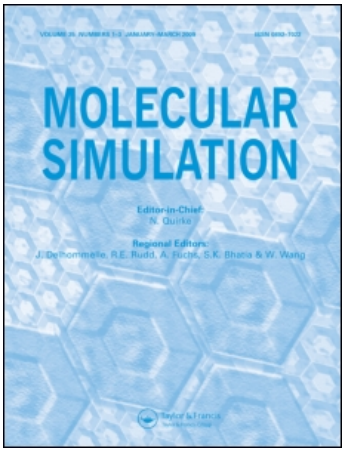

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Molecular Dynamics Simulation of the Hin-Recombinase—DNA Complex

Yuto Komeiji[a]; Masami Uebayasi[b]
[a] Electrotechnical Laboratory, Ibaraki, Japan [b] National Institute of Bioscience and Human-Technology, Ibaraki, Japan

To cite this Article Komeiji, Yuto and Uebayasi, Masami(1999) 'Molecular Dynamics Simulation of the Hin-Recombinase—DNA Complex', Molecular Simulation, 21: 5, 303 — 324
To link to this Article: DOI: 10.1080/08927029908022071
URL: http://dx.doi.org/10.1080/08927029908022071

Molecular Simulation, 1999, Vol. 21, pp. 303–324
Reprints available directly from the publisher
Photocopying permitted by license only

MOLECULAR DYNAMICS SIMULATION OF THE HIN-RECOMBINASE–DNA COMPLEX

YUTO KOMEIJI [a,*] and MASAMI UEBAYASI [b]

[a] *Electrotechnical Laboratory, 1-1-4 Umezono, Tsukuba, Ibaraki 305-8568, Japan;*
[b] *National Institute of Bioscience and Human-Technology 1-1-3 Higashi, Tsukuba, Ibaraki 305-8566, Japan*

(Received February 1998; accepted July 1998)

Molecular dynamics simulation was performed for 1 ns for a peptide:DNA complex formed by Hin recombinase 52mer peptide and a 13 base-pair synthetic *hix*L half-site. The Ewald summation method was employed to treat the electrostatic interaction without the cutoff. Comparison with the results of the conventional cutoff method revealed that proper treatment of the electrostatic interaction was required to produce a stable trajectory. The trajectory thus obtained was used to analyze the interaction between the peptide and the DNA. Both N- and C-terminal regions reside within two regions of the minor groove. The N-terminus-DNA interaction was fairly stable in the current simulation, but the C-terminus-DNA interaction was only marginally stable. These simulation results were consistent with reported experimental data that the N-terminal residues were required for the DNA recognition, while the C-terminal ones were supportive but not necessary.

Keywords: Molecular dynamics; peptide; DNA; Ewald; cutoff; Hin recombinase

1. INTRODUCTION

In this paper, we report molecular dynamics (MD) simulation of a peptide:DNA complex formed by the C-terminal 52mer peptide of Hin recombinase and a 13 base-pair (bp) *hix*L half site.

Hin recombinase is an enzyme that inverts a DNA segment regulating the expression of the flagellin genes of *Salmonella typhimurium* [1], and initiates the sequence of events of the recombination reaction by binding to the two

* Corresponding author. Tel.: 81-298-54-5164, Fax: 81-298-54-2154, e-mail: komeiji@etl.go.jp.

homologous crossover sites, *hix*L and *hix*R, of the supercoiled DNA (see, for example, [2]). The protein consists of 190 amino acid residues, but the carboxyl terminal 52 amino acids were shown to be sufficient for binding to the *hix* sites [3].

The crystal structure of the 52mer:*hix*L complex was resolved to atomic resolution at 1.8 Å ([4], PDB entry 1HCR, Fig. 1). Figure 2 shows the amino acid and base sequences of the 52mer and the synthetic *hix*L half site. The numbering obeys Ref. [4]. The 52mer peptide has three α-helices, the third of which forms the helix-turn-helix motif that binds to the major groove of DNA. In addition, both termini of the peptide reside within two regions of the minor groove. We consider this complex an ideal target for the study of a peptide-DNA interaction because of its simple architecture.

Though there were a number of pioneering studies, reliable MD simulations of nucleotides had long been difficult. The electrostatic cutoff method, often used in biomolecular simulations, causes serious artifacts in the structure and dynamics of the nucleotides because they are highly

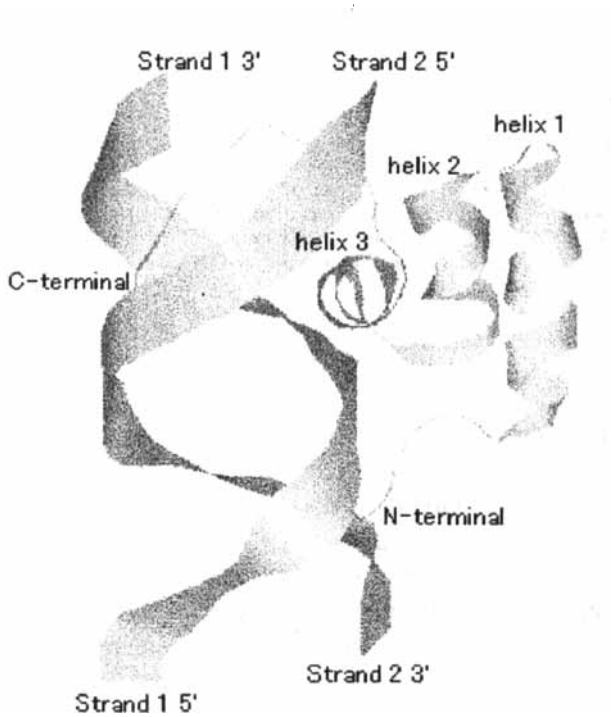

FIGURE 1 The ribbon presentation of the crystal structure of the Hin recombinase 52mer: DNA complex [4].

(a) Amino acid sequence of the 52mer

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

(b) DNA base sequence of the *hix*L 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 2 The sequences of (a) the 52mer peptide and (b) *hix*L half site.

charged molecules (see Appendices C and D). Recently, reliable MD simulations of DNA and RNA have become possible owing to the availability of MD methods without the electrostatic cutoff [5, 6], stimulated by the development of sophisticated algorithms and by the improvement of computer hardware. However the number of MD simulations of protein:DNA/RNA complexes without the electrostatic cutoff is still limited (Glucocorticoid Receptor [7, 8], Homeodomain [9], Estrogen Receptor, [10]). The artifacts arising from the electrostatic cutoff are usually less evident in MD simulations of polypeptides than in those of nucleotides. Nevertheless, several recent studies have shown that avoidance of the cutoff is also desired, or sometimes critical, for MD of polypeptides to obtain reliable results [11–17].

We have recently constructed a high performance system for MD simulation of biomolecules named PEACH-GRAPE [16, 18]. PEACH (Program for Energetic Analysis of bioCHemical molecules) is a software for biomolecular simulation which employs a special-purpose processor named MD-GRAPE (GRAvity piPE for MD [19]) to compute the nonbonded interactions. This system performs MD in spherical or periodic boundary conditions without the electrostatic cutoff [16].

We consider the Hin 52mer:DNA complex a suitable target to simulate by the PEACH-GRAPE system because the electrostatic interaction should play an important role in the DNA recognition by the peptide. We performed MD of the complex by the Ewald method [20], and report of the interactions involved in the DNA recognition. We also produced an MD trajectory by the cutoff method and compared it with that obtained by the Ewald method to demonstrate the improvement (Appendices C and D).

2. RESULTS AND DISCUSSION

2.1. Feasibility of the Generated Trajectory

An MD simulation of the peptide:DNA complex was performed for 1 ns in a periodic box of water as described in Appendix A (see Fig. 3 for the initial configuration). The Ewald method was used to compute the electrostatic interaction (see Appendix B for determination of the Ewald parameters). The trajectory was generated in the microcanonical ensemble, which is suitable to investigate dynamics because it uses no velocity or coordinate scaling.

At first, we studied the total energy and temperatures of the generated trajectory to demonstrate the feasibility of the simulation protocol. We computed the averages and RMS (root mean square) fluctuations of the total energy and temperatures from the 300 – 1000 ps trajectory (Tab. I). The total energy of the system should be conserved in the microcanonical ensemble. The excellent energy conservation (0.05% for a period of 700 ps)

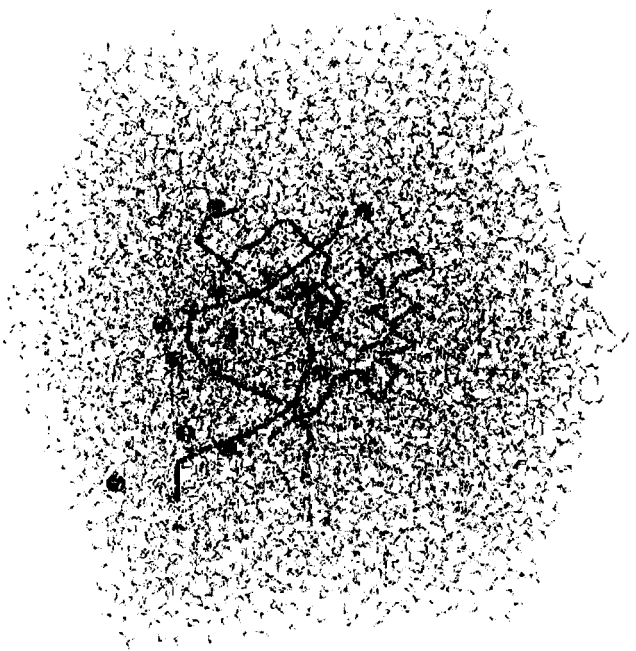

FIGURE 3 The initial configuration for MD of the peptide:DNA complex. The complex was immersed in a cubic box of water ($60.8 \times 60.8 \times 60.8$ Å^3), in which the ions and solvent had been optimized and bad contacts within the solutes had been eliminated (Appendix A). The numbers of atoms were: peptide 875, DNA 859, Na^+ ions 16, and solvent 19,092. In total, 20,812 atoms were simulated.

TABLE I Summary of the MD trajectory (300–1000 ps)

Total Energy (10^4 kcal/mol):	-5.2468 ± 0.0025 (0.05%)	
Temperature (K):	total	301.1 ± 1.4
	solute	302.1 ± 6.1
	ion	301.1 ± 62.4
	solvent	301.0 ± 1.5
RMSD (Å*):	peptide	2.82 (main chain 2.18 side chain 3.31)
	DNA 1	2.72 (main chain 3.17 side chain 2.12)
	DNA 2	2.16 (main chain 2.43 side chain 1.78)

* Time RMS of the RMSD of the trajectory from the crystal structure was computed. Hydrogen atoms were neglected.

indicated the feasibility of the precision of the force calculations and the integration method [21].

Separation of the temperature between the protein and the solvent has been reported in several MD simulations of solvated proteins [16, 22–24]. This phenomenon is caused by the abrupt cutoff of the nonbonded interactions. Our simulation used the high precision Ewald method to compute the electrostatic interaction. Hence, the solute, ion, and solvent had the same temperature of 300 K within acceptable errors (Tab. I). See Appendices C and D for comparison with the result of the cutoff MD.

Next, we investigated the root mean square deviation (RMSD) of the generated structures from the crystal structure. The time evolution of the RMSD is shown in Figure 4, and the RMS of the RMSD's of the 300–1000 ps trajectory are listed in Table I. RMSD of both the peptide and the DNA strands were fairly stable and converged at $< \sim 3$ Å. The time-averaged structure of the trajectory was superimposed onto the crystal structure (Fig. 5). This figure illustrated that the MD structure evolved within the realm of the crystal structure. For the peptide, the main chain atoms had smaller RMSD values than the side chain atoms because the side chains were more exposed to the solvent. In the case of the DNA, the main chains were exposed to the solvent, and were more mobile, and hence had larger RMSDs than the side chains (Tab. I).

2.2. Analyses of the Peptide–DNA Interaction

The energetic quantities and RMSD from the crystal structure thus showed the stability and the feasibility of the generated trajectory. We further analyzed the trajectory to investigate the peptide–DNA interaction.

The RMSD's from the crystal structure of the amino acid residues and DNA bases were determined for the average structure of the 300–1000 ps trajectory (Figs. 6(a) and 7) to investigate which parts of the molecules were

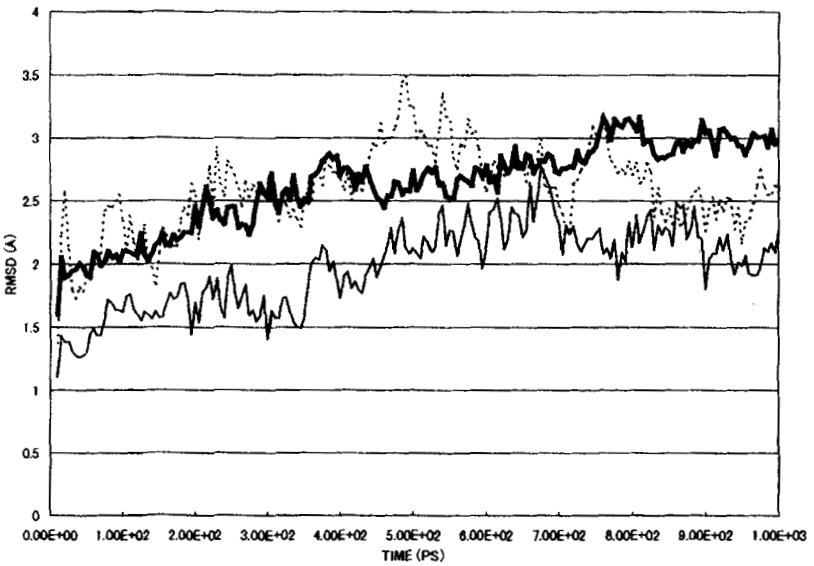

FIGURE 4 Time evolution of RMSD of the simulated structures from the crystal structure. Protein: thick solid line. DNA strand 1: dotted line. DNA strand 2: thin solid line. Hereafter, each generated structure was fitted to the initial crystal structure as described [43]. Hydrogen atoms were neglected.

FIGURE 5 Backbone presentation of the time average of the simulated structure (300 – 1000 ps, gray) superimposed onto the crystal structure (black).

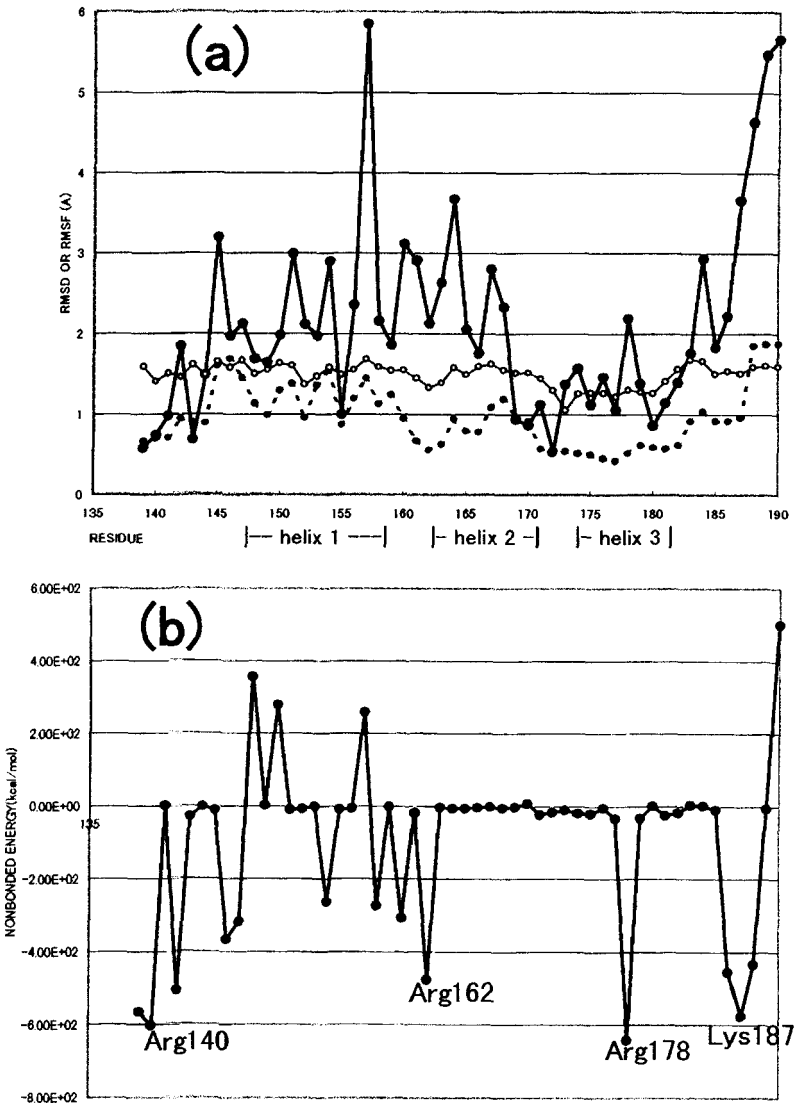

FIGURE 6 (a) RMSD of the time-averaged structure (300–1000 ps) of the peptide from the crystal structure (thick solid line with black circles). The residue RMSFs were also calculated from the trajectory (dotted line with black circles) and crystallographic temperature factors (thin solid line with white circles). (b) Time-averaged nonbonded interaction energies (300–1000 ps) between the peptide residues and the DNA. Several amino acids interacting strongly are labelled. Their interactions with the DNA were further analyzed in Figure 10.

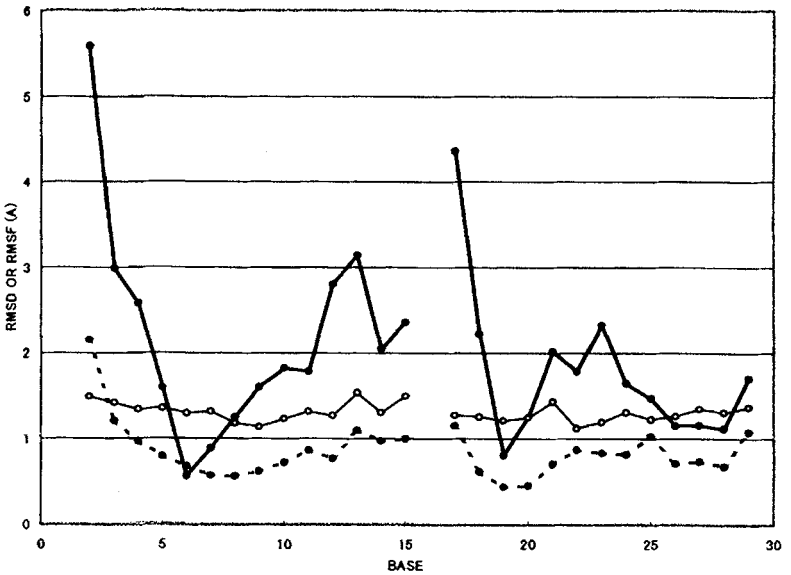

FIGURE 7 RMSD of the time-averaged structure (300–1000 ps) of the DNA from the crystal structure. The RMSFs of the bases were also calculated from the trajectory and crystallographic temperature factors. See legend to Figure 6(a) for definition of the lines.

deviated. RMS fluctuations (RMSF) of the amino acid residues and the DNA bases were also calculated from the MD trajectory and, for comparison, from the temperature factors (B-factors) of the X-ray crystallographic data (Figs. 6(a) and 7). RMSD shows the deviation of the trajectory from the initial structure, while RMSF shows the deviation from the time-average of the trajectory. Several snapshots of the simulated structures were visualized in Figure 8 to illustrate the overall structural fluctuation.

RMSF's from the MD and B-factor were compared (Figs. 6(a) and 7). The patterns agreed to each other qualitatively, but not quantitatively. The reasons for the disagreement between the MD and B-factor have been discussed frequently. Hunenberger *et al.* [25] compared fluctuations from nanosecond-order MD and that from the B-factor in detail and concluded that quantitative agreement between them cannot be expected. In addition, our simulation was intended to mimic the solution structure of the complex, while the B-factors reflected the fluctuations in a crystalline environment. Hence, we did not consider that our MD simulation was inadequate although the RMSF from the MD and B-factor did not agree quantitatively. It was also important to note that, because our simulation was only 1 ns

FIGURE 8 Several snapshots of the backbone structures of the trajectory (300–1000 ps). The simulated structures sampled at every 40 ps period were visualized by the backbone presentation.

long, the RMSF from the MD presumably reflected only fast motions within the complex.

RMSD and RMSF of the peptide indicated that the N-terminus and helix 3 were stable (Fig. 6a). These regions were involved in DNA recognition (Fig. 1). The C-terminal residues were also involved in DNA recognition by interacting with the minor groove (Fig. 1), but they were relatively unstable in this simulation. Helix 1 was more flexible than the other two helices (Figs. 6(a) and 8). This seemed reasonable because helix 1 was exposed to the solvent (Fig. 1) and not directly involved in DNA recognition.

The stability of the hydrogen bonds (H-bonds) present in the crystal were examined for the MD trajectory. H-bonds between the peptide and DNA are schematically presented in Figure 9. Crystallographic H-bonds were defined as those in which donor–acceptor distances (R_{DA}) were < 3.5 Å. The stability of each H-bond during MD was determined by using R_{DA} only (see legend to Fig. 9). Such definitions of H-bond stability were somewhat arbitrary, but were adequate for characterization of the H-bonding pattern. As seen in Figure 9, the N-terminal region, helix 3, and the C-terminal region were making H-bonds with DNA. The H-bonds due to the N-terminal region and helix 3 were mostly stable during MD, but those of the C-terminal region were not. The H-bonds due to the C-terminal three residues, Arg^{188}–Met^{189}–Asn^{190}, were particularly unstable. The stability and instability of the H-bonds were thus consistent with the RMSD (Fig. 6a).

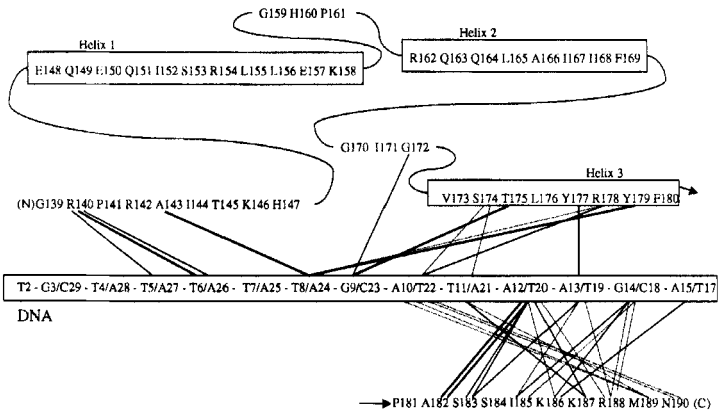

FIGURE 9 Schematic presentation of the H-bonds between the peptide residues and the DNA base pairs. H-bonds were judged by the distance between the donor and the acceptor atoms (R_{DA}). Stability of the hydrogen bonds present in the crystal ($R_{DA} < 3.5$ Å) were examined during the MD simulation. The time average of R_{DH}, during the 300–1000 ps trajectory was used to measure the stability. Stable hydrogen bonds ($R_{DA} < 3.5$ Å) were shown in thick solid lines, marginally stable ones ($3.5 \leq R_{DA} < 5.0$) in thin solid lines, and unstable ones ($R_{DA} \geq 5.0$ Å) in dotted lines.

We also computed the nonbonded interaction energy (electrostatic plus van der Waals) between the peptide and DNA averaged over the trajectory (Fig. 6(b)). Negative interactions were considered to contribute favorably with the formation of the complex. In general, formation of a complex by two molecules depends upon a subtle balance of the enthalpy and entropy of the system including the solvent, and cannot be explained only by the interaction energy between two molecules. However, in an analysis of a substrate bound form of a protein [26], we experienced that the interaction energy gave fairly reasonable accounts for the experimental data. Thus, discussion below based on Figure 6(b) should hold at least qualitatively.

The N-terminal residues, $Gly^{138}-Arg^{140}-Pro^{141}-Arg^{142}$, were interacting favorably with the DNA. Helix 1 had both positively and negatively charged residues (Fig. 2(a)), so the interaction with the DNA, in total, was not strong. Helix 2 consisted of nonpolar amino acids (Fig. 2(a)), and had no contribution to the peptide–DNA interaction. Thus, the main role of helix 2 is probably in the formation of the hydrophobic core of the peptide, rather than in the involvement of DNA recognition. Most of the amino acids of helix 3 were nonpolar, and had little contribution to the peptide–DNA interaction. Presumably, formation of an α-helix that fits well into the major groove of the DNA (Fig. 1) should be of primary importance. The

only exception, Arg^{178}, contributed largely to the interaction with the DNA. Several residues near the C-terminal, $Lys^{186}-Lys^{187}-Arg^{188}$, interacted favorably with the DNA.

We chose four important residues (Fig. 6(b)), and analyzed their interactions with the DNA base pairs in detail (Fig. 10). Naturally, they were interacting the most strongly with nearby base pairs. However, interactions with distant base pairs were not negligible, especially in the case of Arg^{140}.

Some experimental data are available from the literature concerning the roles of the N- and C-terminal regions. The deletion of the N-terminal $Gly^{139}-Arg^{140}$ was reported to cause complete loss of the DNA-binding affinity [27]. As discussed above, the N-terminus was quite stable in our MD simulation, though the contact between the N-terminus and the DNA was rather small (Fig. 1). Base pairs 5–6 (AT–AT) recognized by Arg^{140} were conserved among the enteric bacterial inversion sites (Feng *et al.*, 1994). Consistently, Arg^{140} interacted the most favorably with base pairs 5–6 (Fig. 10). Arg^{140} also had favorable interactions with distant base pairs, suggesting that it was an important residue in the stabilization of the complex.

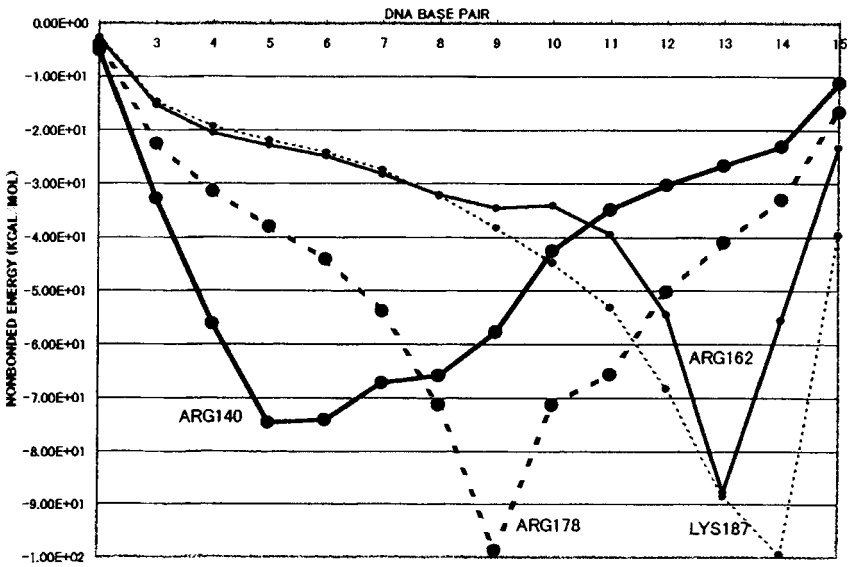

FIGURE 10 Nonbonded interactions between several amino acid residues and the DNA base pairs. Base-pair 3, for example, indicates the base pair formed by bases 3 and 29. See Figure 2 (b) for the base sequences.

As we mentioned above, the crystal structure of the C-terminal region was rather unstable and flexible in the current simulation (Figs. 6(a) and 9). The C-terminal residues were interacting with the terminal base pairs, which were highly flexible and deviated from the initial structure (Figs. 5, 7 and 8). Their instability and fluctuations should be correlated. Even though the C-terminal region was flexible, they remained, on average, within the minor groove (Figs. 5 and 8). It was demonstrated that the Hin peptide without the C-terminal eight residues retained DNA-binding activity, although the affinity was reduced [28]. Hence, we believe that the marginal stability of the C-terminal region of the peptide in our MD trajectory was consistent with the auxiliary role played by the C-terminal region in DNA recognition.

More minute computational methods such as the free energy calculation are necessary for further discussion about specific interactions between the peptide and DNA, but the analyses we have shown so far based on the structural fluctuations and interaction energies qualitatively agreed with the experimental results.

In conclusion, we succeeded in performing a feasible simulation of the Hin-recombinase 52mer peptide:*hix*L complex. The simulation results agreed with the available experimental results. For further analyses of the peptide:DNA interaction, we are now conducting MD simulations of the uncomplexed DNA and peptide, along with the further extension of the simulation of the complex. We expect to reveal the change in dynamics and energetics upon complex formation by comparing the simulation result of the complex with those of the components, as well as the change in solvation and ionic distribution.

Acknowledgement

This study was performed through the special research fund from the Agency of Industrial Science and Technology of the Japanese government.

APPENDICES

A. Methods

All of the computations were performed using the software package PEACH on a DEC personal workstation model 500/500 (0.59 Gflops) connected with the special-purpose computer MD-GRAPE (or ITL-MD one, ITL Corp. Tokyo, Japan; 4 Gflops/board × 2). MD-GRAPE is a

parallel pipeline processor that computes the nonbonded interactions [19]. See [16, 18] for details of the PEACH-GRAPE system. The VME bus adapter needed in the original system was not used. Instead, the MD-GRAPE boards were directly connected to the host workstation using the PCI bus. This improvement increased the data transfer rate between GRAPE and the host.

The crystal structure of the complex composed of the 52mer peptide of the Hin-recombinase and *hix*L half site (PDB entry 1HCR, [4]) was used as the initial structure for the MD (Fig. 1). Hereafter, we call the peptide:DNA complex 'solutes', the water molecules 'solvent', and the counter ions 'ions'. Hydrogen atoms were generated for the crystal structure and optimized by energy minimization (EM). The complex was placed in a cubic box so that the minimum distance between the wall and the solute should become 10 Å. The box was filled with the flexible SPC water [29], and the minimum distance between the solvent and the solute was 3 Å. Sixteen counter ions (Na^+) were generated to neutralize the system by replacing sixteen water molecules. The ions were positioned so as to minimize the solute-ion electrostatic interactions. All of the ions were generated around the DNA, rather than near the peptide, and no ion was placed in the interface between the peptide and DNA. See the legend to Figure 3 for the size of the system.

The latest AMBER all-atom force field [30] was used throughout the study. The Ewald summation of the electrostatic interaction was performed as described [16] in all of the EM and MD simulations. See Appendix B for the determination of the Ewald parameters. For integration in MD simulations a multiple time step method was employed (RESPA, REference System Propagator Algorithm, [31]). The time steps were 0.25 fs for the bond, angle, and torsion interactions (hard force); 1 or 2 fs for the van der Waals (VDW) and the Ewald real-space (r-space) (medium force); and 4 fs for the Ewald k-space (soft force). The medium force time step was 2 fs during the optimization of the solvent and ions, and the subsequent optimization of the solutes, while 1 fs was employed during the production run. The former condition gave 0.13% relative fluctuation of the total energy and the latter gave 0.02% in preliminary MD simulations for as long as 0.4 ps (not shown). The Nose-Hoover thermostat [32] was employed to keep the temperature constant as described [16].

At first, the positions of the ions and the solvent were optimized as follows while the solute atoms were kept frozen. The steepest descent EM was performed for 500 steps. Then MD simulation was performed at 5 K for 0.1 ps. The time constant, τ, for the temperature control was 0.001 ps. The velocity-Verlet method was used with a time step of 0.25 fs during the first

0.1 ps, otherwise RESPA was employed. The temperature was gradually heated to 300 K (τ = 0.1 ps) during the first 6 ps, and the simulation was continued until 50 ps. After 10 ps, τ was reset to 0.5 ps. Then the system was gradually cooled to 5 K for 12 ps. Thus, MD was conducted for 62 ps in total to optimize the solvent and ions. Then the ions and solvent were subjected to EM. Secondly, the solutes were optimized while keeping the solvent and ions fixed. The simulation protocol was similar to that described above, but MD at 300 K was continued only until 30 ps. Hence, MD was conducted for 42 ps in total to optimize the solutes. Then EM was performed for the whole system (the solutes, ions, and solvent). The structure thus obtained was used as the initial structure for the final production run (Fig. 3). After the optimization, most of the ions still resided around the DNA molecules.

The production run was conducted as follows. The heating and equilibration protocol was essentially the same as above, but the heating was performed more slowly (for 9 ps). The temperature control was turned off after 100 ps to perform the microcanonical MD, and the simulation was continued until 1000 ps. The computation time of 4 s was required to perform 1 fs MD. The trajectory of 300 – 1000 ps was used to analyze the structure and energetics. The crystal and simulated structures were visualized by the graphics software RASMOL 2.6 [33].

B. Determination of the Parameters for Ewald Summation by MD-GRAPE

Three parameters should be determined to perform the Ewald summation; R_{cut}, the cutoff radius for the real-space summation (r-space); $K_{\max}$, the maximum of the integer wave number vector (k-space); and η, the parameter to determine the balance between the real-space and k-space summations (see [16] for detailed definitions). Here we describe the process of determination of these parameters for computation by MD-GRAPE. A cubic simulation of side L will be considered.

Below are the r-space (U_r) and k-space (U_k) summations of the electrostatic energies.

$$U_r = \sum_{i>j, r_{ij}<R_{\text{cut}}} q_i q_j \operatorname{erfc}(r_{ij}/\eta)/r_{ij} \tag{B1}$$

$$U_k = \frac{1}{2} \sum_{k<k_{\max}} \frac{\exp\left(-(\pi\eta\mathbf{k}/L)^2\right)}{(\mathbf{k}/L)^3} \left\{ \left(\sum_i \frac{q_i \sin 2\pi\mathbf{k} \bullet \mathbf{r}_i}{L} \right)^2 + \left(\sum_i \frac{q_i \cos 2\pi\mathbf{k} \bullet \mathbf{r}_i}{L} \right)^2 \right\} \tag{B2}$$

First, we consider the fundamental relation among the parameters and the precision, similarly to [34].

The convergence of the r-space and k-space summations mainly depends on $\mathrm{erfc}(r_{ij}/\eta)$ and $\exp(-(\pi\eta k/L)^2)$, respectively. The complimentary error function $\mathrm{erfc}(x)$ becomes $\exp(-x^2)$ at large values of x. Hence, if the relative error of $\varepsilon = \exp(-p)$ is necessary at $r_{ij} = R_{\mathrm{cut}}$ and $k = k_{\max}$, the following should be satisfied:

$$p = \left(\frac{R_{\mathrm{cut}}}{\eta}\right)^2 = \left(\frac{\pi\eta k_{\max}}{L}\right)^2 \tag{B3}$$

It leads to the equations below:

$$k_{\max} = \frac{LR_{\mathrm{cut}}}{\pi\eta^2} \qquad \eta = \frac{R_{\mathrm{cut}}}{\sqrt{p}} \tag{B4}$$

For the r-space summation, MD-GRAPE computes all the interactions between the particles using the minimum image convention. Thus, the cutoff radius should be half of the box size.

$$R_{cut} = \frac{L}{2} \tag{B5}$$

Substituting R_{cut} in Eqs. (B4) with (B5) it follows that

$$\eta = \frac{L}{2\sqrt{p}}, \qquad k_{\max} = \frac{2p}{\pi} \tag{B6}$$

The optimal values for $k_{\max}$ and η using Eqs. (B6) were listed for several values of ε in Table II. We used a $k_{\max}$ value of 10, which should give energy precision better than 10^{-6}.

There may not be any practical problem in using η derived by Eq. (B6), but the above discussion is rough. Also, we have so far considered only the energy and not the force. Hence, we determined η optimal for both energy

TABLE II Precision dependence of the Ewald parameters. See Appendix B for definitions of the parameters

ε	k_{max}	η/L
10^{-3}	4.40	0.190
10^{-4}	5.86	0.165
10^{-5}	7.33	0.147
10^{-6}	8.81	0.134
10^{-7}	10.26	0.125

and force empirically as follows to compare with the value obtained by Eq. (B6).

We used a configuration of the peptide:DNA complex in the cubic water, in which bad contacts had been removed by EM and a short MD simulation. Then we computed both the electrostatic energy ($E(\eta)$) and force ($f_i(\eta)$) for different η values. For a pair of η values (η_1 and η_2), the relative force difference ($Df(\eta_1, \eta_2)$) was defined as follows:

$$Df(\eta_1, \eta_2) = \sqrt{\sum_i |\mathbf{f}_i(\eta_1) - \mathbf{f}_i(\eta_2)|^2 \Big/ \left(\sum_i |\mathbf{f}_i(\eta_1)|^2 + \sum_i |\mathbf{f}_i(\eta_2)|^2 \right) \Big/ 2} \tag{B7}$$

where i is the particle index, and f_i is the force acting on particle i. Similarly, the relative energy difference ($De(\eta_1, \eta_2)$) was computed as below.

$$De(\eta_1, \eta_2) = |\{E(\eta_1) - E(\eta_2)\}/\{E(\eta_1) + E(\eta_2)\}/2| \tag{B8}$$

The meanings of Df and De were simple: the magnitude of the difference between the forces or energies computed for η_1 and η_2 relative to their absolute values. Df and De were computed for $(\eta_1, \eta_2) = (4, 5), (5, 6), (6, 7)$ and so on (Tab. III). We assumed that the optimal η was the η that gave the smallest difference with neighboring η's. Table III showed that $\eta = 7 \sim 8$ Å should be optimal both for the force and energy. L was 62.83 Å in the current simulation, so $\eta = 7 \sim 8$ Å should correspond to $\eta/L = 0.12 \sim 0.13$, close to the value of 0.125 obtained by Eq. (B6) for $k_{max} = 10.3$ (Tab. II). This agreement indicated the feasibility of Eq. (B6), which were derived by purely theoretical considerations.

Thus, we used $\eta = 7$ Å and $k_{max} = 10$ (4169 wave number vectors) for the simulation of the peptide:DNA complex throughout this study, as the values had been shown to be feasible both theoretically and empirically.

TABLE III Difference in Ewald energy and force for neighboring η values. See Appendix B for definitions of the parameters

η_1	η_2 (Å)	Df	De
5	6	1.4×10^{-4}	2.0×10^{-5}
6	7	6.4×10^{-6}	5.0×10^{-7}
7	8	5.9×10^{-7}	1.0×10^{-10}
8	9	3.7×10^{-6}	5.9×10^{-6}
9	10	3.4×10^{-5}	5.1×10^{-5}

C. Comparison of the Ewald Method with the Cutoff Method

To demonstrate the improvement brought about by the Ewald method, another MD trajectory was generated using the electrostatic cutoff method. The cutoff MD simulation was started from the 100 ps configuration of the trajectory obtained by the Ewald method (Appendix A), and was continued for a further 400 ps using a residue-based cutoff radius of 10 Å. We tried the microcanonical ensemble for the cutoff MD, as we did for the Ewald MD, but the system temperature increased up to 2000 K within 100 ps after we stopped the temperature control (not shown). Hence, we gave up the microcanonical ensemble and used the Nose-Hoover thermostat to keep the temperature at 300 K throughout the cutoff MD.

The energetic quantities computed from the last 300 ps trajectory of the cutoff MD were summarized in Table IV, which should be compared with Table I.

The fluctuation of the total energy was large (0.9%) but acceptable because it was not conservative in Nose dynamics. Rather, the Nose Hamiltonian, the Hamiltonian of the extended system including the Nose particle (see [16] for definition), should have been conserved. However, conservation of the Nose Hamiltonian was poor in the cutoff MD (11.5% of relative error, Tab. IV), indicating that the energetic stability of the system suffered from the truncation noise. The temperature separation, not seen in the Ewald MD (Tab. I), was prominent in the cutoff MD (Tab. IV). See Appendix D for further discussion.

Next, we examined the structural stability of the cutoff trajectory. The RMSD from the crystal is shown in Figure 11. The RMSD did not stabilize during the period, and that of DNA strand 1 reached 10 Å at the end of the simulation. The average structure of the cutoff trajectory was superimposed onto the crystal structure (Fig. 12). The structure of the DNA was

TABLE IV Summary of the cutoff MD trajectory (200–500 ps)

Total Energy (10^4 kcal/mol):	-5.1264 ± 0.0438 (0.9%)-5.1264 ± 0.0438 (0.9%)	
Nose Hamiltonian (10^4 kcal/mol):	-4.2660 ± 0.473 (11.1%)	
Temperature (K):	total	300.0 ± 2.9
	solute	331.1 ± 8.5
	ion	486.2 ± 107.7
	solvent	297.0 ± 3.0
RMSD (Å)*:	peptide	5.31 (main chain 4.59 side chain 5.91)
	DNA1	8.35 (main chain 9.40 side chain 7.01)
	DNA2	6.27 (main chain 6.04 side chain 6.53)

* Time RMS of the RMSD of the trajectory from the crystal structure was computed. Hydrogen atoms were neglected.

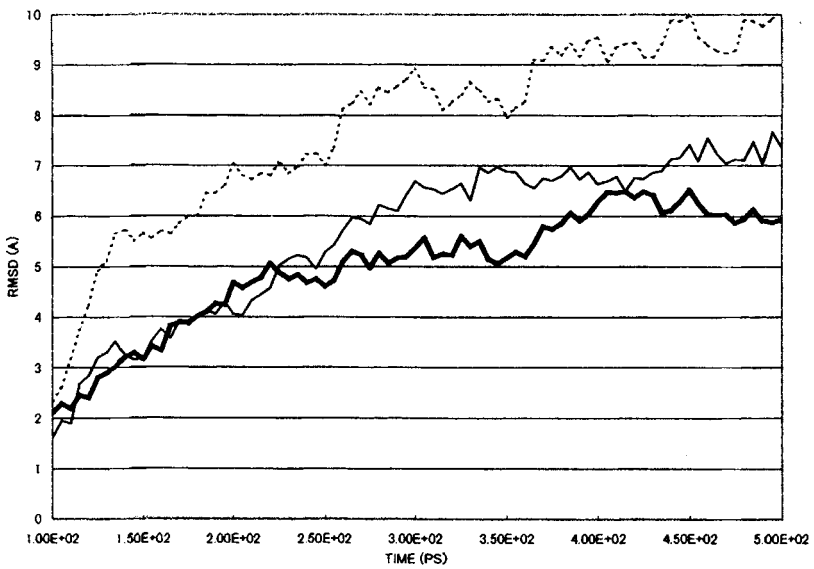

FIGURE 11 Time evolution of RMSD of the simulated structures from the crystal structure in the cutoff MD. The trajectory was obtained by the residue-based nonbonded cutoff scheme (10 Å). The cutoff simulation was started from the 100 ps configuration of the Ewald MD, so this graph begins at 100 ps. See legend to Figure 4 for definitions of the lines. When comparing with Figure 4, note that the scale of Figure 11 is larger than that of Figure 4.

particularly deformed. This result was consistent with several previous findings that crystal structures of DNA and RNA molecules were better conserved when the cutoff was avoided [6, 35].

Thus, the abrupt cutoff scheme brought about undesirable artifacts both to the energetics and to the structures of the trajectory. Hence, considering the improvement of the computer software and hardware, the electrostatic cutoff should be avoided when simulating a highly charged system such as the peptide:DNA complex studied here.

Several algorithms are available to compute the electrostatic interaction without the cutoff. For MD of a non-periodic system of size N, direct summation of the electrostatic interaction is straightforward. Billeter *et al.* [9] applied this method to the MD of a solvated Homeodomain:DNA complex. In direct summation, however, computation time is in proportion to $O(N^2)$, and this method becomes too expensive when $N > 10{,}000$, unless one is equipped with inexpensive special-processors such as GRAPE [16]. For non-periodic systems, multipole expansion methods are more practical, computation by which is proportional to $O(N \ln N)$ (Particle–Particle Cell–

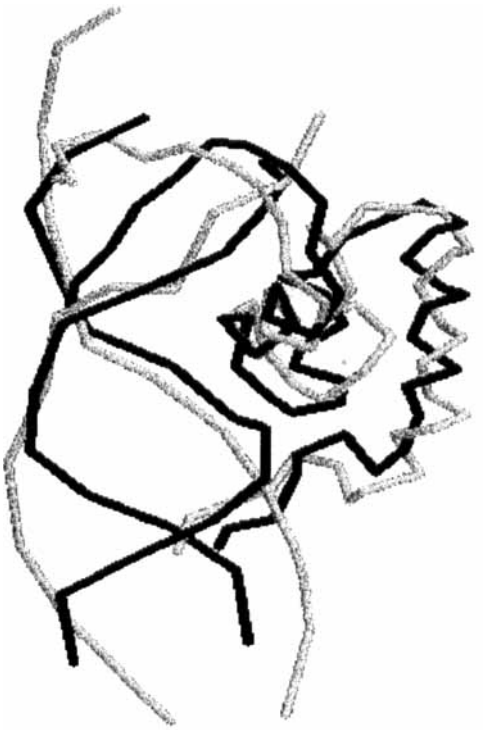

FIGURE 12 Time average of the simulated structure (200–500 ps, gray) superimposed onto the crystal structure (black). The trajectory was obtained by the residue-based nonbonded cutoff scheme (10 Å).

Particle method, PPPC, [11]) or to $O(N)$ (Fast Multipole Method, FMM, [36]). PPPC has been applied to pioneering MD simulations of proteins demonstrating the importance of the electrostatic interaction [11–13]. Schulten's group has been using FMM to perform MD of DNA binding proteins [7, 8, 10]. FMM is also applicable to the periodic boundary. The original Ewald method, which we used in this study, has been recognized as a standard in MD simulation of Coulombic systems in a periodic boundary [20]. The computation time is $O(N^{3/2})$. Particle Mesh Ewald (PME, [37]) is a reformation of the original Ewald method, whose computational cost is $O(N)$. PME has been successfully applied to MD of nucleotides (see, for example, [38] and the references therein).

It is a matter of convenience which algorithm to choose, depending on the size of the simulated system, the boundary condition, the precision needed, the software and hardware environment, and so on. We used the original

Ewald method because it is easy to implement to MD-GRAPE. In the current simulation, the wave-number space summation took only 14% of the total computation time, which suggested that use of PME would not increase the speed significantly. This situation was, of course, due to our special hardware and use of the multi-time step integrator.

D. The Temperature Separation

Finally, we discuss the temperature separation phenomenon in the cutoff MD presented in Appendix C.

As seen in Table IV, the temperature of the ions was 200 K higher than those of the solute and solvent. The solute had 30 K higher temperature than the solvent, but the reported temperature separation phenomena were usually characterized by the hotter solvent [16, 22–24]. We attributed this discrepancy to the difference in the electronic property of the solutes, as explained below.

The temperature separation is a rather complex phenomenon. It also depends on the time-integration and temperature control algorithms [23], not only on the treatment of the nonbonded interactions. However, we would like to give a fairly simple, qualitative explanation. The noise due to the electrostatic cutoff raises the temperature of each molecule, as suggested by the temperature increase in the microcanonical MD, depending on the electrostatic property of the molecule. Therefore, the temperatures of the molecules tend to be: charged > polar > nonpolar. Hence, the ions were as hot as 500 K. Unlike other temperature separation phenomena so far reported, the solutes were highly charged in our simulation ($-25e$ for the DNA and $+9e$ for the peptide). Such molecules should be sensitive to the noise and hence had a higher temperature than the solvent. The solvent molecules were polar, but not charged, and presumably less sensitive to the noise than the solutes, so they had the lowest temperature this time.

The use of a shifting potential was reported to avoid the temperature separation [24]. Though such a potential is a useful expedient, the shifting potential method brought about serious artifacts to the behavior of the solvents and ions [39–41]. Another way to avoid the temperature separation is the use of the separate temperature coupling scheme, such as implemented in AMBER ver. 4.0 [42] or later. Namely, the solute and the solvent are coupled to different temperature baths. This method is useful, for instance, in NMR refinement. Nevertheless, in principle, this method is incapable of preventing the temperature separation 'within' the solute molecule such as reported in [24], and inappropriate when minute dynamics and thermo-

dynamics of the simulated system are concerned. Thus, we conclude that the avoidance of the cutoff is the easiest and most feasible way to prevent the temperature separation.

References

[1] Zieg, J., Silverman, M., Hilmen, M. and Simon, M. (1994). "Recombinational switch for gene expression", *Science*, **196**, 170.

[2] Haykinson, M. J., Johnson, L. M., Soong, J. and Johnson, R. C. (1996). "The Hin dimer interface is critical for Fis-mediated activation of the catalytic steps of site-specific DNA inversion", *Curr. Biol.*, **6**, 163.

[3] Bruist, M. F., Horvath, S. J., Hood, L. E., Steists, T. A. and Simon, M. (1987). "Synthesis of a site specific DNA-binding peptide", *Science*, **235**, 777.

[4] Feng, J.-A., Johnson, R. C. and Dickersoon, R. E. (1994). "Hin recombinase bound to DNA: The origin of specificity in major and minor groove interactions", *Science*, **263**, 348.

[5] Eriksson, M. A. L. and Laaksonen, A. (1992). "A molecular dynamics study of conformational changes and hydration of left-handed $d(CGCGCGCGCGCG)_2$ in a nonsalt solution", *Biopolymers*, **32**, 1035.

[6] Lee, H., Darden, T. and Pedersen, L. (1995). "Accurate crystal molecular dynamics simulations using particle-mesh-Ewald: RNA dinucleotides-ApU and GpC", *Chem. Phys. Lett.*, **243**, 229.

[7] Bishop, T. C. and Schulten, K. (1996). "Molecular dynamics study of Glucocorticoid receptor-DNA binding", *Proteins*, **23**, 115.

[8] Bishop, T. C., Kosztim, D. and Schulten, K. (1997). "How hormone receptor-DNA binding affects nucleosomal DNA. The role of symmetry", *Biophys. J.*, **72**, 2056.

[9] Billeter, M., Guntert, P., Luginbuhl, P. and Wutrich, K. (1996). "Hydration and DNA recognition by Homeodomains", *Cell*, **82**, 1057.

[10] Kosztin, D., Bishop, T. C. and Schulten, K. (1997). "Binding of the Estrogen receptor to DNA. The role of waters", *Biophys. J.*, **73**, 557.

[11] Saito, M. (1992). "Molecular dynamics simulations of proteins in water without the truncation of long-range Coulomb interactions", *Mol. Simul.*, **8**, 321.

[12] Saito, M. (1994). "Molecular dynamics simulations of proteins in solution: artifacts caused by the cutoff approximation", *J. Chem. Phys.*, **101**, 4055.

[13] Saito, M. (1995). "Molecular dynamics/free energy study of a protein in solution with all degree of freedom and long-range Coulomb interactions", *J. Phys. Chem.*, **99**, 17043.

[14] Schreiber, H. and Steinhauser, O. (1992). "Cutoff size does strongly influence molecular dynamics results on solvated polypeptides", *Biochemistry*, **31**, 5856.

[15] Fox, T. and Kollman, P. A. (1996). "The application of different solvation and electrostatic models in molecular dynamics simulations of Ubiquitin: how well is the X-ray structure maintained?", *Proteins*, **25**, 315.

[16] Komeiji, Y., Uebayasi, M., Takata, R., Shimizu, A., Itsukashi, K. and Taiji, M. (1997). "Fast and accurate molecular dynamics simulation of a protein using a special-purpose computer", *J. Comput. Chem.*, **18**, 1546.

[17] Sugita, Y. and Kitao, A. (1998). "Improved protein free energy calculation by more accurate treatment of nonbonded energy: application to Chymotrypsin inhibitor 2, V57A", *Proteins*, **30**, 388.

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

[19] Fukushige, T., Taiji, M., Makino, J., Ebisuzaki, T. and Sugimoto, D. (1996). "A highly parallelized special-purpose computer for many-body simulations with an arbitrary central force: MD-GRAPE", *Astrophys. J.*, **468**.

[20] Allen, M. P. and Tildesley, D. J. (1987). *Computer simulation of liquids*, Oxford Univ. Press, U. K.
[21] Haile, J. M. (1992). *Molecular dynamics simulation–Elementary methods*, John Wiley & Sons, N.Y.
[22] Howard, A. E. and Kollman, P. A. (1992). "Molecular dynamics studies of a DNA-binding protein: I. A comparison of the *trp* repressor and *trp* aporepressor aqueous simulations", *Prot. Sci.*, **1**, 1173.
[23] Guenot, J. and Kollman, P. A. (1993). "Conformational and energetic effect of truncating nonbonded interactions in an aqueous protein dynamics simulation", *J. Comput. Chem.*, **14**, 295.
[24] Oda, K., Miyagawa, H. and Kitamura, K. (1996). "How does the electrostatic force cut-off generate non-uniform temperature distributions in proteins?", *Mol. Simul.*, **15**, 167.
[25] Hunergberger, P. H., Mark, A. E. and van Gunsteren, W. F. (1995). "Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations", *J. Mol. Biol.*, **252**, 492.
[26] Komeiji, Y., Uebayashi, M. and Yamato, I. (1994). "Molecular dynamics simulations of *trp* apo- and holorepressors: domain structure and ligand-protein interaction", *Proteins*, **20**, 248.
[27] Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I. and Dervan, P. D. (1990). "Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin Recombinase", *Biochemistry*, **29**, 6551.
[28] Mack, D. P., Sluka, J. P., Shin, J. A., Griffin, J. H., Simon, M. I. and Dervan, P. B. (1990). "Orientation of the putative recognition helix in the DNA-binding domain of Hin recombinase complexed with the hix site", *Biochemistry*, **29**, 6561.
[29] Dang, L.-X. and Pettitt, B. M. (1987). "Simple intramolecular model potentials for water", *J. Phys. Chem.*, **91**, 3349.
[30] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. and Kollman, P. A. (1995). "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", *J. Am. Chem. Soc.*, **117**, 5179.
[31] Tuckerman, M., Berne, B. J. and Martyna, G. J. (1992). "Reversible multiple time scale molecular dynamics", *J. Chem. Phys.*, **97**, 1990.
[32] Nose, S. (1991). "Constant temperature molecular dynamics methods", *Prog. Theor. Phys. Suppl.*, **103**, 1.
[33] Sayle, R. (1995). RASMOL 2.6, Glaxo Research and Development, U.K.
[34] Fincham, D. (1994). "Optimization of the Ewald sum for large system", *Mol. Simul.*, **13**, 1.
[35] Cheatham, III T. E., Miller, J. L., Fox, T., Darden, T. A. and Kollman, P. A. (1995). "Molecular dynamics simulations on solvated biomolecular systems: The particle mesh Ewald method leads to stable trajectories of DNA, RNA, and Proteins", *J. Am. Chem. Soc.*, **117**, 4193.
[36] Rohklin, V. (1985). "Rapid solution of integral equations of classical potential theory", *J. Comput. Phys.*, **60**, 187.
[37] Darden, T., York, D. and Pedersen, L. (1993). "Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems", *J. Chem. Phys.*, **98**, 10089.
[38] Young, M. A., Ravishanker, G. and Beveridge, D. L. (1997). "A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motion, and solvation", *Biophys. J.*, **73**, 2313.
[39] Smith, P. E. and Pettitt, M. (1991). "Peptides in ionic solutions: a comparison of the Ewald and switching function techniques", *J. Chem. Phys.*, **95**, 8430.
[40] Lau, K. F., Alper, H. E., Thacher, T. S. and Stouch, T. R. (1994). "Effects of switching functions on the behavior of liquid water in molecular dynamics simulations", *J. Phys. Chem.*, **98**, 8785.
[41] Perera, L., Essmann, U. and Berkowitz, M. L. (1995). "Effect of the treatment of long-range forces on the dynamics of ions in aqueous solutions", *J. Chem. Phys.*, **102**, 450.
[42] Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P. and Kollman, P. A. (1991). AMBER ver. 4.0, University of California at San Francisco.
[43] McLachlan, A. D. (1979). "Gene duplications in the structural evolution of Chymotrypsin", *J. Mol. Biol.*, **128**, 49.