2′-Deoxy-3-isoadenosine Forms Hoogsteen-Type Base Pairs with Thymidine in the d(CG[iA]TCG)$_2$ Duplex[1]

Balkrishen Bhat,† Neelima,† Nelson J. Leonard,*,‡ Howard Robinson,§ and Andrew H.-J. Wang*,†,§

Department of Chemistry and Department of Cell and Structural Biology University of Illinois at Urbana-Champaign Urbana, Illinois 61801

Received December 12, 1995

Present-day genetic material, DNA and RNA, consists of five major nucleic acid bases (A, G, T, C, and U) and the sugar−phosphate backbone. In addition to the question of their selection from prebiotic matter, there are more subtle issues, including, in the case of purines, the site of attachment of ribose/deoxyribose at N9 versus N3.[2] The reaction of adenine with protected sugar halides gives a mixture containing more or less equal amounts of N9 and N3 derivatives.[3a]

Strong interaction between 3-isoadenosine and uridine has been observed in the complex of poly(3-isoadenylic acid) with poly(U)[3b] and by the efficient poly(U)-template-directed oligomerization of the imidazolide of 3-isoadenosine 5′-phosphate.[2a] It is probable that the base pairing between oligo(iA) and oligo(U) in these systems is of the Hoogsteen type (involving the 6-NH_2 and N7 of the 3-isoadenosine), but this was not established. Similar interaction can be explained between 2′-deoxy-3-isoadenosine (**1**)[4] and thymidine. Accordingly, we have now addressed the question of interstrand hydrogen bonding (Figure 1B) by examining the 3D structure of a DNA hexamer d(CGATCG) as a control and its iA-substituted analog, d(CG-[iA]TCG), by NMR. This sequence was chosen since it has been used extensively in studies of conformational analysis,[5] ligand−DNA interactions,[6] and restriction-endonuclease recognition.[7]

The modified hexamer d(CG[iA]TCG) was synthesized with difficulty, affording a somewhat limited amount of material.[8] Nonetheless, we were able to obtain NMR data of sufficient quality for structural analysis.[11] The H_2O exchangeable spectrum of d(CGATCG)$_2$ revealed two clear imino proton resonances at 12.95 (G2), 13.73 (T4) ppm plus a rapidly exchanging peak at 13.22 (G6), suggesting Watson−Crick base pairs (Figure 1C).[12] In contrast, d(CG[iA]TCG)$_2$ displayed two clear imino

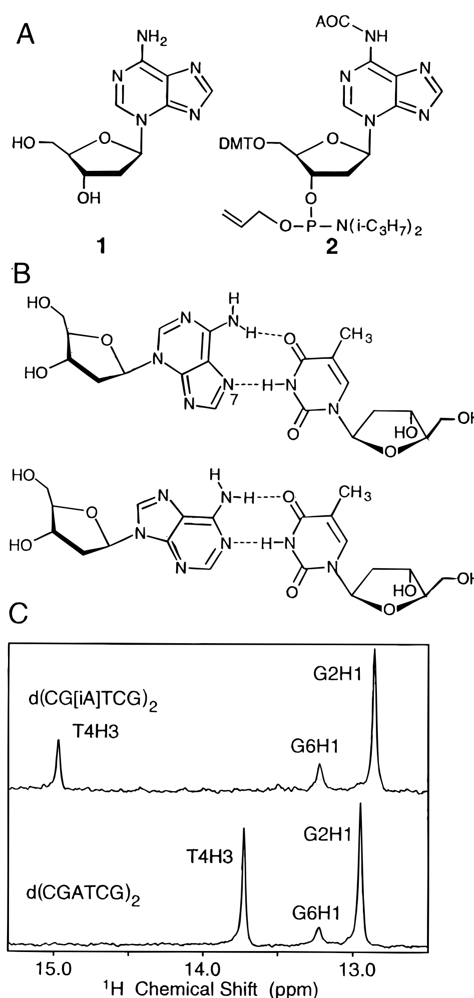

Figure 1. (A) Chemical structure of 2′-deoxy-3-isoadenosine (**1**) and protected nucleoside (**2**). (B) Hoogsteen iA:T and Watson−Crick A:T base pairs. Note that all nucleosides (including iA) are in the *anti* conformation and the C1′−C1′ distances of the two types of base pairs are very similar (~10.5 Å). (C) The imino proton 1D-NMR spectra of d(CG[iA]TCG)$_2$ and d(CGATCG)$_2$ (both at 0.4 mM duplex) at 2 °C.

† Department of Chemistry.
‡ Present address: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125.
§ Department of Cell and Structural Biology.

(1) Abbreviations used: iA, 3-isoadenosine or, in context, 2′-deoxy-3-isoadenosine; AOC, allyloxycarbonyl; DMT, 4,4′-dimethoxytrityl; NMR, nuclear magnetic resonance; 2D-NOESY, two-dimensional nuclear Overhauser effect spectroscopy; TOCSY, total-correlated spectroscopy.

(2) (a) Hill, A. R., Jr.; Kumar, S.; Leonard, N. J.; Orgel, L. E. *J. Mol. Evol.* **1988**, *27*, 91−95. (b) Hill, A. R., Jr.; Kumar, S.; Patil, V. D.; Leonard, N. J.; Orgel, L. E. *J. Mol. Evol.* **1991**, *32*, 447−453 and references therein.

(3) (a) Leonard, N. J.; Laursen, R. A. *Biochemistry* **1965**, *4*, 354−365. (b) Michelson, A. M.; Monny, C.; Laursen, R. A.; Leonard, N. J. *Biochim. Biophys. Acta.* **1966**, *119*, 258−267.

(4) (a) Rasmussen, M.; Leonard, N. J. *J. Am. Chem. Soc.* **1967**, *89*, 5439−5445. (b) Nair, V.; Buenger, G. S.; Leonard, N. J.; Balzarini, J.; DeClercq, E. *J. Chem. Soc., Chem. Commun.* **1991**, *22*, 1650−1651.

(5) (a) Robinson, H.; Wang, A. H.-J. *Biochemistry* **1992**, *31*, 3524−3533. (b) Baxter, S. M.; Greizerstein, M. B.; Kushlan, D. M.; Ashley, G. W. *Biochemistry* **1993**, *32*, 8702−8711. (c) Gao, H.; Yang, M.; Cook, A. F. *Nucleic Acids Res.* **1995**, *23*, 285−292.

(6) (a) Chen, K. X.; Gresh, N.; Pullman, B. *Mol. Pharcmacol.* **1986**, *30*, 279−286. (b) Moore, M. H.; Hunter, W. N.; d'Estaintot, B. L.; Kennard, O. *J. Mol. Biol.* **1989**, *206*, 693−705. (c) Williams, L. D.; Frederick, C. A.; Ughetto, G.; Rich, A. *Nucleic Acids Res.* **1990**, *29*, 5533−5541. (d) Frederick, C. A.; Williams, L. D.; Ughetto, G.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, A. H.-J. *Biochemistry* **1990**, *29*, 2538−2549. (e) Nunn, C. M.; Van Meervelt, L.; Zhang, S.; Moore, M. H.; Kennard, O. *J. Mol. Biol.* **1991**, *222*, 167−177. (f) Gao, Y. G., Wang, A. H.-J. *Anti-Cancer Drug Res.* **1991**, *6*, 137−149. (g) Gallois, B.; d'Estaintot, B. L.; Brown, T.; Hunter, W. N. *Acta Crystallogr.* **1993**, *D49*, 311−317. (h) Lipscomb, L. A.; Peek, M. A.; Zhou, F. X.; Bertrand, J. A.; Van Der Veer, D.; Williams, L. D. *Biochemistry* **1994**, *33*, 3649−3659.

(7) (a) Lynn, S. P.; Cohen, L. K.; Gardner, J. F.; Kaplan, S. *J. Bacteriol.* **1979**, *138*, 505−509. (b) Choi, S. H.; Leach, J. E. *Mol. Gen. Genet.* **1994**, *244*, 383−390.

(8) The incorporation of a 3-iso-dA (**1**) (Figure 1A) unit in the modified hexamer is not simply a matter of isomer substitution in an automated DNA synthesis, because **1** itself is readily susceptible to hydrolysis, as is 3-isoadenosine.[3a] We have devised a scheme for overcoming this problem through our studies of the synthesis of d(iApT) and d(TpiA)[9] by combining standard protection (on an automated synthesizer) with allyl and allyloxycarbonyl protection and hydrogenolytic deprotection.[10] β-Cyanoethylphosphoramidite chemistry was used with the exception of the introduction of dA (for practice) or 3-iso-dA. For the incorporation of the latter in the growing chain from the 3′-end, *freshly-made* N^6-(allyloxycarbonyl)-5′-*O*-(4,4′-dimethoxytrityl)-2′-deoxy-3-isoadenosyl-3′-*O*-(allyl *N,N*-diisopropyl)-phosphoramidite (**2**) was injected into the system by syringe. For the final phase, the fully protected hexanucleotide on the solid support was treated with the palladium(0) reagent, $Pd_2(dba)_3 \cdot CHCl_3$,[10] which removed the allyl and allyloxycarbonyl groups. The partially-deprotected hexanucleotide on resin was then returned to the synthesizer for routine completion of the deprotection scheme followed by release from the solid support. Purification of d(CG[iA]TCG) was accomplished by reverse-phase HPLC on an Altex ODS C18 column in 50 mM sodium phosphate, pH 5.6, at a flow rate of 2 mL/min. A linear gradient of 10−40% of 95% aqueous methanol applied over 30 min provided hexanucleotide. The sample was dried and desalted on the reverse-phase column. The column was eluted for 20 min with water at 2 mL/min followed by 95% methanol over 10 min. The appropriate fraction was dried on a Speed Vac at 20 °C to give the hexamer as a colorless solid: laser desorption ionization MS *m/e* 1793.2 (MH^+) ($C_{56}H_{75}N_{23}O_{34}P_5$ requires 1793.3).

0002-7863/96/1518-3065$12.00/0 © 1996 American Chemical Society

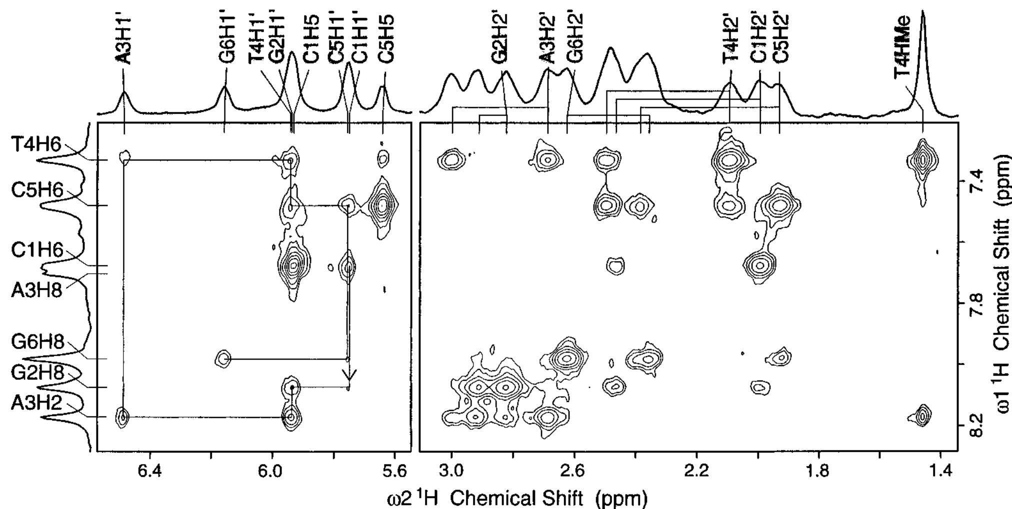

Figure 2. The expanded aromatic to H1′, H2′, H2″, methyl region of the 2D-NOESY spectra of the d(CG[iA]TCG)$_2$ duplex collected at 2 °C in D_2O, showing the nucleotide conformation.

proton resonances at 12.87 (G2), 14.96 (T4) ppm plus a rapidly exchanging peak at 13.23 (G6). The T4H3 imino resonance is further downfield than that of either the normal Watson−Crick A:T or the Hoogsteen A:T base pair in the echinomycin-drug complex[13] or in triplexes.[14] Thus, the strength of the iA:T base pair may be slightly greater than the normal A:T base pair. A strong NOE cross peak between the T4H3 imino proton and the iA3H8, but not iA3H2, proton (data not shown) is consistent with the Hoogsteen-type base pair.

The 2D-NOESY and TOCSY spectra in D_2O were used to assign the resonances of all nonexchangeable protons (Table 1S). Figure 2 shows the aromatic to H1′, H2′, H2″, and methyl cross peak regions of the 2D-NOESY spectrum. The uninterrupted sequential cross peak connectivity of the aromatic protons to the H1′ protons suggested a helical structure, and their intensities indicated that all nucleotides are in the *anti* conformation. The *anti* conformation of the iA3 residue defines how iA pairs with T4 of the opposite strand. Figure 1B compares the Watson−Crick A:T and Hoogsteen iA:T base pairs. It is noteworthy that the C1′−C1′ distance in both base pairs is ∼10.5 Å.[15] Therefore, the Hoogsteen iA:T base pair could be incorporated into a B-DNA duplex with minimum conformational perturbation. A starting model of an iA-modified duplex was built and subsequently energy minimized for NMR refinement. The 3D structure was obtained by a combined SPEDREF[11] and NOE-constrained molecular dynamics refinement.[16] The refined model is of the B-type family (Figure 3). The sugar puckers of all residues in both the d(CGATCG)$_2$ and d(CG[iA]TCG)$_2$ are of the S-type on the basis of the coupling constants measured from the PE-COSY spectra (Table 2S).

In the normal B-DNA of d(CGATCG)$_2$, the pyrimidine ring of the A3 base is stacked significantly with the T4 base. In the structure of the d(CG[iA]TCG)$_2$ duplex, a similar stacking pattern is maintained by adjusting the helical twist angle and the slide between the two iA:T base pairs. This has the effect of bringing the two sugars closer together. Indeed, this is reflected in several NOE cross peaks which are not visible in normal B-DNA. For example, T4 methyl has strong cross peaks to iA3H2′ (3.4 Å), H2″ (4.3 Å), and H3′ (3.5 Å). Another unusual observed cross peak is between iA3H2″ and T4H5′ (2.5 Å).

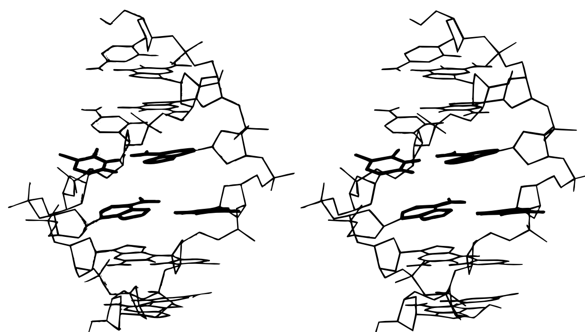

Figure 3. The NOE-restrained refined structure of d(CG[iA]TCG)$_2$. The NMR *R*-factor is 17%. Hydrogen atoms are shown only for the exchangeable protons.

Our results show that DNA incorporating iA can be synthesized by using novel protecting groups that can be removed under mild conditions. The oligonucleotide d(CG[iA]TCG) forms a well-defined B-type duplex with Hoogsteen-type iA:T base pairs flanked by Watson−Crick pairs. The results support the predicted Hoogsteen-type base pairing in the oligo(iA)•-oligo(U) complexes.[2]

Structural studies of alternative base pairs such as this study or alternative backbones (3′−5′ versus 2′−5′)[17] in nucleic acids will help us understand the limiting parameters for early nucleic acid development. A possible scenario deduced from the results here is that nature might have discarded the N3 attachment site for purines due to its chemical instability[4a] instead of structural incompatibility.

Acknowledgment. This work was supported by the NIH grant GM-41612 to A.H.-J.W. The Varian VXR500 NMR spectrometer was supported in part from the NIH shared instrumentation grant 1S10RR06243.

Supporting Information Available: Two figures showing the additional 2D-NOESY data and two tables listing the chemical shifts and coupling constants (6 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA954178A

(9) Leonard, N. J.; Neelima. Manuscript submitted for publication.
(10) Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. *J. Am. Chem. Soc.* **1990**, *112*, 1691−1696.
(11) (a) Robinson, H.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H.-J. *Biochemistry* **1992**, *31*, 10510−10517. (b) Robinson, H.; Wang, A. H.-J. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5224−5228.
(12) States, D. J.; Haberkorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982**, *48*, 286−292.
(13) Gilbert, D.; Feigon, J. *Curr. Opin. Struct. Biol.* **1991**, *1*, 439−445.
(14) Radhakrishnan, I.; Patel, D. J. *Biochemistry* **1994**, *33*, 11405−11416.
(15) This is in contrast to the Hoogsteen A:T base pair where the C1′−C1′ distance is 8.5 Å in the d(CGTACG)−triostine complex: Wang, A. H.-J.; Ughetto, G.; Quigley, G. J.; Hakoshima, T.; van der Marel, G. A.; van Boom, J. H.; Rich, A. *Science* **1984**, *225*, 1115−1121.
(16) Brünger, A. *X-PLOR* (version 3.1); The Howard Hughes Medical Institute and Yale University: New Haven, CT, 1993.
(17) Robinson, H.; Wang, A. H.-J.; Jung, K.-E.; Switzer, C. *J. Am. Chem. Soc.* **1995**, *117*, 837−838.