

Modification of DNA duplexes to smooth their thermal stability independently of their base content for DNA sequencing by hybridization

Hong-Khanh Nguyen, Pascal Auffray⁺, Ulysse Asseline, Daniel Dupret¹ and Nguyen Thanh Thuong*

Centre de Biophysique Moléculaire, CNRS, rue Charles Sadron, 45071 Orléans Cedex 02, France and
¹Appligène-Oncor, Parc d'Innovation, BP 72, 67407 Illkirch, France

Received April 14, 1997; Revised and Accepted June 9, 1997

ABSTRACT

The possibility of equalizing DNA duplex stability is essential for the application of sequencing by hybridization. In this paper we describe a new strategy to obtain DNA duplexes with a thermal stability independent of their base content. Modified *C bases have been developed and incorporated into oligonucleotides. The influence of these modifications on duplex stability has been studied by absorption spectroscopy, thus allowing selection of *N*-4-ethyl-2'-deoxycytidine (d^{4Et}C), which hybridizes specifically with natural dG to give a G^{4Et}C base pair whose stability is very close to that of natural AT base pairs. Duplexes built with AT and/or G^{4Et}C base pairs exhibit thermal stabilities independent of their base content in a classical buffer solution, thus enabling control of the stability of DNA hybrids as a function of their length only.

INTRODUCTION

Specific duplex formation between an oligonucleotide and a DNA sequence is the foundation of nucleic acid analysis and enzymatic labeling of DNA fragments. The success of these techniques implies absolute discrimination of perfect hybrids from ones containing mismatches to avoid false positive results and the design of hybrids having equivalent stabilities, which leads to homogeneous results.

The recently reported reverse hybridization technique is based on detection of perfect hybrids formed between one or several labeled DNA fragments with every oligonucleotide of a given length in a complete set immobilized as a two-dimensional matrix (1–3). This method is based on the possibility of discriminating perfect hybrids from those containing mismatches. This can be performed by increasing the temperature to allow dissociation of hybrids with mismatches before those without mismatches. Reverse hybridization could be a method for the design of a fast analysis technique for large DNA sequences. However, the approaches described in the literature using natural oligonucleotides serving as probes have a serious drawback, due to the base composition dependence of duplex stability. It is well known that

a GC base pair with three hydrogen bonds is more stable than AT or AU base pairs, which have only two hydrogen bonds. A perfect hybrid built with AT-rich sequences would therefore have a similar or even lower stability than do hybrids built with GC-rich sequences involving one mismatch. This leads to false positive or false negative signals depending on the hybridization temperature and washing conditions.

Several techniques have been studied to alleviate this problem but none of them have been successfully completed. Hybridization studies using TMAcI to reduce differential DNA duplex stability according to the base composition have been described and developed by several groups (4–8). However, this process requires multimolar concentrations of TMAcI, which is very viscous, leading to manipulation difficulties, and is not adapted to biochemical procedures involving enzymes, such as random priming or LCR, which are carried out under weak ionic concentrations (9). The proposed concentration variation of each particular oligonucleotide (3) or changing of the oligonucleotide length as a function of its base composition would be technically complex and impractical when applied to a large number of oligonucleotide sequences. Another option, the use of modified bases such as 5-ClIdU and 2-NH₂dA (10), has been examined, but with a great variation in stability. To reduce the stability of GC-rich duplexes, oligonucleotides were built with ribo- and deoxyribonucleotides, but the amount of hybrid obtained greatly decreased as the number of ribo↔deoxyribo transitions increased. Moreover, hybrid thermal dissociation spread over a wide temperature range, making it difficult to discriminate between perfect hybrids and those containing a mismatch (11).

In this paper we describe a new strategy for obtaining DNA duplexes whose thermal stabilities in NaCl solution are independent of their base content. Our approach consists of modification of one of the four natural deoxynucleosides which forms, with the complementary nucleoside, a base pair whose stability is very close to that of the other base pair. To achieve this end, we chose to modify 2'-deoxycytidine (d*C), which hybridizes specifically with natural 2'-deoxyguanosine (dG) to give a G*C base pair having a stability very similar to that of the AT base pair. We describe in this paper the preparation of modified oligonucleotides

*To whom correspondence should be addressed. Tel: +33 2 38 25 55 97; Fax: +33 2 38 63 15 17; Email: asseline@cns-orleans.fr

⁺Present address: Appligène-Oncor, Parc d'Innovation, BP 72, 67407 Illkirch, France

and selection of $d^{4Et}C$, which gives a $G^{4Et}C$ base pair having a stability similar to that of the AT base pair.

MATERIALS AND METHODS

All reagents and solvents were of reagent grade quality and used without further purification. Cytosine arabinoside phosphoramidite was obtained from Glen Research. Unmodified oligonucleotides were from Appligène-Oncon. Absorption spectra were recorded on a UVIKON 860 (Kontron). Absorption studies were carried out on a UVIKON 941 cell changer spectrophotometer (Kontron). Analytical TLC was carried out on Merck 5554 Kieselgel 60F 254 plates and eluted with various eluents: system A, $CH_2Cl_2/MeOH$ (95:5 v/v); system B, $CH_2Cl_2/MeOH$ (90:10 v/v); system C, $CH_2Cl_2/AcOEt/Et_3N$, (45:45:10 v/v/v). Merck 9385 Kieselgel 60 was used for column chromatography. All 4,4-dimethoxytrityl-containing substances were identified as orange colored spots on TLC plates by spraying with 10% perchloric acid solution. HPLC was performed on a Waters 626 E (system controller) equipped with a Waters 996 photodiode array detector. Analysis and purification by ion exchange chromatography were performed with a FPLC apparatus (Pharmacia). NMR experiments were carried out on a Bruker AM 300 WB spectrometer.

Synthesis of modified nucleosides 1

5'-O-Dimethoxytrityl-N-4-ethyl-2'-deoxycytidine **1b**. 1,2,4-Triazole (0.047 mol, 3.1 g) was suspended in anhydrous CH_3CN (50 ml) at 0°C followed by addition of $POCl_3$ (0.01 mol, 0.98 ml) with rapid stirring. Triethylamine (0.052 mol, 7.3 ml) was then added dropwise to the slurry stirred at 0°C for an additional 30 min. *5'-O-Dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-2'-deoxyuridine* (0.031 mol, 2 g), obtained as described in the literature (12), dissolved in 6 ml CH_3CN was added dropwise at 0°C. The ice/water bath was removed and the mixture allowed to react with magnetic stirring for 4 h at room temperature. An ethylamine solution (0.15 mol, 16 M in CH_3CN) was added directly to the crude derivative triazolyl obtained above. The reaction was monitored by TLC analysis. The reaction was complete after 2 h. The solution was concentrated under reduced pressure, solubilized in dichloromethane and washed with 5% sodium hydrogen carbonate solution. After being dried over Na_2SO_4 and concentrated to dryness, the residue was purified on a silica gel column using $CH_2Cl_2/MeOH/Et_3N$ (98:1:1 to 97:2:1 v/v/v) as eluent. Yield 81% (1.7 g), R_f system A 0.26, R_f system B 0.7.

The syntheses of *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-methyl-2'-deoxycytidine*, *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-propyl-2'-deoxycytidine*, *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-allyl-2'-deoxycytidine* and *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-propargyl-2'-deoxycytidine* were carried out as described above by replacing ethylamine by methylamine, propylamine, allylamine and propargylamine respectively (13).

5'-O-Dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-ethyl-2'-deoxycytidine obtained above (0.0025 mol, 1.7 g) was treated with a 1 M tetrabutylammonium fluoride solution (0.005 mol, 5 ml) in THF at room temperature. The reaction, monitored by TLC, was complete after 2 h. The reaction mixture was concentrated under reduced pressure and the residue dissolved in CH_2Cl_2 and washed with 5% sodium hydrogen carbonate solution. After being dried over Na_2SO_4 and concentrated to dryness the

obtained residue was purified on a column of silica gel using $CH_2Cl_2/MeOH$ (98:2 to 95:5 v/v v/v) as eluent. Yield **1b** 78% (1.1 g), R_{f1b} system B 0.55. 1H NMR (DMSO): δ 1.1 (t, 3H, -NH- CH_2 - CH_3), 2 (m, 1H, $H_{2''}$), 2.2 (m, 1H, $H_{2'}$), 3.2 (m, 2H, -NH- CH_2), 3.2–3.3 (m, 2H, $H_{5'}$, $H_{5''}$), 3.7 (s, 6H, -O- CH_3), 3.9 (m, 1H, $H_{4'}$), 4.3 (m, 1H, $H_{3'}$), 5.1 (d, 1H, H_5), 6.2 (m, 1H, $H_{1'}$), 6.8–7.3 (m, 18H Ph), 7.6 (d, 1H, H_6), 7.6 (t, 1H, NH), s: singlet, d: doublet, t: triplet, m: multiplet, Ph: phenyl.

Compounds **1a**, **1c**, **1d** and **1e** were obtained as described for compound **1b** starting from *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-methyl-2'-deoxycytidine*, *5'-O-dimethoxytrityl-1-3'-O-(t-butylidimethylsilyl)-N-4-propyl-2'-deoxycytidine*, *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-allyl-2'-deoxycytidine* and *5'-O-dimethoxytrityl-3'-O-(t-butylidimethylsilyl)-N-4-propargyl-2'-deoxycytidine* respectively. Yield **1a** 78%, R_{f1a} 0.53; yield **1c** 84%, R_{f1c} 0.59; yield **1d** 90%, R_{f1d} 0.63; yield **1e** 72%, R_{f1e} 0.64 system B.

Synthesis of phosphoramidite derivatives 2

5'-O-Dimethoxytrityl-3'-O-(2-cyanoethyl-N,N-diisopropylamido-phosphite)-N-4-ethyl-2'-deoxycytidine **2b**. Compound **1b** was dried by several co-evaporations and left in a desiccator overnight. 2-Cyanoethyl-N,N-diisopropylamidochlorophosphite (0.53 mmol, 0.12 ml) was added dropwise under argon atmosphere to a magnetically stirred mixture of compound **1b** (0.35 mmol, 200 mg) and diisopropylethylamine (1.43 mmol, 0.24 ml) in anhydrous dichloroethane (4 ml) at room temperature. The phosphorylation reaction was monitored by TLC analysis. After 1 h the reaction mixture was diluted with ethylacetate and washed with 10% sodium hydrogen carbonate solution, then with saturated sodium chloride solution. The organic solution was dried over sodium sulfate and concentrated under reduced pressure. The residue was purified on a silica gel column using $CH_2Cl_2/AcOEt/Et_3N$ (60:30:10 v/v/v) as eluent. The collected fractions containing the phosphoramidite compound **2b** were pooled and concentrated under reduced pressure. Compound **2b** was then obtained as a white powder after precipitation in cold hexane (-70°C). This compound was isolated by quick filtration and then dried in a desiccator. Yield **2b** 48% (m = 130 mg), R_{f2b} system A 0.35 and 0.33, R_{f2b} system C 0.68 and 0.61.

The phosphoramidites **2a**, **2c**, **2d** and **2e** were obtained starting from modified nucleosides **1a**, **1c**, **1d** and **1e** respectively, following the procedure described for preparation of **2b**. Yield **2a** 46%, R_{f2a} system A 0.36 and 0.32, R_{f2a} system C 0.52 and 0.43; yield **2c** 47%, R_{f2c} system A 0.39 and 0.34, R_{f2c} system C 0.73 and 0.68; yield **2d** 42%, R_{f2d} system D 0.35 and 0.30, R_{f2d} system C 0.72 and 0.67; yield **2e** 49%, R_{f2e} system A 0.35 and 0.31, R_{f2e} system C 0.75 and 0.69.

Synthesis purification and characterization of modified oligonucleotides

Chain assembly was carried out on a Pharmacia Gene Assembler on solid support CPG (controlled pore glass) functionalized with a nucleoside using phosphoramidite chemistry (14). Syntheses were performed on a 1 μ mol scale using 10 μ mol commercial phosphoramidite or modified phosphoramidite, prepared as previously described, per cycle with a cycle time of 10 min and a coupling time of 1.5 min for the commercial phosphoramidite.

The coupling time of the modified phosphoramidite was increased to 2.5 min. The coupling yields for modified dC phosphoramidites were as good as for the natural ones. The oligonucleotides obtained were then deblocked by treatment with concentrated ammonia overnight at 60°C. After deprotection and extraction of the organic impurities, the crude reaction products were purified by ion exchange chromatography using a Pharmacia FPLC system equipped with a DEAE column (8 µM, 100 × 10 mm) from Waters. A 25 mM Tris-HCl buffer, pH 8, in acetonitrile/water (10:90 v/v) with a linear gradient of NaCl from 0 to 0.45 M over 40 min at 1 ml/min was used as eluent. The fractions were monitored by absorption at 254 nm. $R_t \approx 25$ min for all compounds.

After desalting, the purity of all oligonucleotides described was checked by reverse phase analysis using a Lichrocart system (125 × 4 mm) packed with 5 µm Lichrospher RP 18 from Merck with a linear gradient of acetonitrile from 5 to 20% for 20 min in 0.1 M aqueous triethylammonium acetate buffer, pH 7, with a flow rate of 1 ml/min. The retention times (R_t) of oligonucleotides 5'-d(CG-AYGACGA)-3' involving one modified C at position 4 were as follows: $Y = {}^4\text{MeC}$, R_t 12.1 min; $Y = {}^4\text{EtC}$, R_t 12.2 min; $Y = {}^4\text{PrC}$, R_t 13 min; $Y = {}^4\text{allylC}$, R_t 12.8 min; $Y = {}^4\text{propargylC}$, R_t 12.8 min.

Full deprotection and nucleoside composition of the modified oligonucleotides were ascertained by nuclease degradation. An aliquot of oligonucleotide was digested with snake venom phosphodiesterase (Pharmacia Biotech) and alkaline phosphatase (Boehringer) in 0.1 M Tris-HCl, pH 8.2, for 19 h at room temperature. After inactivation of the enzyme at 90°C for 2 min, the digestion products were analyzed by reverse phase chromatography using a Lichrocart system (125 × 4 mm) packed with Nucleosil 100-5 C18 from Macheray Nagel equilibrated with 0.1 M aqueous triethylammonium acetate buffer, pH 7. The column was eluted at a flow rate of 1 ml/min with 0.1 M aqueous triethylammonium acetate buffer, pH 7, for 15 min and then with a linear gradient of 0–20% acetonitrile in 0.1 M aqueous triethylammonium acetate buffer, pH 7, for 40 min. Detection was performed at 260 nm. All oligonucleotides were totally degraded to nucleosides. In each case four peaks were obtained. Comparison with natural and modified nucleoside samples allowed us to identify the different peaks. Three of them whose retention times were identical for every oligonucleotide hydrolyzed correspond to dC (R_t 3.5 min), dI (R_t 9.1 min) and dG (R_t 11.5 min) respectively. The presence of dI, resulting from deamination of dA, could be due to contamination of the nucleases by adenosine deaminase. Peaks corresponding to the modified dC nucleosides were eluted with the following retention times: ${}^4\text{MeC}$, R_t 7.1 min; ${}^4\text{EtC}$, R_t 14.9 min; ${}^4\text{PrC}$, R_t 25.9 min; ${}^4\text{allylC}$, R_t 19.8 min; ${}^4\text{propargylC}$, R_t 12.6 min.

Melting experiments

Changes in absorbance with temperature of 2 µM duplexes in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA were measured at $\lambda_{260 \text{ nm}}$ in a UVIKON 941 cell changer spectrophotometer equipped with a Huber PD 415 temperature programmer connected to a cryothermostat ministat circulating water bath (Huber). Samples and references were slowly heated at a rate of 0.5°C/min from 0 to 80°C. Melting temperatures (T_m) were taken as the temperature corresponding to half-dissociation of the complexes. The T_m values were determined using the first and second derivatives. The molar extinction coefficients of the sequences were determined as described in the literature (15).

RESULTS AND DISCUSSION

Design and synthesis of modified oligonucleotides

Several approaches can be used to abolish differential binding stability of the hybrids dependent on base composition, among which is modification of one of the four natural 2'-deoxynucleosides, which then forms, with the complementary nucleoside, a base pair showing a stability very close to that of other base pairs. We can also modify both 2'-deoxynucleosides in a base pair to obtain a stability similar to that of the other natural base pairs. Another approach is based on modification of two non-complementary 2'-deoxynucleosides which hybridize with the complementary nucleoside to give base pairs with similar stabilities.

These investigations allow the design of one modified 2'-deoxycytidine, d*C, which hybridizes specifically with natural dG to give a G*C base pair whose stability is very similar to that of an AT base pair. This choice was dictated by the following criteria: it is easier to find a modified GC base pair whose stability is similar to that of an AT natural base pair than to design a modified AT base pair whose stability is close to that of a GC natural base pair; preparation of oligonucleotides containing dC analogs is simpler than that of oligonucleotides built with dG analogs; modification of only one base pair rather than both simplifies the enzymatic preparation of DNA containing one or several modified nucleosides.

We have chosen as dC analogs the araC and dC derivatives, in which one hydrogen of the exocyclic amino group at position 4 is substituted by an alkyl group such as methyl, ethyl, *n*-propyl, allyl or propargyl groups. In fact, it is well known that replacement of one dC by ${}^{\text{ara}}\text{C}$ (16) or ${}^{\text{d}}{}^4\text{MeC}$ (17) in an oligonucleotide induces a decrease in thermal stability of the hybrid formed by this oligonucleotide and the complementary nucleic acid sequence.

Synthesis of the N-4 substituted 2'-deoxycytidine

The synthesis of 5'-*O*-dimethoxytrityl-3'-*O*-(2-cyanoethyl-*N,N*-diisopropylamidophosphite)-*N*-4 substituted-2'-deoxycytidine (**2a**, **2b**, **2c**, **2d** and **2e**) was carried out from commercial deoxyuridine as described in Figure 1. It consists first of protection of the 5'- and 3'-hydroxyl functions of deoxyuridine by dimethoxytrityl and *t*-butyldimethylsilyl groups respectively (12). Conversion of 2'-deoxyuridine to *N*-4 substituted-2'-deoxycytidine was realized by activation at the C₄ position of the protected 2'-deoxyuridine by treatment with phosphorus oxychloride in the presence of 1,2,4-triazole followed by treatment with the primary amines (13). After deprotection of the 3'-hydroxyl function by the action of tetrabutylammonium fluoride, phosphorylation at the 3'-position was achieved using 2-cyanoethyl-*N,N*-diisopropylamidochlorophosphite.

Hybridization properties of oligonucleotides containing a modified d*C

Preparation of duplexes in which every natural dC is replaced by a modified d*C requires a great deal of work. Therefore, studies were first carried out with 9 bp duplexes composed of a triplet repeated three times involving an AT, GC or G*C base pair at position 4 (Table 1). We chose to investigate the stabilities of duplexes composed of a tridecamer and a nonamer to mimic hybrids formed between an oligonucleotide probe and a longer nucleic acid sequence. In fact, it is well known that the presence of dangling arms at both the 3'- and 5'-positions of the

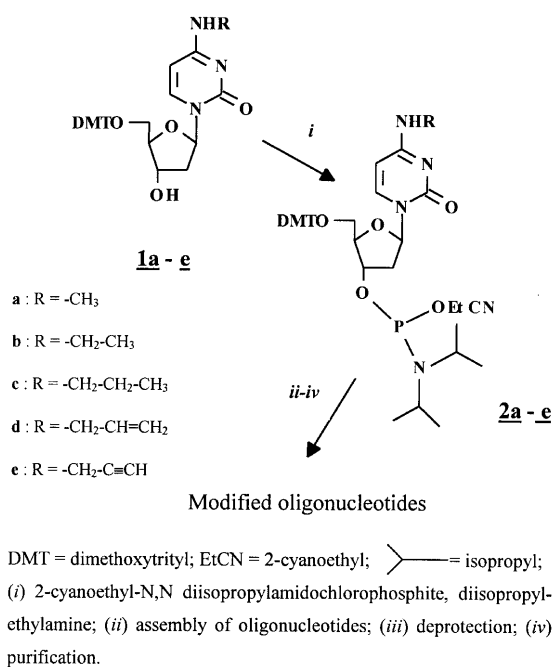


Figure 1. Synthesis of modified oligonucleotides involving *N*-4-substituted-2'-deoxycytidine.

oligonucleotide leads to good stabilization of a duplex which varies with its length (18). The results reported in Table 1 led to the following observations.

For duplexes **3-7** involving *N*-4-substituted dC the T_m variation, measured as described in Materials and Methods, is inversely proportional to the carbon atom number of the alkyl group (Me, Et, *n*-Pr). The T_m decrease is $\sim 2^\circ\text{C}$ when one methylene group is added (Fig. 2). Duplexes **6** and **7**, involving an allyl or propargyl group respectively, led to a T_m slightly higher than that of duplex **5**, involving an *n*-propyl group. This difference could be due to a steric effect or the higher positive inductive effect of the *n*-propyl group versus that of the allyl or propargyl group. Note that the difference in T_m is small and within the margin of error range, which is $\pm 1^\circ\text{C}$. The decrease in thermal stability of duplex **3**, involving one G^{4Me}C base pair, compared with that of duplex **2**, possessing a natural GC base pair at the same position, is small ($\Delta T_m \sim 2.5^\circ\text{C}$). Duplexes **3** and **4**, involving a G^{4Me}C and a G^{4Et}C respectively, are thermally more stable than duplex **1**, having an AT base pair at the same position (ΔT_m 3 and 1°C respectively).

Concerning double-stranded DNA involving natural nucleosides, their stabilities are highly dependent on base composition. Therefore, replacement of one AT base pair at position 4 of hybrid **1** (T_{m1} 46°C) by one GC base pair led to a 5°C increase in T_m (T_{m2} 51°C), whereas replacement of this same AT base pair by one G*C base pair (particularly *C = d^{4allyl}C and d^{4propargyl}C) led to hybrids having a stability very similar to that of duplex **1**.

Specificity of the G*C base pair

To verify the recognition specificity of dG by d*C, thermal denaturation studies were carried out with duplexes involving the mismatches XC and X*C (X = T, C or A; *C = ^{4Me}C, ^{4Et}C, ^{4Pr}C,

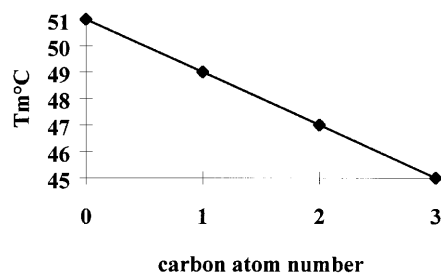


Figure 2. T_m variations of 5'-d (TTTCGTCGTCGTT) -3' / 3'-d (AGCAGYAGC) -5' *C = 4-H(CH₂)_nC as a function of the carbon atom number (*n*) of the substituent.

Table 1. Melting temperatures at λ_{260} nm of natural duplexes and those involving a modified *C at position 4



Duplex	X	Y	T_m (°C)	$\Delta T_m/1$ (°C)	$\Delta T_m/2$ (°C)
1	A	T	46		-5
2	G	C	51	+5	
3	G	^{4Me} C	49	+3	-2
4	G	^{4Et} C	47	+1	-4
5	G	^{4Pr} C	45	-1	-6
6	G	^{4Allyl} C	45.5	-0.5	-4.5
7	G	^{4Propargyl} C	45.5	-0.5	-4.5
8	G	^{ara} C	48	+2	-3

T_m values were determined at an oligomer strand concentration of 2 μM in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA.

4allylC, 4propargylC or ^{ara}C) (Table 2). The results obtained showed the following.

Specificity of G*C base pair formation was maintained. The presence of XC and X*C mismatches at position 4 led, in all cases, to a decrease of $>20^\circ\text{C}$ in T_m value.

The order of stabilities for duplexes involving XC and X*C mismatches was the same:

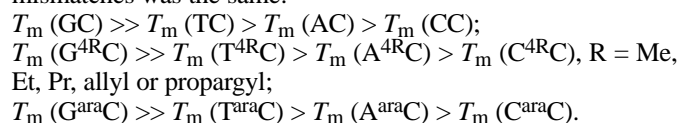


Table 2. Melting temperature values ($^{\circ}\text{C}$) at λ_{260} nm of perfect duplexes and ones containing a mismatch at position 4

5'-d (T T T C G T C X T C G T T) 3'
3'-d (A G C A G Y A G C) 5'
9 8 7 6 5 4 3 2 1

Y X	C	⁴ MeC	⁴ EtC	⁴ Pr C	⁴ Allyl C	⁴ Propargyl C	^{ara} C
G	51	49	47	47	45.5	45.5	48
T	28.3	25	20.5	21.5	20	22.5	22
C	21	17.5	18	22	19.5	20.5	19
A	27	27	25.5	24	24	24	22

T_m values were determined on a sample concentration of $2 \mu\text{M}$ in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA.

The presence of CC or C*C mismatches has a very destabilizing effect. ^{ara}C seems to give a better discrimination than do C and ⁴R C (Table 2).

Choice and hybridization properties of duplexes involving AT and/or G⁴EtC base pairs

Although the previously obtained results for T_m values of duplexes **6** and **7** involving d⁴allylC (T_{m6} 45.5 $^{\circ}\text{C}$) and d⁴propargylC (T_{m7} 45.5 $^{\circ}\text{C}$) respectively were closest to that of duplex **1**, involving an AT base pair (T_{m1} 46 $^{\circ}\text{C}$), we preferred to continue our studies with d⁴EtC, which forms a slightly more stable G⁴EtC base pair (T_{m4} 47 $^{\circ}\text{C}$) with G than it does with the AT base pair (T_{m1} 46 $^{\circ}\text{C}$).

This modification was chosen in order to obtain modified duplexes with a not too low thermal transition and to minimize the possible steric effect of the alkyl group when duplexes were built with several contiguous d*C. We also decided to eliminate ⁴MeC and the ^{ara}C because the first leads to a clearly more stable G⁴MeC base pair (T_{m3} 49 $^{\circ}\text{C}$) than does the AT base pair (T_{m1} 46 $^{\circ}\text{C}$), whereas the triphosphate of the latter could not be used by polymerases (19) for the preparation of a modified DNA fragment, in contrast to N4-substituted dC derivatives, which are accepted by DNA polymerase (20). Since ⁴MeC is present in DNA of certain thermophilic bacteria (21), we can expect that the ⁴EtC derivative has the same physicochemical and biochemical properties. Therefore, studies were carried out with duplexes composed of 9 bp involving various base pairs ranging from nine AT base pairs to nine GC or nine G⁴EtC base pairs (Table 3).

To determine the effect of replacing an AT base pair with a GC or G⁴EtC base pair, studies were carried out on duplexes **9-11**, having the same transition of purine \leftrightarrow pyrimidine (duplexes **10** and **11** correspond to duplex **9** in which the AT base pairs at positions 2, 5 and 8 have been replaced with the GC and G⁴EtC base pairs respectively), and on duplexes **2**, **12**, **13** and **14**, with

Table 3. Melting temperatures of natural and modified duplexes involving *C = ⁴EtC

Duplex		T_m (C)	(AT)	(GC)
9	5-d (T T T A T T A T T A T T T) 3' 3-d (A T A A T A A T A) 5' 8 5 2	20	9	0
10	5-d (T T T G T T G T T G T T T) 3' 3-d (A C A A C A A C A) 5' 3-d (A *C A A *C A A *C A) 5' 5-d (T T T G T T X T T G T T T) 3'	41	6	3
11	X=G	27	6	3
11a	X=T	<10		
11b	X=C	<10		
11c	X=A	<10		
12	3-d (A G *C A G *C A G *C) 5' 5-d (T T T *C G T *C X T *C G T T) 3' X=G	24	3	6
12a	X=T	<10		
13	5-d (T T C G G C G G C G G T T) 3' 3-d (G C C G C C G C C) 5'	59	0	9
14	5-d (T T *C G G *C G G *C G G T T) 3' 3-d (G *C *C G *C *C G *C *C) 5' 3-d (A G C A G C A G C) 5' 5-d (T T T C G T C X T C G T T) 3'	23	0	9
2	X=G	51.5	3	6
2a	X=T	28.5		
2b	X=C	21		
2c	X=A	27		

T_m values were determined at an oligomer strand concentration of $2 \mu\text{M}$ in 10^{-2} M sodium cacodylate buffer, pH 7, containing 1 M NaCl and 2×10^{-4} M EDTA.

different sequences. The results reported in Table 3 led us to the following observations.

Duplexes **11**, **12** and **14**, involving three, six and nine G⁴EtC base pairs respectively, have T_m values (T_{m11} 27 $^{\circ}\text{C}$, T_{m12} 24 $^{\circ}\text{C}$, T_{m14} 23 $^{\circ}\text{C}$) very close to that of duplex **9** (T_{m9} 20 $^{\circ}\text{C}$), composed of nine AT base pairs. These results show that thermal stability of these duplexes does not depend on their base composition, contrary to what was observed with natural duplexes whose T_m values greatly increase with GC base pair number (Figure 3).

As we have previously shown, G⁴EtC base pair formation is specific and, in fact, the T⁴EtC mismatch (duplex **11a**), C⁴EtC mismatch (duplex **11b**) and A⁴EtC mismatch (duplex **11c**) led to very unstable hybrids, making it impossible to determine their T_m values ($T_m < 10^{\circ}\text{C}$).

The T_m of duplex **9** (T_{m9} 20 $^{\circ}\text{C}$), involving nine AT base pairs, is lower than that of the natural duplex having five GC base pairs, three AT base pairs and one TC mismatch (T_{m2a} 28.5 $^{\circ}\text{C}$), CC mismatch (T_{m2b} 21 $^{\circ}\text{C}$) or AC mismatch (T_{m2c} 27 $^{\circ}\text{C}$); under the same conditions the T_m of duplex **9** is higher than that of modified duplexes having five G⁴EtC base pairs, three AT base pairs and one T⁴EtC mismatch ($T_{m11a} < 10^{\circ}\text{C}$), C⁴EtC mismatch ($T_{m11b} < 10^{\circ}\text{C}$) or A⁴EtC mismatch ($T_{m11c} < 10^{\circ}\text{C}$). These results show

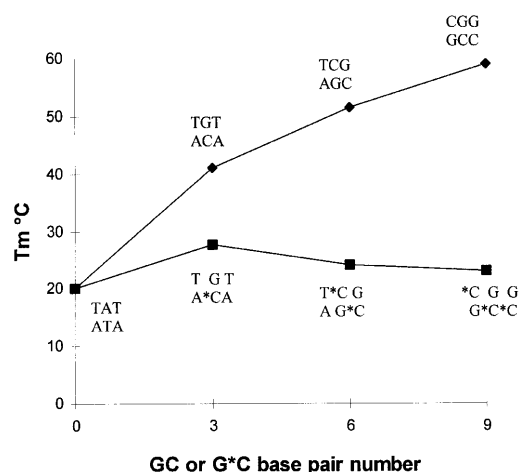
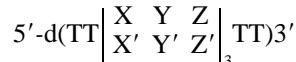


Figure 3. T_m variations of duplexes built with 9 bp



as a function of the GC (◆) or G* C (■) base pair number (*C = 4EtC).

that under the hybridization conditions used it is not possible to discriminate between perfect hybrids built with AT-rich sequences and those built with GC-rich sequences involving a mismatch. This major drawback of the natural double-stranded DNA can be alleviated by using duplexes built with AT and/or $G^{4\text{EtC}}$ base pairs.

The thermal stabilities of modified duplexes **11**, **12** and **14** are slightly higher than that of duplex **9**, having only AT base pairs. This is in accordance with previously obtained results. We note that the difference between the more stable duplex **11** (T_{m11} 27°C) and duplex **9** (T_{m9} 20°C) possessing nine AT base pairs is ~7°C. On the other hand, under the same conditions the difference in stability between natural duplexes involving nine GC (duplex **13**, T_{m13} 59°C) and nine AT base pairs (duplex **9**, T_{m9} 20°C) is ~39°C.

The T_m of duplex **14** (T_{m14} 23°C), containing nine $G^{4\text{EtC}}$ base pairs with three $4\text{EtC}^{4\text{EtC}}$ doublets, was as expected, similar to that of duplexes containing nine AT base pairs (T_{m9} 20°C), six AT and three $G^{4\text{EtC}}$ (T_{m11} 27°C) or three AT and six $G^{4\text{EtC}}$ (T_{m12} 24°C) respectively, where the $d^{4\text{EtC}}$ residues were not consecutive (Table 3 and Fig. 3). Consequently, substitution at position 4 of the C by an ethyl group does not lead to notable steric disturbance. These results were confirmed by the high cooperativity of dissociation of modified duplexes **12** and **14** into single strands, which was similar to that of natural duplexes **2** and **13** (Fig. 4).

For duplexes **9**, **12** and **14**, involving nine AT base pairs, three AT base pairs and six $G^{4\text{EtC}}$ base pairs and nine $G^{4\text{EtC}}$ base pairs respectively, the increase in absorbance recorded at λ_{260} nm at the time of dissociation into single strands decreased when $G^{4\text{EtC}}$ base pair number increased. The same phenomenon was also observed with natural duplexes **2** (three AT base pairs and six GC base pairs) and **13** (nine GC base pairs). These results suggest that the GC and $G^{4\text{EtC}}$ base pairs have similar physicochemical properties. It is highly likely that 4EtC forms a $G^{4\text{EtC}}$ base pair with G having three hydrogen bonds, since the $G^{4\text{EtC}}$ base pair stability would otherwise be weaker, as was shown for a GK base

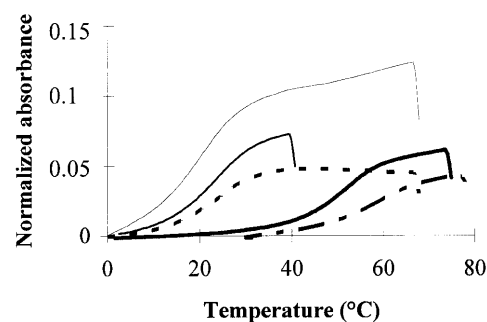


Figure 4. Melting curves of duplexes **14** (----), **12** (—), **13** (— · —), **2** (—) and **9** (---) recorded at λ_{260} nm.

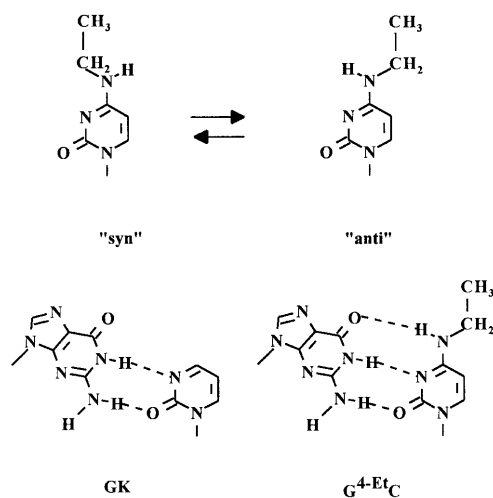


Figure 5. Structure of the 'anti' and 'syn' conformations of 4EtC and of the GK and $G^{4\text{EtC}}$ base pairs.

pair (K being 2-pyrimidinone), which has two hydrogen bonds (22) and is isomorphous with the GC base pair (Fig. 5).

In single-stranded structures the ethyl group in 4EtC has 'syn' and 'anti' conformations, the balance of which can be quickly shifted to the 'anti' configuration by hybridization with G. In fact, variations in optical density as a function of temperature recorded during duplex dissociation into single strands or during association of the latter into duplexes are identical (results not shown).

Several hypotheses could be envisaged to explain the $G^{4\text{EtC}}$ base pair stability decrease compared with that of the natural GC base pair.

Substitution of one hydrogen of the amino group at position 4 of C with one alkyl group (R = Me, Et, Pr, ...) induces a positive inductive effect higher than that of the hydrogen, which makes the other hydrogen less electrophilic leading to a weakening of the hydrogen bond between the H atom at position 4 of C and the oxygen atom at position 6 of G.

The presence of a lipophilic alkyl group whose size is relatively large could modify the conformation and/or hydration of duplexes.

All the results obtained show that the stability of modified duplexes built with AT and/or $G^{4\text{EtC}}$ base pairs is not dependent on their AT/ $G^{4\text{EtC}}$ content ratio. This new system may be very

useful for DNA sequencing by hybridization, allowing discrimination between perfect hybrids and those involving mismatches.

CONCLUSION

We have proven, for the first time, that the stability of duplexes made with oligonucleotides of a given length built with AT and/or G^{4Et}C base pairs is independent of their base content in a classical buffer solution. Specificity of the G^{4Et}C base pair is maintained and the cooperativity of modified duplex dissociation is similar to that of natural DNA duplexes involving natural base pairs. These very useful properties make possible the employment of this new system in reverse hybridization approaches, using a large number of sequences immobilized as a two-dimensional matrix for simple and fast analysis of nucleic acid sequences, or in biochemical techniques, such as random priming, for the preparation of DNA and labeling DNA fragments by an enzymatic method. Work is currently in progress to confirm the hybridization properties of such modified duplexes with large oligonucleotide sequences, to determine the influence of the sequence on thermal stability and to prepare a DNA target fragment using d^{4Et}C triphosphate and DNA polymerases.

ACKNOWLEDGMENT

This work was supported by a BioMed contract (gene-CT 93-0009).

REFERENCES

- Southern, E.M. (1989) International patent WO 89/10977.
- Drmanac, R., Labat, I., Bruckner, I. and Crkvenjakov, R. (1989) *Genomics*, **4**, 114–128.
- Krapko, K.R., Lysov, Y.P., Khorlin, A.A., Ivanov, I.B., Yershov, A.D., Vasilenko, S.K., Florentiev, V.L. and Mirzabekov, A.D. (1991) *J. DNA Sequencing Mapping*, **1**, 375–388.
- Melchoir, W.B., Jr and Von Hippel, P.H. (1973) *Proc. Natl. Acad. Sci. USA*, **70**, 298–302.
- Wood, W.I., Gitschier, J., Ladky, L.A. and Lawn, R.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1585–1588.
- Jacobs, K.A., Rudersdorf, R., Neill, S.D., Dougherty, J.P., Brown, E.L. and Frisch, E.F. (1988) *Nucleic Acids Res.*, **16**, 4637–4650.
- Maskos, U. and Southern, E.M. (1993) *Nucleic Acids Res.*, **21**, 4663–4669.
- Ricelli, P.V. and Benight, A.S. (1993) *Nucleic Acids Res.*, **21**, 3785–3788.
- Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (Eds) (1990) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York, NY.
- Hoheisel, J.D. and Lehrach, H. (1990) *FEBS Lett.*, **274**, 103–106.
- Hoheisel, J.D. (1996) *Nucleic Acids Res.*, **24**, 430–432.
- Fritz, H.J., Frommer, W.B., Kramer, W. and Werr, W. (1982) In Gassen, H.G. and Lang, A. (Eds), *Chemical and Enzymatic Synthesis of Gene Fragments. A Laboratory Manual*. Verlag Chemie, Weinheim, Germany, pp. 43–52.
- Sung, W.L. (1981) *Nucleic Acids Res.*, **9**, 6139–6151.
- Beaucage, S.L. and Caruthers, M.H., (1981) *Tetrahedron Lett.*, **22**, 1859–1862.
- Fasman, G.D. (Ed.) (1975) *Handbook of Biochemistry and Molecular Biology, Nucleic Acids*, 3rd Edn. Vol. 1. CRC Press, Cleveland, Ohio, pp. 589.
- Giannaris, P.A. and Damha, M.J. (1994) *Can. J. Chem.*, **72**, 909–918.
- Butkus, V., Klimasauskas, S., Petrauskienė, L., Maneliene, Z., Janulaitis, A., Minchenkova, L.E. and Schyolkina, A.K. (1987) *Nucleic Acids Res.*, **15**, 8467–8478.
- Williams, J.C., Case-Green, S.C., Mir, K.U. and Southern, E.M., (1994) *Nucleic Acids Res.*, **22**, 1365–1367.
- Mikita, T. and Beardsley, G.P. (1994) *Biochemistry*, **33**, 9195–9208.
- Li, S., Haces, A., Stupar, L., Gebeyehu, G. and Pless, R.C. (1993) *Nucleic Acids Res.*, **21**, 2709–2714.
- Ehrlich, M., Norris, K.F., Wang, R.Y.-H., Kuo, K.C. and Gehrke, C.W. (1986) *Biosci. Rep.*, **6**, 387–393.
- Gildea, B. and McLaughlin, L.W. (1989) *Nucleic Acids Res.*, **17**, 2261–2281.