Pergamon

Tetrahedron Letters, Vol. 38, No. 48, pp. 8337-8340, 1997

0040-4039/97 $17.00 + 0.00

PII: S0040-4039(97)10257-X

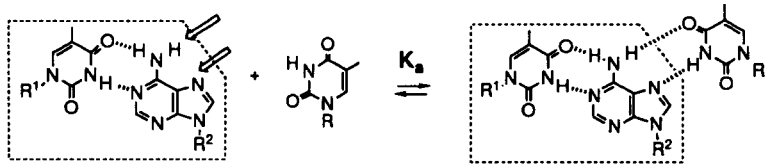
Synthesis of a AT Base Pair Model in DNA and Determination of Hydrogen Bonding Strength on the Formation of Base Triplet T:AT in $CDCl_3$.

Young Lag Cho and Kyu-Sung Jeong*

Department of Chemistry, Yonsei University
Seoul 120-749, Korea

Abstract: A molecular model has been synthesized to determine the hydrogen-bonding strength between AT and T in the base triplet, T:AT. A strong preference (>95%) of Watson-Crick mode along with 91% hydrogen-bonded conformation is observed in the model compound. The association constants of the AT model with 1-propyluracil and glutarimide are 15 M^{-1} and 5.2 M^{-1}, respectively, in $CDCl_3$ at 296 ± 0.5 K. Negative cooperativity by pre-existing hydrogen bonds might be involved in the intermolecular binding events. © 1997 Elsevier Science Ltd.

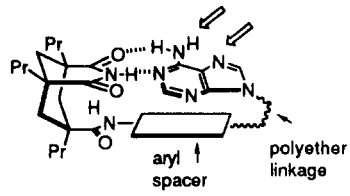
Specific recognition of double-stranded nucleic acids by the formation of base triplets such as T:AT and C^+:GC has been extensively studied over many years.[1] The triple helices are formed by either Hoogsteen or reverse Hoogsteen hydrogen bonds in the major groove of the double helical DNA. Even though various molecular models have been reported to evaluate intermolecular forces of base pair or triplet formations,[2] little is known the model compound[3] which can determine the strength of hydrogen bonds upon forming base triplets between a base pair (AT or GC) and a pyrimidine base (T, U, or C^+). The strength can not be directly measured by merely mixing of adenine with thymine or guanine with cytosine analogue in a nonpolar solvent because predominantly bimolecular 1:1 complexes are formed. Moreover, in a mixture solution of adenine and thymine derivatives, four types of 1:1 complexes, Watson-Crick, reverse Watson-Crick, Hoogsteen, and reverse Hoogsteen hydrogen-bonding modes, exist in a rapid equilibrium without strong preference of any mode.



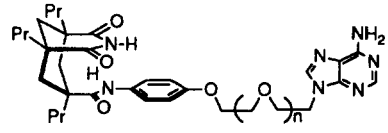
AT base pair **T** **T:AT base triplet**

We report here a model compound of Watson-Crick AT base pair, like in DNA, to determine the hydrogen-bonding strength between AT and T in the base triplet, T:AT. In order to serve as an elegant AT base pair model in DNA, the model compound must fulfill the following two criteria; i) It must exist in a completely hydrogen-bonded state between A and T moieties. ii) The Watson-Crick hydrogen bonds must occur. With these in mind, we designed the model system as shown below, in which a Kemp's acid

derivative with imide functionality[4] is intramolecularly connected with adenine. Intramolecular connection of two hydrogen-bonding partners reduces the translational entropic penalty and thus increases the population of the hydrogen-bonded state. Furthermore, preference of Watson-Crick hydrogen bonds could be controlled by varying the linkage and aryl surface in the model system.



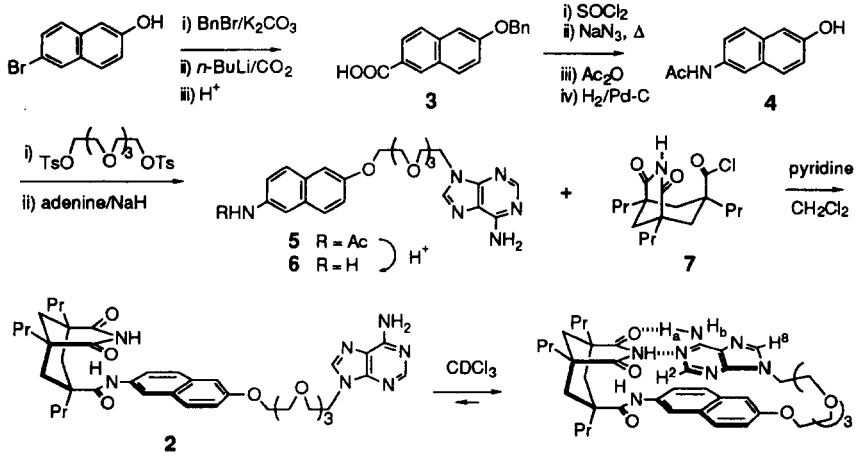
AT base pair model



1 n= 3

A series of podand hosts were previously prepared for studies of preorganization effects on binding of alkali metal cations.[5] Maximum intramolecular hydrogen bonding between imide and adenine moieties was found to occur in the tetraethylene glycol-linked host **1** which showed the ratio 82:18 of Watson-Crick and Hoogsteen mode with 86% cyclic form.[6] In order to enhance both intramolecular hydrogen bonding and Watson-Crick preference, we prepared a new model **2** with naphthalene spacer. Synthesis of **2** is outlined in Scheme 1. Commercially available 6-bromo-2-naphthol was reacted with benzyl bromide/K_2CO_3, then *n*-BuLi/CO_2 and H^+ to give 6-(benzyloxy)-2-naphthoic acid (**3**) in 63% yield. After treatment of the acid **3** with $SOCl_2$, Curtius rearrangement (activated NaN_3 in toluene, reflux)[7], followed by acetylation (Ac_2O/CH_2Cl_2-1N NaOH) and hydrogenolysis (H_2/Pd-C) afforded 6-(acetylamino)-2-naphthol (**4**) in 43% three-step yield. Sequential coupling of tetraethylene glycol ditosylate with **4** (K_2CO_3/DMF, 29%) and adenine (NaH/DMF, 55%) gave the compound **5**. Finally, after deactylation in 1N HCl/H_2O-MeOH , the resulting amine **6** was reacted with acid chloride **7** to give the desired product **2** (44%).[8]

Scheme 1

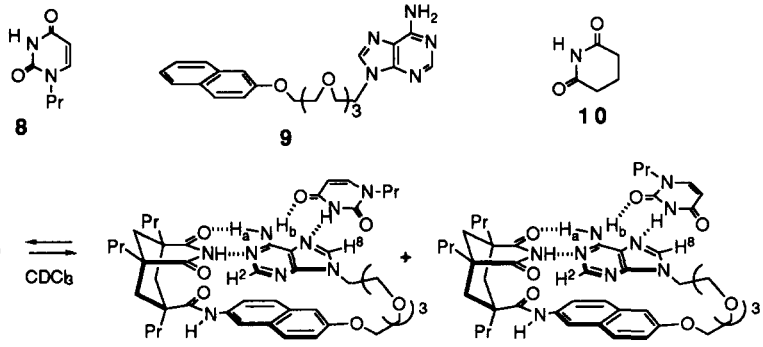


2

Structural features of the model compound **2** were characterized by ^1H NMR spectroscopy. The concentration-independent ^1H NMR signal for the imide NH in **2** was appeared 13.21 ppm at 23 °C in $CDCl_3$, which are highly downfield-shifted compared to that (7.42 ppm) for hydrogen-bonding free imide NH in the analogous molecule.[4b, 9] In addition, H_2 and H_8 signals of the adenine ring in **2** are upfield-shifted

($\Delta\delta$ ~ 0.3 ppm) compared with those in a reference molecule **9**, indicating that stacking occurs between naphthalene and adenine rings. The chemical shift (13.76 ppm at -40 °C) of completely hydrogen-bonded imide NH was obtained by cooling a 2mM solution of **2** in $CDCl_3$ until no further downfield shift was observed. From these limiting values, % cyclic form of **2** as the result of the intramolecular hydrogen bonding is estimated to be 91% at 23 °C in $CDCl_3$. The relative population of Watson-Crick *versus* Hoogsteen type hydrogen bonds was determined by homonuclear NOE difference experiments. Irradiation of the imide NH in **2** resulted in 10.5% and < 0.5% enhancements, respectively, in the H_2 and H_8 of the adenine ring, indicating that Watson-Crick mode is strongly predominant (>95%).

The binding affinities in this study were determined in $CDCl_3$ at 296 ± 0.5 K by at least duplicated ^1H NMR titrations, in which the host concentration was gradually increased from 0 to 50.0 mM while the guest concentration was kept constant (5.00 mM). During the titrations, the guest NH signals were gradually downfield-shifted ($\Delta\delta_{max}$ > 2 ppm) while the imide NH, H_2 and H_8 signals in **2** remained unchanged. Furthermore, the amino ($-NH_2$) proton signal of the adenine ring in **2** is split into two broad singlets by cooling up to - 50 °C; one (H_a) signal appears at 6.91 ppm and the other (H_b) at 4.98 ppm. At the same temperature, addition of 1-propyluracil (**8**, ~ 1 equiv) causes a large downfield shift of the H_b (4.98 to 5.51ppm) but a negligible change ($\Delta\delta$ < 0.03 ppm) in the H_a signal. Finally, in the 1:1 mixture of the host **2** and 1-propyluracil (**8**), irradiation of imide NH in **2** gave 8.6% and < 0.5% enhancements, and irradiation of uracil NH signal gave 0.34% and 0.73% enhancements, respectively, in H_2 and H_8 of adenine ring.[10] All of these observations indicate that intramolecular Watson-Crick hydrogen bonding is preserved and complexation occurs largely via intermolecular Hoogsteen or reverse Hoogsteen hydrogen bonding as shown below. The 1:1 complex formations of the hosts **1** and **2** with 1-propyluracil (**8**) was confirmed by Job's plots[11] which showed maximum complexations at molar ratio 0.5.



The association constants (K_a ± 5%) of a $CDCl_3$-soluble guest **8** with the hosts **1** and **2** are 19 M^{-1} and 17 M^{-1}, respectively, which were obtained by nonlinear squares fitting method of saturation curves of the guest NH signals. Since the hosts **1** and **2** exist in 86% and 91% cyclic forms, respectively, with a strong preference of Watson-Crick hydrogen bond (82% in **1** and >95% in **2**), the association constant, $K_a(c)$, of pure cyclic hosts with guest through intermolecular Hoogsteen hydrogen bonds could be estimated by the following equations:

$$\Delta G^o(obsd) = f(ac) \times \Delta G^o(ac) + f(c) \times \Delta G^o(c), \text{ and thus}$$

$$\text{Log } K_a(obsd) = f(ac) \times \text{Log } K_a(ac) + f(c) \times \text{Log } K_a(c)$$

where f(ac) and f(c) stand for fractions of acyclic and cyclic forms of the hosts, and $K_a(obsd)$ for observed K_a from titrations, $K_a(ac)$ and $K_a(c)$ for K_as of acyclic and cyclic forms. The value (68 M^{-1}) of $K_a(ac)$ is

obtained from titration of 1-propyluracil (**8**) with a reference molecule **9**. From equations described above, $K_a(c) = 15 \pm 0.3\ M^{-1}$ which is ~ one fourth or fifth of $K_a(ac) = 68\ M^{-1}$. The same trend was observed when glutarimide (**10**) was employed as a guest. The association constants of **10** with the host **2** and the reference **9** are 6.0 M^{-1} and 24 M^{-1}, respectively, and thus $K_a(c) = 5.2\ M^{-1}$. Previous studies [2a,12] have demonstrated that the strengths of Watson-Crick and Hoogsteen hydrogen bonds are almost equal in the bimolecular 1:1 complexes between adenine and uracil analogues. The calculated $K_a(c)$ values (15 M^{-1} and 5.2 M^{-1}) are smaller than the expected values (34 M^{-1} for **8** and 12 M^{-1} for **10**) by statistical correction. Any of steric strains between host and guest residues could not be noticeable from CPK molecular modelings. Therefore, these lower values suggest that pre-existing Watson-Crick hydrogen bonding might exert a mild negative cooperativity[13] on the intermolecular Hoogsteen hydrogen-bonding formations.

In conclusion, a model compound for AT base pair in DNA has been synthesized by intramolecular connection of Kemp's acid derivative with adenine, and the hydrogen bonding strength between AT base pair and T(U) nucleobase has been for the first time determined using simple molecular model.

Acknowledgment This work was financially supported in part by the Korea Science and Engineering Foundation (Project No, 96-0501-04-01-03) and by the Organic Chemistry Research Center-KOSEF.

References and Notes

1. For reviews see: a) Moser, H. E.; Dervan, P.B. *Science* **1987**, *238*, 645-650. b) Thuong, N. T.; Hélène, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666-690. c) Sun, J.-s.; Garestier, T.; Hélène, C. *Curr. Opin. Struct. Biol.* **1996**, *6*, 327-333. d) Doronina, S. O.; Behr, J.-P. *Chem. Soc, Rev.* **1997**, 63-71. e) Nielsen, P. E.; Haaima, G. *Chem. Soc, Rev.* **1997**, 73-70.
2. a) Askew, B.; Ballester, P.; Buhr, C.; Jeong, K.-S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1989**, *111*, 1882-1090. b) Adrian, J. C., Jr.; Wilcox, C. S. *J. Am. Chem. Soc.* **1989**, *111*, 8055-8057. c) Hamilton, A. D. In *Inclusion Phenomena and Molecular Recognition*; Atwood, J. L., Ed.; Plenum: New York, 1990; pp 57-64. d) Park, T. K.; Schroder, J.; Rebek, J., Jr. *Tetrahedron* **1991**, *47*, 2507-2518. e) Kuroda, Y.; Lintuluoto, J. M.; Ogoshi, H. *Tetrahedron Lett.* **1994**, *35*, 3729-3732. f) Lonergan, D. G.; Riego, J.; Deslongchamps, G. *tetrahedron Lett.* **1996**, *37*, 6109-6112.
3. a) Hisatome, M.; Ikeda, K.; Kishibata, S.; Yamakawa, K. *Chem. Lett.* **1993**, 1357-1360. b) Sasaki, S.; Nakashima, S.; Nagatsugi, F.; Tanaka, Y.; Hisatome, M.; Maeda, M. *Tetrahedron Lett.* **1995**, *36*, 9521-9524.
4. a) Jeong, K.-S.; Muehldorf, A. V.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 6144-6145. b) Jeong, K.-S.; Tjivikua, T.; Muehldorf, A. V.; Deslongchamps, G.; Famulok, M; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 201-209.
5. Jeong, K.-S; Cho, Y. L.; Pyun, S. Y. *Tetrahedron Lett.* **1995**, *36*, 2827-2830.
6. The chemical shift of the imide NH in **1** is shown at 12.90 ppm. The compound **1** was reported[5] to exist in 95% cyclic form which value was estimated from two limiting values of 7.6 ppm for free NH and 13.2 ppm for 100% hydrogen-bonded NH with adenine moiety. The value of 86% is here recalculated by employing the new limiting values of 7.42 and 13.76 ppm.
7. Smith, P. A. S. In *Organic Reactions*; Adams, R., Ed.; John Wiley & Sons: New York, 1959; Vol. 3. Chapter 9.
8. Physical properties of **2**:: Mp 108-110 °C; IR (thin film) 739, 1121, 1232, 1467, 1540, 1606, 1651, 1702, 2872, 2953, 3204, 3329 cm^{-1}; ^1H NMR (250 MHz, $CDCl_3$), δ 0.86 (t, 3H, J = 7.0 Hz), 0.97 (t, 6H, J = 6.9 Hz), 1.31-1.55 (m, 13H), 1.97-2.07 (m, 2H), 2.28 (d, 1H, J = 13.3 Hz), 2.64 (d, 2H, J = 14.5 Hz), 3.66-3.78 (m, 10H), 3.90-3.95 (m, 2H), 4.14-4.20 (m, 4H), 5.92 (s, br, 2H, NH_2-adenine) 6.69 (dd, 1H, J = 8.8, 2.3 Hz), 6.72 (d, 1H, J = 2.3 Hz), 7.06 (dd, 1H, J = 8.7, 2.1 Hz), 7.12 (d, 1H, J = 8.8 Hz), 7.24 (s, 1H, amide NH), 7.35 (d, 1H, J = 8.7 Hz), 7.77 (s, 1H, H^8-adenine), 8.03 (s, 1H, H^2-adenine), 8.08 (d, 1H, J = 2.1 Hz), 13.21 (s, 1H, imide NH); Anal. Calcd for $C_{41}H_{55}N_7O_7$: C, 64.97; H, 7.31; N, 12.94. Found: C, 64.98; H, 7.32; N, 12.87.
9. Jeong, K.-S.; Tjivikua, T.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 3215-3217.
10. The NOE difference experiment was performed at 0 °C and a little proton exchanges were observed between the host imide NH, uracil NH, and a trace of water. The same experiment using a reference compound **9** and 9-propyluracil (**8**) gave 3.0% and 2.2% enhancements in H_2 and H_8 of the adenine ring, respectively.
11. Connors, K. A. *Binding Constant*; John Wiley & Sons: New York, 1984; p 24.
12. a) Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1967**, *57* 250-257. b) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983; pp 126-131.
13. For discussions on the cooperativity in hydrogen bonding, see: Jeffrey, G, A.; Saenger, W. Hydrogen Bonding in Biological Structures; Springer-Verlag: New-York, 1991; pp 36-39.

(*Received in Japan* 28 *July* 1997; *revised* 24 *September* 1997; *accepted* 26 *September* 1997)