Free-Energy Maps of Base−Amino Acid Interactions for DNA−Protein Recognition

Fabio Pichierri,[†,‡] Misako Aida,[§,||] M. Michael Gromiha,[†] and Akinori Sarai*[,†]

Contribution from the Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba 305-0074, Japan, and National Cancer Center Research Institute, Tsukiji, Tokyo 104-0045, Japan

Received November 30, 1998

Abstract: DNA−protein recognition plays a central role in gene expression and regulation. Despite increasing structural data on DNA−protein complexes, the molecular mechanism of DNA−protein recognition is not well understood yet, partly because of the considerable extent of redundancy in the base−amino acid interactions as well as of the structural flexibility present within the same interaction pair. To understand the specificity of such interactions, we should examine the interaction energetics by taking account of the structural flexibility. We describe a strategy for elucidating the specificity of DNA−protein interactions by computer simulation, in which free energies of interactions between the amino acid side chains and base pairs are computed by extensive conformational sampling. The simulations enable us to estimate thermodynamic quantities, such as the interaction free energy, enthalpy, and entropy, for each given position of the C_α atom of the amino acid side chain by conformational averaging, and to evaluate the free-energy map around base pairs. We report the results for the interactions of Asn with base pairs. In the case of an A-T base pair, we observed a curved valley-shaped region of free-energy minima on the major-groove side of A. Inspection of global-minimum energy configurations of Asn-A interactions in this valley region shows specific double hydrogen bonds, N−H · · · N7 and C═O · · · HN6, commonly observed in DNA−protein complex structures. On the other hand, we also observed other kinds of double and single hydrogen bonds depending on the C_α position in this valley region. Furthermore, the Asn side chain at various positions of C_α is flexible enough to readjust its conformation to form similar interactions with the base pair. These results demonstrate the importance of structural flexibility for the specific interactions involved. The free-energy map shows subtle differences from the interaction energy map, revealing the role of entropy in the specificity. The comparison of the free-energy maps for G-C and A-T revealed significant differences in their patterns. These results suggest that our method can be used to dissect the mechanism of base discrimination by amino acids.

Introduction

DNA−protein recognition plays a central role in gene expression of living organisms.[1] Much experimental and theoretical effort has been made to understand its molecular basis. A large body of experimental data collected so far by X-ray crystallography and NMR spectroscopy has revealed several general features on the mode of DNA−protein interaction.[2,3] DNA-binding proteins often use certain motifs to achieve a sequence-specific fit with DNA. For example, the helix−turn−helix (HTH) motif,[4] which consists of a core of two α helices, inserts one "recognition helix" into the major groove of DNA. Thus, a "shape complementarity" between architectural elements plays an important role in DNA sequence recognition. Inspection of the DNA−protein interfaces at atomic resolution has revealed specific hydrogen bond patterns between chemical groups of amino acid side chains and base-pair atoms.[5,6] Nonspecific contacts of an electrostatic nature between amino acid side chains and the sugar−phosphate backbone of DNA have been identified to favor the correct orientation and anchorage of the protein.[3] In their classical paper, Seeman et al.[7] were the first to recognize the formation of "point contacts", particularly double hydrogen bonds, being required for amino acids to discriminate among different bases. Although the structural analysis of several DNA−protein complexes has revealed some frequently occurring interactions, such as Asn-A and Lys-G, there exists a considerable degree of redundancy in the possible interactions between amino acids and bases. As a matter of fact,

* Corresponding author: RIKEN Tsukuba Life Science Center. Phone: +81-298-36-9082. Fax: +81-298-36-9080. E-mail: sarai@rtc.riken.go.jp.

† The Institute of Physical and Chemical Research.

‡ Present address: Computational Science Laboratory, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi 351-0198, Japan.

§ National Cancer Center Research Institute.

|| Present address: Faculty of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526 Japan.

(1) See, for example: (a) Saenger, W.; Heinemann, U. *Protein-nucleic Acid Interaction*; Macmillan Press Ltd.: Basingstoke, 1989. (b) Travers, A. *DNA-Protein-Interactions*; Chapman and Hall: London, 1993.

(2) (a) Harrison, S. C. *Nature* **1991**, *353*, 715−719. (b) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem*. **1984**, *53*, 293−321. (c) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem*. **1992**, *61*, 1053−1095. (d) Freemont, P. S.; Lane, A. N.; Sanderson, M. R. *Biochem. J.* **1991**, *278*, 1−23. (e) Ollis, D. L.; White, S. W. *Chem. Rev.* **1987**, *87*, 981−995.

(3) Lilley, D. M. J. *DNA-Protein: Structural Interactions;* Oxford University Press: Oxford, 1995.

(4) (a) Brennan, R. G.; Matthews, B. W. *J. Biol. Chem.* **1989**, *264*, 1903−1906. (b) Harrison, S. C.; Aggarwal, A. K. *Annu. Rev. Biochem*. **1990**, *59*, 933−969. (c) Wintjens, R.; Rooman, M. *J. Mol. Biol.* **1996**, *262*, 294−313.

(5) (a) Mandel-Gutfreund, Y.; Schueler, O.; Margalit, H. *J. Mol. Biol*. **1995**, *253*, 370−382. (b) Mandel-Gutfreund, Y.; Margalit, H. *Nucleic Acids Res*. **1998**, *26*, 2306−2312.

(6) Kono, H.; Sarai, A. *Proteins* **1999**, *35*, 114−131.

(7) Seeman, N. C.; Rosenberg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804−808.

10.1021/ja984124b CCC: $18.00
Published on Web 06/17/1999

the same amino acid can interact with different bases, and different amino acids can interact with the same base.[5,6] Therefore, there is no one-to-one code-like correspondence between amino acids and bases,[5,6,8] although base−amino acid preferences have been identified for a specific binding motif family.[9] Furthermore, there are quite a few degrees of structural flexibility within the same interaction pair; e.g., the position of Lys is widely distributed around G in the structural data.[6] How can we understand the specificity of such interactions, and how can we quantify the redundancy in the recognition process? It will require not only structural information but also energetic information to resolve such questions.

The purpose of the present study is to establish a strategy for clarifying the mechanism of base discrimination by amino acids, where redundant atomic interactions are involved. Due to its enormous complexity, the recognition problem needs to be dissected at different levels of complexity. Here, we take a "bottom-up" approach, starting with the simplest system and extending to larger systems step-by-step by incorporating various factors. We chose to start with the study of the intrinsic interactions between an amino acid side chain and a DNA base pair. Such interactions have been investigated by means of quantum-chemical calculations.[10,11] However, those methods can deal with only a small number of conformations, whereas numerous configurations of base−amino acid contacts exist in the structures of DNA−protein complexes so far structurally characterized. Certain conformational averaging would be necessary to derive quantities which can be compared with experimental data. There have been some attempts to extract the empirical interaction potentials from statistical analyses of structural database[6] and binding data.[12] The former analysis has shown that the statistical potentials with respect to the C_α position of amino acid against base, when applied to the prediction of target sequences of transcription factors, has the capability for discriminating natural target sequences against random sequences.[6] In this study, therefore, we first attempt to derive the corresponding interaction potentials by averaging side-chain conformations around the C_α position. We combined force-field potential functions, derived from quantum-chemical calculations, with a systematic sampling of side-chain conformations for the given C_α position and computed a large number of interaction configurations. Thermodynamic quantities, such as the interaction free energy and conformational entropy, are then calculated and their maps constructed. Comparison of these maps for different combinations of amino acids and base pairs, together with the experimental information, will provide a physical basis for how they can discriminate. Here we applied our method to the interaction between Asn and base pairs. The results demonstrate the feasibility of the method for understanding the base discrimination by amino acids.

Method of Calculation

We formulated an algorithm to sample amino acid conformations and their interactions with base pairs and to derive the interaction free-energy surfaces. At first, an amino acid is positioned around the base pair, by placing its C_α atom at grid points on the base plane, to mimic the situation where the C_α position is fixed by the protein main chain. Although proteins would impose some restrictions on the ranges of the C_α position and side-chain orientation, we avoided imposing arbitrary restrictions in this study. Then, an initial side-chain orientation is generated by specifying polar angles θ and ϕ formed by the direction cosines of the $C_\alpha-C_\beta$ bond vector, as depicted in Figure 1. Side-chain rotamers are generated by systematically varying the torsion angles (χ_1 and χ_2 for Asn; the amide plane is fixed). For each conformation of the amino acid side chain, bonded and nonbonded interaction energies are computed. This process is repeated at each point of the grid. The increments $\Delta\theta$ and $\Delta\phi$ are adjusted so as to produce a uniform distribution of C_β positions on the spherical surface spanned by the $C_\alpha-C_\beta$ bond vector. For simplicity, we assume that there is no preferential direction of the $C_\alpha-C_\beta$ bond vector. For each grid point, about 1 million conformations are generated and the Boltzmann averaging is performed to calculate thermodynamic quantities. Finally, the maps for free energy, entropy, etc. are constructed by interpolating the values calculated at each grid point. These maps represent adiabatic potentials for the given C_α position of Asn around base pairs, which may be compared with the C_α distribution obtained from the structural database.

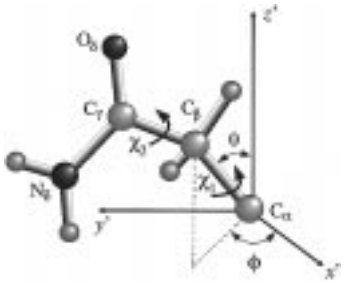

Figure 1. Side-chain rotation angles of Asn. C_α of Asn is placed in the relative coordinate system on the base plane.

The molecular structures of Asn and base pairs were obtained by carrying out ab initio MO calculations at the HF/6-31G* level of theory[13] with the HONDO program.[14] In this study, the effect of sugar−phosphate backbones was simplified by taking the fiber structure of the B-DNA conformation.[15] We constructed one unit of the B-DNA structure on computer graphics and replaced the base pair by the above optimized base-pair structure after superimposing the two base-pair structures. In the present calculations, we do not consider main-chain atoms of the amino acid, to avoid introducing possible bias into the interaction energy, since we focus on the intrinsic interactions by amino acid side chains. The main coordinate system was set by placing the *x* axis from N1 of pyrimidines to N9 of purines (see Figures 2 and 3 for atom numbers), with the origin at the middle point, the *y* axis pointing toward the major groove in the base plane (for simplicity, propeller twist is not introduced in this calculation), and the *z* axis pointing toward the helical axis.

We used the AMBER 4.1 force field of Kollman and co-workers.[16] The total energy (E_{tot}) is partitioned into bonded (E_b) and nonbonded (E_{nb}) energy terms. For E_b only the energy contribution due to dihedral angles is considered. The E_{nb} energy term is computed by the 6-12 potential and the Coulomb term to account for van der Waals and electrostatic interactions, respectively. The distance-dependent effective dielectric constant ($\epsilon = R_{ij}$, where R_{ij} is the distance between atoms) was included in the calculation to take account of the dielectric damping effect of the Coulomb interaction.

For the amino acid side chain at a given grid point (x, y), we calculated the partition function, $Z = \sum\exp(-E_{tot}/kT)$, where the

(8) Matthews, B. W. *Nature* **1988**, *335*, 294−295.
(9) (a) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11168−11172. (b) Choo, Y.; Klug, A. *Curr. Op. Struct. Biol.* **1997**, *7*, 117−125. (c) Suzuki, M.; Yagi, N. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12357-12361. (d) Suzuki, M. *Structure* **1994**, *2*, 317−326.
(10) (a) Sarai, A.; Saito, M. *Int. J. Quantum Chem.* **1984**, *25*, 527−533. (b) Sarai, A.; Saito, M. *Int. J. Quantum Chem.* **1985**, *28*, 399−409.
(11) Otto, P.; Clementi, E.; Ladik, J.; Martino, F. *J. Chem. Phys*. **1984**, *80*, 5294−5304.
(12) (a) Lustig, B.; Jernigan, R. L. *Nucleic Acids Res.* **1995**, *23*, 4707−4711. (b) Lustig, B.; Arora, S.; Jernigan, R. L. *Nucleic Acids Res.* **1997**, *25*, 2562−2565.
(13) (a) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222. (b) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFree, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(14) Dupuis, M.; Marquez, A.; Davidson, E. R. HONDO 96, Quantum Chemistry Program Exchange, Indiana University, 1996.
(15) Arnott, S.; Hukins, D. W. L. *J. Mol. Biol.* **1973**, *81*, 93−105.
(16) 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.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

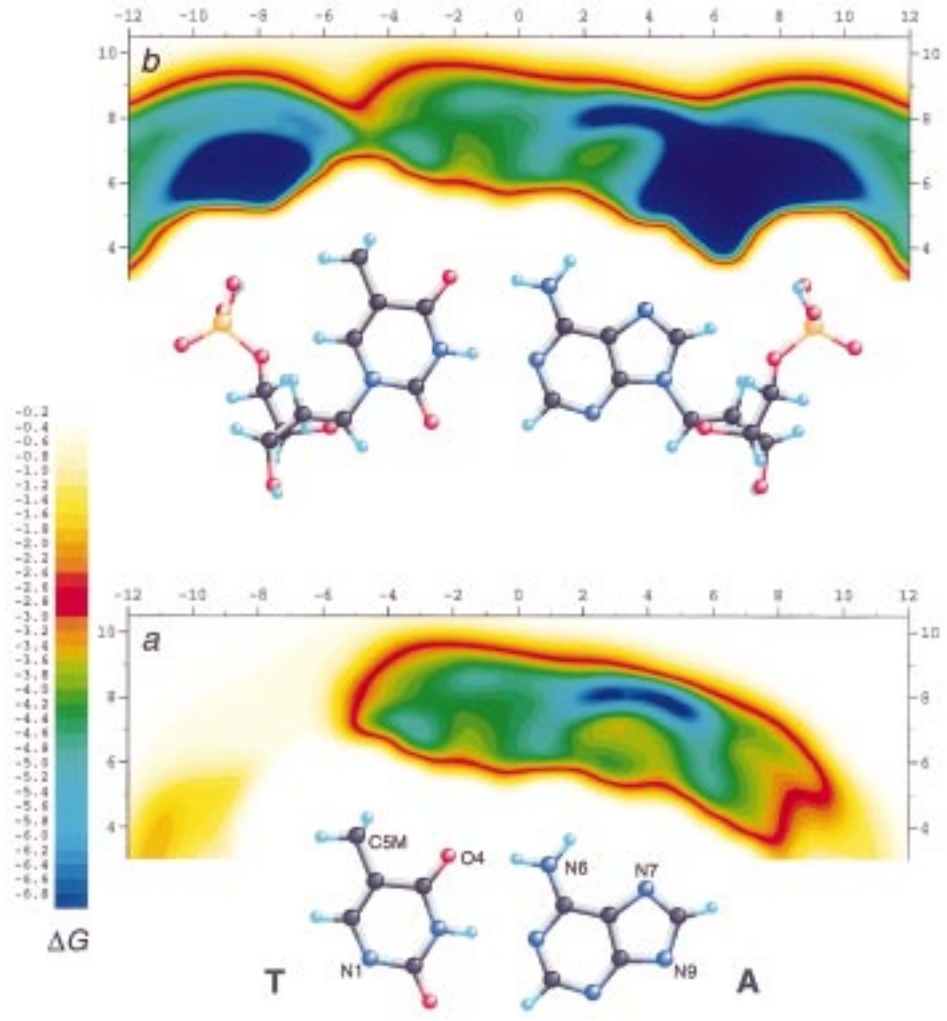

Figure 2. Free-energy maps of interactions between Asn and A-T: (a) base pair only and (b) base pair plus neutralized sugar−phosphate backbone.

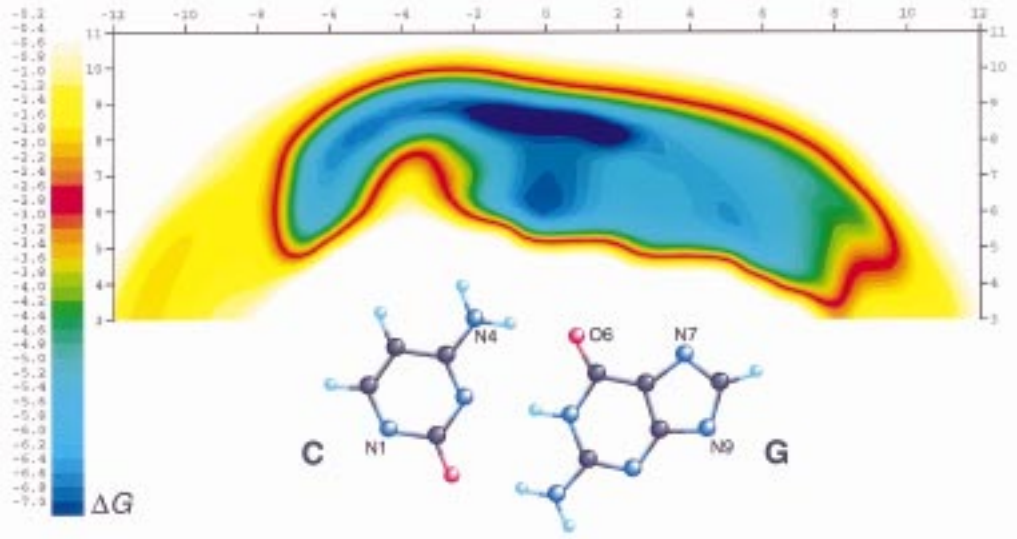

Figure 3. Free-energy map of interactions between Asn and G-C (base pair only).

summation is taken over the whole angle space (see Figure 1). Then we calculated the free energy, enthalpy, and entropy by $G = -kT \ln Z$, $H = \{\sum E_{tot}\exp(-E_{tot}/kT)\}/Z$, and $S = (H - G)/T$, respectively. The same quantities were also calculated for a large separation between amino acid and base pair as a reference (C_α at $x = 0$ Å, $y = 100$ Å), to evaluate the differences ΔG, ΔH, and ΔS due to the formation of the base−amino acid complex. The global-minimum energy, E_{min}, and the corresponding base−amino acid configuration at each grid point were derived during the sampling process.

Results

Free-Energy Maps for the Interactions between Asn and Base Pairs. We first applied our approach to the interactions

Table 1. Calculated Thermodynamic Quantities at Selected Grid Points[a]

model[b]	x	y	z	ΔG	ΔH	$-T\Delta S$	ΔE_{min}
M1	2.5	8.0	0.0	−8.54	−13.31	4.77	−14.02
M2	2.5	8.0	0.0	−10.63	−15.52	4.89	−16.17
M3	2.5	8.0	0.0	−9.16	−13.97	4.82	−14.56
M3	2.5	8.0	0.5	−8.98	−13.80	4.81	−14.40
M3	2.5	8.0	1.0	−8.78	−13.52	4.74	−13.95
M3	2.5	8.0	−0.5	−9.23	−14.15	4.92	−14.66
M3	2.5	8.0	−1.0	−9.16	−14.08	4.92	−14.60

[a] Coordinates are in units of Å, and energies in units of kcal/mol. T = 298 K. [b] M1: base pair only. M2: base pair plus sugar−phosphate backbones. M3: base pair plus neutralized sugar−phosphate backbones.

between an Asn side chain and an A-T base pair. A total of 673 920 conformations of the Asn side chain have been generated in the conformational space spanned by the polar and dihedral angles, with increments of 10° for each dihedral angle. Then, we computed E_{min} and the corresponding conformation of the Asn side chain, Z, ΔG, ΔH, and ΔS, at each grid point (C_α position). A total of 833 grid points with a 0.5 Å interval were considered in our simulation, to cover a sufficient portion of the major groove of base pairs ($-12 \le x \le +12$ Å; $+3 \le y \le +11$ Å). We checked the effect of the dihedral angle increment on these thermodynamic quantities by varying it in the range from 1° to 30° and found that, for example, the ΔG value changed only by 0.5% when the increment increases from 1° to 10°.

The ΔG maps obtained by interpolating the calculated ΔG values at the grid points are shown in Figures 2a and 3. The areas in blue are the lowest values of ΔG, which indicates favorable interactions between the Asn side chain and the base pairs. In the case of A-T, the ΔG minima are localized in a slightly curved valley-shaped region around A reaching between N6 and N7. The ΔG values in this range are between −8 and −9 kcal/mol (see also Table 1). On the other hand, Asn tends to be more broadly distributed around G-C, as shown in Figure 3. The lowest ΔG values are located in the middle of G-C. In contrast to A-T, the low-ΔG region extends toward C, which does not have a methyl group. Thus, the specificity of interaction of Asn toward A-T or G-C appears quite different.

Effect of Sugar−Phosphate Backbones on the ΔG Map. We examined the effects of the presence of sugar−phosphate backbones on the ΔG map by considering three different molecular models. The first model (M1) considers only base-pair atoms and the total atomic charge is equal to zero; in the second model (M2), the atoms of the sugar−phosphate moieties are included in the simulation and the total charge is set to −2; and in the third model (M3), the phosphate groups are included but the total charge is set to zero simply by adding +0.25 charge unit to each phosphate oxygen atom, to mimic the screening effect exerted by the counterions. Figure 2b shows the effect of backbones on the ΔG map for the case of A-T. The presence of backbones without net charges introduces additional regions of ΔG minima, concentric to the phosphate groups, in which the NH group of Asn forms hydrogen bonds with nonbridging oxygen atoms of the backbones. The presence of the charged backbones enlarges the regions of ΔG minima (not shown). These regions introduced by backbones partially overlap with the region of ΔG minima due to the base pair. Therefore, the sugar−phosphate backbones of DNA will interfere with some of the interactions between the Asn side chain and base pairs when Asn is near the DNA backbones. Otherwise, the backbones have little effect on the interactions between Asn and an A-T base pair, as shown in Table 1.

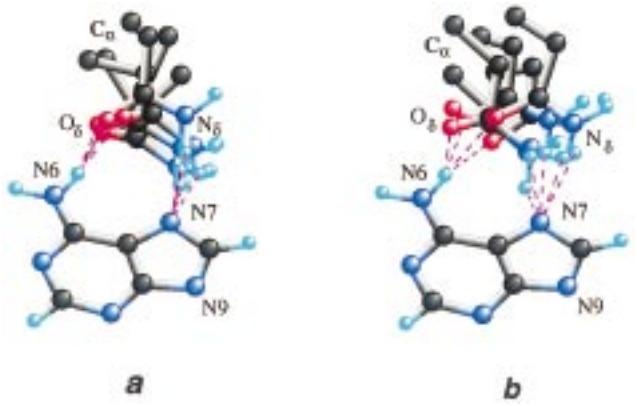

Figure 4. Comparison between simulated and experimental interaction configurations of Asn and A-T: (a) five global-minimum energy configurations at the grid points (x, y, z) = (+3.0, +8.0, +1.0), (+2.0, +7.5, 0.0), (+2.5, +8.0, −0.5), (+4.5, +8.0, −1.0), and (+4.0, +8.0, +0.5) and (b) experimental structures extracted from the following protein−DNA complexes deposited in the Brookhaven Protein Database (PDB):[19] 1APL(182),[20] 1MSE(183),[21] 1OCT(151),[22] 1UBD(369),[23] and 4RVE(185),[24] where the numbers in parentheses indicate the chosen amino acid residue.

Global-Minimum Energy Configurations. The present calculations enabled us to obtain the global-minimum energy and corresponding configuration of the Asn side chain interacting with base pairs at each grid point. Figure 5b shows the landscape of global-minimum energy. The ΔE_{min} map for A-T exhibits a curved valley-shaped region of energy minima similar to that of the ΔG map (Figure 2a). However, the ΔE_{min} map has a wider area of low-energy region (light-blue region). This difference between ΔE_{min} and ΔG maps results from the entropic effect of the amino acid side chain (see below). The valley-shaped region of minima spans the HN6 and N7 atoms of A, suggesting that Asn forms specific interactions with these atoms. We found several configurations in which Asn forms double hydrogen bonds, i.e., C═O · · · HN6 and NH · · · N7, as shown in Figure 4a. These configurations are frequently found in the Asn-A pair in the experimental structures of DNA−protein complexes (see Figure 4b). On the other hand, we also observed other configurations; e.g., the NH moiety of Asn strongly interacts with O4 of T, while C═O interacts with HN6 of A; Asn forms only one hydrogen bond with HN6 or N7, with the amide plane of Asn nearly perpendicular to the base plane. Also we noticed that Asn forms a bridging hydrogen bond between N7 of A and the oxygen atoms of the sugar−phosphate backbones, when the C_α atom position comes close to the backbone. In the case of G-C, the ΔG map shows minima between G and C. In this region, Asn forms double hydrogen bonds with N4 of C and O6 of G. In the other broad area of ΔG minima around G-C, Asn exhibits a variety of interaction configurations with G and C, mostly forming single hydrogen bonds with N4 of C and O6 and N7 of G. These results suggest the existence of structural variability within the same base−amino acid pair.

Effect of Asn Displacement from the Base Plane. The above calculations were carried out for C_α positions on the base plane, i.e., z = 0 Å. Since for the real interactions the C_α position of Asn is not restricted to the base plane, we examined the effect of out-of-plane displacement of the C_α atom of Asn. The C_α atom was displaced from the plane in the following range: $-1 \le z \le +1$ Å. The displacement of Asn enhanced or reduced the effect of the DNA backbone due to the variation in the distance between the amino acid and phosphate groups. Otherwise, the essential characteristics of the ΔG map remained the same as those corresponding to the in-plane case (z = 0).

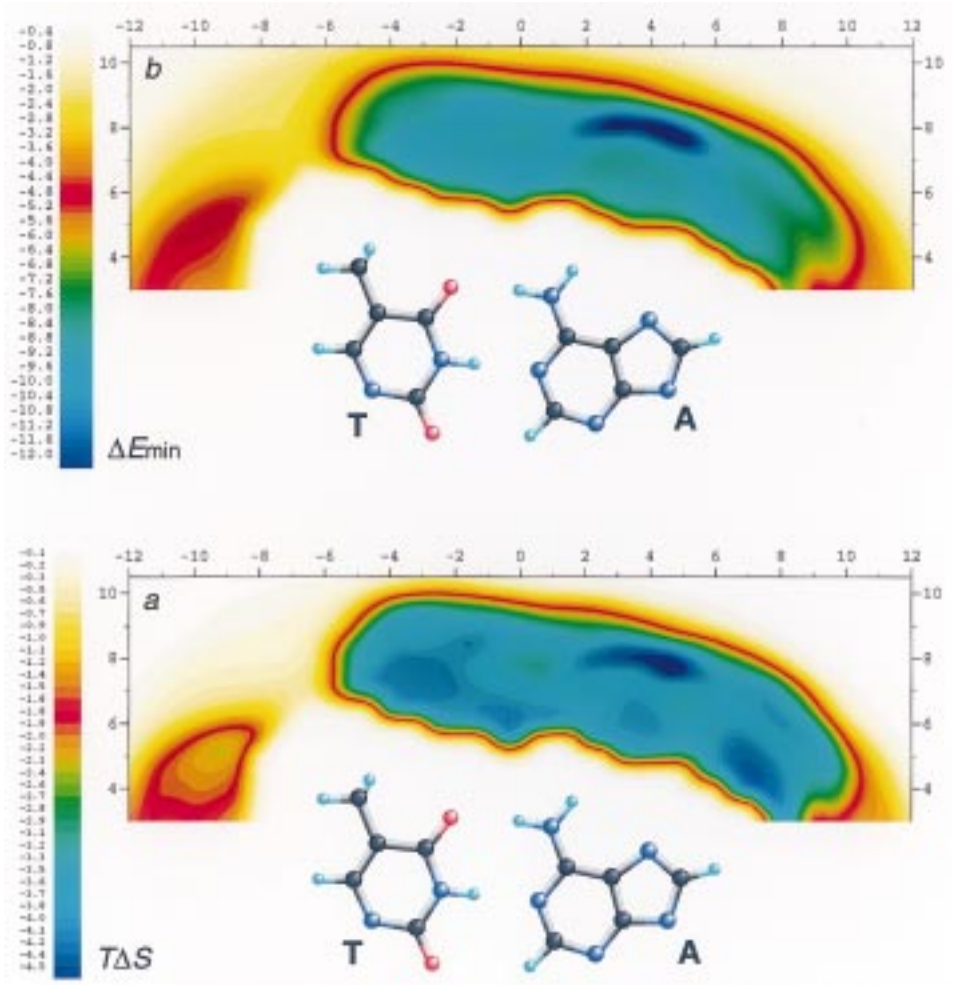

Figure 5. Maps of $T\Delta S$ (a) and ΔE_{min} (b) for the interactions between Asn and A-T (base pair only).

The ΔG values at the positions where Asn forms double hydrogen bonds with A are little affected by the out-of-plane displacement of C_α (see Table 1), suggesting that the Asn side chain might be flexible enough to readjust its conformation to form similar double hydrogen bond contacts upon displacement (see Figure 4).

Conformational Entropy Effect. In general, the conformational entropy associated with the amino acid side chains can contribute to the free energy of protein folding or protein association with other molecules, since the conformational degrees of freedom of the side chains may be severely restricted in the folded or complexed state. The present calculations yield the entropy change of the Asn side chain associated with the interaction with an A-T base pair. Figure 5a shows the map of $T\Delta S$, which exhibits several regions of minima. The $T\Delta S$ map shows a curved valley-shaped region of energy minimum corresponding to that in the ΔE_{min} map, where the Asn mostly forms double hydrogen bonds, N−H · · · N7 and C═O · · · HN6, with A. In this region, the loss of conformational entropy of the Asn side chain relative to its free state amounts to nearly 5 kcal/mol (Table 1). This is reasonable because the double hydrogen bonds restrict the movement of the side chain and lower its conformational entropy. Inspection of the global-minimum configuration at the other minimum regions shows that the Asn side chain forms at least two contacts with the atoms of A-T: NH_2 · · · O4(T) and $C_\beta H$ · · · CH_3(T); NH_2 · · · O4(T) and O_δ · · · $N6H_2$-(A); and O_δ · · · $N6H_2$(A) and $C_\beta H$ · · · N7(A). Thus, the movement of an Asn side chain is severely restricted at the grid points located in these regions. Some of these entropic effects appear to counterbalance the effect of energy (enthalpy) and enhance the specificity of the ΔG map (see Figures 5b and 2a).

Discussion

In this study, we examined the free-energy maps of the interactions between base pairs and amino acids by extensive conformational sampling. The ΔG maps for Asn revealed significant differences between A-T and G-C. In the case of A-T, the ΔG map exhibited a specific narrow region of minima around A (Figure 2a). This map closely resembles the corresponding map obtained from the statistical analysis of protein−DNA complex structures (Figure 6a of ref 6), indicating the feasibility of the present method. In contrast with A-T, the ΔG map for G-C showed a much larger low-ΔG region around G and C (Figure 3). This result indicates that the interaction of Asn is more specific to A-T than G-C, and for a given C_α position, Asn can discriminate either A against G or T against C. In fact, the interaction of Asn with A is most frequently observed among other bases in the DNA−protein complex structures.[5,6]

Inspection of the Asn side-chain conformations and their contacts with A-T shows not only typical double hydrogen bond interactions with A but also a variety of other types of interactions depending on the C_α position. These results suggest that the interaction between Asn and A-T is quite flexible. In real DNA−protein complexes, the position of C_α relative to the base pair may be constrained, so that additional factors should be considered. However, the intrinsic interactions between Asn and A-T, revealed by the present calculation,

indicate that numerous interaction configurations with similar values of free energy are possible in a specific region. Also, similar interactions can be formed from different directions and positions of C_α (Figure 4) despite how short the side chain of Asn is. Thus, a consideration of structural flexibility is essential for the quantification of specificity in the base−amino acid interactions.

The present free-energy values include contributions from both enthalpic and side-chain entropic terms. Comparison of the ΔG and ΔE_{min} (or ΔH) maps (Figures 2a and 5b) shows that there is a subtle difference between them. The origin of this difference is due to an entropic effect. The $T\Delta S$ map (Figure 5a) shows several minima where the degrees of freedom of Asn are restricted by multiple contacts with A-T. The lowest value of $T\Delta S$ is about −5 kcal/mol at 298 K. We find a similar value for G-C. This value is considerably higher than those generally estimated for protein folding.[17] The loss of side-chain conformational entropy of Asn ($-T\Delta S$) upon protein folding has been estimated[17] to be of the order of ∼1 kcal/mol, which is only 18% of the total side-chain entropy. The larger loss in our result may be partly due to the rigid conformation of Asn in its interaction with A, whereas the conformation of Asn in folded proteins may be more flexible. Also, in the present calculations we are moving the direction of the $C_\alpha-C_\beta$ vector freely, which will overestimate the absolute entropy value, in comparison to the contribution due only to the variation of dihedral angles. The present results suggest that the conformational entropy of the side chains should contribute significantly to the interaction free energy and specificity even for the case of the relatively short Asn side chain. Hence, we expect that amino acids with longer side chains will have even more pronounced effects while interacting with base pairs of DNA.

The strategy of the present study is based on a bottom-up approach, starting with the simplest system, the intrinsic interactions between base pairs and amino acids. Therefore, there are still many factors to be considered in the analysis. The structures of base pairs and DNA backbone were fixed in the calculation, but in reality they are not rigid. The effects of their flexibility and relaxation in bond length, angle, and dihedrals on the interaction process should be taken into account. The effects of multiple base-pair steps, counterions, and solvents should also be considered. The dielectric damping effect was considered as the distance-dependent dielectric constant, but its validity in the protein−DNA complex remains to be tested. In the present calculations, we did not impose any restrictions on the position of C_α and the orientation of the side chain, but in reality proteins would impose some restrictions to these ranges. Since the present ΔG map represents adiabatic potentials for a given C_α position against base pairs, this may be compared with the C_α distribution obtained from the structural database, and some potentials of mean force to mimic the effect of protein main chain could be imposed on the interaction. It would also be possible to introduce some restrictions on the $C_\alpha-C_\beta$ bond orientation when more experimental data become available for comparison. The present approach will be extended to other combinations of amino acids and base pairs, to compare the specificity of a given amino acid toward an A-T or G-C base pair as well as the specificity of a given base pair for all amino acids. In that stage, more efficient conformational sampling algorithms such as the multi-canonical Monte Carlo sampling method[18] would be required for the simulations of more complex systems.

(17) Doig, A. J.; Sternberg, M. J. E. *Protein Sci.* **1995**, *4*, 2247−2251.

Conclusions

We have presented a novel approach for analyzing the DNA−protein recognition process in terms of the free-energy map of interaction between amino acids and base pairs. By applying our method to the interactions between an Asn side chain and base pairs, we were able to show the difference in interaction specificity toward A-T and G-C. We also highlighted the importance of both structural flexibility and entropic effects in the specific interactions between bases and amino acids. We anticipate that the systematic analyses of base−amino acid interactions by these calculations would provide a physical basis of how amino acids and base pairs can discriminate with each other.

Acknowledgment. The authors thank Dr. J. Shen for his help at the initial stage of this study and Dr. H. Kono for providing the experimental Asn-A structures. The Japan Society for advancement of Technology (JST) is acknowledged for the use of their computer resources. M.M.G. thanks RIKEN for financial support. F.P. is grateful to the Science and Technology Agency (STA) of Japan for a grant fellowship.

JA984124B

(18) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9−12.

(19) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E.; Bryce, M. D.; Rogers, J. R.; Kennard, O.; Shikanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

(20) (a) Wolberger, C.; Vershon, A. K.; Liu, B.; Johnson, A. D.; Pabo, C. O. *Cell* **1991**, *67*, 517−528. (b) Wolberger, C.; Vershon, A. K.; Liu, B.; Johnson, A. D.; Pabo, C. O. *J. Mol. Biol.* **1991**, *217*, 11−13.

(21) (a) Ogata, K.; Morikawa, S.; Nakamura, H.; Sekikawa, A.; Inoue, T.; Kanai, H.; Sarai, A.; Ishii, S.; Nishimura, Y. *Cell* **1994**, *79*, 639−648. (b) Ogata, K.; Hojo, H.; Aimoto, S.; Nakai, T.; Nakamura, H.; Sarai, A.; Ishii, S.; Nishimura, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6428−6432.

(22) Klemm, J. D.; Rould, M. A.; Aurora, R.; Herr, W.; Pabo, C. *Cell* **1994**, *77*, 21−32.

(23) Houbaviy, H. B.; Usheva, A.; Shenk, T.; Burley, S. K. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13577−13582.

(24) (a) Winkler, F. K. *Curr. Op. Struct. Biol.* **1992**, *2*, 93−99. (b) Winkler, F. K.; D'Arci, A.; Bloecker, H.; Frank, R.; van Boom, J. H. *J. Mol. Biol.* **1991**, *217*, 235−238.