AN *AB INITIO* STUDY ON THE HYDROGEN BOND ENERGIES OF NUCLEIC ACID BASEPAIRS: ADENINE-URACIL WATSON-CRICK AND GUANINE-URACIL WOBBLE BASEPAIRS

Shun-ichi Kawahara and Tadafumi Uchimaru*

Department of Physical Chemistry, National Institute of Materials and Chemical Research, Agency of Industrial Science and Technology, MITI, Tsukuba Science City 305-8565, Japan. E-mail: t_uchimaru@home.nimc.go.jp

Abstract – The hydrogen bond energies of adenine-uracil Watson-Crick and guanine-uracil wobble basepairs were evaluated using molecular orbital theory. The hydrogen bond energies were found to be almost the same for these two systems, thus uracil can distinguish adenine (A) and guanine (G) only when in combination with the backbone of a nucleotide duplex.

Introduction

Nucleic acid bases form several combinations of basepairs through two or three hydrogen bonds. The formation of Watson-Crick type basepair, which is responsible for formation of duplex of DNA or RNA molecules, is essential for transmission of genetic information.[1] The processes of transcription from DNA to mRNA,[2] and of translation from mRNA to protein *via* tRNA[3] are also based on the formation of Watson-Crick type basepairs.

Meanwhile, several kinds of non-Watson-Crick type basepairs are frequently found in tRNA. The wobble basepair, which is the basepair between guanine (G) and uracil (U) in an unusual manner (G-Uw in Figure 1),[4,5] is the one of the most importance among non-Watson-Crick type basepairs. G-Uw is formed when an anticodon in tRNA accepts a different codon (the degeneracy of genetic code).[6] G-Uw is also found in

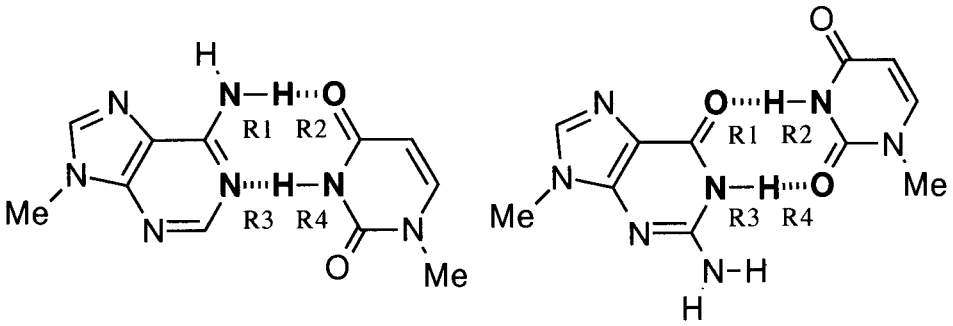

Figure 1. Basepair Structures in This Study

the acceptor stem of yeast $tRNA^{Phe}$. The unique structure of $tRNA^{Phe}$ results from the formation of G-Uw, as well as from that of the other unusual basepairs.[7] Wobble basepairs are also considered as models of mutation.[8]

U forms both Watson-Crick type basepairing with A and wobble type basepairrig with G. It is known that the stability of normal complementary duplex declines upon substitution of G for adenine (A). The destabilization due to this type of substitution is less significant than that caused by another types of substitution in which wobble basepairs cannot be formed.[9] Moreover, poly-U and poly-G form a relatively stable duplex, although the duplex is less stable than a duplex of poly-U and poly-A.[10] It is therefore likely that G-Uw can be formed even in normal nucleic acid duplex, however, the stability of this basepair should be less than A-U.

The differences in behavior between A and G should arise from several factors. The difference of hydrogen bond energies between A-U and G-Uw should be one of the major factors. It is thus essential to evaluate hydrogen bond energies of the A-U and G-Uw basepairs. Duplex stability can be examined experimentally by measuring the melting temperature (Tm). Change in Tm upon substituting a basepair for a different basepair will provide an indication of the difference in hydrogen bond energy between the two basepairs. However, the difference in Tm also includes other factors such as the effect of the strain of the backbone and so it is difficult to directly evaluate the hydrogen bonding stability by these experiments.

On the other hand, quantum mechanical calculations will provide hydrogen bonding energies between nucleic basis. There are many theoretical studies about Watson-Crick type basepairs.[11] However, very few reports regarding non-Watson-Crick type basepairs have appeared in the litarature. In particular, to the best of our knowledge, only one *ab initio* study comparing the hydrogen bond energies A-U and G-

Uw has been reported.[12] Hobza *et al.* reported the hydrogen bond energies of 29 different basepairs,[13] but G-Uw was not included in their study.

We report herein, an *ab initio* study on the basis set effect on hydrogen bond energies of A-U and G-Uw. In addition, we will discuss the difference in the hydrogen bond energy between A-U and G-Uw.

Computational Method

In most theoretical studies on Watson-Crick type basepairs, hydrogen bond energies are evaluated at the second order Møller-Plesset (MP2) level using basis sets with DZP quality. Sponer *et al.*[14] found that the contribution of higher level electron correlation was small on hydrogen bond energy, and that MP2 interaction energies were close to the CCSD(T) data. Hydrogen bond energy is mainly characterized by electrostatic contribution,[15] which can be well reproduced even at the Hartree-Fock (HF) level, and so the contribution of electron correlation should be relatively small. Thus, the conclusion of Sponer *et al.* would be quite reasonable and also be generally applicable to various types of hydrogen bonding systems. Moreover, they also reported that relatively small basis sets such as 6-31G* and ccpVDZ underestimated the hydrogen bond energies as compared with the larger aug-ccpVTZ basis set. However, due to computational resources available, it is difficult to apply a basis set as large as aug-ccpVTZ to MP2 calculations of chemical species like nucleic acid basepairs.[16]

Considering the findings of Sponer *et al.*,[14] we evaluated the hydrogen bond energies of A-U and G-Uw at the MP2 theoretial level. However, we were unable to employ a basis set as large as aug-ccpVTZ. Thus, we systematically examined the effects of the basis set utilized. Starting from the basis sets 6-31G* and 6-311G*, we added successive polarization functions and diffuse functions to the basis sets. We carried out calculations on 12 types of basis set. The largest basis set utilized in the present study was 6-311++G**. The number of basis functions utilized was 531 and 553 for A-U and G-Uw, respectively.

The structure of A-U and G-Uw basepair, as well as nucleic acid bases A, G, and U were optimized at the HF theoretical level using each basis set. In all cases, *Cs* symmetry was observed: all atoms, except for two hydrogen atoms in the methyl group, were placed on the plane of symmetry. In addition, the energy of the optimized structures was also evaluated with single-point calculations at the MP2 theoretical level. The basis set super position error (BSSE) for hydrogen bond energies (ΔE^{HB}) was evaluated by using the counterpoise method.[17] The atom distance between the oxygen/nitrogen atom and hydrogen atom

constructing each hydrogen bond is named as R1 – R4 (see Figure 1). These atom distances were most influenced by the basis set.

Result and Discussion

A-U Watson-Crick Basepair

ΔE^{HB}(A-U) and BSSE are shown in Figure 2 and Table 1. HF and MP2 values for ΔE^{HB}(A-U) range from -9.65 to -9.10 kcal•mol^{-1} and from -12.66 to -11.87 kcal•mol^{-1}, respectively. These ΔE^{HB}(A-U) values at the MP2 level well reproduced the results of Raben *et al.*[18] At the HF level, the basis set effect on ΔE^{HB}(A-U) was small: the variation was 0.55 kcal•mol^{-1} at the maximum. The values for ΔE^{HB}(A-U) calculated using double-ζ basis sets of 6-31+G* and 6-31++G* and using triple-ζ basis sets from 6-311G* to 6-311++G** were almost the same. Relatively large BSSE values were observed for basis sets of 6-31G*, 6-31G**, 6-311G* and 6-311G**. However, BSSE was reduced effectively upon adding a single set of diffuse functions on heavy atoms. At the MP2 level, the values for both ΔE^{HB}(A-U) and BSSE were larger by 2-3 kcal•mol^{-1} than the HF results, as shown in Figure 2 MP2. The fluctuation of ΔE^{HB}(A-U) was slightly larger (0.79 kcal•mol^{-1} at maximum) than at the HF level. 6-31G based double-ζ

Table 1. ΔE and BSSE (kcal•mol^{-1})

	HF				MP2			
	A-U		G-Uw		A-U		G-Uw	
Basis Set	ΔE	BSSE	ΔE	BSSE	ΔE	BSSE	ΔE	BSSE
6-31G*	-9.44	2.42	-11.44	2.22	-12.57	5.20	-13.60	4.77
6-31G**	-9.65	2.25	-11.60	2.06	-12.67	4.96	-13.65	4.56
6-31+G*	-9.13	1.10	-11.30	1.17	-12.53	3.23	-13.82	3.47
6-31++G*	-9.13	1.12	-11.32	1.19	-12.56	3.36	-13.85	3.61
6-31+G**	-9.38	0.84	-11.49	0.90	-12.64	2.81	-13.88	3.02
6-31++G**	-9.37	0.86	-11.49	0.92	-12.66	2.93	-13.90	3.13
6-311G*	-9.12	2.05	-11.10	1.81	-11.87	4.18	-12.55	3.75
6-311G**	-9.19	2.04	-11.09	1.84	-12.02	4.17	-12.58	3.80
6-311+G*	-9.10	0.81	-11.18	0.87	-12.15	2.53	-13.05	2.88
6-311++G*	-9.10	0.81	-11.18	0.89	-12.16	2.61	-13.07	3.00
6-311+G**	-9.18	0.74	-11.19	0.80	-12.30	2.41	-13.06	2.74
6-311++G**	-9.18	0.74	-11.19	0.81	-12.31	2.48	-13.07	2.84

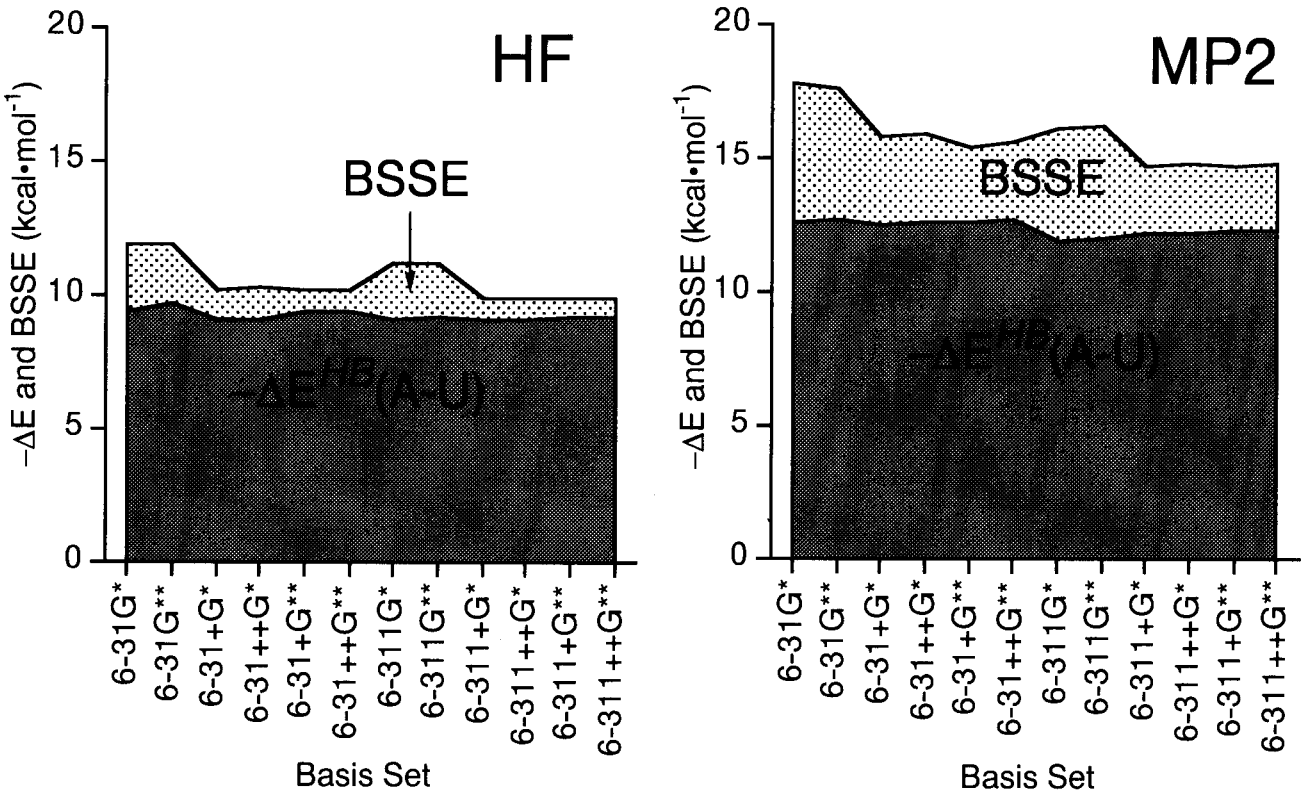

Figure 2. $-\Delta E^{HB}$(A-U) and BSSE at the HF and MP2 Level

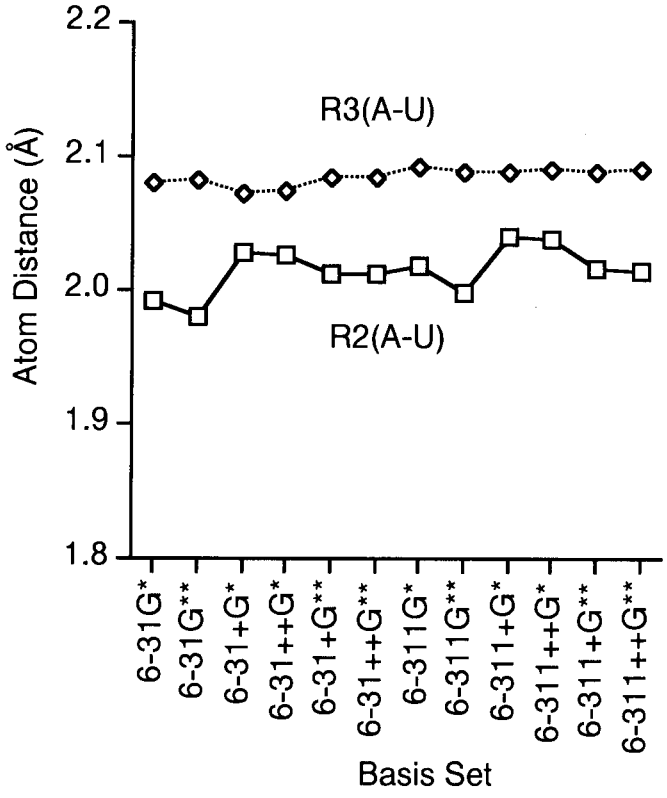

Figure 3. Atom Distances of Hydrogen Bonds

Table 2. Atom Distances Neighboring Hydrogen Bond

Basis Set	R1	R2	R3	R4
6-31G*	1.0006	2.0794	1.9925	1.0136
6-31G**	0.9990	2.0822	1.9808	1.0134
6-31+G*	1.0011	2.0729	2.0275	1.0134
6-31++G*	1.0011	2.0734	2.0263	1.0134
6-31+G**	0.9994	2.0838	2.0125	1.0129
6-31++G**	0.9994	2.0849	2.0115	1.0129
6-311G*	0.9958	2.0922	2.0178	1.0080
6-311G**	0.9982	2.0883	1.9982	1.0119
6-311+G*	0.9966	2.0881	2.0391	1.0082
6-311++G*	0.9966	2.0891	2.0388	1.0082
6-311+G**	0.9986	2.0884	2.0158	1.0118
6-311++G**	0.9986	2.0891	2.0143	1.0119

basis sets gave slightly higher estimates for ΔE^{HB}(A-U) than 6-311G based triple-ζ basis sets. On the other hand, we added the polarization functions and/or diffuse functions to 6-311G based triple-ζ basis set, which resulted in a slight increase in ΔE^{HB}(A-U). In consequence of these two opposing effects, ΔE^{HB}(A-U) observed using 6-31+G* was close to the result using 6-311++G** (+0.22 kcal•mol^{-1}, 1.8% larger than the result of 6-311++G**). Similar to the HF results, when basis sets without diffuse functions on heavy atoms were used, BSSE was large, however, BSSE could be reduced by using diffuse functions. Nonetheless, the influence of BSSE at the MP2 level (5 – 6.5 kcal•mol^{-1}) was greater than at the HF level.

Figure 3 and Table 2 show the atom distances of neighboring hydrogen bonds (R1 – R4 in Figure 1). As compared with other atom distances, R3 (hydrogen bond length between imide proton of U and N^1 of A) showed a greater degree of variation (0.06 Å at maximum) depending on whether or not the polarization functions on hydrogen atoms were set. There was a relationship between the differences of R3 and the differences of ΔE^{HB}(A-U), however, the effect was negligible because ΔE^{HB}(A-U) varied only slightly depending on the polarization functions on the hydrogen atoms.

G-U Wobble Basepair

Figure 4 and Table 1 show ΔE^{HB}(G-Uw) and BSSE. HF and MP2 values for ΔE^{HB}(G-Uw) range from -11.60 to -11.10 kcal•mol^{-1} and from -13.90 to -12.55 kcal•mol^{-1}, respectively. Although the tendencies of the ΔE and the BSSE were very similar to those observed for A-U basepair, the fluctuation of ΔE^{HB}(G-Uw) (depending on the basis set) was larger (1.35 kcal•mol^{-1} at maximum) than ΔE^{HB}(A-U) at the MP2 level. 6-31G based double-ζ basis sets estimated ΔE^{HB}(G-Uw) larger than 6-311G based triple-ζ basis sets. ΔE^{HB}(G-Uw) showed the same tendency as ΔE^{HB}(A-U): a slight increase in ΔE^{HB}(G-Uw) was observed upon setting the polarization functions and/or diffuse functions to heavy atoms and/or hydrogen atoms. ΔE^{HB}(G-Uw) was calculated to be almost constant when 6-311+G* or a larger basis set was employed.

Figure 5 and Table 3 show the atom distances of neighboring hydrogen bonds (R1 – R4 in Figure 1). The fluctutaion in R1 and R4 depends mainly on whether or not the polarization functions on hydrogens were set. Larger fluctuation were observed in the 6-311G based triple-ζ basis sets than for the double-ζ basis sets. Calculated values for ΔE^{HB}(G-Uw) at the MP2 level were not sensitive to fluctuations in R1 – R4. Thus, we can discuss hydrogen bond energy of G-Uw by using MP2 level energies of structures optimized at the HF level.

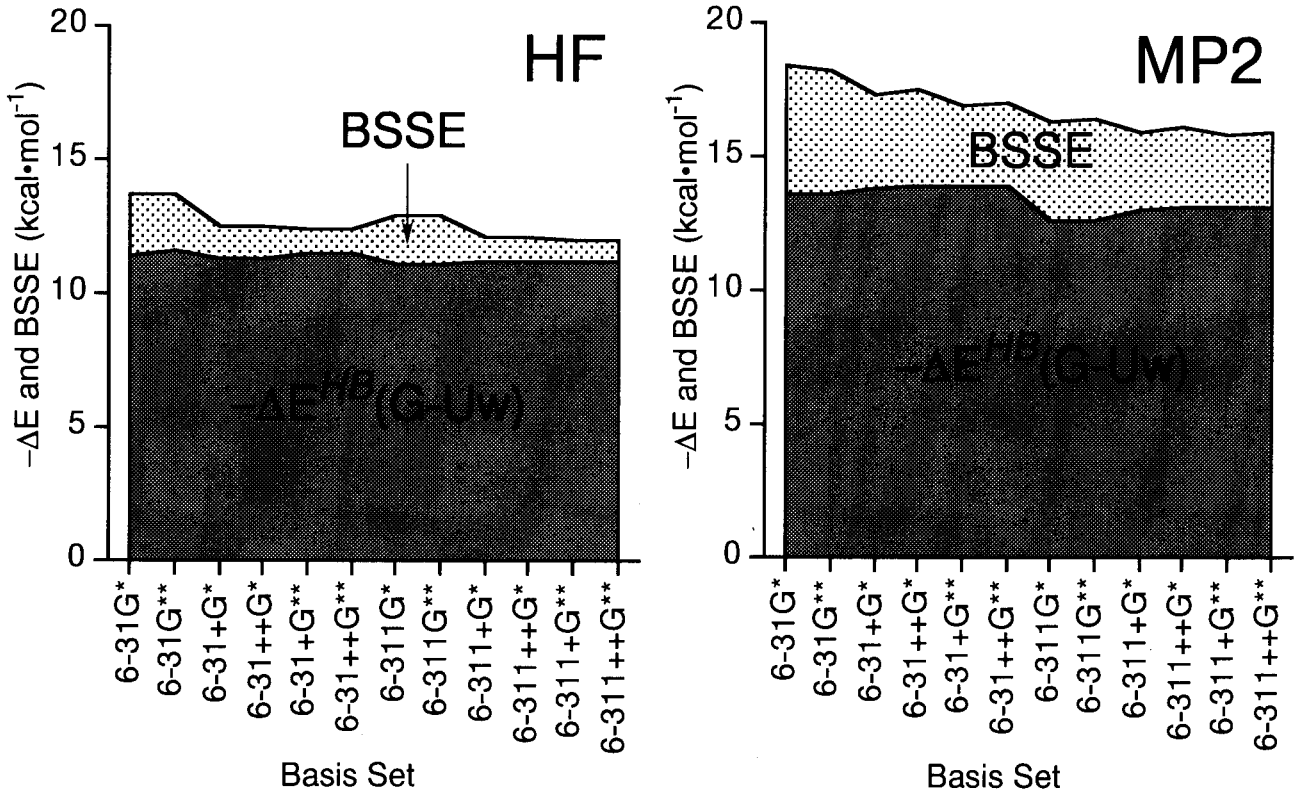

Figure 4. $-\Delta E^{HB}$(G-Uw) and BSSE at the HF and MP2 Level

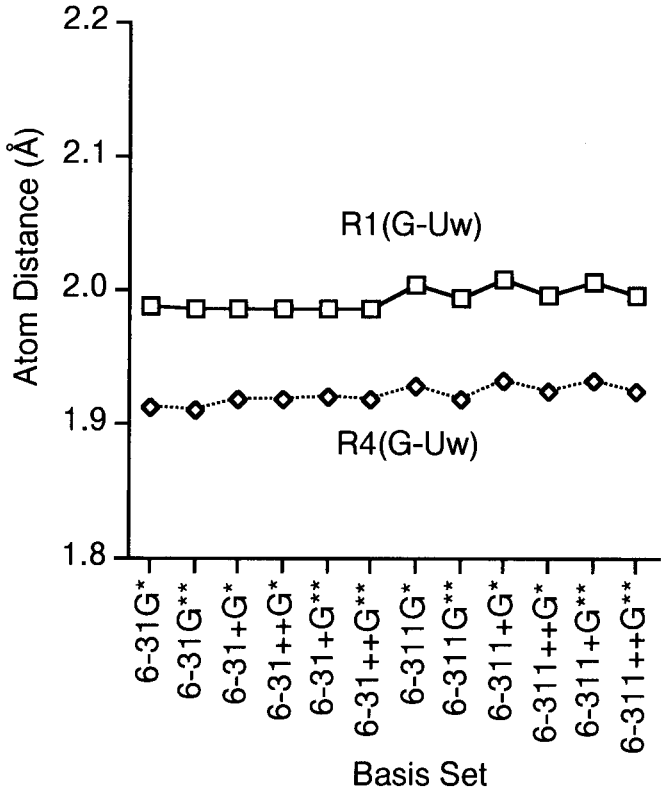

Figure 5. Atom Distances of Hydrogen Bonds

Table 3. Atom Distances Neighboring Hydrogen Bond

Basis Set	R1	R2	R3	R4
6-31G*	1.9871	1.0051	1.0115	1.9116
6-31G**	1.9855	1.0042	1.0108	1.9098
6-31+G*	1.9863	1.0123	1.0059	1.9188
6-31++G*	1.9862	1.0122	1.0059	1.9185
6-31+G**	1.9867	1.0111	1.0048	1.9190
6-31++G**	1.9866	1.0111	1.0048	1.9186
6-311G*	2.0040	1.0060	1.0002	1.9273
6-311G**	1.9937	1.0094	1.0033	1.9182
6-311+G*	2.0081	1.0067	1.0011	1.9328
6-311++G*	1.9960	1.0097	1.0038	1.9236
6-311+G**	2.0056	1.0068	1.0038	1.9321
6-311++G**	1.9954	1.0097	1.0038	1.9237

Hydrogen Bond Formation Between U and A or U and G

Regardless of computational level utilized in the present work, ΔE^{HB}(G-Uw) was found to be slightly more negative (0.5 – 1.3 kcal•mol^{-1}, 0.76 kcal•mol^{-1} in MP2/6-311++G**//HF//6-311++G**) than ΔE^{HB}(A-U). This is in accord with the fact that both hydrogen bond lengths of G-Uw (R1 and R4) are shorter than those of A-U (R2 and R3). However, the difference between the calculated values for ΔE^{HB}(A-U) and ΔE^{HB}(G-Uw) was less than 1.3 kcal•mol^{-1}. We cannot discuss which basepair (A-U or G-Uw) is the more stable, however, we can safely conclude that the hydrogen bond stability of G-Uw should be almost the same as that of A-U.

Our results indicate that ΔE^{HB}(A-U) and ΔE^{HB}(G-Uw) are comparable. Thus, the flexible backbone structure of the anticodon loop of tRNA allows the formation of both A-U and G-Uw at the 5' end of the anticodon. The substitution of G for A in normal complementary nucleic acid duplex causes decline in duplex stability,[9] despite the result that U can form equally stable hydrogen bonds with G and A. This decline in stability should be due to the strain of the backbone accompanied by the formation of a wobble basepair. Our results reconfirm the importance of the backbone, which should orientate the bases at the positions suitable for Watson-Crick type basepair formation.[19]

Recently, molecular recognition ability based on Watson-Crick type basepair formation has been applied to supermolecular formation and template synthesis.[20] However, outcome of this study indicates that U can distinguish A and G only in the case when in combination with the backbone, which should be considered in the design of an artificial supermolecular system.

Conclusion

1) ΔE^{HB}(A-U) and ΔE^{HB}(G-Uw) are almost converged when 6-311+G* or larger basis sets are used.

2) ΔE^{HB}(A-U) and ΔE^{HB}(G-Uw) are converge better in the case when BSSE is corrected, therefore BSSE should be corrected.

3) ΔE^{HB}(A-U) and ΔE^{HB}(G-Uw) are almost the same. The reason why U can distinguish A and G is attributable to the backbone supporting the bases in the correct position.

ACKNOWLEDGMENT

The services and computational time made available by Tsukuba Advanced Computing Center (TACC) have been essential to this study and are gratefully acknowledged. We also thank the Science and Techno-

logy Agency for their financial support.

REFERENCES AND NOTES

1. J. D. Watson and H. C. Crick, *Nature*, 1953, **171**, 757.

2. S. Brenner, F. Jacob, and M. Meselson, *Nature*, 1961, **190**, 576.

3. J. J. Hopfield, *Proc. Nat. Acad. Sci. USA*, 1974, **71**, 4135.

4. There are another types wobble basepair between inosine (I)-A. I-U and I-C. However, I is special base found in only tRNA. J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner, 'Moleculaer Biology of the Gene', Benjamin/Cummings Pub. Co. Inc. pp. 438-440.

5. Hereafter, we refer to A-U Watson-Crick type basepair and G-U wobble type basepair just simply as A-U and G-Uw.

6. Spring Harber *Symp. Quant. Biol.*, 1966, XXXI.

7. a) J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church, and S. -H. Kim, *J. Mol. Biol.*, 1978, **123**, 607; b) S. -H. Kim, *Prog. Nucl. Acid. Res. Mol. Biol.*, 1976, **17**, 181.

8. a) F. H. C. Crick, *J. Mol. Biol.*, 1966, **19**, 548; b) D. Söll, J. D. Cherayil, and R. M. Bock, *J. Mol. Biol.*, 1967, **29**, 97; c) D. A. Kellogg, B. P. Doctor, J. E. Loebel, and M. W. Nirenberg, *Proc. Nat. Acad. Sci. USA*, 1966, **55**, 912; d) A. J. Lomant and J. R. Fresco, *Prog. Nucl. Acid. Res. Mol. Biol.*, 1975, **15**, 185.

9. P. J. Romaniuk, D. W. Hughes, R. J. Gregoire, T. Neilson, and R. A. Bell, *J. Chem. Soc., Chem. Commun.*, 1979, 559.

10. A. G. Lezius and E. Domin, *Nature New Biol.*,1973, **244**, 169.

11. a) S. Scheiner and C. W. Kern, *Chem. Phys. Lett.*, 1978, **57**, 331; b) W. B. Person, K. Szczepaniak, M. Szczesniak, J. S. Kwiatkowski, L. H. Czerminski, and R. Czerminski, *J. Mol. Struct.*, 1989, **194**, 239; c) A. Le´s and L. Adamowicz, *J. Phys. Chem.*, 1989, **93**, 7078; d) R. Czerminski, K. Szczepaniak, W. B. Person, and J. S. Kwiatkowski, *J. Mol. Struct.*, 1990, **237**, 151; e) A. Le´s and L. Adamowicz, *J. Am. Chem. Soc.*, 1990, **112**, 1504; f) M. Sabio, S. Topiol, and S. W. C. Lumma Jr., *J. Phys. Chem.*, 1990, **94**, 1366; g) A. R. Katritzky and M. Szafran, *J. Chem. Soc., Perkin Trans. 2*, 1990, 871; h) A. R. Katritzky and M. Karelson, *J. Am. Chem. Soc.*, 1991, **113**, 1561; i) J. Leszczy´nski and K. Lammertsma, *J. Phys. Chem.*, 1991, **95**, 3128; j) K. I. Trollope, I. R. Gould, and I. H. Hiller, *Chem. Phys. Lett.*, 1993, **209**, 113; k) I. R. Gould and P.

A. Kollman, *J. Am. Chem. Soc.*, 1994, **116**, 2493; l) A. Destexhe, J. Smets, L. Adamowicz, and G. Maes, *J. Phys. Chem.*, 1994, **98**, 1506; m) P. Hobza, J. Sponer, and M. Polásek, *J. Am. Chem. Soc.*, 1995, **117**, 792; n) J. Sponer, J. Leszczynski, and P. Hobza, *J. Biomol. Strct. Dyn.*, 1996, **14**, 117; o) J. Sponer and P. Hobza *Chem. Phys.*, 1996, **204**, 365.

12. R. Czerminski, K. Szczepaniak, W. B. Person, and J. S. Kwiatkowski, *J. Mol. Struct.*, 1990, **237**, 151.
13. P. Hobza and C. Sandorfy, *J. Am. Chem. Soc.*, 1987, **109**, 1302.
14. J. Sponer and P. Hobza, *Chem. Phys.Lett.*, 1997, **267**, 263.
15. a) K. Morokuma, *Acc. Chem. Res.*, 1977, **10**, 294; b) A. P. L. Rendell, G. B. Bacskay, and N. S. Hush, *Chem. Phys. Lett.*,1985, **117**, 400; c) W. Chen and M. S. Gordon, *J. Phys. Chem.*, 1996, **100**, 14316.
16. For example, nubmer of basis to calculate A-U in aug-ccpVTZ is 1219.
17. S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
18. P. R. Raben, J. W. Lockman, and W. L. Jorgensen, *J. Phys. Chem. A*, 1998, **102**, 3782.
19. W. Fuller and A. Hodgson, *Nature*, 1976, **263**, 289.
20. a) J. L. Sessler and R. Wang, *Angew. Chem., Int. Ed. Engl.*, 1998, **37**, 1726; b) J. J. González, P. Prados, and J. de Mendoza, *Angew. Chem., Int. Ed. Engl.*, 1999, **38**, 525; c) T. Tjivikua, P. Ballester, and J. Rebek, Jr., *J. Am . Chem. Soc.*, 1990, **112**, 1249; d) J. S. Nowick, Q. Feng, T. Tjivikua, P. Ballester, and J. Rebek, Jr., *J. Am. Chem. Soc.*, 1991, **113**, 8831; e) F. M. Menger, A. V. Eliseev, N. A. Khanjin, and M. J. Sherrod, *J. Org. Chem.*, 1995, **60**, 2870.

Received, 2nd August, 1999