BSSE-corrected Three-body Interaction Energy in the Recognition of GC Base Pair by Asparagine

Tomoki Yoshida and Misako Aida*
Center for Quantum Life Sciences, and Department of Chemistry, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526

(Received October 13, 2006; CL-061212; E-mail: maida@hiroshima-u.ac.jp)

We investigate the total binding energy and many-body components in the recognition of guanine (G)–cytosine (C) base pair by asparagine residue (N). There are two possible bridged recognition structures for GCN trimers; one is in the major groove and the other is in the minor groove. The hydrogen-bonding pattern is the same for both trimers, while the binding energies and the components are significantly different. Three-body interaction energy depends on the binding site of N on GC pair. The difference in the relative direction between the dipole moments of GC base pair and N may partly account for the difference in the two-body interaction energies of the trimers.

Sequence specific DNA–protein interaction plays an important role in biological processes, e.g., DNA replication, translation, and transcription. In spite of growing number of available structural data in DNA–protein interaction, the molecular mechanism of the specific recognition is still unclear. The interaction between amino acid side chains and base pairs is one of fundamental factors in DNA–protein recognition.

A base pair in DNA can form various hydrogen-bonding patterns with an amino acid side chain,[1–4] not only in the major groove but also in the minor groove of DNA double helix. Asparagine residue (N) is one of amino acid residues which are found frequently in binding sites of DNA–protein recognition. N can form the same bridged hydrogen-bonding pattern with GC base pair in the major groove side as in the minor groove side, where two hydrogen bonds are formed, one between G and N, and one between C and N. In this letter, we compare two GCN complexes, i.e., the major groove recognition complex (complex-I) and the minor groove complex (complex-II). We calculate the total binding energies of the complexes together with the deformation energies of the monomers and two- and three-body interaction energies.

Although the evaluation of many-body contributions to intermolecular binding energy is an important subject in molecular interactions,[5–8] those contributions in DNA–protein recognition have not been taken into account yet. This letter presents the first evaluation; we show that the interaction energy and its components in the major groove recognition complex of GCN are different from those in the minor groove recognition complex.

The geometries of two GCN complexes are optimized with the MP2/6-31G* level of theory without frozen core approximation. This level of theory is important for reliable prediction of the charge distribution, dipole moments, and geometries of base pairs.[9] N is modeled by propionamide. To evaluate the deformation energies, the geometries of G, C, the isolated GC base pair and N with all-trans conformation are also optimized with the same level of theory. Each of the optimized geometries is confirmed to be the local minimum by the normal mode analysis.

The binding energy of XYZ trimer, $E_{\mathrm{bind}}^{\mathrm{X,Y,Z}}$, is defined as follows.

$$E_{\mathrm{bind}}^{\mathrm{X,Y,Z}} = E^{\mathrm{XYZ}} - (E_0^{\mathrm{X}} + E_0^{\mathrm{Y}} + E_0^{\mathrm{Z}}). \quad (1)$$

Here, E^{XYZ} is the total energy of XYZ trimer, E_0^{X} is the total energy of the isolated monomer X. $E_{\mathrm{bind}}^{\mathrm{X,Y,Z}}$ is comprised of the following terms.

$$E_{\mathrm{bind}}^{\mathrm{X,Y,Z}} = E_1^{\mathrm{X,Y,Z}} + E_2^{\mathrm{X,Y,Z}} + E_3^{\mathrm{X,Y,Z}}. \quad (2)$$

Here, $E_1^{\mathrm{X,Y,Z}}$ collects the distortion energies of monomers in the trimer, $E_2^{\mathrm{X,Y,Z}}$ is the total of the interaction energies of dimers in the trimer, and $E_3^{\mathrm{X,Y,Z}}$ is the remaining contribution to $E_{\mathrm{bind}}^{\mathrm{X,Y,Z}}$.

$$E_1^{\mathrm{X,Y,Z}} = E^{\mathrm{X}} + E^{\mathrm{Y}} + E^{\mathrm{Z}} - (E_0^{\mathrm{X}} + E_0^{\mathrm{Y}} + E_0^{\mathrm{Z}}). \quad (3)$$

$$E_2^{\mathrm{X,Y,Z}} = E^{\mathrm{XY}} + E^{\mathrm{YZ}} + E^{\mathrm{ZX}} - 2(E^{\mathrm{X}} + E^{\mathrm{Y}} + E^{\mathrm{Z}}). \quad (4)$$

$$E_3^{\mathrm{X,Y,Z}} = E^{\mathrm{XYZ}} - (E^{\mathrm{XY}} + E^{\mathrm{YZ}} + E^{\mathrm{ZX}}) + E^{\mathrm{X}} + E^{\mathrm{Y}} + E^{\mathrm{Z}}. \quad (5)$$

For XY dimer, $E_{\mathrm{bind}}^{\mathrm{X,Y}}$, $E_1^{\mathrm{X,Y}}$, and $E_2^{\mathrm{X,Y}}$ are defined as follows.

$$E_{\mathrm{bind}}^{\mathrm{X,Y}} = E_1^{\mathrm{X,Y}} + E_2^{\mathrm{X,Y}}. \quad (6)$$

$$E_1^{\mathrm{X,Y}} = E^{\mathrm{X}} + E^{\mathrm{Y}} - (E_0^{\mathrm{X}} + E_0^{\mathrm{Y}}). \quad (7)$$

$$E_2^{\mathrm{X,Y}} = E^{\mathrm{XY}} - (E^{\mathrm{X}} + E^{\mathrm{Y}}). \quad (8)$$

The basis set superposition error (BSSE) to E_2 and E_3 are treated hierarchically.[5,10,11] All energy terms are evaluated with the basis set of XYZ trimer. The BSSE-corrected binding energy is accounted with the sum of E_1, and BSSE-corrected E_2 and E_3.

The optimized geometries of the GCN trimers are shown in Figure 1. Their Cartesian coordinates are given in the Supporting Information.[14] The BSSE-corrected binding energies and their

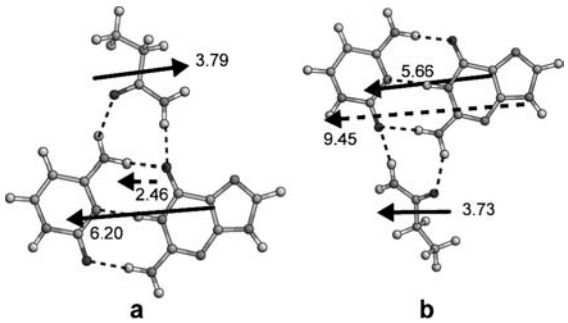

Figure 1. Optimized geometries of GCN trimers (MP2/6-31G*). a) Major groove recognition complex (complex-I); b) minor groove recognition complex (complex-II). In each case, dotted arrow shows the dipole moment of the trimer; solid arrow shows the dipole moment of N or GC base pair. The magnitude of each dipole moment is shown near the arrow in debye.

Table 1. BSSE-corrected total binding energy and its components in kcal/mol for the major groove recognition complex (complex-I)

X	Y	Z	$E_{\mathrm{bind}}^{\mathrm{X,Y,Z}}$	$E_1^{\mathrm{X,Y,Z}}$	$E_2^{\mathrm{X,Y,Z}}$	$E_3^{\mathrm{X,Y,Z}}$
G	C	N	−36.69	5.84	−39.40	−3.14
G	C	—	−22.88	5.02	−27.89	—
G	N	—	−2.29	3.80	−6.09	—
C	N	—	−2.55	2.86	−5.42	—
GC	N	—	−12.73	1.91	−14.64	—

Table 2. BSSE-corrected total binding energy and its components in kcal/mol for the minor groove recognition complex (complex-II)

X	Y	Z	$E_{\mathrm{bind}}^{\mathrm{X,Y,Z}}$	$E_1^{\mathrm{X,Y,Z}}$	$E_2^{\mathrm{X,Y,Z}}$	$E_3^{\mathrm{X,Y,Z}}$
G	C	N	−30.09	5.08	−35.52	0.35
G	C	—	−23.24	4.67	−27.91	—
G	N	—	−0.39	3.01	−3.39	—
C	N	—	−1.73	2.48	−4.21	—
GC	N	—	−6.09	1.16	−7.25	—

components of complex-I and complex-II are summarized in Tables 1 and 2, respectively.

Because of the same hydrogen-bonding pattern in complex-I and complex-II, it is often regarded that both complexes may have similar binding energy. However, complex-I is actually more stable than complex-II by around 7 kcal/mol (−36.7 vs −30.1 kcal/mol; see Tables 1 and 2) and the two-body interaction energy is ca. 4 kcal/mol more negative (−39.4 vs −35.5 kcal/mol) and the three-body interaction energy is ca. 3 kcal/mol more negative (−3.14 vs +0.35 kcal/mol) in the major groove recognition than in the minor groove recognition complex. Both two- and three-body interaction energies contribute almost equally to the difference in the binding energies between the major groove and the minor groove recognition complexes. This is the first explicit evaluation to indicate the importance of the three-body interaction in the DNA-protein recognition. It should be emphasized that most of potential energy functions do not account for this effect: care may be needed in the assessment of simulations using those potential energy functions.

It is interesting to note that the three-body interaction energy can be expressed as follows only with the two-body interaction energies of dimers in the trimer. The derivation of eq 9 is shown in the Supporting Information.[14]

$$E_3^{\mathrm{G,C,N}} = E_2^{\mathrm{GC,N}} - (E_2^{\mathrm{G,N}} + E_2^{\mathrm{C,N}}). \quad (9)$$

GCN trimer can be viewed as a dimer of GC base pair and N. The three-body interaction energy in the trimer is equal to the difference between the two-body interaction energies of GC–N and the sum of those for the constituent pairs (G–N and C–N). In complex-I, this difference is large: i.e., GC base pair is different from just the sum of G and C when N is bound in the major groove site. On the other hand, GC base pair is actually the same as the sum of G and C, when N is bound in the minor groove site. N brings about different induction effect on GC base pair depending on its binding position. The dipole moment of GC base pair in GCN trimer reflects this difference in the induction effect: the dipole moment of GC base pair in complex-I is larger than that in complex-II (Figure 1).

The two-body interaction energy between GC base pair and N, $E_2^{\mathrm{GC,N}}$ is about two times more negative in complex-I than in complex-II (−14.6 vs −7.25 kcal/mol). This is partly originated from the difference in the relative direction of the dipole moments of GC base pair and N. In complex-I, the dipole moments of GC base pair and N are in antiparallel, which brings about the stabilization. On the other hand, in complex-II, the dipole moments of GC base pair and N are in parallel with the same direction, which brings about destabilization (Figure 1).

The electron densities at the critical points of the two hydrogen bonds between GC base pair and N are higher in complex-I than in complex-II (Tables S14 and S15).[14] Those hydrogen-bond lengths are shorter in complex-I than those in complex-II.[14] These are consistent with the result that the interaction between GC base pair and N is stronger in complex-I than in complex-II.

In this letter, we have shown that there is a significant difference in the total binding energies for GCN trimers between the major and the minor groove recognitions. The difference originates from both two- and three-body interaction energies. It should be stressed that the amount of three-body interaction energy depends on the binding site. Three-body interaction may play an important role in DNA–protein recognition.

The critical points and the laplacian of the electron density are calculated using the AIM2000 program.[12] All geometries and energies are calculated using the Gaussian 03 program package[13] on the PC cluster system of the Center for Quantum Life Sciences (QuL*i*S) and the NEC SX7 at the Research Center for Computational Science.

This study was partly supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References and Notes

1 Y. Mandel-Gutfreund, O. Schueler, H. Margalit, *J. Mol. Biol.* **1995**, *253*, 370.
2 H. Kono, A. Sarai, *Proteins* **1999**, *35*, 114.
3 N. M. Luscombe, R. A. Laskowski, J. M. Thornton, *Nucleic Acids Res.* **2001**, *29*, 2860.
4 T. Yoshida, T. Nishimura, M. Aida, F. Pichierri, M. M. Gromiha, A. Sarai, *Biopolymers* **2001**, *61*, 84.
5 J. C. White, E. R. Davidson, *J. Chem. Phys.* **1990**, *93*, 8029.
6 S. S. Xantheas, *J. Chem. Phys.* **1994**, *100*, 7523.
7 P. Jurečka, P. Nachtigall, P. Hobza, *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
8 A. Chaudhari, P. K. Sahu, S.-L. Lee, *J. Chem. Phys.* **2004**, *120*, 170.
9 V. I. Danilov, V. M. Anisimov, *J. Biomol. Struct. Dyn.* **2005**, *22*, 471.
10 S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
11 P. Valiron, I. Mayer, *Chem. Phys. Lett.* **1997**, *275*, 46.
12 F. Biegler-König, J. Schönbohm, D. Bayles, *J. Comput. Chem.* **2001**, *22*, 545.
13 M. J. Frisch et al., *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford CT, **2004**.
14 Supporting Information is available electronically on the CSJ-Journal Web site, http://www.csj.jp/journals/chem-lett/index.html.
15 The figure in Graphical Abstract was created using the 'VMD – visual molecular dynamics' program. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.

Published on the web (Advance View) December 9, 2006; doi:10.1246/cl.2007.124