Theoretical Studies on Hydrogen Bonding Interactions between Peptide Units#

Misako AIDA
Biophysics Division, National Cancer Center Research Institute, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104
(Received May 13, 1993)

The hydrogen bonding interaction energies between peptide units were calculated using the *ab initio* MO method with the STO-3G, 7,3/2,1 and 6-31G basis sets and a correction of the basis set superposition errors (BSSE). The same systems were also investigated using an *ab initio* potential which has recently been developed and the AMBER potential, which is widely used. The energy profiles calculated using the *ab initio* potential are similar to those calculated using the 6-31G basis set. The distance and angle dependencies of the hydrogen bonding interactions are well represented by the *ab initio* potential. In contrast, they are not reproduced by the AMBER potential.

Hydrogen bonding interactions play important roles in maintaining both the secondary and tertiary structures of biological molecules. Especially in proteins, important structural elements, such as the α-helix and β-sheet, cannot be formed without hydrogen bonding interactions.

To elucidate the conformational characteristics of proteins, the conformations of model dipeptides have been studied using *ab initio* molecular orbital (MO) methods.[1—4] They are of great importance, since dipeptides can mimic segments of proteins. For protein structures, however, not only the nearest-neighbor interactions in dipeptide, but also the interactions between the distant residues, are important. It is thus necessary to describe the characteristics of hydrogen bonding interactions between peptides.

Hydrogen bonding interactions have been one of the major topics in quantum chemistry. A number of studies have been performed concerning the chemical properties of the hydrogen bond by means of *ab initio* MO methods.[5] *Ab initio* MO approaches have also been applied to hydrogen bonding interactions between small peptide fragments.[6—8] However, only little detailed information is available concerning the interactions between the peptide units, with the correction of superposition errors (BSSE). It is very important to take into account BSSE in obtaining reliable intermolecular interaction energies.[9]

To clarify the properties of small molecules, the *ab initio* MO method serves as an important tool. It is impracticable, however, to apply this method to large biological molecules, such as proteins and nucleic acids, because of the size and complexity of biological systems. To describe the static and dynamic properties of biological molecules, calculations using potential energy functions[10—12] are often performed. To parameterize those potential energy functions, experimental data from crystallography and spectroscopy are used with various combinations of *ab initio* MO calculations. Although these potentials are widely used in describing the conformations of large molecules, no quantitative assessment is available concerning the reliability of the calculations using those functions. A comparison of the structures obtained using these potential functions has shown that they are not sufficiently sensitive to reproduce very specific interactions.[13,14]

Recently, a new set of *ab initio* potentials has been derived for interactions of biological molecules based on only *ab initio* MO computations.[15] For the derivation, thirty three kinds of model molecules were used, which are fundamental and repeating units of DNA and proteins. The atoms in nucleic acids and proteins were classified according to their chemical environments. An effective charge, a modification of the charge obtained from a Mulliken population analysis,[16] was introduced and used to represent the electrostatic energy. More than thirty thousand SCF interaction energies with the correction of BSSE using the 7,3/2,1 basis set[17] were calculated between the model molecules in order to provide reference data for the fitting procedure to obtain parameters for 76 classes. The standard deviation was 1.61 kcal mol^{-1} (1 kcal mol^{-1}=4.184 kJ mol^{-1}) for interaction energies spanning the range from about -220 kcal mol^{-1} to $+20$ kcal mol^{-1}.

In this paper, the characteristics of hydrogen bonding interactions between the peptide units in various forms are described based on the *ab initio* MO method using several levels of basis sets with the correction of BSSE, and also with the *ab initio* potential[15] and AMBER potential.[18] It is shown how calculations with the potential energy functions can simulate the distance and angle dependencies of the hydrogen bonding energies obtained with the *ab initio* MO method.

Methods and Results

Ab initio MO calculations have been performed using the programs KGNMOL,[17] HONDO 7,[19] and HONDO 8,[20] and the basis sets of 7,3/2,1,[17] STO-3G,[21] 3-21G,[22] and 6-31G.[23] The counterpoise method[24] was applied to calculations of the interaction energies in order to reduce the basis set superposition errors (BSSE).

The calculations with potential energy functions were carried out using the KGNMD program in MOTECC.[25] Two kinds of parameter sets were used:

#This paper is dedicated to the late Professor Hiroshi Kato.

One was the *ab initio* potential;[15] the other was the AMBER potential,[18] which is used widely in the field of biology.

In the current work, CH_3CO–$NHCH_3$ (*N*-methylacetamide) was used as the smallest model molecule of the peptide unit. The geometry of this molecule (*cis* form) was optimized using the 3-21G basis set with C_s symmetry. The optimized bond lengths and bond angles are given in Fig. 1. The total energies of this molecule using various basis sets are summarized in Table 1.

Two kinds of the intermolecular conformations were taken for the hydrogen bonding CH_3CO–$NHCH_3$ molecules: System I and System II. The former represents the hydrogen bonding interaction in the planar form (Fig. 2), while the latter represents the perpendicular form (Fig. 3). The hydrogen bonding energies were calculated with the *ab initio* MO method by varying R, which is the distance between the hydrogen bonding H and O atoms (see Figs. 2 and 3), using the STO-3G, 7,3/2,1 and 6-31G basis sets, while taking account of BSSE. For the same systems, the interaction energies were calculated using the *ab initio* potential and the AMBER potential. The interaction energies are summarized in Tables 2 and 3 for System I and System II, respectively. They are shown graphically in Figs. 4 and 5 for System I and System II, respectively.

For the model of an anti-parallel β-sheet in protein we used a hydrogen bonding pair of the unit peptide: CH_3CO–$NHCH_2CO$–$NHCH_3$ (2-(acetylamino)-*N*-methylacetamide). The geometry of this unit peptide was optimized using the STO-3G basis set with C_s symmetry. The optimized bond lengths and bond angles are given in Fig. 6. The geometry corresponding to the conformation of an anti-parallel β-sheet in protein was constructed using the bond lengths and bond angles in this planar form and with -139.0 for ϕ and 135.0 for ψ. Here, ϕ is the dihedral angle for C'_{i-1}–N_i–$C_{\alpha i}$–C'_i, and ψ for N_i–$C_{\alpha i}$–C'_i–N_{i+1} (see Fig. 7). The total energies of this molecule in these two conformations are summarized in Table 4 along with two kinds of basis sets.

The intermolecular conformation for the anti-parallel β-sheet was constructed so as to be in the C_2 symmetry (System III; Fig. 8). The hydrogen bonding energies were calculated with the *ab initio* MO method by varying R, which is the distance between the hydrogen bonding H and O atoms (see Fig. 8), using the STO-3G and 7,3/2,1 basis sets, while taking account of BSSE. For the same system, the interaction energies were calculated using the *ab initio* potential and the AMBER potential. The interaction energies are summarized in Table 5 for System III. They are also shown graphically in Fig. 9.

Table 1. Total Energies (E_T) of CH_3CO–$NHCH_3$[a] with Various Basis Sets

Basis set	E_T/a.u.
STO-3G	−243.852662
7,3/2,1	−246.024050
3-21G	−245.629666
6-31G	−246.895658

a) The geometry was optimized with the 3-21G basis set. The bond lengths and bond angles are shown in Fig. 1.

Discussion

Biological molecules are so large that applying the *ab initio* MO method is impractical. Therefore, many kinds of analytical pair potentials have been proposed. These potential functions can be applied to any large biological molecule. However, only a few detailed analyses have been carried out concerning whether they can represent the distance dependency and the angle dependency of many kinds of interactions which are formed in biological molecules.

No estimate of the qualities of potential energy functions can be made through a comparison of the structures of biological molecules calculated using the potential energy functions with experimental observations. The structures of biological molecules vary with the experimental conditions. Since it is very difficult to take account of all of the experimental conditions in calculations, a straightforward comparison between the calculated and experimental structures may lead to misunderstandings. Any assessment should therefore be based on how the potential energy functions can simulate the computational results using the *ab initio* MO method with a basis set of good quality.

Hydrogen bonding interactions play important roles in biological systems. Potential functions should be able to represent the distance and angle dependencies of the hydrogen bonding interactions, so that those functions can be used to obtain information concerning biological molecules, such as the energy-minimized structures of peptides or proteins. The purpose of the current work was to obtain an assessment of the potential energy functions concerning the hydrogen bonding interaction between peptide units. It is impractical, however, to use an extended basis set for hydrogen bonding peptide units. Instead of using an extended basis set, we can use a basis set of moderate size, if we take into account BSSE.[9] For hydrogen bonding systems, single-determinant (Hartree–Fock) calculations are acceptable for investigating the geometry.[26]

Figure 4 shows the hydrogen bonding interaction energy profiles for a pair of small model peptides in the planar form, using various levels of methods. The hydrogen bonding energies calculated with the STO-3G basis set were quite underestimated, because of the limited flexibility of this basis set. The 7,3/2,1 basis set gives interaction energies of better quality, even though it is called "minimal basis set". This basis set gives almost the same value (R=2.0 Å) for the optimum distance of the hydrogen bond as is obtained with the 6-

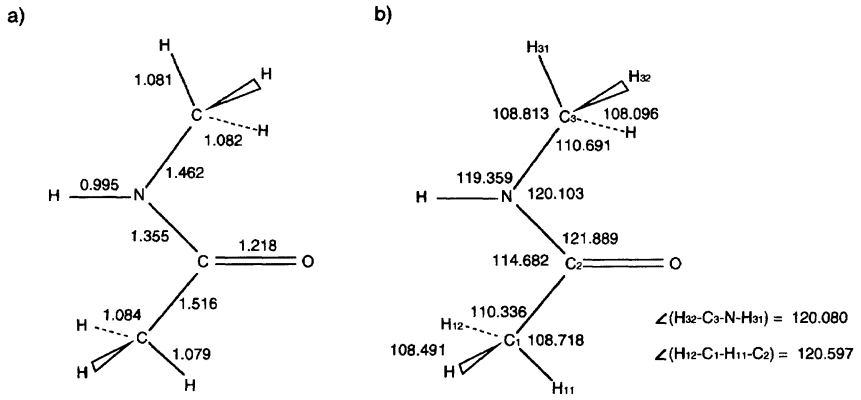

Fig. 1. Optimized geometry of CH_3CO–$NHCH_3$ in C_s symmetry with the 3-21G basis set. a) Bond lengths in Å; b) Bond angles in degrees.

Table 2. Interaction Energies (ΔE) of System I[a)]

R/Å[b)]	ΔE/kcal mol^{-1}				
	*ab initio*MO (6-31G)	*ab initio*MO (7,3/2,1)	*ab initio*MO (STO-3G)	*ab initio* potential	AMBER potential
1.40		7.17		3.75	491.34
1.50	1.09	1.44	6.18	−2.36	207.36
1.65		−2.97		−5.80	59.05
1.75	−5.43	−4.29	−0.34	−6.48	24.66
1.85	−6.34	−4.88	−1.43	−6.62	8.79
2.00	−6.80	−5.05	−2.26	−6.36	−0.77
2.10	−6.75	−4.92	−2.48	−6.06	−3.15
2.25	−6.41	−4.58	−2.54	−5.55	−4.43
2.40	−5.92	−4.17	−2.43	−5.04	−4.58
2.50	−5.57	−3.89	−2.31	−4.71	−4.44
3.00	−3.97	−2.70	−1.67	−3.40	−3.23
4.00	−2.14	−1.43	−0.90	−1.91	−1.69
5.00	−1.32	−0.87	−0.55	−1.18	−1.01
8.00	−0.46	−0.29	−0.19	−0.39	−0.35

a) Planar hydrogen bonding system of two CH_3CO–$NHCH_3$ molecules. See Fig. 2. b) R is the distance between hydrogen bonding H and O. See Fig. 2.

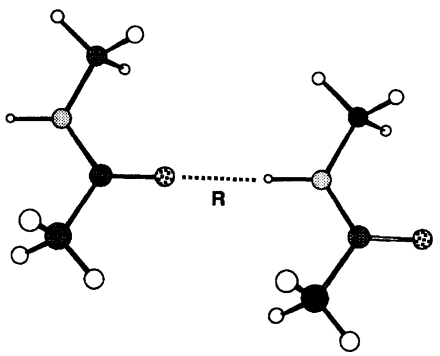

Fig. 2. System I: the planar form of the hydrogen bonding CH_3CO–$NHCH_3$ molecules. R is the distance between hydrogen bonding H and O atoms.

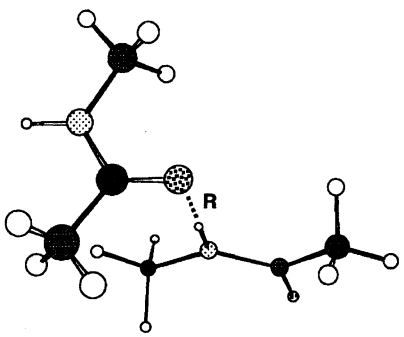

Fig. 3. System II: the perpendicular form of the hydrogen bonding CH_3CO–$NHCH_3$ molecules. R is the distance between hydrogen bonding H and O atoms.

31G basis set, which is a double-zeta level. The STO-3G basis set gives a slightly larger value (R=2.25 Å) for the optimum distance of the hydrogen bond.

For the same system, the interaction energies calculated with the *ab initio* potential are close to those with the 6-31G basis set, even though the optimum hydrogen bond distance is a slightly shorter (R=1.85 Å). The *ab initio* potentials were obtained from the *ab initio* MO interaction energies with the 7,3/2,1 basis set. In this

Table 3. Interaction Energies (ΔE) of System II[a)]

R/Å[b)]	ΔE/kcal mol^{-1}				
	*ab initio*MO (6-31G)	*ab initio*MO (7,3/2,1)	*ab initio*MO (STO-3G)	*ab initio* potential	AMBER potential
1.50				7.39	244.89
1.75	4.20	4.30	6.98	−0.81	36.01
1.85	1.58	1.96	4.33	−1.92	15.73
2.00	−0.79	−0.05	1.84	−2.72	2.42
2.15	−1.98	−1.01	0.48	−2.97	−2.46
2.25	−2.38	−1.32	−0.04	−2.99	−3.77
2.40	−2.62	−1.51	−0.47	−2.89	−4.46
2.50	−2.64	−1.53	−0.61	−2.78	−4.49
2.75	−2.43	−1.39	−0.72	−2.44	−4.02
3.00	−2.08	−1.17	−0.66	−2.09	−3.36
4.00	−1.00	−0.55	−0.35	−1.07	−1.50
5.00	−0.54	−0.29	−0.19	−0.56	−0.73
8.00	−0.13	−0.06	−0.04	−0.10	−0.14

a) Perpendicular hydrogen bonding system of two CH_3CO–$NHCH_3$ molecules. See Fig. 3.
b) R is the distance between hydrogen bonding H and O. See Fig. 3.

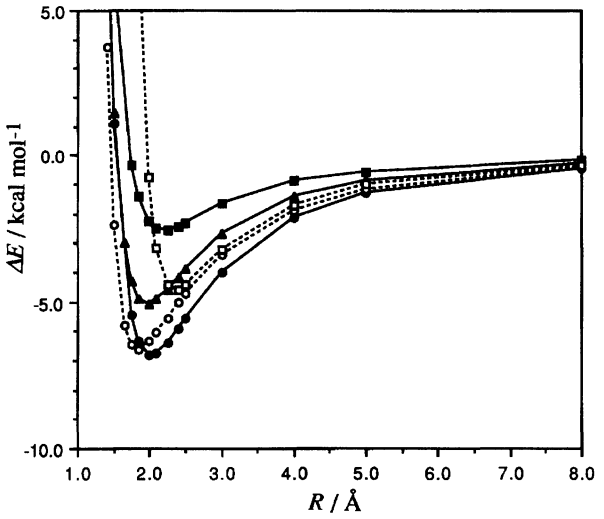

Fig. 4. Plots of the interaction energies (ΔE) of System I calculated with various methods. ●, *ab initio* MO (6-31G); ▲, *ab initio* MO (7,3/2,1); ■, *ab initio* MO (STO-3G); ○, *ab initio* potential; □, AMBER potential.

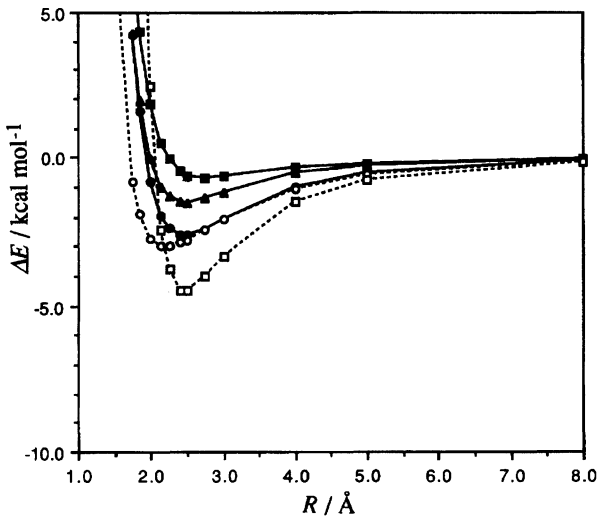

Fig. 5. Plots of the interaction energies (ΔE) of System II calculated with various methods. ●, *ab initio* MO (6-31G); ▲, *ab initio* MO (7,3/2,1); ■, *ab initio* MO (STO-3G); ○, *ab initio* potential; □, AMBER potential.

case, however, the interaction energy profile calculated with the *ab initio* potential resembles that calculated with the 6-31G basis set.

Figure 5 shows the hydrogen bonding interaction energy profiles in the perpendicular form. This conformation is far away from the optimal hydrogen bonding interaction. The *ab initio* MO interaction energies are smaller than those for the planar form (shown in Fig. 4), with any basis set. Both the 6-31G and the 7,3/2,1 basis sets give a similar optimum distance of the hydrogen bond in this conformation (R=2.5 Å). Stabilization by a hydrogen bonding interaction is not obtained with the STO-3G basis set in this conformation.

For the perpendicular system, the interaction energy profile calculated with the *ab initio* potential resembles that calculated with the 6-31G basis set. Here again, the optimum hydrogen bond distance is a slightly shorter (R=2.25 Å).

For the planar form of the CH_3CO–$NHCH_3$ molecules (System I; Table 2 and Fig. 4), the optimum hydrogen bond distance with the AMBER potential is about 2.4 Å with an interaction energy of −4.58 kcal mol^{-1}. For the perpendicular form (System II; Table 3 and Fig. 5), the optimum hydrogen bond distance with the AMBER potential is about 2.5 Å with an interaction energy of −4.49 kcal mol^{-1}. As shown in Figs. 4 and 5, the energy profiles for both the systems are quite similar. These results mean that the AMBER potential is not able to

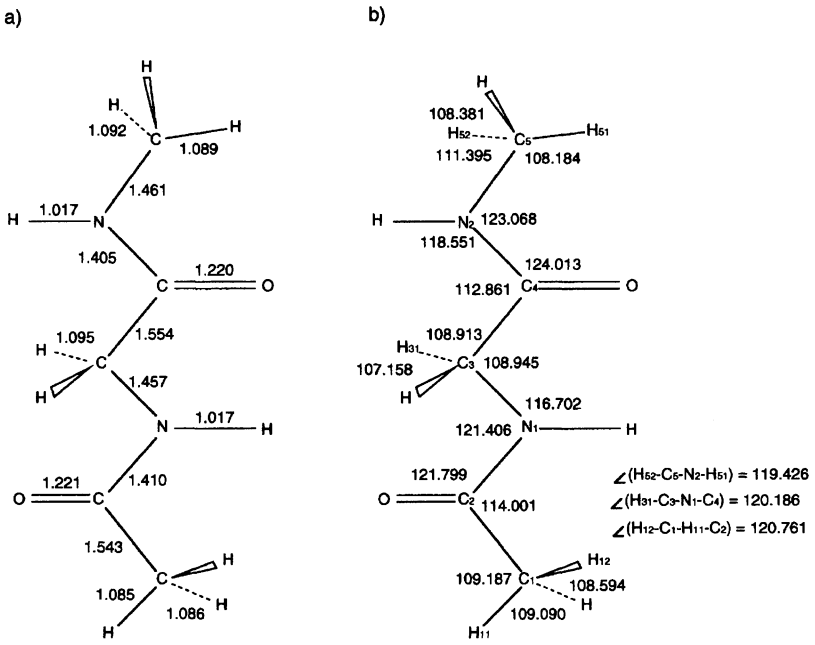

Fig. 6. Optimized geometry of CH_3CO–$NHCH_2CO$–$NHCH_3$ in C_s symmetry with the STO-3G basis set. a) Bond lengths in Å; b) Bond angles in degrees.

Table 4. Total Energies (E_T) of CH_3CO–$NHCH_2CO$–$NHCH_3$[a] with Various Basis Sets

Basis set	E_T/a.u.
STO-3G	−447.993320[b]
STO-3G	−447.987523[c]
7,3/2,1	−452.106539[c]

a) The bond lengths and bond angles optimized in C_s symmetry with STO-3G basis set are shown in Fig. 6. b) Total energy corresponding to the geometry (Fig. 6) optimized in C_s symmetry with STO-3G basis set. c) Total energy corresponding to the geometry in the conformation of anti-parallel β-sheet, using the bond lengths and bond angles shown in Fig. 6.

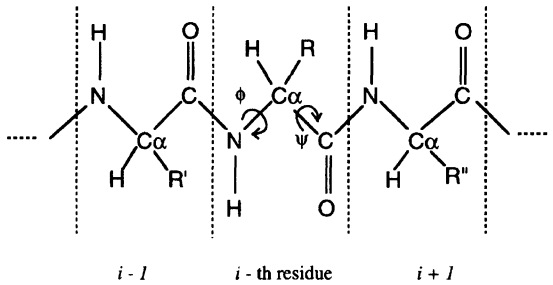

Fig. 7. Schematic representation of the protein main chain.

Table 5. Interaction Energies (ΔE) of System III[a]

R/Å[b]	ΔE/kcal mol^{-1}			
	*ab initio*MO (7,3/2,1)	*ab initio*MO (STO-3G)	*ab initio* potential	AMBER potential
1.40			10.88	1015.45
1.50	5.72	15.27	−1.03	434.40
1.60	−0.95	8.23	−6.32	192.29
1.70	−4.62	3.68	−8.50	85.91
1.80	−6.46	0.78	−9.17	37.10
1.90	−7.20	−1.00	−9.11	13.99
2.00	−7.32	−2.05	−8.69	2.85
2.10	−7.10	−2.61	−8.11	−2.50
2.20	−6.71	−2.85	−7.47	−4.97
2.30	−6.22	−2.88	−6.84	−5.97
2.50	−5.20	−2.63	−5.65	−6.07
2.75	−4.03	−2.12	−4.40	−5.12
3.00	−3.07	−1.64	−3.43	−4.06
4.00	−1.10	−0.57	−1.27	−1.49
5.00	−0.46	−0.22	−0.47	−0.58
7.00	−0.10	−0.04	−0.02	−0.11

a) Hydrogen bonding system of two CH_3CO–$NHCH_2CO$–$NHCH_3$ molecules in anti-parallel β-sheet conformation. See Fig. 8. b) R is the distance between hydrogen bonding H and O. See Fig. 8.

distinguish System II from System I.

As shown in Fig. 4, the optimum hydrogen bond distance calculated with the AMBER potential is about 0.5 Å longer than that calculated with the 6-31G and 7,3/2,1 basis sets. At distances where the calculations with the *ab initio* MO method show a large stabilization by a hydrogen bonding interaction (i.e., R of around 2.0 Å), the calculations with the AMBER potential show a large repulsion. This implies that the AMBER potential cannot properly describe the conformations of the hydrogen bonding molecules.

As shown in Figs. 4 and 5, the energy profiles calculated with the *ab initio* potential resemble those calculated with the 6-31G basis set. The *ab initio* potentials were obtained from the *ab initio* MO interaction ener-

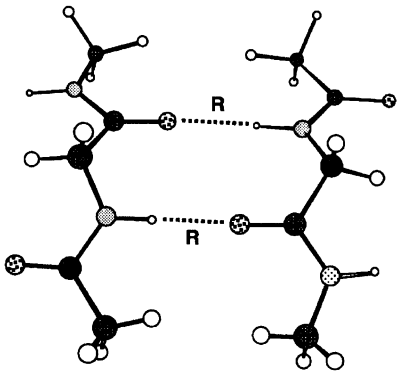

Fig. 8. System III: hydrogen bonding CH_3CO–$NHCH_2CO$–$NHCH_3$ molecules in the conformation of anti-parallel β-sheet. R is the distance between the hydrogen bonding H and O atoms.

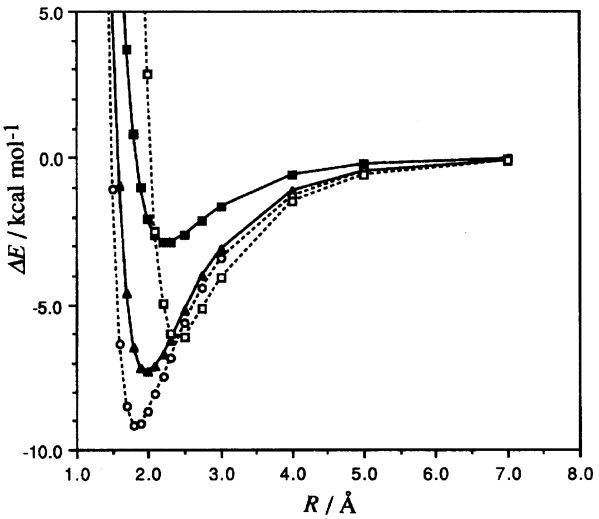

Fig. 9. Plots of the interaction energies (ΔE) of System III calculated with various methods. ▲, *ab initio* MO (7,3/2,1); ■, *ab initio* MO (STO-3G); ○, *ab initio* potential; □, AMBER potential.

gies with the 7,3/2,1 basis set by the fitting.[15] For deriving the *ab initio* potentials, thirty three kinds of model molecules were used, and a total of more than thirty thousand interaction energies were calculated and used for the fitting. It should be stressed that the molecules and conformations which were used in the current work were not included in the course of the derivation of the *ab initio* potentials.

The energy profile calculated with the STO-3G basis set for System III is almost the same as that for System I, in spite of the different number of hydrogen bonds formed in the systems. This implies the incapability of the STO-3G basis set to represent the intermolecular interactions. In contrast with the STO-3G basis set, the energy profile calculated with the 7,3/2,1 basis set for System III is deep and narrow compared with that for System I, in correspondence with the different number of hydrogen bonds formed in the systems. This indicates that the 7,3/2,1 basis set is suitable for calculations of the interaction energies between large molecules.

For System III, the optimum hydrogen bond distance calculated with the AMBER potential is about 0.5 Å longer than that calculated with the 7,3/2,1 basis set. The hydrogen bonding interaction in the form of an anti-parallel β-sheet is stable at a distance (R) of around 2.0 Å, where the calculations with the AMBER potential show a repulsion. The AMBER potential seems to be inadequate for the hydrogen bonding system. The tendency of the hydrogen bonding interaction energies is represented well by the *ab initio* potential.

The bond lengths and bond angles in both of the interacting molecules vary upon hydrogen bond formation. Here, however, they were not taken into consideration, since the purpose of the current work was not to obtain the exact geometries of the hydrogen bonding molecules, but to assess potential energy functions. The geometries of the interacting molecules are kept fixed in order to compare the hydrogen bonding energies calculated with the various levels of methods.

The numerical calculations were carried out on the IBM Powerstation at the National Cancer Center Research Institute, and on the HITAC M680H at the computer center of the Institute for Molecular Science. This work was supported in part by a Grant-in Aid for Scientific Research from the Ministry of Education, Science and Culture.

References

1) L. Schäefer, C. Van Alsenoy, and J. N. Scarsdale, *J. Chem. Phys.*, **76**, 1439 (1982).
2) J. N. Scarsdale, C. Van Alsenoy, V. J. Klimkowski, L. Schäefer, and F. A. Momany, *J. Am. Chem. Soc.*, **105**, 3438 (1983).
3) S. J. Weiner, U. C. Singh, T. J. O'Donnell, and P. A. Kollman, *J. Am. Chem. Soc.*, **106**, 6243 (1984).
4) A. Perczel, J. G. Ángyán, M. Kajtár, W. Viviani, J. -L. Rivail, J. -F. Marcoccia, and I. G. Csizmadia, *J. Am. Chem. Soc.*, **113**, 6256 (1991).
5) For review: P. A. Kollman, "Applications of Electronic Structure Theory," ed by H. F. Schaefer, III, Plenum Press, New York (1977), pp. 109—152.
6) A. Johansson, P. A. Kollman, S. Rothenberg, and J. McKelvey, *J. Am. Chem. Soc.*, **96**, 3794 (1974).
7) J. F. Hinton and R. D. Harpool, *J. Am. Chem. Soc.*, **99**, 349 (1977).
8) A. M. Sapse, L. M. Fugler, and D. Cowburn, *Int. J. Quantum Chem.*, **29**, 1241 (1986).
9) E. Clementi, "Computational Aspects for Large Chemical Systems," Lecture Notes in Chemistry, Springer -Verlag, New York (1980), Vol. 19.
10) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
11) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Am.*

Chem. Soc., **106**, 765 (1984).
12) L. Nilsson and M. Karplus, *J. Comput. Chem.*, **7**, 591 (1986).
13) I. K. Roterman, K. D. Gibson, and H. A. Scheraga, *J. Biomol. Struct. Dyn.*, **7**, 391 (1989).
14) I. K. Roterman, M. H. Lambert, K. D. Gibson, and H. A. Scheraga, *J. Biomol. Struct. Dyn.*, **7**, 421 (1989).
15) M. Aida, G. Corongiu, and E. Clementi, *Int. J. Quantum Chem.*, **42**, 1353 (1992).
16) R. S. Mulliken, *J. Chem. Phys.*, **23**, 1833 (1955).
17) E. Clementi, G. Corongiu, and S. Chakravorty, "Modern Techniques in Computational Chemistry: MOTECC-90," ed by E. Clementi, ESCOM Science Publishers, Leiden (1990), Chap. 7.
18) S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comput. Chem.*, **7**, 230 (1986).
19) M. Dupuis, J. D. Watts, H. O. Viller, and G. J. B. Hurst, HONDO 7, converted from Quantum Chemistry Program Exchange, No. 544 by Mutsumi Aoyagi and Shigeki Obata.
20) M. Dupuis, A. Farazdel, S. P. Karna, and S. A. Maluendes, "Modern Techniques in Computational Chemistry: MOTECC-90," ed by E. Clementi, ESCOM Science Publishers, Leiden (1990), Chap. 6.
21) W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).
22) J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.*, **102**, 939 (1980).
23) W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56**, 2257 (1972).
24) S. F. Boys and F. Bernardi, *J. Mol. Phys.*, **19**, 553 (1970).
25) E. Clementi, G. Corongiu, M. Aida, U. Niesar, and G. Kneller, "Modern Techniques in Computational Chemistry: MOTECC-90," ed by E. Clementi, ESCOM Science Publishers, Leiden (1990), Chap. 17.
26) W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, "*Ab initio* Molecular Orbital Theory," John Wiley & Sons, New York (1986), p. 226.