Organic & Biomolecular Chemistry

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Specific recognition of cytosine by hypoxanthine in pyrrolidinyl peptide nucleic acid†

Cite this: *Org. Biomol. Chem.*, 2013, **11**, 2310

Chotima Vilaivan,[a] Wimonmas Srinarang,[a] Nattawut Yotapan,[a] Woraluk Mansawat,[a] Chalothorn Boonlua,[a] Junji Kawakami,[b,c] Yoshie Yamaguchi,[b,c] Yuko Tanaka[b,c] and Tirayut Vilaivan*[a]

Received 31st October 2012,
Accepted 30th January 2013

DOI: 10.1039/c3ob27129c

www.rsc.org/obc

Hypoxanthine is an unnatural base that can potentially pair with all natural nucleobases. While hypoxanthine in DNA exhibits marginal preference for pairing with cytosine (C), little is known about its pairing behavior in other DNA analogues. In this study, we synthesized a hypoxanthine-containing monomer and incorporated it into pyrrolidinyl peptide nucleic acid with α/β-peptide backbone derived from D-prolyl-(1*S*,2*S*)-2-aminocyclopentanecarboxylic acid (acpcPNA). DNA binding studies clearly revealed that hypoxanthine in acpcPNA behaves like G-analogue because it can specifically form a stable base pair with dC in DNA. The ability to replace G by hypoxanthine without affecting the base pairing properties of acpcPNA can solve a number of problems associated with G-rich acpcPNA including difficult synthesis, formation of secondary structures and fluorescence quenching.

Introduction

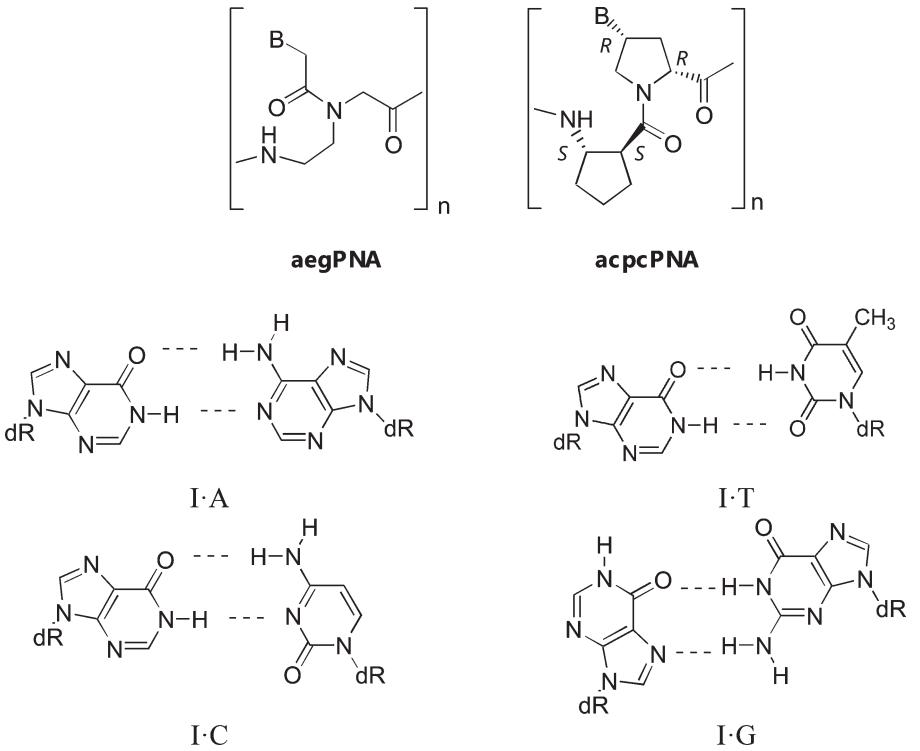
Hypoxanthine is a purine base often found in the first anticodon position of tRNA.[1] From its structure, there exist possibilities of formation of reasonable hydrogen-bonded base pairs with all canonical nucleobases in DNA and RNA (Scheme 1), leading to a hypothesis that it can behave as universal base. Indeed, applications of oligonucleotides containing inosine (the nucleotide form of hypoxanthine, often designated as I) as degenerate probes and primers have been realized.[2] The pairing behavior of hypoxanthine in DNA was by no means universal, but was fairly selective for pairing with C. The base pairs containing I are generally destabilized in solution compared to conventional Watson–Crick base pairs, with the I·C pairs being the least destabilized and the order of stability of the base pairs is I·C > I·A > I·G ~ I·T.[3] However, a recent stability study suggested that I-containing base pairs are at least as stable as the A·T pairs in the gas phase.[4] The base-pairing behavior of hypoxanthine in other unnatural analogues of DNA and RNA is less well characterized. The base-pairing properties of hypoxanthine incorporated into locked nucleic acid (LNA)[5] and peptide nucleic acid (PNA)[6] – a DNA analogue in which the deoxyribose phosphate backbone is completely replaced by a peptide-like structure such as *N*-2-aminoethylglycine (also known as aegPNA)[7] – had been investigated. While hypoxanthine in LNA can be regarded as G analogue because it can preferentially pair to C albeit with lower stability and selectivity than G,[5] hypoxanthine in aegPNA may pair to either C or A, depending on the identity of flanking nucleobases.[6] Base pairing properties of aegPNA bearing other unnatural nucleobases including xanthine,[8] cyanuric acid[9] as well as other potential universal nucleobase systems[10] have also been reported. In view of superior binding characteristic of our new pyrrolidinyl PNA consisting of an α/β peptide backbone derived from D-proline/(1*S*,2*S*)-2-aminocyclopentanecarboxylic acid (acpcPNA) (Scheme 1)[11] in terms of binding affinity as well as base-pairing and directional specificity compared to aegPNA,[12] we were interested in synthesizing and evaluating the DNA-pairing properties of acpcPNA carrying hypoxanthine as an unnatural base. The knowledge of pairing rules of hypoxanthine (will be designated as I from now on) in acpcPNA should be useful in expanding the scope of applications of this acpcPNA system.

[a]*Organic Synthesis Research Unit, Department of Chemistry, Faculty of Science, Chulalongkorn University, Phayathai Road, Patumwan, Bangkok 10330, Thailand. E-mail: vtirayut@chula.ac.th; Fax: +66-2-2187598*
[b]*Department of Nanobiochemistry, FIRST, Konan University, 7-1-20 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan*
[c]*Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 7-1-20 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan*
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c3ob27129c

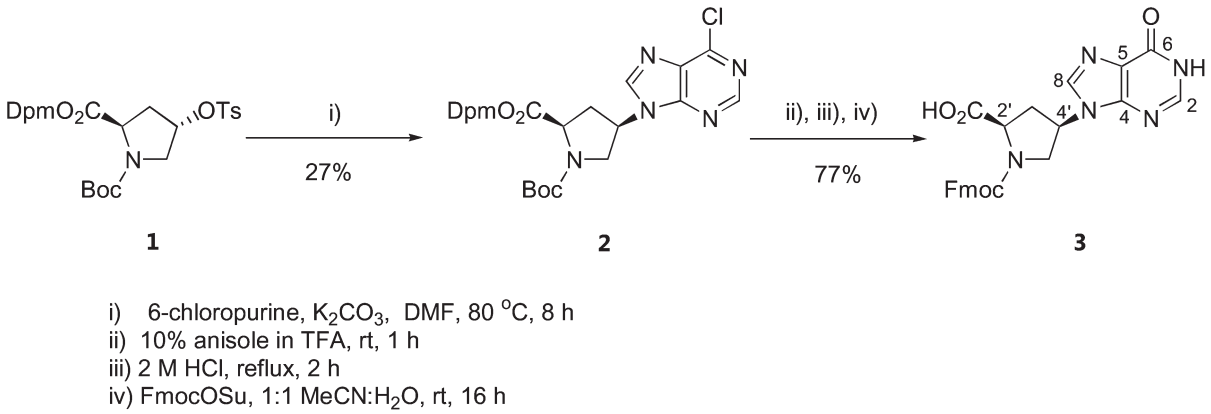
Results and discussion

Synthesis of hypoxanthine acpcPNA monomer

Direct S_N2 displacement of the tosylate group in the known intermediate (**1**) with hypoxanthine under standard conditions successfully employed for the synthesis of the protected



Scheme 1 Structures of **aegPNA** and **acpcPNA** (top) and hypothetical base pairing schemes between hypoxanthine and all four canonical nucleobases (bottom).



Scheme 2 Synthesis of hypoxanthine acpcPNA monomer (**3**).

adenine and cytosine monomers[13] did not give the expected substitution product. A related study had shown that regioselectivity (most notably *O*-alkylation) was also problematic with direct alkylation of hypoxanthine by Mitsunobu reaction.[14] An alternative method starting from 6-chloropurine as a masked hypoxanthine equivalent as shown in Scheme 2 was therefore employed.[15] In the first step, 6-chloropurine was reacted with **1** in the presence of K_2CO_3 to give the 6-chloropurine derivative (**2**). The N^9 regioselectivity was assumed based on literature precedent in a closely related system.[16] The Boc and diphenylmethyl (Dpm) groups were simultaneously removed by treatment with trifluoroacetic acid (TFA). Subsequent hydrolysis of the 6-chloro group to give the unprotected hypoxanthine monomer was accomplished by refluxing with 2 M aqueous HCl. Finally, the Fmoc group was introduced using FmocOSu under aqueous basic conditions to give the Fmoc-protected hypoxanthine acpcPNA monomer (**3**). The N^9 substitution pattern in the final product was confirmed by ^1H–^{13}C HMBC experiment (Fig. S3†) whereby correlations between H4′ (δ_H 5.11 ppm)–C(4) (δ_C 148.3 ppm) and between H4′–C(8)H (δ_C 138.8 ppm), but not between H4′–C(5) (δ_C 124.2 ppm), were observed.

Synthesis of hypoxanthine-containing acpcPNA

The Fmoc-protected monomer (**3**) could be incorporated into any desired positions of acpcPNA sequences employing the

Table 1 Sequence and mass spectrometric characterization (MALDI-TOF) data of the acpcPNA synthesized in this work

Entry	PNA	Sequence (N→C)[a]	*m/z* (calcd)	*m/z* (found)
1	**PNA1**	TTTTITTTT	3189.4	3190.9
2	**PNA2**	CCTTĀIACATC	3821.1	3820.1
3	**PNA3**	CCTTTITCATC	3803.1	3801.0
4	**PNA4**	CCTTCICCATC	3773.1	3772.5
5	**PNA5**	CCTTGIGCATC	3853.1	3853.9
6	**PNA6**	CCTTAGACATC	3836.2	3836.9
7	**PNA7**	CCTTTGTCATC	3818.1	3816.0
8	**PNA8**	AATTTICATCA	3860.2	3860.8
9	**PNA9**	AATTTĠCATCA	3875.2	3873.0
10	**PNA10**	Flu-O-ICTGCTTCACT[b]	4178.4	4180.8
11	**PNA11**	Flu-O-ĠCTGCTTCACT[b]	4193.4	4196.5

[a] All acpcPNA were C-terminal capped with L-lysinamide and *N*-acetylated. [b] The ACPC spacer at the N-terminus was omitted from these two sequences. Flu = 5(6)-carboxyfluorescein; O = aminoethoxyethoxyacetyl linker.

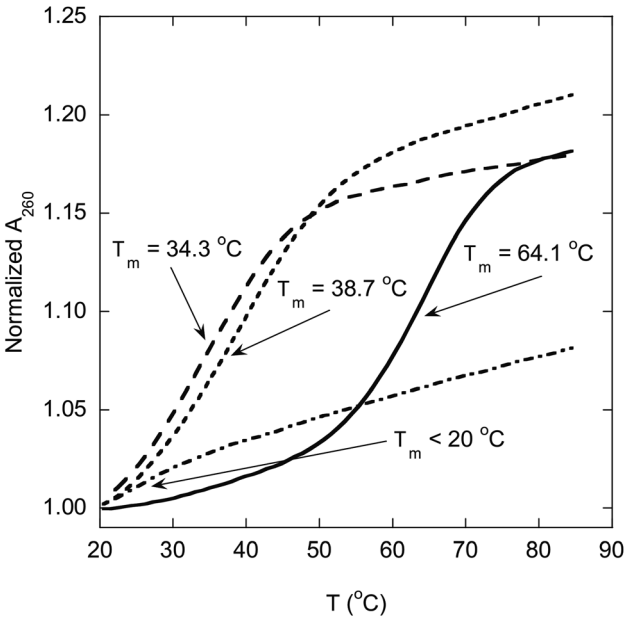


Fig. 1 Melting curves of hybrids between **PNA1** and dA_4NA_4 (N: — C, –·–· G, – – T, ····· A). *Conditions*: 1.0 μM PNA, 1.0 μM DNA, 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl. The curves were smoothed by the boxcar algorithm using the smooth macro in KaleidaGraph 4.1 employing a window size = 5.

standard protocol for Fmoc-based PNA synthesis using standard HATU/DIEA activation.[11] The sequences synthesized include a homothymine acpcPNA with the base T at the middle position of the strand was replaced by I (**PNA1**), four 11mer mix-base acpcPNA with an I residue in the middle flanked by each of the four different standard nucleobases (**PNA2–PNA5**), two hypoxanthine-modified mix-base acpcPNA (**PNA8** and **PNA10**). The **PNA10** containing hypoxanthine at the N-terminal position was further modified at the N-terminus with 5(6)-carboxyfluorescein. Selected acpcPNA sequences having the base I replaced by G were also synthesized as controls (**PNA6**, **PNA7**, **PNA9** and **PNA11**). The identity of all acpcPNA was confirmed by MALDI-TOF mass spectrometry and the purity was determined to be >90% by reverse phase HPLC. The sequences and characterization data of the acpcPNA synthesized are summarized in Table 1.

Hybridization properties of hypoxanthine-containing acpcPNA

We initially investigated the base pairing behavior of **PNA1**, which has a homothymine sequence with I in the middle position. We had earlier demonstrated that homothymine acpcPNA only form duplex with complementary DNA instead of the triplex usually observed with aegPNA and other PNA systems.[11] It is therefore reasonable to assume that the **PNA1** should also form duplex with DNA. As shown in Fig. 1, the T_m of the hybrid of **PNA1** with dA_4CA_4 (pI·dC pair) was 64.1 °C, which was much higher than the other three hybrids (dA_4GA_4: <20 °C; dA_4TA_4: 34.3 °C; dA_9: 38.7 °C). The higher T_m of the hybrid containing pI·dC pair by >25–30 °C clearly revealed that the base hypoxanthine in **PNA1** showed a strong preference in pairing with C. Hence, the base I behaves as a G analogue rather than as a universal base. This behavior is analogous to natural DNA[3] and LNA,[5] although the dI·dC pairing in acpcPNA described here is much more selective.

As the stabilities of I-containing base pairs in DNA tend to vary depending on the sequence,[17] it is important to determine whether the preference for pairing of base I in acpcPNA with C in DNA is a general behavior or limited to certain sequences. The **PNA2–PNA5** with the same mix-base sequence but carrying four different base triads (NIN; N = A, T, C, G) in the middle position of the acpcPNA strand were synthesized and their binding properties with DNA studied by thermal denaturation experiments. These particular sequences were chosen because the corresponding data for aegPNA system are available for direct comparison.[6] As shown in Table 2 and Fig. S16,† the T_m data obtained from these mixed sequence acpcPNA again confirmed the preferential pairing of pI with dC. In all four cases the pI·dC pairs were the most stable (T_m with complementary DNA of **PNA2**: 70.2; **PNA3**: 66.1; **PNA4**: 48.0; **PNA5**: 60.8 °C). The unusually high thermal stability of A+T rich acpcPNA hybrids has previously been observed in this acpcPNA system.[11] Thermodynamic parameters analyses indicated that the formation of A·T base pairs in the acpcPNA·DNA hybrids are enthalpically more favorable over G·C pairs.[11b] Interestingly, the thermal stabilities of acpcPNA·DNA hybrids carrying pI·dC were generally higher than the corresponding acpcPNA·DNA hybrids with pG·dC pairs (Fig. S17†). For example, the hybrid of **PNA6**, bearing the same sequence as **PNA2**, but with I replaced by G, with its complementary DNA showed a T_m of 62.3 °C which was 7.9 °C lower than that of the corresponding **PNA2** hybrid. Likewise, the hybrid of **PNA7** with its complementary DNA showed a T_m of 64.4 °C which was 1.7 °C lower than that of the corresponding DNA hybrid of **PNA3**. Significant reduction in T_m values was observed in acpcPNA·DNA hybrids carrying "mismatched" pI·dT, pI·dG and pI·dA pairs (ΔT_m **PNA2**: −22.0 to −27.2; **PNA3**: −24.1 to −34.9; **PNA4**: −11.9 to −16.2; **PNA5**: −7.7 to −18.6 °C). The

Table 2 T_m and thermodynamic parameters of hypoxanthine-containing **PNA2–PNA5** and guanine-containing **PNA6–PNA7**[a,b]

Entry	PNA	DNA[c]	T_m (°C)	$\Delta H°$ (kcal mol^{-1})	$\Delta S°$ (cal mol^{-1} K^{-1})	$\Delta G°_{298}$ (kcal mol^{-1})
1	**PNA2**	...TCT...	70.2	−82.0 ± 0.5	−210 ± 1.5	−19.4 ± 0.7
2	(...AIA...)	...TTT...	43.0	−55.0 ± 0.7	−145 ± 2.1	−11.7 ± 0.9
3		...TGT...	44.0	−57.0 ± 0.8	−151 ± 2.6	−12.0 ± 1.1
4		...TAT...	48.0	−66.8 ± 1.8	−179 ± 5.8	−13.4 ± 2.5
5	**PNA6**	...TCT...	62.3	−74.6 ± 0.8	−194 ± 2.4	−16.9 ± 1.1
6	(...AGA...)	...TTT...	42.3	−55.4 ± 0.6	−147 ± 2.0	−11.7 ± 0.9
7		...TGT...	39.1	−48.2 ± 0.2	−125 ± 0.7	−10.8 ± 0.3
8		...TAT...	37.6	−52.9 ± 1.5	−141 ± 4.7	−10.8 ± 2.0
9	**PNA3**	...ACA...	66.1	−69.1 ± 0.5	−175 ± 1.5	−17.0 ± 0.7
10	(...TIT...)	...ATA...	42.0	−53.3 ± 0.5	−140 ± 1.6	−11.5 ± 0.7
11		...AGA...	31.2	−44.0 ± 0.5	−116 ± 1.7	−9.5 ± 0.7
12		...AAA...	37.7	−46.6 ± 0.5	−121 ± 1.5	−10.5 ± 0.7
13	**PNA7**	...ACA...	64.4	−68.6 ± 0.6	−174 ± 1.5	−16.6 ± 0.8
14	(...TGT...)	...ATA...	46.4	−60.6 ± 0.5	−161 ± 1.6	−12.7 ± 0.7
15		...AGA...	29.0	−41.1 ± 0.7	−107 ± 2.1	−9.2 ± 0.9
16		...AAA...	25.5	−39.6 ± 0.6	−104 ± 1.9	−8.7 ± 0.8
17	**PNA4**	...GCG...	48.0	−50.4 ± 0.5	−128 ± 1.4	−12.2 ± 0.6
18	(...CIC...)	...GTG...	36.1	−51.2 ± 0.7	−137 ± 2.4	−10.5 ± 1.0
19		...GGG...			Non-two-state melting curve	
20		...GAG...	31.8	−49.7 ± 0.6	−134 ± 2.1	−9.7 ± 0.9
21	**PNA5**	...CCC...	60.8	−64.0 ± 0.6	−163 ± 1.9	−15.5 ± 0.8
22	(...GIG...)	...CTC...	47.7	−58.6 ± 0.5	−154 ± 1.5	−12.8 ± 0.6
23		...CGC...	42.2	−49.8 ± 0.5	−129 ± 1.6	−11.3 ± 0.7
24		...CAC...	53.1	−68.3 ± 2.0	−181 ± 6.0	−14.5 ± 2.7

[a] The melting experiments were carried out in 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl, concentration of PNA and DNA = 1.0 μM, monitored by UV-Vis spectrophotometry at 260 nm, heating rate 1 °C min^{-1}. [b] Both T_m and thermodynamics parameters were determined from thermal denaturation curves by nonlinear curve-fitting as described in the Experimental section. [c] The symbols ...NNN... denote the three bases in the middle position of the DNA sequences (5′→3′) *e.g.* ...TCT... = GATGTCTAAGG; ...AAA ... = GATGAAAAAGG *etc*.

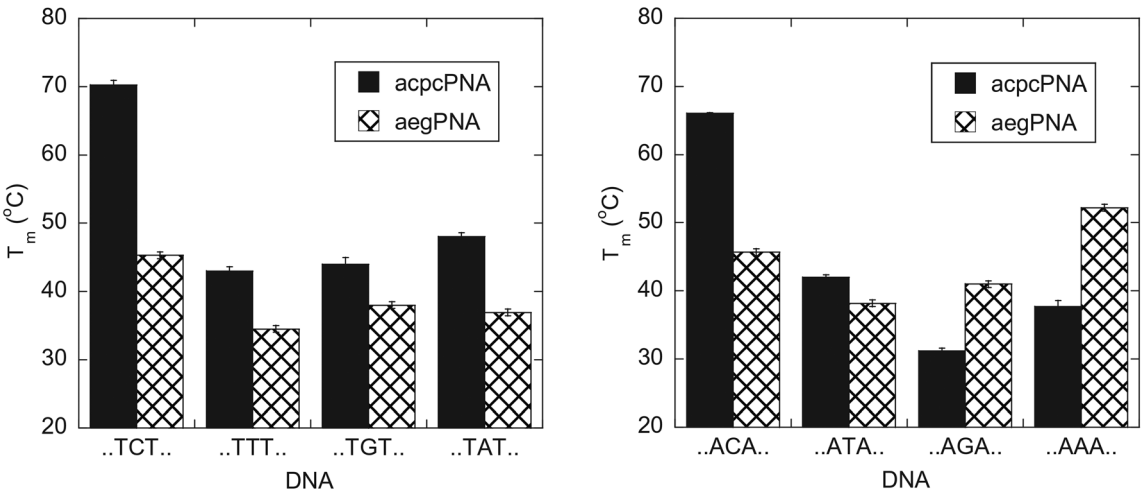


Fig. 2 Comparison of melting temperatures (T_m) of DNA hybrids of hypoxanthine-containing acpcPNA **PNA2** (AIA: ■) and aegPNA (▨) (left) and hypoxanthine-containing acpcPNA **PNA3** (TIT: ■) and aegPNA (▨) (right). The symbols …NNN… in the *x*-axis denote the three base in the middle position of the DNA sequences. *Conditions*: 1.0 μM PNA, 1.0 μM DNA, 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl. T_m data of aegPNA are taken from ref. 6.

corresponding ΔT_m figures for **PNA6** and **PNA7** are −20.0 to −24.7 and −18.0 to −38.9 °C, respectively. These results clearly indicated that the base I in acpcPNA could behave as a G analogue by pairing strongly and specifically with the base C in DNA regardless of the identity of flanking bases.

The results are in sharp contrast to aegPNA whereby a similar study revealed that the pairing properties of the base I were highly dependent on the flanking bases.[6] I in aegPNA prefers pairing to A when the flanking bases are T. On the other hand, when the flanking bases are A, I in aegPNA prefers pairing to C (Fig. 2). Based on a quantum chemical calculation, it was suggested that the dA in the pI·dA pair adopted a *syn*-conformation leading to the formation of Hoogsteen type base pairing rather than the standard Watson–Crick base pairing.[6] In the case of acpcPNA, however, the base I with flanking A or T shows the same preference for pairing to C with almost the same level of mismatch discrimination observed in the base G. The same preference for pairing to C

is also observed in sequences having the base I with flanking C or G (Table 2). Among mismatched pairs, the smallest decrease of T_m was generally observed when the base opposite to the base I was A. In the worst case (pI·dA mismatch in **PNA5**), the ΔT_m was only −7.7 °C compared to the pI·dC pair derived from the same PNA. This ΔT_m figure is rather small for acpcPNA, but can still be considered relatively large compared to DNA and other analogues.

Other thermodynamic data demonstrate more detailed circumstances. As shown in Table 2, the thermodynamic data of pI·dN duplexes and pG·dN duplexes (N = C, T, G, and A) share a feature with significant similarity. Especially when N = T, all the parameters of **PNA2** with pI·dT mismatch pair are in absolute agreement with those of **PNA6** with pG·dT pair. The relatively stable pG·dT mismatches in DNA duplexes form a wobble base pairs in which 2-amino group of guanine does not contribute, and therefore pI·dT pairs are likely to be in a corresponding base geometry. On the other hand in the case of N is a purine base (A or G) and a large purine–purine base pair arise, the difference of enthalpy change from the values of the most stable pI·dC (or pG·dC) pair ($\Delta\Delta H°$) was almost the same. The $\Delta\Delta H°$ of pI·dG or pG·dG pair of **PNA2**, **PNA3**, **PNA6** and **PNA7** was 25.0, 25.1, 26.4, and 27.5 kcal mol^{-1}, respectively. When N = A, the decrease of enthalpic binding energy was similar or slightly smaller than that of pI·dG (or pG·dG) pair. These energetic similarity, including the case of N = T, strongly suggest that the binding mode and the structure of pI·dN pairs do not have drastic change from those of pG·dN pairs. Despite the lack of a hydrogen bond between 2-amino group of purine and carbonyl oxygen of cytosine, enthalpic interaction of pI·dC pair is rather stronger than the pG·dC case. This finding suggests the situation in which the pI·dC pair stands at appropriate position for superior base stacking, compensating for the absence of hydrogen bond. This idea is completely consistent with our previous study, *i.e.*, the difference of the enthalpy change between G·C and A·T pairs was rather small and the result came from favorable stacking rather than hydrogen bonding for acpcPNA·DNA duplexes.[11b]

Potential applications of acpcPNA carrying hypoxanthine base

Although it has been known for some time that hypoxanthine behave rather like guanine in DNA,[3] the stability of the dI·dC pair is generally less than that of the canonical dG·dC pair and the difference between the stability of dI·dC pair over other "mismatched" pairs is not large enough to be generally useful. The strong affinity and high specificity of pI·dC pairing over pI·dG, pI·dA and pI·dT in acpcPNA system is therefore quite unexpected and could lead to several practical applications. In order to evaluate the generality of the strong and selective pairing between pI and dC in acpcPNA, the **PNA8** and **PNA9** were synthesized and their hybridization with various DNA targets were compared. The only difference between **PNA8** and **PNA9** is the base at the position 6 (I in **PNA8** and G in **PNA9**) with T/C as neighboring bases. Results from thermal denaturation experiments (Table 3) confirmed that the complementary DNA hybrid of **PNA8** containing pI·dC pair was at least as strong as the corresponding DNA hybrid of **PNA9** (with pG·dC pair). Furthermore, the specificity of the pI·dC pairing in **PNA8** remains very high, and is more uniform (ΔT_m −21.9 to −25.8 for mismatched pairs) compared to that of pG·dC pairing in **PNA9** (ΔT_m −16.1 to −30.6 for mismatched pairs). Circular dichroism (CD) spectra of DNA hybrids of **PNA8** and **PNA9** suggested that the overall helical conformations of the duplexes were similar according to their virtually identical CD spectra (Fig. 3).

Table 3 Comparison of T_m of DNA hybrids of hypoxanthine-containing **PNA8** and guanine-containing **PNA9**

Entry	PNA (N→C) AATTTRCATCA	N in DNA (5′→3′) TGATGNAAATT	T_m [a] (°C)	ΔT_m [b] (°C)
1	**PNA8**	C	74.9	N/A
2	(R = I)	T	53.0	−21.9
3		G	49.1	−25.8
4		A	49.9	−25.0
5	**PNA9**	C	71.7	N/A
6	(R = G)	T	55.6	−16.1
7		G	46.8	−24.9
8		A	41.1	−30.6

[a] T_m were measured at 1.0 μM duplex concentration, 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl. [b] $\Delta T_m = T_m$ (pR·dN) − T_m (pR·dC), when R = I or G and N is the DNA base under consideration.

From the results obtained above, we may confidently propose that hypoxanthine can be used as a perfect mimic of guanine in acpcPNA under a variety of sequence contexts. The ability to change the base from G to I without compromising the binding affinity and sequence specificity of acpcPNA can be considered very significant because there are a number of problems associated with G-rich acpcPNA sequences. First of all, G monomer of acpcPNA is difficult to synthesize, requiring expensive starting materials and long, low-yielding synthetic sequences.[13] Secondly, DNA as well as PNA with G-rich sequences are known to form secondary structures such as G-quadruplexes[18] that can prevent them form binding with complementary DNA targets by standard Watson–Crick base pairing. Since hypoxanthine cannot form the hydrogen bond network in the same way as guanine, replacement of some guanine residues in G-rich sequences acpcPNA should disfavor the formation of G-quadruplexes, and thus facilitate the DNA binding. Thirdly, the unique redox properties of guanine may complicate certain applications such as electrochemical assays.[19] Hence, the ability to replace G with its less readily oxidized counterpart without changing the base pairing property is essential for development of PNA or DNA probes to be used in these applications.[20] Finally, guanine and G·C base pair are known to quench several common fluorescent labels.[21,22] The quenching effect by guanine has been utilized for certain applications[23] but needed to be avoided in other situations. To demonstrate this effect, the fluorescence spectra of the 5(6)-carboxyfluorescein-labeled **PNA10** and **PNA11** were compared. The **PNA10**, with hypoxanthine as the immediate base adjacent to the fluorescein label showed a significantly stronger fluorescence emission (1.54 times) compared to the **PNA11** which

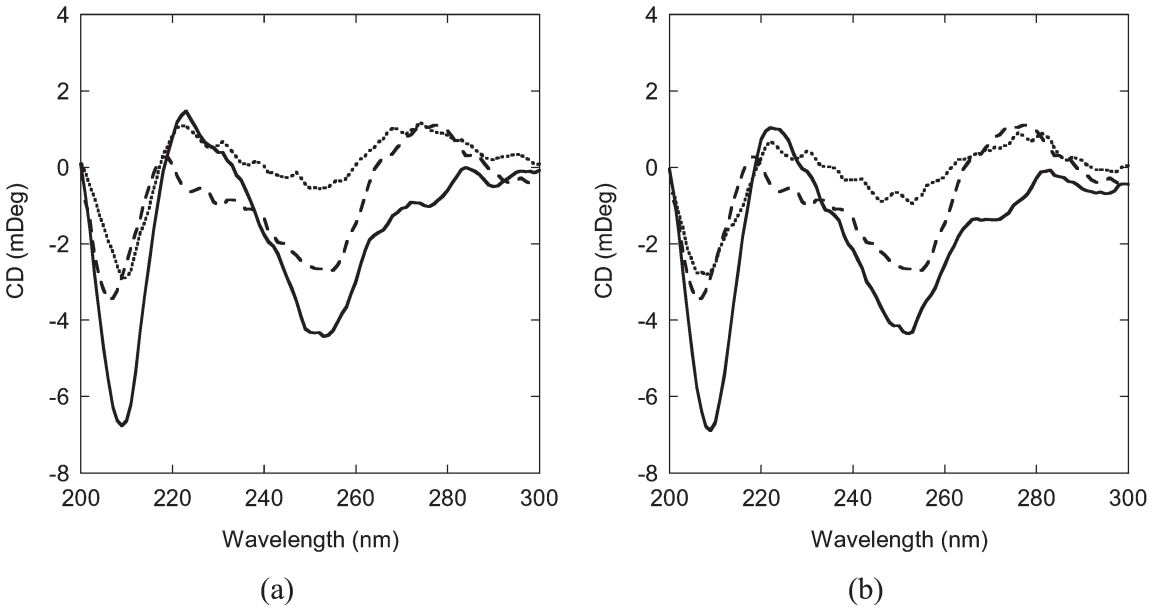


Fig. 3 CD spectra of **PNA8** (···) (a) and **PNA9** (···) (b) together with their complementary DNA hybrids (—). The CD spectrum of the single stranded complementary DNA (– –) was also included for comparison. *Conditions*: 1.0 μM PNA, 1.0 μM DNA, 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl.

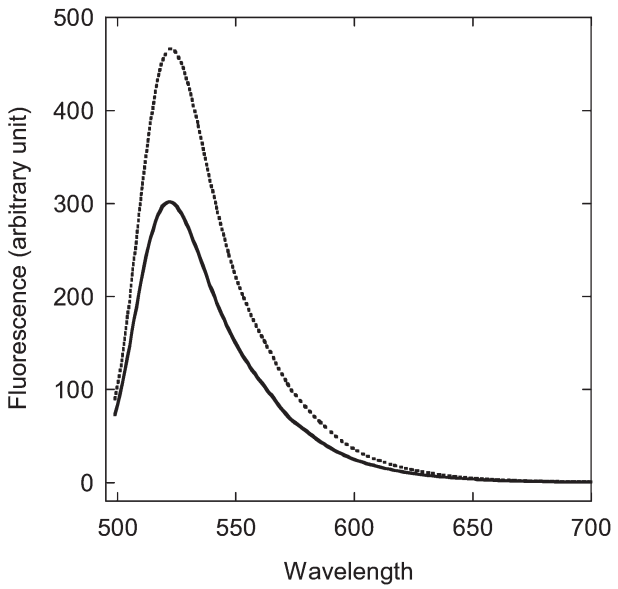


Fig. 4 Comparison of steady state fluorescence spectra of fluorescein-labeled **PNA10** (···) and **PNA11** (—). *Conditions*: 0.05 μM PNA, 10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl, λ_{ex} = 485 nm (λ_{em} = 522 nm).

bears guanine at the same position under similar conditions (Fig. 4).

Experimental

General

Chemicals and solvents were purchased from commercial suppliers and were used without further purification. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker Avance 400 NMR spectrometer operating at 400 MHz for ^{1}H and 100 MHz for ^{13}C. MALDI-TOF mass spectra were obtained on a Microflex MALDI-TOF mass spectrometry (Bruker Daltonics) using α-cyano-4-hydroxycinnamic acid (CCA) as the matrix. High resolution mass spectra were recorded in positive ESI mode on a Bruker Daltonics micrOTOF mass spectrometer. IR spectra were recorded on a Nicolet iS10 FT-IR spectrometer.

Oligonucleotides were purchased from Biodesign (Pathumthani, Thailand) and were used as received. The concentration of oligonucleotides and acpcPNAs was determined from the absorbance at 260 nm using the reported molar extinction coefficients at 260 nm (ε_{260}) for the corresponding DNA monomers.[7c] The molar extinction coefficient of I was obtained as described for deoxyinosine-containing DNA.[3b]

N-tert-Butoxycarbonyl-(4′*R*)-(6-chloro-9*H*-purin-9-yl)-(2′*R*)-proline diphenylmethyl ester (2)

A mixture of tosylate **1**[13] (0.5863 g, 1.06 mmol), 6-chloropurine (0.246 g, 1.60 mmol) and anhydrous K_2CO_3 (0.35 g, 2.50 mmol) in anhydrous DMF (5 mL) was stirred under nitrogen at 80 °C for 8 h. Water (30 mL) was added to the reaction and the suspension was extracted with dichloromethane (3 × 30 mL). The organic phase was washed with water, dried over $MgSO_4$ and evaporated to give the crude product, which was purified by column chromatography eluting with hexanes : ethyl acetate 1 : 3 on silica gel to obtain compound **2** as a white amorphous solid (0.1583 g, 27%). TLC R_f = 0.12 (EtOAc : hexanes 1 : 1); $[\alpha]_D^{25}$ = +7.86 (*c* 1.02 in $CHCl_3$); ν_{max}/cm^{-1} 2977, 1748, 1704, 1590, 1561, 1400; δ_H (400 MHz, $CDCl_3$) 1.32 and 1.49 (9H, s, C*H*$_3$ Boc), 2.53 and 2.95 [2H, m C*H*$_2$(3′)], 4.00–4.16 and 4.58–4.69 [2H, m, C*H*$_2$(5′)], 4.17 [1H, m, C*H*(2′)],

5.17 [1H, m, C*H*(4′)], 6.85 (1H, s, C*H*Ph$_2$), 7.19–7.29 (m, 10H, Ph), 8.12 and 8.19 [1H, m, purine C(2)*H* rotamers], 8.64 [1H, s, purine C(8)*H*]; δ_C (100 MHz, $CDCl_3$) 28.0 and 28.3 (*C*H$_3$ Boc rotamers), 34.6 and 35.8 [*C*(3′)H$_2$ rotamers], 50.0 and 50.6 [*C*(5′)H$_2$ rotamers], 52.7 and 53.3 [*C*(4′)H rotamers], 57.5 [*C*(2′)H], 77.7 and 78.1 (*C*HPh$_2$ rotamers), 81.2 (*C*CH$_3$ Boc), 127.1, 128.4 and 128.5 (Dpm Ar *C*H), 131.8 (purine *C*5), 139.2 (Dpm *C*), 143.5 [purine *C*(8)H], 151.5 (purine *C*4), 151.7 [purine *C*(2)H], 153.4 and 153.6 (*C*O Boc rotamers), 170.4 (*C*O proline); HRMS (ESI-TOF) *m*/*z* calcd for $C_{28}H_{29}ClN_5O_4Na$ [M + Na]$^+$ 556.1727; found 556.1768.

N-9*H*-Fluoren-9-ylmethoxycarbonyl-(4′*R*)-[6-oxo-1*H*-purin-9(6*H*)-yl]-(2′*R*)-proline (3)

Compound **2** (0.1583 g, 0.29 mmol) was dissolved in trifluoroacetic acid containing 10% anisole (5 mL) at room temperature. After 1 h, the solvent was removed by a gentle stream of nitrogen. The residue was triturated and washed with diethyl ether. The white solid formed was collected by filtration and heated with 2 M HCl (10 mL) under reflux for 2 h. The solvent was removed by rotary evaporation and the residue dissolved in 1 : 1 H_2O : MeCN (2 mL) and treated with solid $NaHCO_3$ until pH 8. FmocOSu (0.223 g, 0.6 mmol) was then added in small portions with stirring. After stirring at room temperature overnight, the solvent was removed by rotary evaporation. The residue was diluted with water (10 mL) and extracted with diethyl ether (3 × 20 mL). After purging the extracted aqueous layer to remove the dissolved ether with a gentle stream of N_2, the pH was adjusted to 2 with conc. HCl. The white precipitate was collected by filtration, washed with water, diethyl ether and dried under vacuum to afford **3** as a white solid (0.109 g, 77%). TLC R_f = 0.14 (EtOAc : MeOH 9 : 1 containing 0.1% TFA); $[\alpha]_D^{24}$ = −7.04 (*c* 0.50 in DMF); ν_{max}/cm^{-1} 1698, 1450, 1420, 1352 cm^{-1}; δ_H (400 MHz, DMSO-d$_6$) δ_H 2.55–2.68 and 2.85–3.02 [2H, 2 × m, C*H*$_2$(3′) rotamers], 3.80–3.88 [1H, m, 1 × C*H*$_2$(5′)], 4.10–4.34 and 4.52–4.56 [5H, m, 1 × C*H*$_2$(5′), C*H*(2′), Fmoc C*H* and C*H*$_2$ rotamers], 5.11 [1H, m, C*H*(4′)], 7.35 (2H, m, Fmoc Ar C*H*), 7.43 (2H, m, Fmoc Ar C*H*), 7.69 (2H, d *J* = 7.3 Hz, Fmoc Ar C*H*), 7.91 (2H, d *J* = 7.4 Hz, Fmoc Ar C*H*), 8.10 [1H, br m, hypoxanthine C(2)*H*], 8.22 [1H, s, hypoxanthine C(8)*H*], 12.40 (1H, s, N*H*), 12.80 (1H, br s, COOH); δ_C (100 MHz, DMSO-d$_6$) 33.8 and 34.9 [*C*H$_2$(3′) rotamers], 46.6 (Fmoc aliphatic *C*H), 49.8 and 50.3 [*C*(5′)H$_2$ rotamers], 51.9 and 52.5 [*C*(4′)H rotamers] 57.1 and 57.4 [*C*(2′)H rotamers], 66.9 and 67.3 (Fmoc *C*H$_2$ rotamers), 120.1 (Fmoc Ar *C*H), 124.2 (hypoxanthine *C*5), 125.2 (Fmoc Ar *C*H), 127.2 (Fmoc Ar *C*H), 127.7 (Fmoc Ar *C*H), 138.8 [hypoxanthine *C*(8)H], 140.7 (Fmoc Ar *C*), 143.6 (Fmoc Ar *C*), 145.6 [hypoxanthine *C*(2)H], 148.3 [hypoxanthine *C*(4)], 153.8 (Fmoc *C*O), 156.5 [hypoxanthine *C*(6)O], 172.4 and 172.9 (proline *C*O rotamers); HRMS (ESI-TOF) *m*/*z* calcd for $C_{25}H_{22}N_5O_5$ [M + H]$^+$ 472.1621; found 472.1592.

PNA synthesis

All acpcPNAs were synthesized from the Fmoc-protected acpcPNA monomers (A, T, C, G, I) and ACPC spacer by manual Fmoc-solid phase peptide synthesis on a Tentagel S-RAM resin carrying the acid-labile Rink amide linker.[11] The A, T and C monomers and the spacer were incorporated as Fmoc/pentafluorophenyl ester whereas the G and I monomers were incorporated *via* HATU activation of the corresponding Fmoc acid. L-Lysinamide was included at the C-termini to improve water solubility of the PNA. The PNA was end-capped by acetylation (**PNA1–PNA9**) or by reaction with 5(6)-carboxyfluorescein *N*-succinimidyl ester. After ether precipitation and HPLC purification, the identity and purity were determined by MALDI-TOF mass spectrometry and by reverse phase HPLC using the following conditions: C18 column 4.6 × 50 mm, particle size 3 μm; mobile phase A = 0.1% TFA in water; B = 0.1% TFA in MeCN; gradient = 10–90% B over 25 min with 5 min equilibration time at a flow rate of 0.5 mL min^{-1}.

Thermal denaturation experiments

UV melting experiments were performed on a CARY 100 UV Spectrophotometer (Varian Inc., Australia) equipped with a Peltier temperature controller and a thermal analysis software. The samples for thermal denaturation measurement were prepared by mixing calculated amounts of stock oligonucleotide, PNA, buffer solutions and deionized water to give the final concentration of PNA = 1.0 μM, DNA = 1.0 μM in 10 mM sodium phosphate pH 7.0 and 100 mM sodium chloride at a total volume of 1.0 mL. The sample in a 10 mm quartz cell with Teflon stopper was equilibrated at 20 °C for 10 min. The A_{260} was recorded in steps from 20 to 90 °C with a temperature increment of 1.0 °C min^{-1}. The recorded temperature was corrected by a linear equation obtained from the relationship between the recorded temperature and actual temperature measured by a built-in temperature probe. Thermodynamic parameters were obtained from melting curves by nonlinear curve fitting.[24] The data obtained are average values from at least three melting curves for each PNA sample at a concentration of PNA = DNA = 1.0 μM (c_t = 2.0 μM). The $\Delta G°$ at 298 K and T_m were calculated from the $\Delta H°$ and $\Delta S°$ values.

Fluorescence experiments

Fluorescence experiments were performed on a CARY Eclipse Spectrofluorometer (Varian Inc., Australia). The samples were prepared by mixing calculated amounts of stock PNA, buffer solutions and deionized water to give the final concentration of PNA = 0.05 μM in 10 mM sodium phosphate pH 7.0 and 100 mM sodium chloride at a total volume of 1.0 mL. The fluorescence spectra were measured in a 10 mm quartz cell with Teflon stopper at 20 °C using the following instrument parameters: Excitation slit = 5 nm, Emission slit = 5 nm, Excitation filter = Auto, Emission filter = Auto, Smoothing = Moving average.

CD spectroscopy

Circular dichroism experiments were performed on a JASCO J810 Spectropolarimeter (Jasco, Japan). The samples were prepared by mixing calculated amounts of stock oligonucleotide, PNA, buffer solutions and deionized water to give the final

concentration of PNA = 1.0 μM, DNA = 1.0 μM in 10 mM sodium phosphate pH 7.0 and 100 mM sodium chloride at a total volume of 1.0 mL. The sample in a 10 mm quartz cell with Teflon stopper was equilibrated at 20 °C for 10 min. The CD spectra were measured at 20 °C from 200 to 400 nm with scanning speed 200 nm min^{-1} and averaged 4 times then subtracted from a spectrum of the blank (10 mM sodium phosphate buffer pH 7.0, 100 mM NaCl) under the same conditions.

Conclusion

We have demonstrated that hypoxanthine in acpcPNA showed a remarkable preference for pairing with C over other natural DNA bases (A, G, T) under a variety of sequence contexts. The study revealed that the base G in acpcPNA can be replaced with hypoxanthine without affecting the binding affinity and specificity of pairing to its DNA target. This suggests potential use of this strategy to solve many problems associated with G-rich acpcPNA such as synthetic difficulty, formation of stable secondary structures and fluorescence quenching.

Acknowledgements

Financial support to this work from the Thailand Research Fund (RTA5280002), the Thai government stimulus package 2 (TKK2555, SP2) under the Project for Establishment of Comprehensive Center for Innovative Food, Health Products and Agriculture, the National Research University of CHE and the Ratchadaphiseksomphot Endowment Fund (AM1006A), the Ratchadaphiseksomphot Endowment Fund from Chulalongkorn University (postdoctoral fellowship to WM) and the Thailand Research Fund and Chulalongkorn University through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0316/2550, to NY and TV).

References

1 F. H. C. Crick, *J. Mol. Biol.*, 1996, **19**, 548.
2 (*a*) F. H. Martin, M. M. Castro, F. Aboul-ela and I. Tinoco Jr, *Nucleic Acids Res.*, 1985, **13**, 8927; (*b*) E. Ohtsuka, S. Matsuki, M. Ikehara, Y. Takahashi and K. Matsubara, *J. Biol. Chem.*, 1985, **260**, 2605.
3 (*a*) S. C. Case-Green and E. M. Southern, *Nucleic Acids Res.*, 1994, **22**, 131; (*b*) N. E. Watkins Jr and J. SantaLucia Jr, *Nucleic Acids Res.*, 2005, **33**, 6258.
4 X. Sun and J. K. Lee, *J. Org. Chem.*, 2010, **75**, 1848.
5 A. A. Koshkin, *J. Org. Chem.*, 2004, **69**, 3711.
6 Z. Timár, S. Bottka, L. Kovács and B. Penke, *Nucleosides Nucleotides*, 1999, **18**, 1131.
7 (*a*) P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497; (*b*) M. Egholm, O. Buchardt, P. E. Nielsen and R. H. Berg, *J. Am. Chem. Soc.*, 1992, **114**, 1895; (*c*) M. Egholm, P. E. Nielsen, O. Buchardt and R. H. Berg, *J. Am. Chem. Soc.*, 1992, **114**, 9677.
8 R. E. Hudson, M. Goncharenko, A. P. Wallman and F. Wojciechowski, *Synlett*, 2005, 1442.
9 (*a*) G. J. Sanjayan, V. R. Pedireddi and K. N. Ganesh, *Org. Lett.*, 2000, **2**, 2825; (*b*) R. Vysabhattar and K. N. Ganesh, *Tetrahedron Lett.*, 2008, **49**, 1314.
10 (*a*) K. F. MacKinnon, D. F. Qualley and S. A. Woski, *Tetrahedron Lett.*, 2007, **48**, 8074; (*b*) H. Challa, M. L. Styers and S. A. Woski, *Org. Lett.*, 1999, **1**, 1639; (*c*) H. Challa and S. A. Woski, *Tetrahedron Lett.*, 1999, **40**, 8333.
11 (*a*) T. Vilaivan and C. Srisuwannaket, *Org. Lett.*, 2006, **8**, 1897; (*b*) C. Vilaivan, C. Srisuwannaket, C. Ananthanawat, C. Suparpprom, J. Kawakami, Y. Yamaguchi, Y. Tanaka and T. Vilaivan, *Artif. DNA PNA XNA*, 2011, **2**, 50; (*c*) C. Suparpprom, C. Srisuwannket, P. Sangvanich and T. Vilaivan, *Tetrahedron Lett.*, 2005, **46**, 2833.
12 C. Ananthanawat, T. Vilaivan, V. P. Hoven and X. D. Su, *Biosens. Bioelectron.*, 2010, **25**, 1064.
13 G. Lowe and T. Vilaivan, *J. Chem. Soc., Perkin Trans. 1*, 1997, 547.
14 V. Antle and C. A. Caperelli, *Nucleosides Nucleotides*, 1999, **18**, 1911.
15 (*a*) S. Hikishima, M. Isobe, S. Koyanagi, S. Soeda, H. Shimeno, S. Shibuya and T. Yokomatsu, *Bioorg. Med. Chem.*, 2006, **14**, 1660; (*b*) R. A. Tromp, R. F. Spanjersberg, J. K. von Frijtag Drabbe Knzel and A. P. IJzerman, *J. Med. Chem.*, 2005, **48**, 321; (*c*) B. R. Baker and P. M. Tanna, *J. Org. Chem.*, 1965, **30**, 2857.
16 M. L. Peterson and R. Vince, *J. Med. Chem.*, 1991, **34**, 2787.
17 L. R. Rutledge and S. D. Wetmore, *Phys. Chem. Chem. Phys.*, 2012, **14**, 2743.
18 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402.
19 E. Paleček, *Electroanalysis*, 1996, **8**, 7.
20 J. Wang, G. Rivas, J. R. Fernandes, J. L. L. Paz, M. Jiang and R. Waymire, *Anal. Chim. Acta*, 1998, **375**, 197.
21 C. A. M. Seidel, A. Schulz and M. H. M. Sauer, *J. Phys. Chem.*, 1996, **100**, 5541.
22 I. Nazarenko, R. Pires, B. Lowe, M. Obaidy and A. Rashtchian, *Nucleic Acids Res.*, 2002, **30**, 2089.
23 A. O. Crockett and C. L. Wittwer, *Anal. Biochem.*, 2001, **290**, 89.
24 J. Kawakami, Y. Tanaka and K. Kishimoto, *Nucleic Acids Symp. Ser.*, 2009, **53**, 227.