# Properties of pseudo-complementary DNA substituted with weakly pairing analogs of guanine or cytosine

Georges Lahoud[1], Victor Timoshchuk[2], Alexandre Lebedev[2], Khalil Arar[3], Ya-Ming Hou[1] and Howard Gamper[1,*]

[1]Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA 19107, [2]TriLink BioTechnologies, 9955 Mesa Rim Road, San Diego, CA 92121, USA and [3]Sigma-Proligo, 1 Rue Delaunay, Paris 75011, France

## ABSTRACT

A straightforward enzymatic protocol for converting regular DNA into pseudo-complementary DNA could improve the performance of oligonucleotide microarrays by generating readily hybridizable structure-free targets. Here we screened several highly destabilizing analogs of G and C for one that could be used with 2-aminoadenine (nA) and 2-thiothymine (sT) to generate structure-free DNA that is fully accessible to complementary probes. The analogs, which included bioactive bases such as 6-thioguanine (sG), 5-nitrocytosine (NitroC), 2-pyrimidinone (P; the free base of zebularine) and 6-methylfuranopyrimidinone (MefP), were prepared as dNTPs and evaluated as substrates for T7 and Phi29 DNA polymerases that lacked editor function. Pairing properties of the analogs were characterized by solution hybridization assays using modified oligonucleotides or primer extension products. P and MeP did not support robust primer extension whereas sG and NitroC did. In hybridization assays, however, sG lacked discrimination and NitroC paired too strongly to C. The dNTPs of two other base analogs, 7-nitro-7-deazahypoxanthine (NitrocH) and 2-thiocytosine (sC), exhibited the greatest promise. Either analog could be used with nA and sT to generate DNA that was nearly structure-free. Hybridization of probes to these modified DNAs will require the development of base analogs that pair strongly to NitrocH or sC.

## INTRODUCTION

Secondary structure in single-stranded DNA and RNA is a significant impediment to the use of short oligonucleotide probes in hybridization assays (1–5). Under conditions where oligonucleotides form stable hybrids, long single-stranded nucleic acid targets are highly structured and many sequences are inaccessible to short probes. This phenomenon becomes increasingly restrictive as the length of a probe is reduced (6–9). If oligonucleotides <20 nt in length could readily access any sequence in a given nucleic acid target without loss of specificity, the detection of allelic variants by hybridization would benefit. For example, the availability of structure-free nucleic acid samples would significantly improve the sensitivity, resolution and reliability of high density oligonucleotide microarrays that are used to detect single nucleotide polymorphisms in whole genome association studies.

Pseudo-complementary DNA targets would be ideally suited for use with short oligonucleotide probes since they are by definition structure-free molecules that can participate in hybridization reactions. This strategy requires the development of two complementary sets of bases, with one set used in the target DNA and the other set used in the probes. Importantly, the set of bases used in the target DNA should not interact with one another thus leading to a structure-free target. The bases present in the probe can form complementary pairs with each other since secondary structure is less of an issue. In one format the probes contain regular bases and the target DNA contains modified bases. Alternatively, both the target and probe molecules can be prepared with a mixture of regular and modified bases. Regardless of which strategy is employed, practical application of the technology requires that pseudo-complementary DNA be readily prepared by primer extension or asymmetric PCR from the sample of interest. In either case the template DNA must contain regular or modified bases that support uptake of pseudo-complementary base analogs into the DNA product.

There is an ongoing need for analogs of G and C that can be used with 2-aminoadenine (nA) and 2-thiothymine (sT) to generate pseudo-complementary DNA. nA and sT

*To whom correspondence should be addressed. Tel: +1 215 503 9798; Fax: +1 215 503 4954; Email: howard.gamper@jefferson.edu

form stable nA-T and A-sT base pairs but together function as a mismatch (10). The dNTPs of these bases are readily incorporated into DNA (11). We recently reported that 7-ethyl-7-deazaguanine (EtcG) and $N^4$-ethylcytosine (EtC) together with the corresponding methyl derivatives provide a route to pseudo-complementary DNA (12). These analogs, which are weakly to moderately destabilizing when base paired to C or G, interact poorly with one another. DNA substituted with nA, sT, EtcG and EtC is structure-free but is relatively hydrophobic and challenging to synthesize. Here we examine several highly destabilizing analogs of G and C for one that can confer pseudo-complementarity to DNA in the presence of nA and sT.

Several of these analogs have pharmacological significance including 6-thioguanine (sG) (13), 2-pyrimidinone (P) (14), 6-methylfuranopyrimidinone (MefP) (15) and 5-nitrocytosine (NitroC) (16). sG is an anticancer agent commonly used for treating leukemia. P is a potent inhibitor of DNA methylation and may have application in the epigenetic therapy of cancer. Both NitroC and certain derivatives of MefP possess antiviral properties. While sG and P must be incorporated into DNA to exhibit activity this may not be the case of MefP and NitroC. The structure-activity relationships reported here provide an opportunity to compare the enzymatic uptake and base pairing properties of these bases.

## EXPERIMENTAL PROCEDURES

### Materials

Unmodified oligodeoxynucleotides were purchased from IDT (Coralville, IA) and locked nucleic acids were a gift from Sigma Proligo (Paris, France). Modified oligodeoxynucleotides that contained H or sG were synthesized using phosphoramidites from Glen Research (Sterling, VA). Vent (exo-) DNA polymerase was obtained from New England Biolabs (Ipswich, MA). Exo- T7 DNA polymerase (Sequenase) was purchased from USB (Cleveland, OH). The N62D mutant of φ29 DNA polymerase was obtained from M. Salas (Instituto de Biologia Molecular 'Eladio Viñuela', Madrid, Spain).

### Deoxynucleoside triphosphates

dNTPs of nA, sT, hypoxanthine (H), sG, 2-thiocytosine (sC), 7-deazaguanine (cG) and P are commercially available from TriLink Biotechnologies (San Diego, CA). Deoxynucleosides of 7-deazahypoxanthine (cH) and 6-methylfuranopyrimidinone (MefP) were purchased, respectively, from ChemGenes (Wilmington, MA) and Berry & Associates (Dexter, MI). Deoxynucleosides of 7-iodo-7-deazahypoxanthine (IcH), 7-nitro-7-deazahypoxanthine (NitrocH), 7-nitro-7-deazaguanine (NitrocG) and 5-nitrocytosine (NitroC) were synthesized as described in Supplementary Data. Nucleosides were phosphorylated by stirring 0.3–0.4 mmol quantities with 1.5–3 equivalents of $POCl_3$ in 2–3 ml of dry trimethylphosphate for several hours at 0°C. Reactions were terminated by adding 5 equivalents of tributylammonium pyrophosphate (0.5 M in acetonitrile) and 4 equivalents of tributylamine

with vigorous stirring. After 5 min 5–10 ml of 1 M triethylammonium bicarbonate (TEAB) was added with stirring for an additional 30 min. Each dNTP was isolated and purified by a combination of anion exchange (water–1 M TEAB) and reverse phase (100 mM TEAB–acetonitrile) chromatography. The isolated triethylammonium salt of the dNTP was converted into the Li salt by precipitation from 2% lithium perchlorate in acetone. Isolated yields were 15–30%. The synthesis of 7-nitro-7-deaza-2′-deoxyguanosine 5′-triphosphate required the presence of 1,8-bis(dimethyl-amino)naphthalene (Proton Sponge, 2 equiv.) to prevent degradation of the nucleoside base (17).

### Primer extension

Most primer extension reactions were conducted at 30°C for 30 min in the presence of 0.6 μM $^{32}$P end-labeled primer, 1.2 μM template, 200 μM each dNTP, and 0.05 μg/μl φ29 N62D DNA polymerase in 40 mM Tris–HCl, pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 4 mM $(NH_4)_2SO_4$ (12). Reactions that contained T7 instead of φ29 DNA polymerase were carried out as previously described (11). Labeled DNA products were resolved by electrophoresis of reaction aliquots through a 7 M urea–12% polyacrylamide gel. Since the product of primer extension differed in length from its template, denaturing gel electrophoresis afforded pure single-stranded target for use in hybridization reactions. Primer extension products with a hairpin configuration were not gel purified prior to use as hybridization targets. Rather, these reactions were diluted 15-fold with water and heated 2 min at 95°C in the presence of an oligonucleotide (0.4 μM) that was complementary to the 3′ overhang of the template. Upon quick cooling in an ice bath the template and product strands collapsed into separate hairpins. The oligonucleotide prevented reannealing of template and product hairpins by blocking hybridization between their complementary single-stranded tails.

### Hybridization studies

The hybridization assays used here have also been described in previous publications (11,18). Briefly, hybridizations were conducted in 5 mM $MgCl_2$, 25 mM NaCl, 20 mM HEPES, pH 7.5 using synthetic oligonucleotides as probes and end-labeled primer extension products or chemically synthesized oligomers as targets. Apparent melting temperatures of hairpins were determined by incubating a trace amount of labeled hairpin with 5 μM of a probe equal in length and complementary to one arm of the hairpin. Incubations were conducted over a wide range of temperatures using 5°C increments. After 5 min each reaction aliquot was quenched in an ice bath and analyzed by gel mobility shift through a 12% non-denaturing gel run in the presence of 5 mM $MgCl_2$ in a cold room. A somewhat different protocol was used to determine the apparent melting temperatures of intermolecular Watson–Crick hybrids. Probes (1 μM) were hybridized to a trace amount of radiolabeled single-stranded target by raising the temperature to 90°C and then lowering it to 4°C in a thermocycler. Aliquots of the hybrids were again incubated over a wide temperature range. After 5 min

free probe was sequestered by adding a competitor oligonucleotide to 10 μM and the reaction was quenched in an ice bath. Competitor oligonucleotides contained regular bases and were identical in backbone and complementary in sequence to the probe. Each reaction set was then subjected to gel mobility shift analysis as described above to construct a melting curve. A similar protocol was followed for the determination of specificity. In this case the single-stranded target was hybridized to one perfect-match and three mismatch probes at a fixed temperature under stringent conditions. These reactions were usually quenched by adding a competitor oligonucleotide that was complementary to the perfect-match probe. Tiling experiments consisted of hybridizing 5 μM probe to radiolabeled target for 5–10 min at the indicated temperature. No competitor oligonucleotide was added to the tiling reactions prior to electrophoretic analysis.

## RESULTS

### Selection and evaluation of base analogs

Candidate bases (Figure 1A) were synthesized as dNTPs since efficient enzymatic incorporation into DNA was deemed an important attribute. Primer extension was used to prepare single-stranded and hairpin DNAs that contained each of the analogs either alone or in combination with nA and sT (Figure 1B). Apparent melting temperatures of the modified DNA products were determined by gel mobility shift assay and allowed us to ascertain the stability and specificity of nonnatural G-C base pairs. A 25 bp DNA hairpin (HP25) was used extensively in these studies and its $T_m$ provided a convenient way to measure hybrid stability in the presence of an analog of G or C. Tiled sets of 12-mer DNA and 8-mer LNA probes were hybridized to the modified hairpins as a test of whether a given set of bases rendered the target structure-free.

The analogs investigated here were known or expected to pair very weakly with their standard base complements. cH, IcH and NitrocH lack the C2 exocyclic amino group of G and pair with C through two hydrogen bonds. By analogy with the corresponding cG derivatives, cH should be less stable than H (the free base of inosine) while IcH and NitrocH should be more stable (12,19–21). The 7-deaza substitution in these analogs has the advantage of eliminating self-association and reducing the likelihood of mispairing to G (22). We have previously shown that DNA substituted with nA, sT and H is pseudo-complementary but does not exhibit specificity in its pairing due to degeneracy of H (11,23). P [the free base of zebularine; (24)] and MefP (25) are analogs of C that lack the C4 exocyclic amino group. Pairing of these bases to G is limited to two hydrogen bonds. A third analog of C, NitroC (16,26), can form three hydrogen bonds with G. However, the strongly electron withdrawing nitro group should weaken the central hydrogen bond. In combination with nA and sT any one of these analogs should reduce the secondary structure of DNA provided that the respective dNTPs are utilized by DNA polymerase.

Also of interest were sG (27–30) and sC (31). The sG–C and G–sC base pairs are weakened by steric clash between an amino group on one base and a thioketo group on the other base (Figure 1C). In this respect they bear similarity to the nA–sT mismatch. While either of these couples would be expected to work in concert with the nA–sT couple to prevent intramolecular pairing in DNA, successful hybridization might require the use of modified probes that hydrogen bond to sG or sC without clash. For instance, analogs of P or H might be developed for stable and specific pairing to sG or sC, respectively. Several groups have reported that dsGTP is incorporated into DNA using regular templates even though the sG–C pair is weak and distorted (32–36). The properties of dsCTP have not been described and its uptake into DNA might require the presence of an 'adaptor' base in the template DNA.

### Enzymatic incorporation of dGTP and dCTP analogs

Each analog was substituted for dGTP or dCTP in primer extension reactions catalyzed by Sequenase. This engineered version of T7 DNA polymerase lacks proofreading function and is noted for its ability to utilize a variety of base analogs (37). Many of the analogs prematurely terminated synthesis if the template contained a stable hairpin element. For instance, when dsGTP was substituted for dGTP, Sequenase was unable to synthesize a full length 75 mer that contained an embedded 25 bp hairpin (Figure 2). Synthesis terminated midway into the hairpin element after addition of two consecutive sG residues. We postulate that synthesis was blocked because the 3'-end of the nascent DNA product was displaced from the template strand by the homologous strand of the hairpin element. In this situation strand displacement would be thermodynamically driven since two weak sG–C base pairs would be replaced by two much stronger G–C base pairs. We cannot, however, exclude the possibility that two consecutive sG's at the primer terminus might fray the duplex such that Sequenase, which does not have 3'-exonuclease activity, would not be able to add the next dNTP.

We reduced this block to synthesis by switching to φ29 DNA polymerase (DNAP), an enzyme which has potent strand displacement activity and can readily extend a primer through a double-stranded DNA. The catalytic center of this enzyme resides within a 'tunnel' that accommodates the template and nascent product strands (38) and conceivably reduces fraying by stabilizing the nascent hybrid. The non-template strand of a dsDNA template is excluded from the active site thus also preventing strand displacement as described for Sequenase. Consequently, φ29 DNAP could utilize weakly pairing dNTPs to synthesize both single-stranded and hairpin products. The uptake of such dNTPs could be further enhanced by using the N62D mutant of φ29 DNAP (39). This mutation severely depresses the proofreading function of wild-type enzyme without impairing its ability to carry out strand displacement synthesis. By nearly eliminating 3'–5' exonuclease activity, possible hydrolysis of weak pairing bases from the 3'-end of the nascent DNA was minimized.
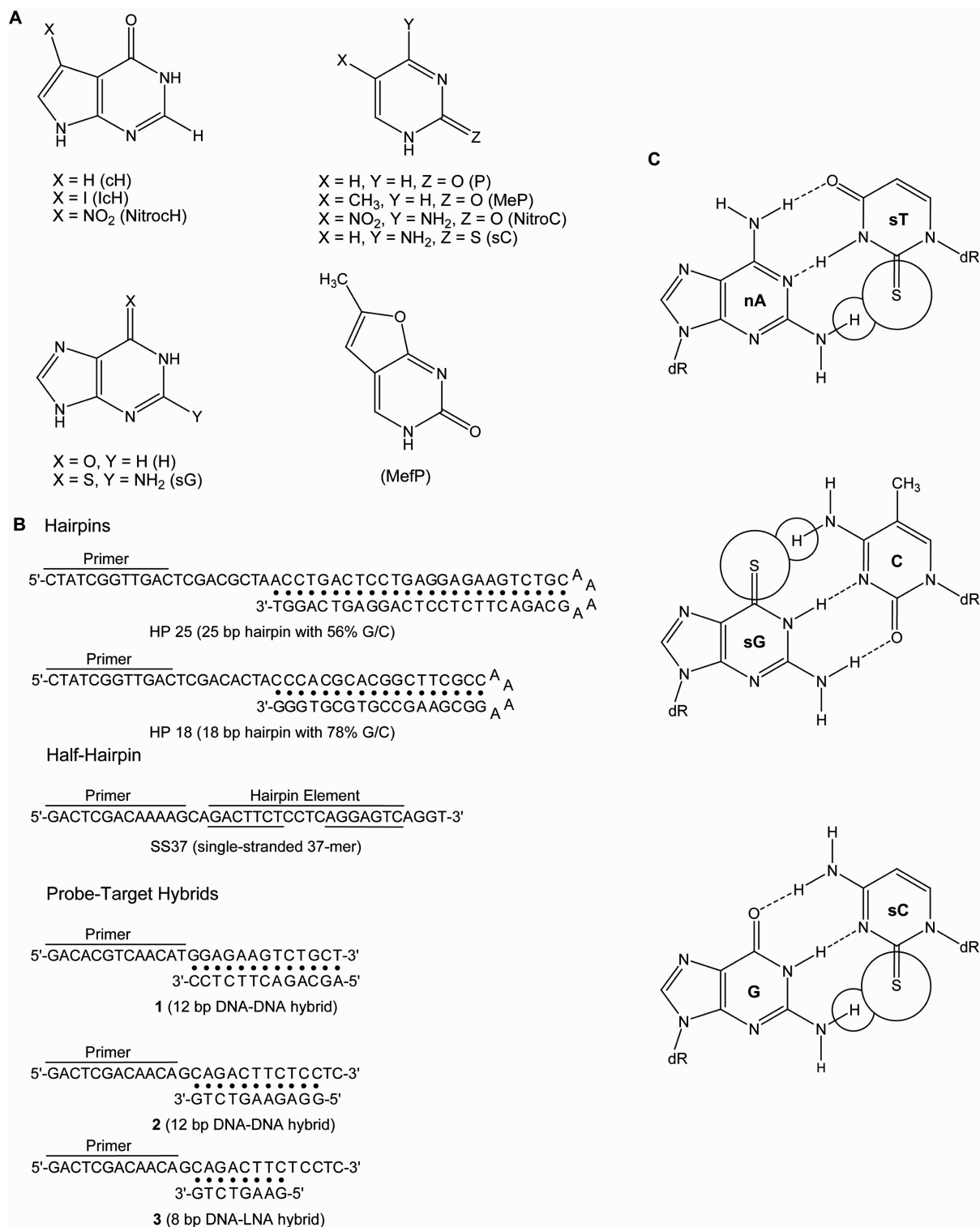
**Figure 1.** (**A**) Analogs of G and C evaluated in this study; (**B**) Primer extension products prepared using Phi29 N62D DNA polymerase in the presence of regular and base-modified dNTPs. Specificity of pairing was determined by hybridizing gel-purified 26-mers to perfect-match and mismatch probes under stringent conditions; (**C**) Base pairing in nA–sT, sG–C and G–sC couples is weakened by steric clash between a thioketo group on one base and an amino group on the other.

In our screening of dNTP analogs we used the N62D mutant of φ29 DNAP.

Table 1 summarizes the yields of single-strand (SS37) and hairpin (HP25) primer extension products (Figure 1)

synthesized with different analogs of dGTP or dCTP. This table also contains the results for a parallel set of reactions that also contained dnATP and dsTTP in place of dATP and dTTP. Anticipating the probable use of
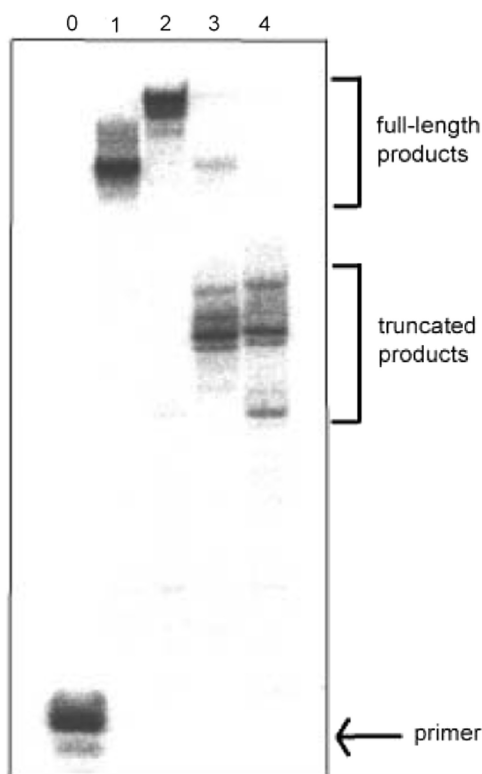
**Figure 2.** Premature termination of primer extension in the presence of dsGTP. Sequenase was used to synthesize a 25 bp hairpin (HP25) in the presence of different combinations of dNTPs. Extension of the radiolabeled primer was monitored by electrophoresis in a 12% sequencing gel. dNTPs of the following bases were used: (i) A,T,G,C; (ii) nA,sT,G,C; (iii) A,T,sG,C and (iv) nA,sT,sG,C. The unmodified hairpin in lane 1 retained a native conformation in 7 M urea and so ran faster than the fully denatured nA/sT-substituted hairpin in lane 2.

**Table 1.** Relative uptake of dGTP and dCTP analogs into SS37 and HP25 primer extension products by φ29 N62D DNA polymerase

|  | Single-strand | Hairpin | Single-strand | Hairpin |
|---|---|---|---|---|
| dGTP analogs | dNTPs of A,T,C | | dNTPs of nA,sT,C | |
| G | 1.00 | 1.00 | 1.12 | 0.84 |
| cG | 1.07 | 1.07 | 1.11 | 0.82 |
| NitrocH | 0.92 | 1.17 | 1.16 | 0.66 |
| IcH | 1.14 | 0.35 | 1.27 | 0.58 |
| H | 1.03 | 1.17 | 1.06 | 0.81 |
| cH | 1.04 | 0.89 | 1.01 | 0.69 |
| sG | 0.98 | 0.98 | 0.98 | 0.10 |
| dCTP analogs | dNTPs of A,T,cG | | dNTPs of nA,sT,cG | |
| C | 1.07 | 1.07 | 0.93 | 0.96 |
| NitroC | 0.91 | 0.88 | 1.03 | 0.62 |
| sC | 0.69 | 0.03 | 0.33 | 0.02 |
| MefP | 0.87 | 0.37 | 0.52 | 0 |
| P | 0.01 | 0 | 0.01 | 0 |
| sC[a] | ND | 0.87 | ND | 0.78 |

[a]Template for synthesis contained H in place of G.
ND, not determined.

a deazaguanine derivative in preparing pseudo-complementary DNA, the reactions with modified dCTPs contained dcGTP in place of dGTP. This analog of G is readily incorporated into DNA and does not

self-associate (19,20,40–42). Yields in each case were determined relative to the amount of each primer extension product obtained when using regular dNTPs.

All of the dGTP analogs functioned effectively with A, T and C or nA, sT and C when preparing single-stranded DNAs. However, synthesis of the hairpin was less reliable. In the presence of A, T and C the IcH analog gave reduced yield while in the presence of nA, sT and C all of the analogs were less effective. This was particularly true for sG where synthesis of full-length hairpin was reduced by 90%. Uptake of H and NitrocH was further investigated using Vent (exo-) DNA polymerase to carry out one cycle of asymmetric PCR in the presence of nA and sT. Whereas NitrocH supported synthesis of a 600-mer product that contained some highly G/C-rich secondary structure elements, H did not. In the presence of the more weakly pairing H analog, primer extension was blocked by these elements.

Incorporation of the dCTP analogs into SS37 and HP25 was highly variable. NitroC gave moderate to good yields of all four primer extension products. Conversely, P did not support synthesis of any full-length product and sC only supported synthesis of single-stranded DNAs. MefP improved upon sC by also supporting limited synthesis of hairpin in the absence of nA and sT. In an attempt to improve the utilization of dsCTP, we investigated whether using a template in which H was substituted for G could direct incorporation of sC into the complementary hairpin product. This strategy was supported by an earlier study which had shown that incorporation of rsCTP into RNA was more efficient when templated by H instead of G (43). With steric clash now eliminated by using H as the templating base, sC was incorporated into the hairpin product along with cG or nA, sT and cG. Although use of H in a primer extension reaction is a cause for concern, polymerases appear to maintain specificity in the presence of this base (44). Using an analogous approach, the synthesis of hairpin substituted with sG, nA and sT might be made more robust by using a template in which C has been enzymatically replaced by a readily incorporated analog of MefP.

## Relative pairing strengths of G/C analogs

Melting properties of the radiolabeled hairpins prepared using Phi29 DNA polymerase provided a convenient read-out of how each analog decreased hybrid stability either alone or in combination with nA and sT. Apparent melting temperatures were determined using a gel mobility shift assay. Briefly, each hairpin was incubated with a large molar excess of 25-mer probe at temperatures ranging from 10°C to 80°C. The probe, which was complementary to the distal arm of the hairpin stem, was expected to hybridize to the hairpin at a temperature approximating its true $T_m$. Melting profiles were obtained by quenching each hybridization reaction in an ice bath and analyzing for hybrid formation by non-denaturing gel electrophoresis. The midpoint in the melting curve was taken as equal to an apparent $T_m$. When hybridizing to hairpins that contained A and T, the 25-mer probe was synthesized with a chimeric LNA–DNA backbone to

**Table 2.** Apparent melting temperatures of the HP25 hairpin substituted with analogs of G or C

| Substitution | $T_m$ (°C) | $\Delta T_m$/analog |
|---|---|---|
| None | 76 | |
| nA + sT | 40 | |
| G Analogs | | |
| cG | 68 | −0.57 |
| NitrocH | 64 | −0.86 |
| IcH | 61 | −1.1 |
| H | 55 | −1.5 |
| cH | 50 | −1.9 |
| sG | 46 | −2.1 |
| C Analogs | | |
| NitroC | 57 | −1.4 |
| sC | 35 | −2.9 |

**Table 3.** Apparent melting temperatures of hybrids substituted in one strand with analogs of G or C

| Substitution | $T_m$ (°C) | $\Delta T_m$/analog |
|---|---|---|
| C analogs (12 bp DNA–DNA hybrid 1) | | |
| C | 54 | – |
| NitroC | 47 | −3.5 |
| sC | 45 | −4.5 |
| MefP | 36 | −9 |
| G analogs (12 bp DNA–DNA hybrid 2) | | |
| G | 58 | – |
| NitrocG | 53 | −2.5 |
| cG | 52 | −3 |
| H | 47 | −5.5 |
| NitrocH | 46 | −6 |
| IcH | 45 | −6.5 |
| cH | 43 | −7.5 |
| G analogs (8 bp LNA–DNA hybrid 3) | | |
| NitrocG | 69 | +9 |
| G | 60 | – |
| cG | 60 | – |
| NitrocH | 51 | −9 |
| cH | 45 | −15 |
| H | 44 | −16 |
| IcH | 43 | −17 |
| sG | 42 | −18 |

enhance stability of the probe-target hybrid. Probes with a locked nucleic acid (LNA) backbone hybridize with high affinity to complementary sequences (45). Since enhanced stability was unnecessary when hybridizing to hairpins substituted with nA and sT, the probe to these hairpins contained a DNA backbone. Strand invasion of the modified hairpins by either probe is always a possibility and would result in underestimation of the true $T_m$.

The apparent melting temperatures listed in Table 2 provide a ranking of base pairing strength for the G/C analogs in the absence of nA and sT. Among the G analogs, the sG–C couple with its inherent steric clash decreased the $T_m$ by 30°C. The H–C base pair, with only two hydrogen bonds, lowered the $T_m$ by 21°C. Relative to this base pair the cH–C couple was more destabilizing while the IcH–C and NitrocH–C couples were less destabilizing. Only two C analogs were successfully incorporated into the hairpin. Based on the melting analysis, the NitroC–cG base pair reduced the $T_m$ by 19°C while the sC–cG couple, with its associated steric clash, reduced it by 41°C.

With the exception of NitrocH and NitroC all of the other testable analogs [IcH, H, cH, sG and sC] eliminated secondary structure in the HP25 hairpin when used in conjunction with nA and sT. Hybridization of the 25-mer DNA probe to these hairpins was complete at 10°C. By comparison, HP25 substituted with nA + sT or with nA + sT + cG had $T_m$'s of 40°C and 26°C, respectively. Hairpin substituted with NitrocH + nA + sT had a $T_m$ of 18°C. The NitroC analog was unimpressive as a candidate for destabilizing DNA since it did not contribute to decreased hybrid stability when used in conjunction with nA and sT. Hairpins substituted with nA + sT + NitroC or nA + sT + NitroC + cG had $T_m$'s of 48°C and 44°C, respectively.

The results obtained with hairpins were extended by determining apparent melting temperatures of short Watson–Crick hybrids in which one strand was substituted with an analog of G or C. Each modified strand was prepared by primer extension of a radiolabeled primer and then purified by electrophoresis in a 7 M urea–12% polyacrylamide gel. The modified strands

were hybridized to a complementary 12-mer DNA or 8-mer LNA probe to form the hybrids depicted in Figure 1B. Apparent melting temperatures were determined using a variation of the gel mobility shift assay described earlier. The $T_m$'s listed in Table 3 confirm the relative base pairing strengths of the C analogs and further demonstrate that MefP paired much more weakly to G than did sC or NitroC. The hypoxanthine bases were nearly equivalent to one another in destabilizing the 12 bp DNA hybrid. However, this was not the case for the 8-bp LNA–DNA hybrid where NitrocH exhibited much greater pairing strength than H, cH and IcH. Interestingly, NitrocG also exhibited enhanced pairing strength when one strand of the duplex had an LNA backbone. sG, which was only tested in the LNA–DNA hybrid, was the most destabilizing of the G analogs.

Hybridization of oligonucleotide probes to DNA substituted with sG or sC will probably require probes that are themselves modified to eliminate steric clash with the thioketo group of these base analogs. For example, sG–MeP and H–sC lack the steric clash of sG–C and G–sC and might be more stable. Indirect evidence suggested that this was the case for sG–MeP, which was found to be isostable with A–T (46). We evaluated base pair stability by determining $T_m$'s of a 12 bp DNA hybrid in which two G–C base pairs were replaced by sG–C, sG–MeP, G–MeP, G–sC, H–sC or H–C couples. The results in Table 4 show that sG–MeP and G–MeP, which had similar pairing strengths, were much less stable than sG–C. This suggests that hydrogen bonding between the thioketo group of sG and the amino group of C stabilized the sG–C base pair even in the presence of steric clash. Substitution of MeP for C or G for sG removed steric clash but also eliminated a hydrogen bond thus leading to further destabilization. In contrast, virtually no difference in stability was observed

between the G–sC, H–sC and H-C base pairs. Here steric clash may have prevented hydrogen bonding between the thioketo group of C and the amino group of G. If so substitution of H for G or C for sC would eliminate clash without any loss of base pairing strength. The nA–sT couple provides yet a third example of steric clash. In this case the analogous clash-free couples of A–sT and A–T were much more stable than nA–sT even though they only possessed two hydrogen bonds (10). It is likely that altered electronic and stacking properties of the respective base pairs also have an important role in determining their relative stabilities. The three examples demonstrate that

**Table 4.** Apparent melting temperatures of 12 bp DNA hybrids substituted with weakly pairing analogs of G and C

| X | Y | $T_m$ (°C) | $\Delta T_m$ |
|---|---|---|---|
| sG and MeP[a] | | | |
| G | C | 52 | – |
| sG | C | 41 | −11 |
| sG | MeP | 34 | −18 |
| G | MeP | 32 | −20 |
| H and sC[b] | | | |
| G | c | 54 | – |
| G | sC | 45 | −9 |
| H | sC | 44 | −10 |
| H | C | 45 | −9 |

[a]5′–**X**CA**X**ACTTCTCC
 3′–**Y**GT**Y**TGAAGAGG
[b]5′–**X**CA**X**ACTTCTCC
3′–T**Y**GT**Y**TGAAGAGGTACAACTGCACAG

elimination of clash does not necessarily lead to a more stable base pair.

### Specificity of base pairing by G/C analogs

The short single-stranded primer extension products described previously were used to evaluate whether otherwise promising G/C analogs discriminated against incorrect bases. For this analysis the modified DNAs were incubated under stringent conditions with four 12-mer DNA probes each of which presented a different base to the analog of interest. Reactions were quenched by adding an unlabeled competitor oligonucleotide to sequester free probe followed by rapid cooling in an ice bath. Hybrid formation was detected by gel mobility shift and the ratio of mismatched to matched hybrid was used as a proxy for specificity. Using this methodology we previously showed that nA and sT are comparable to A and T in their discrimination of mismatched bases (11). In the current study specific pairing was observed for NitroC and sC (Figure 3A) but not for any of the G analogs (data not shown). The low yield of perfect-match hybrid in the sC panel may be attributable to strand displacement of radiolabeled target from the duplex by unmodified competitor oligonucleotide.

Pairing properties of the hypoxanthine analogs were further investigated using probes that hybridized under more stringent conditions. The naturally high binding affinity of LNA and its preferential stabilization of the NitrocH-C base pair were sufficient to impart specificity to this analog when using 8-mer LNA probes (Figure 3B). By comparison, some of the same LNA probes formed
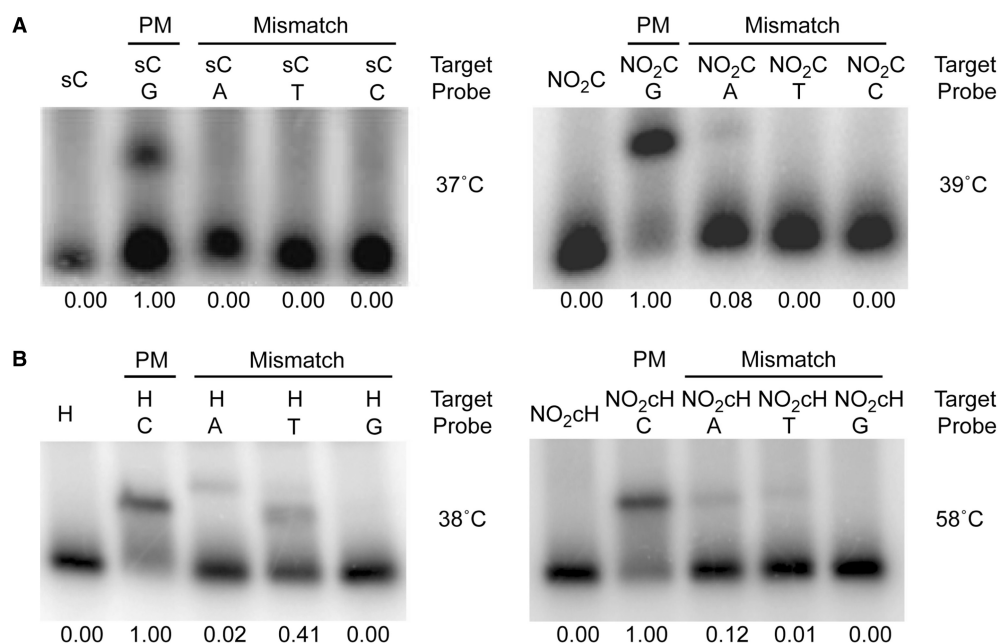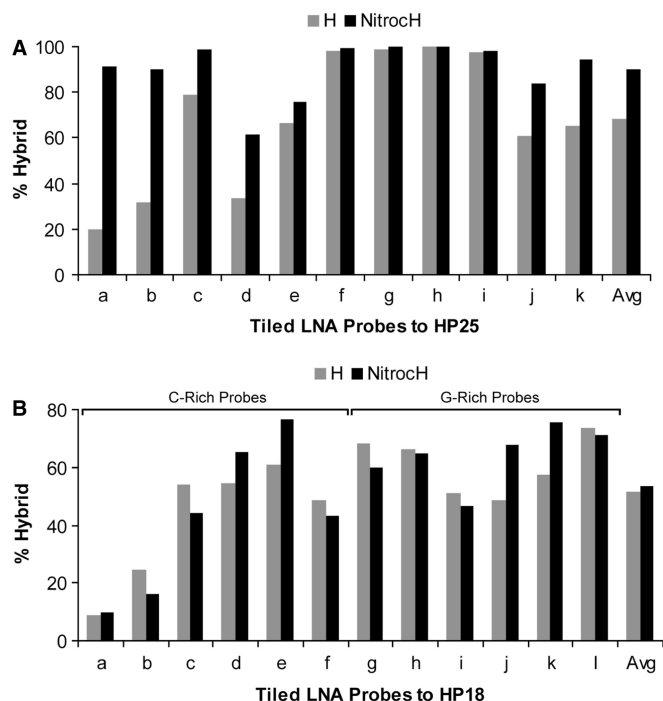


**Figure 3.** Specificity of base pairing by selected base analogs. Analogs of (**A**) cytosine and (**B**) guanine were evaluated using probe-target hybrids 2 and 3, respectively (Figure 1B). Radiolabeled targets substituted with the indicated analog of G or C were prepared by primer extension and hybridized under stringent conditions to probes that presented each of the regular bases to a centrally positioned C or G analog. Reactions were quenched by sequestering free probe with unlabeled competitor oligonucleotide after which probe-target hybrid was resolved by gel mobility shift assay. DNA targets substituted with NitrocH or H also contained nA and sT while DNA substituted with NitroC or sC did not. The hybridization temperature is indicated next to each reaction series; apparent $T_m$'s of the corresponding perfect-match hybrids were 45°C (sC), 47°C (NitroC), 43°C (H) and 63°C (NitrocH).

mismatched hybrids when the DNA target contained H, cH or IcH. Attempts to enhance the specificity of NitrocH by utilizing 12-mer DNA probes substituted with stabilizing analogs of C were unsuccessful. For example, even though the tricyclic aminoethyl-phenoxazine analog of C [(47); the so-called 'G-clamp'] formed a very stable base pair with NitrocH and increased hybrid stability by 12°C, mispairing of the standard bases to NitrocH was still observed under stringent conditions.

### Hybridization of tiled probes to modified hairpins

The HP25 hairpin provided an opportunity to test whether highly structured DNA substituted with nA, sT and H or NitrocH could hybridize to 8-mer LNA probes under non-denaturing conditions. The probes were tiled in two base increments across the 3′ arm of the hairpin stem. Each probe-hairpin combination was incubated for 5–10 min at 30°C in the presence of 5 μM probe and limiting hairpin target. Gel mobility shift analysis of the reactions showed that unmodified hairpin hybridized with an average efficiency of only 3.5% to the probe set. Introduction of nA and sT into the hairpin increased hybridization efficiency to 34%, underscoring the need to destabilize G-C pairing as well. NitrocH was emphasized in this study since it had the best uptake and pairing properties of the analogs studied. For comparison, we also prepared hairpins that were substituted with H. In the presence of nA and sT, both G analogs promoted robust hybridization with the LNA probes (Figure 4A). Identical results were obtained using a parallel set of tiled 12-mer DNA

probes (data not shown). The ready accessibility of the hairpin sequences indicates that the modified bases had eliminated secondary structure.

The ability of NitrocH to destabilize a more highly G/C rich duplex was tested by hybridizing a tiled set of 8-mer LNA probes to both arms of an 18 bp hairpin (HP18; Figure 1B) that was also substituted with nA and sT. With a G/C content of 78% and an estimated $T_m$ of 96°C the unmodified HP18 control hairpin hybridized poorly to the probes. When the same set of probes was hybridized to hairpin substituted with nA, sT and NitrocH at 30°C, the average hybridization level was only 33%, suggesting the presence of residual secondary structure in the hairpin target. This hypothesis was confirmed by demonstrating that the extent of hybridization increased to 50% when the temperature was raised to 45°C (Figure 4B). At this temperature NitrocH and H were nearly interchangeable. Interestingly, G-rich probes/hybridized to a greater extent than C-rich probes, possibly reflecting a decrease in hybrid stability that was proportional to the number of NitrocH–C or H–C base pairs in the duplex. Significant probe-to-probe variation was also observed, with one C-rich probe giving <10% hybrid. A more uniform hybridization profile was obtained by using a parallel set of tiled 12-mer DNA probes that contained LNA C residues (data not shown).

## DISCUSSION

We have previously shown that pseudo-complementary DNA can be generated from regular DNA by carrying out primer extension with the dNTPs of nA, sT, H and cytosine (11). DNA composed of these bases can be synthesized by DNA polymerases and is nearly structure-free since nA–sT is a mismatch and H–C is a very weak base pair. Hybridization of this DNA to unmodified probes is driven by the formation of nA–T, A–sT, G–C and H–C base pairs. Unfortunately, H exhibits degenerate pairing and cannot be used in hybridization assays where specificity is required. In this study we have evaluated the uptake and base pairing properties of several highly destabilizing analogs of G and C that might function similarly to H but with greater specificity.

Several of the base analogs evaluated in this study have anti-viral or anti-cancer activity. When incorporated into DNA, P is potent inhibitor of cytosine methyltransferase (14). Our results suggest that this is probably a rare event since P/MeP did not form a stable base pair with G and was a poor substrate for primer extension. The anti-viral activity of MefP type bases does not depend upon their incorporation into DNA (15). This is consistent with our observations that MefP paired weakly to G and was a marginal substrate for primer extension. The anti-cancer agent dsGTP was readily incorporated into DNA as long as the template lacked secondary structure. sG paired weakly to C and mispaired to other bases. It has been shown that mis-incorporation of bases opposite sG contributes to the cytotoxicity of this agent (48). NitroC exhibits anti-viral activity by binding tightly to viral polymerases (16). In our study this base was less destabilizing



**Figure 4.** Hybridization of tiled 8-mer LNA probes to (**A**) the HP25 hairpin at 30°C and (**B**) the H18 hairpin at 45°C. Each hairpin was substituted with nA, sT and NitrocH or H. Probes to HP25 were tiled across the 3′ arm of the hairpin stem while probes to HP18 were tiled across both arms of the hairpin stem. Gel mobility shift assay was used to determine the fraction of hairpin that was hybridized to each probe.

than the other analogs and was readily incorporated into DNA where it exhibited good specificity of pairing.

In contrast to these base analogs, NitrocH proved to be a promising candidate for use as a pseudo-complementary base. As expected, the electron-withdrawing nitro group strengthened base pairing such that NitrocH–C was more stable than H–C. Consequently, the dNTP of NitrocH was less prone than H to inhibit primer extension through G/C-rich secondary structure features. When incorporated into DNA along with nA and sT, NitrocH was nearly as effective as H in eliminating secondary structure. Furthermore, long primer extension products substituted with nA, sT and NitrocH could be purified from template DNA by electrophoresis through a formamide agarose gel (data not shown). Importantly, NitrocH was less degenerate in its base pairing than H and exhibited good discrimination in the presence of LNA probes. A contributing factor to the performance of LNA probes was the unusually high stability of NitrocH-C base pairs in LNA–DNA hybrids. This phenomenon was also observed with 7-nitro-7-deazaguanine, which exhibited much greater base pairing strength than G in those hybrids. Since LNA–DNA duplexes are less hydrated than DNA duplexes (49), the polar nitro group may preferentially stabilize the mixed hybrid by attracting water molecules to the major groove. Of course, LNA–DNA hybrids differ in many respects from dsDNA and other factors such as base stacking, hydrogen bonding and steric constraints could account for the different responses of the two hybrids to 7-nitro-7-deazapurine bases.

Since unmodified DNA probes were unable to discriminate between matched and mismatched NitrocH bases, probes that contain G clamp or LNA C residues were also evaluated. Although both modifications exhibited a high affinity for NitrocH, indiscriminate pairing was still observed. A more specific version of NitrocH is clearly needed but how to design such an analog is not obvious.

Among the analogs of C that we evaluated, sC has potential for use as a pseudo-complementary base. It forms a very weak pair with G due to steric clash yet does not accommodate mismatching to A, T or C. Efficient incorporation of dsCTP into DNA is made possible by the use of 'adaptor' templates in which G has been replaced by H. DNA substituted with nA, sT and sC is structure-free but does not form stable hybrids with short unmodified probes. The challenge is to find an analog of G that specifically pairs to sC by forming three hydrogen bonds without steric clash. Such an analog could be substituted for G in probes. While inosine can be utilized as a template base in primer extension reactions, it is too indiscriminate to be used in probes. We are actively pursuing the design and evaluation of novel heterocycles capable of pairing to sC and in parallel are characterizing the fidelity of converting regular DNA into pseudo-complementary DNA by using dNTPs of nA, sT and sC in asymmetric PCR.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## REFERENCES

1. Hacia,J.G., Woski,S.A., Fidanza,J., Edgemon,K., Hunt,N., McGall,G., Fodor,S.P. and Collins,F.S. (1998) Enhanced high density oligonucleotide array-based sequence analysis using modified nucleoside triphosphates. *Nucleic Acids Res.*, **26**, 4975–4982.
2. Gamper,H.B., Cimino,G.D. and Hearst,J.E. (1987) Solution hybridization of crosslinkable DNA oligonucleotides to bacteriophage M13 DNA. Effect of secondary structure on hybridization kinetics and equilibria. *J. Mol. Biol.*, **197**, 349–362.
3. Ratushna,V.G., Weller,J.W. and Gibas,C.J. (2005) Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. *BMC Genomics*, **6**, 31.
4. Kierzek,E., Kierzek,R., Turner,D.H. and Catrina,I.E. (2006) Facilitating RNA structure prediction with microarrays. *Biochemistry*, **45**, 581–593.
5. Nguyen,H.K. and Southern,E.M. (2000) Minimising the secondary structure of DNA targets by incorporation of a modified deoxynucleoside: implications for nucleic acid analysis by hybridisation. *Nucleic Acids Res.*, **28**, 3904–3909.
6. Kushon,S.A., Jordan,J.P., Seifert,J.L., Nielsen,H., Nielsen,P.E. and Armitage,B.A. (2001) Effect of secondary structure on the thermodynamics and kinetics of PNA hybridization to DNA hairpins. *J. Am. Chem. Soc.*, **123**, 10805–10813.
7. Armitage,B.A. (2003) The impact of nucleic acid secondary structure on PNA hybridization. *Drug Discov. Today*, **8**, 222–228.
8. Ramdas,L., Cogdell,D.E., Jia,J.Y., Taylor,E.E., Dunmire,VR., Hu,L., Hamilton,S.R. and Zhang,W. (2004) Improving signal intensities for genes with low-expression on oligonucleotide microarrays. *BMC Genomics*, **5**, 35.
9. Chou,C.C., Chen,C.H., Lee,T.T. and Peck,K. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res.*, **32**, e99.
10. Kutyavin,I.V., Rhinehart,R.L., Lukhtanov,E.A., Gorn,V.V., Meyer,R.B. Jr. and Gamper,H.B. Jr. (1996) Oligonucleotides containing 2-aminoadenine and 2-thiothymine act as selectively binding complementary agents. *Biochemistry*, **35**, 11170–11176.
11. Gamper,H.B. Jr., Arar,K., Gewirtz,A. and Hou,Y.M. (2006) Unrestricted hybridization of oligonucleotides to structure-free DNA. *Biochemistry*, **45**, 6978–6986.
12. Lahoud,G., Timoshchuk,V., Lebedev,A., de Vega,M., Salas,M., Arar,K., Hou,Y.M. and Gamper,H. (2008) Enzymatic synthesis of structure-free DNA with pseudo-complementary properties. *Nucleic Acids Res.*, **36**, 3409–3419.
13. Pui,C.H. and Jeha,S. (2007) New therapeutic strategies for the treatment of acute lymphoblastic leukaemia. *Nat. Rev. Drug Discov.*, **6**, 149–165.
14. Marquez,V.E., Kelley,J.A., Agbaria,R., Ben-Kasus,T., Cheng,J.C., Yoo,C.B. and Jones,P.A. (2005) Zebularine: a unique molecule for an epigenetically based strategy in cancer chemotherapy. *Ann. N Y Acad. Sci.*, **1058**, 246–254.

15. Robins,M.J., Miranda,K., Rajwanshi,V.K., Peterson,M.A., Andrei,G., Snoeck,R., De Clercq,E. and Balzarini,J. (2006) Synthesis and biological evaluation of 6-(alkyn-1-yl)furo[2,3-d]pyr-imidin-2(3H)-one base and nucleoside derivatives. *J. Med. Chem.*, **49**, 391–398.

16. Harki,D.A., Graci,J.D., Galarraga,J.E., Chain,W.J., Cameron,C.E. and Peterson,B.R. (2006) Synthesis and antiviral activity of 5-sub-stituted cytidine analogues: identification of a potent inhibitor of viral RNA-dependent RNA polymerases. *J. Med. Chem.*, **49**, 6166–6169.

17. Guo,M.J., Hildbrand,S., Leumann,C.J., McLaughlin,L.W. and Waring,M.J. (1998) Inhibition of DNA polymerase reactions by pyrimidine nucleotide analogues lacking the 2-keto group. *Nucleic Acids Res.*, **26**, 1863–1869.

18. Gamper,H.B. Jr., Gewirtz,A., Edwards,J. and Hou,Y.M. (2004) Modified Bases in RNA Reduce Secondary Structure and Enhance Hybridization. *Biochemistry*, **43**, 10224–10236.

19. Seela,F. and Driller,H. (1985) Solid-phase synthesis of the self-complementary hexamer d(c7GpCpc7GpCpc7GpC) via the O-3′-phosphoramidite of 7-deaza-2′-deoxyguanosine. *Nucleic Acids Res.*, **13**, 911–926.

20. Mizusawa,S., Nishimura,S. and Seela,F. (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.*, **14**, 1319–1324.

21. McDougall,M.G., Sun,L., Livshin,I., Hosta,L.P., McArdle,B.F., Samols,S.B., Fuller,C.W. and Kumar,S. (2001) Analogs of guanine nucleoside triphosphates for sequencing applications. *Nucleosides Nucleotides Nucleic Acids*, **20**, 501–506.

22. Watkins,N.E. Jr. and SantaLucia,J. Jr. (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.*, **33**, 6258–6267.

23. Martin,F.H., Castro,M.M., Aboul-ela,F. and Tinoco,I. Jr. (1985) Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.*, **13**, 8927–8938.

24. Kim,C.-H., Marquez,V.E., Mao,D.T., Haines,D.R. and McCormack,J.J. (1986) Synthesis of pyrimidin-2-one nucleosides as acid-stable inhibitors of cytidine deaminase. *J. Med. Chem.*, **29**, 1374–1380.

25. Robins,M. and Barr,P. (1981) Nucleic acid related compounds. 31. Smooth and efficient palladium-copper catalyzed coupling of term-inal alkynes with 5-iodouracil nucleosides. *Tetrahedron Lett.*, **22**, 421–424.

26. Fox,J. and Van Praag,D. (1961) Pyrimidine nucleosides. VIII. Synthesis of 5-nitrocytidine and related nucleosides. *J. Org. Chem.*, **26**, 526–532.

27. Marathias,V.M., Sawicki,M.J. and Bolton,P.H. (1999) 6-Thioguanine alters the structure and stability of duplex DNA and inhibits quadruplex DNA formation. *Nucleic Acids Res.*, **27**, 2860–2867.

28. Bohon,J. and de los Santos,C.R. (2005) Effect of 6-thioguanine on the stability of duplex DNA. *Nucleic Acids Res.*, **33**, 2880–2886.

29. Bohon,J. and de los Santos,C.R. (2003) Structural effect of the anticancer agent 6-thioguanine on duplex DNA. *Nucleic Acids Res.*, **31**, 1331–1338.

30. Somerville,L., Krynetski,E.Y., Krynetskaia,N.F., Beger,R.D., Zhang,W., Marhefka,C.A., Evans,W.E. and Kriwacki,R.W. (2003) Structure and dynamics of thioguanine-modified duplex DNA. *J. Biol. Chem.*, **278**, 1005–1011.

31. Rogers,G. and Ulbricht,T. (1970) Transfer RNA components. I. Direct synthesis of nucleosides of cytosine and 2-thiocytosine. *J. Chem. Soc.*, **8**, 1109–1113.

32. Ling,Y.H., Chan,J.Y., Beattie,K.L. and Nelson,J.A. (1992) Consequences of 6-thioguanine incorporation into DNA on poly-merase, ligase, and endonuclease reactions. *Mol. Pharmacol.*, **42**, 802–807.

33. Yoshida,S., Yamada,M., Masaki,S. and Saneyoshi,M. (1979) Utilization of 2′-deoxy-6-thioguanosine 5′-triphosphate in DNA synthesis in vitro by DNA polymerase alpha from calf thymus. *Cancer Res.*, **39**, 3955–3958.

34. Maybaum,J., Bainnson,A.N., Roethel,W.M., Ajmera,S., Iwaniec,L.M., TerBush,D.R. and Kroll,J.J. (1987) Effects of incor-poration of 6-thioguanine into SV40 DNA. *Mol. Pharmacol.*, **32**, 606–614.

35. Ling,Y.H., Nelson,J.A., Cheng,Y.C., Anderson,R.S. and Beattie,K.L. (1991) 2′-Deoxy-6-thioguanosine 5′-triphosphate as a substrate for purified human DNA polymerases and calf thymus terminal deoxynucleotidyltransferase in vitro. *Mol. Pharmacol.*, **40**, 508–514.

36. Rappaport,H.P. (1993) Replication of the base pair 6-thioguanine/ 5-methyl-2-pyrimidine with the large Klenow fragment of Escherichia coli DNA polymerase I. *Biochemistry*, **32**, 3047–3057.

37. Tabor,S. and Richardson,C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl Acad. Sci. USA*, **84**, 4767–4771.

38. Kamtekar,S., Berman,A.J., Wang,J., Lazaro,J.M., de Vega,M., Blanco,L., Salas,M. and Steitz,T.A. (2004) Insights into strand displacement and processivity from the crystal structure of the protein-primed DNA polymerase of bacteriophage phi29. *Mol. Cell*, **16**, 609–618.

39. de Vega,M., Lazaro,J.M., Salas,M. and Blanco,L. (1996) Primer-terminus stabilization at the 3′-5′ exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. *EMBO J.*, **15**, 1182–1192.

40. Seela,F. and Driller,H. (1989) Alternating d(G-C)3 and d(C-G)3 hexanucleotides containing 7-deaza-2′-deoxyguanosine or 8-aza-7-deaza-2′-deoxyguanosine in place of dG. *Nucleic Acids Res.*, **17**, 901–910.

41. Seela,F. and Roling,A. (1992) 7-Deazapurine containing DNA: efficiency of c7GdTP, c7AdTP and c7IdTP incorporation during PCR-amplification and protection from endodeoxyribonuclease hydrolysis. *Nucleic Acids Res.*, **20**, 55–61.

42. Seela,F., Tran-Thi,Q.H. and Franzen,D. (1982) Poly(7-deazaguanylic acid), the homopolynucleotide of the parent nucleoside of queuosine. *Biochemistry*, **21**, 4338–4343.

43. Faerber,P. (1974) The synthesis of the alternating copolymer poly (r(G-s2C) by RNA polymerase of E. coli. A polynucleotide with an extremely high thermal stability. *FEBS Lett.*, **44**, 111–116.

44. Mills,D.R. and Kramer,F.R. (1979) Structure-independent nucleo-tide sequence analysis. *Proc. Natl Acad. Sci. USA*, **76**, 2232–2235.

45. Jepsen,J.S., Sorensen,M.D. and Wengel,J. (2004) Locked nucleic Acid: a potent nucleic Acid analog in therapeutics and biotechnol-ogy. *Oligonucleotides*, **14**, 130–146.

46. Rappaport,H.P. (1988) The 6-thioguanine/5-methyl-2-pyrimidinone base pair. *Nucleic Acids Res.*, **16**, 7253–7267.

47. Linn,K.-Y. and Matteucci,M.D. (1998) A cytosine analog capable of clamp-like binding to a guanine in helical nucleic acids. *J. Am. Chem. Soc.*, **120**, 8531–8532.

48. Yuan,B. and Wang,Y. (2008) Mutagenic and cytotoxic properties of 6-thioguanine, S6-methylthioguanine, and guanine-S6-sulfonic acid. *J. Biol. Chem.*, **283**, 23665–23670.

49. Pande,V. and Nilsson,L. (2008) Insights into structure, dynamics and hydration of locked nucleic acid, (LNA) strand-based duplexes from molecular dynamics simulations. *Nucleic Acids Res.*, **36**, 1508–1516.