DNA Recognition

Stable and Selective Recognition of Three Base Pairs in the Parallel Triple-Helical DNA Binding Motif**

*Sabrina Buchini and Christian J. Leumann**

Selective artificial control of gene expression is a longstanding dream of researchers in biotechnology and human therapy. Oligonucleotides seem perfectly suited to this purpose because of their unique base–base recognition properties. The triplex approach uses oligonucleotides that sequence-selectively bind genomic, double-stranded DNA and interfere with transcription and the DNA processing machinery through formation of a triple helix. This approach allows not only transcriptional activation and deactivation but also gene knock-out, targeted mutagenesis, targeted recombina-

[*] Dipl.-Chem. S. Buchini, Prof. C. J. Leumann
Department of Chemistry and Biochemistry
University of Bern, Freiestrasse 3
3012 Bern (Switzerland)
Fax: (+41) 31-631-4355
E-mail: leumann@ioc.unibe.ch

[**] The authors thank Christine Brotschi, Alain Mayer, and Dorte Renneberg for advice on experimental procedures. Financial support from the Swiss National Science Foundation (Grant no. 200020-100178) is gratefully acknowledged.

 DOI: 10.1002/ange.200460159

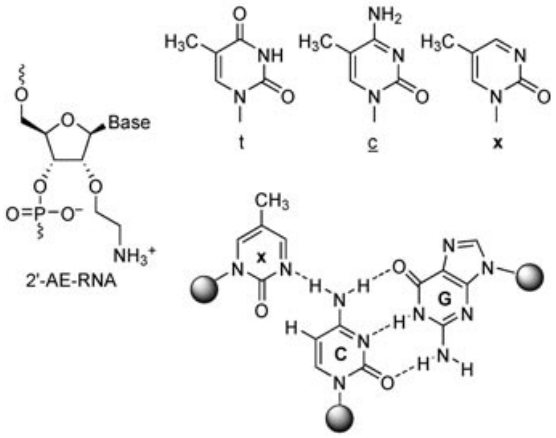
tion, and sequence-selective manipulation of genomic DNA.[1–3]

In the parallel or pyrimidine binding motif a homopyrimidine triplex-forming oligonucleotide (TFO) oriented parallel to the purine strand of a DNA molecule binds to the major groove of the target duplex by forming Hoogsteen C^+·G–C and T·A–T base triples. However, formation of the C^+·G–C base triple is pH sensitive and disfavored at neutral pH. Furthermore, recognition of duplex DNA is constrained to homopurine/homopyrimidine sequence tracts, and the third-strand affinities of the oligonucleotides to their targets are generally low. These factors severely restrict the applicability of TFOs in vitro and in vivo.

Numerous attempts have been made to chemically modify TFOs to remove these constraints, but with limited success.[4] Conformationally constrained locked nucleic acids (LNAs)[5–7] and 2′-aminoethyl-oligoribonucleotides (2′AE-RNAs)[8,9] are milestones in this field. The latter molecules recognize duplex DNA not only through base–base contacts but also by formation of a salt bridge between the protonated amino function of the oligonucleotide and a phosphodiester unit of the target purine strand (dual recognition). Both modifications proved to dramatically increase TFO affinity for duplex targets.

It seems plausible that a combination of general-affinity-enhancing modifications of the TFO sugar units or backbone and inclusion of new base analogues that selectively recognize pyrimidine bases could alleviate restriction of the target sequence. This idea has been explored with remarkable success by Imanishi and co-workers, who used LNAs containing a 2-pyridone base designed to selectively recognize a C–G base pair.[10,11] We recently investigated selective C–G recognition by oligodeoxynucleotides containing single 2′-aminoethylribonucleosides with a 5-methyl-2-pyrimidinone base (Scheme 1, **x**).[12] Herein we present experimental evidence that such a nucleoside within the context of a fully modified 2′AE-RNA is able to recognize up to five C–G inversions in a 15-mer DNA target duplex (33% pyrimidine content) with high selectivity and a third-strand affinity matching or even excelling the affinity of the two strands of the target duplex for one another.

The synthesis of the modified nucleoside **x**[12] and the corresponding fully modified AE-TFOs by standard phosphoramidite chemistry will be described in detail elsewhere. All TFOs contained a deoxynucleoside unit at their 3′ end for synthetic ease. All TFOs were analyzed by ESI-MS and found to have masses corresponding to the expected values. To analyze the effect of up to five 5-methyl-2-pyrimidinone (**x**) substitutions in a 15-mer we prepared the TFOs and duplex targets listed in Table 1 and measured the corresponding T_m values by standard UV/melting-curve techniques. The melting curves corresponding to entry 1 are shown in Figure 1 as a representative example.



Scheme 1. Top: Chemical structure of a 2′-aminoethylribonucleotide unit and the bases thymine (t), 5-methylcytosine (c), and 5-methyl-2-pyrimidone (**x**). Bottom: Anticipated **x**·C–G base triplet.

Table 1: T_m data for third-strand melting, taken from UV/melting curves ($\lambda = 260$ nm).[a] T_m values for duplex melting are indicated in parentheses.

Entry	Triplex sequence	T_m value of the third strand (duplex) [°C]			
		Y/Z: C/G	G/C	A/T	T/A
1	5′-GCTAAAAAGA**Y**AGAGAGATCG-3′	77	66	66	65
	3′-CGATTTTTCT**Z**TCTCTCTAGC-3′	(63)	(62)	(61)	(61)
	5′-ttttt**c**t**x**t**c**t**c**t**c**T-3′				
2	5′-GCTAAAAA**Y**A**Y**A**Y**AGAGATCG-3′	67	26	26	<10
	3′-CGATTTTT**Z**T**Z**T**Z**TCTCTAGC-3′	(64)	(62)	(60)	(58)
	5′-ttttt**x**t**x**t**x**t**c**t**c**T-3′				
3	5′-GCTAA**Y**AA**Y**A**Y**A**Y**AGA**Y**ATCG-3′	67	n.d.[b]	n.d.[b]	n.d.[b]
	3′-CGATT**Z**TT**Z**T**Z**T**Z**TCT**Z**TAGC-3′	(67)	(63)	(59)	(55)
	5′-tt**x**tt**x**t**x**t**x**t**c**t**x**T-3′				

[a] $c = 1.6$ μM in 10 mM sodium cacodylate, 100 mM NaCl, 0.25 mM spermine, pH 6.75 ± 0.25. [b] No T_m value could be determined for third-strand melting.

As can be seen from the T_m data in Table 1, triplex formation occurs preferentially at the C–G inversion site of the DNA duplex. In the case of a single substitution (entry 1), all other arrangements are less stable and have T_m values at least 11 °C lower than that observed with a C–G inversion site. All the triplexes containing one **x** unit melt at higher temperatures than the underlying Watson–Crick duplexes. Triple substitution of the target duplex (entry 2) results in increased destabilization at non-C–G sites, as characterized by a difference in the T_m value of more than 40 °C between molecules with C–G sites and those without. Again, the thermal stability of the triplex formed by third (Hoogsteen)-strand binding to the purine C–G target at neutral pH is higher than that of the Watson–Crick duplex. The melting curves of the triplexes with five substitutions (Table 1, entry 3) revealed only one transition in each case, with T_m values in the same range as those of the underlying duplexes. It was difficult to conclude from this T_m analysis alone whether triplex formation occurred at all in the presence of five substitutions. Clarification came from a gel

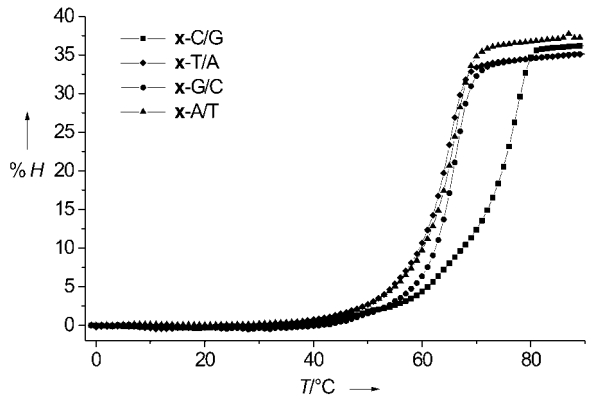


Figure 1. UV/melting curves (260 nm) of the triplexes listed in Table 1, entry 1. Experimental conditions are described in Table 1.

shift assay (Figure 2), which clearly indicated the presence of a triplex containing a C–G inversion site and the absence of triplex formation by all other duplexes. We thus conclude that

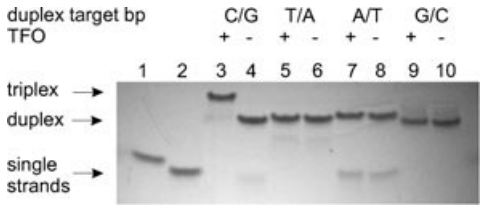


Figure 2. Nondenaturing 20% polyacrylamide gel of triplexes with five **x** substitutions (Table 1, entry 3). Lanes 1 and 2: single-strand controls (the strands correspond to those present in the target duplexes); lanes 3–10: stoichiometric mixtures (0.8 nmol) of the strands of the target duplexes, with and without TFO (2 equiv). Buffer: tris(hydroxymethyl)aminomethane borate (90 mM), $MgCl_2$ (10 mM), pH 7.0, 4 °C. Visualization by exposure to UV light.

the transitions observed in the melting curves correspond to triplex-to-single-strand melting in the case of the C–G target duplex and duplex-to-single-strand transitions in all other cases. The higher hyperchromicity of the transition with the C–G target duplex compared to the hyperchromicities of the transitions involving the other targets corroborates this interpretation. The same gel experiment was also conducted with the triplexes given in entry 2 of Table 1 (data not shown). Again, triplex formation was only observed with the C–G duplex target.

The next step was to explore different sequences. We prepared a further series of TFOs and target duplexes with the sequences given in Table 2 and determined the corresponding T_m values for the melting of the triplex. In the first sequence context investigated (Table 2, entry 1), three **x** units interrupt a stretch of six contiguous methylcytidine units in the TFO. Use of this sequence, rich in contiguous c nucleotides, leads to triplexes with a lower stability than the target duplex. When the **x** units are placed between thymidine and methylcytidine residues and contiguous c residues are avoided (Table 2, entry 2), the third strand affinity is similar to that observed when **x** is located between two thymidine residues (for example, Table 1, entry 2). The reason for the decreased affinity in the first case seems to be local repulsion of the accumulated positive charges in the pile of protonated c residues rather than the effect of the modification **x**. Sequence effects on AE-RNA/DNA interactions within the context of canonical base recognition (t·A-T, c^+·G-C) have not been addressed so far. However, it is well-known that contiguous cytosine units in a TFO decrease the affinity for the third strand,[13] both for RNA and for DNA TFOs. Three contiguous **x** residues in an AE-TFO (Table 2, entry 3) lead to a strongly bound triplex, which corroborates the view that not the modification **x** but the contiguous protonated cytosine sequence is mostly responsible for the observed reduction in stability.

The triplex stabilities found for the systems described herein are remarkable given that **x** recognizes the cytosine base by forming only one hydrogen bond, possibly assisted by a favorable electrostatic C–H···O interaction (Scheme 1). These stabilities suggest that the dual recognition scheme, in which the protonated ammonium function on the 2′-side chain of the third strand unselectively recognizes a phosphate oxygen atom on the purine strand of the target duplex, can energetically replace the loss of a hydrogen bond at the base interface without a loss of overall selectivity. This hypothesis seems plausible since formation of a salt bridge is strongly distance-dependent and would not tolerate the structural isomorphism expected in base mismatching. Indeed, we have previously found that the presence of an **x** unit in a TFO with a DNA backbone does not lead to an increase in stability such as that observed in the context of the fully modified AE-TFOs, which points to an energetically less significant contribution of the implied salt bridge to triplex stability.[12] The lack of impact made by formation of a salt bridge on stability is probably the result of differences in third-strand conformation (DNA versus RNA) and thus differences in the distance between the relevant phosphodiester backbones in the triplex.

Within the context of 2′-*O*-aminoethyl-oligoribonucleotides, the unit **x** fulfills the criteria of selectivity and high affinity for C–G inversion sites and expands the recognition

Table 2: T_m data for third-strand melting, taken from UV/melting curves ($\lambda = 260$ nm).[a]

Entry	Triplex sequence	T_m duplex [°C]	T_m value of third strand [°C]
1	5′-GCTAAAG**C**GG**C**GG**C**GAAATCG-3′ 3′-CGATTTC**G**CC**G**CC**G**CTTTAGC-3′ 5′-ttt**x**cc**x**cc**x**cttT-3′	73	20[b]
2	5′-GCTGGAG**C**AG**C**AG**C**AGAATCG-3′ 3′-CGACCTC**G**TC**G**TC**G**TCTTAGC-3′ 5′-cct**x**tc**x**tc**x**tctT-3′	72	ca. 60[c]
3	5′-GCTGAAGA**CCC**GAGGAGATCG-3′ 3′-CGACTTCT**GGG**CTCCTCTAGC-3′ 5′-ctt**xxx**ctcctcT-3′	71	ca. 62[c]

[a] $c = 1.6$ μM in 10 mM sodium cacodylate, 100 mM NaCl, 0.25 mM spermine, pH 6.5. [b] pH 6.0. [c] Third strand and duplex melting partially overlaid.

scheme in the parallel triplex binding motif from two to three duplex base pairs. It now remains to define the limitations of the binding of **x** to target DNA and to show whether **x** can be used productively for selective recognition and manipulation of cellular targets, for example in applications such as targeted gene knock-out.[14,15] It will be a challenge for future research to find new bases that, within the context of AE-TFOs, selectively recognize T–A inversion sites.

Received: March 30, 2004 [Z460159]

Keywords: antigenes · DNA recognition · DNA triplex · intercalation · oligonucleotides

[1] R. V. Guntaka, B. R. Varma, K. T. Weber, *Int. J. Biochem. Cell Biol.* **2003**, *35*, 22 – 31.
[2] M. M. Seidman, P. M. Glazer, *J. Clin. Invest.* **2003**, *112*, 487 – 494.
[3] S. Buchini, C. J. Leumann, *Curr. Opin. Chem. Biol.* **2003**, *7*, 717 – 726.
[4] J. Robles, A. Grandas, E. Pedroso, F. J. Luque, R. Eritja, M. Orozco, *Curr. Org. Chem.* **2002**, *6*, 1333 – 1368.
[5] S. Obika, Y. Hari, T. Sugimoto, M. Sekiguchi, T. Imanishi, *Tetrahedron Lett.* **2000**, *41*, 8923 – 8927.
[6] S. Obika, T. Uneda, T. Sugimoto, D. Nanbu, T. Minami, T. Doi, T. Imanishi, *Bioorg. Med. Chem.* **2001**, *9*, 1001 – 1011.
[7] H. Torigoe, Y. Hari, M. Sekiguchi, S. Obika, T. Imanishi, *J. Biol. Chem.* **2001**, *276*, 2354 – 2360.
[8] M. J. J. Blommers, F. Natt, W. Jahnke, B. Cuenoud, *Biochemistry* **1998**, *37*, 17714 – 17725.
[9] B. Cuenoud, F. Casset, D. Hüsken, F. Natt, R. M. Wolf, K.-H. Altmann, P. Martin, H. E. Moser, *Angew. Chem.* **1998**, *110*, 1350 – 1353; *Angew. Chem. Int. Ed.* **1998**, *37*, 1288 – 1291.
[10] S. Obika, Y. Hari, M. Sekiguchi, T. Imanishi, *Angew. Chem.* **2001**, *113*, 2149 – 2151; *Angew. Chem. Int. Ed.* **2001**, *40*, 2079 – 2081.
[11] S. Obika, Y. Hari, M. Sekiguchi, T. Imanishi, *Chem. Eur. J.* **2002**, *8*, 4796 – 4802.
[12] S. Buchini, C. J. Leumann, *Tetrahedron Lett.* **2003**, *44*, 5065 – 5068.
[13] S. Hildbrand, A. Blaser, S. P. Parel, C. J. Leumann, *J. Am. Chem. Soc.* **1997**, *119*, 5499 – 5511.
[14] N. Puri, A. Majumdar, B. Cuenoud, F. Natt, P. Martin, A. Boyd, P. S. Miller, M. M. Seidman, *J. Biol. Chem.* **2001**, *276*, 28991 – 28998.
[15] N. Puri, A. Majumdar, B. Cuenoud, F. Natt, P. Martin, A. Boyd, P. S. Miller, M. M. Seidman, *Biochemistry* **2002**, *41*, 7716 – 7724.

