CHEMCOMM
Communication
www.rsc.org/chemcomm

# Incorporation of a novel nucleobase allows stable oligonucleotide-directed triple helix formation at the target sequence containing a purine·pyrimidine interruption

**Dominique Guianvarc'h,[a] Rachid Benhida,[a] Jean-Louis Fourrey,*[a] Rosalie Maurisse[b] and Jian-Sheng Sun*[b]**

[a] *Institut de Chimie des Substances Naturelles, CNRS, Avenue de la Terrasse 91198. Gif-sur-Yvette Cédex, France. E-mail: fourrey@icsn.cnrs-gif.fr; Fax: +33-1-69077247; Tel: +33-0169283053*
[b] *Laboratoire de Biophysique, UMR 8646 CNRS-Muséum National d'Histoire Naturelle, INSERM U201, 43 rue Cuvier 75231. Paris Cedex 05, France. E-mail: sun@mnhn.fr; Fax: +33-1-69077247; Tel: +33-01407937*

***Received (in Cambridge, UK) 25th April 2001, Accepted 21st July 2001***
***First published as an Advance Article on the web 23rd August 2001***

**Thermal denaturation experiments have established that an oligonucleotide incorporating the artificial nucleobase S, does form a stable triplex with a double stranded DNA which exhibits a pyrimidine interruption within the oligopurine sequence.**

For many years, triple helix-forming oligonucleotides (TFOs) have been known to be able to bind in the major groove of oligopyrimidine·oligopurine sequences of double-stranded DNA (dsDNA). TFOs establish specific hydrogen bonds with the oligopurine strand of dsDNA through the formation of T·A×T and C·G×C$^+$ base triplets[1] in Hoogsteen (pyrimidine-motif, Fig. 1) or T·A×A and T·A×T, and C·G×G triplets in reverse Hoogsteen configuration (purine- or mixed-motif).[2] Hence, dsDNA sequence recognition by TFOs has potential applications in gene expression modulation and in gene targeting technologies.[3] Unfortunately, these interesting applications must be restricted to long oligopyrimidine·oligopurine sequences only (> about 15 bp) since any interruption by even a single A·T or G·C base pair strongly destabilizes triple helix formation.[4] Consequently, during the last decade, considerable efforts have been devoted, so far with limited success, to circumvent such a sequence limitation.[5] Two approaches have been undertaken to design and synthesize: (1) new base analogs featuring an extended heterocyclic ring system for achieving specific hydrogen bonds with all hydrogen bond-forming sites available in the major groove side of the inverted A·T and G·C base pairs;[6] (2) nucleobases capable of stabilizing, non-sequence-specifically, triple helix formation at the purine·pyrimidine interruption sites, either by the attachment of an intercalating agent in conjunction with an appropriate nucleobase, or by a nucleobase alone.[7] The second strategy (universal base approach) has been more successful than the first (specific base approach). In the literature, a 4-(3-benzamidophenyl)imidazole (**$D_3$**) has been shown to equally stabilize the triplex at both A·T and G·C sites.[6a] Subsequent NMR studies showed that the binding mode is intercalation[6b] which is consistent with the lack of base pair discrimination.

In this work, a new base analog (**S**) has been synthesized and incorporated into a TFO in an attempt to achieve better triplex stabilization than the **$D_3$** nucleobase at the purine·pyrimidine sites within its cognate dsDNA sequence. Compared to **$D_3$** the **S** nucleobase consists of two unfused aromatic rings which are linked to 2′-deoxyribose by an acetamide motif instead of a three ring construct attached directly to 2′-deoxyribose (Fig. 2). The synthesis of the required phosphoramidite **5** to serve for the incorporation of **S** in TFOs is straightforward. Thus, compound **2**, readily obtained by catalytic hydrogenation of 2-acetamido-4-(3-nitrophenyl)thiazole **1**,[8] was acylated with 2-(2-deoxy-5-*O*-dimethoxytrityldeoxyribosyl)acetic acid **3**[9] using 2-chloro-1-methylpyridinium iodide. Finally, the resulting derivative **4** was phosphitylated in the usual manner to give the desired phosphoramidite **5** (Scheme 1).

The capacity of triple helix stabilization of the novel nucleobase **S** was assessed in a model system where **S** was incorporated in the middle of a 18-mer TFO (at position Z)[10] and was screened against all four possible base pairs (X·Y = T·A, C·G, A·T or G·C) in the target oligopyrimidine·oligopurine sequence (Table 1). The thermal denaturation experiments[11] indicated that the $T_m$ value of the triplex containing an A·T×**S** triplet ($T_m$ = 50 °C) was very close to those of perfect triplexes without any interruption of oligopyrimidine·oligopurine sequences (T·A×T or C·G×C$^+$, $T_m$ = 51 or 50 °C, respectively). It was noted that the use of **S** base provides a 5–8 °C triplex stabilization as compared to the best base triplet made of natural bases (A·T×**S** *vs.* A·T×G; G·C×**S** *vs.* G·C×T), respectively. In terms of triplex stability, the novel **S** nucleobase is at least as good as the previously reported **$D_3$** base. The main difference

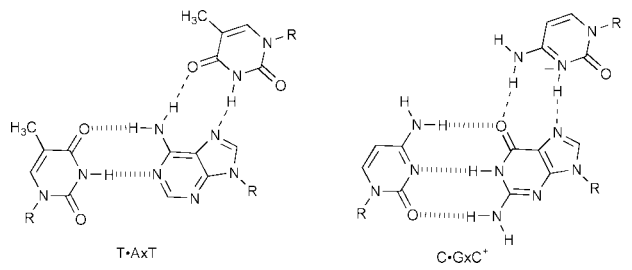



**Fig. 1** Canonical T·A×T and C·G×C$^+$ base triplets in Hoogsteen configuration (pyrimidine-motif). R = 2′-deoxyribosyl.

† Electronic supplementary information (ESI) available: experimental details. See http://www.rsc.org/suppdata/cc/b1/b103743a/

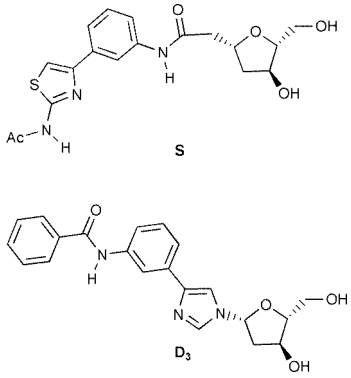
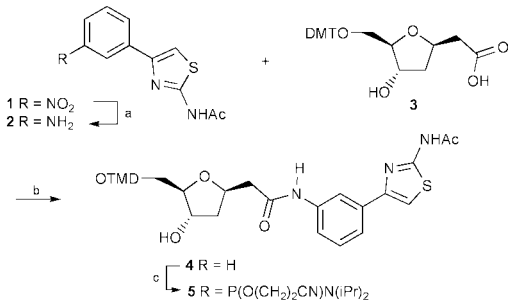



**Fig. 2** Structures of the 2′-deoxynucleosides featuring nucleobases **S** and **$D_3$**.

 DOI: 10.1039/b103743a




**Scheme 1** *Reagents and conditions*: (a) $H_2$, Pd/C, EtOH–AcOH, 95%. (b) 2-Chloro-1-methylpyridinium iodide, $NEt_3$, $CH_2Cl_2$, 60 °C, 90%. (c) 2-Cyanoethyl diisopropylchlorophosphoramidite, *N*,*N*-diisopropylethylamine, $CH_2Cl_2$, rt, 75%.

between **S** and $\mathbf{D_3}$ is the observation that the **S** nucleobase can discriminate, to some extent, an A·T base pair from others (namely A·T×**S** *vs*. G·C×**S**, T·A×**S** and C·G×**S**). It is worth noting that there is evidence that the aminothiazole moiety of **S** is involved in the A·T base pair recognition as the replacement of the heterocyclic moiety of **S** by aniline caused a 14 °C decrease in $T_m$ and abolished base pair discrimination. However, additional studies[12] are needed to obtain the definitive structural arguments which would validate the recognition pattern proposed in Fig. 3. In this scheme the triple between the A·T base pair and **S** is characterized by the establishment of three hydrogen bonds to the N7 atom and the 6-amino group of adenine, and to the 4-oxo group of thymine in a co-planar arrangement. According to this scheme, steric hindrance between the C5 methyl of T in the oligopurine-rich target strand and the deoxyribose moiety of **S** can be anticipated. Indeed, when T was replaced by U in the target strand, a further stabilization of the triplex featuring an A·U×**S** triplet was observed with the $T_m$ increasing by 3 °C. Such an observation is consistent with the proposed mode of recognition (Fig. 3).

In conclusion, this work shows that a novel **S** nucleobase when incorporated into a pyrimidine-motif TFO can effectively circumvent a purine·pyrimidine base pair interruption in an oligopyrimidine·oligopurine sequence. This outstanding property of **S** opens new perspective as it could be exploited as a lead compound, in both universal base or specific base approaches, to develop new nucleobases capable of achieving sequence-specific recognition of further extended dsDNA sequences by oligonucleotide-directed triple helix formation.

**Table 1** Sequence of the triple helices studied in this work. Melting temperature ($T_m$) of all combinations of base triplets at the X·Y×Z site. Estimated accuracy of the melting temperatures is ±1 °C

3′-CGTA-TTTTCTTCTCTT**X**TTCTT-AGTG-5′ **I**

5′-GCAT-AAAAGAAGAGAA**Y**AAGAA-TCAC-3′ **II**

5′-TTTTCTTCTCTT**Z**TTCTT-3′ **III**

| | Z | | | |
|---|---|---|---|---|
| X·Y | T | C | G | S |
| T·A | 51 | 31 | 31 | 42 |
| C·G | 40 | 50 | 31 | 41 |
| A·T | 33 | 33 | 45 | 50 |
| G·C | 38 | 35 | 35 | 46 |

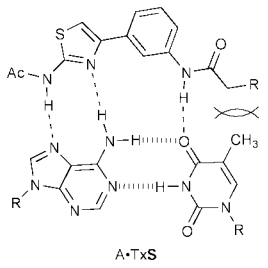



**Fig. 3** Model proposed for the specific recognition of an A·T base pair by nucleobase **S**. R = 2′-deoxyribosyl. Putative steric interaction in the plane of the A·T×**S** triple is indicated (see text).

## Notes and references


1 The symbols · and × stand for Watson–Crick and Hoogsteen-like hydrogen bonding, respectively.

2 (*a*) T. Le Doan, L. Perrouault, D. Praseuth, N. Habhoud, J. L. Decout, N. T. Thuong, J. Lhomme and C. Hélène, *Nucleic Acids Res*., 1987, **15**, 7749; (*b*) H. E. Moser and P. B. Dervan, *Science*, 1987, **238**, 645. For reviews see (*c*)–(*e*) (*c*) N. T. Thuong and C. Hélène, *Angew. Chem., Int. Ed. Engl*., 1993, **32**, 666; (*d*) M. T. Frank-Kamenetskii and S. M. Mirkin, *Annu. Rev. Biochem*., 1995, **64**, 65; (*e*) S. Neidle, *Anti-Cancer Drug Des*., 1997, **12**, 433.

3 (*a*) J. L. Maher III, *Cancer Invest*., 1996, **14**, 66; (*b*) C. Giovannangeli and C. Hélène, *Antisense Nucleic Acid Drug Dev*., 1997, **7**, 413; (*c*) K. M. Vasquez and J. H. Wilson, *Trends Biochem. Sci*., 1998, **23**, 4; (*d*) D. Praseuth, A. L. Guieysse-Peugeot and C. Hélène, *Biochim. Biophys. Acta*, 1999, **1489**, 181.

4 (*a*) J. L. Mergny, J. S. Sun, M. Rougée, T. Garestier, F. Barcelo, J. Chomilier and C. Hélène, *Biochemistry*, 1991, **30**, 9791; (*b*) W. A. Greenberg and P. B. Dervan, *J. Am. Chem. Soc*., 1995, **117**, 5016; (*c*) G. C. Best and P. B. Dervan, *J. Am. Chem. Soc*., 1995, **117**, 1187.

5 (*a*) Reviews: J. S. Sun and C. Hélène, *Curr. Opin. Struct. Biol*., 1993, **3**, 345; (*b*) S. O. Doronina and J. P. Behr, *Chem. Soc. Rev*., 1997, 63; (*c*) D. M. Gowers and K. R. Fox, *Nucleic Acids Res*., 1999, **27**, 1569 and references cited therein.

6 (*a*) L. C. Griffin, L. L. Kiessling, P. A. Beal, P. Gillespie and P. B. Dervan, *J. Am. Chem. Soc*., 1992, **114**, 7976; (*b*) K. M. Koshlap, P. Gillespie, P. B. Dervan and J. Feigon, *J. Am. Chem. Soc*., 1993, **115**, 7908; (*c*) C. Y. Huang, C. D. Cushman and P. S. Miller, *J. Org. Chem*., 1993, **58**, 5048; (*d*) C. Y. Huang, G. Bi and P. S. Miller, *Nucleic Acids Res*., 1996, **24**, 2606; (*e*) T. E. Lehmann, W. A. Greenberg, D. A. Liberles, C. K. Wada and P. B. Dervan, *Helv. Chim. Acta*, 1997, **80**, 2002; (*f*) I. Prévot-Halter and C. J. Leumann, *Bioorg. Med. Chem. Lett*., 1999, **9**, 2657.

7 (*a*) S. Kukreti, J. S. Sun, T. Garestier and C. Hélène, *Nucleic Acids Res*., 1997, **25**, 4264; (*b*) S. Kukreti, J. S. Sun, D. Loakes, D. M. Brown, C. H. Nguyen, E. Bisagni, T. Garestier and C. Hélène, *Nucleic Acids Res*., 1998, **26**, 2179; (*c*) D. A. Gianolio and L. W. McLaughlin, *J. Am. Chem. Soc*., 1999, **121**, 6334.

8 C. D. Hurd and N. Karasch, *J. Am. Chem. Soc*., 1946, **68**, 658.

9 J. Hovinen and H. Salo, *J. Chem. Soc., Perkin Trans. 1*, 1997, 3017.

10 The identity and homogeneity of the 18-mer was confirmed by MALDI-TOF [M calc. 5502.8; found *m*/*z* 5501.5 $(M - H)^-$] and RP-HPLC.

11 DNA thermal denaturation and renaturation experiments were carried out by first mixing **I** and **II** strands (1.2 and 1.0 μM, respectively), then adding 1.5 μM of TFO (**III**) in a 10 mM cacodylate buffer (pH 5.9) containing 100 mM NaCl, 10 mM $MgCl_2$ and 0.5 mM spermine.

12 A comprehensive NMR study is under way with an intramolecular system consisting of a 31-mer in order to establish the recognition mode either within the A·T×**S** triple and/or between the adjacent bases. Results will be reported in due course.