ORGANIC LETTERS

2000
Vol. 2, No. 19
2931−2934

# Synthetic Receptors for CG Base Pairs

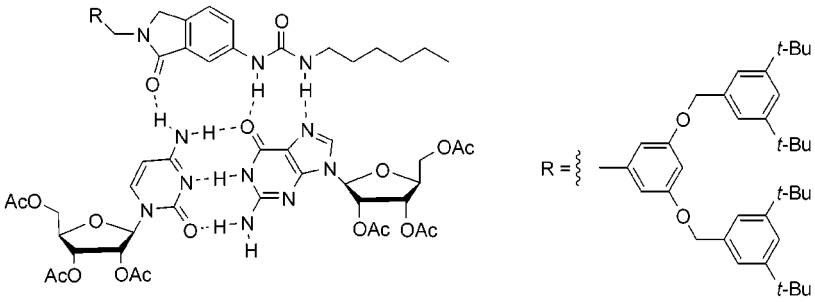
**Eric Mertz, Sebastiano Mattei, and Steven C. Zimmerman***

*Department of Chemistry, 600 South Mathews Avenue, University of Illinois, Urbana, Illinois 61801*

*sczimmer@uiuc.edu*

**Received June 4, 2000**

## ABSTRACT




**Hydrogen-bond-mediated complexation of a CG base pair by a hexylureido phthalimide and a hexylureido isoindolin-1-one was studied by $^1$H NMR spectroscopy in an organic solvent. Chemical shift data indicate that both receptors effectively bind the CG base pair from the major groove side.**

The numerous examples of biomolecular recognition mediated through hydrogen bonds have inspired the design of nonnatural hydrogen-bonding systems capable of functioning as receptors for molecules and assemblies of biological interest. Because DNA and RNA contain multiple hydrogen-bonding sites on the edges of the nucleic acid bases, individual bases as well as oligonucleotides are ideal targets for recognition through hydrogen bonding.[1,2] The major groove side of a cytosine−guanine (CG) base pair presents a noncontiguous donor−acceptor−acceptor (DAA) hydrogen-bonding array (Figure 1). This immediately suggests that a receptor capable of binding the base pair could be designed from a complementary hydrogen-bonding module. In addition to strong binding, the three hydrogen bonds formed between such a receptor and the CG base pair should provide selective binding because the thymine−adenine, guanine−

(1) Previous examples of host−guest chemistry involving recognition of nucleic acid bases include the following: (a) Feibush, B.; Saha, M.; Onan, K.; Karger, B.; Giese, R. *J. Am. Chem. Soc.* **1987**, *109*, 7531−7533. (b) Adrian, J. C., Jr.; Wilcox, C. S. *J. Am. Chem. Soc.* **1989**, *111*, 8055−8057. (c) Zimmerman, S. C.; Wu, W. *J. Am. Chem. Soc.* **1989**, *111*, 8054−8055. (d) Rebek, J., Jr. *Acc. Chem. Res.* **1990**, *23*, 399−404. (e) Hamiliton, A. D. *J. Chem. Educ.* **1990**, *67*, 821−828. (f) Hisatome, M.; Maruyama, N.; Furutera, T.; Ishikawa, T.; Yamakawa, K. *Chem. Lett.* **1990**, 2251−2254. (g) Seel, C.; Vögtle, F. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 442−444. (h) Furuta, H.; Magda, D.; Sessler, J. L. *J. Am. Chem. Soc.* **1991**, *113*, 978−985. (i) Ogoshi, H.; Hatakeyama, H.; Kotani, J.; Kawashima, A.; Kuroda, Y. *J. Am. Chem. Soc.* **1991**, *113*, 8181−8183. (j) Inouye, M.; Kim, K.; Kitao, T. *J. Am. Chem. Soc.* **1992**, *114*, 778−780. (k) Schwartz, E. B.; Knobler, C. B.; Cram, D. J. *J. Am. Chem. Soc.* **1992**, *114*, 10775−10784. (l) Nowick, J. S.; Chen, J. S.; Noronha, G. *J. Am. Chem. Soc.* **1993**, *115*, 7636−7644. (m) Eliseev, A. V.; Schneider, H.-J. *J. Am. Chem. Soc.* **1994**, *116*, 6081−6088. (n) Čudić, P.; Žinić, M.; Tomišić, V.; Simeon, V.; Vigneron, J.-P.; Lehn, J.-M. *J. Chem. Soc., Chem. Commun.* **1995**, 1073−1075. (o) Bell, T. W.; Hou, Z.; Zimmerman, S. C.; Thiessen, P. A. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2163-2165. (p) Mathew, J.; Buchardt, O. *Bioconjugate Chem.* **1995**, *6*, 524−528. (q) Kickham, J. E.; Loeb, S. J.; Murphy, S. L. *Chem. Eur. J.* **1997**, *3*, 1203−1213. (r) Spivak, D.; Gilmore, M. A.; Shea, K. J. *J. Am. Chem. Soc.* **1997**, *119*, 4388−4393. (s) Asanuma, H.; Ban, T.; Gotoh, S.; Hishiya, T.; Komiyama, M. *Macromolecules* **1998**, *31*, 371−377. (t) Piantanida, I.; Tomišić, V.; Žinić, M. *J. Chem. Soc., Perkin Trans. 2* **2000**, 375−383.

(2) Selected reviews and reports on artificial ligands for sequence specific recognition of nucleic acids include the following: (a) Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666−690. (b) Lehmann, T. E.; Greenberg, W. A.; Liberles, D. A.; Wada, C. K.; Dervan, P. B. *Helv. Chim. Acta* **1997**, *80*, 2002−2022. (c) Kool, E. T. *Chem. Rev.* **1997**, *97*, 1473−1487. (d) Doronina, S. O.; Behr, J.-P. *Chem. Soc. Rev.* **1997**, 63−71. (e) Kool, E. T. *New J. Chem.* **1997**, *21*, 33−45. (f) Nielsen, P. E. *Chem. Eur. J.* **1997**, *3*, 505−507. (g) Dervan, P. B.; Bürli, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688−693. (h) Nielsen, P. E. *Acc. Chem. Res.***1999**, *32*, 624−630.

(3) Zimmerman, S. C. Schmitt, P. *J. Am. Chem. Soc.* **1995**, *117*, 10769−10770.

(4) (a) Sasaki, S.; Nakashima, S.; Nagatsugi, F.; Tanaka, Y.; Hisatome, M.; Maeda, M. *Tetrahedron Lett.* **1995**, *36*, 9521−9524. (b) Cho, Y. L.; Jeong, K.-S. *Tetrahedron Lett.* **1997**, *38*, 8337−8340. (c) Lecubin, F.; Benhida, R.; Fourrey, J.-L.; Sun, J.-S. *Tetrahedron Lett.* **1999**, *40*, 8085−8088. (d) Saito, A.; Kuroda, R. *Tetrahedron Lett.* **1999**, *40*, 4837−4840. (e) López de la Paz, M.; González, C.; Vicent, C. *Chem Commun.* **2000**, 411−412.

10.1021/ol006157d CCC: $19.00 
**Published on Web 08/19/2000**

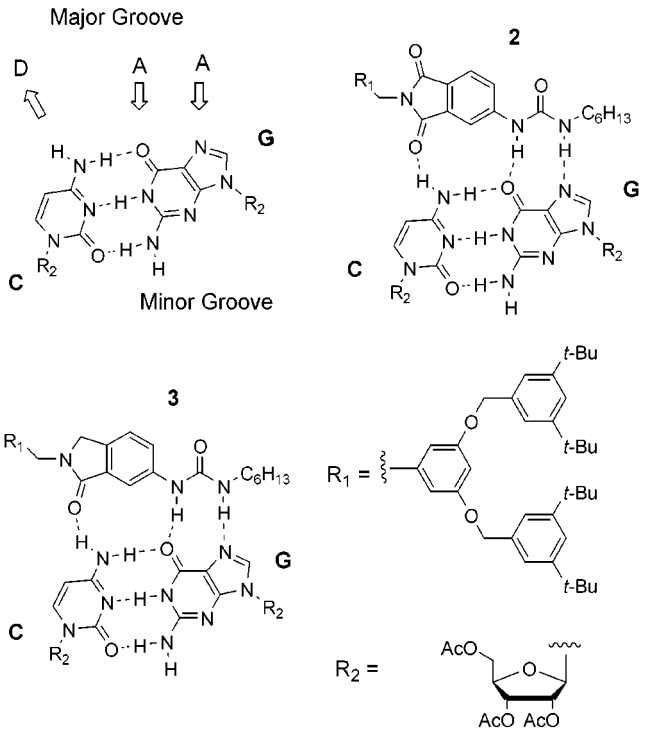
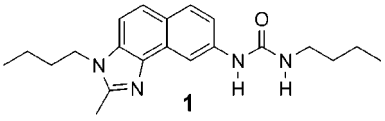



**Figure 1.** The major groove side of the CG base pair as a donor−acceptor−acceptor (DAA) hydrogen-bonding array and the complementary phthalimide (**2**) and isoindolin-1-one (**3**) receptors complexed to CG by the formation of three hydrogen bonds.

cytosine, and adenine−thymine base pairs present different arrays of hydrogen-bond donor and acceptor sites. We have previously used $^1$H NMR complexation studies as a convenient method for quantifying binding between the ureido naphthimidazole receptor **1** and a CG base pair.[3] Other



**1**

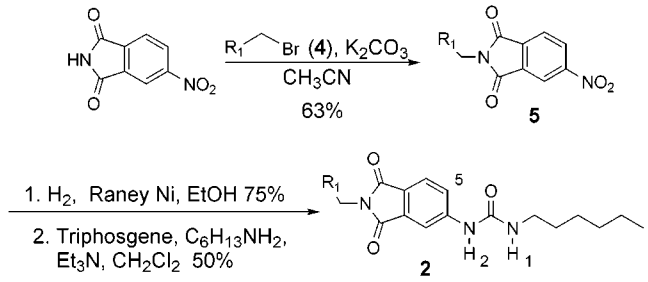
groups have subsequently used this approach to examine the binding of other artificial and natural base pair binders.[4]

Receptor **1** was found to complex CG with an association constant ($K_{assoc}$) of 160 $M^{-1}$ through formation of three hydrogen bonds; however, this $K_{assoc}$ value is relatively low considering the strength of some triply hydrogen bonded complexes having a DAA−ADD motif.[5]

Furthermore, we viewed the studies in organic solvents as a logical first step toward developing nucleotide bases capable of recognizing CG base pairs within triplex DNA. The nucleotide analogue of **1** was incorporated into oligonucleotides, but it did not effectively bind CG in the pyrimidine−purine−pyrimidine triplex motif.[2b,6] To find base pair receptors capable of stronger binding to CG, the 4-(*N*-hexylureido)-phthalimide **2** and the 4-(*N*-hexylureido)-isoindolin-1-one **3** were designed to form three hydrogen bonds with the CG base pair (Figure 1). Molecular modeling suggested a better fit for **2**·CG and **3**·CG in comparison to **1**·CG. The synthesis of receptors **2** and **3** and recognition studies with CG in chloroform solution are detailed herein.

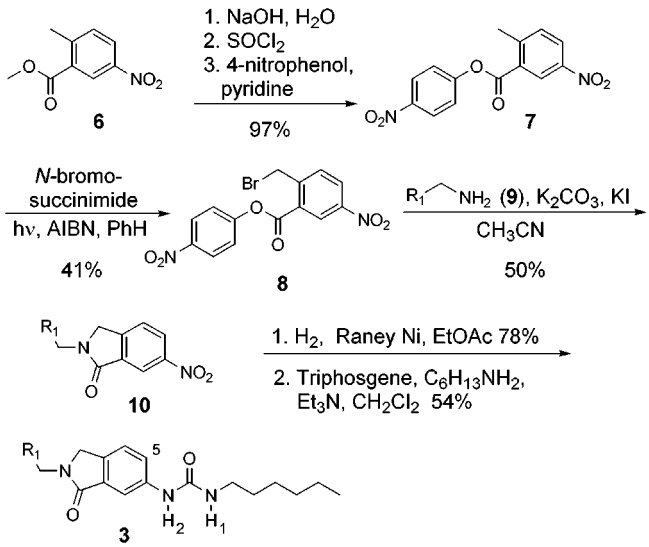
Phthalimide receptor **2** was synthesized through *N*-alkylation of 4-nitrophthalimide[7] with the dendritic bromide **4** (Scheme 1). This Fréchet-type dendron[8] provides solubility

**Scheme 1**




in organic solvents and a simple $^1$H NMR spectrum, making it ideal for NMR binding analyses. Subsequent Raney nickel mediated hydrogenation of the nitro adduct **5** followed by condensation with triphosgene and hexylamine produced the desired urea **2**. The synthesis of **3** proceeded through a key cyclization of nitrobenzoate **8** with dendritic amine **9** (Scheme 2).[9] Base precursor **8** was synthesized by conversion of methyl 6-methyl-3-nitrobenzoate (**6**)[10] to the corresponding 4-nitrophenyl ester **7**. Free radical bromination of **7** afforded **8** in moderate yield. Treatment of **8** with **9** and base afforded **10**, which was readily converted into the desired urea **3** using the methods described for the synthesis of **2**.

**Scheme 2**




(5) (a) Kelly, T. R.; Zhao, C.; Bridger, G. J. *J. Am. Chem. Soc.* **1989**, *111*, 3744−3745. (b) Murray, T. J.; Zimmerman, S. C. *J. Am. Chem. Soc.* **1992**, *114*, 4010−4011. (c) Zimmerman, S. C.; Corbin, P. S. *Struct. Bonding (Berlin)* **2000**, *96*, 64−94.

(6) Zimmerman, S. C.; Schmitt, P., unpublished results

The binding studies were performed in $CDCl_3$ solutions by titrating receptors **2** or **3** with a 1:1 mixture of tri-*O*-acetylcytidine and tri-*O*-acetylguanosine and observing the changes in $^1$H NMR chemical shifts of H-1, H-2, and H-5 (Figures 2 and 3). Strong binding between C and G ($K_{assoc} \sim 10^5$ $M^{-1}$)[11] allowed a 1:1 mixture of the two nucleosides to be treated as a single component.[3] In a representative study of the **2**·CG complex, increasing [CG] from 0.54 to 6.7 mM resulted in significant downfield shifts in the $^1$H NMR signals for H-1 ($\Delta\delta$ = 2.1 ppm), H-2 ($\Delta\delta$ = 1.5 ppm), and H-5 ($\Delta\delta$ = 0.37 ppm) of receptor **2**. Titration of **3** with CG (0.26−6.5 mM) resulted in similar downfield shifts in signals for H-1 ($\Delta\delta$ = 2.4 ppm), H-2 ($\Delta\delta$ = 1.5 ppm), and H-5 ($\Delta\delta$ = 0.40 ppm) of receptor **3**. For both receptors, plots of $\Delta\delta$ vs [CG] for the aforementioned protons were fit to a nonlinear binding equation appropriate for a 1:1 binding model.[12,13] Thus, association constants ($K_{assoc}$) of 1016 (±57) $M^{-1}$ and 736 (±155) $M^{-1}$ were determined for **2**·CG and **3**·CG, respectively (Figure 2).

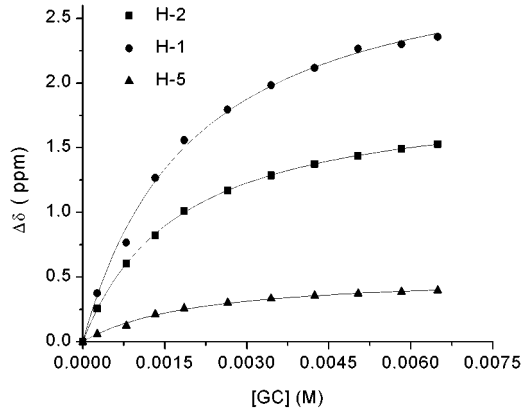
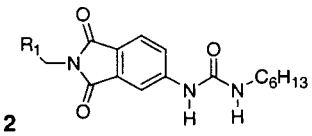
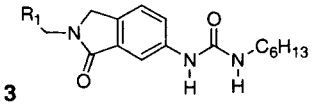
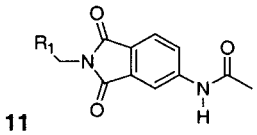



**Figure 2.** Representative plot of $\Delta\delta$ vs [CG] for titration of a synthetic receptor with CG. Shown is titration to form the complex **3**·CG. $\Delta\delta$ for only H-1, H-2, and H-5 are shown; data for these three protons were fit to a nonlinear binding equation to calculate $K_{assoc}$.

In the case of phthalimide **2**, self-association studies showed the receptor to weakly dimerize ($K_{dimer}$ = 18 $M^{-1}$). Such minimal self-association precludes dimer formation at 1.9 mM, the fixed concentration of **2** used in the complexation analysis. Interestingly, **3** was found to self-associate much more strongly than **2**, possibly due to the increased hydrogen-bond-acceptor potential of the lactam carbonyl group in **3** compared to the imide carbonyl group in **2**. Accordingly, binding studies of **3**·CG were conducted with a sufficiently low fixed concentration of **3** (0.130 mM) so that the **3**·**3** dimer could be neglected.

The downfield shifts in both urea protons of the two receptors indicate the formation of the two hydrogen bonds between the urea group and the base pair. The weaker complexation observed for the amide analogue **11**·CG complex ($K_{assoc}$ = 70 $M^{-1}$) confirms the importance of both urea group hydrogen bonds (Table 1). The downfield shift

**Table 1.** Summary of Association Constants for Complexation of Synthetic Receptors with CG

| receptor | $K_{assoc}$ ($M^{-1}$) |
|---|---|
|  **2** | 1016 ± 155 |
|  **3** | 736 ± 57 |
|  **11** | 70 |

in H-5 observed for both **2** and **3** is also consistent with complexation occurring as shown in Figure 1. Uncomplexed **2** and **3** both have free rotation about the C−N bond linking the aromatic ring and the urea group. Such free rotation allows conformations **3a** and **3b** to be populated in the case of uncomplexed **3** (Figure 3); similar conformations can exist for uncomplexed **2**. Formation of the three hydrogen bonds

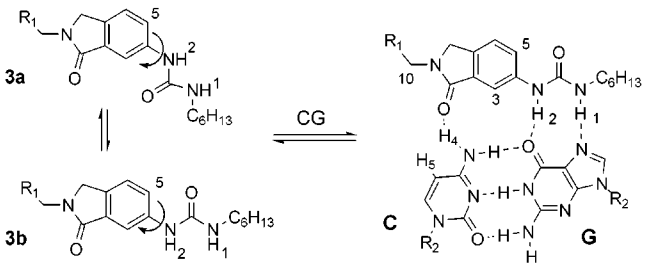



**Figure 3.** Formation of the **3**·CG complex by freezing of the *anti* conformation.

(7) Huntress, E. H.; Shriner, R. L. In *Organic Syntheses*; Blatt, A. H., Ed.; Wiley and Sons: New York, 1943; Collect. Vol. 2, p 459.

(8) (a) Hawker, C. J.; Fréchet, J. M. J. *J. Am. Chem. Soc.* **1990**, *112*, 7638−7647. (b) Zeng, F.; Reichert, D. E. C.; Kolotuchin, S. V.; Zimmerman, S. C. *Science* **1996**, *271*, 1095−1098.

(9) This cyclization approach is analogous to those used in the following: (a) Almansa, C.; Gomez, L. A.; Cavalcanti, F. L.; Rodriguez, R.; Carceller, E.; Bartroli, J.; Garcia-Rafanell, J.; Forn, J. *J. Med. Chem.* **1993**, *36*, 2121−2133. (b) Egbertson, M. S.; Naylor, A. M.; Hartman, G. D.; Cook, J. J.; Gould, R. J.; Holahan, M. A.; Lynch, J. J., Jr.; Lynch, R. J.; Stranieri, M. T.; Vassallo, L. M. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1835−1840.

(10) Phadnis, A. P.; Nanda, B.; Patwardhan, S. A.; Gupta, A. S. *Indian J. Chem.; Sect. B* **1984**, *23*, 1098−1102.

(11) Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969**, *179*, 10−17.

(12) Wilcox, C. S. In *Frontiers in Supramolecular Organic Chemistry and Photochemistry*; Schneider, H. J., Durr, H., Eds.; VCH: New York, 1991; pp 123−143.

(13) Job's plot analyses for the **2**·CG and **3**·CG complexes were consistent with 1:1 stoichiometry. Job, P. *Ann. Chim. Ser. 10* **1928**, *9*, 113−134.

between the receptors and the major groove side of CG requires the conformation **3b** wherein H-5 becomes fixed within the deshielding region of the urea carbonyl group.

Consistent with this model, complexation studies also showed upfield shifts in the signals for H-3 in both **2** and **3**. Titration of **3** with CG also caused the $^1$H NMR singlet for the methylene group protons (H-10) to become an AB quartet. These protons become diastereotopic upon binding to chiral CG. As the concentration of **3** was increased from 0.27 to 6.5 mM, $\Delta\nu$ for the AB quartet increased from 6.9 to 35.4 Hz. A plot of $\Delta\nu$ vs [CG] was also fit to a nonlinear binding equation. The association constant ($K_{assoc} = 709\ M^{-1}$) thus determined was comparable to that found through chemical shift analysis of protons H-1, H-2, and H-5.

Further evidence of the **2**·CG and **3**·CG complexes was obtained in reverse titrations. Thus, titration of 1.88 mM solutions of CG with **2** (0.57−5.1 mM) resulted in a large downfield shift in the $^1$H NMR signal for H-4 of C ($\Delta\delta$ = 0.836 ppm, see Figure 3) and a smaller downfield shift in the signal for H-5 of C ($\Delta\delta$ = 0.07 ppm). Increasing [**3**] from 0.70 to 7.0 mM in solutions with [CG] fixed at 2.0 mM resulted in similar downfield shifts in H-4 of C ($\Delta\delta$ = 0.826 ppm) and H-5 of C ($\Delta\delta$ = 0.16 ppm). A plot of $\Delta\delta$ vs [**2**] fit a nonlinear equation describing a 1:1 binding model giving $K_{assoc} = 800\ M^{-1}$, a value similar to that found for titration of **2** with CG. Data for the titration of CG with **3** did not fit a 1:1 binding model. This result may be attributed to dissociation of the CG base pair upon addition of an excess of **3**.[14] For titration of CG with both **2** and **3**, the observed downfield shifts in H-4 of C are consistent with the formation of a hydrogen bond between the cytidine amino group and the imide or lactam carbonyl group of the receptors. Downfield shifts in H-5 of C also support the formation of such a hydrogen bond as H-5 moves into the deshielding region of the imide or lactam carbonyl group.

In summary, $^1$H NMR binding studies indicate that synthetic receptors **2** and **3** can bind CG strongly through formation of three hydrogen bonds on the major groove side of the base pair. The stability of these base triplets is about 10-fold higher (in terms of $K_{assoc}$) than that for the CG binder **1**. Complexation in organic solvents does not necessarily guarantee that a receptor will recognize a base pair within a target DNA double helix.[2b,6] Nonetheless, the strong complexation of the CG base pair by receptors **2** and **3** bodes well for the development of nonnatural nucleosides capable of recognizing CG base pairs in the pyrimidine−purine−pyrimidine DNA triple helix motif.[2a−e] A nucleoside analogue of receptor **3** has been successfully incorporated into oligonucleotides to test its ability to recognize CG base pairs within double helical DNA. The results of these studies will be reported separately.

**Acknowledgment.** We thank the National Institutes of Health (GM38010) for funding of this work and the University of Illinois for fellowship assistance.

OL006157D

(14) To determine the extent to which **3** can dissociate a CG base pair, complexes of **3** and the individual nucleosides (**3**·C and **3**·G) were studied by $^1$H NMR. Complexation was observed in the case of **3**·C ($K_{assoc} = 319\ M^{-1}$), but no complexation was observed for **3**·G. We have not attempted to fit these data to a multiequilibrium model, which in any event would be complicated by the absence of a precise dissociation constant for the CG base pair.