Sequence Dependencies of DNA Deformability and Hydration in the Minor Groove

Yoshiteru Yonetani[†] and Hidetoshi Kono[†‡§*]

[†]Computational Biology Group, Quantum Beam Science Directorate, Japan Atomic Energy Agency, Kyoto, Japan; [‡]Quantum Bioinformatics Team, Center for Computational Science and e-Systems, Japan Atomic Energy Agency, Kyoto, Japan; and [§]PRESTO, Japan Science and Technology Agency, Saitama, Japan

ABSTRACT DNA deformability and hydration are both sequence-dependent and are essential in specific DNA sequence recognition by proteins. However, the relationship between the two is not well understood. Here, systematic molecular dynamics simulations of 136 DNA sequences that differ from each other in their central tetramer revealed that sequence dependence of hydration is clearly correlated with that of deformability. We show that this correlation can be illustrated by four typical cases. Most rigid basepair steps are highly likely to form an ordered hydration pattern composed of one water molecule forming a bridge between the bases of distinct strands, but a few exceptions favor another ordered hydration composed of two water molecules forming such a bridge. Steps with medium deformability can display both of these hydration patterns with frequent transition. Highly flexible steps do not have any stable hydration pattern. A detailed picture of this correlation demonstrates that motions of hydration water molecules and DNA bases are tightly coupled with each other at the atomic level. These results contribute to our understanding of the entropic contribution from water molecules in protein or drug binding and could be applied for the purpose of predicting binding sites.

INTRODUCTION

Deformability of DNA depends on the base sequence. This sequence-dependent conformational deformability has been considered to be highly associated with the regulation of gene expression. For example, transcription regulatory proteins that specifically bind to a unique sequence require in some cases the bending of DNA to achieve suitable DNA-protein contacts (e.g., TATA box binding proteins). In this case, a sequence-dependent difference in DNA deformability can be one of the determinants for realizing favorable DNA-protein contacts, and it is thus critical for the formation of DNA-protein complexes, which is known as a major factor of indirect readout. Another situation in which the DNA conformational deformability plays an important role is global nuclear organization. DNA in eukaryotes globally bends to wrap a histone and form a compact chromatin structure. Some DNA sequences are known to strongly favor a nucleosome structure (1). In viruses, DNA bends to be stored in capsid. A large number of studies concerning the formation of such peculiar structures indicate that the physical property of DNA, as well as its sequence, is essential for understanding DNA-related biological events. Valuable information about the sequence dependence of DNA deformability has been reported (2,3). By analyzing many crystal structures of DNA-protein complex, Olson et al. (3) found that different basepair steps have significantly different deformability, e.g., the AT step is very rigid, whereas the TA step is very flexible. Similar trials of the deformability characterization have been made using computational approaches (4–10). The sequence dependence of DNA deformability has been investigated thoroughly (4,8,9) from molecular dynamics (MD) simulations of 136 unique DNA sequences that consider all sequences allowed for a tetramer. The computationally deduced data not only made up for the shortage of sequence variation in experimental data but also provided more detailed information on basepair step deformability. In addition to the characters of 10 unique dimer steps, the influence by their neighboring bases was clarified (3,7). Such knowledge about the deformability has been applied to the prediction of DNA-drug interactions (11).

In addition to DNA deformability, the hydration pattern appearing on the DNA surface is also dependent on the sequence. It is known that a well-defined hydration pattern of the DNA minor groove, the hydration spine, appears in particular sequences (12–14). The hydration spine, which was first found in the crystal structure of B-DNA composed of 5′CGCGAATTCGCG3′, determined by Drew and Dickerson (12), consists of two layers of ordered water molecules aligned along the DNA minor groove. Later, higher-resolution analyses showed a hydration structure that was extended up to four layers (15,16). However, analyses using different sequences showed that the DNA hydration pattern was sensitive to sequence, and a different hydration pattern was observed for a sequence with a wider minor groove (17–19).

Such a sequence-dependent difference in DNA hydration can have a nonnegligible effect on biological processes such as DNA-protein interactions. When a protein binds to DNA, hydration water molecules are excluded from the DNA surface. Because the release of water molecules leads to an increase in entropy, it will be a driving force for association with other molecules (20). If water molecules to be excluded

Submitted January 7, 2009, and accepted for publication May 29, 2009.

*Correspondence: kono.hidetoshi@jaea.go.jp

Editor: David P. Millar.

0006-3495/09/08/1138/10 $2.00

doi: 10.1016/j.bpj.2009.05.049

upon protein/ligand binding are highly structured on the DNA surface, a large entropy gain will be obtained by release of these water molecules. This can make an essential contribution to binding affinity and specificity. A series of thermodynamic experiments by Privalov et al. (21,22) on DNA-protein association clearly showed this entropic effect caused by releasing hydration water molecules.

As mentioned above, both deformability and hydration of DNA are sequence-dependent, which raises the question of how they are related to each other. The purpose of this work is to reveal this relationship, and to further understand its underlying mechanism. Crystallographic data of DNA hydration water have continued to accumulate since the first report by Drew and Dickerson (12). Also, with increased resolution of experimental measurements, accurate description of DNA hydration patterns has become possible. However, because there is limited variability in DNA sequence in the structure data reported in the literature, presumably due to the difficulty of crystallization, these data are not adequate to describe the details of sequence-dependent hydration. Unlike the experimental approach, MD simulation is not subject to this crystallization difficulty and can thus be used to investigate the sequence effect systematically. In addition, not only time-averaged structures of DNA and hydration water but also their dynamics can be derived from MD trajectories. In this way, MD simulation provides an atomic-level picture of the deformability-hydration relationship that helps us to understand the mechanical linkage of both behaviors.

In this work, sequence dependence of DNA deformability and hydration were extensively investigated by MD simulation. Different basepair steps exhibited different behavior not only in deformability but also in their minor-groove hydration patterns. By comparing the two properties, it was found that an explicit correlation exists between the sequence dependences of deformability and minor groove hydration. All sequence behaviors considered here were well demonstrated by four representative cases. The obtained correlation between deformability and hydration was also characterized by atomic-level motions of basepair steps and hydration water, which are tightly coupled with each other. Meanings of the obtained knowledge about the sequence dependence of DNA hydration are discussed in the context of protein-DNA interactions.

METHODS

DNA deformability and hydration were evaluated for 10 unique dimeric steps: AT, TA, AC, CA, CG, AA, GA, GC, GG, and AG. These steps were extracted from the central dimeric step N_2N_3 of the B-form DNA whose sequence is $5'CGCGN_1N_2N_3N_4CGCG3'$, which was subjected to MD simulations. N_x denotes any base A, T, G, or C, so the total number of DNA was 136. Because the DNA sequences differ from each other not only in the central dimeric step but also in their flanking bases, the dependency on flanking base of central dimeric step properties was evaluated.

Each DNA structure was constructed in a canonical B-form using the AMBER Nucgen module (23). MD simulations were carried out by the AMBER Sander module (23); 10 ns trajectories were obtained for each DNA system by numerical integration with a time step of 1 fs. The pressure and temperature were set to 1 atm and 300 K, respectively. Trajectories of the last 8 ns of simulations were used for the analysis of DNA deformability and hydration. Details of the system and the simulations are given in Section S1 in Supporting Material.

DNA deformability expressed as basepair step fluctuations

Deformability of basepair steps was quantified by their fluctuation. Using the MD trajectories, the deformability was evaluated for each dimeric step N_2N_3 (Fig. 1; see also Fig. S1 in the Supporting Material). In this evaluation, we used basepair step parameters that described the relative configurations between two basepairs, N_2 and N_3. The parameters were represented by the six variables θ_i ($i = 1, 2, \ldots, 6$) of shift, slide, and rise for translational displacements and tilt, roll, and twist for rotational displacements (24,25). Then, the quantity V_{step}, representing the degrees of fluctuation in the six-variable space, was calculated. The V_{step} is given as

$$V_{\text{step}} = \prod_{i=1}^{6} \sqrt{\lambda_i}, \tag{1}$$

where λ_i is the eigen value for the covariance matrix $\mathbf{M}$ having the components $m_{ij} = \langle(\theta_i - \langle\theta_i\rangle)(\theta_j - \langle\theta_j\rangle)\rangle$. Since $\mathbf{M}$ is positive definite, it can be diagonalized using an orthonormal transformation matrix $\mathbf{R}$ for $\mathbf{M}$,

$$\mathbf{R}^t\mathbf{M}\mathbf{R} = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_6). \tag{2}$$

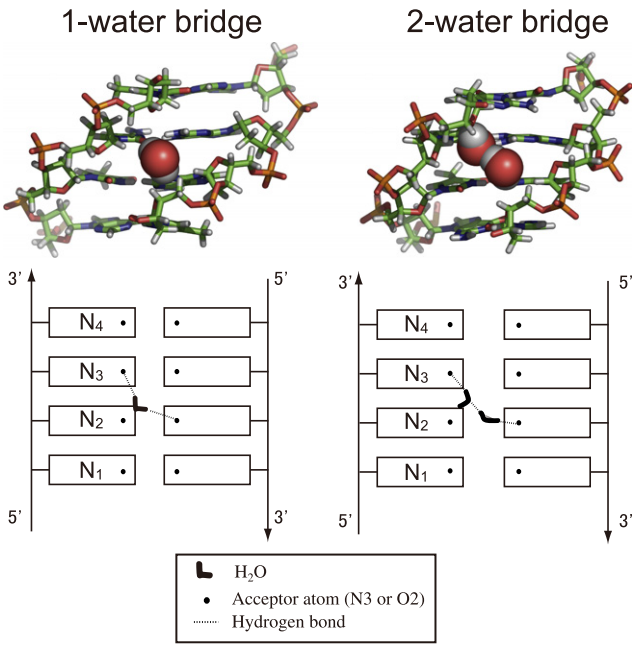

FIGURE 1 Typical hydration patterns observed in the minor grooves, shown by snapshots taken from the simulations in this study (*upper*) and their schematic illustration (*lower*). In the one-water bridge, one water molecule is positioned between two bases that are diagonally opposite each other, and each hydrogen atom of the water molecule makes a hydrogen bond with the corresponding acceptor atom (N3 or O2). This water molecule participates in the formation of the first layer of a hydration pattern known as spine hydration (12–14). In the case of the two-water bridge, two water molecules connected by a hydrogen bond participate in the bridge construction. The remaining hydrogen atom of each water molecule bonds to the corresponding acceptor atom of the bases.

If V_{step} is large, it indicates that the basepair step is highly deformable, and vice versa. V_{step} is a useful quantity for characterizing DNA deformability, as shown in previous studies (3,9).

Evaluation of hydration in DNA minor grooves

Hydration patterns in the minor groove of the central dimeric step N_2N_3 were analyzed. Both experimental (19,26) and computational (17,27) hydration data have suggested that two distinct patterns of hydration can exist in the minor grooves. We also confirmed in a previous study on AATT and TTAA hydration (28) that the two hydration patterns indeed appear in the MD trajectories. The atomic configurations of these two patterns shown in Fig. 1 are considered from a stereochemical point of view. We refer to such hydration connecting two distinct chains of DNA as one-water and two-water bridges. Any kind of base has sites of negative charge in the minor groove—N3 atoms of A or G or O2 atoms of T or C—that can serve as acceptors for hydrogen atoms of a water molecule. One- and two-water bridges connect two distinct DNA chains by using one or two water molecules as mediators. As shown later, deformability was highly correlated with the hydration described by these two types of bridge formation. We evaluated bridge formation for every snapshot of MD trajectories to obtain the probabilities of one- or two-water bridge formations, P_1 and P_2. We assumed that the water bridge is formed when all associated hydrogen bonds (i.e., two for one-water and three for two-water bridges (Fig. 1)) are formed. For hydrogen bonds, the same criterion was used as in Auffinger and Westhof (29): the distance of A···H is <2.5 Å and the angle of A···H-O is >135°, where A denotes the acceptor atom.

RESULTS

Before showing the calculated results of deformability and hydration, we should mention that because bases A, T, G, and C differ in their geometry, the positions of the acceptor atoms (N3 or O2) are not exactly identical. This geometric effect can cause different hydration patterns to be made of different bases. If this is the case, it will become difficult to purely extract the contribution from basepair fluctuations, and we cannot then discuss the relationship between conformational deformability and hydration. It was, however, found that this geometrical difference of acceptor sites is subtle. The geometric comparison demonstrates that the distance between two acceptor atoms differed only slightly among base types, settling in a narrow range of 4.40–4.71 Å, as shown in Fig. S2, where the canonical B-form structures of various basepair steps are compared. Accordingly, assuming that the geometric effect is approximately the same among different basepair steps, we focused on the analysis of the sequence-dependent relation between DNA hydration and deformability.

Relationship between DNA basepair step deformability and water-bridge formations

In Fig. 2, we show the calculated probabilities of one- and two-water bridge formation, P_1 and P_2, respectively, plotted together with the deformability V_{step}. Each point corresponds to a central dimeric step of 136 DNA sequences, and thus 136 points are seen in the figure. These points are colored according to the value of basepair step deformability, V_{step}. Flexible (*red*) and rigid (*blue*) basepair steps are distributed

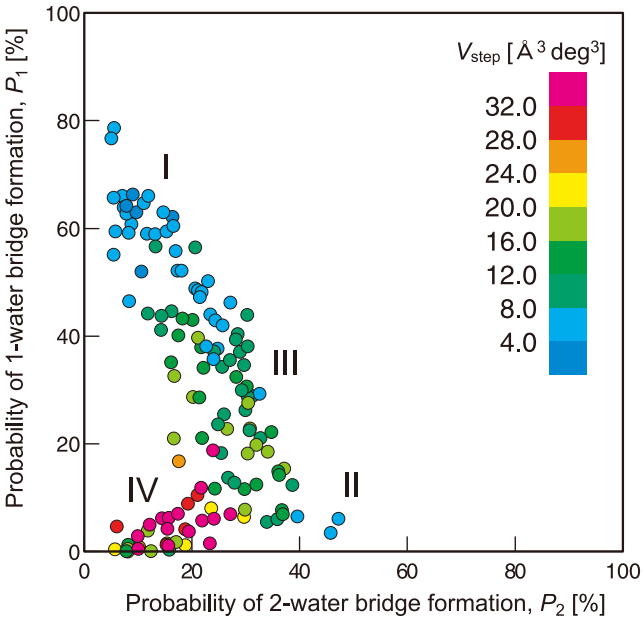

FIGURE 2 Relation between DNA deformability and minor-groove hydration. The probabilities of one- and two-water bridge formation and the basepair step deformability, V_{step}, for central dimeric steps of 136 DNA sequences are shown. Labels I–IV are employed to denote the different types of behavior. (See also Section S3 in the Supporting Material for values of P_1, P_2, and V_{step}.)

in distinct regions of the P_1-P_2 plot, indicating that a clear correlation exists between deformability and water-bridge formation. To explain this correlation, we here characterize the distribution based on the four regional types with different behaviors: Most of the rigid basepair steps exist in the region with high P_1 (~60%) and low P_2 (~10%) (type I), but a few rigid basepair steps appear in the opposite side, i.e., low P_1 (~10%) and high P_2 (~50%) (type II). Basepair steps with medium deformability are distributed in the intermediate region with P_1 of ~30% and P_2 of ~30% (type III). Very flexible basepair steps are found only in the region where both P_1 and P_2 have low values, ~10% for P_1 and ~20% for P_2 (type IV). This tendency can be summarized as follows: rigid steps are likely to form either one- or two-water bridges, and these are principal hydration patterns for minor grooves of rigid DNA. On the other hand, flexible steps do not have any bridge.

It may be intriguing to elucidate the causality of both properties. One scenario is that water bridge formation governs the basepair step motion. To form a water bridge, acceptor atoms of bases must be precisely positioned, which may be allowed for some bases and not for others. Nonetheless, if any stable bridges are formed, the basepair step will be rigidified. Another scenario is that the basepair step deformability governs water-bridge formation. Different bases have different interactions with neighboring bases. If a basepair step is stably positioned by such interactions, it becomes rigid, inducing an ordering of water molecules. Here, which hydration pattern—a one- or two-water bridge—appears

depends on the positioning of the associated base-acceptor sites. We will return to these hypotheses in the Discussion.

To show the differences among basepair steps, all points of the P_1-P_2 plot were divided into 10 cases of different dimeric steps (Fig. 3). The points in each panel show the results for a particular dimeric step, but with different flanking bases. Thus, the distribution of the points shows the influence of the flanking bases. Fig. 3 clearly shows that different basepair steps have different distribution patterns. The AT step is very rigid irrespective of the flanking bases, and the points are localized in the region with high P_1 and low P_2, showing that a one-water bridge stably forms. On the other hand, all pyrimidine-prine steps (TA, CA, and CG steps) are concentrated in the region with low P_1 and low P_2, indicating that they do not have any stable hydration patterns and are flexible. The other steps—GA, GG, AC, AA, GC, and GA—are significantly influenced by the flanking bases. For example, the GA step has the potential to make either one- or two-water bridges (Fig. 3), but we found that which of these hydration patterns is formed depends on the flanking bases: the observed hydration pattern for AGAT is mainly the one-water bridge, whereas CGAC predominantly forms the two-water bridge.

Atomic-level observation of the motion of basepair steps and hydration water

To obtain a clear picture of the deformability-hydration correlation shown in Fig. 2, we observed time-dependent motions of DNA basepair steps and bridge-forming water molecules at the atomic levels (see Fig. S3). The GA ($V_{step} = 6.5$ Å^3 deg^3) and TA ($V_{step} = 47.5$ Å^3 deg^3), which correspond to the central steps of AGAT and ATAA samples, respectively, were chosen as representatives of the rigid and flexible cases. In the rigid GA step, the six variables settle around their equilibrium values. In contrast, the flexible TA step exhibits larger fluctuations. In this case, remarkable changes are observed in the variables shift, slide and twist.

We next checked the behavior of water molecules to clarify its relationship to the above basepair step motion. To monitor the motion of basepair steps, we employed, instead of the six variables defined above, the distance between two acceptor atoms, $d_{A\text{-}A}$, because we noticed that this distance is closely related to the dynamics of bridge water molecules, and it is useful for capturing the correlated motions.

The acceptor-acceptor distance, $d_{A\text{-}A}$, and the probabilities of water-bridge formation are shown in Fig. 4, where four typical cases of types I–IV are shown. The presented probabilities are evaluated for each 100-ps interval. Note that these probabilities are defined differently from those in Figs. 2 and 3, where probabilities P_1 and P_2 were evaluated using 8-ns trajectories. In the basepair step of type I, the distance $d_{A\text{-}A}$ is kept at ~3.8 Å throughout the simulation, and then a one-water bridge almost always forms. The step of type II is also stable, but shows a slightly larger value ($d_{A\text{-}A} \sim 4.4$ Å). In this case, a two-water bridge is formed. Compared to these two cases, the step of type III shows a large fluctuation. The distance $d_{A\text{-}A}$ takes either of ~3.8 or ~4.4 Å, and often transits between the two values. A water bridge suited to the distance (i.e., a one-water bridge for $d_{A\text{-}A}$ ~3.8 Å or a two-water bridge for $d_{A\text{-}A}$ ~4.4 Å) is formed, which reveals that both motions of basepair steps and water molecules are highly correlated. The very flexible basepair step of type IV has a significant fluctuation of $d_{A\text{-}A}$, and in this case, the hydration pattern also fluctuates significantly. The one- or two-water bridge is rarely formed, but another complicated hydration pattern appeared in which three or four water molecules participate in the acceptor-acceptor connection (see snapshot in Fig. 4). As demonstrated here, the dynamics of basepair steps and water molecules are linked through the positions of acceptor atoms. Fig. 2 clearly shows this dynamic correlation. It should be noted that both water-bridge formation and basepair step motion are related to the minor groove width. Typically, a one-water bridge appears when the groove width is narrow, as reported previously (17,19,30). We also examined the minor groove dynamics and their relation to the hydration water and basepair step deformability. They are discussed in Sections S4 and S5 in the Supporting Material.

Correspondence between simulation results and experimental hydration data

We compared the current MD results with the hydration data obtained by x-ray crystal structure analyses. With regard

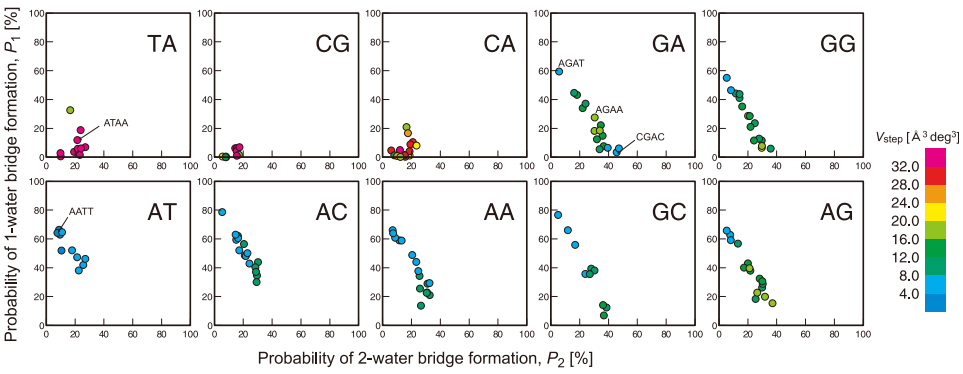

FIGURE 3 Sequence dependence of DNA deformability and minor-groove hydration. All points from Fig. 2 are decomposed into 10 groups according to their central dimeric step. Indices are given for the sequences cited in the text.

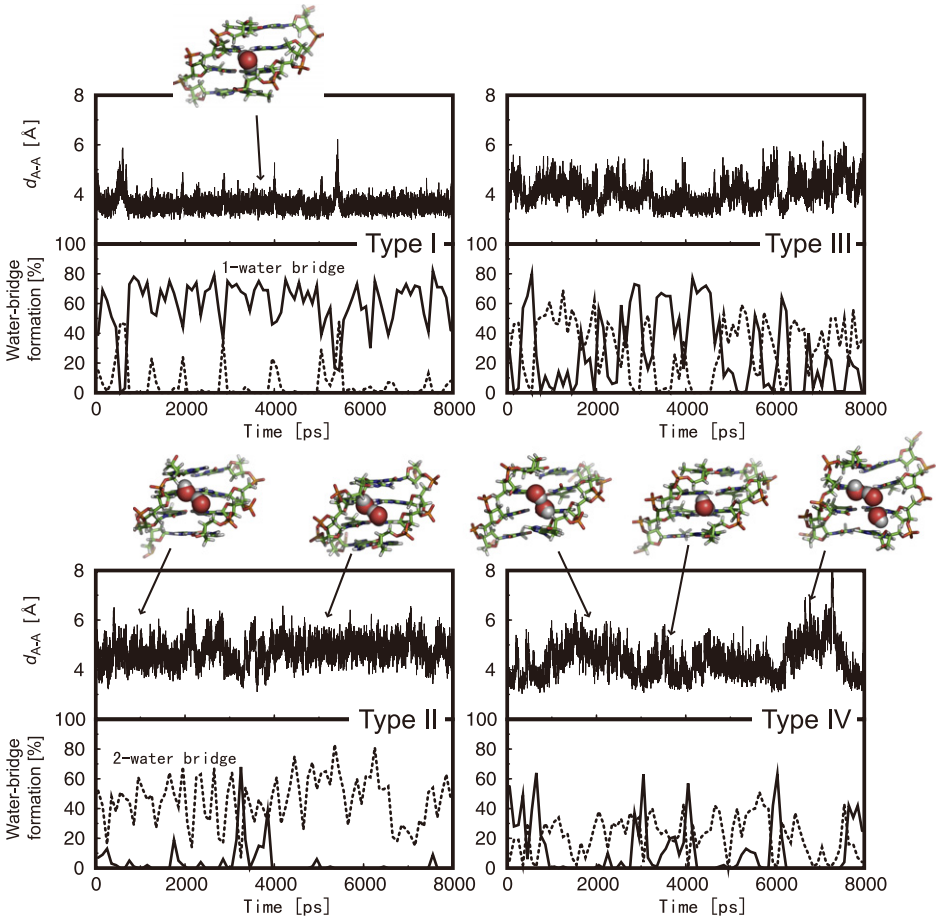

FIGURE 4 Differences in the dynamics of DNA and hydration water among sequences: type I, GA step of AGAT; type II, GA step of CGAC; type III, GA step of AGAA; and type IV, TA step of ATAA. The acceptor-acceptor distance, $d_{A\text{-}A}$, shown in the upper portion of each panel indicates basepair step motions. Probabilities of one- and two-water bridge formation are indicated by solid and dotted lines, respectively.

to sequence dependence of basepair step deformability, a good correspondence between MD and experimental results was confirmed previously (9). We here check the result of hydration patterns. Table 1 shows the sequence preferences of one- and two-water bridges obtained from crystal structures and our simulations. The 21 crystal structures of B-form DNA with resolution $\leq$2.5 Å were used (for details of the procedure, see Table 1 note). We should mention the differences in physical meaning between these statistics. The experimental results are obtained by averaging over the various crystal samples, whereas the MD results involve not only the sample average (i.e., over the same basepair step with different flanking bases) but also the average over the time-dependent variation. Though some uncertainties are involved due to these differences in physical meaning, we were able to summarize the several correspondences in four points.

1. In both MD and experiment, the one-water bridge is more likely to appear than the two-water bridge.
2. In the experimental results, AT, AC, and AA steps are very likely to form the one-water bridge (see, for example, the AT step shown in Fig. 5 *a*), which is consistent with the MD results.
3. The two-water bridge is indeed observed experimentally, but it is difficult to experimentally characterize the sequence preference of the two-water bridge because of the infrequent appearance of this hydration pattern and the resultant shortage of data. The GA step seems to be most favorable for two-water bridge formation, but it can also make a one-water bridge. In crystal structures of the GA step, water molecules are found at either site of the one- or two-water bridges (Fig. 5, *c* and *d*). That is, the GA step can adopt a one- or two-water bridge formation, and its hydration pattern is dependent on the flanking bases. As confirmed in Fig. 3, the same tendency is found in our calculations.
4. Finally, the hydration of very flexible basepair steps (TA, CA, and CG of type IV) rarely has a well-defined pattern in our MD simulations. Consistent with the calculation results, no hydration patterns for these steps are found in the crystal structures (Table 1 and Fig. 5 *b*).

DISCUSSION

We have clarified the relationship between two sequence-dependent properties, DNA deformability and hydration. Our results confirm that they are strongly correlated with each other: rigid basepair steps favor ordered states of hydration water, whereas flexible steps do not. Such

TABLE 1 Formation of one-water and two-water bridges in the minor groove

	MD*		Experiment†		
	P_1 (%)	P_2 (%)	One-water bridge (%)	Two-water bridge (%)	N‡
TA	9.1	19.8	33	0	6
CA	5.6	15.2	0	18	11
CG	2.8	12.5	0	0	12
GA	22.6	30.4	15	15	13
GG	26.5	21.0	0	0	6
AT	53.6	16.3	41	0	17
AC	50.5	21.4	29	0	7
AA	39.8	22.0	51	3	35
GC	38.1	25.5	0	0	16
AG	37.5	22.2	60	0	5

*Probabilities of formation of one- and two-water bridges, P_1 and P_2. Values were averaged over DNA samples with the same basepair step at the center of the sequence.

†Experimental data were obtained from the Nucleic Acids Database (May 8, 2008, release) (47). Of 3826 deposited nucleic acid structures, 28 B-DNA structures were selected, which were solved at $\leq$2.5 Å resolution and without any mismatch or modification. From those 28 structures, seven sequence-redundant structures were excluded, leaving 21 structures to be analyzed. Furthermore, to remove the effect of crystal packing, the basepair steps having a close contact with neighboring DNA in the crystal states were not considered, leaving 128 basepair steps in the final analysis (see Table S3 in Section S6 in the Supporting Material). Hydration patterns (i.e., one- or two-water bridges, or no bridge) of the 128 basepair steps were checked with the distance measurement of PyMOL (48). Because of the lack of hydrogen positions in the crystal data, hydrogen bonds were identified using 3.4 Å as the oxygen-oxygen or oxygen-nitrogen distance.

‡Number of basepair steps used for the experimental averages.

sequence-dependent hydration is important in the context of biological functions of DNA. A discussion follows of the effect of hydration upon protein or drug binding.

Entropic effect of water molecules released upon protein or drug binding

Water molecules structurally ordered on the DNA surface can affect the affinities in DNA-protein associations, because their release from the DNA surface contributes to the binding entropy (20,31). In fact, it has been shown experimentally that protein binding in the minor groove is accompanied by a large increase of entropy (21,22), suggesting that the driving force of the binding comes from the release of the water molecules. The amount of entropy gain can be determined by the degree of ordering of water molecules bound on the DNA surface, and hydration is thus thought to be an important factor for binding. As an example, we cite the binding of AT-hook (32), which specifically binds on the minor groove surface of the AATTT sequence (Fig. S5). This binding is known to be an entropically driven process (33), which leads us to expect that the hydration water of the minor groove region makes an essential contribution to the binding. We evaluated the hydration of this AATTT region using the MD data of hydration presented in this article (Fig. 3), and found that the water molecules are indeed highly ordered (one- and two-water bridge-forming probabilities (P_1, P_2) are (64,7), (64,8), (66,7), and (23,31) for the four sites of the sequence, respectively. Therefore, the release of the ordered water is thought to be a possible cause of the large entropy gain.

We next consider the binding of a small DNA-binding drug, Hoechst 33258 (34) (Fig. 6 *a*, *right*). This molecule binds to DNA from the minor groove in a manner similar to that of AT-hook, described above, which is driven entropically (35,36) (see more details in Section S8 of the Supporting Material). This is a good example of the effect of sequence-dependent hydration, because the binding affinities for the various DNA sequences have been systematically measured (37,38). The Hoechst 33258 can bind to any A- and T-rich sequence, but the affinity varies substantially with the base arrangement. According to the fluorescence titrations by Breusegem et al. (37), the sequence preference is in the order of AATT (−11.70) > TAAT (−9.90) ≈ ATAT (−9.91) > TATA (−8.61) ≈ TTAA (−8.71), where the values in parentheses indicate the binding free energy (kcal/mol). The MD data presented here regarding the probabilities of water-bridge formation ($\overline{P_1}, \overline{P_2}$) show that the degrees of ordering of minor groove water should be, from order to disorder, AATT (64,7) > ATAT (46,18) > TAAT (37,19) > TATA (25,24) > TTAA (10,21). $\overline{P_1}$ and $\overline{P_2}$ denote the average over the three bridge formation sites involved in the tetramers (Fig. 6 *a*, *left*): $\overline{P_i} = \sum_{\text{site}=1\sim3} P_i^{\text{site}}$. Both properties of binding affinity and hydration are well correlated with each other. This will be understood in the following way: the higher the ordering of the water bound on the DNA surface, the larger the resultant entropy gain from the release of the water. Accordingly, a sequence with highly ordered water (e.g., AATT) has a high affinity with the Hoechst molecule. In this way, the simulation results provide quite a reasonable explanation for Hoechst binding.

In the DNA-Hoechst binding, entropy gain is of course the main contribution to the binding free energy (35,36). It should be noted, however, that the binding free energy is determined by the balance of entropy and enthalpy, so keeping enthalpy cost low is also important for realizing favorable binding (see Section S8 of the Supporting Material). In the DNA-Hoechst complex structure where the Hoechst is binding to the minor groove of AATT (34), the Hoechst suitably forms hydrogen bonds with the hydrophilic atoms that reside at the floor of the minor groove. In addition, the Hoechst structure just fits into the narrow minor groove. These favorable conditions are thought to lower the cost of binding enthalpy and these contacts are probably established due to the rigidity of the AATT conformation and the narrowness of the groove. In contrast, the minor groove of TTAA may be unfavorable for realizing such hydrogen bonds and a tightly packed form, because the basepair step fluctuation is large and the groove structure is fluctuating and wider (28). Differences in the DNA-Hoechst binding affinity among DNA sequences (37) are partly attributed to such different contributions in terms of enthalpy.

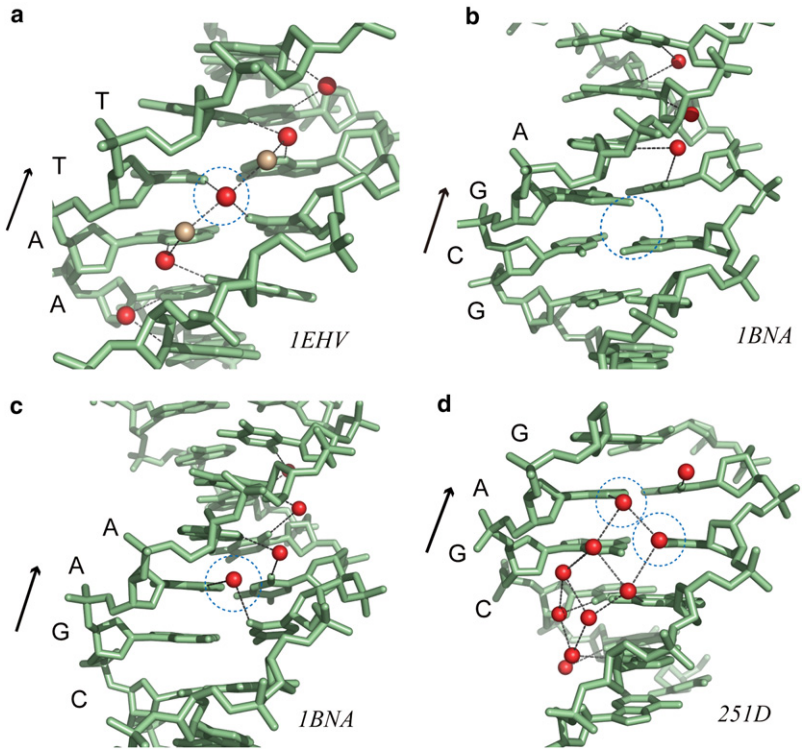

FIGURE 5 Hydration patterns in the crystal structures of DNA. Experimentally deduced water oxygen sites are denoted by red and white spheres, which correspond to the first and second hydration layers, respectively. Blue dotted circles indicate the regions of interest. (*a*) One-water bridge in the AT step from AATT (PDB code 1EHV). (*b*) No bridge in the CG step from GCGA (PDB code 1BNA). (*c*) One-water bridge in the GA step from CGAA (PDB code 1BNA). (*d*) Two-water bridge in the GA step from CGAG (PDB code 251D).

It should be noted that for a major-groove binding protein, Watkins et al. recently reported that the ordering of water molecules around DNA is correlated with the binding affinity (39). One interpretation of this result could be that water ordering at the minor groove region that does not make contact with this protein in the DNA-protein complex is correlated with the binding affinity. This might be another factor to consider in the contribution of water molecules to the binding free energy.

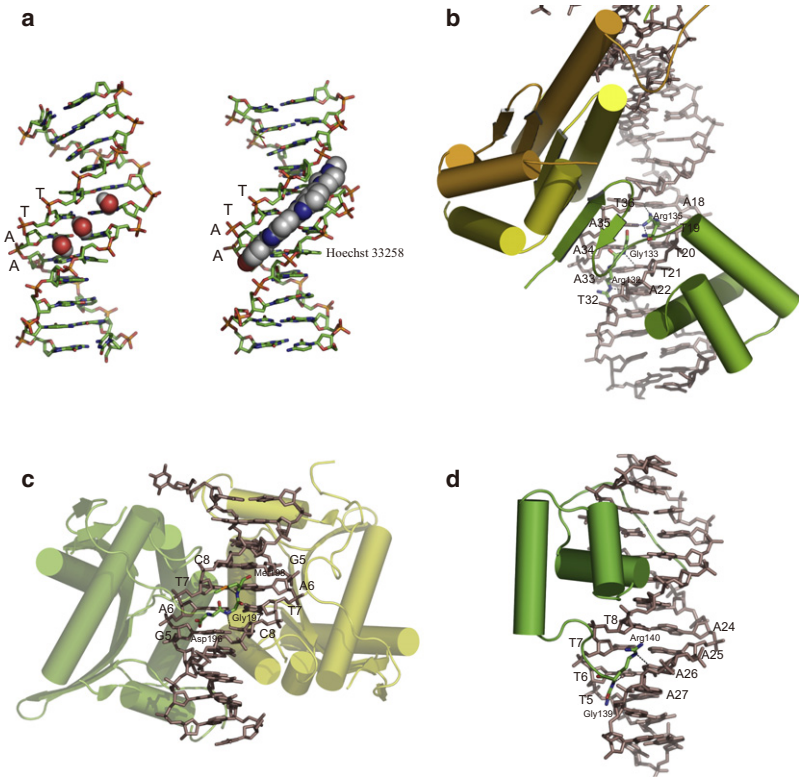

FIGURE 6 Interactions in the minor groove. (*a*) Interactions with a small molecule, showing hydration in the minor groove of the AATT sequence in this study (*left*) and the crystal structure with Hoechst 33258 (34) (PDB code 1D43) (*right*). (*b–d*) Interactions with three proteins. (*b*) MATα2/MCM (PDB code 1MNM). Six hydrogen bonds are formed in the minor groove between TAAAT and residues: (Arg^{135})N-H1$\cdots$O2(T^{36}), (Arg^{135})N-H1$\cdots$O2(T^{19}), (Arg^{135})N-H2$\cdots$N3(A^{35}), (Arg^{135})N-H2$\cdots$O2(T^{20}), (Gly^{133})N-H$\cdots$O2(T^{21}), and (Arg^{132})N-H$\cdots$O2(T^{22}). (*c*) Endonuclease *Bam*HI (PDB code 1BHM). Two hydrogen bonds are formed in the minor groove between GATC and residues: (Gly^{197})N-H$\cdots$O2(T^{7}) and (Asp^{196})N-H$\cdots$O2(C^{8}). (*d*) Hin recombinase (PDB code 1HCR). Two hydrogen bonds are formed in the minor groove between TTTT and residues: (Arg^{140})N-H$\cdots$N3(A^{26}) and (Arg^{140})N-H$\cdots$O2(T^{6}). The MD results from this study (Fig. 3) suggest that the minor grooves shown have strongly ordered hydration in the protein-free state (see text).

Hydration-water-based prediction of DNA-protein interaction sites

Knowledge about DNA hydration can be used to predict protein-DNA interaction sites. Analysis of the crystal structures of various DNA-protein complexes has revealed the following relation: positions where protein residues make hydrogen bonds with DNA in the complex state are in good correspondence with the hydration sites of the DNA in the free state (40). Based on this rule, the data of hydration presented here (Fig. 3) can be applied to predicting the hydrogen-bonding sites of DNA-protein complexes. According to Fig. 3, AT and AC steps are likely to make stable hydrogen bonds to the water. This means that these steps have the potential to serve as a hydrogen-bonding site when proteins bind to the DNA. On the other hand, TA, CA, and CG steps are not expected to behave in such a way, because they are not likely to make any water bridges in the free state.

Among various DNA-protein complexes listed by Selvaraj et al. (41), we found cases in which hydration sites of the DNA minor groove are replaced by the hydrogen-bonding sites formed upon protein binding. MATα2/MCM has a contact with the DNA minor groove of TAAAT (42). The hydration data obtained by this study indicate that this minor groove region is indeed well hydrated (probabilities of one- and two-water bridges forming (P_1, P_2) are (6,24), (26,26), (66,7) and (64,8) for the four sites). The water molecules at these sites are displaced upon protein binding, and the N-terminal residues of the protein, Arg^{132}, Gly^{133}, and Arg^{135}, then intrude into the minor groove (Fig. 6 *b*). In this process, eight hydrogen bonds involved in four one-water bridges are lost, and six hydrogen bonds are created by binding of the protein residues (Fig. 6 *b* and legend). The crystal structure of endonuclease *Bam*HI (43) with DNA is also considered to have hydrated water molecules replaced by protein in the same way. Three residues of this C-terminal region bind to the GATC minor groove (Fig. 6 *c*). From the hydration data in Fig. 3, this region must be highly hydrated, and in fact, the probabilities of forming one- and two-water bridges (P_1, P_2) are (45,16), (65,11), and (45,16) for the three sites. Upon binding of *Bam*HI, the hydration water molecules are replaced by the residues in the C-terminal region, which form two hydrogen bonds with DNA (Fig. 6 *c* and legend).

DNA and proteins make a rather smaller number of contacts in the minor groove than in the major groove, as seen in the examples above. However, in some cases, such a small number of contacts can be essential for DNA-protein complex formation. The base (44) or amino acid (45) mutation experiments for Hin recombinase, which specifically binds to the minor groove of TTTT using its N-terminal region (Fig. 6 *d* and (46)), showed drastic changes in binding affinity, and the interactions in the minor groove were identified as an essential contribution to this binding. Again, this contribution can be predicted by the hydration data presented here (Fig. 3): probabilities of water-bridge formations for the TTTT region are (P_1,P_2) = (59,12), (59,12), and (21,33), showing that this minor groove is a highly hydrated site.

Mechanical linkage between the motions of DNA basepair steps and hydration water

We have shown that an explicit correlation exists between the behavior of basepair steps and that of hydration water. Furthermore, we have obtained a detailed picture of this correlation, i.e., the motions of basepair steps and water molecules were tightly coupled at the atomic level. A question yet to be resolved is the causality between the two behaviors. We now consider the question of what produces the deformability-hydration correlation. Some say that DNA basepair step motions determine the bridge-water dynamics, and others that the water-bridge formation determines the basepair step motion. To address which hypothesis is more likely, we conducted an additional MD simulation using a DNA sequence containing a tetramer of AATT at the center. The central dimeric step AT was originally rigid and was suited to the formation of a one-water bridge (Fig. 3). We artificially modified the charges of the acceptor atoms, changing the original atomic charge of the two acceptors (O2 of T) from −0.588e to −0.088e, so that no water bridge would be formed. If the AT step were to become flexible by inducing collapse of the water bridge, we could conclude that the rigidity was a result of the water-bridge formation.

The MD result showed that the water bridge was completely broken by modifying the acceptor sites (the probability of one-water bridge formation, P_1, decreased from 64% to 2%). On the other hand, the deformability of this basepair step did not change fundamentally (the resultant V_{step} was 2.0 $Å^3\,deg^3$, which was approximately the same as the original value of 2.7 $Å^3\,deg^3$). It can be interpreted from this result that DNA deformability is not affected by hydrogen-bonding of hydration water. However, it is difficult to simply say so, because we observed a different kind of water molecule localization in the minor groove in the trajectory (see Section S9 of the Supporting Material). Furthermore, there are other possibilities to be considered. Though breaking one water bridge did not change the basepair step deformability, multiple breaks may have a large impact on deformability. Another possibility is the effect of the flanking basepair steps, which may be related to the deformability of the central step. In the case of the AATT sequence, not only the central AT step but also the flanking steps, AA and TT, are both rigid and suitable to form the one-water bridge. Such a cooperative effect of the flanking basepair step with hydration remains to be systematically examined.

Higher layers of ordered water molecules

High-resolution crystal structure analyses revealed that a third and a fourth layer of ordered water molecules exist (15,16). In this study, we did not observe such higher layers, probably because our MD simulations were performed at

300 K, whereas the structure determinations were carried out at a much lower temperature (120 K and 163 K). The low temperature suppresses the dynamics of water molecules and enhances the organization of higher layers of ordered water molecules. In fact, MD simulations at low temperatures showed the higher layers (see Section S10 of the Supporting Material).

CONCLUSION

The sequence dependences of basepair step deformability and minor-groove hydration patterns were obtained from MD simulations for all 136 samples of DNA. It was found that the two sequence-dependent behaviors are directly related to each other. This was explained by picking up four typical cases. Most of the rigid basepair steps are very likely to form a one-water bridge, but there are a few exceptions that favor another hydration pattern, the two-water bridge. Basepair steps with medium deformability can make either hydration pattern, but frequently transit between the two states. Very flexible basepair steps do not have any stable hydration patterns.

A detailed picture of this deformability-hydration correlation is presented. It is shown that the correlation comes from coupling of the motions between base acceptors and hydration water molecules. The causality of the deformability-hydration correlation remains to be resolved, but it is expected to be clarified by another simulation with an implicit solvent model.

We have discussed two scenarios in which the sequence-dependent difference of hydration is thought to be an essential component. One is the entropic gain from water molecules released upon protein or drug binding. Another is the practical use for predicting critical interactions in DNA-protein recognition. The knowledge about sequence-dependent hydration provided by this study will be helpful for understanding protein-DNA interactions at an atomic level.

SUPPORTING MATERIAL

Additional text, figures, tables, and references are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(09)01107-2.

We thank Prof. Akinori Sarai and Dr. Satoshi Fujii for their help at the initial stage of this work and Prof. Nobuhiro Go for providing many invaluable comments.

This work was supported by a Grant-in-Aid for Scientific Research to H.K. (18031042) from the Ministry of Education, Culture, Sports, Science and Technology in Japan. Y.Y. was supported by a Japan Society for the Promotion of Science Fellowship for Young Scientists.

REFERENCES

1. Widom, J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* 34:269–324.
2. Okonogi, T. M., S. C. Alley, A. W. Reese, P. B. Hopkins, and B. H. Robinson. 2000. Sequence-dependent dynamics in duplex DNA. *Biophys. J.* 78:2560–2571.
3. Olson, W. K., A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA.* 95: 11163–11168.
4. Beveridge, D. L., G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, III, et al. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(C_pG) steps. *Biophys. J.* 87:3799–3813.
5. Flatters, D., and R. Lavery. 1998. Sequence-dependent dynamics of TATA-box binding sites. *Biophys. J.* 75:372–381.
6. Lankaš, F., J. Šponer, J. Langowski, and T. E. Cheatham, III. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* 85:2872–2883.
7. Matsumoto, A., and W. K. Olson. 2002. Sequence-dependent motions of DNA: a normal mode analysis at the basepair level. *Biophys. J.* 83:22–41.
8. Araúzo-Bravo, M. J., S. Fujii, H. Kono, S. Ahmad, and A. Sarai. 2005. Sequence-dependent conformational energy of DNA derived from molecular dynamics simulations: toward understanding the indirect readout mechanism in protein-DNA recognition. *J. Am. Chem. Soc.* 127:16074–16089.
9. Fujii, S., H. Kono, S. Takenaka, N. Go, and A. Sarai. 2007. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.* 35:6063–6074.
10. Pérez, A., F. Lankas, F. J. Luque, and M. Orozco. 2008. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.* 36:2379–2394.
11. Araúzo-Bravo, M. J., and A. Sarai. 2008. Indirect readout in drug-DNA recognition: role of sequence-dependent DNA conformation. *Nucleic Acids Res.* 36:376–386.
12. Drew, H. R., and R. E. Dickerson. 1981. Structure of a B-DNA dodecamer III. Geometry of hydration. *J. Mol. Biol.* 151:535–556.
13. Mack, D. R., T. K. Chiu, and R. E. Dickerson. 2001. Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts. *J. Mol. Biol.* 312: 1037–1049.
14. Shatzky-Schwartz, M., N. D. Arbuckle, M. Eisenstein, D. Rabinovich, A. Bareket-Samish, et al. 1997. X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.* 267:595–623.
15. Shui, X., C. C. Sines, L. McFail-Isom, D. VanDerveer, and L. D. Williams. 1998. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry.* 37:16877–16887.
16. Tereshko, V., G. Minasov, and M. Egli. 1999. A ''Hydrat-ion'' spine in a B-DNA minor groove. *J. Am. Chem. Soc.* 121:3590–3595.
17. Chuprina, V. P., U. Heinemann, A. A. Nurislamov, P. Zielenkiewicz, R. E. Dickerson, et al. 1991. Molecular dynamics simulation of the hydration shell of a B-DNA decamer reveals two main types of minor-groove hydration depending on groove width. *Proc. Natl. Acad. Sci. USA.* 88:593–597.
18. Wahl, M. C., S. T. Rao, and M. Sundaralingam. 1996. Crystal structure of the B-DNA hexamer d(CTCGAG): model for an A-to-B transition. *Biophys. J.* 70:2857–2866.
19. Privé, G. G., K. Yanagi, and R. E. Dickerson. 1991. Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *J. Mol. Biol.* 217:177–199.
20. Dunitz, J. D. 1994. The entropic cost of bound water in crystals and biomolecules. *Science.* 264:670.
21. Crane-Robinson, C., A. I. Dragan, and P. L. Privalov. 2006. The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem. Sci.* 31:547–552.
22. Privalov, P. L., A. I. Dragan, C. Crane-Robinson, K. J. Breslauer, D. P. Remeta, et al. 2007. What drives proteins into the major or minor grooves of DNA? *J. Mol. Biol.* 365:1–9.

23. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, J. Wang, et al. 2002. AMBER7. University of California, San Francisco.

24. Dickerson, R. E., M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, et al. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 208:787–791.

25. Lu, X.-J., and W. K. Olson. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31:5108–5121.

26. Berman, H. M., and B. Schneider. 1999. Nucleic acid hydration. *In* Oxford Handbook of Nucleic Acid Structure. S. Neidle, editor. Oxford University Press, New York. 295–312.

27. Feig, M., and B. M. Pettitt. 1999. Modeling high-resolution hydration patterns in correlation with DNA sequence and conformation. *J. Mol. Biol.* 286:1075–1095.

28. Yonetani, Y., H. Kono, S. Fujii, A. Sarai, and N. Go. 2007. DNA deformability and hydration studied by molecular dynamics simulation. *Mol. Simulat.* 33:103–107.

29. Auffinger, P., and E. Westhof. 2000. Water and ion binding around RNA and DNA (C,G) oligomers. *J. Mol. Biol.* 300:1113–1131.

30. Strahs, D., and T. Schlick. 2000. A-tract bending: insights into experimental structures by computational models. *J. Mol. Biol.* 301:643–663.

31. Jayaram, B., and T. Jain. 2004. The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* 33:343–361.

32. Huth, J. R., C. A. Bewley, M. S. Nissen, J. N. S. Evans, R. Reeves, et al. 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol.* 4:657–665.

33. Dragan, A. I., J. R. Liggins, C. Crane-Robinson, and P. L. Privalov. 2003. The energetics of specific binding of AT-hooks from HMGA1 to target DNA. *J. Mol. Biol.* 327:393–411.

34. Quintana, J. R., A. A. Lipanov, and R. E. Dickerson. 1991. Low-temperature crystallographic analyses of the binding of Hoechst 33258 to the double-helical DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *Biochemistry.* 30:10294–10306.

35. Han, F., N. Taulier, and T. V. Chalikian. 2005. Association of the minor groove binding drug Hoechst 33258 with $d(CGCGAATTCGCG)_2$: volumetric, calorimetric, and spectroscopic characterizations. *Biochemistry.* 44:9785–9794.

36. Haq, I., J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins, and J. B. Chaires. 1997. Specific binding of Hoechst 33258 to the $d(CGCAAATTTGCG)_2$ duplex: calorimetric and spectroscopic studies. *J. Mol. Biol.* 271:244–257.

37. Breusegem, S. Y., R. M. Clegg, and F. G. Loontiens. 2002. Base-sequence specificity of Hoechst 33258 and DAPI binding to five $(A/T)_4$ DNA sites with kinetic evidence for more than one high-affinity Hoechst 33258-AATT complex. *J. Mol. Biol.* 315:1049–1061.

38. Abu-Daya, A., P. M. Brown, and K. R. Fox. 1995. DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucleic Acids Res.* 23:3385–3392.

39. Watkins, D., C. Hsiao, K. K. Woods, G. B. Koudelka, and L. D. Williams. 2008. P22 c2 repressor-operator complex: mechanisms of direct and indirect readout. *Biochemistry.* 47:2325–2338.

40. Woda, J., B. Schneider, K. Patel, K. Mistry, and H. M. Berman. 1998. An analysis of the relationship between hydration and protein-DNA interactions. *Biophys. J.* 75:2170–2177.

41. Selvaraj, S., H. Kono, and A. Sarai. 2002. Specificity of protein-DNA recognition revealed by structure-based potentials: symmetric/asymmetric and cognate/non-cognate binding. *J. Mol. Biol.* 322:907–915.

42. Tan, S., and T. J. Richmond. 1998. Crystal structure of the yeast MATα2/MCM1/DNA ternary complex. *Nature.* 391:660–666.

43. Newman, M., T. Strzelecka, L. F. Dorner, I. Schildkraut, and A. K. Aggarwal. 1995. Structure of *Bam* HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science.* 269:656–663.

44. 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.

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

46. Feng, J.-A., 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.

47. Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, et al. 1992. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63: 751–759.

48. DeLano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA.