Change in Conformation by DNA-Peptide Association: Molecular Dynamics of the Hin-Recombinase–*hixL* Complex

Yuto Komeiji* and Masami Uebayasi#
*National Institute for Advanced Interdisciplinary Research, Electrotechnical Laboratory, and #National Institute of Bioscience and Human Technology, AIST, Tsukuba, Ibaraki, Japan

ABSTRACT The Hin-DNA complex is a molecular complex formed by the C-terminal 52mer peptide of the Hin-recombinase and a synthetic 13-bp *hixL* DNA. The peptide has three α-helices, the second and third of which form the helix-turn-helix motif to bind to the major groove. Both termini of the peptide reside within the minor groove. Three molecular dynamics simulations were performed based on the crystal structure of the Hin-DNA complex: one for the free Hin peptide, one for the free *hixL* DNA, and one for the complex. Analyses of the trajectories revealed that the dynamic fluctuations of both the Hin peptide and the *hixL* DNA were lowered by the complex formation. The simulation supported the experimental observation that the N-terminus and the helix-turn-helix motif were critical for formation of the complex, but the C-terminus played only a supportive role in DNA recognition. The simulations strongly suggested that the binding reaction should proceed by the induced fit mechanism. The ion and solvent distributions around the molecules were also examined.

INTRODUCTION

The question examined by this study is a simple one: how does the formation of the Hin-DNA complex affect the structure and dynamics of the constituent peptide and DNA?

DNA binding by polypeptides plays important roles in a variety of biomolecular systems. Gene expression, replication, recombination, repair, and so on are all controlled by DNA binding reactions. Although the physical basis for these recognition processes is not yet fully understood, x-ray crystallography and NMR have been used to characterize a vast number of three-dimensional structures of DNA-peptide complexes, thus providing us with a wealth of knowledge on DNA recognition (see, for example, the recent reviews of Chang and Varani, 1997, and Schwabe, 1997). Based on these structural studies, we here attempted to characterize the change in conformation and dynamics of a DNA and a peptide upon complex formation by means of molecular dynamics (MD) simulations.

MD simulation is a theoretical tool used to investigate the time evolution of a molecular system by solving the classical equations of motion of the particles comprising the system. MD has been extensively applied to biological macromolecules, such as proteins, to understand their dynamics, model their structures, compute their free energy, and so on (Karplus and Petsko, 1990; Kollman, 1996; Merz, 1997; Tidor, 1997). Over the past half-decade, a large number of MD simulations of nucleotides have appeared in the literature, stimulated by advances in the computation algorithms of the electrostatic interaction (see Berne and Straub, 1997, for a review of such methods), as well as by the development of a sophisticated force field (Cornell et al., 1995). See Jayaram and Beveridge (1996) and Auffinger and Westhof (1998) for recent reviews of MD analyses of nucleic acids. The number of MD simulations of protein-DNA (or RNA) complexes is still limited, but increasing. Avoidance of the electrostatic cutoff will soon be standard in the simulation of DNA-peptide complexes, because the cutoff contributes serious artifacts to the results of such highly charged molecules (See Komeiji and Uebayasi, 1999, and references therein).

We chose a DNA-peptide complex formed by the C-terminal 52mer peptide of the Hin-recombinase and the half-site of *hixL* DNA (Fig. 1 *A*; Feng et al., 1994) as the target of our study.

The Hin-recombinase (Hin) of *Salmonella typhimurium* belongs to the class of DNA-cleaving enzymes known as invertases. Hin catalyzes a DNA inversion reaction of a 1-kb segment of the chromosome to control the alternate expression of two flagellin genes (Zieg et al., 1977; Hughes et al., 1992). Hin proteins initiate the sequence of recombination reactions by binding to two 26-bp homologous sequences, *hixL* and *hixR*, that flank the invertible segment. Two dimers of a Fis protein bind to a region of DNA called the recombinational enhancer. Then the *hixL* and *hixR* sites with their bound Hin proteins form a complex with the enhancer with its bound Fis proteins. The two recombination sites are aligned by this complex formation, and the Hin protein initiates the exchange of DNA strands, leading to inversion of the intervening DNA. See Hughes et al. (1992) for details of the recombination procedure.

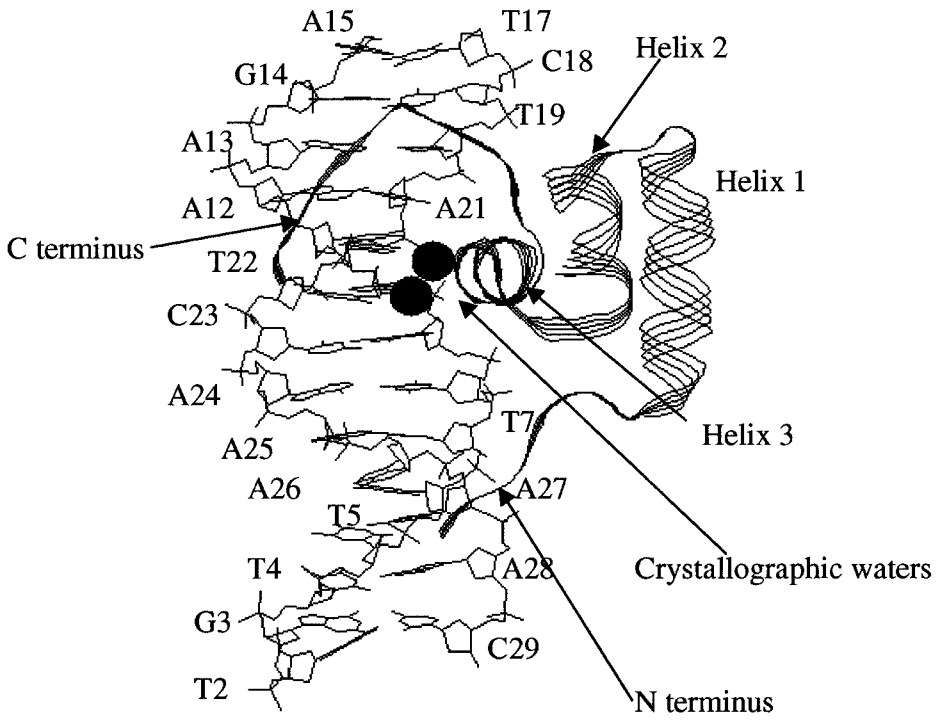
Hin is a 190-amino acid polypeptide with an N-terminal catalytic domain and a C-terminal DNA binding domain. The C-terminal 52mer peptide (Fig. 1 *B*) has been shown to specifically bind to the *hixL* site with a moderate affinity (dissociation constant: $K_d = 10^{-7}$, corresponding to a Gibbs free energy change of $\Delta G = -9.5$ kcal/mol; Bruist et al.,

Received for publication 31 December 1998 and in final form 22 March 1999.

Address reprint requests to Dr. Yuto Komeiji, National Institute for Advanced Interdisciplinary Research, 1-1-4 Higashi, Tsukuba, Ibaraki, Japan 305-8562. Tel.: +81-298-54-5164; Fax: +81-298-54-3364; E-mail: komeiji@etl.go.jp.

0006-3495/99/07/123/16 $2.00

A. Crystal Structure of the HIN-*hixL* complex



B. Amino acid sequence of the 52mer

```
139       148             162          173       181        190
|         |               |            |         |          |
GRPRAITKH-EQEQISRLLEK-GHP-RQQLAIIF-GIG-VSTLYRYF-PASSIKKRMN
α-helix  |---(1)----|     |--(2)--|     |--(3)--|
```

C. DNA base sequence of the *hixL* half site

```
                2  3  4  5  6  7  8  9 10 11 12 13 14 15
Strand 1  5'    T  G  T  T  T  T  T  G  A  T  A  A  G  A  3'
Strand 2  3'       C  A  A  A  A  A  C  T  A  T  T  C  T  5'
                  29 28 27 26 25 24 23 22 21 20 19 18 17
```

FIGURE 1 (*A*) Schematic representation of the crystal structure of the Hin-DNA complex (Feng et al., 1994). Labeled are the bases of the DNA (synthetic *hixL* half site), and the termini and three α-helices of the peptide. Also shown are two crystallographic waters involved in water-mediated H-bonds between the peptide and DNA. (*B*) Amino acid sequence of the C-terminal 52mer peptide of the Hin-recombinase. The residues shown in bold are identical in at least three of the four enteric invertases: Hin, Gin, Cin, and Pin (Feng et al., 1994). (*C*) DNA base sequence of the *hixL* half-site. The base pairs shown in bold are important because of their recognition by invertases (Feng et al., 1994).

1987). The three-dimensional structure of the complex formed by the 52mer and a synthetic 13-bp *hixL* half-site was determined to atomic resolution (Feng et al., 1994; Fig. 1 *A*). The peptide consisted of three α-helices, the second and third of which made up the helix-turn-helix (HTH) DNA binding motif. Helix 3 was inserted in the major groove of the DNA. The N- and C-termini of the peptide resided within two regions of the minor groove. The overall structure was intermediate between a typical bacterial HTH and the eukaryotic homeodomain. Deletion of the N-terminal Gly^{139}-Arg^{140} caused complete loss of the DNA-binding affinity (Sluka et al., 1990), whereas deletion of the eight C-terminal eight residues resulted only in a reduced affinity (Mack et al., 1990). The *hixL* half-site is an AT-rich DNA (Fig. 1 *C*). The crystal structure of the *hixL* half-site adopted the B-form conformation. Two crystallographic waters were detected as mediating the water-mediated hydrogen bonds (Fig. 1 *A*). Participation of water in the complex formation was also demonstrated by an experiment utilizing osmotic stress (Robinson and Sligar, 1996).

The Hin 52mer peptide-DNA complex is fairly small, but retains the essential features of DNA recognition; hence, we have chosen it as a model system of DNA binding by a peptide. We have recently reported a 1-ns MD trajectory for the Hin 52mer-DNA complex (Komeiji and Uebayasi, 1999; hereafter referred to as paper I). In paper I, we investigated the peptide-DNA interaction and suggested the importance of both termini of the peptide in DNA recognition. We also compared MD simulations with and without the truncation of the electrostatic interaction and concluded

that the electrostatic interaction should not be truncated for such a highly charged molecular system. In the present paper, we extended our investigation in paper I by adopting a strategy similar to that of both Eriksson et al. (1995) and Tang and Nilsson (1998). Namely, we performed MD simulations of the free DNA and peptide separately and compared the results with those for the complex to estimate the change in dynamics and conformation upon binding. We generated 2.5-ns trajectories of the Hin 52mer peptide and the *hixL* half-site DNA separately. We also continued the MD simulation of the complex in paper I up to 2.5 ns to increase the statistical significance. Thus three trajectories, that of the complex, the free peptide, and the free DNA, were generated. The conformation and dynamics of the components in the free and complexed states were compared. The behaviors of the ions and solvent molecules were also analyzed. The DNA-peptide interface was examined in detail. Finally, speculations about the binding process were made based on the computational results.

METHODS

Simulation protocol

Three MD simulations (MD-pep, MD-dna, and MD-com; Table 1) were performed based on the crystal structure of the complex composed of the Hin 52mer and the *hixL* half-site (PDB entry 1HCR; resolution 1.8 Å; Feng et al., 1994). MD-pep was started from the peptide moiety of the complex, MD-dna from the DNA moiety, and MD-com from the whole complex. The simulation protocol of MD-com has already been reported in paper I, and MD-pep and MD-dna simulations were performed using the same protocol. Hence, only an outline is given below. The procedures for MD simulations and analyses are schematically presented in Fig. 2.

An in-house software package, PEACH (Program for Energetic Analysis of Biochemical Molecules; Komeiji et al., 1995, 1997) was used throughout this study. The AMBER95 all-atom force field (Cornell et al., 1995) was employed for the peptide, DNA, and Na^+ and Cl^- ions. A flexible version of SPC water (Dang and Pettitt, 1987) was used to model the water molecules. The Ewald summation of the electrostatic interaction was performed in all of the energy minimization (EM) and MD simulations by using a special processor named MD-GRAPE (gravity pipe for MD; Fukushige et al., 1996). The van der Waals (VDW) interaction was also computed by MD-GRAPE. The multiple time step method of Tuckerman et al. (1992) was used to perform the time integration. The time steps were 0.25 fs for the bond, angle, and torsion interactions (hard force); 1 or 2 fs for the VDW and Ewald real-space (r-space) (medium force); and 4 fs for the Ewald wave number-space (soft force). The medium force time step was 2 fs for the optimization of the solvent and ions and for the following optimization of the solutes (Fig. 2, stage 1), whereas 1 fs was employed during the production run (Fig. 2, stage 2). A Nose-Hoover thermostat (Nose, 1991) was employed as described (Komeiji et al., 1997) whenever temperature control was needed.

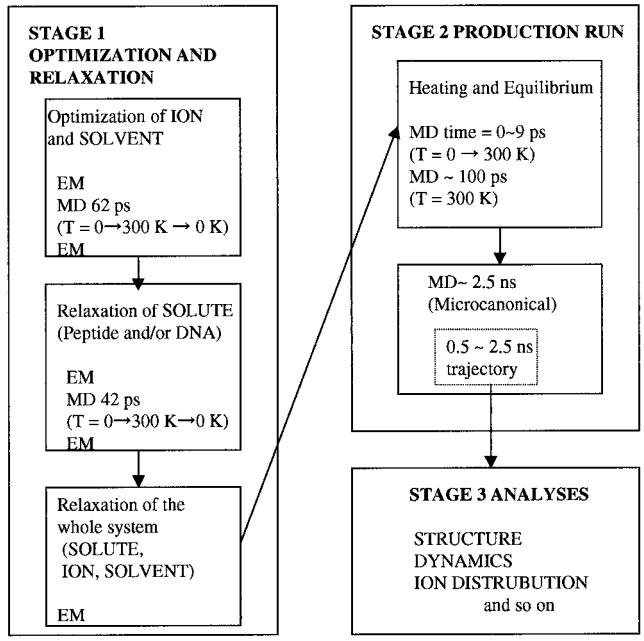


FIGURE 2 Simulation protocol. The protocol has been described in detail by Komeiji and Uebayasi (1999).

Hereafter, the peptide and/or DNA are referred to as "solute," the counterions as "ion," and the water molecules as "solvent," as in paper I. The solute molecules were immersed in a cubic box filled with the water molecules. The shortest distance between the solute and the wall was 10 Å. Enough Na^+ or Cl^- ions to neutralize the system were generated. See Table 1 for the sizes of the simulated systems. Optimization and relaxation were performed (Fig. 2, stage 1) before the production run. The positions of the solvent molecules and ions were optimized to the solute by EM and MD for 62 ps. Then the solute molecules were subjected to EM and MD for 42 ps while the solvent and ions were kept frozen. Finally, the whole system was subjected to EM (Fig. 2, end of stage 1; see Fig. 3 for the configurations obtained at this stage), and the production MD run was started (Fig. 2, stage 2). Each simulation was performed under an isothermal condition at 300 K for 100 ps. Microcanonical MD was then started; that is, the temperature control was turned off. Each simulation was continued for 2500 ps (2.5 ns). CPU seconds needed per femtosecond of MD simulation were 2.2 for MD-pep, 2.8 for MD-dna, and 4.3 for MD-com. Trajectories of the 0.5–2.5-ns period were mainly used to analyze the structure and dynamics (Fig. 2, stage 3), as described in the following subsection. The crystal and simulated structures were visualized by RASMOL 2.6 (Sayle, 1995).

TABLE 1 Sizes (number of atoms) of the simulated systems

	MD-pep*	MD-dna#	MD-com§
Solute	875	859	1734
Na^+ ions	—	25	16
Cl^- ions	9	—	—
Solvent	13,065	16,014	19,062
Box dimension (Å)	53.4	57.0	60.8
Total	13,949	16,898	20,812

The MD simulations were started from the peptide,* the DNA,# and the whole complex.§

Analyses of the trajectories

The structure and dynamics of the three MD trajectories were analyzed as follows. Hereafter, the time average of quantity A is designated as $\langle A \rangle$.

The main purpose of this study was to compare the dynamics of the DNA in the presence and absence of the peptide, and that of the peptide in the presence and absence of the DNA. Hence, we extracted the peptide coordinates and the DNA coordinates separately from the trajectory of MD-com and compared them with the trajectories of MD-pep and MD-dna, respectively. Whenever necessary, the coordinates were fitted to the initial configuration according to the method of McLachlan (1979).

Root mean square deviations (RMSDs) of the generated trajectories from

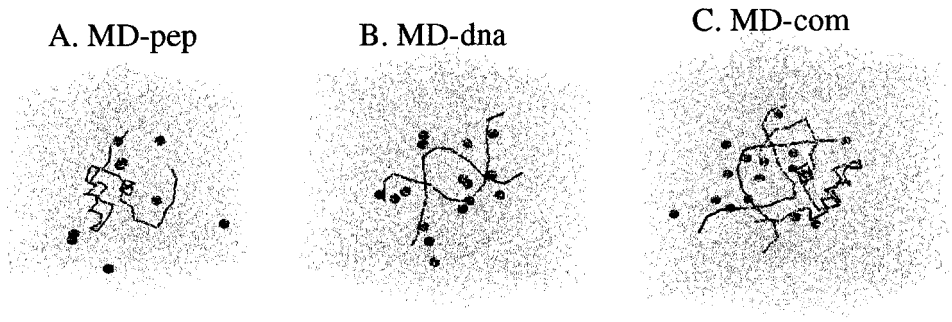


FIGURE 3 Initial configurations used for the production runs of (*A*) MD-pep, (*B*) MD-dna, and (*C*) MD-com obtained at the end of stage 1 in Fig. 2.

the initial crystal structure were calculated. Namely, RMSD at time t is

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}i(t) - \mathbf{r}i^{\mathrm{init}})^2} \quad (1)$$

where i is the atom index and N is the total number of atoms. The instantaneous and initial coordinates of atom i are $r_i(t)$ and r_i^{init}, respectively.

Root mean square fluctuations (RMSFs) were calculated with respect to the amino acid residues and DNA bases as follows:

$$\mathrm{RMSF}_j = \sqrt{\left\langle \frac{1}{N_j}\sum_{i=1}^{N_j}(\mathbf{r}i(t) - \langle \mathbf{r}i\rangle)^2 \right\rangle} \quad (2)$$

Here, N_j is the number of atoms belonging to residue or base j. The Hydrogen atoms were excluded from RMSD and RMSF.

Several parameters concerning the DNA structures were computed by the Curves program (Lavery and Sklenar, 1988) included in the analytical software package Toolchest 1.0 (Ravishanker and Beveridge, 1997).

The solvent-accessible surface area (SASA) of each coordinate set was computed according to the method of Shrake and Rupley (1973) and averaged over the MD trajectories. The radius of the probing solvent was 1.4 Å, and VDW radii of the solute atoms were taken directly from the AMBER force field.

The densities of the ions and solvent molecules around the solute molecules were represented by several methods. Before the analyses, each ion or solvent molecule in each time frame of the trajectory was transferred to its closest solute atom within the periodic boundary. The position of each atom was represented by the atom center.

The normalized ion (or solvent) density distribution was presented as $G(r)$, which was defined as follows:

$$G(r) = \frac{\Delta N(r)}{D\Delta V(r)} \quad (3)$$

The distance r from the surface of the solute molecule was defined as the distance of the ion (or solvent) from the nearest solute atom. $\Delta N(r)$ was the number of ions located in the $r - \frac{1}{2}\Delta r \approx r + \frac{1}{2}\Delta r$ region from the solute surface, where Δr is a small increment of distance. D was the number density of the ions in the box. $\Delta V(r)$ was the volume of the $r - \frac{1}{2}\Delta r \approx r + \frac{1}{2}\Delta r$ region. To compute $\Delta V(r)$, the surface area at distance r ($S(r)$) was calculated, analogously to SASA. That is, all of the solute atoms were assigned a radius of zero, and $S(r)$ was computed for the time-averaged structure of the solute by using a probe solvent of radius r. Then $\Delta V(r)$ was computed as follows:

$$\Delta V(r) = S(r)\Delta r \quad (4)$$

Thus $G(r)$ is an extension of the radial distribution function, computed to describe the atom-atom distribution, to a solute-atom distribution. In computation of $G(r)$, Hydrogen atoms of the solutes were neglected, and each solvent molecule was represented by the water oxygen only.

The cylindrical density distribution of the ions from the DNA axis, $G_{\mathrm{cyl}}(r)$, was calculated in a manner similar to that for $G(r)$, using Eq. 3. However, $\Delta N(r)$ was the number of ions residing in the $r - \frac{1}{2}\Delta r \approx r + \frac{1}{2}\Delta r$ region from the DNA axis, and $\Delta V(r)$ was the volume of the cylinder of the $r - \frac{1}{2}\Delta r \approx r + \frac{1}{2}\Delta r$ region. Before calculation of $G_{\mathrm{cyl}}(r)$, the coordinate sets from the trajectory of MD-DNA were fitted to a canonical B-DNA whose axis was the z axis in the orthogonal space.

RESULTS AND DISCUSSION

Energy conservation and temperatures of the trajectories

We first analyzed the total energy and the temperatures of the generated trajectories to confirm that the systems reached thermodynamic stability (Table 2).

All of the simulations were performed in the microcanonical ensemble after 100 ps. In the microcanonical ensemble, the total energy should be conserved. The conservation of the total energy was excellent (fluctuation $\fallingdotseq$ 0.1% for the 0.5–2.5-ns period), considering the length of the trajectories. Temperature is not conserved in the microcanonical ensemble, but it should be stabilized at the equilibrium. The total temperature was 300 K at the end of the isothermal simulations (100 ps), and the average of the 0.5–2.5-ps trajectory was 3–6 K higher (Table 2). We considered the temperature increase of 5 K to be negligible. The artifact of the temperature separation often seen under

TABLE 2 Summary of the total energy and temperatures of MD trajectories (0.5–2.5 ns)

	MD-pep	MD-dna	MD-comp
Total energy (10^4 kcal/mol)	−3.2907 ± 0.0035 (0.11%)	−4.4749 ± 0.0037 (0.09%)	−5.2380 ± 0.0061 (0.12%)
Temperature (K)			
All	304.6 ± 1.9	305.1 ± 1.6	304.2 ± 2.4
Solute	304.8 ± 8.3	305.3 ± 8.4	305.0 ± 6.4
Ion	302.9 ± 89.7	306.0 ± 51.8	303.8 ± 64.0
Solvent	304.6 ± 2.0	305.1 ± 1.7	304.0 ± 2.5

the electrostatic cutoff (see paper I and references therein) did not occur in the present simulations; the solute, ions, and solvent had similar temperatures within acceptable errors (Table 2).

Structure and dynamics of the solute molecules

The fluctuations of the solute molecules were illustrated by superposing snapshots of the solute backbones (Fig. 4). The overall difference in dynamics between MD-com and MD-pep (or MD-dna) was well visualized. Both the peptide and DNA had larger fluctuations in the free states (Fig. 4, *A* and *B*) than in the complex state (Fig. 4 *C*), and the differences were larger in the regions forming the complex. Below we discuss in detail the structure and dynamics of the peptide and DNA in the free and complexed states to estimate the change in dynamics and conformation caused by the complex formation.

Peptide

The time evolution of RMSD of the peptide in MD-com was compared with that in MD-pep (Fig. 5 *A*). We have previously shown the results for MD-com up to 1 ns (paper I), and in the present simulation RMSD of the peptide portion did not show a significant change after 1 ns. The structural deviation increased for the first 0.5 ns and then fluctuated about 2.5–2.8 Å for the rest of the simulation (Fig. 5 *A*, *dotted line*). In MD-pep, however, RMSD continued to increase up to 5–6 Å for 0.5 ns and then stabilized (Fig. 5 *A*, *solid line*). The difference between MD-com and MD-pep was rather small (2–3 Å) but significant. The residue RMSD was computed for the crystal structure and the average of the trajectory (Fig. 6 *A*). In MD-com, the N-terminal residues deviated only slightly from the crystal structure, but they deviated largely in MD-pep. As discussed in paper I, the C-terminal residues were only marginally stable in MD-com, but comparison with the results for MD-pep indicated that the C-terminus was also greatly stabilized by interaction with DNA.

The difference in RMSF between MD-pep and MD-com was also clear (Fig. 7 *A*; see also Fig. 4, *A* and *C*). Fluctuation was decreased by the complex formation in both termini and in helix 3. This result was natural, because these parts of the peptide were in direct contact with the DNA (Fig. 1 *A*). The difference was quite notable, especially in the two termini of the peptide, again indicating the importance of these termini in DNA recognition. Most of the portions of helix 2 were not directly involved in DNA recognition (Fig. 1 *A*), but the RMSF showed stabilization of helix 2 by the complex formation (Fig. 7 *A*). Helix 1 showed small fluctuation in both MD-pep and MD-complex (Fig. 7 *A*).

DNA

The explanation of the time evolutions of RMSD of the DNA in MD-com and MD-dna (Fig. 5 *B*) was not as simple as that for the peptide. Although on average the RMSD was somewhat larger in MD-dna than in MD-com, the time evolution did not show that the crystal structure was always more stable in the latter. In MD-dna, RMSD gradually increased up to 4.5 Å for the first 0.7 ns, but then decreased and became smaller than 3 Å at 2 ns. In MD-com, RMSD was rather stable at 2.5 Å for the first 1.2 ns, but increased and then stabilized at 3.5 Å. Thus the DNA molecules underwent structural fluctuations and transitions in both the presence and absence of the peptide.

RMSD between the crystal structure and the time averages of the trajectories were computed with respect to the

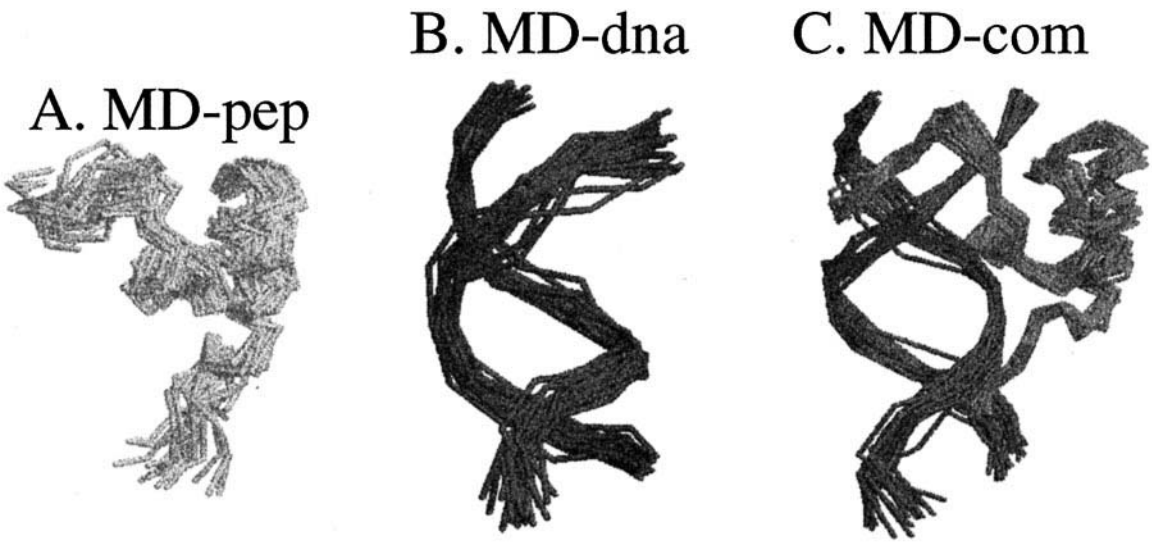


FIGURE 4 Superimposed snapshots of the solute molecules from (*A*) MD-pep, (*B*) MD-dna, and (*C*) MD-com. Snapshots were sampled every 100 ps from the 0.5–2.5-ns trajectory.

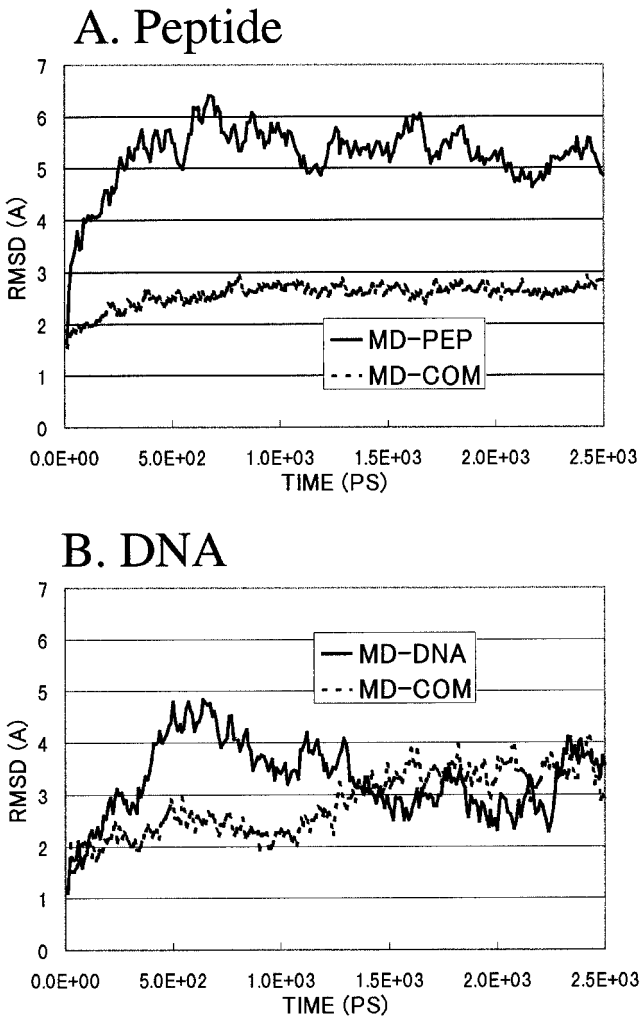


FIGURE 5 Time evolutions of RMSD (Eq. 1) of the simulated structures from the crystal structure for (*A*) peptide and (*B*) DNA. In computation of RMSD of the peptide and the DNA in MD-com, the peptide and DNA moieties were extracted independently from the trajectory and fitted to the crystal structures of the peptide and DNA, respectively.

DNA bases to analyze which portion of the DNA deviated the most and which portions showed large differences between MD-com and MD-dna (Fig. 6 *B*). Here, the averaged structures were computed for the 0.5–2.5-ns, 0.5–1.0-ns, and 1.5–2.5-ns trajectories, considering the total RMSD in Fig. 5 *B*. The RMSD between the averaged structures of MD-dna and MD-com is also shown (Fig. 5 *C*).

RMSD was large near the 5′ ends of the two strands, both in MD-dna and in MD-com. Large structural deviation and fluctuation of the ends of DNA have often been reported (Young et al., 1997b; Noberto de Souza and Ornstein, 1997). In particular, base T2 is unpaired, and its location was hard to determine by x-ray crystallography (Feng et al., 1994). Thus the large deviation and fluctuation of T2 should be unavoidable. Computation of RMSD for only the central 11 bp resulted in a decrease in RMSD of ~0.5 Å for both MD-com and MD-dna (data not shown). Prominent differences between MD-dna (Fig. 6 *B*, *solid lines*) and MD-com (Fig. 6 *B*, *dotted lines*) were seen in the T6–G9 and T19–A21 regions, where the crystal structures were better conserved in MD-com than in MD-dna. Similar differences can be seen in Fig. 6 *C*, in which the difference between MD-dna and MD-com was directly compared. These regions make contact with the peptide (Fig. 1 *A*) and should be stabilized by formation of the complex.

The 0.5–1.0-ns (*thin lines*) and 1.5–2.5-ns (*medium lines*) trajectories were compared for both MD-dna and MD-com to analyze the structural change over the 1.0–1.5-ns period (Fig. 5 *B*). The structural change was larger in strand 2 (T17–C29) in both MD-dna and MD-com. In MD-dna, the difference between the 0.5–1.0-ns and 1.5–2.5-ns trajectories was prominent in the G9–A13 and A21–C23 regions. These regions of the DNA were in contact with helix 3 of the peptide (Fig. 1 *A*). Because MD-dna was started from the DNA portion of the complexed structure, this portion of strand 2 largely deviated for the first 0.7 ns (Fig. 5 *B*) and then shifted back slightly in the direction of the original structure. In MD-com, the difference between the 0.5–1.0-ns and 1.5–2.5-ns trajectories was largest in the G3–T4 and T22–A25 regions. These regions were not directly involved in the peptide binding.

Thus the DNA molecule underwent slight conformational transition during the 1.0–1.5-ns period in both MD-dna and MD-com, but in MD-dna the structural deviation occurred mainly in the peptide binding region, whereas in MD-com it occurred in the region not directly involved in the peptide binding.

Next we compared the groove widths of the crystal structure and those averaged over the trajectories (Fig. 8). There are several ways to define the widths of the major and minor grooves. Here we used the distance between atom centers of two phosphorous atoms appropriate for the canonical B-DNA structure, consistent with the Toolchest program (Ravishanker and Beveridge, 1997).

The major groove width in MD-com (Fig. 8 *A*, *dotted lines*) was in close agreement with that of the crystal structure (*thin solid line*) in the T5/C23–G9/T19 region. This region of the major groove makes direct contact with helix 3 (Fig. 1 *A*). The major groove width was larger in MD-dna than in the crystal in all of the regions, indicating that the DNA structure was relaxed and widened in MD-dna. Thus the stabilization effect of the attachment of the HTH motif (helix 3) and the major groove was demonstrated.

Usually, the minor groove width of an AT-rich DNA, such as *hixL* (Fig. 1 *C*), is smaller than the canonical value (Drew et al., 1981; Nelson et al., 1987; Schui et al., 1998). In the crystal of the Hin-*hixL* complex used in the present study, however, the minor groove width was close to the canonical value in most of the regions (Feng et al., 1994). Feng et al. attributed this discrepancy to the occupation of the minor groove by the termini of the peptide. Comparison of the present MD-com and MD-dna results supports their idea. The minor groove width was larger in MD-com than in MD-dna in most of the regions, suggesting that the presence of the peptide enlarged the minor groove.

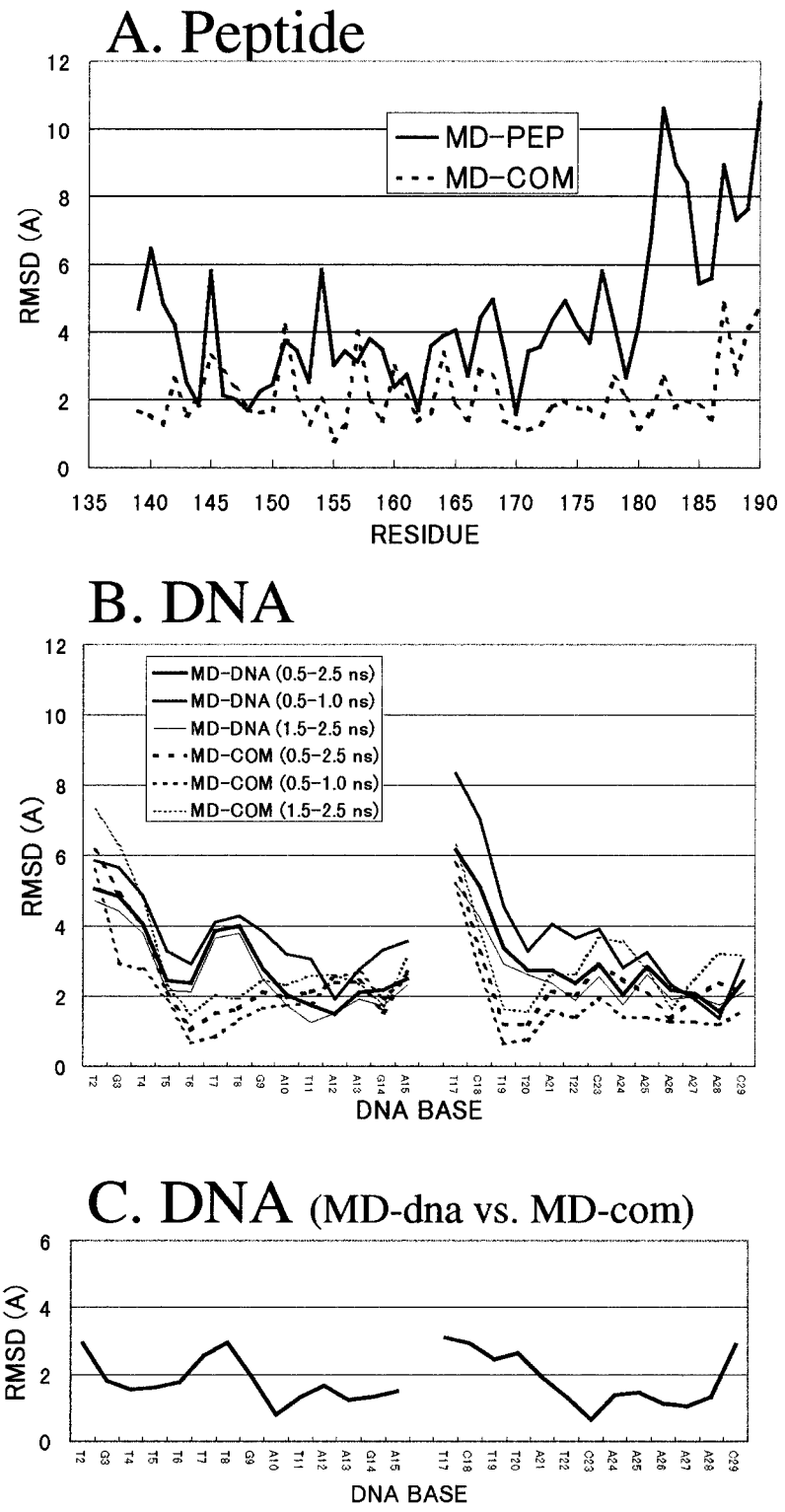


FIGURE 6 RMSD (Eq. 1) was calculated between the crystal structures and the simulated structures averaged over 0.5–2.5 ns with respect to the amino acid residues of the peptide (*A*) or the bases of the DNA (*B*). The 0.5–1.0-ns and 1.5–2.5-ns trajectories were also considered for the DNA portion (*B*). RMSD between the averaged structures of MD-dna and MD-com (0.5–2.5 ns) is also shown (*C*).

There must be some argument, however, about the minor groove width obtained in the present MD simulations. The minor groove width obtained by MD-com was, on the average, ~2 Å larger than that of the crystal structure (Fig. 8 *B*). Although the width obtained for MD-dna was smaller than that obtained for MD-com, it was still larger than that obtained for the crystal structure in most of the regions. We cannot eliminate the possibility that the simulation protocol including the force field was inappropriate, but we are inclined to believe that the simulation-induced expansion of the minor groove was a result of the end effects. Usually the ends of DNA show large fluctuations and deviations (Noberto de Souza and Ornstein, 1997), and these are sometimes withheld from the analyses so as not to introduce

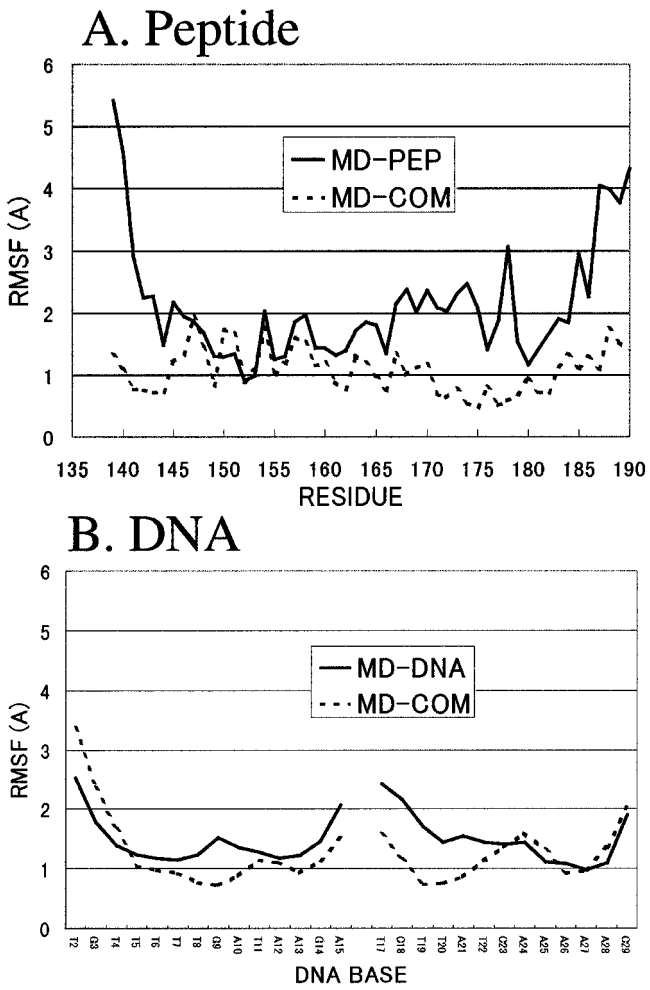


FIGURE 7 RMSF (Eq. 2) of the amino acid residues (*A*: peptide) and the nucleotides (*B*: DNA). RMSF was calculated for the 0.5–2.5-ns trajectories.

artifacts into the simulation results. In our simulations, however, the AT-rich regions in question were located near both ends of the DNA (Fig. 1 *A*). A larger DNA molecule free from the end effects would be needed to test the ability of an MD simulation to reproduce the narrowing of the minor groove of an AT-rich DNA. Apart from the problem of the end effects, our results qualitatively supported the idea that the minor groove is widened by the bound peptide.

The difference in the RMSF of DNA between MD-dna and MD-com was rather small (Fig. 7 *B*) compared to that of the peptide (Fig. 7 *A*), but the fluctuations of T8–T11 of strand 1 and T17–T22 of strand 2 were larger in MD-dna than in MD-com. Bases T8–T11 were in contact with helix 3, and bases T17–T22 were in contact with both helix 3 and the C-terminal residues of the peptide. Hence the difference in fluctuations was reasonable. As discussed in the previous subsection, the N-terminus of the peptide was firmly attached to the DNA. Its influence on the fluctuation of the DNA structure was not prominent, however, presumably because of the small size of the area attached by the N-terminus.

In the current simulations, we employed the AMBER95 force field. The A ↔ B conformational transitions of several DNA molecules under the several conditions seen in the experiments have previously been well produced by this force field (Miller and Kollman, 1997; Cieplak et al., 1997; Cheatham and Kollman, 1997; Sprous et al., 1998). In a few previous cases, this force field failed to reproduce the observed A-to-B transition (Sprous et al., 1998), but this did not seem to be a serious drawback in the present simulation.

Among the conformational and helicoidal parameters of DNA defined by Ravishanker et al. (1989), two of the axis-base parameters, X-displacement (XDP) and Inclination (INC), were the principal diagnostic parameters distinguishing between the A and B forms of DNA. The parameter XDP was definitely closer to the canonical B for the crystal and MD structures (Fig. 9 *A*). The parameter INC's of the crystal and MD-com structures were in between, although the values were closer to those of the B form (Fig. 9 *B*). Interestingly, in MD-dna, the parameter INC adopted values nearly identical to to those of the canonical B form. That MD-dna was closer to the B form than the crystal, and MD-com was also seen in the case of XDP (Fig. 8 *A*), although less prominently. The results for XDP and INC suggested that the peptide-binding region of the *hixL* DNA was slightly distorted from the canonical B form because of the complex formation.

Ion distribution

In this subsection we report on the ion distribution around the solute molecules.

The normalized distributions ($G(r)$, Eq. 3) of the ions around the solutes are shown in Fig. 10, and the high ion density points are visualized in Fig. 11 *A–C*. MD-pep and MD-dna were first analyzed as references, and then MD-com was investigated based on the references.

First, the behavior of Cl^- ions in MD-pep was examined. The Cl^--peptide distribution was simple, with a single peak at 3.4 Å (Fig. 10 *A*). The ions underwent diffusion around the peptide, and none of them were considered to be bound to or localized in specific regions of the peptide (not shown). However, the Cl^- ion density was highest in the vicinity of the N-terminal region of the peptide (Fig. 11 *A*), indicating the positively charged nature of this region. This fact was consistent with the strong attachment of the N-terminus to DNA, a negatively charged molecule, in MD-com.

Next, the ion distribution around the DNA in MD-dna was analyzed in light of Manning's counterion condensation (CC) theory (Manning, 1978) and were compared with recent related analyses, for instance, Young et al. (1997b).

The ion behavior around a polyelectrolyte such as DNA has been well explained by the CC theory. The theory is summarized as follows. When a polyelectrolyte is present, a certain number of ions condense to the polyelectrolyte. The number of condensed ions depends solely on the linear

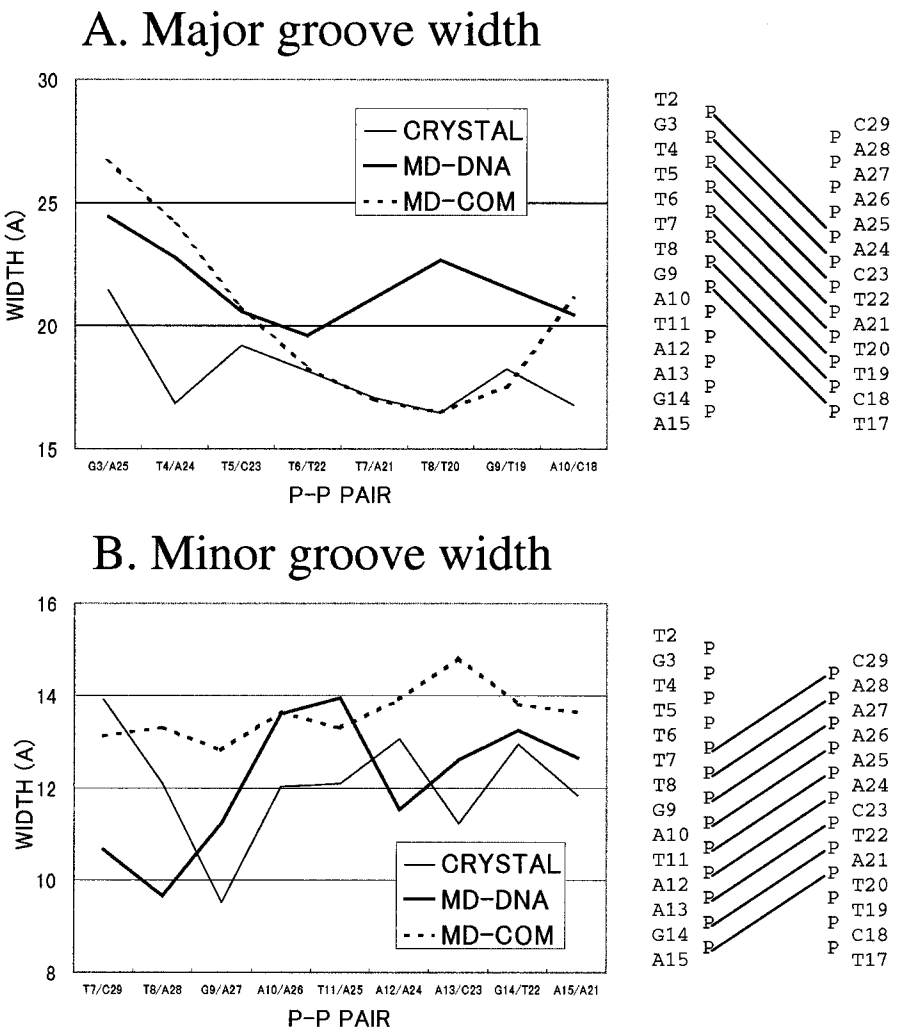


FIGURE 8 Time-averaged (*A*) major and (*B*) minor groove widths of the DNA in MD-dna and in MD-com. The groove widths were also shown for the crystal structure. The groove width was defined as the P-P distance between appropriate bases for the canonical B DNA, consistent with the software Toolchest (Ravishanker and Beveridge, 1997). The P-P pairs actually used were indicated in the right side of the graphs.

charge density of the polyelectrolyte and on the valence of the ions and is independent of the ion concentration. The ions condense within a certain distance from the solute (Manning radius). In the case of B-DNA, enough monovalent ions to screen 76% of the DNA charge are predicted to condense within a Manning radius of 17 Å from the axis of the DNA (Manning, 1978; Jayaram et al., 1996).

We computed the cylindrical distribution of the Na^+ ions around the axis of the DNA and the running coordination number (Fig. 12) from the 0.5–2.5-ns trajectory. To minimize the end effects, only the central 8 bp were considered. The cylindrical distribution had two broad peaks at 8 Å and 13 Å. The former corresponds to the ions within the grooves, especially the minor groove, and the latter corresponds to those residing near the phosphate backbones. The initial positions of the ions were determined to minimize the ion-solute interaction (paper I); hence one may suspect that our initial configuration favored the ion condensation. However, the shape of the cylindrical distribution in Fig. 12 is similar to that of Young et al. (1997b), in which the initial positions of the ions were randomly assigned. This agreement suggests the generality of our result.

If we define the inflection point at 17 Å in Fig. 12 as the Manning radius, only 60% of the DNA charge is condensed, a value smaller than the predicted 76%. Originally, the CC theory was constructed for infinite polyelectrolytes. Subsequently, Fenly et al. (1990) extended the condensation theory for the polyelectrolyte with finite length and concluded that the condensation should become weak for a short polyelectrolyte. Although the ion concentration in MD-dna in the present study was high (0.22 mol/L), and hence the size limit of the DNA molecule should have been fairly small, the predicted ion condensation of 76% must be an overestimate. The small size of the simulation box could be another source of artifact. The box was a cube with a dimension of 57 Å (Table 1), which is only three times longer than the Manning radius. In conclusion, the ion behavior in MD-dna was not inconsistent with the CC theory. Nevertheless, the general applicability of the theory to MD results must be reconsidered by performing simulations of longer DNA molecules in larger boxes under various concentrations of ions.

The analysis of the cylindrical ion distribution was not appropriate for MD-com, in which the solutes were not

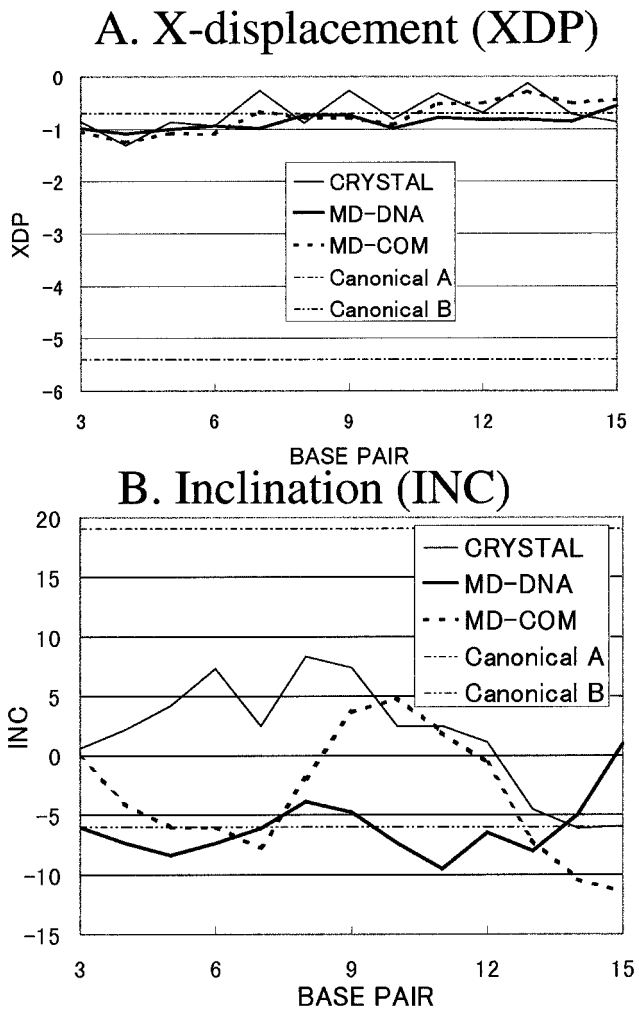


FIGURE 9 The axis-base parameters (*A*: XDP and *B*: INC) defined by Ravishanker et al. (1989), distinguishing A and B forms of DNA. For MD-dna and MD-com, average values of the 0.5–2.5-ns trajectories are shown.

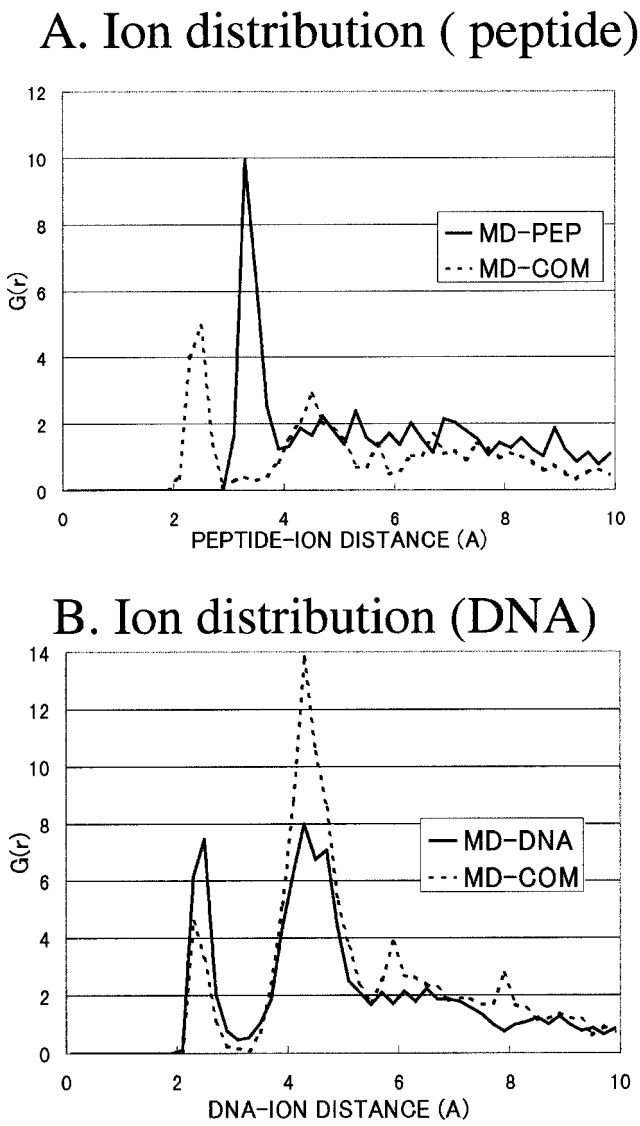


FIGURE 10 Normalized ion density distribution ($G(r)$, Eq. 3) as a function of the distance from the surface atom of the solutes: (*A*) peptide and (*B*) DNA. Cl^- ions were examined for MD-pep in *A*; otherwise, Na^+ ions were examined.

cylindrical. Hence, the normalized ion density ($G(r)$, Eq. 3) as a function of the distance of the ion from the surface atom of the solutes was computed to compare the overall ion distributions in MD-dna and MD-com (Fig. 10 *B*).

The Na^+-DNA distribution in MD-dna was characterized by two sharp peaks, one at 2.3 Å and the other at 4.5 Å. The former corresponds to those making direct contact with the solute atom and the latter to those making water-mediated contact. The ion density was especially high along the minor groove and then around the phosphate backbones (Fig. 11 *B*). Similar peaks were seen in MD-com, but the first peak was much smaller than that in MD-dna. This difference could be a statistical variation due to the short simulation time and the small number of ions. However, we could also attribute this to the occupation of the minor groove by the peptide in MD-com.

Na^+ ions were also found near the peptide in MD-com (Fig. 11 *C*), but the density was lower than that near the DNA. This was natural because the peptide was positively charged (+9e).

We should keep in mind that an MD trajectory as much as a few nanoseconds in duration is too short to equilibrate and sample the ionic behavior adequately (Young et al., 1997b; Feig and Pettitt, 1998). Thus the results and discussion presented so far should only be considered semiquantitative or qualitative.

Solvent distribution

Solvent distribution around the solute molecules was also examined. The normalized solvent distribution ($G(r)$) of water oxygens was computed (Fig. 13). In the case of the solvent, the difference in distribution around the DNA (or peptide) was extremely small and was negligibly different for MD-dna (or MD-pep) versus MD-com. All four curves were characterized by two peaks at ~2.8 Å and ~3.8 Å.

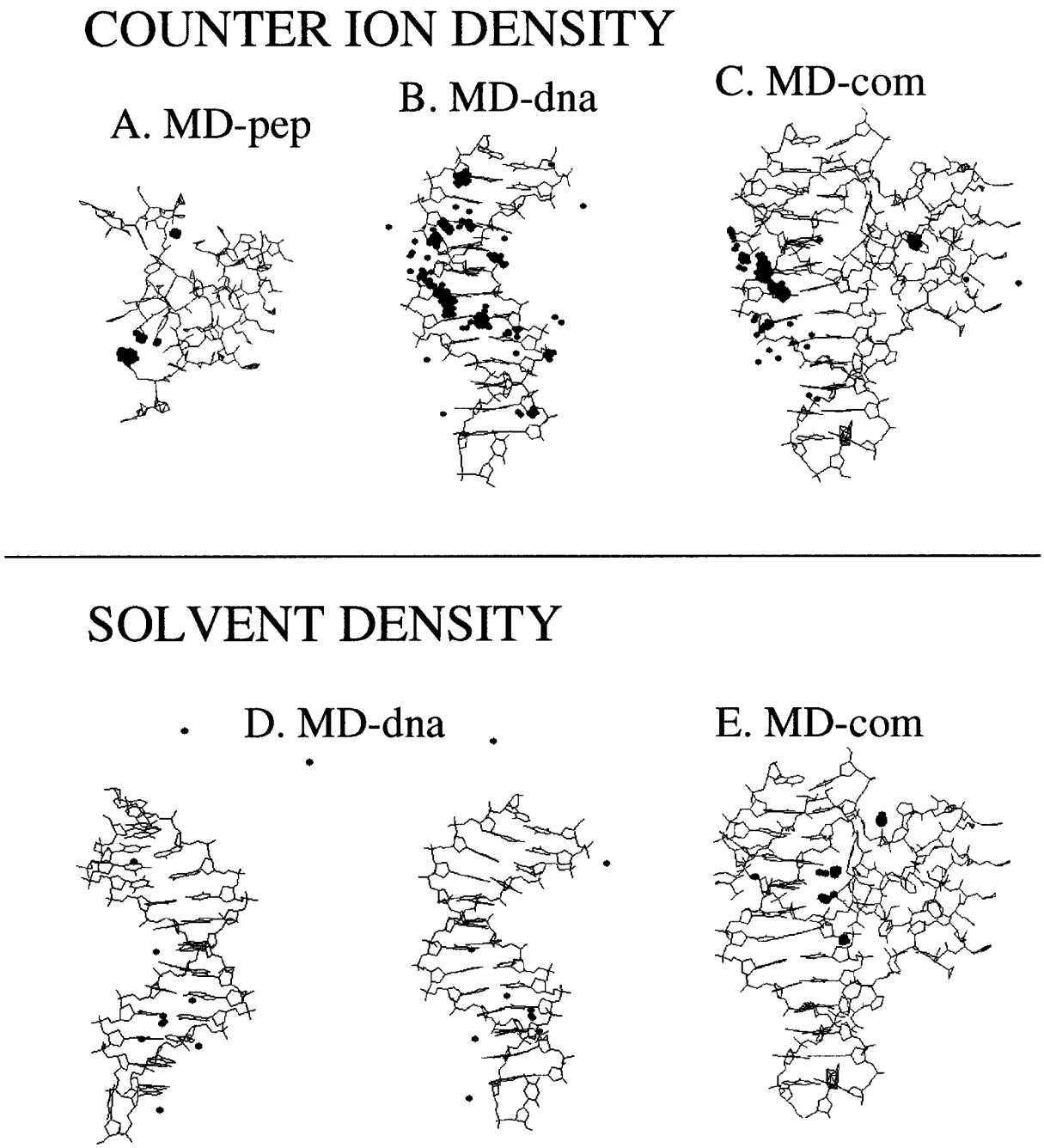


FIGURE 11 Ion density maps in (*A*) MD-pep (Cl^- ions), (*B*) MD-dna (Na^+ ions), and (*C*) MD-com (Na^+ ions), and solvent density maps in (*D*) MD-dna and (*E*) MD-com. The area was divided by mesh points (0.5 Å × 0.5 Å × 0.5 Å), and the average ion or solvent (water oxygen) density was computed for each mesh point for the 0.5–2.5-ns trajectories. Those mesh points with a density larger than 0.02 ion/Å^3 (*A*, *B*, *C*) or 0.4 solvents/Å^3 (*D*, *E*) were visualized by black points along with the average structures of the solutes.

The former corresponds to those solvent molecules making direct hydrogen bonds (H-bonds) with charged atoms of the solutes, whereas the latter corresponds to those forming the hydrophobic shell around nonpolar atoms (Komeiji et al., 1993; Norin et al., 1994; Makarov et al., 1998). The first peak was sharper around the DNA (Fig. 13, *thin lines*) than around the peptide (*thick lines*). Presumably, this difference was attributable to the highly charged nature of the DNA; that is, the DNA attracted polar solvents more strongly than did the peptide.

High density points in MD-dna (Fig. 11 *D*) and in MD-com (Fig. 11 *E*) were also visualized. The points in Fig. 11, *D* and *E*, had a solvent density larger than 0.4 molecules/Å^3, ~15 times denser than the average. No such high-density point was seen in MD-pep, so solvent density in MD-pep was not included in Fig. 11.

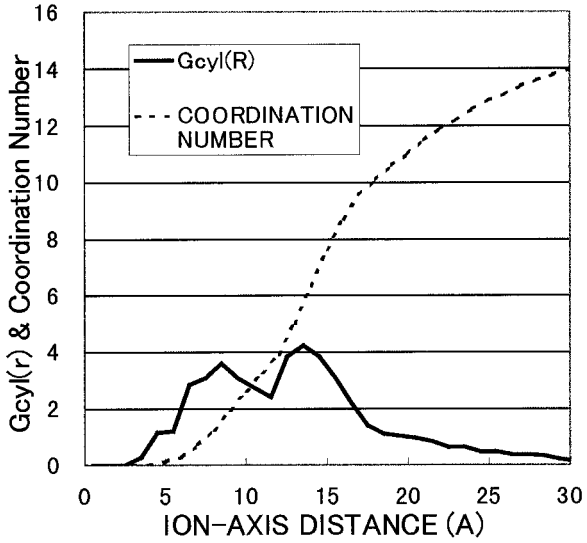


FIGURE 12 The cylindrical ion distribution ($G_{cyl}(r)$) around DNA in MD-dna (——) and the running coordination number (– – –). Only the central 8 bp were considered, to reduce the end effects. The coordination number of 16 corresponds to 100% ion condensation.

As already stated, the ion density along the minor groove was higher along bp 8–14 (Fig. 11 *B*). As if compensating for this high ion density, the solvent density along the minor groove was higher along bp 4–7 (Fig. 11 *D*). We do not know whether this difference in ion and solvent densities along the minor groove was an artifact of the short sampling time of 2 ns or a result of the small difference in the sequence (Fig. 1 *C*). This question can only be resolved by making an MD simulation that is one or two orders of magnitude longer. The statistical significance of the current solvent density maps was rather low; for instance, there were a few high-density points far from the solutes (Fig. 11 *D*), but they seemed to be only statistical fluctuations. Here we can only conclude that the spine of water along the minor groove and the intruding ions

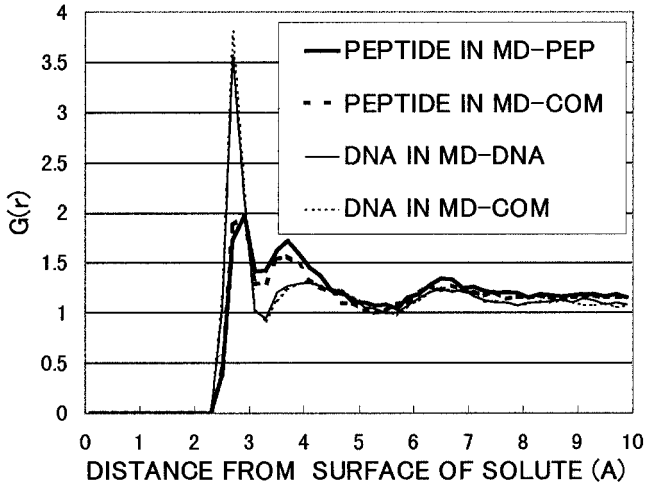


FIGURE 13 Normalized solvent density distribution ($G(r)$, Eq. 3) as a function of the distance from the surface atom of the solute. Only oxygen atoms of the water molecules were considered.

were also reproduced in our simulation, which is consistent with several previous reports (Duan et al., 1997; Young et al., 1997a; De Winter et al., 1998).

In MD-com, most of the high-density points of the solvent were seen around the DNA-helix 3 interface (Fig. 11 *E*). Some high-density points were also seen in the DNA-C terminus interface. Some of the solvent molecules located in these regions were considered to be involved in water-mediated H-bonds. We will discuss this point in the following subsection.

To characterize the overall distribution of the solvents, SASA and the number of hydrating water molecules were computed (Table 3). In computing the hydrating solvent molecules, we considered those residing within 3.5 Å of the atom center of the surface atom of the solute molecules. Such a definition is arbitrary, but it should be a good approximation, considering the solvent distribution in Fig. 13. The results of Table 3 give a rough picture of solvent behavior upon complex formation. When the complex is formed, the solutes lose SASA of 2.4×10^3 Å^2 and release 84 solvent molecules. About 27 solvent molecules remain in the DNA-peptide interface. This simulation result was qualitatively consistent with the experimental data of Robinson and Sligar (1996), in which the participation of water in DNA recognition was demonstrated.

Peptide-DNA interactions

In this subsection we discuss the peptide-DNA interaction of the Hin-*hixL* complex by focusing on the trajectory of MD-com.

In paper I, we gave a preliminary report on the dynamics of the peptide-DNA interface. We mainly discussed the stability of the H-bonds present in the crystal structure by analyzing the first 1-ns trajectory of MD-com. The N-terminal, helix 3, and C-terminal regions of the peptide were involved in DNA recognition via H-bonds. The H-bonds due to the N-terminus and helix 3 were stable during MD simulation, but those due to the C-terminus were less stable.

In the present paper, we also investigated H-bonds not present in the crystal structure. In addition, we included the solvent-mediated H-bonds in our analyses. See the legend to Table 4 for the definitions of an ordinary H-bond and a water-mediated H-bond. The stability of H-bonds was measured by the ratio (%) of their appearance during the 0.5–2.5-ns trajectory.

The majority of the H-bonds present in the crystal structure were fairly well reproduced during the MD simulation for the N-terminus (amino acids 140, 143) and helix 3 (172–182). Nevertheless, preservation of the crystallographic H-bonds was extremely poor in the C-terminal region (183–190), although the backbone of the region mostly resided in the minor groove throughout MD-com (Fig. 4 *C*). Two of the 16 crystallographic waters formed water-mediated H-bonds between the peptide and DNA (Feng et al., 1994; Fig. 1 *A*). One of the two, that between

TABLE 3 Solvation of the solute molecules in MD trajectories (average of 0.5–2.5 ns)

	A MD-pept	B MD-dna	C MD-com	A + B − C
Hydrating water molecules*	194	280	390 (27)[#]	84
SASA (Å^2)[§]	4,798 ± 156	5,472 ± 43	7,872 ± 89	2,398 ± 185

*Average number of water molecules residing within 3.5 Å of the solute molecules were counted.
[#]Water molecules residing in the interface between the peptide and DNA (within 3.5 Å of both peptide and DNA).
[§]The radius of the probe solvent was 1.4 Å. Hydrogen atoms were neglected.

R178 and A21 and T22, was also seen in MD-com (Table 4). The water-mediated H-bond was reflected in the high water density of the region (Fig. 11 *E*). The other water-mediated H-bond, however, that residing between S174 and A21, was not reproduced.

Thus, not all of the H-bonds in the crystal structure were stable, but this result did not necessarily mean that the H-bonds between the peptide and the DNA were weakened or destroyed by MD. During MD, several new H-bonds were formed. Moreover, some of the ordinary H-bonds changed to form water-mediated H-bonds (S174-A10, S174-T11, and R178-T8; Table 4). The number of ordinary H-bonds in the crystal structure and the average of MD-com were compared for the amino acid residues of the peptide and the base pairs of the DNA (Table 5). It was noteworthy that the total number of the H-bonds was 22, in both the crystal structure and the MD-com trajectory. This suggested that even though the H-bonds went through formation and disappearance along with the fluctuation of the complex, the strength of H-bonds between the peptide and DNA remained similar during MD simulation. Even some of those C-terminal residues in which the original H-bonds had mostly disappeared during MD formed H-bonds with different base pairs. Table 5 indicates that, although the H-bonds present in the crystal were not always stable, other H-bonds appeared to compensate for the strength of the intermolecular H-bonds. Such a dynamic formation and disappearance of several H-bonds in the peptide-DNA in-

TABLE 4 Stability of H-bonds present in the crystal structure

Peptide residue	DNA base	H-bonding atom pairs		Stability during MD-com (%)*	
				Ordinary[#]	Water-mediated[#]
Ordinary H-bonds in crystal[#]					
R140[§]	T6	N	02	79	9
	A26	NE	N3	0	18
	A27	N	N3	67	0
A143	T8	N	01P	99	0
G172	**G9**	N	02P	47	0
S174	**A10**	0G	N7	18	72
	T11	0G	04	44	32
T175	G9	0G1	02P	100	0
Y177	**T19**	0H	02P	0	0
R178	T8	NH1	02P	0	77
	G9	NH2	N7	18	0
Y179	T8	0H	01P	100	0
A182	**T20**	N	02P	95	1
S183	**T20**	N	01P	3	15
I185	G14	0	N2	5	0
K186	**T20**	NZ	03′	0	2
K187	**T20**	N	02	0	0
R188	G14	NH2	01P	3	4
		NH2	05′	1	3
M189	**T22**	N	04′	0	0
N190	**A10**	ND2	N3	0	1
	T22	N	02	1	1
Water-mediated H-bonds in crystal[#]					
S174	**A21**	0	N7	0	1
		0	N6	0	0
R178	**A21**	NE	N7	0	85
	T22	NE	04	0	20

*Stability of H-bonds were measured by the percentage of their presence in 400 snapshots during the 0.5–2.5-ns trajectory of MD-com.
[#]Ordinary H-bonds were defined as those with a donor-acceptor distance $R_{DA} < 3.5$ Å. Water-mediated H-bonds were those that do not satisfy the above, but both atoms form H-bonds with a water oxygen.
[§]Important amino acids and DNA bases (see legend to Fig. 1) are shown in bold.

TABLE 5 Number of H-bonds between peptide residues and DNA base pairs*

	DNA													
Peptide	3GC	**4TA**	**5TA**	6TA	7TA	8TA	**9GC**	**10AT**	**11TA**	**12AT**	**13AT**	14GC	15AT	Subtotal
G139	−/−	−/0.4	−/0.1	−/0.2	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/0.7
R140	−/−	−/−	−/0.8	1/0.8	2/−	−/0.2	−/−	−/−	−/−	−/−	−/−	−/−	−/−	3/1.8
R142	−/−	−/−	−/−	−/−	−/1.1	−/0.9	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/2.0
A143	−/−	−/−	−/−	−/−	−/−	1/1.0	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/1.0
R162	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/0.8	−/−	−/−	−/0.8
N163	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/0.1	−/−	−/0.1
G172	−/−	−/−	−/−	−/−	−/−	−/−	1/1.5	−/−	−/−	−/−	−/−	−/−	−/−	1/1.5
S174	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/1.2	1/1.3	−/−	−/−	−/−	−/−	2/2.5
T175	−/−	−/−	−/−	−/−	−/−	−/−	1/1.6	−/−	−/−	−/−	−/−	−/−	−/−	1/1.6
Y177	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/1.5	−/−	−/−	1/1.5
R178	−/−	−/−	−/−	−/−	−/−	1/−	1/2.7	−/1.0	−/−	−/−	−/−	−/−	−/−	2/3.7
Y179	−/−	−/−	−/−	−/−	−/−	1/1.0	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/1.0
A182	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/1.2	−/−	−/−	−/−	1/1.2
S183	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/−	−/−	−/−	−/−	1/−
I185	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/0.1	−/−	1/0.1
K186	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/−	−/−	−/−	−/−	1/−
K187	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/0.4	−/0.3	−/1.2	−/−	1/1.9
R188	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	−/−	2/0.2	−/−	2/0.2
M189	−/−	−/−	−/−	−/−	−/−	−/−	−/−	1/−	−/0.1	−/−	−/−	−/−	−/−	1/0.1
N190	−/−	−/−	−/−	−/−	−/−	−/−	−/−	2/0.3	−/0.2	−/−	−/−	−/−	−/−	2/0.5
Subtotal	−/−	−/0.3	−/0.9	1/1.0	2/1.2	3/3.1	3/5.8	4/2.4	1/1.6	4/1.6	1/2.6	3/1.5	−/−	22/22.0

*Number of H-bond pairs present in the crystal (left) or average of 0.5–2.5-ns trajectory of MD-com (right) between peptide residues and DNA base pairs, namely, (crystal)/(MD). Only ordinary H-bonds were considered. For example, 1/1.5 means that one H-bond was present in the crystal structure and the average number of H-bonds during MD-com was 1.5. A minus sign indicates that the average number was <0.1. Important amino acids and DNA bases (see legend to Fig. 1) are shown in bold.

terface, seen here in MD-com, has also been seen in a homeodomain, in which the dynamic structure was in an equilibrium between two or three H-bond-forming sites (Schwabe, 1997).

The amino acid residues of the peptide and base pairs of the DNA, chosen from the viewpoint of conservation among the Hin recombinase family and its cognitive sequences (Feng et al., 1994), are shown in bold in Tables 4 and 5. The C-terminal residues of the peptide (S183-N190) were not conserved, and their H-bonds were consistently unstable. The N-terminal region and HTH motif have several conserved amino acids, but there seemed to be little correlation between the stability of their H-bonds and their importance.

Formation of an H-bond is a rather short-range interaction, but the electrostatic interaction is long-range. H-bond interaction is implicitly treated by VDW plus electrostatic interactions in the AMBER95 force field. Thus we also analyzed the nonbonded interactions (VDW and electrostatic) between the peptide and DNA in the complex. A contour plot of nonbonded interactions is presented in Fig. 14, in which favorable interactions are shown in reddish colors and unfavorable ones in bluish colors. The total nonbonded energy of the crystal structure and the average of MD-com were similar (-4.8×10^3 versus -4.6×10^3 kcal/mol). No essential difference was seen between the map of the crystal (*A*) and that of MD-com (*B*). Roughly four parts of the peptide interacted favorably with the DNA: the N-terminus, the loop between helices 1 and 2, helix 3, and the C-terminus. The residues in the N-terminus and helix 3 interacted strongly with all parts of the DNA. The C-terminal residues (185–189) also interacted favorably in the crystal and, though a little weakened, in MD. Although the H-bonds were unstable, the C-terminal residues were still important in the binding to the minor groove.

To understand molecular association, it is not enough to analyze only the nonbonded interactions, including H-bonds, between the associated molecules. The change in free energy for this complex formation was only −9.5 kcal/mol, but the computed interaction energy was larger by three orders of magnitude. Although this difference should be compensated for by including the interactions with the surrounding medium, the hydrophobic effect, and the loss of entropy due to the complex formation (Schwabe, 1997), such effects are not within the scope of the current simulations. Therefore, the analyses of the nonbonded interactions including H-bonds so far presented (Tables 3 and 4, Fig. 14) showed only the relative strengths of interactions between individual amino acids and DNA base pairs. In summary, the H-bonds formed by the peptide excluding the C-terminal residues (amino acids 139–180) were relatively stable, but those formed by the peptide including the C-terminal residues (181–190) were not. However, the C-terminal residues still interacted favorably via nonbonded interactions. This suggests that the C-terminal residues contribute to the peptide-DNA association via nonspecific interactions. Some of the H-bonds formed via the N-terminus and HTH motif were not rigid; rather, the H-bonding pattern in the crystal was in a dynamic equilibrium with other H-bonding sites, including the water-mediated H-bonds.

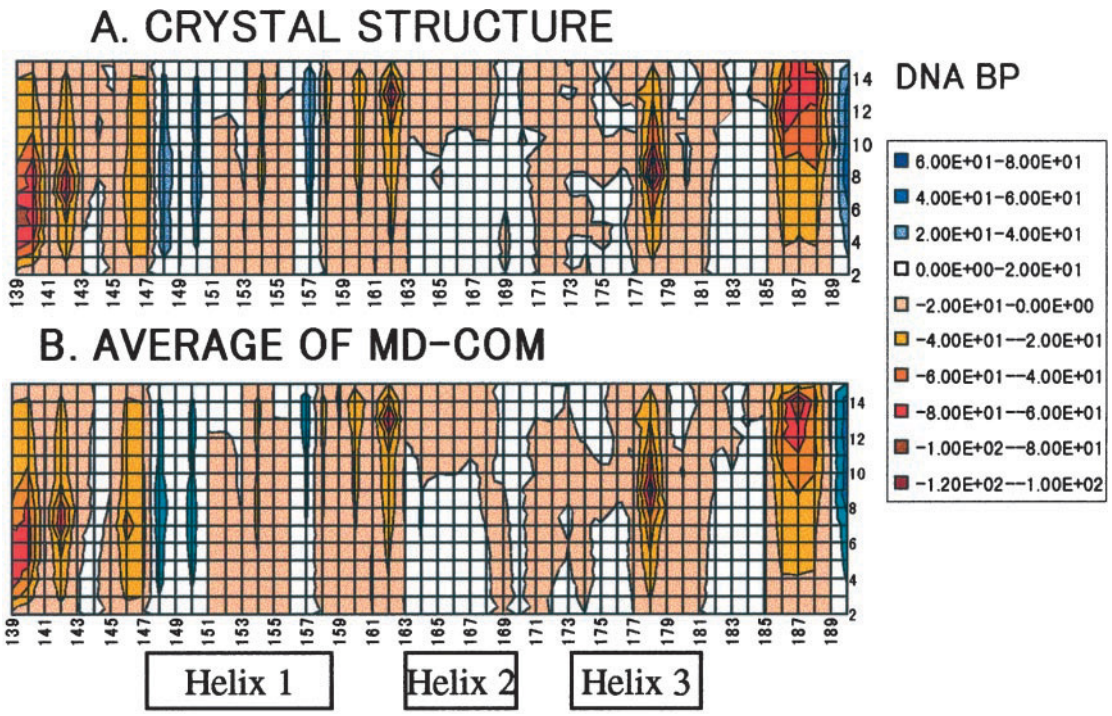


FIGURE 14 Contour plots of nonbonded interaction (electrostatic plus van der Waals) between amino acid residues of the peptide and DNA base pairs (*A*) in the crystal structure (obtained at the end of stage 1, Fig. 2) and (*B*) averaged over a 0.5–2.5-ns trajectory of MD-com. Negative interactions are shown in reddish colors, and positive ones in white or blue. The values in the box are given in kcal/mol. The summations of the interactions in the counterplots were -4.79×10^3 kcal/mol for the crystal structure and -4.58×10^3 kcal/mol for the average of MD-com.

CONCLUSION

Molecular dynamics simulations of the Hin-recombinase–DNA complex, the peptide, and the DNA were performed separately. The obtained trajectories showed good stability, as judged by their energetic quantities.

Both the peptide and DNA showed larger fluctuation in the free state than in the complex, but the stabilization due to the complex formation was larger for the peptide than for the DNA. The overall fluctuation patterns are well illustrated in Fig. 4. The N-terminal region and HTH motif of the peptide were stabilized largely by the interaction with the DNA. The C-terminus was also stabilized by the complex formation, but only marginally. The effect of the bound peptide on the DNA conformation was relatively small, but the bound peptide was suggested to enlarge the minor groove and to distort the DNA slightly from the canonical B-conformation. The central major groove was rigidly fixed by the interaction with the HTH motif. The ions and solvents preferentially stayed in the minor groove of the free DNA.

Our simulations were not intended to mimic the binding process itself, but some speculation may be possible. For the peptide, the simulation results strongly support the induced fit model of binding, in which the peptide changes conformation upon binding to DNA. When the peptide binds to DNA, both termini fit into the minor groove by replacing the solvent and ions. The DNA may also change its conformation slightly by widening the minor groove. The specific interaction between the peptide and the DNA should be mostly governed by the N-terminus and the HTH motif via numerous H-bonds, including solvent-mediated ones. Many of the H-bonds in the DNA-peptide interface may not be static, but rather may exist in a dynamic equilibrium among several binding sites. The C-terminus should participate in nonspecific DNA binding by loosely associating with the minor groove via the electrostatic interaction. In conclusion, comparison of the MD trajectory of the peptide-DNA complex, the free peptide, and the free DNA has provided a valuable contribution to the understanding of the DNA recognition process.

Visual inspection of the animations of the trajectories made by Dr. Yutaka Ueno of the Electrotechnical Laboratory was of critical importance to the current study. Dr. G. Ravishanker of Wesleyan University is acknowledged for provision of Toolchest 1.0. The AMBER 95 force field was obtained from the home page of Prof. P. A. Kollman of the University of California at San Francisco. Finally, Tsukuba Advanced Computing Center is acknowledged for computer support.

REFERENCES

Auffinger, P., and E. Westhof. 1998. Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* 8:227–236.

Berne, B. J., and J. E. Straub. 1997. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opin. Struct. Biol.* 7:181–189.

Bruist, M. F., S. J. Horvath, L. E. Hood, T. A. Steists, and M. Simon. 1987. Synthesis of a site specific DNA-binding peptide. *Science.* 235:777–780.

Chang, K.-Y., and G. Varani. 1997. Nucleic acids structure and recognition. *Nature Struct. Biol.* 4:854–858.

Cheatham, T. E., III, and P. A. Kollman. 1997. Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]$_2$ in the presence of hexaamminecobalt (III). *Structure.* 15:1297–1311.

Cieplak, P., T. E. Cheatham, III, and P. A. Kollman. 1997. Molecular dynamics simulations find that 3′ phosphoramidate modified DNA duplexes undergo a B to A transition and normal DNA duplexes an A to B transition. *J. Am. Chem. Soc.* 119:6722–6730.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Dang, L.-X., and B. M. Pettitt. 1987. Simple intramolecular model potentials for water. *J. Phys. Chem.* 91:3349–3354.

De Winter, H., E. Lescrinier, V. Van Aerschot, and P. Herdewijn. 1998. Molecular dynamics simulation to investigate differences in minor groove hydration of HNA/RNA hybrids as compared to HNA/DNA complexes. *J. Am. Chem. Soc.* 120:5381–5394.

Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1981. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA.* 78:2179–2183.

Duan, Y., P. Wilkosz, M. Crowley, and J. M. Rosenberg. 1997. Molecular dynamics simulation study of DNA dodecamer d(CGCGAATTCGCG) in solution: conformation and dynamics. *J. Mol. Biol.* 272:553–572.

Eriksson, M. A. L., T. Haerd, and L. Nilsson. 1995. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.* 68:402–426.

Feig, M., and B. M. Pettitt. 1998. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* 75:134–149.

Feng, J.-A. R., R. C. Johnson, and R. E. Dickerson. 1994. Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science.* 263:348–355.

Fenly, M. O., G. S. Manning, and W. K. Olson. 1990. Approach to the limit of counterion condensation. *Biopolymers.* 30:1191–1203.

Fukushige, T., M. Taiji, J. Makino, T. Ebisuzaki, and D. Sugimoto. 1996. A highly parallelized special-purpose computer for many-body simulations with an arbitrary central force: MD-GRAPE. *Astrophys. J.* 468:51–65.

Hughes, K. T., P. C. W. Gaines, J. E. Karlinsey, R. Vinayak, and M. I. Simon. 1992. Sequence-specific interaction of the *Salmonella* Hin recombinase in both major and minor grooves of DNA. *EMBO J.* 11: 2695–2705.

Jayaram, B., and D. M. Beveridge. 1996. Modeling DNA in aqueous solutions: theoretical and computer simulations studies on the ion atmosphere of DNA. *Annu. Rev. Biophys. Biomol. Struct.* 25:367–394.

Karplus, M., and G. A. Petsko. 1990. Molecular dynamics simulations in biology. *Nature.* 347:631–639.

Kollman, P. A. 1996. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Chem. Acc. Res.* 29:461–469.

Komeiji, Y., and M. Uebayasi. 1999. Molecular dynamics simulation of the Hin-recombinase-DNA complex. *Mol. Simul.* 21:303–324.

Komeiji, Y., M. Uebayasi, R. Takata, A. Shimizu, K. Itsukashi, and M. Taiji. 1997. Fast and accurate molecular dynamics simulation of a protein using a special-purpose computer. *J. Comput. Chem.* 18: 1546–1563.

Komeiji, Y., M. Uebayasi, and I. Yamato. 1993. A molecular dynamics study of solvent behavior around a protein. *Proteins.* 16:268–277.

Komeiji, Y., H. Yokoyama, M. Uebayasi, M. Taiji, T. Fukushige, D. Sugimoto, R. Takata, A. Shimizu, and K. Itsukashi. 1995. A high performance system for molecular dynamics simulation of biomolecules using a special-purpose computer. *In* Pacific Symposium on Biocomputing '96. L. Hunter and T. E. Klein, editors. World Scientific, Singapore. 472–487.

Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

Mack, D. P., J. P. Sluka, J. A. Shin, J. H. Griffin, M. I. Simon, and P. B. Dervan. 1990. Orientation of the putative recognition helix in the DNA-binding domain of Hin recombinase complexed with the hix site. *Biochemistry.* 29:6561–6567.

Makarov, V. A., M. Feig, B. K. Andrews, and B. M. Pettitt. 1998. Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophys. J.* 75:150–158.

Manning, G. S. 1978. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* 11:179–246.

McLachlan, A. D. 1979. Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* 128:49–79.

Merz, K. M., Jr. 1997. Molecular dynamics simulations of lipid bilayers. *Curr. Opin. Struct. Biol.* 7:511–517.

Miller, J. L., and P. A. Kollman. 1997. Observation of an A-DNA to B-DNA transition in a nonhelical nucleic acid hairpin molecule using molecular dynamics. *Biophys. J.* 73:2702–2710.

Nelson, H. C. M., J. T. Finch, B. F. Luisi, and A. Klug. 1987. The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature.* 330:221–226.

Noberto de Souza, O., and R. L. Ornstein. 1997. Effect of warmup protocol and sampling time on convergence of molecular dynamics simulations of a DNA dodecamer using AMBER 4.1 and particle-mesh Ewald method. *J. Biomol. Struct. Dyn.* 14:607–611.

Norin, M., F. Haeffner, K. Hult, and O. Edholm. 1994. Molecular dynamics simulations of an enzyme surrounded by vacuum, water, or a hydrophobic solvent. *Biophys. J.* 67:548–559.

Nose, S. 1991. Constant temperature molecular dynamics methods. *Prog. Theor. Phys. Suppl.* 103:1–46.

Ravishanker, G. S., and D. L. Beveridge. 1997. Toolchest, Version 1.0. Wesleyan University, Middletown, CT.

Ravishanker, G. S., S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar. 1989. Conformational and helicoidal analysis of 30 ps of molecular dynamics on the d(CGCGAATTCGCG) double helix: "curves," dials and windows. *J. Biomol. Struct. Dyn.* 6:669–699.

Robinson, C. R., and S. G. Sligar. 1996. Participation of water in Hin recombinase-DNA recognition. *Protein Sci.* 5:2119–2124.

Sayle, R. 1995. RASMOL, Version 2.6. Glaxo Research and Development, Hertfordshire, U.K.

Schui, X., L. McFail-Isom, G. G. Hu, and L. D. Williams. 1998. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry.* 37:8341–8355.

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

Shrake, A., and J. A. Rupley. 1973. Environment and exposure of solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79:351–371.

Sluka, J. P., S. J. Horvath, A. C. Glasgow, M. I. Simon, and P. D. Dervan. 1990. Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin recombinase. *Biochemistry.* 29:6551–6561.

Sprous, D., M. A. Young, and D. L. Beveridge. 1998. Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: theoretical considerations of the A↔B transition. *J. Phys. Chem. B.* 102:4658–4667.

Tang, Y., and L. Nilsson. 1998. Interaction of Human SRY protein with DNA: a molecular dynamics study. *Proteins.* 31:417–433.

Tidor, B. 1997. Molecular dynamics simulations. *Curr. Biol.* 7:525–527.

Tuckerman, M., B. J. Berne, and G. J. Martyna. 1992. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 97:1990–2001.

Young, M. A., B. Jayaram, and D. L. Beveridge. 1997a. Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* 119:59–69.

Young, M. A., G. Ravishanker, and D. L. Beveridge. 1997b. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys. J.* 73:2313–2336.

Zieg, Z., M. Silverman, M. Hilmen, and M. Simon. 1977. Recombinational switch for gene expression. *Science.* 196:170–172.