Theor Chem Acc (2000) 104:96–108
DOI 10.1007/s002149900106

Theoretical
Chemistry Accounts

Recognition determinants in a T4 ← G4 mutant derived from a 5′-GCGTGGGCGT-3′ oligomer in a zinc finger 268–DNA complex.

A molecular dynamics study of the fully charged complex in water

G. Roxström, M. Paulino, O. Tapia

Department of Physical Chemistry, Uppsala University, Box 532, 85121 Uppsala, Sweden

Received: 21 June 1999 / Accepted: 19 October 1999 / Published online: 14 March 2000

Abstract. The 5′-GCGTGGGCGT-3′ (T4) oligomer found in the zinc finger 268–DNA complex was mutated into the sequence 5′-GCGGGGGCGT-3′ (G4). A 3D model was constructed from the T4 sequence using an X-ray structure as a template. Molecular dynamics simulations were used to test the thermal stability of the model. A 500-ps trajectory was obtained for the fully charged complex in water using GROMOS87. The complex and the G4 sequence are found to have dynamically stationary behavior. Comparisons made with a previous T4 sequence molecular dynamics simulation show both systems have similar thermal stability. The structure of DNA appears to be maintained by its global interactions with the protein although the mutated site does not contribute with its full potential for binding. The protein structure shows some small differences compared to the T4 simulation. The simulation provided evidence for the role of a chloride ion interacting with the protein and helping in the recognition process.

Key words: Molecular dynamics simulations – Zinc finger – DNA-binding proteins – Protein – DNA contacts

1 Introduction

Molecular modeling and molecular dynamics (MD) simulations are used to obtain information concerning the structure and fluctuations of a DNA oligomer bound to an 85-residue-long zinc finger (Zif) protein in a bath of simple-point-charge (SPC) water and counterions. Although the structure and activity cannot be dissociated, simulation approaches can be used to obtain information that is complementary to experiments. Special attention is given to the study of protein–DNA recognition determinants measured by atom–atom fluctuation patterns, while the question concerning the role of cations in molecular recognition [1, 2] is outside the scope of this paper.

MD simulations of highly charged subsystems interacting among themselves and counterions have represented, according to a review by Louise-May et al. [3], a challenge to which partial solutions have recently been found [4, 5]. The structural equilibrium of DNA represented by different force fields has been studied by Feig and Pettitt [5]. Successful simulations have been published for DNA oligomers in double- [4] and triple-helical arrangements [3, 5]. For protein–DNA complexes with fully charged molecules results are more sparse [6]. A symmetrized counterion model for the full Zif complex [7, 8] showed a reasonable degree of stability for a 1.1-ns trajectory. Although in the last few years collective variable modeling of nucleic acids has been developed [9], atomic resolution simulations are still a complementary means to achieve a description taking into account atomic contacts and the sequence-dependent heterogeneity that contributes to protein–nucleic acid recognition. Two strategies can be found in the literature concerning the protein–DNA representation: one where special features and new accurate parameter sets are introduced to represent the system [5]; the other, used here, aims to retain the general atom potential representation valid for any model molecule and is able to describe the system from a vacuum to condensed phases. In the spirit of the latter strategy, we are examining the results obtained with the parameter set found in GROMOS87 [10] and GROMOS90 [11]. This approach, extensively used in our group, has been successful in representing proteins in vacuo [12, 13] and in solution and, without further change, it is being used to study protein–DNA complexes [7, 8]. The numerical results suggest that, besides the treatments of ions interacting with phosphate groups (P ions), the generic force field used in GROMOS87 does not require any external constraints on the solute on a nanosecond time scale to get reasonable fluctuation

Permanent address: M. Paulino, Department of Quantum Chemistry, Facultad de Química, Universidad de la República, 11800 Montevideo, Uruguay

Correspondence to: O. Tapia

patterns. It is worth noting that the SPC solute interaction in the early version of this model was at fault [14]. After correcting solvent-to-solute interactions that, pictorially speaking, transformed what was before a hydrophilic carbon atom into a hydrophobic one without affecting water-to-water interactions [14], an effort to simulate the coion distribution was made by using particular forces on the P ions. Coions are to be understood as the presence of an ionic atmosphere [15] that might be surrounding the complex in solution.

For the DNA in Zif268 [16] two consensus sequences have been reported: 5′-GCGTGGGCGT-3′ and 5′-GCGGGGGCGT-3′ [17], referred to as T4 and G4 sequences, respectively. We consider the Zif268–DNA complex (Zif complex) for which X-ray data is available [18, 19] and for which MD simulations of the fully charged complex in water have been carried out [7, 8]. The full structure is illustrated in Fig. 1. The G4 sequence coordinates were obtained by the molecular replacement

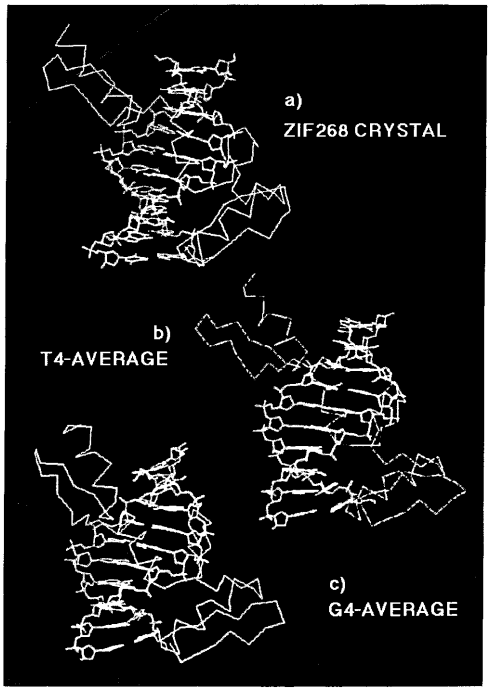

Fig. 1. **a** The first (T4) model that was constructed from the X-ray structure of the zinc finger (Zif) 268 peptide–DNA complex solved by Pavletich and Pabo [18]. The crystal structure of DNA consists of ten base pairs plus an extra base in the 5′ end of each strand [18]. The two extra bases were removed in the T4 model. The protein domain (Zif), which is 85 residues long, contains three tandemly repeated Zif motifs (named F-1, F-2 and F-3) that are arranged in a semicircular (C shaped) structure that fits into the major groove of DNA. Each Zif consists of an antiparallel β sheet and an α helix that are held together by a zinc ion (coordinating two cysteines and two histidines) and a hydrophobic core. The five linker residues between the fingers are denoted L_{12} and L_{23} respectively. **b** T4 simulation: average and regularized structure of 1100 ps of simulation. The two submolecules are separately least-squares fitted against crystal structure **a** The average is taken over 110 frames (each 10 ps). **c** G4 simulation: average and regularized structure of 500 ps of simulation. The two submolecules are separately least-squares fitted against crystal structure **a** The average is taken over 50 frames (each 10 ps)

T4 ← G4 in the 10 base pairs found in the Zif complex reported by Pavletich and Pabo [18] and Elrod-Eriksson et al. [19]. One of the aims of this work was to test the stability of the molecular model to thermal fluctuations. To this end, a 500-ps-long trajectory was obtained and analyzed. Here, we report changes in the fluctuation pattern of recognition sites and intramolecular interactions in both partners to help sense the effects of local DNA–protein interactions. The results are compared with our previous 1.1-ns trajectory of the Zif complex (i.e. the T4 sequence) [7, 8]. The model system used here contains about 5000 SPC water molecules [20–22] and so the effect of an ionic atmosphere was introduced via a weak position constraint (force constant 0.58 kcal mol^{-1} $Å^{-2}$) acting on the P ions after relaxation as described in our previous work [4]. This model does not solve the general problem of representing the ionic atmosphere around the fully charged DNA although, in practice, it yields reasonable results. Questions such as those raised by univalent cations [2] that appear to mediate specific interactions with DNA thereby helping mediate recognition [23] are not treated here.

From a modeling perspective, simulating the G4 sequence may be of interest for the study of another family of Zif proteins, a two-zinc finger protein related to the high-affinity binding site of the negative regulator mediating carbon catabolism repression in *Aspergillus nidulans* (CreA) [24] for which the consensus DNA sequence can be approximated by the G4 sequence. The results found here may help modeling of this protein–DNA complex.

2 Model and computation details

2.1 Setup and equilibration

The computer mutation of DNA in its complex with the Zif protein was done with mdFRODO [25]. The refine option was used to geometry-optimize the mutated base pair in order to sense the initial effects of local DNA–protein interactions. The initial geometry of the P ions (eight Na^+) was taken from the initial positions used for the T4 sequence simulation [7, 8]. The cubic box, 7.056-nm length, and octahedral periodic boundary conditions were used to generate a 5175 SPC water sample [26]. The solvent configuration, generated with the PROBOX routine of GROMOS87 [10], was regularized with an energy minimization, keeping the solute and ion positions restricted. This step converged after 84 steps. The full complex including P ions has a global state of charge of +1. This charge was neutralized with a chloride ion, positioned with the help of PROION [10], which put it between Arg70 and Arg80 belonging to the third Zif. Histidines are not protonated while arginines and lysines are positively charged, and aspartates and glutamates each have negative charge. The cutoffs are the same as used previously: 12 and 8Å for Coulomb (long-range) and Lennard-Jones interactions, respectively. This system was submitted to a 5-ps MD simulation with the solute and ions restrained to construct a starting point with the same structure as in the T4 simulation. Full

solvent relaxation was carried out before computing 10 ps of equilibration for the whole system. For such large systems, the solute temperature equilibrates rather fast [7, 8]. The global system also equilibrates fast, although potential-energy fluctuations between solute and solvent are manifest. For the sake of comparison with the T4 simulation, the trajectory was analyzed after the 10 ps used to equilibrate the whole system; hence, any nonequilibrium behavior can be sensed directly.

2.2 Model differences

The system simulated differs from the previously reported one on few points:

1. Mutation of the thymine at position 4 by a guanine (T4 ← G4).
2. The global charge of +1 for the zinc coordination motif used with the T4 sequence is here scaled down to zero by using +1 on the zinc and −1/2 on each sulfur atom of the coordinated cysteines. This gives a multipolar interaction of zinc with its surroundings that is weaker than the preceding one.
3. Instead of four randomly positioned chloride ions (positioned far away from the surface of the solute) there is only one initial chloride ion positioned between Arg70 and Arg80. The chloride ion was allowed to move freely in the simulation.

2.3 Binding site definition

Measuring from the 5′ end, the binding site for finger 3 (F-3) corresponds to the base G1-C2-G3 sequence, finger 2 (F-2) recognizes the sequence G4-G5-G6 (i.e. the mutated binding site), while finger 1 (F-1) interacts with G7-C8-G9. As indicated in Scheme 1, the recognition sites are named m3, S6 and S3 [27], respectively.

	CONTACTS		
	F-1	**F-2**	**F-3**
S 3	Arg18--->G9	Arg46--->G6	Arg74--->G3
S 6	Glu21 C8	His49···>G5	Glu77 C2
m 3	Arg24--->G7	Thr52 G4	Arg80--->G1

Scheme 1

2.4 Initial structure of the mutated binding site

The G4-G5-G6 sequence interacting with F-2 in the initial model is shown in Fig. 2a. It involves the direct S3

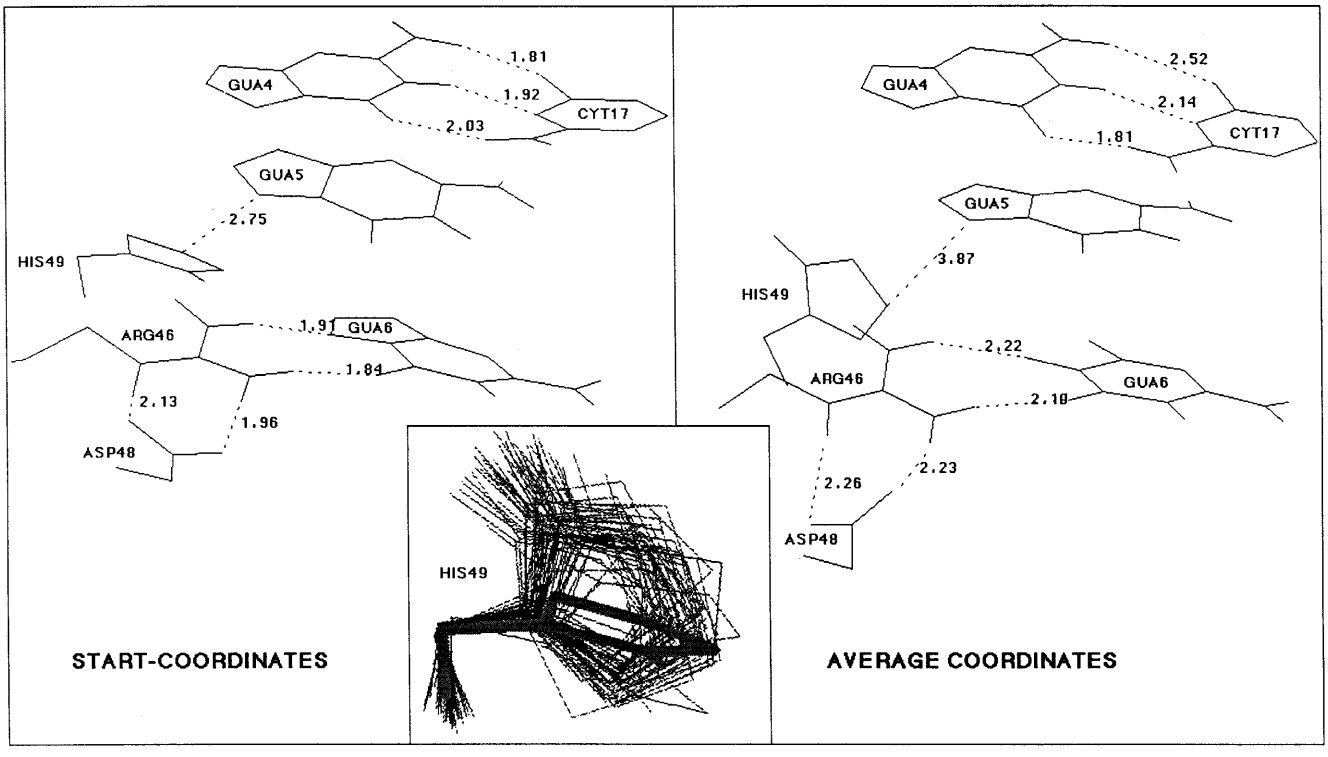

Fig. 2a–c. The network of interactions involved in the contacts of F-2 is here illustrated by the side chains of residues Arg46, Asp48 and His49 in the protein and the two bases in G5 and G6 in the guanine-rich strand of DNA. The mutated bases in G4 and C17 are also illustrated. **a** The start coordinates after mutation in mdFRODO [25] and energy minimization as described in the text. **b** The 500-ps time average (not regularized) coordinates yield contact distances below 2.30 Å, except for the contact with His49 and the base pair contact between N2 of G4 and O2 of C17. However, the shortening of the distance in the third contact of the same base pair leads to an opening and a negative stretch [31] with respect to the setup coordinates (**a**). **c** An overlay of 50 snapshots (each 10 ps) of the side chain of His49. Initial orientation is colored *red*

contact between Arg46 and Gua6. The weaker S6 contact between Gua5 and His49 (dotted arrow in Scheme 1) is left here without initial hydrogen bonding. In the X-ray structure, His49 is oriented so that the Nε hydrogen interacts with the N7 of Gua5. Here, the hydrogen was put at Nδ, thereby weakening the interaction with Gua5. The m3 contact is absent, while in F-1 and F-3, the m3 contacts are made via arginines; the corresponding position in F-2 (residue number 6 in the α helix) is occupied by a threonine (Thr52). Thr52 is far away and only rather weak van der Waals interactions may be effective.

Some intramolecular interactions in the protein have been suggested to act as stabilizers for the recognition determinants. In Fig. 2a, Asp48 is near Arg46 and may lead to stabilization of the S3 interaction. This is an issue to be monitored in the simulation. Note also that the mutated base pair (Gua4/Cyt17) is rather far away from this important S3 contact.

To sum up the initial model, nothing has been done to enhance the interactions of finger F-2 with the DNA. We assume that the dominant interaction is made between Arg46 and Gua6 (S3 contact), probably assisted by Asp48 as was the case in the T4 sequence simulation.

3 Results and discussion

3.1 Average structure features of the mutated binding site

In Fig. 2b, the 500-ps time-average structure (nonregularized coordinates) shows that the S3-binding and the S3-stabilizing salt bridges between Asp48 and Arg46 are conserved. The same figure illustrates that the S6 contact is not re-formed. His49 has an average orientation that is tilted 144° about the Cβ—Cγ bond with respect to the initial structure. Pavletich and Pabo [18] noticed that a 180° rotation about the Cβ—Cγ bond in His49 would allow direct contact between the Nε hydrogen and the O6 atom (instead of the N7 atom) in the same guanine. In the simulation (initial Nε—N7 distance of 2.8 Å) the distance increases and so no hydrogen bonds could be formed with either N7 or O6.

3.2 Contact distance fluctuations in the mutated binding site

The Gua4-Cyt17 Watson–Crick contacts fluctuating about a value not far from the initial configuration are shown in Fig. 3a. As is apparent from Fig. 2b the hydrogen-bonding pattern seems to have a fluctuation in the opening angle as defined in the work of Lavery and Sklenar [28].

The S3 contact of F-2 (Fig. 3b) and the S3-stabilizing intramolecular contact between Arg46 and Asp48 (Fig. 3c) fluctuate around their initial values. In the vicinity of the S3-stabilizing contact there is one contact between Asp48 and Cyt14 (the complementary base to Gua7) that may contribute to the recognition. Pavletich and Pabo [18] pointed out that the geometry of these

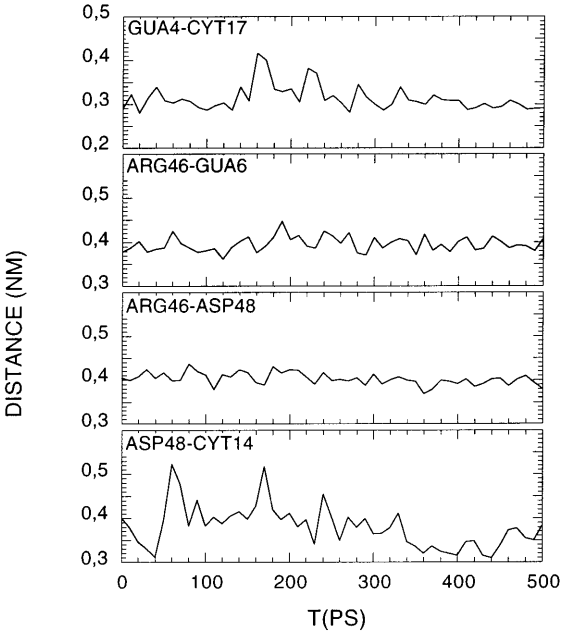

Fig. 3. **a** The Watson–Crick interactions in the mutated base pair (G4:C17) were measured as the distance-average over the three hydrogen bonds. Since hydrogen atoms are not visible in the X-ray data the distances are measured between the following (heavy) atoms N2(G4):O2(C17), N1(G4):N3(C17) and O6(G4):N4(C17). **b** The S3 contact was monitored as the average distance between the CZ of Arg46 and the two acceptor groups (O6 and N7) in G6. **c** The S3-stabilizing intramolecular interaction between Arg46 and Asp48 was measured as the average between the two hydrogen-donating groups (CZ and NE) in Arg46 and the accepting CG in Asp48. **d** The S3-stabilizing intermolecular contact between Asp48 acid and C14 was detected by measuring the distance between the donating N4 of C14 and the accepting CG in Asp48

hydrogen bonds was not favorable and assumed they do not play a major role in the recognition. This contact is totally lost after 60 ps (Fig. 3d) but reappears after 350 ps.

The structural fluctuations of His49 are shown in Fig. 2c. This is an overlay of 50 snapshots (from each 10 ps) of the side-chain coordinates. The picture indicates that the histidine does not rotate freely in spite of the fact that there is no hydrogen bonding with N7(Gua5). The time evolution of rotational angle about the Cβ—Cγ bond of His49 (not shown) reveals that the space available to His49 does not allow it to rotate freely in our model. After the first 80 ps the imidazole ring librates at a value between 120° and 180° from the initial site without hydrogen bonding to N7 (and there is no hydrogen-bond acceptor close enough to attract the Nδ hydrogen of His49). In the T4 simulation, His49 found its way back to a regime of stationary fluctuation close to the initial orientation after 700 ps. It should be noted that in the T4 sequence the S6 contact was complemented by the direct interaction between His49 and Thy4. This contact was very well conserved during the T4 simulation and it might have contributed to the recovery of the original orientation of His49; however, calculations of the potential of the mean force by Norberg and Nilsson [29, 30] show such direct van der Waals

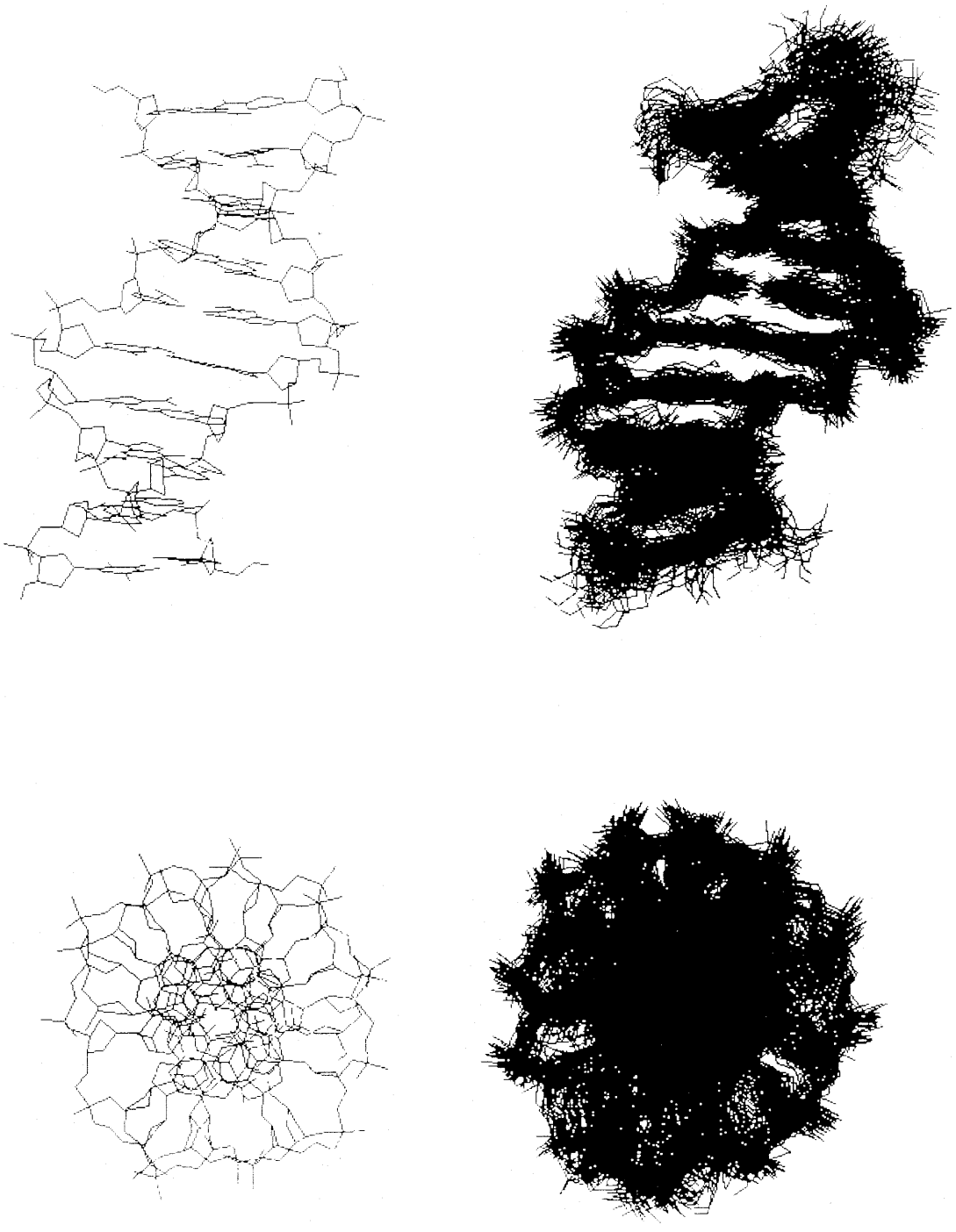

Fig. 4. a The DNA structures (*right*) during the simulation illustrated by 50 frames, each 10 ps, least-squares fitted to all DNA atoms in the starting structure. The reference starting structure is on the *left*. **b** Time series of global base–base parameters for G4 and its complementary base (*solid line*) and the average taken over all ten base pairs (*dashed line*). The averaged X-ray value is indicated in *parentheses*. **c** Time series of the global inter-base-pair parameters (*first three rows*) and the global base-pair axis parameters (*last two rows*) for G4/C17 (*solid line*) and the all-ten average (*dashed line*)

interaction with the methyl group of thymine to be rather weak. These results indicate that in the T4 sequence the rotation of the imidazole ring is hindered because of the bulky methyl group rather than the van der Waals attraction to it.

As the G4 sequence model was biased towards a weak S6 contact, the simulation suggests that possible hydrogen bonding between Nε and N7(Gua5) would marginally contribute to the site stability.

3.3 Distance fluctuations in the nonmutated binding sites

The recognition contact distances involving F-1 and F-3 (i.e. the two nonmutated binding sites) fluctuate very

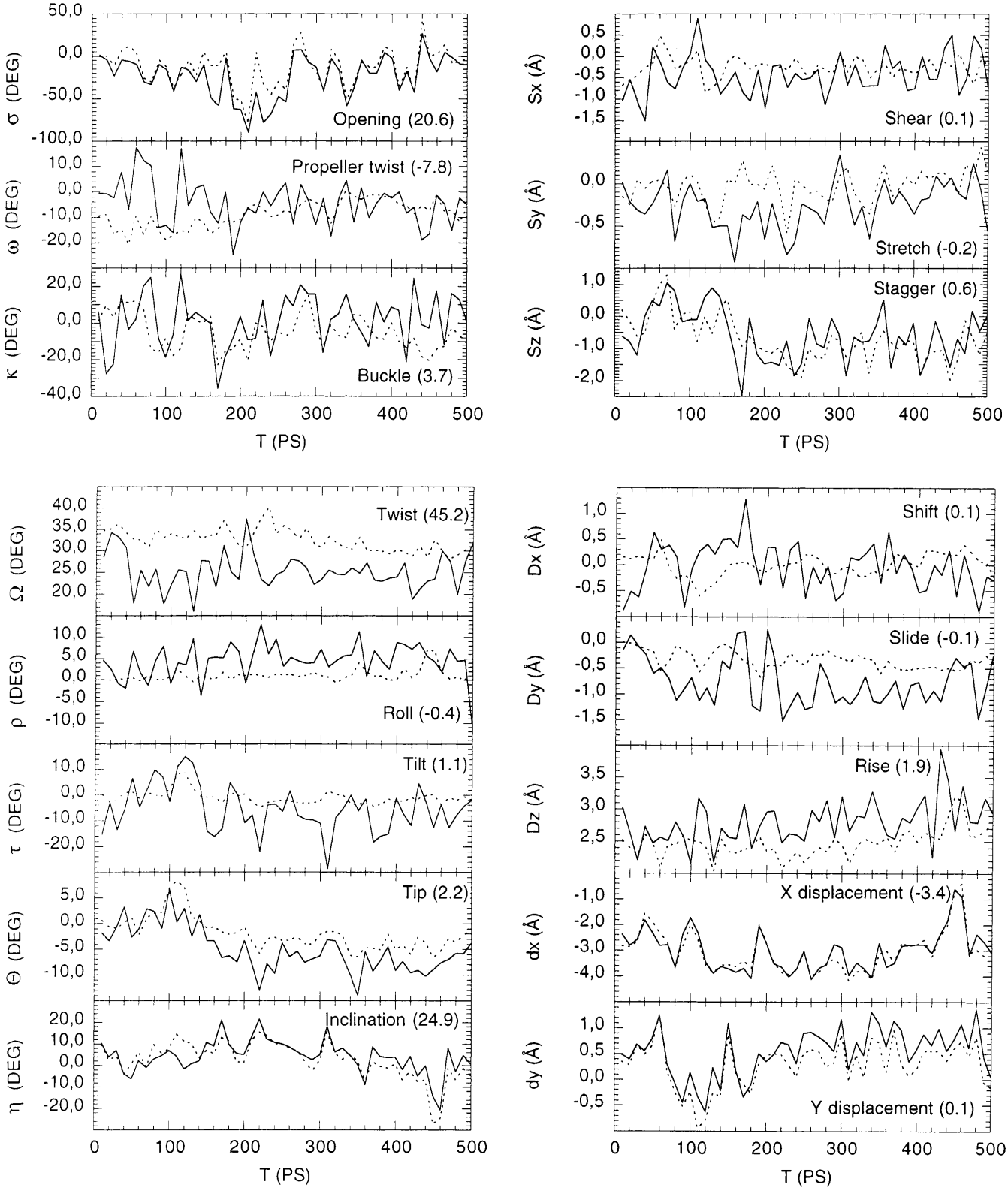

Fig. 4. (Continued)

similarly to the T4 simulation: both trajectories produce weak, direct S3 interactions with F-1 (i.e. the average distances between heavy atoms are above 4.5 Å) and stronger S3 contact with F-3. It is also significant the total lost of the intramolecular S3-stabilizing contact in F-1, shown in the T4 simulation as well as in the G4 simulation, while the same contact distances in F-3 fluctuate with remarkably low amplitudes (lasting 400 ps in the T4 simulation and at least 500 ps in the G4 simulation).

The S3 contact in F-3 involves the base neighbor of the mutated Gua4; this seems to have no effect in the trend. Note that the distance between Arg74 and Asp76 increased after 400 ps in the T4 simulation. This change is not observed here, which suggests that this local structure may be less strained than the T4 structure.

The m3 contact was totally lost for both F-1 and F-3 in the T4 simulation. The same thing happens in this simulation for F-1, while the m3 contact distance for F-3 keeps on fluctuating around 4 Å. A close analysis shows that the difference may be assigned to the position of the chloride ion. In the G4 simulation it is located between Arg70 and Arg80 and fluctuates there at about 4-Å distance. No constraint acts on this ion. The result appears to be a rather stable m3 interaction for F-3. In contrast, in the T4 simulation, no chloride was there to hold the side chains and the m3 contact vanished. A parallel G4 simulation where the chloride ion drifted away resulted in the same behavior as in the T4 simulation.

3.4 Global properties

Figure 1 depicts the optimized crystal form, the T4 average structure and the G4 average coordinates. The 500-ps-averaged and regularized coordinates of G4 (Fig. 1c) elicit a well-conserved global structure for both the DNA (all atoms) and the protein (α carbons) when compared to the crystal structure (Fig. 1a). Comparison with the 1.1-ns average structure of T4 (Fig. 1b) shows that the two simulations behave about the same in this respect.

Fig. 5. **a** The full complex, overlay of 50 frames, one each 10 ps, from the simulation. For DNA, all atoms defined in the molecular dynamics program are drawn, while the protein structure is illustrated by the α carbons. Each frame is least-squares fitted to the crystal structure with respect to the DNA atoms only. **b** Fifty frames of the protein. The structures are least-squares fitted to the F-2 atoms. The pictures illustrate the α carbons

3.4.1 Behavior of DNA

Figure 4 illustrates how the DNA molecule fluctuates close to an average structure. The time series of the root-mean-square deviation (RMSD) from the starting structure of DNA (not shown) ends up with a value below 1.6 Å. We used the program Curves [28] to get more refined information.

The time series for complementary base parameters (opening, propeller twist, buckle, shear, stretch and stagger) are displayed in Fig. 4b. Except for the propeller twist, the behavior of the average and the mutated base pair is strongly correlated. During the first 150 ps the propeller twist shows relatively large fluctuations: at least they are different when compared to the value averaged over all pairs. As already noted by Pabo and coworkers, the X-ray structure is a special variant of the canonical B form ($B_{enlarged\ groove}$). The advantage of this MD simulation is to show the high degree of stability shown by these parameters when plotted as a function of time. The modeling has not introduced structural faults. This can be appreciated in the analyses of neighboring base pair parameters (see later).

A significant analysis is provided by global inter-base-pair parameters such as twist, roll, tilt, tip and inclination [31]. The time series for G4 and its complementary base are reported together with the average value taken over the 10 base pairs in Fig. 4c. We also include the global base-pair axis parameters: shift, slide, rise, x-displacement and y-displacement. The X-ray value (average) is indicated in parentheses.

The X-ray structure is not a canonical B structure. The twist value is about 45, which is 10 points larger that the standard B form (35.9). The twist related to our mutated pair fluctuates around 25, while the average is higher that this figure and it tends towards a B-form value. The roll for the X-ray structure is negative (about −0.4); the average roll fluctuates in the vicinity of zero, while the mutated pair fluctuates slightly above the average value. This angle in the simulation is above the standard B form (about −1.17).

3.4.2 Properties of the protein

Fifty frames of the complex (all atoms in DNA and α carbons in the protein), taken from each 10 ps, are shown in Fig. 5a. The picture illustrates well-conserved secondary and tertiary structures of the system, although the protein seems to fluctuate more than the DNA. The final RMSD for the full protein is 2.2 Å, while each Zif has a final value below 1.5 Å. The protein coordinates are least-squares fitted to the F-2 atoms in Fig. 5b. The F-1/F-3 motion can be pictorially described as if the Zifs were hugging the DNA. From the Zif protein in the absence of DNA (with and without water, data not shown) the motion is still present, although with a smaller amplitude. The presence of SPC water molecules enhances this motion as the solvent prevents implosion, which is a common feature of vacuum simulations.

Comparisons with the T4 simulation reveals that the finger–finger orientations were most likely affected by the rescaling of the coordination sphere from +1 to 0. It should be noted that the relative distances between the zinc coordination spheres are larger than the Coulomb cutoff. The effect may be due to interactions with the P ions or to intramolecular interactions with the lysines found in the linkers.

The linker between F-1 and F-2 (L_{12}) consists of the residues Thr-Gly-Gln-Lys-Pro, with the lysine having a net charge of +1. The five residues between fingers F-2 and F-3 (L_{23}) make a H-C link [32] or a Kruppel-type linker (Thr-Gly-Glu-Lys-Pro) [33] that includes a salt bridge between the lysine and the glutamate. These structures are important as they convey flexibility to the external fingers (F-1 and F-3) [34, 35] and may be responsible for the different behavior in the T4 and G4 simulations. We have found that the relative finger–finger displacement is conspicuously related to rotation about the linker backbone and the first residues of the β sheet. Moreover, as the zinc coordination sphere interacts with the charged residues in the linker (distances shorter than the Coulomb cutoff), one could expect different behavior of the fingers when the zinc net charge is rescaled.

Another reason for the different relative orientation of F-1 and F-3 is probably the change in repulsion from a symmetrized positively charged counterion (Na^+) that is positioned between F-1 and the DNA. The ion was put there since the crystal structure showed a very small contact surface between the two submolecules. We propose that F-1 might have been molded by crystal interactions so that when freed in the simulation it could move away. In the T4 simulation the repulsion force between Zn^{2+} and the counterion was large and since the counterion in the long run cannot move freely, the finger finally found an alternative conformation that had less contact with DNA. In the present simulation the counterion had the same constraint, but rescaling the charge of zinc led to a less stressed situation, where F-1

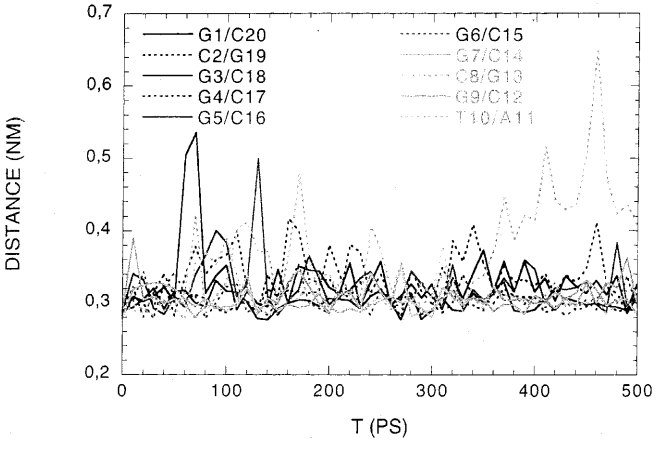

Fig. 6. Time series (each 10 ps) of the distance between the complementary bases. The persistence of contacts in Watson–Crick bridges are detected via the distances between three hydrogen bonds in G:C contacts and two hydrogen bonds in A:T contacts. Since hydrogen atoms are not visible in the X-ray data the distances are measured between the following (heavy) atom types N2(G):O2(C), N1(G):N3(C), O6(G):N4(C) for the G:C contacts and N3(T):N1(A), O4(T):N6(A) for the A:T contacts

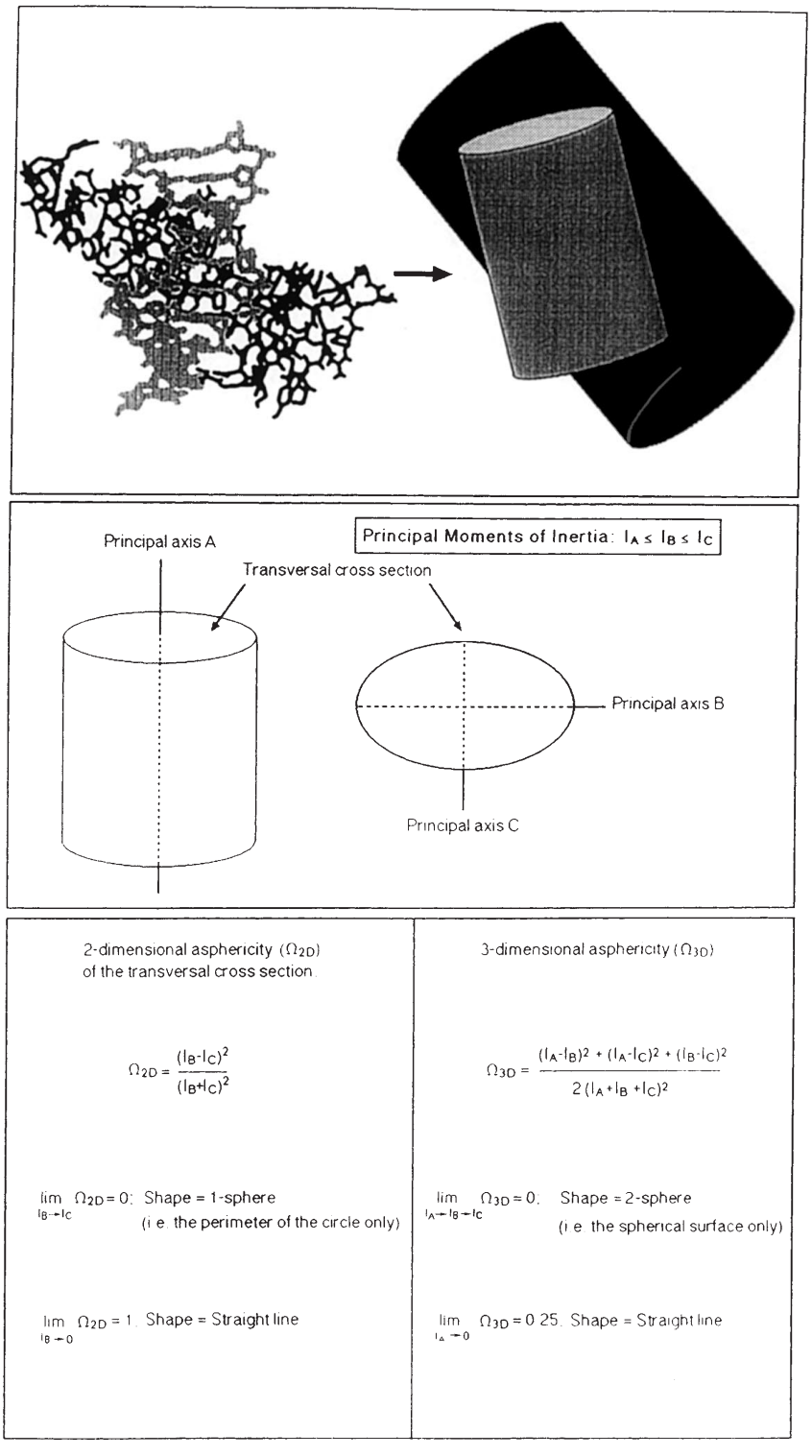

Fig. 7. **a** A schematic picture of DNA and the Zif in the Zif complex as a description of the principal moments of inertia analysis of the molecules. Note that the two axes are not comparable. **b** Both molecules have one longer principal axis (*A*) and two significantly shorter principal axes (*B* and *C*) that indicate a cylindrical mass distribution. The moment of inertia about principal axis *A* (I_A) increases with increasing diameter of the cylinder (the transversal cross section). The shape of the transversal cross section is a 1-sphere (i.e. the perimeter of a circle) when the moments of inertia about *B* and *C* (I_B and I_C) are equal. **c** The 2D asphericity (Ω_{2D}) is a 2D scale-independent shape descriptor that is applied on principal moments of inertia data in planes. We have studied the Ω_{2D} of the transversal cross section (here defined by principal axes *B* and *C*) in the two molecules. A Ω_{2D} of 0 corresponds to a 1-sphere. The maximum value of 1 corresponds to a straight line (collinear with principal axis *B*) and with zero moment of inertia about *B* (the moment of inertia about *C* is nonzero). **d** A 3D descriptor of the mass distribution is the 3D asphericity (Ω_{3D}). The extreme values are 0, that corresponds to a spherical surface and 0.25, which corresponds to a straight line with zero moments of inertia. The two other principal moments of inertia (I_B, I_C) are nonzero

is less molded (discussed later). The conformational search led to more open conformers with respect to finger–finger orientations.

The overall results including the T4 simulation suggest that F-1 has a behavior differing from the other two fingers. The different behavior is a combination of a less tight fit into the major groove of DNA and a higher mobility. It is tempting to assume that these differences are caused by a more flexible linker (remember that L_{12} lacks a salt bridge); however, other Zifs have been reported to show high motional freedom in the presence of the same linker residues as in L_{23} [34].

3.5 Watson–Crick pair conservation

The overlay of DNA structures from each 10 ps (Fig. 4) illustrates how the bases stay in the region that defines

the Watson–Crick bridges, especially in the central region where the mutation was made and where the intermolecular interactions with F-2 take place. The application of a distance criterion of a maximum 3.15 Å for heavy atoms involved in the crystal structure Watson–Crick bridge leads to an average hydrogen-bond persistence of 64%. Moreover, 90% of the bases keep an average distance below 3.65 Å, which is taken to be the maximum distance criterion for a base pair that is still able to make a hydrogen bond. The same analysis in the T4 simulation gave 64 and 89%, respectively. The time series of hydrogen-bonding distances is presented in Fig. 6a. Except for the first 5′ (G1/C20) and last (T10/A11) 3′ base pairs the fluctuations show a well-conserved hydrogen-bonding pattern. It is only the T10/A11 pair which actually breaks away. We expected the base pairs in the center of DNA to keep the Watson–Crick bridges since that had been the trend in previous simulations of free DNA (T4 sequence) and the Zif complex. DNA had a tendency to unwind and open up in the 5′ and 3′ ends both simulated free and in a complex with Zif. The overall base-pair stabilizing tendency in the presence of the protein led to better contact for DNA in the complex than the simulation of free DNA in water [4]. Monitoring the 3D trajectories in mdFRODO (not shown) suggests that the intermolecular direct contacts are most conserved in the center of DNA, just like in the T4 simulation. This interaction contributes to stabilize the base-pair hydrogen bonds. These comparisons suggest that the DNA structure is helped by its interaction with the protein, a result agreeing with experiment [19] and previous MD simulations of other protein–DNA complexes [36, 37].

3.6 Time series of the shape descriptor

A sensitive analytical tool is provided by shape descriptors [38, 39]. The asphericity descriptor is defined in Fig. 7. The relative orientation of DNA and the Zif protein is sketched as cylinders to help the discussion.

A cross-comparison of the 2D asphericity (Ω_{2D}) and the moment of inertia (I_A) about the principal axis, A, is presented in Fig. 8. The DNA molecule (Fig. 8a) starts from an almost circular shape for the transversal cross section and slowly moves towards a more elliptic shape.

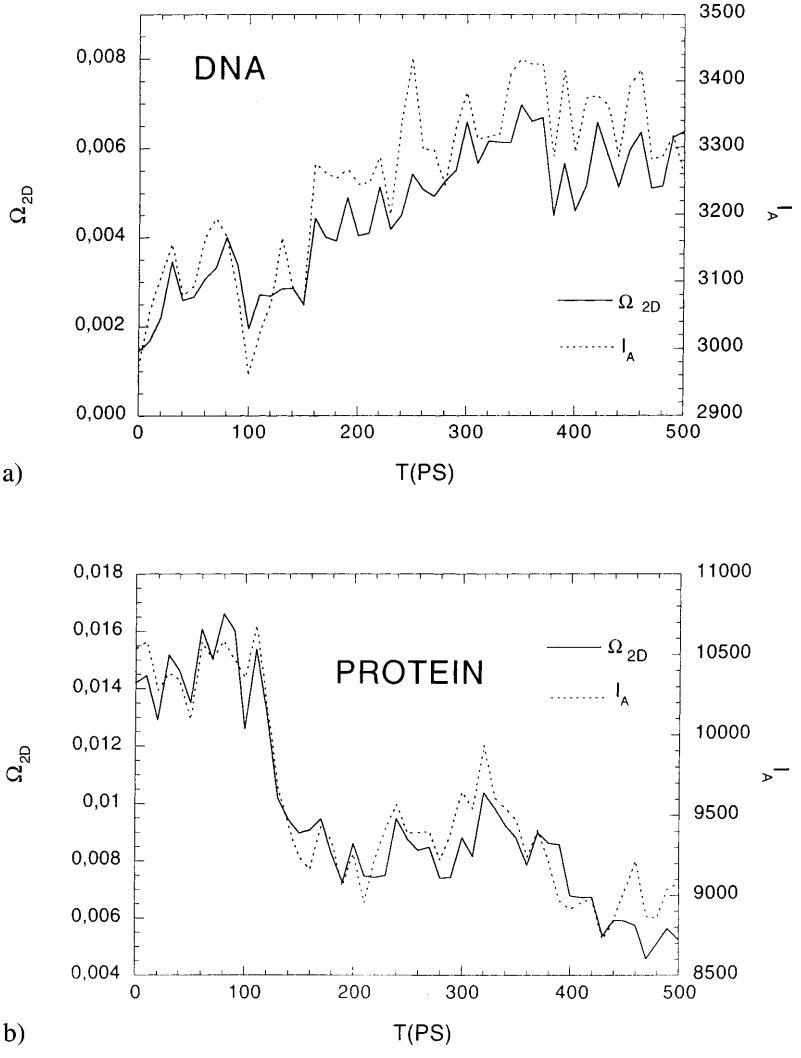

Fig. 8. Time evolution the Ω_{2D}, (*solid line*), which describes the shape of the transversal cross section, and the principal moments of inertia (I_A, *dashed line*) about the longest principal axis, A, which corresponds to the size of the transversal cross section as described in Fig. 7: **a** DNA molecule; **b** Zif (protein domain) molecule

I_A shows a similar time evolution. The high correlation between the two parameters for the transversal cross section (Ω_{2D} describes the shape and I_A describes the size) indicates distance fluctuations of the diameter in the same direction (i.e. principal axis *B* in Fig. 7) all the time. Although the changes in absolute value are small, the trend is what matters most here.

The protein shape mirrors the behavior of DNA as shown in Fig. 8b. Note that the axes of the Zif are not comparable to those of DNA as Fig. 7 illustrates. The protein shows a clear trend towards a more circular cross sectional shape, the opposite of the DNA behavior.

The 3D shape descriptor (Ω_{3D}) results are depicted in Fig. 9 with comparison to the results from the T4 simulation. The G4 trajectory (Fig. 9a) in the 500-ps range shows convergence in the shape of both partners. This analysis describes a behavior that might be a fine sensor of intermolecular interactions. The T4 trajectory in Fig. 9b presents similar behavior with a slower rate of convergence. It also seems that the two molecules are more in phase in the T4 simulation.

This shape descriptor is also useful for sensing subtle differences in the behavior of the three fingers. The time series of Ω_{3D} for each finger in the trajectories of G4 and T4 are illustrated in Fig. 10a and b, respectively. It is apparent that F-1 is different from the other two: F-1 is always less spherical than F-2 and F-3, which are basically very similar in both simulations. It should be noted that the time series of the principal moments of inertia (not shown) for F-1 is extremely well conserved in the present model. There is an indication that the repulsion between the symmetrized counterion and F-1 is less dominating in the present simulation.

A final comparison, in so far as DNA is concerned, can be made between the G4 and T4 simulations by using the width, depth, diameter and path length/end-to-end distance as shown in Fig. 11. The minor groove presents in both cases a fairly stationary type of fluctuation. The major groove reveals larger fluctuations that can be traced back to the interactions with the protein. The path length and end-to-end distance, for each case, are pretty much the same. This indicates a DNA fluctuation without bending. The diameter of DNA is also fairly stable. Thus, the DNA region exposed to the solvent is fairly stable (in terms of rather stationary fluctuations), while the region exposed to

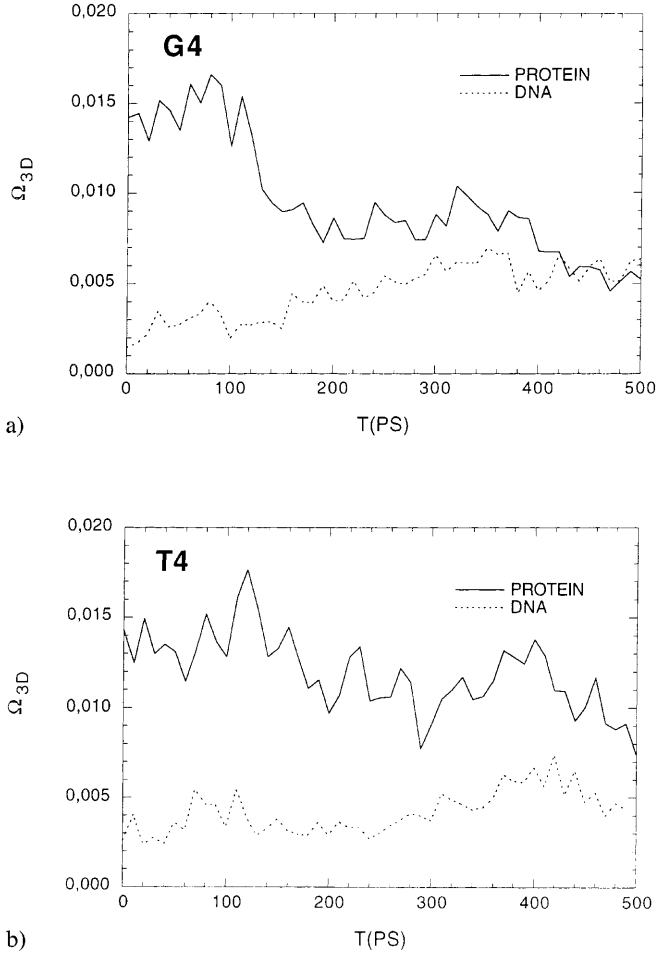

Fig. 9. Time evolution of Ω_{3D}, which describes the shape of the transversal cross section as described in Fig. 7: **a** for DNA (*solid line*) and Zif (*dashed line*) in the G4 simulation; **b** for DNA (*solid line*) and Zif (*dashed line*) in the T4 simulation

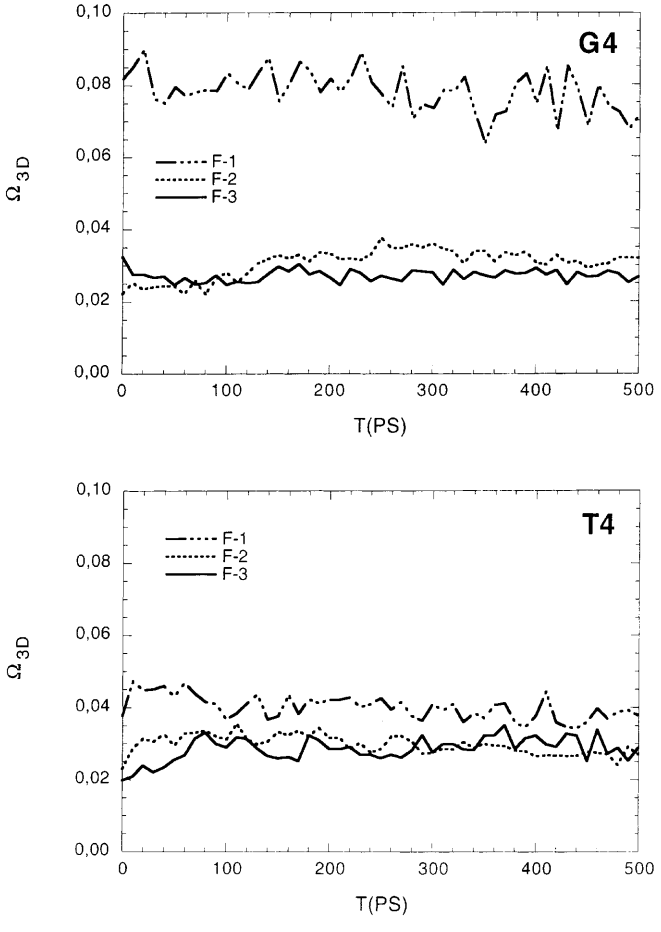

Fig. 10. Time evolution of Ω_{3D}, which describes the shape of the three Zif motifs: **a** from the T4 simulation; **b** from the G4 simulation. The crystal values are 0.697, 0.124 and 0.075 for fingers F-1, F-2 and F-3, respectively

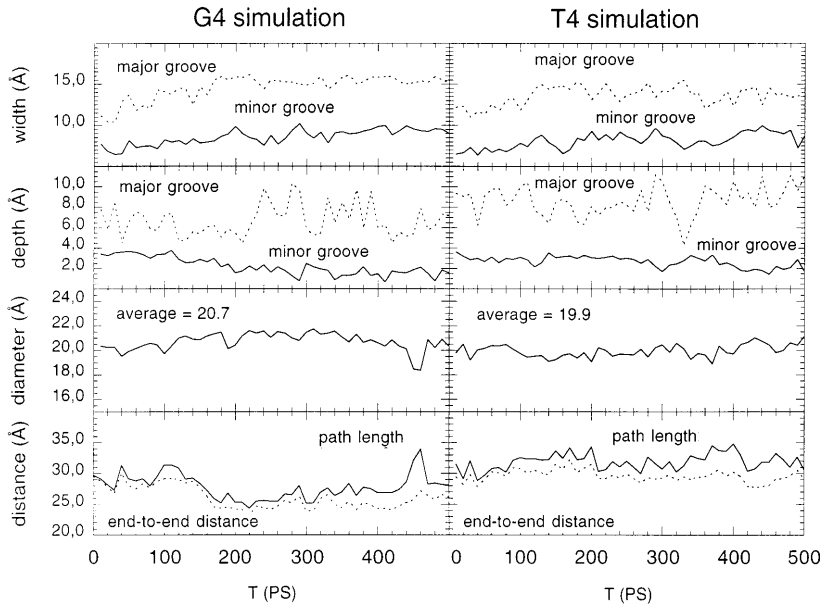

Fig. 11. The G4 and T4 average width and depth of the grooves (minor groove: *solid line*; major groove *dashed line*) as a function of time are compared in the *first two rows*. The time series of the average diameter and path length end-to-end distances are compared in the *last two rows*

protein interactions responds with large fluctuations. Observe that both minor and large groove features are influenced by P ions. It can be concluded that these ions do not freeze the DNA structure at all. In this sense, they provide a practical model to represent such a complex system.

4 Final remarks

Modeling of the G4 consensus sequence in a complex with the Zif protein was submitted to a MD study. Fully charged protein, DNA and a SPC water bath of about 5000 molecules were used. The results were compared to X-ray data obtained with the T4 consensus sequence as well as with previous MD simulation of the T4 model.

A hug motion was detected in the Zif when the relative motion between the fingers was monitored. The hugging is a collective property of the protein. Differences in hugging were found for the G4 and the T4 simulations. The picture is of a protein fluctuating around its DNA structure. The fluctuations are as large as those found for the T4 simulation. The three fingers do not behave in the same manner. Fingers F-2 and F-3 keep the recognition sites within reasonable boundaries during their thermal fluctuations; F-1 retains less of that information.

The distribution of P ions in our model should not be mixed up with the distributions of the neutralizing ions in nucleic acid crystals. The simulation model setup used here can now be used as a starting point for further studies concerning issues such as the role of cations on DNA structure. These issues have recently been reviewed by McFail-Isom et al. [23]. A further application is in the construction of 3D structures for the CreA Zif DNA complexes [24], for which only sequence information is available. The results obtained so far suggest that the F-2 and F-3 of Zif268 may provide suitable templates to model this two-zinc finger protein.

References

1. Westhof E, Patel D (1999) Curr Opin Struct Biol 9: 293
2. Young MA, Jarayam B, Beveridge DL (1997) J Am Chem Soc 119: 59
3. Louise-May S, Auffinger P, Westhof E (1996) Curr Opin Struct Biol 6: 289
4. Tapia O, Velazquez I (1997) J Am Chem Soc 119: 5934
5. Feig M, Pettitt M (1998) Biophys J 75: 134
6. Nilsson L (1997) In: Scheleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III (eds) Encyclopedia of computational chemistry. Wiley, New York, Volume 3 pp 2220
7. Roxström G, Velázquez I, Paulino M, Tapia O (1998) J Phys Chem 102: 1828
8. Roxström G, Velázquez I, Paulino M, Tapia O (1998) J Biomol Struct Dyn 16: 301
9. Lafontaine I, Lavery R (1999) Curr Opin Struct Biol 9: 170
10 van Gunsteren WF, Berendsen HJC (1987) Groningen Molecular Simulation (GROMOS) Library Manual. BIOMOS B.V., Groningen, The Netherlands
11. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP,Tironi IG (1996) GROMOS 96, BIOMOS B.V., Zürich
12. Reinmann CT, Velazquez I, Tapia O (1998) J Phys Chem B 102: 2277
13. Reinmann CT, Velazquez I, Tapia O (1998) J Phys Chem B 102: 9344
14. Daura X, Oliva B, Querol E, Avilés FX, Tapia O (1996) Proteins 25: 89
15. Richards EG (1980) An introduction to physical properties of large molecules in solution. Cambridge University Press, Cambridge
16. Christy BA, Lau LF, Nathans D (1988) Proc Natl Acad Sci USA 85: 7857
17. Christy B, Nathans D (1989) Proc Natl Acad Sci USA 86: 8737
18. Pavletich NP, Pabo CO (1991) Science 252: 809
19. Elrod-Eriksson M, Rould MA, Nekludova L, Pabo CO (1996) Structure 4: 1171
20. Postma J, Berendsen H, Haak J (1982) Faraday Symp Chem Soc 17: 55

21. Smith PE, van Gunsteren WF (1993) Chem Phys Lett 215: 315
22. Smith PE, van Gunsteren WF (1994) J Chem Phys 100: 3169
23. McFail-Isom L, Sines C, Williams L (1999) Curr Opin Struct Biol 9: 298
24. Cubero B, Scazzocchio C (1994) EMBO J 13: 407
25. Nilsson O (1990) J Mol Graph 8: 192
26. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, p 331
27. Jacobs GH (1992) EMBO J 11: 4507
28. Lavery R, Sklenar H (1988) J Biomol Struct Dyn 6: 63
29. Norberg J, Nilsson L (1995) J Am Chem Soc 117: 10832
30. Norberg J, Nilsson L (1995) Biophys J 69: 2277
31. Babcock MS, Pednault EP, Olson WK (1994) J Mol Biol 237: 125
32. Schuh R, Aicher W, Gaul U, Côté S, Preiss A, Maier D, Seifert E, Nauber U, Schröder C, Kemler R, Jäckle H (1996) Cell 47: 1025
33. Choo Y, Klug A (1993) Nucl Acids Res 21: 3341
34. Brüscheweiler R, Liao X, Wright P (1995) Science 268: 886
35. Foster M, Wuttke D, Radhakrishnan I, Case D, Gottsfield J, Wright P (1997) Nat Struct Biol 4: 605
36. Harris LF, Sullivan MR, Popken-Harris PD, Hickok DF (1994) J Biomol Struct Dyn 12: 249
37. Eriksson MA, Härd T, Nilsson L (1995) Biophys J 68: 402
38. Arteca GA (1996) Macromolecules 29: 7594
39. Arteca GA, Velázquez I, Reinmann CT, Tapia O (1999) Phys Rev E 59: 5981