Model Studies Directed toward a General Triplex DNA Recognition Scheme: A Novel DNA Base That Binds a CG Base-Pair in an Organic Solvent

Steven C. Zimmerman* and Philippe Schmitt

Department of Chemistry, University of Illinois Urbana, Illinois 61801

Received July 27, 1995

Oligonucleotides are capable of binding to regions of double helical DNA through the formation of localized triple helical (triplex) structures.[1] Recently, this recognition process has attracted considerable interest because of its potential use in regulating gene expression,[1c] selectively cleaving DNA,[2] and affecting a range of other biological processes.[3] However, for these applications to be realized, triplex formation must occur selectively at any chosen sequence, and yet to date, the DNA target sites are mostly limited to homopurine tracts. This limitation arises from the mechanism of recognition which involves specific Hoogsteen-type hydrogen bonds between the bases of the third strand and purine bases in the major groove of the DNA duplex.[4] In the best-studied case, the third strand contains pyrimidines which bind the duplex by forming the base triplets shown in Figure 1a,b. Within this pyr–pur–pyr motif, G will selectively recognize the TA base-pair with intermediate affinity, but none of the four naturally occurring bases shows selectivity for the CG base-pair.[5] Thus, there is a need for nonnatural bases that selectively bind the CG base-pair with high affinity in the pyr–pur–pyr motif.

The difficulty in designing a base to selectively and tightly bind the CG base-pair is apparent from Figure 1c, where a single hydrogen bond donor group is presented by the cytosine base. One logical way to achieve affinity and selectivity is to use an extended base that can simultaneously contact the cytosine 4-NH group and the Hoogsteen hydrogen bonding site on the guanine base (Figure 1c).[6] The novel DNA nucleoside 3-(2-deoxy-β-D-ribofuranosyl)-2-methyl-8-(*N'*-*n*-butylureido)naphth[1,2-*d*]imidazole (**1**) was designed to contain a complementary acceptor–donor–donor hydrogen bonding array to the CG base-pair and to form the base-triplet shown in Figure 1d. An important design criterion was that **1** contain a glycosidic bond compatible with the geometrical constraints imposed by the phosphodiester backbone within the pyr–pur–pyr motif.

Herein we describe the synthesis of **1** and show that its alkyl analog **5** binds a CG base-pair in chloroform through the formation of simultaneous hydrogen bonds to both bases. To our knowledge, this is the first study to examine an isolated, nonnatural base-triplet in an organic solvent.[7] This type of model study allows the quality of the hydrogen bonding interactions to be evaluated without restrictions from base-stacking or spatial constraints from preferred backbone conformations.

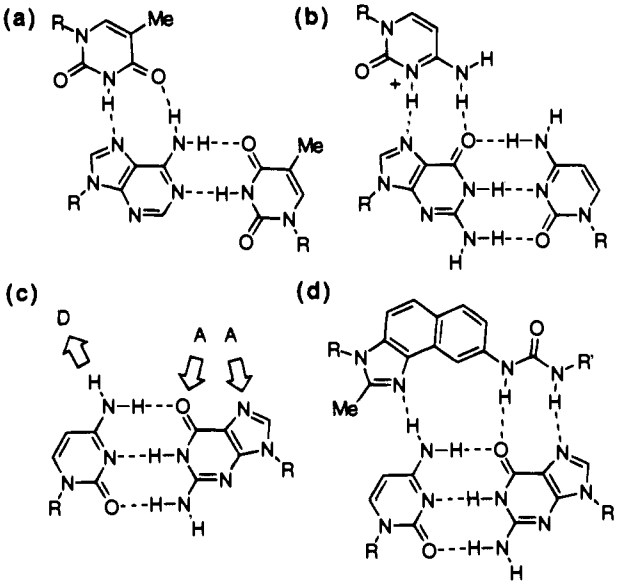
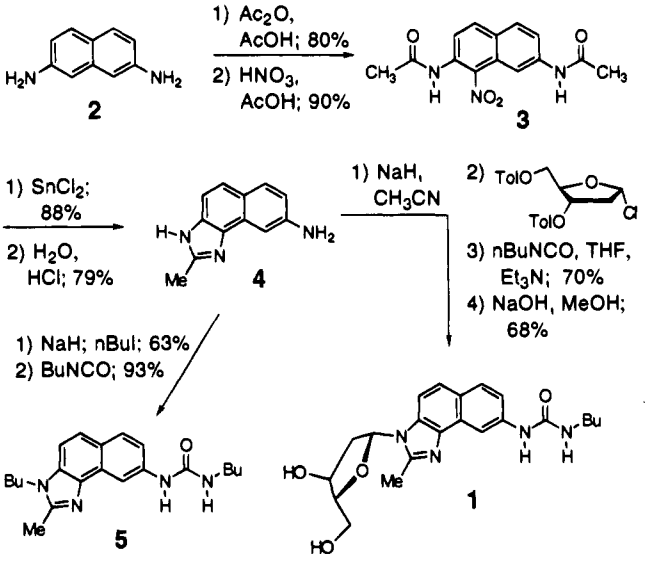


Figure 1. (a) TAT base-triplet, (b) C^+GC base-triplet, (c) major groove hydrogen bonding sites of CG base-pair, and (d) **1**(**5**)–CG base-triplet.

Scheme 1



The synthesis of **1** is outlined in Scheme 1.[8] The starting material, **2**, was made in multigram quantities by a Bücherer reaction of commercially available 2,7-dihydroxynaphthalene.[9] Acetylation, nitration, reductive cyclization, and hydrolysis provided base **4** in 50% overall yield. Glycosidation of **4**, treatment with *n*-butyl isocyanate, and deprotection afforded **1**, whose stereo- and regiochemistry were confirmed by ^1H NMR coupling constants and NOE studies. The overall synthesis is efficient, although none of the yields have been optimized. To have a simpler analog for chloroform binding studies, **4** was alkylated with *n*-butyl iodide and converted to urea **5** as outlined in Scheme 1.

Initially, a qualitative complexation study was performed in chloroform-*d* by titrating a 1:1 mixture of tri-*O*-acetylguanosine

(1) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650. (b) LeDoan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thoung, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749–7760. (c) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456–459.

(2) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639–1642.

(3) For leading references, see: Thoung, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 6666–6690.

(4) Early studies: Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024. Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822–823. Miller, J. H.; Sobell, H. M. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 1201–1205. Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. T. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 93–101.

(5) Griffin, L. C.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *245*, 967–971. Mergny, J.; Sun, J.; Rougée, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, M.; Hélène, C. *Biochemistry* **1991**, *30*, 9791–9798.

(6) Similar approaches in triplex studies: Huang, C.-Y.; Cushman, C. D.; Miller, P. S. *J. Org. Chem.* **1993**, *58*, 5048–5049. Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908–7909. Huang, C.-Y.; Miller, P. S. *J. Am. Chem. Soc.* **1993**, *115*, 10456–10457.

(7) CGG triplet: Williams, L. D.; Chawla, B. Shaw, B. R. *Biopolymers* **1987**, *26*, 591–603.

(8) New compounds gave a correct elemental analysis or HRMS and had spectral data in full accord with their structures.

(9) Bell, F.; Gibson, J. A.; Wilson, R. D. *J. Chem. Soc.* **1956**, 2335–2340.

0002-7863/95/1517-10769$09.00/0

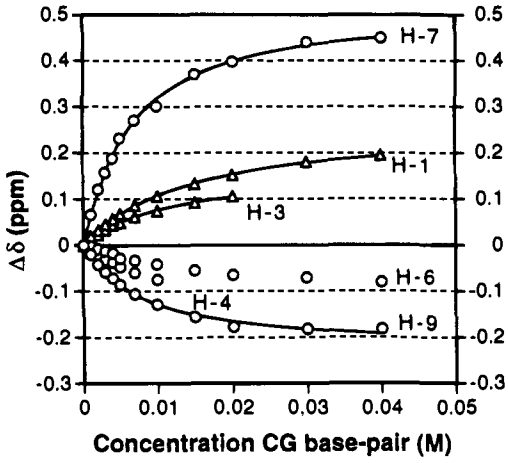


Figure 2. Plot of chemical shift of selected protons in **5** (○) and **6** (△) as a function of concentration of CG base-pair. Theoretical curves are nonlinear least-squares curve fitting to a 1:1 (**5** or **6** to CG) binding isotherm.

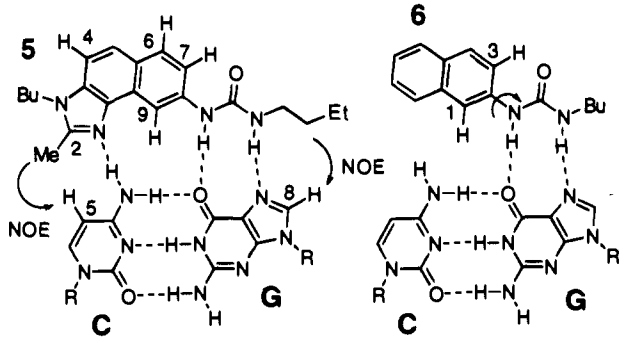


Figure 3. Numbering system and NOEs observed for **5**−CG and **6**−CG base-triplets.

and tri-*O*-acetylcytidine (2 mM) with **5** (1 → 20 mM). This experiment and the reverse titration (vide infra) rely on the high stability of the CG base-pair in chloroform ($K_{assoc} > 10^4-10^5$ M^{-1}),[10] which under the conditions of the titration allows it to be treated as a single, associated species. That the base-pair remains intact is supported by the observation that the imino proton of the GC base-pair shifted minimally ($\Delta\delta \leq 0.15$ ppm) during the titration. Furthermore, one of the amino groups shifted ($\Delta\delta > 1$ ppm), while the other remained fixed ($\Delta\delta \approx 0.0$ ppm), consistent with hydrogen bonding at the cytosine 4-amino group of an intact CG base-pair.

The reverse titration was carried out by titrating **5** (2 mM) with a 1:1 mixture of tri-*O*-acetylguanosine and tri-*O*-acetylcytidine (1 → 20 mM). During the titration of **5** by the CG base-pair, the aryl and aliphatic urea N−H protons of **5** shift downfield by 2.06 and 1.84 ppm, respectively. These shifts indicate that both urea N−H protons are engaged in hydrogen bonds, presumably to the Hoogsteen site of G within the CG base-pair. Most informative are the magnitude and direction of the aryl proton shifts (Figure 2) and the analogous shifts seen in an identical titration of naphthylurea **6** (see also, Figure 3). The small shifts seen in H-4 and H-6 of **5** indicated that, as expected, the primary mode of complexation does not involve π-stacking. The large downfield shift of H-7 and the moderate upfield shift of H-9 can best be explained by a freezing of the rotation of the aryl to urea nitrogen bond with the urea carbonyl oriented toward H-7 in the complex. Loss of this free rotation would occur if an additional interaction locked the relative positions of the urea and aryl groups, as would be expected by simultaneous hydrogen bonding to the imidazole and urea groups (Figure 3). Support for these arguments can be seen in the titration of naphthylurea **6**, where the H-1 and H-3 protons—analogous, respectively, to H-9 and H-7 of **5**—are both shifted downfield to a small extent (~0.1−0.2 ppm), i.e., no loss of bond rotation.

Further support for the base-triplet structure shown in Figure 1d is the higher K_{assoc} for the **5**−CG complex (K_{assoc} = 160 M^{-1}) than for the **6**−CG complex (K_{assoc} = 70 M^{-1}), which is consistent with the additional intermolecular hydrogen bond. The increase in K_{assoc} from the additional hydrogen bond is balanced by (1) the loss of the aryl−urea bond rotation in the **5**−CG complex, (2) a lengthening and slight bending of the hydrogen bonds between **5** and G upon formation of the additional contact, as indicated by modeling, (3) the potentially weaker hydrogen bond donor ability of the urea in **5** versus **6**,[11] and (4) the fact that the **6**−CG complex can form in two ways. In any event, the K_{assoc} values represent lower limits because the CG base-pair dimerizes under the conditions of the titration experiments (K_2 = 125 M^{-1}). Compound **5** also aggregates in chloroform-*d* above a critical concentration of 18 mM, but its dilution curve could not be fit to a simple dimer, trimer, or single *n*-mer model, suggesting a polymeric aggregation process. Self-association was negligible at 2 mM, the concentration where the quantitative complexation studies were performed.

Finally, difference NOE experiments were performed on a 1:1:1 mixture of 9-octylguanine, 1-propylcytosine, and **5** at 20 mM in 5% DMSO-d_6−chloroform-*d*. The addition of DMSO was necessary to reduce the degree of overlap in the aliphatic portion of the ^{1}H NMR. Under these conditions, weak intermolecular contacts are seen between the methyl group of **5** and H-5 of cytosine and between H-8 of guanine and the β-methylene (but not α) protons of the urea butyl group (Figure 3). These contacts were the only ones predicted from modeling and are fully consistent with the base-triplet structure in Figure 3. No additional intermolecular contacts were seen in a ROESY experiment conducted on a 15 mM 1:1:1 mixture of tri-*O*-acetylguanosine, tri-*O*-acetylcytidine, and **5** in chloroform-*d*.

Described herein is the synthesis of novel DNA nucleoside **1**. Model studies on alkyl analog **5** document its ability to bind a CG base-pair in chloroform solution, forming a stable base-triplet with six hydrogen bonds. These types of simple studies provide important information about the quality of the hydrogen bonding contacts in isolation from base-stacking and backbone conformational considerations. As such, they may provide an important screen for candidates to be incorporated into oligonucleotides. Of course, the ability to recognize CG base-pairs within the triplex motif is the ultimate test. Novel base **1** has been incorporated into oligonucleotides, and its recognition properties within triplex DNA will be reported in due course.

Acknowledgment. Partial funding of this work by the National Institutes of Health (GM38010) is gratefully acknowledged. We thank Prof. J. Wand for assistance with the ROESY experiment and Paul Thiessen for the modeling studies.

JA952519E

(10) Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969**, *179*, 10−17. Zimmerman, S. C.; Murray, T. J. In *Computational Approaches in Supramolecular Chemistry*; Wipff, G., Ed.; NATO ASI Series 426; Kluwer: Amsterdam, 1994; pp 109−115 and references therein.

(11) Wilcox, C. S.; Kim, E.; Romano, D.; Kuo, L. H.; Burt, A. L.; Curran, D. P. *Tetrahedron* **1995**, *51*, 621−634.