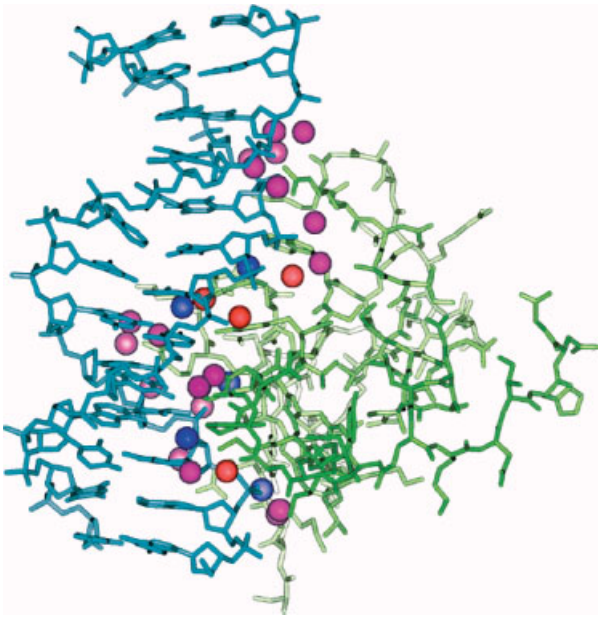

Figure 5. Typical pattern of hydration of the INT–DBD–DNA interface. A snapshot taken at 2.5 ns is used for illustration. Water molecules which are simultaneously closer than 2.6 Å from at least one protein atom and one DNA atom are shown as balls. 'Long-life' waters and 'cavity' waters are colored in blue and red, respectively (see text for definitions). Other transient water molecules are colored in magenta. Protein (green) and DNA (blue) are shown in a stick representation.

affinity, Tyr40, is hydrogen-bonded to the base of either Ade20 or Cyt21, the latter interaction being the dominating interaction according to NMR data. The simulation correctly describes its behavior. In the starting structure, the hydroxyl group of Tyr40 donates a hydrogen bond to Ade20. In the course of the simulation, however, the side chain becomes more frequently interacting with Cyt21. The transient loss of the base-specific Lys28/Gua3 bond is consistent with the absence of this interaction in a number of NMR conformers calculated without artificial hydrogen bonds as input restraints and without electrostatic and hydrogen-bonding terms in the energy function (Wojciak *et al.*, 1999). The fluctuating direct and water-mediated interactions of Arg20 in the major groove substantiate the observed insufficient number of proximal hydrogen atoms to define the side-chain conformation. The simulation demonstrates the 'active' role of water molecules in formation of the binding interface. It is worth noting that the trajectory was started from the 'dry' structure of the complex. Although the number of water molecules at the interface equilibrated very early (after $\sim$10–20 ps), it is likely that the interaction network in the second half of the simulation is a more representative structural description of the energetic determinants of the INT–DBD–DNA complex. Finally, the simulation validates our prediction that the interface is substantially hydrated, as we have previously proposed based on calorimetric analysis of the binding heat capacity and enthalpy (Milev *et al.*, 2003a,b).

Altogether, the simulation results suggest that structural flexibility is essential for the formation of the sequence-specific INT–DBD complex. Close positioning and optimal mutual orientation of side chains and DNA groups is achieved in part by reorganization of intrinsically flexible segments of the protein domain. Many positively charged side chains are involved in a highly dynamic network of hydrogen bonds and van der Waals interactions with DNA. Isolated water molecules and water clusters help in optimizing of polar interactions. Backbone and side chain flexibility at a 'loose' binding site is entropically favored and might thus counteract the entropic loss from dehydration, albeit incomplete.

Acknowledgements

We thank Emanuele Paci for advice, Robert T. Clubb for providing experimental NOEs, Urs Haberthür and Terence Hale for computer support, and Hans Rudolf Bosshard for critical reading of the manuscript. This work was supported in part by the Swiss National Science Foundation and in part by the Swiss National Center for Competence in Research in Structural Biology (NCCR). The coordinate files of the simulations are available upon request (e-mail: iljel@bioc.unizh.ch).

REFERENCES

Billeter M, Qian YQ, Otting G, Muller M, Gehring W, Wuthrich K. 1993. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. *J. Mol. Biol.* **234**: 1084–1093.

Billeter M, Guntert P, Luginbuhl P, Wuthrich K. 1996. Hydration and DNA recognition by homeodomains. *Cell* **85**: 1057–1065.

Bogusz S, Cheatham T, Brooks B. 1998. Removal of pressure and free energy artifacts in charged periodic systems via net charge corrections to the Ewald potential. *J. Chem. Phys.* **108**: 7070–7084.

Brooks BR, Bruccorleri RE, Olafson BD, David JS, Swaminathan S, Karplus M. 1983. CHARMM: a program for macromelecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **4**: 187–217.

Chillemi G, Castrignano T, Desideri A. 2001. Structure and hydration of the DNA-human topoisomerase I covalent complex. *Biophys. J.* **81**: 490–500.

Clewell DB, FIlanagan SE. 1993. In *Bacterial Conjugation*, Clewell ED (ed.). Plenum Press: New York.

Clewell DB, Jaworski DD, Flannagan SE, Zitzow LA, Su YA. 1995. The conjugative transposon Tn916 of *Enterococcus faecalis*: structural analysis and some key factors involved in movement. *Dev. Biol. Stand.* **85**: 11–17.

Connolly KM, Wojciak JM, Clubb RT. 1998. Site-specific DNA binding using a variation of the double stranded RNA binding motif. *Nat. Struct. Biol.* **5**: 546–550.

Connolly KM, Ilangovan U, Wojciak JM, Iwahara M, Clubb RT. 2000. Major groove recognition by three-stranded beta-sheets: affinity determinants and conserved structural features. *J. Mol. Biol.* **300**: 841–856.

Darden TA, York DM, Pedersen LJ. 1993. Particle mesh Ewald: an *N*.log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**: 10089–10092.

Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**: 1097–1113.

Duan J, Nilsson L. 2002. The role of residue 50 and hydration water molecules in homeodomain DNA recognition. *Eur. Biophys. J.* **31**: 306–316.

Dyson HJ, Wright PE. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**: 54–60.

Gorfe AA, Jelesarov I. 2003. Energetics of sequence-specific protein–DNA association: computational analysis of integrase Tn916 binding to its target DNA. *Biochemistry* **42**: 11568–11576.

Gruschus JM, Ferretti JA. 2001. Quantitative measurement of water diffusion lifetimes at a protein/DNA interface by NMR. *J. Biomol. NMR* **20**: 111–126.

Janin J. 1999. Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. *Struct. Fold. Des.* **7**: R277–R279.

Jen-Jacobson L, Engler LE, Jacobson LA. 2000. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **8**: 1915–1923.

Jorgensen WL, Charndrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of Simple potential Functions for Simulating liquid water. *J. Chem. Phys.* **79**: 926–935.

Labeots LA, Weiss MA. 1997. Electrostatics and hydration at the homeodomain–DNA interface: chemical probes of an interfacial water cavity. *J. Mol. Biol.* **269**: 113–128.

Lynch TW, Kosztin D, McLean MA, Schulten K, Sligar SG. 2002. Dissecting the molecular origins of specific protein-nucleic acid recognition: hydrostatic pressure and molecular dynamics. *Biophys. J.* **82**: 93–98.

MacKerell JAD, Bashford D, Bellott M, Dunbrack JRL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J,

Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. 1998. Empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**: 3586–3616.

Madrid M, Lukin JA, Madura JD, Ding J, Arnold E. 2001. Molecular dynamics of HIV-1 reverse transcriptase indicates increased flexibility upon DNA binding. *Proteins* **45**: 176–182.

Milev M, Gorfe AA, Karshikoff A, Clubb RT, Bosshard HR, Jelesarov I. 2003a. Energetics of sequence-specific protein–DNA association: conformational stability of the DNA binding domain of integrase Tn916 and its cognate DNA duplex. *Biochemistry* **42**: 3492–3502.

Milev S, Gorfe AA, Karshikoff A, Clubb RT, Bosshard HR, Jelesarov I. 2003b. Energetics of sequence-specific protein–DNA association: binding of integrase Tn916 to its target DNA. *Biochemistry* **42**: 3481–3491.

Nadassy K, Wodak SJ, Janin J. 1999. Structural features of protein–nucleic acid recognition sites. *Biochemistry* **38**: 1999–2017.

Otting G, Liepinsh E, Halle B, Frey U. 1997. NMR identification of hydrophobic cavities with low water occupancies in protein structures using small gas molecules. *Nat. Struct. Biol.* **4**: 396–404.

Otwinowski Z, Schevitz RW, Zhang RG, Lawson CL, Joachimiak A, Marmorstein RQ, Luisi F, Sigler PB. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**: 321–329.

Pedersen JS, Otzen DE, Kristensen P. 2002. Directed evolution of barnase stability using proteolytic selection. *J. Mol. Biol.* **323**: 115–123.

Rader AJ, Hespenheide BM, Kuhn LA, Thorpe MF. 2002. Protein unfolding: rigidity lost. *Proc. Natl Acad. Sci. USA* **99**: 3540–3545.

Reddy CK, Das A, Jayaram B. 2001. Do water molecules mediate protein–DNA recognition? *J. Mol. Biol.* **314**: 619–632.

Rykaert JP, Ciccotti G, Berendesen HJC. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**: 327–341.

Schwabe JW. 1997. The role of water in protein–DNA interactions. *Curr. Opin. Struct. Biol.* **7**: 126–134.

Scott JR, Churchward GG. 1995. Conjugative transposition. *A. Rev. Microbiol.* **49**: 367–397.

Sen S, Nilsson L. 1999. Structure, interaction, dynamics and solvent effects on the DNA–*Eco*RI complex in aqueous solution from molecular dynamics simulation. *Biophys. J.* **77**: 1782–1800.

Sunnerhagen M, Denisov VP, Venu K, Bonvin AMJJ, Carey J, Halle B, Otting G. 1998. Water molecules in DNA recognition I: hydration lifetimes of trp operator DNA in solution measured by NMR spectroscopy. *J. Mol. Biol.* **282**: 847–858.

Tsui V, Radhakrishnan I, Wright PE, Case DA. 2000. NMR and molecular dynamics studies of the hydration of a zinc finger–DNA complex. *J. Mol. Biol.* **302**: 1101–1117.

Wojciak JM, Connolly KM, Clubb RT. 1999. NMR structure of the Tn916 integrase–DNA complex. *Nat. Struct. Biol.* **6**: 366–373.

