A New Pyrimidine Nucleoside ($m^{5ox}C$) for the pH-Independent Recognition of G-C Base Pairs by Oligonucleotide-Directed Triplex Formation

Guobing Xiang, Walid Soussou, and Larry W. McLaughlin*

Department of Chemistry, Merkert Chemistry Center
Boston College, Chestnut Hill, Massachusetts 02167

Received August 16, 1994

Pyrimidine oligodeoxynucleotides will bind to double-stranded DNA containing polypurine tracts by interaction with the Hoogsteen functional groups of the target purine residues.[1] In this motif, the pyrimidine strand interacts in a parallel orientation with the purine residues, and recognition occurs through the formation of T-A-T[2] and C^+-G-C base triplets.[1,3] The pH dependence of the C^+-G-C base triplet, as well as NMR studies, suggests that the N^3-nitrogen of the C residue must be protonated in order to form the desired hydrogen bond in the C^+-G-C base triplet.[4] The use of m^5C in place of C provides additional triplet stability under neutral or slightly basic pH conditions,[5] but the use of multiple C or m^5C residues can result in overall complex destabilization, presumably due to charge repulsion effects.[6] Some nucleoside analogues including pseudoisocytosine,[7] two related pyrazolopyrimidinone derivatives,[8] and 8-oxoadenine,[9] have been developed to eliminate this pH dependence and permit effective triplex formation at pH values above 7. In addition to nucleoside analogues, G can interact with G-C base pairs from an antiparallel orientation,[10] but this recognition motif may depend upon the nature of the target sequence.[9a,10b] We wish to report on a pyrimidine nucleoside that permits the recognition of G-C base pairs by the parallel-stranded recognition motif under neutral, mildly acidic, and mildly basic pH conditions.

Our design criteria for a nucleoside analogue included (i) a pyrimidine-like ring system to minimize anomalous conformational changes in the backbone of the triplex strand, (ii) a bidentate hydrogen bond donor to interact effectively with the

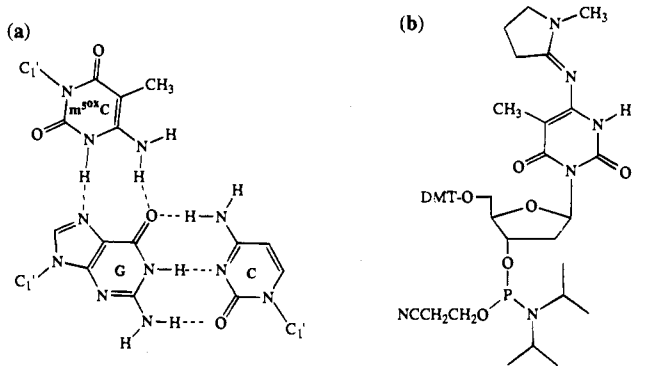


Figure 1. (a) The $m^{5ox}C$ nucleoside analogue, which was designed to mimic m^5C in its interaction with target G-C base pairs but provided instead two pH-independent hydrogen bond donors for interaction with the O^6-oxygen and N^7-nitrogen of G. (b) Structure of the fully protected phosphoramidite derivative of $m^{5ox}C$ suitable for DNA synthesis.

O^6-oxygen and N^7-nitrogen of the target G-C, and (iii) a 5-methyl group (effective in providing additional stability in the m^5C^+-G-C triplet) to contribute to the formation of a potentially important, stabilizing, hydrophobic spine.[5b] One analogue that fits these three criteria is 4-amino-1-(β-2-deoxy-D-*erythro*-pentofuranosyl)-5-methyl-2,6-[*1H,3H*]-pyrimidione ($m^{5ox}C$),[11] which is simply the 6-keto derivative of m^5C and is shown in Figure 1a in the putative $m^{5ox}C$-G-C base triplet. Of the possible tautomers for this base residue, N^3-H or the O^6-H, the former is the desired structure for triplex formation by the present design, and it appears to be the preferred form based upon single crystal x-ray analysis of the related 4-amino-1-(β-D-ribofuranosyl)-2,6-[*1H,3H*]-pyrimidione.[12]

The heterocyclic base 6-aminothymine was prepared by a simple condensation of urea with ethyl 2-cyanopropionate.[13] Glycosidation of the silylated 6-aminothymine resulted in two regiochemical products.[14] The major nucleoside produced was the N^3-glycosidation[15] product, $m^{5ox}C$ (α and β anomers),[16] while the minor product appeared to be the N^1-glycosidation product, 6-aminothymidine (α and β anomers).[17] These glycosidation products could be differentiated[16,17] on the basis of the NMR characteristics of the amino resonances as have been described[18] for this general ring system. Glycosidation (or simple alkylation) of silyl-protected trisubstituted pyrimidiones has been described[18,19] to occur primarily at the nitrogen between

(1) For early examples, see: (a) Felsenfeld, G.; Davies, D. R.; Rich A. *J. Am. Chem. Soc.* **1957**, *79*, 2023−2025, (b) Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822−825. For recent seminal examples, see: (c) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645−649. (d) Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am Chem. Soc.* **1988**, *110*, 7927−7928. (e) Praseuth, D.; Perrouault, L.; Le Doan, T.; Chassignol, M.; Thuong, N. T.; Lhomme, J.; Helene, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1349−1352.

(2) All letter abbreviations refer to the 2′-deoxynucleosides.

(3) See, for example: (a) Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 7286−7288. (b) Syamichev, V. I.; Mirkin, S. M.; Frank-Kamenetskii, M. D.; Cantor, C. R. *Nucleic Acids Res.* **1988**, *16*, 2165−2178. (c) Francois, J. C.; Saison-Behoaras, T.; Thuong, N. T.; Helene, C. *Biochemistry* **1989**, *28*, 9617−9626.

(4) (a) Rajagopal, P.; Feigon, J. *Nature* **1989**, *239*, 637−639. (b) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859−7864. (c) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282−7290.

(5) (a) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids Res.* **1984**, *12*, 6603−6614. (b) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059−3061. (c) Collier, D. A.; Thuong, N. T.; Helene, C. *J. Am. Chem. Soc.* **1991**, *113*, 1457−1458.

(6) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P. *Biochemistry* **1984**, *23*, 3277−3281.

(7) (a) Ono, A.; Ts'o, P. O. P.; Kan, L.-S. *J. Am. Chem. Soc.* **1991**, *113*, 4032−4033. (b) Ono, A., Ts'o, P. O. P.; Kan, L.-S. *J. Org. Chem.* **1992**, *57*, 3225−3230.

(8) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470−1478.

(9) (a) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3761−3764. (b) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788−6793. (c) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249−3254.

(10) (a) Cooney, M.; Czernuszesicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456−459. (b) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360−1363. (c) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9246−9255.

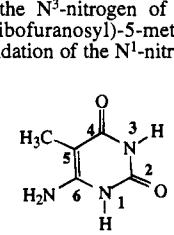
(11) This abbreviation reflects the fact that the 4-amino-1-(β-2′-deoxy-D-*erythro*-pentofuranosyl)-5-methyl-2,6-[*1H,3H*]-pyrimidione nucleoside can be viewed as an oxidized version of m^5C.

(12) Gorski, J.; Tollin, P. *Cryst. Struct. Commun.* **1982**, *11*, 543−546.

(13) (a) *Chem. Abstr.* **1946**, *70*, P78006w. For the corresponding 6-aminouracil, see: (b) Cain, C. K.; Mallette, M.; Taylor, E. C., Jr. *J. Am. Chem. Soc.* **1946**, *68*, 1996−1999.

(14) The nucleoside was prepared using α-1-chloro-3,5-*O*-tolyl-β-2-deoxy-D-*erythro*-pentofuranose (Hoffer, M. *Chem. Ber.* **1960**, 2777−2781).

(15) Glycosidation of the N^3-nitrogen of 6-aminothymine generates 4-amino-1-(β-D-2-deoxyribofuranosyl)-5-methyl-2,6-[*1H,3H*]-pyrimidione ($m^{5ox}C$), while glycosidation of the N^1-nitrogen generates 6-aminothymidine.



(16) After resolution and deprotection of the nucleosides, ^1H NMR (DMSO-d_6) δ 6.11 (s, 2H, $-NH_2$, β-anomer), 6.17 (s, 2H, $-NH_2$, α-anomer).

(17) After deprotection of the nucleosides, ^1H NMR (DMSO-d_6) δ 7.09, 7.11 (s,s, 2H,2H, $-NH_2$, α + β-anomers).

(18) Müller, C. E. *Tetrahedron Lett.* **1991**, *32*, 6539−6540.

(19) (a) Winkley, M. W.; Robins, R. K. *J. Chem. Soc.* **1969**, 791−796. (b) Noell, C. W.; Robins, R. *J. Heterocycl. Chem.* **1964**, *1*, 34−41.

0002-7863/94/1516-11155$04.50/0

Table 1. Tm Values for Native and Analogue-Nucleoside Triplexes

```
          5' T T T X T T T T X T X T X T T
5' G C G C G A A A G A A A A G A G A G A A C C C G G
3' C G C G C T T T C T T T T C T C T C T T G G G C C
```

		triplex Tm values (°C)[a]				
entry	sequence	pH 6.4	pH 7.0	pH 7.5	pH 8.0	pH 8.5
1	X = C	39	32	20	16	14
2	X = m^5C	42	38	31	24	22
3	X = $m^{5ox}C$	28	28	27	27	26

[a] The duplex Tm values in all cases occurred between ~67 °C (pH 6.4) and ~64 °C (pH 8.5).

the two carbonyls, and in the present system, a similar preference was observed. Chromatographic separation of the anomeric N^3-glycosylated derivatives was easily achieved after reaction with *o*-nitrobenzenesulfenyl chloride, but this protecting group was not stable to the conditions of DNA synthesis. It was subsequently removed, and the exocyclic amino group was protected as the cyclic *N*-methyl-2-pyrrolidine amidine.[20] The amidine derivative[21] of the $m^{5ox}C$ was then converted to its (4,4'-dimethoxytrityl)phosphoramidite derivative by conventional procedures to yield a derivative suitable for DNA synthesis (see Figure 1b).

A number of assays of triplex stability were made using a 25-mer double-stranded target sequence and a series of 15-mer probes.[22] The triple helices formed from these sequences each contain four target G-C base pairs (Table 1). Triplex stability was assayed at five different pH values, including mildly acidic (pH 6.4), neutral (pH 7.0), and mildly basic (pH 7.5, 8.0, and 8.5) conditions. Absorbance vs temperature plots resulted in two transitions, with the higher transition occurring between 64 and 67 °C. Analysis of the duplex alone resulted in this same transition. The second transition was present at a lower temperature, with a T_m value dependent upon the nature of the triplex formed and the pH of the solution. This transition was interpreted to reflect cooperative denaturation of the triplex strand from the duplex target, and this observation is in agreement with other reports[8,9b] documenting triplex formation using relatively short oligodeoxynucleotides. An example of the two transitions observed for the $m^{5ox}C$-containing triplex at pH 8.5 is illustrated in Figure 2.

The T_m values for the C-containing 15-mer (entry 1, Table 1) decreased from 39.0 °C at pH 6.4 to only 14.0 °C at pH 8.5. Replacement of the four C residues by m^5C (entry 2, Table 1) resulted in more stable triplexes, as documented in the literature,[5] but the pH dependence of triplex formation was still readily apparent. By comparison, the $m^{5ox}C$-containing 15-mer (entry 3, Table 1) resulted in largely pH-independent transitions, varying by only about 2 °C over the range of pH values examined.

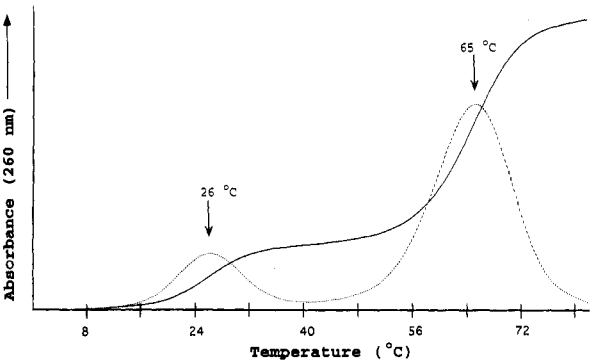


Figure 2. Absorbance vs temperature plot for the $m^{5ox}C$-containing triplex (see Table 1) in 25 mM Tris·HCl, pH 8.5, 10 mM $MgCl_2$, and 50 mM NaCl with a heating rate of 1 °C/min (solid line). First-order differential of the absorbance vs temperature plot (dashed line).

The T_m values for the C-containing and the m^5C-containing 15-mers, complexed to their Watson–Crick complementary sequence, were 44 and 48 °C, respectively (pH 7.0). No corresponding helix-to-coil transition was observed for the $m^{5ox}C$-containing strand, suggesting the presence of duplex destabilizing interactions. This observation is consistent with an $m^{5ox}C$ residue preferentailly adopting the N^3-H tautomeric form such that the N^3-hydrogen is opposed to the N^1-hydrogen of the complementary guanine residue in the Watson–Crick duplex.

The T_m values for the C-containing and m^5C-containing triplexes are significantly higher than those observed for the $m^{5ox}C$-containing triplex at pH 6.4 and 7.0. This observation likely reflects, in part, the differences in hydrogen-bonding characteristics. Each C (or m^5C) will form one hydrogen bond in the triplex complex that contains a charged H-bonding partner, while the $m^{5ox}C$ residues form only H-bonds involving uncharged partners. Charged hydrogen-bonding partners generally result in enhanced complex stability as has been documented for enzyme–substrate complexes[23] as well as for simpler systems.[24] As the pH value of the solution increases to 8.0 and 8.5, the extent of protonation of C (or m^5C) decreases, and the $m^{5ox}C$-containing triplex, with neutral H-bonds, dominates in stability. Differences in base stacking characteristics are also likely to account for variations in triplex stability, as described recently for the propynyl-modified U residues,[25] but stacking effects for $m^{5ox}C$ cannot be adequately interpreted at present.

Acknowledgment. This work was supported in part by a grant from the NSF (MCB-9206528). L.W.M. is the recipient of an American Cancer Society Faculty Research Award (FRA-384).

Supplementary Material Available: Text describing experimental methods (10 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

(20) (a) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *29*, 2953–2956. (b) McBride, L. J.; Kierzek, R.; Beaucage, S. L.; Caruthers, M. H. *J. Am. Chem. Soc.* **1986**, *108*, 2040–2048.

(21) Preparation of the amidine derivatives directly from the N^3 α- and β-anomeric nucleosides obtained from the initial coupling reaction did not result in effective diastereomeric separation.

(22) The 15-mers containing $m^{5ox}C$ were prepared by standard techniques (Matteucci, M. D.; Caruthers, M. *J. Am. Chem. Soc.* **1981**, *103*, 3185–3191). Coupling efficiencies for the analogue were comparable to those of common nucleoside phosphoramidites. Deprotection of the *N*-methyl-2-pyrrolidine amidine necessitated a 16 h reaction in concentrated aqueous ammonia at 60 °C. Oligomers were digested with S1 nuclease/calf intestinal alkaline phosphatase, and under standard HPLC conditions (see Mazzarelli et al. *Biochemistry* **1992**, *31*, 5925–5936), $m^{5ox}C$ eluted with a retention time of 20.2 min and T at 21.2 min (0–35% methanol over 3 h, pH 5.5).

(23) Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* **1985**, *314*, 235–238.

(24) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(25) (a) Froehler, B. C.; Wadwani, S.; Terhorst, T. J.; Gerrard, S. R. *Tetrahedron Lett.* **1992**, *33*, 5307–5310. (b) Colocci, N.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 785–786.